Chip Scale Review®

ChipScaleReview.com

The Future of Semiconductor Packaging

Volume 25, Number 1

January • February 2021

Emerging process and assembly challenges in electronics manufacturing

page 12

- Wafer-scale S-MCM
- 600mm WLFO on PLP
- Revealing invisible defects on large 600mm panels
- Going beyond traditional temporary bonding materials
- Heterogeneous integration prompts test content to “shift left”
- High-throughput flexible direct imaging for packaging/MEMS fabrication
- Extreme Si thinning and nano-TSVs to advance 3D heterogeneous integration

PALOMAR®
TECHNOLOGIES

Making the connected world possible™

Palomar Die Bonders

High accuracy, high flexibility eutectic and epoxy die bonding specializing in precision assembly of complex packages

Innovation Centers

Specializing in contract manufacturing and process development to deliver rapid ROI for new product introductions

Palomar Wire & Wedge Bonders

High-speed, robust chip-to-die interconnections, ball-and-stitch, ball bumping, wedge and ribbon bonding

SST Vacuum Reflow Systems

Unique combination of vacuum, pressure and heat to create highly reliable, void-free solder

Advanced Photonic & Microelectronic Packaging Solutions delivering improved yield, reduced assembly times and rapid ROI

www.PalomarTechnologies.com

CONTENTS

January • February 2021
Volume 25, Number 1

Cover image represents bare die on a wafer. These die were used in the assembly of a MEMS package, which was integrated into a pressure sensor. MEMS are responsible for the sensing element of heterogeneous integration and are one of many application types that will advance tremendously by leveraging intricate multi-die designs to deliver greater performance in a more compact footprint.

Photo courtesy of Universal Instruments

DEPARTMENTS

EVG®

invent
innovate
implement

www.EVGroup.com

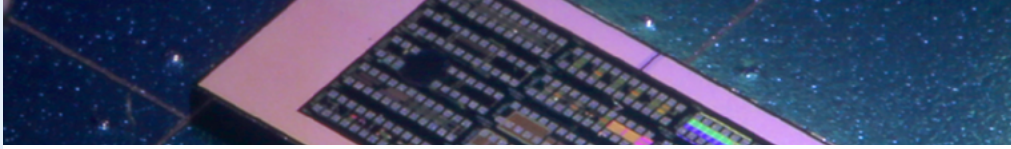
DIE-TO-WAFER (D2W) BONDING SOLUTIONS

- Fusion and hybrid bonding for next-generation heterogeneous integration
- Collective D2W bonding enabled by extensive knowledge in carrier preparation and die handling
- Direct placement D2W activation and cleaning complete solution with EVG®320 D2W
- Production-ready equipment solutions for successful integration of chiplets
- Heterogeneous Integration Competence Center™ serving as leading-edge incubation center for customers and partners

EVG® 320 D2W

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

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



Volume 25, Number 1
January • February 2021

Subscriptions in the U.S. are available without charge to qualified individuals in the electronics industry.
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 Gilroy, Calif., and additional offices.

POSTMASTER: Send address changes to Chip Scale Review magazine
P.O. Box 2165
Morgan Hill, CA 95038
Tel: +1-408-846-8580
E-Mail: subs@chipscalereview.com

Printed in the United States

FEATURE ARTICLES

The wonders of quantum computing

By Debra Vogler, Sr. Technical Editor

When I first laid eyes on the IBM Q System One computer (an "integrated universal approximate quantum computing system" according to IBM's news release) at SEMICON West 2019, I was stunned. I had seen photos of it beforehand, but in person—it was a "wow" moment. I felt the same way I did as a kid when President Kennedy challenged the nation to land a man on the moon—and then watched it happen less than 10 years later. When quantum computing (QC) was discussed at IWLPC 2019, I felt challenged as senior technical editor of this magazine to solicit content from industry experts that would tackle the impact of QC and applications such as artificial intelligence (AI) on the future of packaging and test. (While AI also uses classical computing, it will benefit from QC capabilities as that technology is further developed.)

Industry experts heeded *Chip Scale Review*'s call for QC/AI/next-generation computing content, and as a result, we were able to publish several relevant articles in 2020 that included:

1) "Post-Moore's Law electronics: now, until quantum electronics," (R. Tummala/Georgia Tech, Mar/Apr);

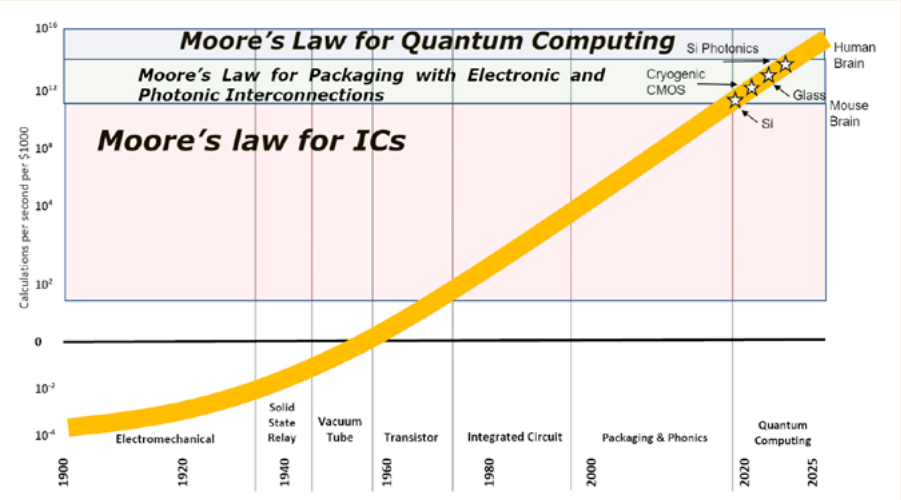
2) AI's impact on 3D packaging: heterogeneous integration," (S. Kumar/Yole Développement, Korea, May/June);

3) "Electronic packaging for future electronic systems," (M. Töpper, T. Braun, R. Aschenbrenner/Fraunhofer IZM, Jul/Aug);

4) "Enabling AI with heterogeneous integration," (N. Fan/ASM Pacific Technology Ltd., Sept/Oct);

5) "A deep-learning solution for heterogeneous package inspection," (S. Chitchian/INTEKPLUS Corporation Ltd., Sept/Oct);

6) "Enabling AI with heterogeneous integration," (A. Kumar, M. Farooq/IBM Research, Nov/Dec).



Computing performance driven by Moore's Law and post-Moore's Law technologies. SOURCE: "Post-Moore's Law electronics: now, until quantum electronics," (R. Tummala, *Chip Scale Review*, Mar/Apr 2020)

In 2021, *CSR* will be asking industry experts to, once again, put forth the vision that will drive the packaging and test sectors. Heterogeneous integration will continue to play a major role and, as semiconductor processes continue to show signs of blurring the boundaries between front-end-of-line, middle-end-of-line, and back-end-of-line, there will undoubtedly be more blurring of the lines among our standard coverage topics. QC will continue to be a topic of significant interest. In our first issue of the new year, members of the Quantum Information and Integrated Nanosystems Group at MIT Lincoln Laboratory (R. Das, V. Bolkhovsky, A. Wynn, R. Rastogi, S. Zarr, L. Johnson) present their strategy for using laser direct write (LDW) and optical lithography to fabricate 200mm wafer-scale superconducting multi-chip modules (S-MCM) for interconnecting multiple active superconducting flux quantum (SFQ) chips for next-generation cryogenic processing systems. The authors note that, "the demand for superconducting computing scalability beyond arrays of a few superconducting chips is driving the need for greater wiring densities and more functionality onto a single cryogenic package."

As the industry continues implementation of QC and AI, I look forward to receiving abstracts from our long-time authors, as well as those who have not written for *CSR* before. Our readers need articles that shed light on the challenges with respect to packaging qubits, such as: reducing electromigration (EM) loss, the impact of dielectric loss on how to scale wiring layers, the impact of a low thermal budget for Josephson junctions (JJs), and developing superconducting through-silicon vias (TSVs), among others (ref.: "Fabricating Quantum Technologies," G. Ribeill/Raytheon Technologies, SEMICON West 2020).

Here's looking to innovation (and to better times) in 2021.

INTEK PLUS
Integrated Measurement System

KOSDAQ LISTED COMPANY

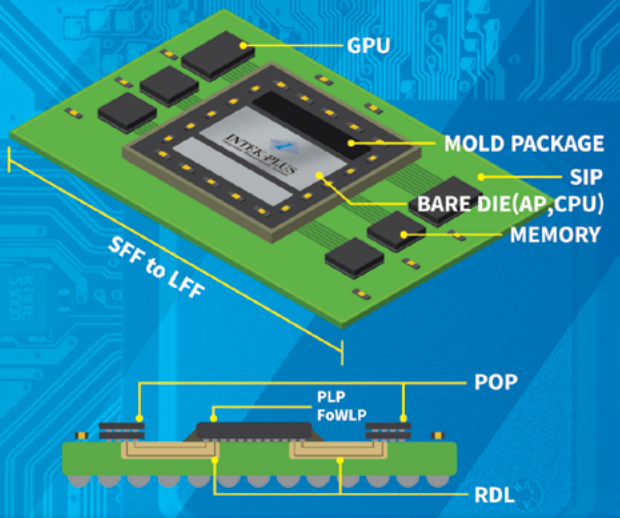
Semiconductor Package Inspection

INTEK PLUS
Integrated Measurement System

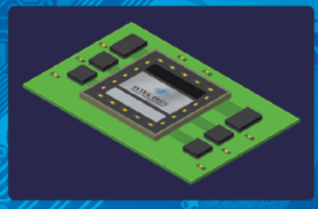
Fast, Accurate, and the most Advanced

INTEKPLUS specializes in 2D and 3D Surface Inspection and Metrology for Semiconductor Package Products. We provide the World's fastest Inspection Capability to achieve High Productivity.

High-end Package Trends



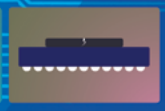
Supreme Vision Solution



Heterogeneous Integration

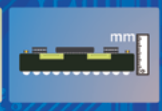
AI Deep Learning

Shadow Free



Micro Crack (Die Mold)

SFF to LFF Size Coverage



Total Height

Out-Tray Inspection

iPIS-140 / 240 / 240+ / 340
* Hexa Model for 4 Sides

In-Tray Inspection

iPIS-380 / 560 / 580
* Hexa Model for 4 Sides

Inspection + Tape & Reel + ARC

iPIS-380TR / 560TR / 580TR
* TR Model for Tape & Reel
* ARC Ext. for Full Automation

Looking for Sales Representative in Europe, America, Southeast Asia(Singapore, Malaysia, Philippines, Thailand, Vietnam)

INTEKPLUS CO., LTD.
#263, Techno 2-ro, Yuseong-Gu, Daejeon, 34026 Korea
Tel : +82-42-930-9900 Fax : +82-42-930-9999

For more information
www.intekplus.com / sales1@intekplus.com

smiths interconnect

Beyond Connectivity

Smiths Interconnect test sockets and probe cards offer superior quality and reliability in semiconductor test applications.

Leading-edge engineering, development and simulation expertise supporting automated, system-level and development test platforms for area array, wafer level, peripheral and PoP devices.

Area Array Test

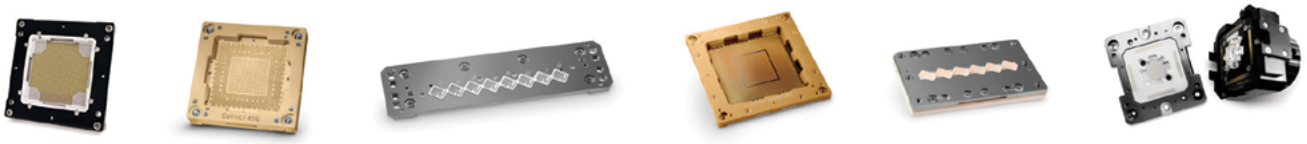
- In-house maintenance and contact replacement
- Entirely shielded signal path*
- Controlled impedance*
- High speed test to 67GHz Analog & 56Gbps Digital*
- System simulation modeling

*DaVinci Series

Wafer Level Test

- Reliable RF signal integrity
- Excellent compliance and contact force
- Simple maintenance
- Quick installation

Best In Class



more > smithsinterconnect.com

Heterogeneous integration prompts test content to "shift left"

By Dave Armstrong [Advantest America, Inc.]

Heterogeneous integration and the resulting need for known-good die (KGD) are driving the transition to a new test flow, best described as "shift left." With this flow, test functions once performed at system-level test are moving to final test, or a KGD test step that occurs after thin-bump sawing. Similarly, final-test functions are shifting left to the KGD step or to the wafer-probing step (**Figure 1**).

The need for this shift-left flow arises because many companies today are starting to ship KGD—including some very complex artificial intelligence (AI) devices— and others are shipping memory stacks. In these examples, there is no longer a packaged-device ship location. Shipping parts in die form represents a new ship location, and you have a critical need for more test content prior to that ship location to make sure the devices are of sufficient quality for subsequent heterogeneous integration. Indeed, you may not have a profitable product if you do not push your quality upstream. Pushing quality upstream gets you profit downstream.

Consider the traditional test flow. At the wafer-test step, automatic test equipment and a wafer prober find hard rejects and perform scan and some functional test at one temperature. At the KGD test step, automatic test equipment (ATE) and a singulated die prober confirm scan and functional tests at a second temperature. At final test, ATE and a device/die handler with active thermal control perform packaged device test, extended scan test, parametric performance test, at-speed test, high-power test, and stress tests. And finally, a system-level tester repeats packaged device tests and performs boot-up tests and fuse blowing.

The flow described above may have been sufficient five or six years ago when multi-die integrations might have had one large logic part and four high-bandwidth memory (HBM) devices. But today, people are moving to heterogeneous devices that have 30, 40, or 50 devices on them, and the error elements of each one of those die multiply together. If you have 0.99% good devices and you multiply all the error elements together, you could very easily end up with an assembly that has only a 60% probability of being good. So the fundamental challenge is that you need to get not just 0.99% good parts, but 0.9999% good parts. This is where people suggest that we are moving past the parts per million into an industry where parts per billion is the new norm.

As a result of the new quality requirements noted above, you have to do a lot more testing at wafer, or perhaps after thin bump sawing, where you can actually perform active-thermal-control thermal testing, where you can do full-power testing of your part, and where you can do at-speed testing of your part. People are finding that by being able to do at-speed, at-power test of KGD, they can do pretty much all the testing that they used to do at final test. In turn, final test is becoming more of a system-level test step—performing boot-up testing, for example, and checking whether your software and firmware are working.

A big challenge for heterogeneous integration is simple continuity. You might have 40,000 bumps on your logic part, and you may have ten of those logic parts, so you easily could have a half a million bumps that have to make good contact. And then you have other parts with other I/O, so you have a very large amount of continuity checks and interface tests to perform. And that's where some of the IEEE standards come into play, because you can't necessarily

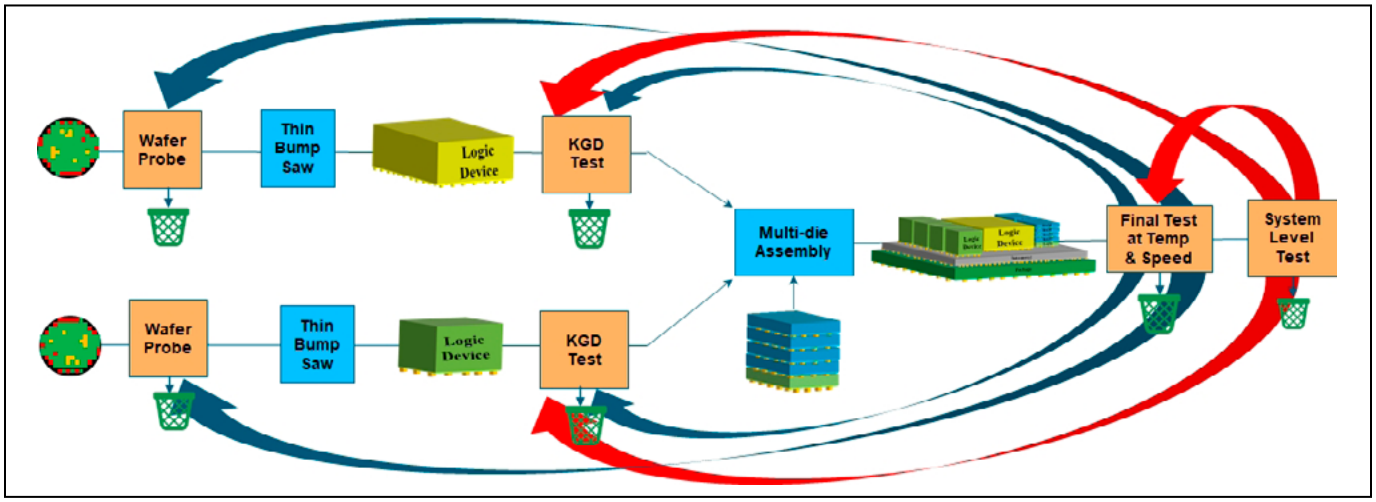


Figure 1: Test content is shifting left, from system-level test to final test and KGD test—and from final test to KGD test and wafer-level test.

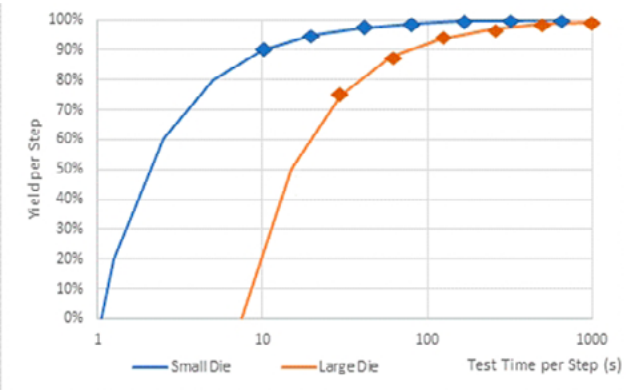


Figure 2: This graph shows yield per step vs. test time per step for successive test steps; half the remaining faults were detected in the next step, which takes twice as long as the previous step.

check those externally. Consequently, IEEE P1500 and related I/O tests are critical.

There are some misperceptions in the industry that need to be addressed. For example, you may ask, "If I get 90% yield, have I got a 90% good product?" The answer is no. You only know that 90% of the tests you have run yield positively, but the tests you have run may not cover 90% of the potential faults. And then the next question is, "If I spent ten seconds to get to 90%, how long do I need to spend to get the rest of the percentage?" And here you face the law of diminishing returns. You can approach 100% coverage exponentially, but some faults you are never going to find.

Math examples

Consider the math behind a sequence of KGD test steps, based on some reasonable assumptions regarding test time and quality of product. Assume half of remaining faults are detected in each successive test step, which takes twice as long as the preceding step (Figure 2). Further, assume a small 10x10mm device fabricated in 10nm technology with 650 die per wafer and a four-site test with 90% yield at wafer probe. As shown in Figure 3, the shift-left strategy increases manufacturing cost by 9% but reduces unfound failures shipped by 51%. The walkaway here is that you can increase quality by finding 51% of the previously unfound faults before shipment with a small investment in additional testing.

For a larger logic part, you will have to work harder to obtain observability and controllability, and test time will be longer. Nevertheless, the same metric seems to apply as for the smaller part. Consider single-site testing of a 10nm device with 85 die per wafer. Assume that without using a shift-left flow, this device requires a 30s wafer-probe time at 75% yield and a 60s KGD test time for 87.5% yield. With the shift-left flow, manufacturing cost increases 16%, but unfound failures shipped are reduced by 53% (Figure 4). Finally, consider a multi-chip module that, without left shift, requires a 520s final test for 72.8% yield and a 1,040s system-level test for 85.4% yield. The shift left flow reduces manufacturing cost 28% and increases parts shipped by 37% (Figure 5).

Test system changes

For an effective shift-left strategy, test systems will require some changes. One key point is that active thermal control is becoming important, and it is something you cannot do on wafer probers. With wafer probers' chucks and high thermal mass, you can just set them at a temperature and hope it stays there.

	Without Shift-Left					With Shift-Left				
	# Out	# Fail	Test Time	Yield	$$	# Out	# Fail	Test Time	Yield	$$
	650	0	-	-	1x	650	0	-	-	1x
Wafer Probe	585	65	10	90%	1.14x	556	94	30	85.5%	1.24x
Thin Bump Saw	579	6	-	99%	1.15x	550	6	-	99%	1.25x
KGD Test	550	29	20	95%	1.25x	536	14	40	97.5%	1.36x
	550	27	-	-	-	536	13	-	-	-

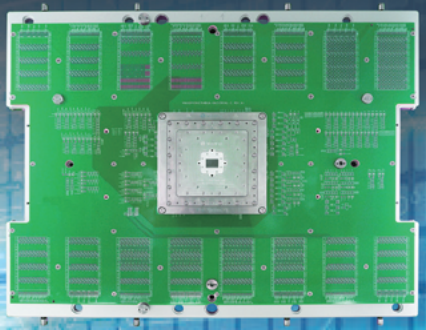
Figure 3: For a small device, the shift-left strategy increases manufacturing cost by 9%, but reduces unfound failures shipped by 51%.

	Without Shift-Left					With Shift-Left				
	# Out	# Fail	Test Time	Yield	$$	# Out	# Fail	Test Time	Yield	$$
	85	0	-	-	7.4x	85	0	-	-	7.4x
Wafer Probe	64	21	30	75%	10.2x	56	29	90	65.6%	12.4x
Thin Bump Saw	63	1	-	99%	10.3x	55	1	-	99%	12.5x
KGD Test	55	8	60	87.5%	12.4x	52	3	120	93.8%	14.3x
	55	7	-	-		52	3	-	-	

Figure 4: For a large logic part shipping as a KGD using the shift-left strategy, manufacturing cost increases 16%, but unfound failures shipped are reduced by 53%.

WinWay Technology

Your trusted partner in IC testing

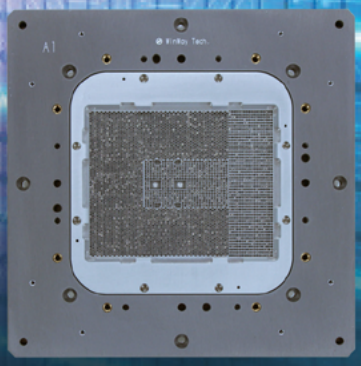


Vertical Probe Card

- Hign pin count > 30,000pins
- C4 pad pitch down to 75um

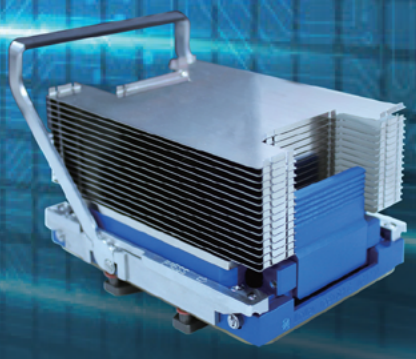
WLCSP Probe Card

- Maximum pin count up to 3000
- Minimum pitch down to 120um



Coaxial Socket

- SerDes 112Gbps PAM4
- Pitch down to 0.35mm



Burn-in Socket

- 850W heat dissipation capability
- Creative lid design

Thermal Control System

- 800W heat dissipation solution
- Mass production for handler

sales@winwayglobal.com

www.winwayglobal.com

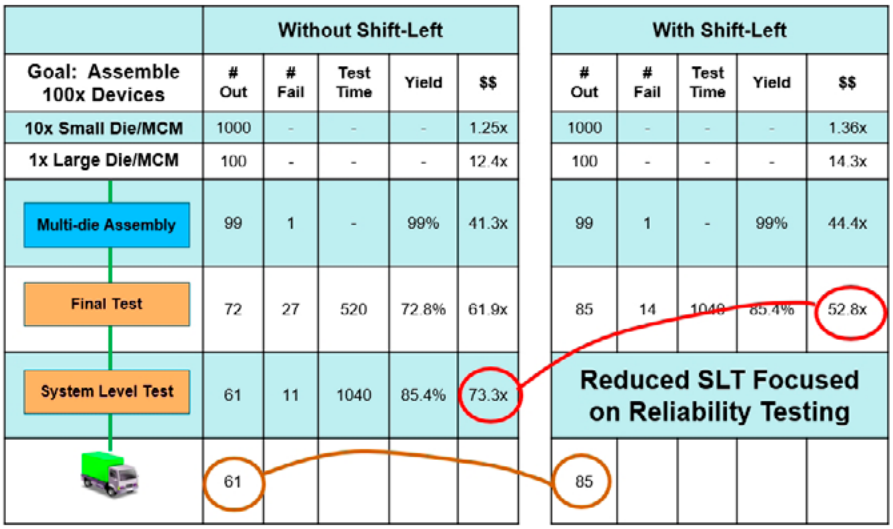


Figure 5: For a multi-chip module, the shift-left strategy reduces manufacturing cost 28% and increases parts shipped by 37%.

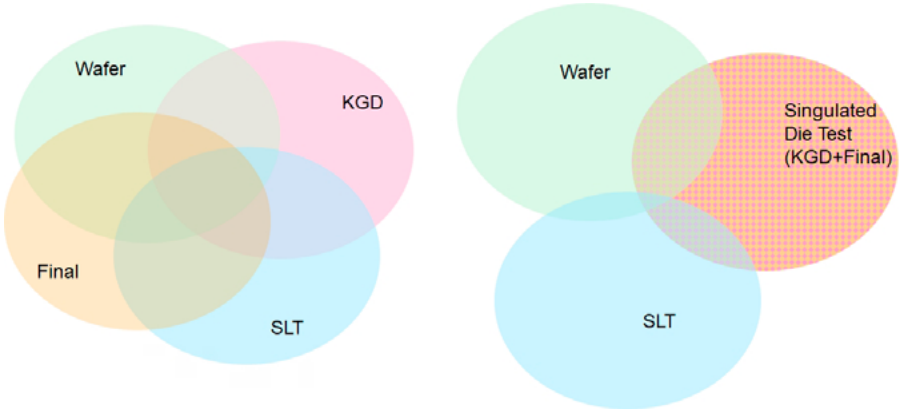


Figure 6: The move from: a) (left) a traditional test flow to b) (right) a shift-left test flow eliminates one test insertion and reduces over-testing.

Active thermal control (ATC) is feasible at the KGD test step, however, using a system such as the Advantest HA1000 die-level handling and probing system for singulated die testing.

In addition to ATC (with junction-temperature feedback per the device under test [DUT]), a KGD tester is able to take on final-test functions such as at-speed testing, which will require high-speed instruments, high-power supplies, high-frequency probes, and high current-carrying-capacity probes.

As KGD testers take on final-test functions, some traditional KGD tests, including two-temperature testing, will shift left to wafer test. And finally, system-level test functions, including packaged device test, boot-up tests, and fuse-blowing, will shift left and can be performed on either a traditional final-test system or on a system-level-test system. What is required at this step is a system-focused environment with system-focused code that can boot up the device under test and run its firmware and software. A key benefit of the shift left flow is that it can eliminate one test insertion. One less test insertion means one less test cell—one less handler—bringing about significant financial benefits.

A related issue the industry is contending with is over-testing. If today we have wafer test, KGD test, final test, and system-level test, the reality is that a lot of tests are run four times, or at least two or three times. If we remove one test insertion, then by definition you are not going to be doing any tests four times over. In general, the Venn diagram regions in **Figure 6** will move away from each other, resulting in less overlap.

Summary

Several conclusions arise regarding this overview of the shift-left concept. First, test content itself is an incredibly valuable resource, and it is a resource that we can move around. For example, an f_{MAX} test can be done at the wafer level, it can be done at final test, or it can be done at system-level test. What we need to look at is where is the best place for us to do this important test for each device.

If you are shipping KGD, prior to shipment you will need to screen for obvious failures, confirm functionality at temperature extremes, find assembly-induced problems, and perform speed and power binning, full built-in self-test (BIST) and scan, high-speed I/O test, and fuse blowing.

In addition, this overview of shift left reconfirms what may be obvious to anybody in the industry: more testing sooner will increase the quality of your product for a small incremental cost increase. You will have to pay for test in any event, and making that payment sooner will allow you to: 1) save packaging cost, 2) reduce the number of good parts scrapped because of another part's problem in a shared multi-die assembly, and 3) it will allow you to end up with an ultimately lower cost, more profitable product.

Biography

Dave Armstrong is Director of Business Development at Advantest America, Inc., San Jose, CA, and is also Chairman of the Test Technology Working group for the Heterogeneous Integration Roadmap. Prior to Advantest, he spent over two decades in HP/Agilent's IC test group, and before that he worked in the semiconductor industry in areas of IC and system design, product/yield engineering, as well as test engineering. He has degrees in Electrical, Computer, and Environmental Engineering from the U. of Michigan. Email d.armstrong@advantest.com

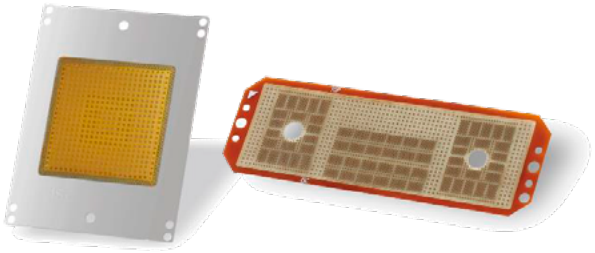
ABOUT I.S.C

Global No.1! Total Test Solution Provider!

ISC provides all the answers to the 5G semiconductor test solutions.
ISC is here as the future approaches with IT, BT, Autonomous Car, and 5G.

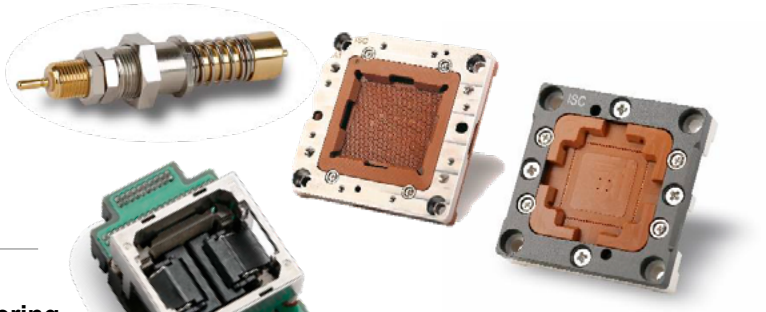
ISC is a total test solutions based on a product portfolio that can test all semiconductor IC and IT products, providing the highest quality products in the shortest time to realize customer satisfaction and impression

ELASTOMER SOCKETS & INTERPOSERS



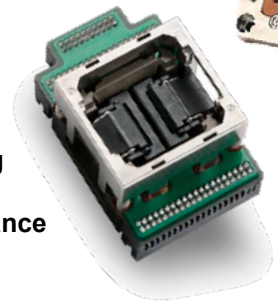
- High performance and competitive prices
- High speed & RF device capability
- No load board pad damage & no contact trace on the ball
- Customized design to meet challenging budget constraints
- Full thermal and electrical simulation

POGO SOCKET SOLUTIONS



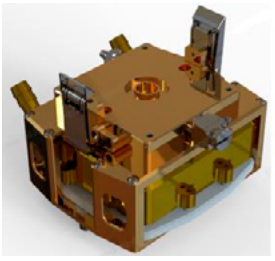
- Various design available
- Excellent gap control & long lifespan
- High bandwidth & low contact resistance

BURN-IN SOCKET SOLUTIONS



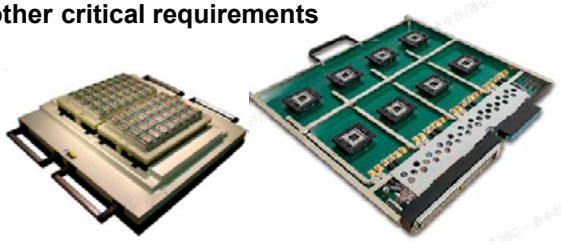
- No ball damage
- Direct inserting on Burn In Board without soldering
- No damage on Burn In Board land
- Simple structure without sub PCB & easy maintenance

THERMAL CONTROL UNIT



- Extreme active temperature control
- Customized design to meet challenging requirements
- Price competitiveness through self-designing and fabrication
- Safety auto shut-down temperature monitoring of the device & thermal control unit
- Full FEA analysis for strength, deflection, air flow and any other critical requirements

Hi-fix & Burn In Board



- High performance and competitive price
- Test fine pitch, high speed device at hot & cold temperature
- Customized design to meet individual requirements

CONTACT ISC CO., LTD

ISC HQ
Seong-nam, Korea

ISC International
Silicon-valley, CA

Tel: +82-31-777-7675 / **Fax:** +82-31-777-7699
Email: sales@isc21.kr / **Website:** www.isc21.kr

Emerging process and assembly challenges in electronics manufacturing

By Glenn Farris [Universal Instruments]

The semiconductor and semiconductor equipment industries expect to see a strong upturn in the next few years, with advanced packaging technologies a significant beneficiary of the market's strength. 5G, artificial intelligence, edge computing, persistent memory, integrated power management, and the transition to sub-5nm silicon technology are all driving the need for innovative packaging solutions. These solutions integrate silicon produced with disparate process nodes and deliver maximum performance at optimal cost.

Heterogeneous integration (**Figure 1**), which utilizes a multitude of interconnect methodologies (from fan-out to silicon interposer, to chiplet), addresses this challenge but requires unique solutions for efficient, cost-effective die placement. High-speed, high precision multi-die placement, directly and efficiently extracted from a range of different sized wafers, is critical to enable cost-effective assembly.

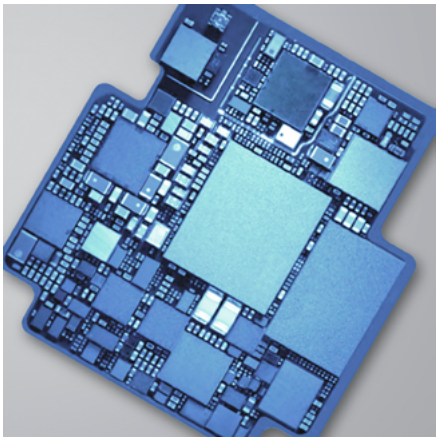


Figure 1: Example of heterogeneous integration application.

Complex multi-die architectures support a wide range of applications, including: insulated-gate bipolar transistor (IGBT), antennae-in-package (AiP), microelectromechanical systems (MEMS), high-performance computing (HPC), and advanced packaging products, each of which benefit from this technology. These architectures also create several challenges for efficient, cost-effective assembly. Accurate die placement, efficient changeover from picking one die type, or wafer type, to a different type, and the ability to mix wafer, tape, and tray fed material are critical to enable mass production solutions.

Alternatives for implementing multi-die architectures

Multiple alternatives have been proposed to implement the multi-die architectures noted above. Some of these examples are described below.

EMIB. Embedded multi-die interconnect bridge (EMIB) is an elegant and cost-effective approach to the in-package high-density interconnect of heterogeneous chips. EMIB uses a very small bridge die with multiple routing layers. This bridge die is embedded as part of the substrate fabrication process.

FOPLP (fan-out panel-level packaging). One of the latest packaging trends in microelectronics is FOPLP, which has a high potential for significant package miniaturization concerning package volume, but also with respect to its thickness. The technological core of FOPLP is the formation of a reconfigured molded wafer combined with a thin-film redistribution layer (RDL) to yield a surface-mount device (SMD)- compatible package.

FOWLP (fan-out wafer-level packaging). FOWLP is an integrated circuit packaging technology, and an enhancement of standard wafer-level packaging (WLP) solutions. In FOWLP, the wafer is diced first, but then the dies are very precisely re-positioned on a carrier wafer or panel, with space for fan-out kept around each die. The carrier is then reconstituted by molding, followed by making a RDL atop the entire molded area (both atop the chip and atop the adjacent fan-out area), and then forming solder balls on top.

InFO (integrated fan-out wafer-level packaging). InFO is an innovative wafer-level system integration technology platform, featuring high-density RDL and through-InFO via (TIV) for high-density interconnect and performance for various applications, such as mobile, high-performance computing, etc.

CoWoS® (chip-on-wafer-on-substrate). CoWoS® is a 2.5D wafer-level multi-chip packaging technology that incorporates multiple dies side-by-side on a silicon interposer in order to achieve better interconnect density and performance. Individual chips are bonded through micro-bumps on a silicon interposer forming a chip-on-wafer (CoW). The CoW is then subsequently thinned such that the through-silicon via (TSV) perforations are exposed. This is followed by C4 bumps formation and singulation. A CoWoS® package is completed through bonding to a package substrate.

SOIC (small outline integrated circuit). SOIC is a surface-mounted integrated circuit (IC) package that occupies an area about 30–50% less than an equivalent dual in-line package (DIP), with a typical thickness being 70% less.

SiWLP (system in wafer-level package). A SiWLP is fabricated using "RDL-first" technology for FOWLPs and provides high chip-I/O density, design flexibility, and package miniaturization.

2.5D. 2.5D is a packaging methodology for including multiple die inside the same package.

eWFO (embedded wafer fan-out). eWFO is fabricated in either wafer fabs using back end of line (BEOL) tools, materials and processes, or at outsourced semiconductor assembly and test suppliers (OSATS) using their built-up fabs and tools.

Heterogeneous integration

Moore's law is still providing a reduction in transistor size by a factor of two each year, but in the most advanced nodes we are no longer getting a corresponding reduction in cost. The combination of the need for performance advancement at lower cost is leading to new architectural paradigms. Printed circuit board assembly (PCBA)-based systems need to shrink into microelectronic form factors to address the needs of edge computing. SoC semiconductor devices are disaggregating to optimize process nodes by function.

The net result of the disaggregation noted above is a new approach to device packaging: heterogeneous integration. Rather than trying to cram functionality into a smaller package, the world is moving to optimize the performance of the chip with the performance of the package. This results in complex packaging assembly needs, and the need to support much thinner die handling. These resulting solutions require the combination of multiple die part numbers transferred from different wafer types (see **Figure 2**).

In the case of InFO, two different die types are evident, including SoC and dynamic random-access memory (DRAM). Each of these may utilize a different feeding source, potentially with one device fed direct from the wafer and another from a Joint Electron Device Engineering Council (JEDEC) standard tray. The key advantages of InFO are higher compute density and faster training time.

In the case of CoWoS®, three die types are evident, an SoC, a DRAM, and a silicon interposer (Si). These span a very wide die size range, requiring highly flexible placement platform capability.

TSMC's innovative CoWoS® advanced packaging technology (**Figure 2**) integrates logic computing and memory chips in a 3-D manner for advanced products targeting artificial intelligence, cloud computing, data center, and super computer applications. This revolutionary 3-D integration facilitates power-efficient, high-speed computing while reducing heat and CO_2 emissions.

The transition to either silicon-based chip-to-chip interconnects or lithographically-processed interconnects requires extreme placement accuracy (**Figure 3**). At these pad pitches, placement accuracy below 10µm is required. Assembly

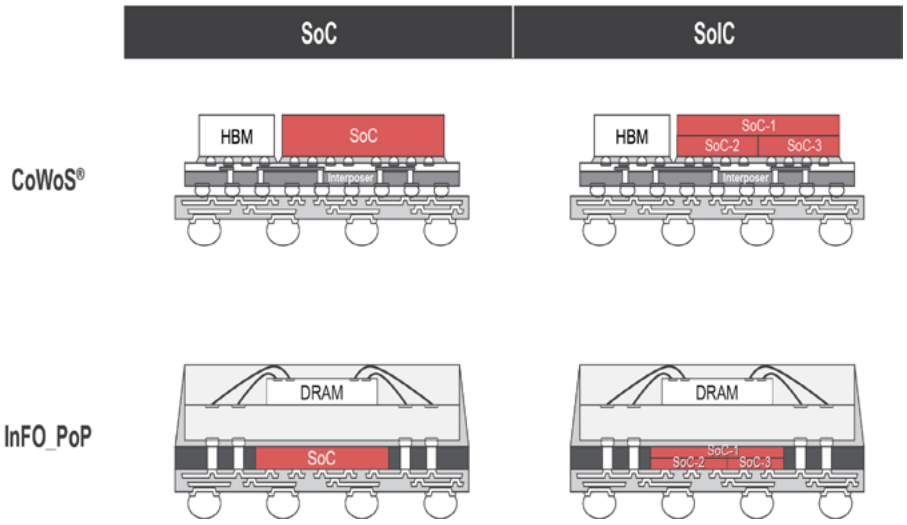


Figure 2: Example of CoWoS® architecture (top) and InFO_PoP architecture (bottom). Both applications are represented with and without SOIC integration.

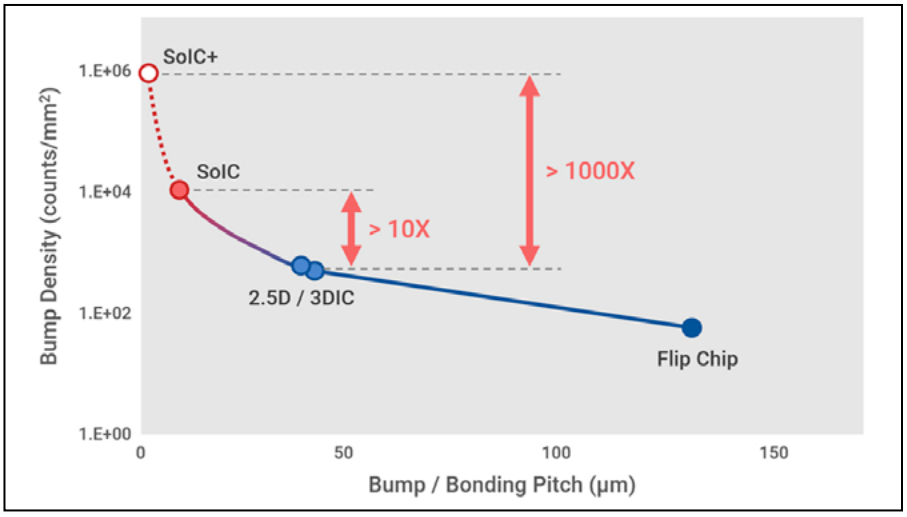


Figure 3: Representation of typical interconnect pad pitch for 2.5D and 3D structures. The graph highlights pad pitch for 2D vs. SOIC.

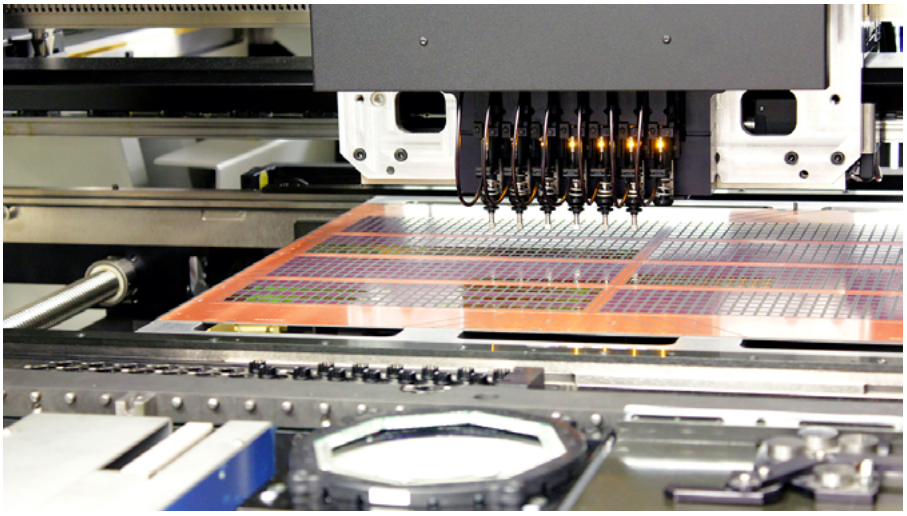


Figure 4: Wafer-level fan-out application on a 600mm x 600mm substrate.

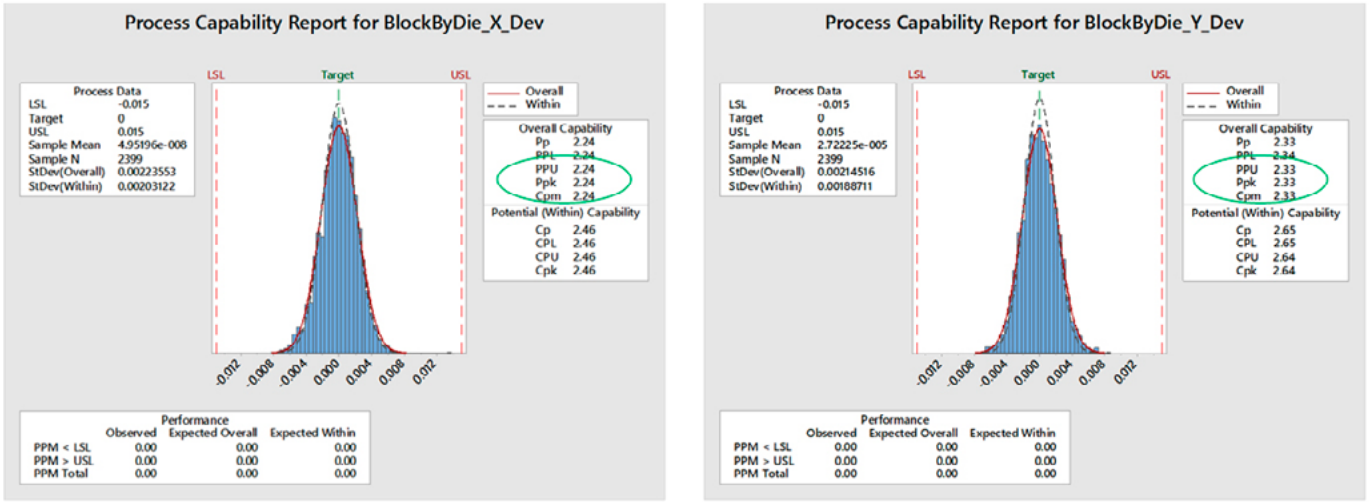
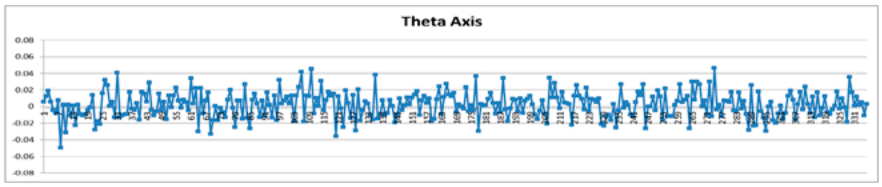
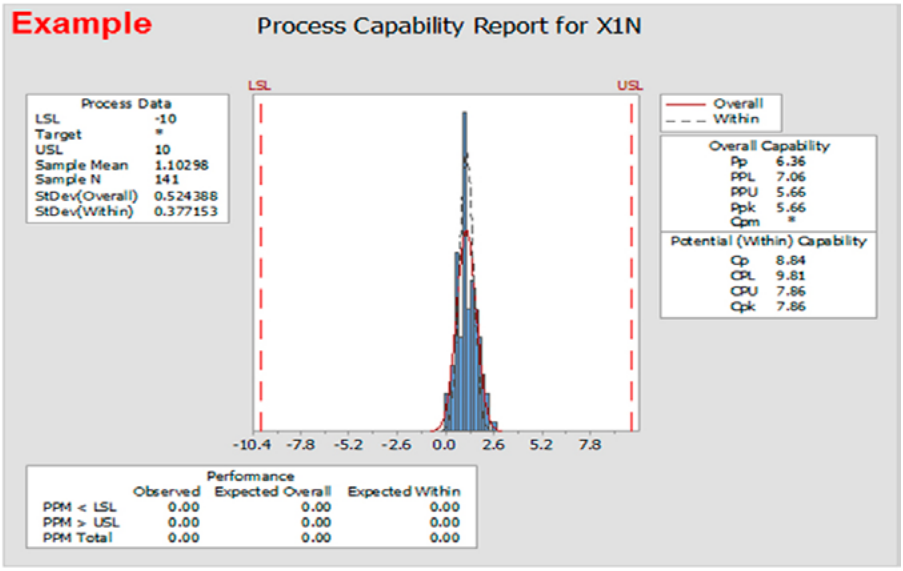


Figure 5: X-axis and Y-axis placement accuracy data at 16Kcph shows a standard deviation of <2.3µm.



ALL ROTATIONS	X-AXIS	Y-AXIS	THETA
Min	-7.34	-6.58	-0.04949
Max	7.01	6.96	0.04672
STDev	2.83	2.50	0.02
Avg	-0.84	-0.41	0.00
Cpk	1.08	1.28	1.56

Figure 6: Theta accuracy statistics for 40mm x 40mm die on interposer.



CPK	X1N	Y1N	X2N	Y2N
1.33	±3.19 µm	±3.01 µm	±3.16 µm	±2.88 µm
1.67	±3.73 µm	±3.58 µm	±3.83 µm	±3.61 µm

Figure 7: XY scatterplot of AMS results. The table shows specification limits corresponding to Cpk values of 1.33 and 1.67 based on the above means and standard deviations. Data indicates system accuracies of <3.2µm @ Cpk 1.33.

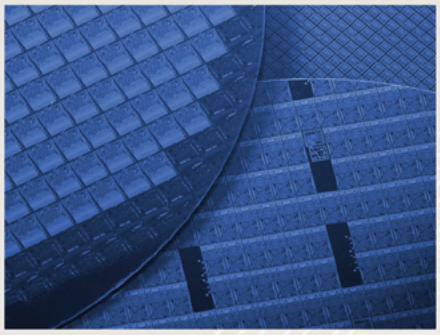
of these heterogeneous integration structures in a FOWLP, FOPLP, SLP or embedded process requires that this accuracy must be maintained over a placement area as large as 600mm x 600mm (**Figure 4**). To determine the feasability for these requirements, we assembled over 12 panels with an approximate 6mm x 6mm die with sub-50µm pitch copper pillar bumps utilizing our FuzionSC™ platform.

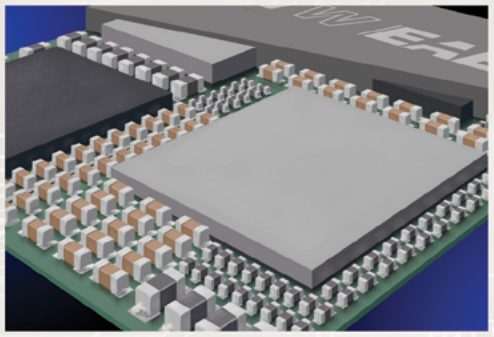
Four panels were built to establish a baseline and to validate trimming requirements by spindle and by placement location. These results were then incorporated into the placement map of the system. Six panels were then assembled and all placements were measured for X, Y and Theta variation (**Figure 5**). As can be seen from the data, the system was validated to be able to place die at speeds >16Kcph, with an accuracy of <2.3µm standard deviations.

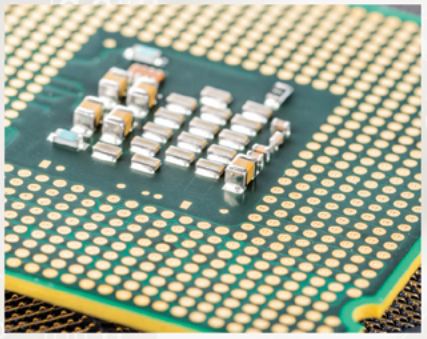
As die sizes for high-performance computing applications grow, the theta accuracy also becomes critical for precision pad to bump alignment. A study was completed assembling 40 large bumped die on an interposer, with all results measured (**Figure 6**). The results of this study demonstrated a capability of <0.075deg @ Cpk 1.56.

In high-volume manufacturing, active monitoring and control is required to maintain accuracy as a function of temperature, time, or number of placements, which prevent drift. An example solution is an accuracy management system (AMS), which monitors the placement performance of each spindle using a standard high-precision slug, and actively adjusts

MPM | Camalot | Electrovert | Vitronics Soltec | Despatch







Printing, Dispensing, Reflow, Cleaning and Thermal Processing Equipment for Semiconductor Packaging

ITW EAE is advancing innovation and development of next generation technology for semiconductor packaging. New technology that dramatically improves productivity and yields.

The MPM® Edison™ is the most accurate printer in the market. The Camalot® Prodigy™ dispenser features advanced technology like the Dynamic Dual Head™ that can synchronously dispense with two pumps regardless of part-to-part rotation. Vitronics Soltec reflow systems provide unmatched reliability. Electrovert® centrifugal and in-line cleaning systems provide high-performance cleaning of advanced packages. Despatch® offers high-performance ovens for polyimide curing and more.

Designed for semiconductor yield improvement

Electronic Assembly Equipment

ITW EAE

Learn more at www.itweae.com

A division of Illinois Tools Works

spindle height to control for any variation identified.

Studies were conducted to determine baseline system performance using the AMS, such that any spindle variation over time could then be controlled (**Figure 7**). As can be seen from the results, AMS can resolve down to system accuracies of <3.2µm @ Cpk 1.33.

Substrate and interconnect dimensions require placement capability over large substrates up to 600mm x 600mm at sub-10µm XY accuracy and sub-0.075° theta accuracy to achieve high yield. Placement rates of >16K have been demonstrated at this performance level—a key aspect to minimizing the cost of assembly.

Additionally, as die thicknesses are reduced below 100µm, there is an increased potential for cracking, chipping or uncontrolled warpage when handling and transporting die in tape or tray. Taping or tray transfer also has negative impacts on lead time and work in process (WiP). By transitioning to direct pick from wafer with a high-speed placement platform, die damage can be eliminated, while lead-time and WiP are reduced. Direct pick from wafer, however, presents several challenges that must be addressed. First, the process of stretching and unstretching a wafer can also lead to chipping, and must be eliminated. Second, with multiple die types and wafer types, a single placement solution needs to efficiently handle multiple die types picked from different wafers by minimizing the "changeover" time from one wafer type to a different wafer type.

A solution that can stretch and store wafers in an assembly system, asynchronous from pick operations in a wafer table, can eliminate stretch/unstretch cycles as well as maximize overall system utilization. Such a solution has been developed and evaluated utilizing a patented cartridge storage and transfer system for wafers. Data collected has demonstrated that such a system eliminates the need to unstretch a wafer until consumed. It also results in a reduction in wafer "changeover" time of approximately 50%. For large die applications, where the overall time to pick all die on a single wafer can be under 120 seconds, this can result in an overall system throughput improvement on the order of 25%. Likewise, for heterogeneous integration applications where a wafer exchange may occur in as few as 30 seconds, this can lead to overall throughput improvements of up to 100%.

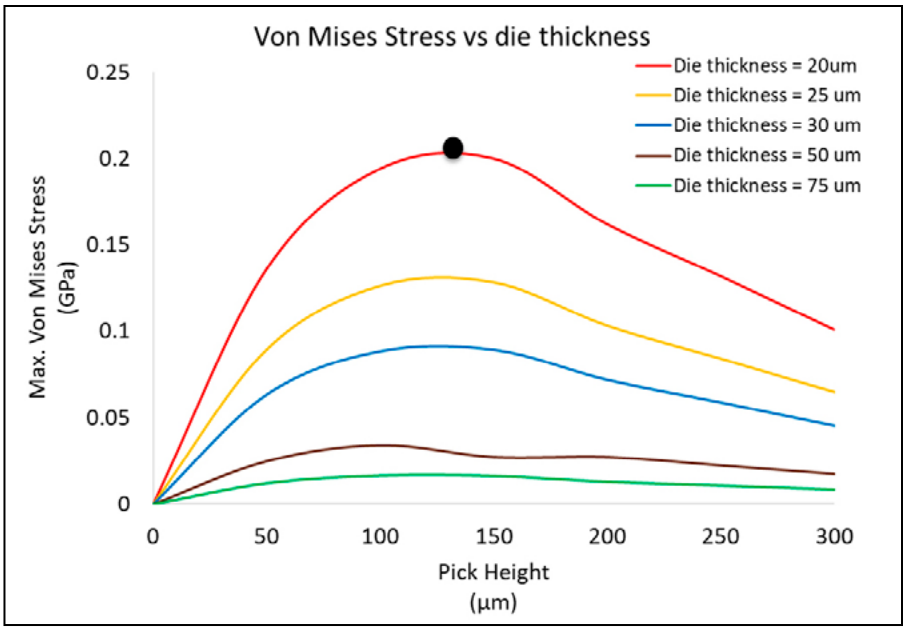


Figure 8: Study on key ejection parameters affecting die stress to ensure optimal ejector performance.

Key benefits of the patented cartridge system includes: 1) minimizing de-expansion events; 2) managements of multiple wafer sizes; 3) reduction of downtime for tooling changeover; 4) accommodation of difficult die types (thin, large, high aspect ratio); and 5) elimination of die chipping with programmable stretch.

Key features of the online expansion system include: 1) eliminates the need to pre-expand wafers; 2) supports four 13-slot wafer cassettes or two 25-slot cassettes; and 3) four slots per side allow for management of four cartridges each.

Another feature of this solution is a high-precision (sub-micron X,Y,Z) servo-driven ejector that precisely releases the die from the wafer, allowing for fast wafer to placement handoff. In order to determine the correct ejector operation, our Advanced Process Lab (APL) undertook a study to examine the key ejection parameters affecting die stress. Results of this study are presented in **Figure 8**.

Summary

In conclusion, advanced semiconductor packaging applications are growing in volume and complexity and require new assembly solutions to ensure high yield at the best overall cost per placement. Innovative solutions have been assessed and proven to be viable, delivering the required accuracy over the SEMI standard large panel format. These solutions have additionally demonstrated the speed and utilization required for efficient and economic assembly operations. Heterogeneous integration delivers the ability to optimize the performance of the chip with the performance of the package.

Biography

Glenn Farris is VP Marketing and Corporate Strategy at Universal Instruments, Conklin, NY USA. He joined Universal in 2013 as VP, Marketing and has been instrumental in leading the company into developing strategic relationship with some of the world's largest technology leaders. He holds an MBA in Marketing and Finance from Santa Clara U., an MS in Engineering from Stanford U., and a BS in Engineering from Purdue U. Email glenn.farris@uic.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

Wafer-scale superconducting multi-chip module

By Rabindra N. Das, Vladimir Bolkhovsky, Alex Wynn, Ravi Rastogi, Scott Zarr, Leonard M. Johnson
[Quantum Information and Integrated Nanosystems Group, MIT Lincoln Laboratory]

This paper describes a strategy to combine laser direct write (LDW) and optical lithography (I-line) to fabricate 200mm wafer-scale superconducting multi-chip modules (S-MCM) for interconnecting multiple active superconducting flux quantum (SFQ) chips for next-generation cryogenic processing systems. The packaging strategy includes the development of S-MCM (48mm X 48mm) using large single I-line reticles, followed by reticle stitching to fabricate nearly the largest possible stitched S-MCM (96mm X 96mm) using a four mask/layer process. The stitching process starts with sequential exposure of multiple I-line photomasks – with small overlap (stitched area) – to realize larger combined circuit areas for design-critical S-MCM layers with minimum linewidths of 0.8-1µm. The process also utilizes laser direct write (LDW) lithography to make wider (>1µm) features such as fan-out circuits, extending the stitched circuit area to include the entire 200mm wafer as a single S-MCM.

Introduction

As CMOS reaches the end of Moore's Law, scaling and power consumption continue to be a challenge, thereby driving the need to develop "beyond-CMOS" device technologies to advance high-performance computing. Superconducting electronics using Josephson junctions (JJs) as active devices are a promising candidate for high-performance computing because of their extremely low gate energies, fast clock speeds, and lossless signal propagation for data transport [1]. A major technical challenge facing superconducting circuit technology is achieving a very large scale of integration (VLSI). Developing a VLSI capability (10^7 or more JJs) with 100s of SFQ chips in proximity to one another along with auxiliary semiconductor electronics (e.g., power supplies, clock generators, output amplifiers) in a single system is highly desirable to realize lossless circuit functionalities required for superconducting computing architectures. However, such VLSI capability has yet to be demonstrated.

This paper presents a system-on-wafer approach for integrating a large number of SFQ chips onto a full 200mm wafer S-MCM. This approach increases the circuit complexity (number of JJs) that can be integrated within a given cryogenic space by producing high chip-to-chip connectivity. Connecting superconductor electronics components using our approach enables a parallel scaling path. Connectivity of individual chips through superconducting wiring offers significant advantages relative to the equivalent integration methods for semiconductor-based electronics. With low-loss superconducting lines, power requirements for drivers and receivers for inter-chip communication can be equivalent to on-chip communication. This kind of communication is in stark contrast to CMOS-based integration where the number of I/Os can quickly dominate the power budget for a similar system. We present a niobium-indium microbump scheme to produce a variety of stitched and wafer-scale S-MCMs. This scheme enables the heterogeneous integration of known good chips to increase circuit density, functionality, and reduce the circuit footprint. In addition, we discuss thermocompression bonding of niobium-indium microbumps and their electrical performance at cryogenic temperatures.

Development of wafer-scale S-MCM proceeded in three steps. In subsequent sections, we will discuss the fabrication of a single reticle S-MCM, multiple reticles based stitched S-MCM, and stitched reticles with laser direct writing based wafer-scale S-MCM.

Single reticle superconducting MCM

As an intermediate step towards demonstrating a large-area stitched S-MCM, a 48 x 48mm^2 S-MCM fabrication using a single I-line mask set (without stitching) was evaluated. The S-MCM comprised four superconducting metal (Nb) layers and one resistor layer, allowing for several impedance-controlled clock and data lines. The MCM stack-up [2] requires a 0.8µm wide line to achieve 50 Ohms impedance. For S-MCM, all the critical layers with tight impedance-controlled lines are fabricated by I-line lithography. The primary goal of this design is to evaluate minimum feature sizes for 48mm x 48mm circuits with a single mask exposure. The design includes snake/comb test structures and critical dimension (CD) cells (0.8µm lines) around the 48mm x 48mm periphery to evaluate the maximum S-MCM size within the limitations of the I-line photolithography tool; patterning aberrations may occur near the edge of the full reticle field, which is monitored using these structures. Second, the design includes perimeter and interior array interconnects that cover full reticle (20mm x 20mm) to be tested with SFQ chip flip-chip bonding. Scanning electron microscopy (SEM) results indicate that it is possible to create 48mm x 48mm critical circuits with a single I-line mask exposure, with the potential to be extended to the entire I-line field for stitched MCMs. In the next section, we will show stitching of 48mm x 48mm I-line masks (four masks/layer) to produce 96mm x 96mm MCMs with 0.8µm CDs for Nb routing layers.

Prior to the stitching demonstrations, the flip-chip bonding process was tested using a two-component daisy chain structure. The first half of the structure consists of a single 200mm diameter S-MCM with many 20mm x 20mm bumped areas. Each bumped area contains many short and discontinuous wire segments terminated at a bump on either end of the segment. The second half of the structure is created by passive 20mm x 20mm Si chips that contain complementary wire segments terminated at bump pads; when the two halves are bonded together, they form a

DISCOVER THE HSWF

Universal Instruments

Advanced Packaging enables the world's most innovative products for 5G, AI, Edge Computing, AiP, and Integrated Power Management.

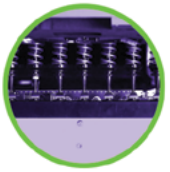
Consumer, Automotive and IoT devices are getting smaller and thinner, 2.5D and SiP multi-die packages are becoming more complex, InFO and Panel Fan-Out are driving large-format batch processing, and volume demands are rising.

These new challenges demand a comprehensive solution that breaks traditional boundaries for efficient multi-die assembly.

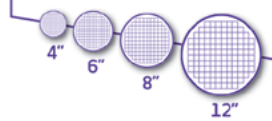
Introducing the **High-Speed Wafer Feeder** from Universal Instruments – **The world's fastest rapid-exchange multi-die feeder**. Combined with Universal's FuzionSC™ Platform, it is the **ultimate multi-die solution for heterogeneous integration**.

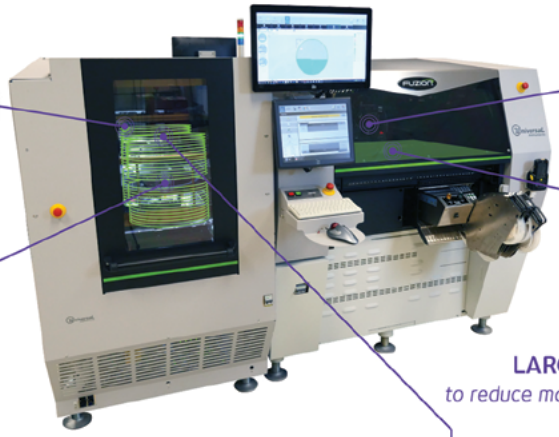
- 16K cph dual-wafer tables, low-force pick heads
- Up to 52 unique wafers types at once; 4", 6", 8", 12"
- One-step "wafer-to-placement" handoff
- Synchronous wafer stretch and storage
- Maximum panel sizes up to 635mm x 610mm
- Maximum die size range with thinnest die

WAFER CAPACITY
to minimize replenishment rate



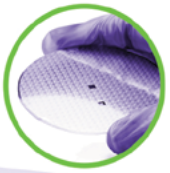
SPEED
to meet volume requirements



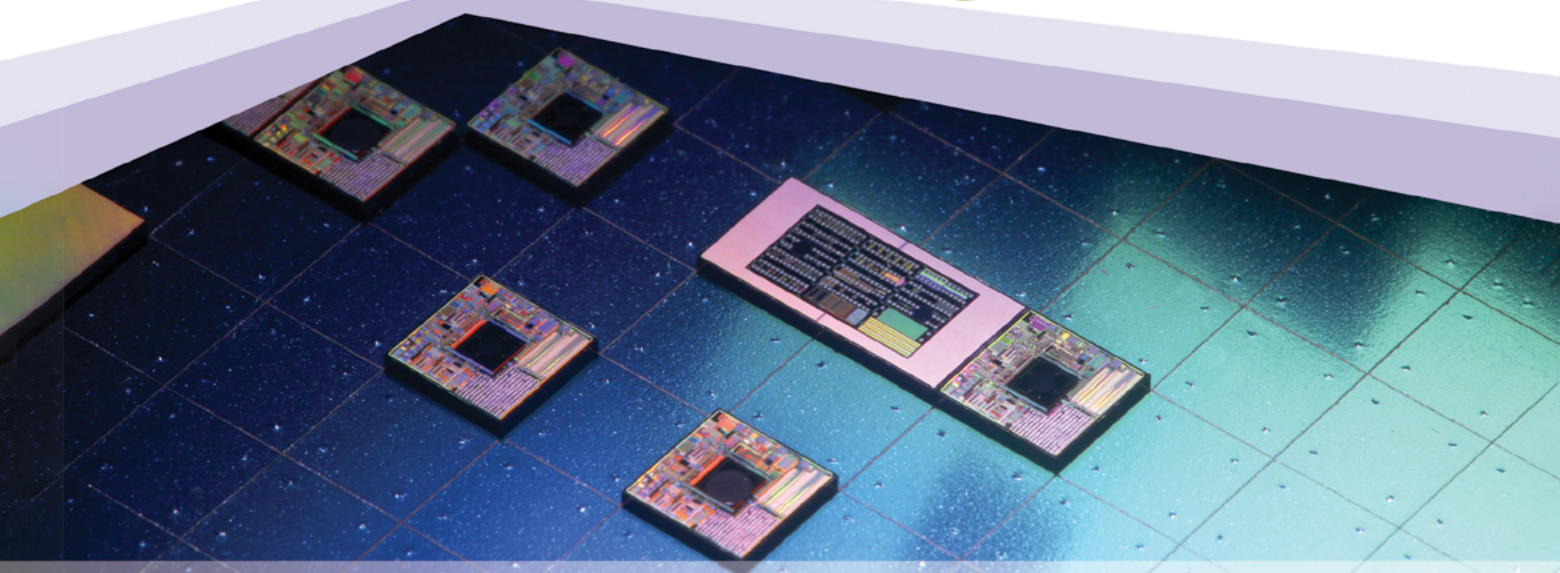


LARGE SUBSTRATE
to reduce manufacturing costs

MULTIPLE DIE TYPES
to maximize utilization



THIN DIE
to maximize sub 100-micron yields



uic.com

complete daisy chain that can be tested for continuity and total resistance. These devices are surrogates for a superconducting wafer-scale MCM and the associated superconducting chips to be connected. The current microbump fabrication process uses a single I-line liftoff process on the superconducting MCM. An interconnect layer (1000nm Nb, 20nm Ti, 50nm Pt, 150nm Au, 2000nm In) and an underbump metalization (UBM) layer (1000nm Nb, 20nm Ti, 50nm Pt, 150nm Au) are evaporated onto the S-MCM, and superconducting chips, respectively. The I-line photo process is important for reducing wafer-scale photoresist defects, defining bump diameter, and improved alignment to maintain electrical properties and minimize bump-bonding issues.

Figure 1 shows flip-chip S-MCMs bonded with 20mm x 20mm and 5mm x 5mm superconducting chips. We have fabricated 48mm x 48mm S-MCMs with niobium-indium (Nb-In) microbumps. Thermocompression (TC) bonding of Nb-In microbumps provides sufficient mechanical strength, creates low-resistance electrical interconnects, and produces minimum spacing between the chip and S-MCM. Furthermore, large Si-chip bonding reduces the total number of bonding cycles and simplifies the assembly process. For example, Figure 1 shows a 2-chip (20mm x 20mm) S-MCM and a 16-chip (5mm x 5mm) S-MCM, which used 2, and 16 TC bonding cycles, respectively to complete each S-MCM. Additionally, because of minimum spacing and edge keep-out regions for dicing and in-line metrology, 2-chip (20mm x 20mm) S-MCMs can have a significantly higher active circuit area compared to a 16-chip S-MCM for a given lithography process.

As a case study of larger area flip-chip interconnections, a 25mm x 25mm S-MCM with niobium-indium microbumps and a 20mm x 20mm superconducting chip were bonded together to fabricate a daisy chain structure. By alternating daisy chains in the lay-up prior to bonding, the Nb-In microbumps electrically connect the daisy chains. The S-MCMs include a 20mm x 20mm flip-chip daisy chain structure prepared with approximately 10,000 or 100,000 niobium-indium microbumps. To assess the electrical performance of the bump bonded full reticle chips, the S-MCMs with 20mm x 20mm chips were attached to a PCB card and wire bonded to measure I-V characteristics of bump interrupted niobium at 4.2K [2]. Niobium-indium microbumps not only showed very low resistance in the range of 0.05-0.1 milliohm at 4.2K, but also maintained high niobium critical current. We measured 20 mm x 20 mm flip-chip daisy chains ranging in number of interconnect segments from a few thousand to tens of thousands of microbumps in series. The I-V curve of approximately 24,000 Nb-In microbumps series flip-chip daisy-chain had a niobium critical current in excess of 50mA at 4.2K. The large number of microbumps per chip, compact bump geometry, high critical current of niobium, and high current-carrying capacity of the microbumps enable this process to be suitable for building VLSI flip-chip structures and developing complex superconducting computing systems.

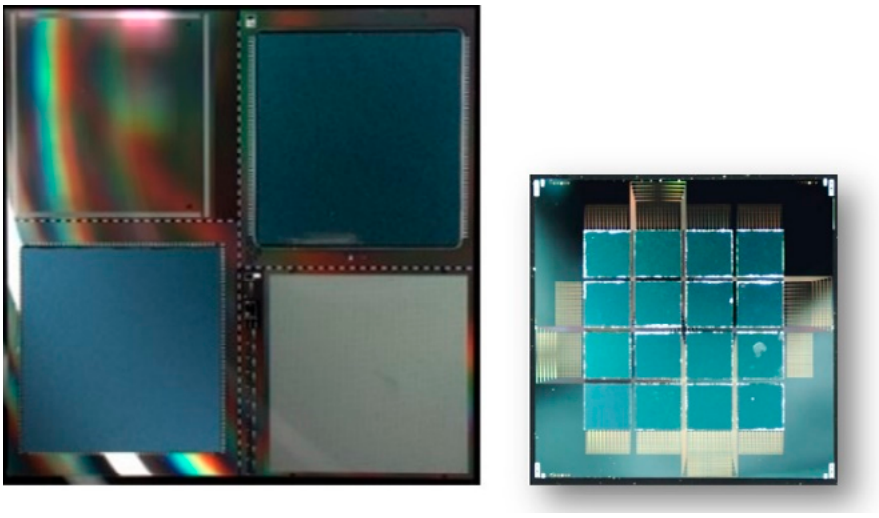


Figure 1: Superconducting multichip modules (S-MCM) with attached superconducting test chips: a) (left) Optical image of a 48mm x 48mm S-MCM bonded with two 20mm x 20mm superconducting test chips; b) (right) Optical image of 32mm x 32mm S-MCM bonded with sixteen 5mm x 5mm superconducting test chips. One 20mm X 20mm superconducting test chip and sixteen 5mm X 5mm chips have the same circuit area.

Stitched S-MCM

In this section, we detail the fabrication processes for reticle stitching. The previously described single reticle-based S-MCM fabrication process was modified to create a larger area stitched S-MCM. Figure 2 represents the stitching process where circuit lines of individual reticles are stitched at a stitch boundary. Four I-line photomasks (A, B, C, D) were joined together to create a stitched field. Photomask A and photomask B were stitched in the X-direction, while photomask A and photomask C were stitched in the Y-direction. Similarly, photomask B and photomask D were stitched in the Y-direction, while Photomask C and photomask D were stitched in the X-direction. The individual size of each photomask reticle (A, B, C, D) will determine the overall stitched field size. For example, a 35mm x 35mm reticle and a 48mm x 48mm reticle will produce 70mm x 70mm and 96mm x 96mm stitched S-MCMs respectively, for a four masks per layer process. Figures 2c and d also show an enlarged stitching area between photomask A and photomask B. Mask A has a circuit line extending into the chrome area. The extensions in the chrome area are defined as the overlap length where the line will expose twice (i.e., a double exposure). We optimize the overlap length as 0.25µm. So, for a stitched line with a 0.25µm overlap length on each side, a minimum 0.5µm long line at the stitch boundary will expose photoresist twice for stitching. A double-exposed resist line will distort the linewidth at the stitch boundary. An optimal 0.25µm overlap length for each side provided the least amount of linewidth distortion at the stitch boundary. Additionally, corresponding SEM micrographs of a 0.8µm line going through the stitch boundary (in the X-direction) shows a representative example. SEM images (Figure 2e) show an approximately 70-110nm linewidth variation within a 1µm

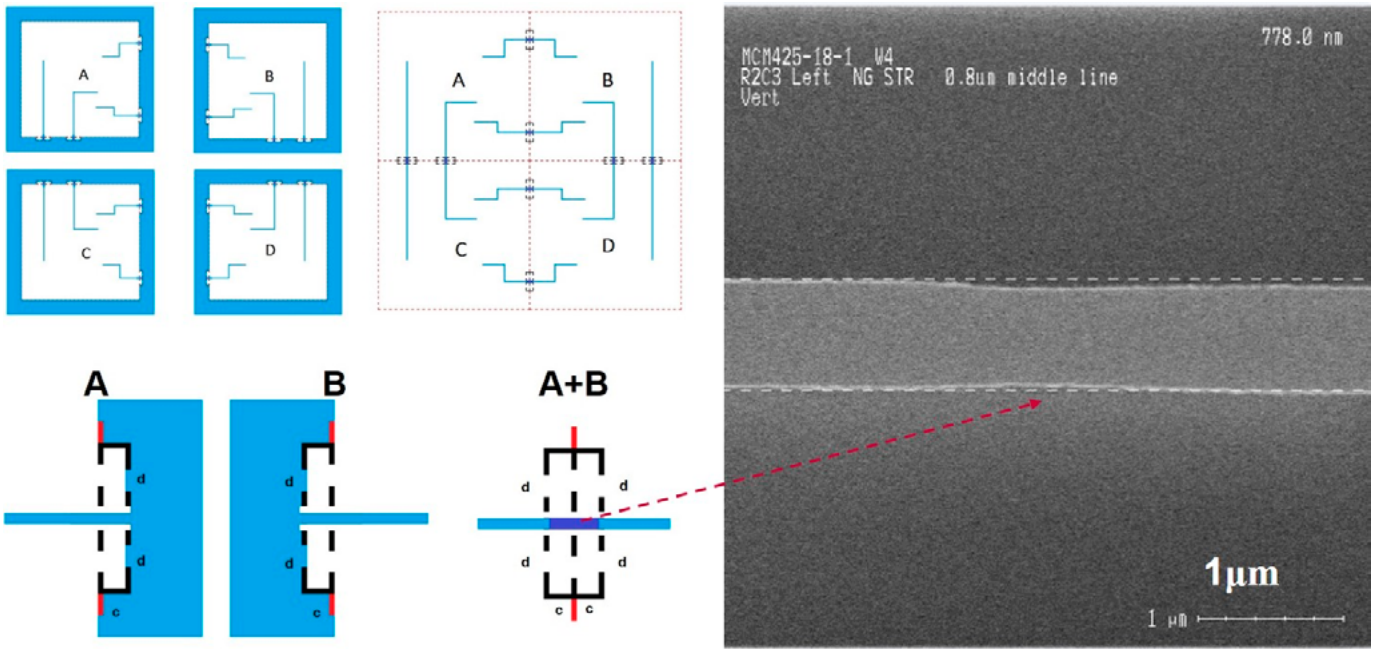


Figure 2: Illustration of the stitching process with I-line clear field photomasks. The blue color defined in the images is a chrome layer in the mask: a) (top left) Four individual (A, B, C, D) pre-stitched reticles and their orientation prior to stitching; b) (top right) A stitched reticle. Stitching of the four individual reticles produces a single stitched field. For example, four 35mm x 35mm individual reticles will produce a 70mm x 70mm-stitched field. The bottom images represent an enlarged stitch area: c) (bottom left) This image represents a circuit line within individual reticles prior to stitching, and d) (bottom right) represents a circuit line stitched in the stitched reticle. e) SEM micrographs of a 50 Ohms (0.8µm) line at the stitch boundary with a 0.25µm overlap.

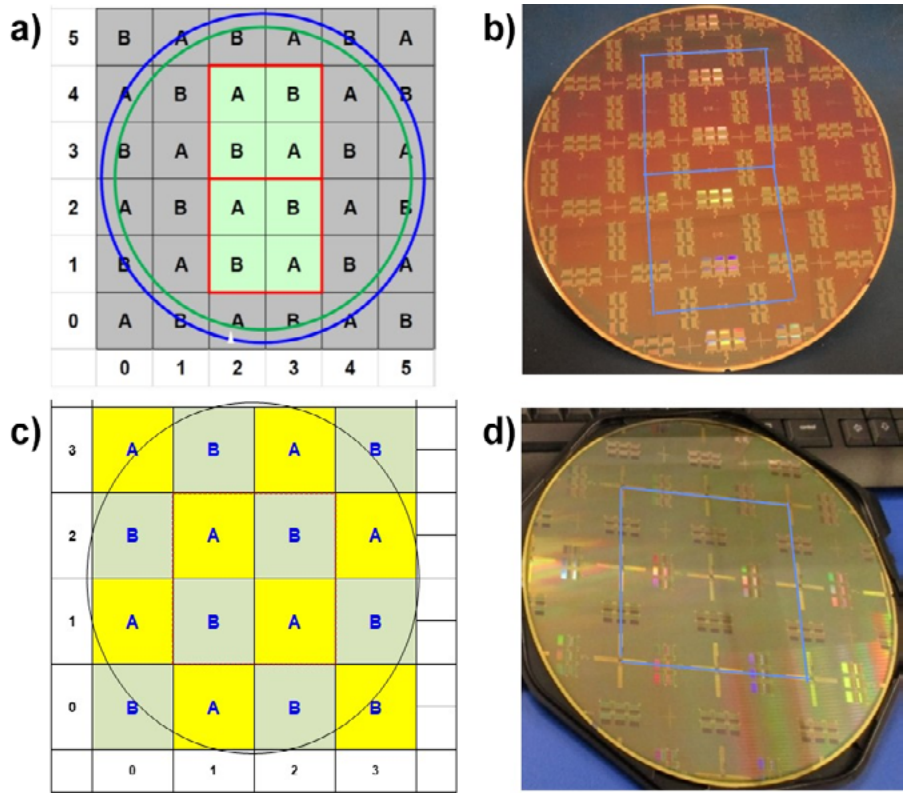


Figure 3: A stitched S-MCM wafer before dicing. Each design targeted a different reticle size for stitching: a) (top left) Schematic of a 35mm X 35mm I-line reticle stitching approach to create two 70mm X 70mm stitched S-MCMs; b) (top right) A corresponding image of two 70mm x 70mm stitched S-MCMs on a 200mm wafer; c) (bottom left) Schematic of a 48mm x 48mm I-line reticle stitching approach to create a single 96mm x 96mm stitched S-MCM; and d) (bottom right) Corresponding image of a 96mm x 96mm stitched S-MCM on a 200mm wafer.

length at the stitch boundary for a 0.25µm overlap. Electrical simulation shows that these kind of linewidth variations at the stitch boundary are electrically small and have little impedance effect below 1THz [3].

We have designed, fabricated, and tested two stitched S-MCMs. **Figure 3** shows examples of stitched S-MCM based on 35mm X 35mm and 48mm X 48mm reticles. A stitched design includes a variety of impedance controlled lines, resonators, snake/combs, and via chains going back and forth between one reticle to the other and measured linewidth distortion at the stitch boundary. For example, a variety of snake/comb lines with linewidth/space ranging from 0.8µm/1µm to 2µm/2µm that travel back and forth through the stitch boundary were evaluated and compared to reference structures that were located away from the stitch boundary and did not see the stitching process.

Wafer-scale room-temperature electrical testing was used to evaluate reticle stitching. Each S-MCM wafer has a total of 384 test structures and each structure consists of snake/combs with LW/LS in the range of 0.8µm/1µm and larger. Out of 384 test structures, 96 test structures are stitched. As desired,

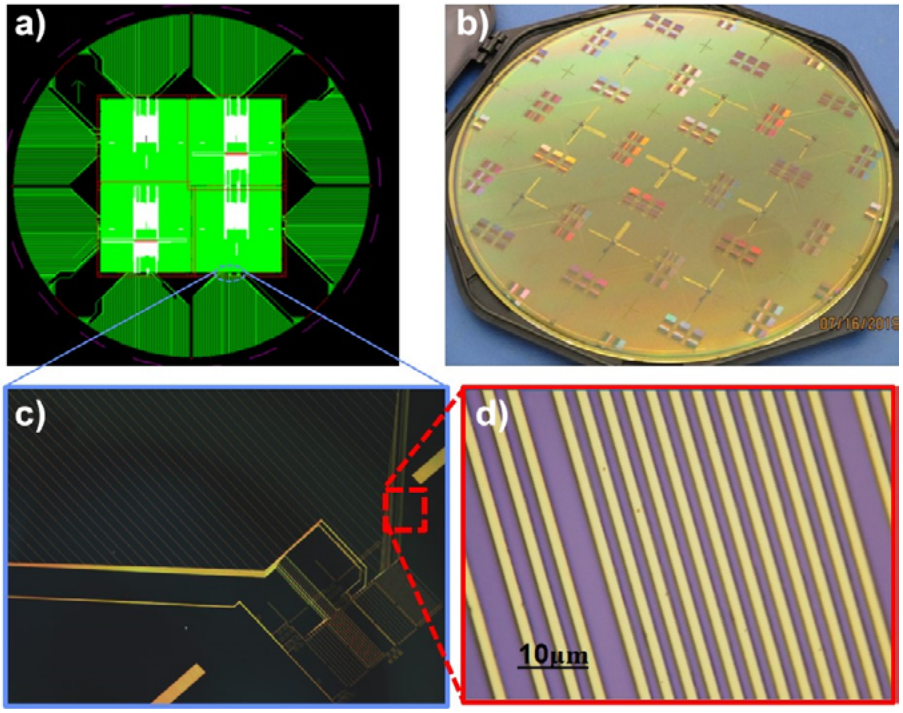


Figure 4: A 200mm wafer-scale S-MCM fabrication approach with a combination of laser direct write and I-line photolithography: a) GDS layers to create wafer-scale S-MCM; b) A corresponding image of wafer-scale S-MCM; and c-d) Enlarged optical micrographs of laser direct write based fan-out circuits.

the test structure shows no abnormal variation after stitching. Furthermore, the I-V characteristics of the snake/combs structures show wafer-to-wafer consistency. The measured RT resistance of stitched niobium snake/combs are similar to the non-stitched niobium snake/combs.

Wafer-scale S-MCM

Traditional wafer-scale MCM fabrication approaches use laser direct write or contact photolithography. Both approaches produce wider lines than the I-line process and require thicker dielectric in order to achieve the desired 50Ω impedance. Our approach uses reticle stitching for critical layers to achieve a 50Ω impedance without changing the dielectric thickness and laser direct write provides additional wafer-scale fan-outs. Figure 3 shows that stitching increases S-MCM sizes and decreases the total number of S-MCMs per wafer. For example, 35mmX35mm and 48mmX48mm reticles will produce 16 S-MCMs and 6 S-MCMs per 200mm wafer, respectively. Whereas stitching of 35mmX35mm and 48mmX48mm reticles will produce two S-MCMs and one S-MCM per 200mm wafer, respectively. Wafer real estate usage from a 35mmX35mm MCM to a 96mmX96mm MCM was reduced from 62.3% to 29.3%, respectively, and remaining area of the wafer will be diced out.

We have fabricated 200mm wafer-scale S-MCMs by combining I-line and laser direct write methods. The process uses stitching of multiple I-line photomasks to produce critical design layers with minimum feature sizes around 0.8µm. For example, Figure 4 uses a 48mmX48mm reticle to create 96mmX96mm large stitched layers. Subsequently, applied laser direct write lithography (LDW) for connecting reticles with wider (>1µm) lines [3], adding fan-out circuits and extending circuit features to the entire wafer real estate. This kind of wafer-scale S-MCM with a wafer-interposer-wafer (WIW) configuration [4] offers many advantages. The wafer-scale S-MCM eliminates the need for substrates, printed

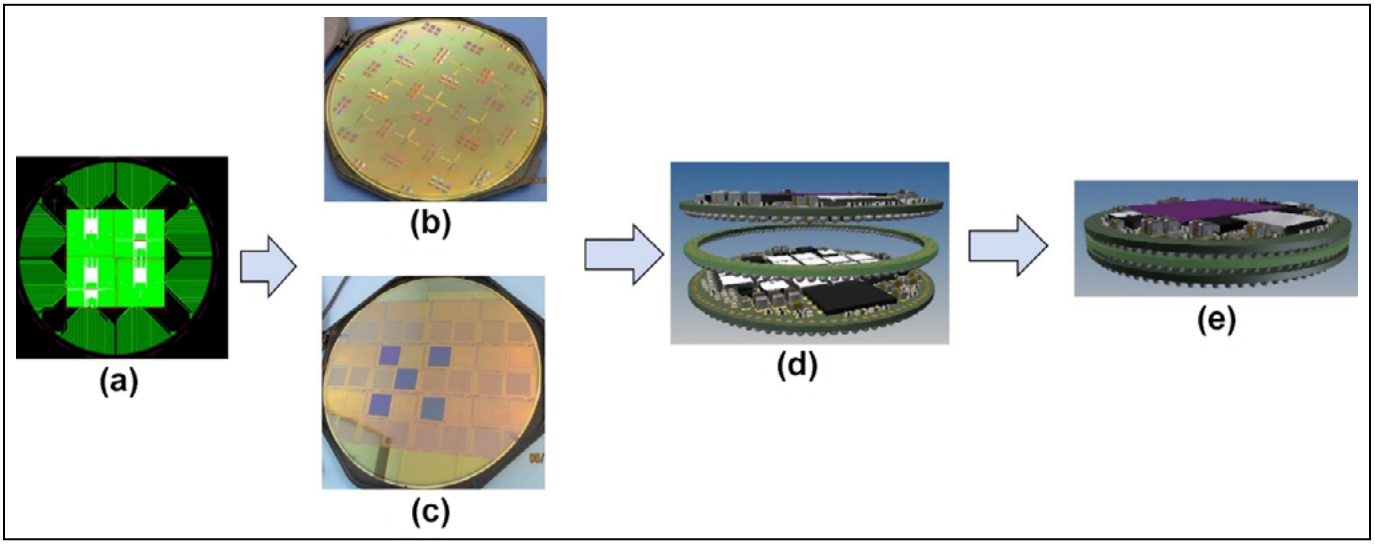


Figure 5: A wafer-scale assembly process with a superconducting multichip module (S-MCM): a) I-line and LDW GDS layers for wafer-scale S-MCM; b) A corresponding wafer-scale S-MCM; c) Flip-chip wafer-scale bonded S-MCM with 20mm x 20mm superconducting test chip; and d) Schematic of wafer-to-wafer assembly options; (e) Schematic of a future 3D wafer-interposer-wafer (WIW) integration strategy. The superconducting chip, interposer, and wafer-scale S-MCMs are fabricated separately and joined together using flip-chip connection.

circuit boards (PCBs) and associated assembly processing. In this approach, the superconducting test chip joined using niobium-indium or indium μ-bumps that provide both electrical connectivity and mechanical stability between the different chips. The combination of I-line and LDW reduces the total number of masks. The use of single photoresist for UBM and micro-bump fabrication not only reduces additional processing steps, but also minimize/eliminate bump-related defects suitable for wafer-scale S-MCM fabrication. Furthermore, LDW lithography utilizes full-wafer real estate and provides fan-out circuits for attaching connectors and cables to connect with the next level of semiconducting electronics. Wafer-scale S-MCM circuits will a create a loss-less, superconducting path for chip-to-chip communication, whereas the traditional approach will go through a multilevel of assemblies with many normal metal interrupted superconducting paths.

The niobium-indium microbump thermocompression bonding approach can be extended to attach multiple 20mm x 20mm chips to a single, large superconducting MCM wafer, which is a valuable capability as we consider scaling to larger systems, such as wafer-scale S-MCM demonstrations. The use of wafer-scale S-MCM technology for the base superconducting module enables a combination of active and passive interconnect-based circuits; such an S-MCM could be used both for superconducting chips (single-flux-quantum [SFQ], rapid single-flux-quantum [RSFQ], quantum flux parametron [QFP]) in proximity to one another, along with auxiliary semiconductor electronics (e.g., power supplies, clock generators, output amplifiers) in a single system.

As illustrated in **Figure 5**, the multiple wafer-to-wafer assembly approach creates a system that can accommodate a large number of superconducting chips. In **Figures 5d** and **e**, we also show a schematic view of how these wafer-scale S-MCM signals will be routed through a high-density interposer to produce wafer-interposer-wafer (WIW) configurations. In general, three types of signals are envisioned: DC, 50 Ohm characteristic impedance for clock line, and 10-20 Ohm characteristic impedance for the data line. As we look further out, we also consider a superconducting PCB to assemble a WIW structure for various superconducting cryogenic packaging solutions.

Summary

The demand for superconducting computing scalability beyond arrays of a few superconducting chips is driving the need for greater wiring densities and more functionality onto a single cryogenic package. One way to address this demand is with the use of flip-chip integration of wafer-scale S-MCMs with microbump-based interconnects to electrically interface multiple superconducting chips. Niobium-indium microbump-based flip-chip interconnects are capable of maintaining very low interconnect resistance, which are advantageous for superconducting packaging.

We have demonstrated large S-MCM (48 x 48mm^2) fabrication with single I-line mask exposure and thermocompression bonding to attach up to 20mm x 20mm superconducting chips to the S-MCM module. We have shown that the sequential exposure of two photomasks (e.g., A and B,) with small overlap (stitched) can produce larger S-MCM substrates up to 96 x 96mm^2 with four masks/layer combined. I-line and laser direct write photolithography combinations demonstrated full-wafer S-MCM fabrication capability. These combined features show the potential to scale S-MCM technology with high-density, low-resistance interconnects to wafer-size substrates, thereby enabling system integration for next-generation superconducting VLSI systems.

Acknowledgments

We gratefully acknowledge M.A. Gouker, S.K. Tolpygo, R. Lambert, R. D'Onofrio, C. Stark, and P. Murphy for useful discussions, and K. Magoon, P. Baldo, T. Weir, M. Townsend, M. Augeri, J. Liddell, B. Osadchy, M. Hellstrom, C. Thoummaraj, and J. Wilson for valuable technical assistance. This research was funded by the Defense Advanced Research Projects Agency (DARPA). Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

References

1. D. S. Holmes, A. L. Ripple, M. M. Manheimer, "Energy efficient superconducting computing-power budgets and requirements," IEEE Trans. on Applied Superconductivity, 23(3)1701610 (2013).
2. R. N. Das, V. Bolkhovsky, S. K. Tolpygo, P. Gouker, L. M. Johnson, E. A. Dauler, M. A. Gouker, "Large-scale cryogenic integration approach for superconducting high-performance computing," IEEE Electronic Comp. & Tech. Conf. (ECTC) Proc., 675-683 (2017).
3. R. N. Das, V. Bolkhovsky, C. Galbraith, D. Oates, J. J. Plant, R. D. Lambert, et al.,"Interconnect scheme for die-to-die and die-to-wafer-level heterogeneous integration for high-performance computing," IEEE ECTC Proc., 1611-1621 (2019).
4. R.N. Das, F. D. Egitto, J. Lauffer, B. Bonitz, B. Wilson, M. D. Poliks, et al., "3D-interconnect approach for high end electronics," IEEE ECTC Proc., 1333-1339 (2012).

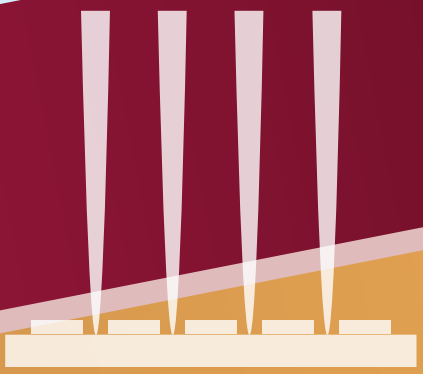
Biographies

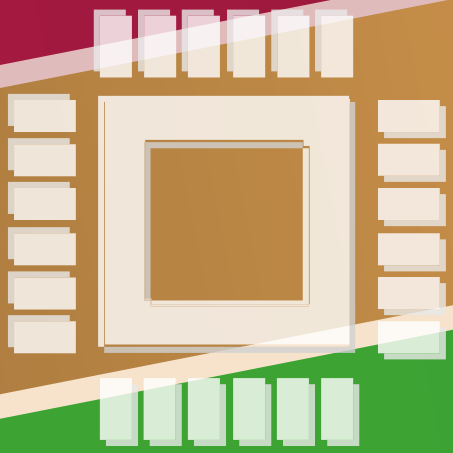
Rabindra N. Das is a Member of the Technical Staff in the Quantum Information and Integrated Nanosystems Group, MIT Lincoln Laboratory, Lexington, MA USA. Prior to MIT, he was a Principal Engineer at Endicott Interconnect Technologies (formerly IBM Endicott). Dr. Das has 18 years of experience in microelectronics packaging development for applications ranging from HPC to medical to quantum electronics. He holds 48 patents and more than 100 publications. Email Rabindra.Das@ll.mit.edu

Vladimir Bolkhovsky is a Member of the Technical Staff in the Quantum Information and Integrated Nanosystems Group, MIT Lincoln Laboratory, Lexington, MA USA. He has more than 40 years of experience in the semiconductor industry and in superconductor process development. He received an MS in Chemical Engineering and holds multiple patents and has more than 20 publications.

international TEST SOLUTIONS

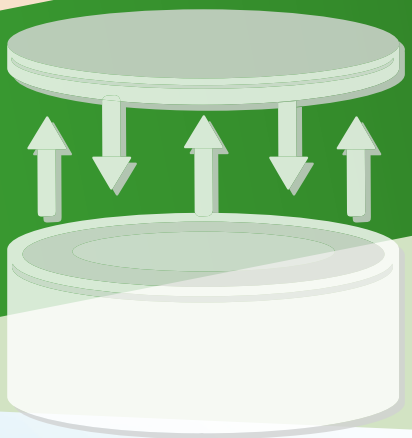
PROBE
CARD
CLEANING





TEST
CONTACTOR
CLEANING

CHUCK
CLEANING
WAFER



ENGINEERED CLEANING SOLUTIONS
FOR THE SEMICONDUCTOR INDUSTRY:

Improving yield, reducing cost.

www.inttest.net

600mm wafer-level fan-out on panel-level processing with 6-sided die protection

By Jacinta Aman Lim, YunMook Park, Byung Cheol Kim, Edil Devera [nepes]

Currently, a 300mm carrier wafer for fan-out wafer-level packaging (FOWLP) is the mainstream format used for power management integrated circuits (PMICs), radio frequency (RF) and other single-die applications. As the volume of these devices continues to rise, the need for migration to panel sizes larger than 300mm becomes a necessity for cost reduction and capacity.

The fastest adoption of fan-out technology is now in 5G, automotive and healthcare. Traditional applications such as audio codecs, PMICs, micro-controller units (MCU) and RF continue to use FOWLP as an alternative to wafer-level chip-scale packaging (WLCSP) due to its 5-sided or 6-sided die protection. As fan-out packaging becomes mainstream and in order to get broader adoption of it, the need for driving down the cost continues to be at the forefront of fan-out suppliers.

The 600mm x 600mm format utilized in this study leverages existing backend processing equipment used on 200mm and 300mm wafers for cost savings. Utilizing existing equipment enabled the panel to be singulated into 4x300mm or 9x200mm square segments to enable probe testing.

Coupling the 6-sided die protection process M-Series™ with 600mm x 600mm panel-level processing paves the way for innovative methods for fan-out processing. New photolithography processing that utilizes laser direct imaging (Adaptive Patterning™) to auto scale for die-shift mitigation is heavily dependent on segmentation of the panel. In this instance, the 600mm panel is either segmented into 4x300mm, 9x200mm or 1x600mm for the photolithography steps. Depending on the number of fiducials used during the photolithography steps, capital expenditure and exposure accuracy would be highly dependent on the segmentation chosen. In addition, new metrology tools and panel warpage management will need to be considered for quality assurance.

This paper will present a case study of utilizing 600mm x 600mm panel size to process a single die with 6-sided die protection. Considerations for repassivation, redistribution layer and solder ball placement will be discussed. Challenges pertaining to large panel processing through the repassivation and redistribution layer will be presented; additionally, panel-level inspection considerations post mold cure, reliability considerations, and the future of 600mm x 600mm panel-level processing for 6-sided die protection will be summarized.

Introduction to wafer-level packaging

There are several types of wafer-level packaging (WLP) in the industry. The mainstream is WLCSP, or fan-in, followed by variations of FOWLP. In the example shown in **Table 1**, we look at two fan-out processes in comparison with WLCSP. In comparison with processing complexity, the left-most column (M-Series™) ranks highest, with WLCSP being the least complex. In terms of package reliability, 6-sided die protection ranks highest compared to WLCSP as shown in the Weibull plot (**Figure 1**). In the plot, a 6mm x 6mm WLCSP test chip, which was also used on a 6.25mm x 6.25mm fan-out package, was tested with board-level temperature cycling (TC) on a 1mm board, with SAC405. The bump pitch was at 500µm with no under bump metallization (UBM). There were no TC failures up to 1000 cycles. The plot in **Figure 1** shows >200% improvement for TC over WLCSP [1].

Fan-out packaging drivers

As can be seen from **Figure 2**, some of the major market drivers are RF, audio codec, PMIC, radar, 5G and high-speed computing [2].

5G adoption is central to large data transfers to enable a new user experience, and enabling better cloud-based business management as well as increasing our level of interaction with one another. In comparison to flip-chip ball grid array (FCBGA) packages, the interconnect length between the integrated circuit (IC) and the antenna is shorter, thereby reducing the signal loss from the radio frequency (RF) chip in fan-out packaging. Leading players such as TSMC and Mediatek have shown that fan-out packaging can enable low transmission loss and high antenna performance for mmWave system integration [2].

As for high-performance computing (HPC), a surge in the number of Internet

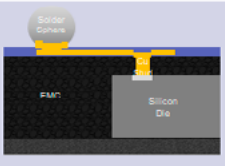
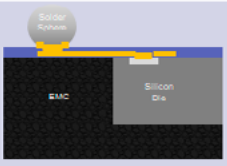
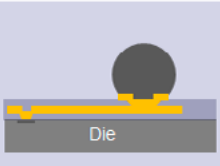
	M-Series	eWLB	WLCSP
Package Structure			
Features	• Die first, face up • Cu stud + Fully encapsulated PKG • Adaptive Patterning™ > Die Drift • Backside epoxy	• Die first, face down • 5-sided protection with MC	• Package Size = Die Size • Pitch of interconnect is limited by die size • Backside laminate option is available
Process	• Cu stud needed • Create GDS file per panel • Backside epoxy	• Warpage adjust for critical process step	• Standard thin film processing
Die Drift Control	• Low Accuracy Chip Attach Process & Adaptive Patterning™ • Pitch Compensation	• High Accuracy Pick and Place Process • Pitch Compensation	• None
Warpage Control	• Use of Backside epoxy	• Warpage Adjust for critical process steps	• None, die picked from Film Frame

Table 1: Attributes between FOWLP and WLCSP.

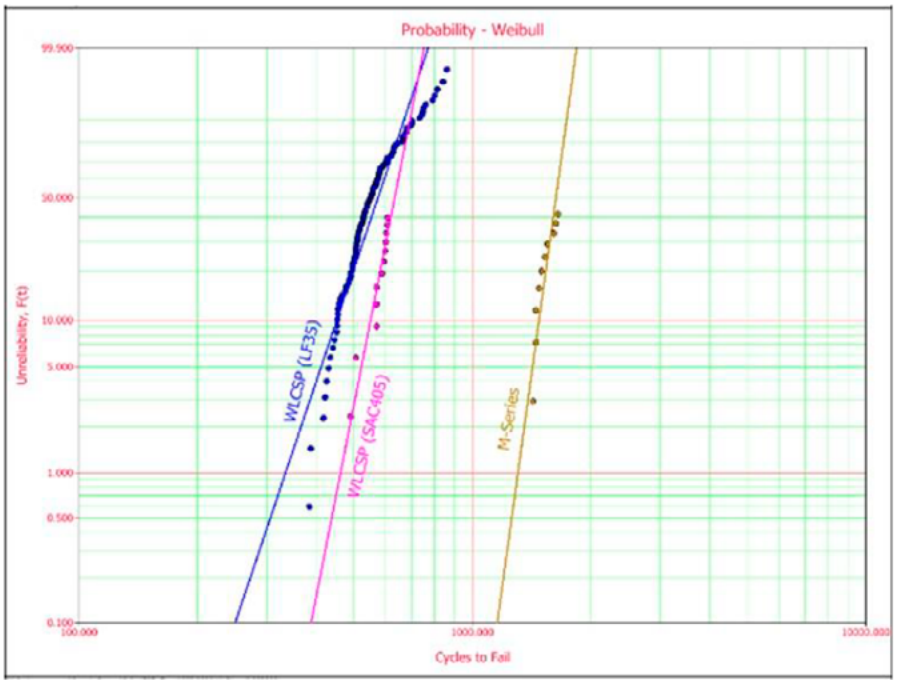


Figure 1: Weibull plot for WLCSP vs. M-Series™. SOURCE: DECA

of Things (IoT) devices is boosting the demands for HPC systems and data centers. Higher functionality coupled with smaller package sizes are main drivers for fan-out system in package (SiP). The industry is relying more on heterogeneous solutions to integrate dies with shorter and denser interconnections [2]. Based on early stage testing, high-bandwidth and high-speed SERDES signal is positively validated by fan-out packaging from many fabless players like NVIDIA, Mediatek, Nephos and HiSilicon.

Lastly, for radar at 77GHz, advanced driver assistance systems (ADAS) are paving the way for full autonomy in the automotive industry [2,4]. A 77GHz automotive radar system offers key advantages over a 24GHz radar and light detection and ranging (LiDAR) because of its robustness in detection, range resolution and simultaneous multiple-depth detection under severe environmental conditions. Fan-out packaging is already established in radar and 77GHz performance is proven to be better than other packaging platforms such as flip-chip ball grid array (FCBGA). This improvement is because of the routing of redistribution layers (RDLs) from thin-flim processing in FOWLP, thereby resulting in very small layer dimensions and tolerances. Consequently, the fine line/space from RDL processing enables low-loss wiring and superior RF performance [2,4].

Production volume of fan-out packaging

In terms of production volume, both FOWLP and FOPLP are growing, though FOPLP will grow significantly faster from 2021 and beyond (**Figure 3**). We do note that the COVID-19 pandemic (Coronavirus) has caused a contraction of the overall economic activity, negatively impacting semiconductor development and production as seen in 2019 and 2020 [2,5]. FOWLP revenue will continue to experience some decline in 2020, reflecting the COVID-19 impact for mobile and consumer applications.

Data-driven end systems, however, continue to grow in delivering more data functionalities paving the way to 5G adoption. FOWLP revenue for

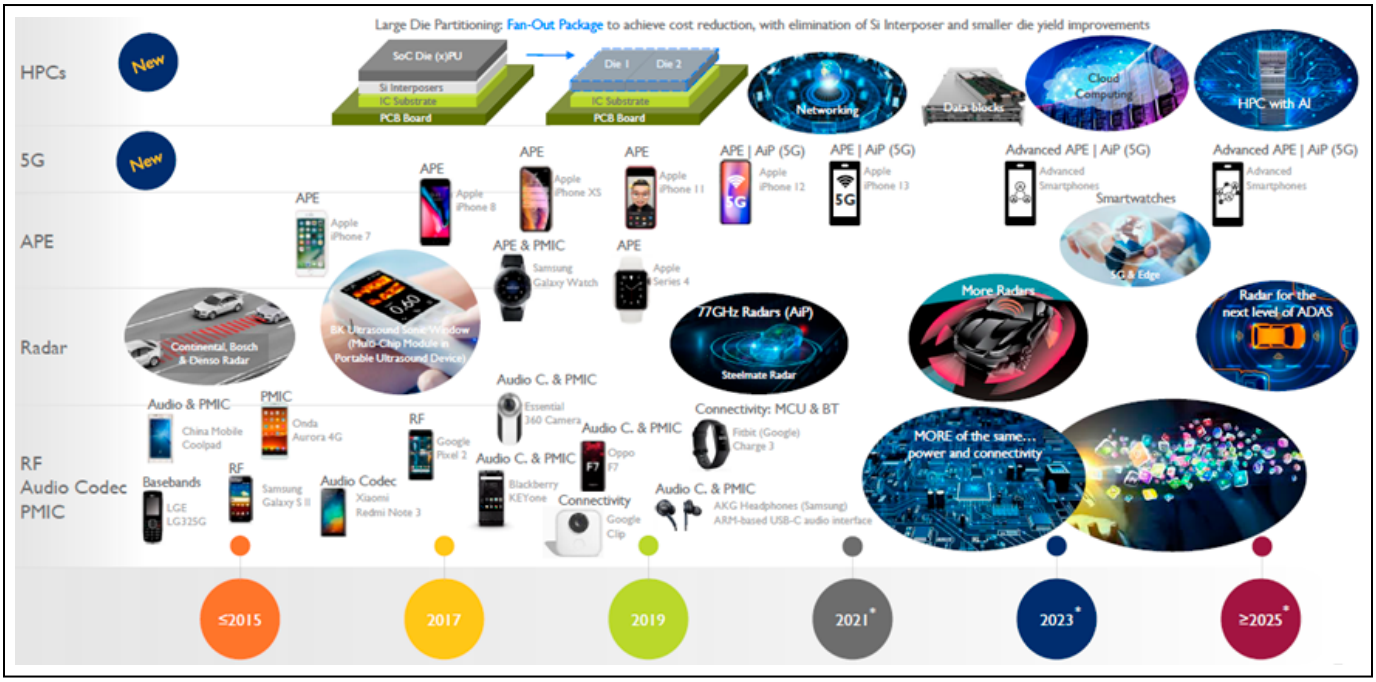


Figure 2: Fan-out packaging drivers. SOURCE: YDR20078 Fan-out Packaging 2020 Yole Report, Yole Développement

LEENO

Total Interface Solutions

Since 1978

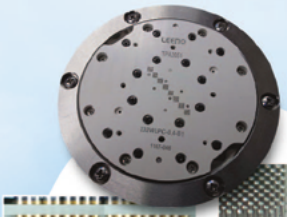
GLOBAL LEADER

www.leeno.com

Spring Contact Probe

90um Pitch ~

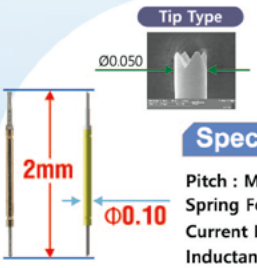
Probe Head



90um Pitch~

RF Probe for Fine Pitch Probe Head

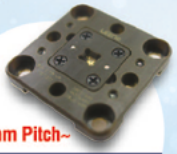
Tip Type



Specification

Pitch : Min.120um
Spring Force : 8.0g@250um
Current Rating : 1.3A
Inductance : 0.3nH

RF Coaxial Spring Probe & Impedance Controlled Socket



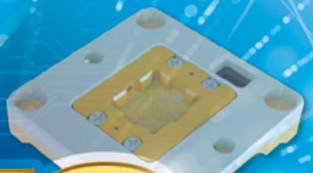
0.18mm Pitch~

Logic Test Socket

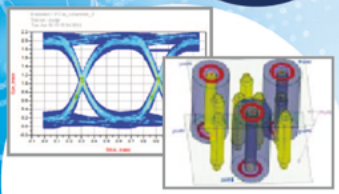
High Speed Automatic Coaxial Probe

Frequency : >20GHz
VSWR : < 1.2

5G

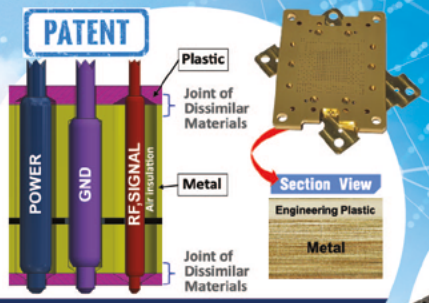


Electrical Analysis

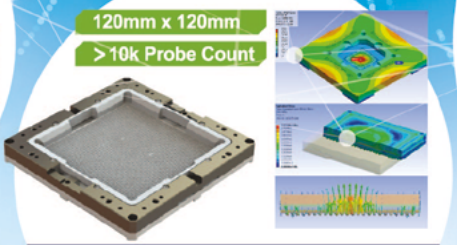


CCC Test, HFSS, TDR
Eye Diagram
4Port VNA Test

MP Socket



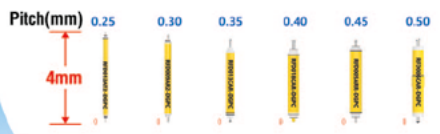
Large Device Socket



Coaxial Probe for 100GHz

Specification

Frequency : 80GHz(BGA),
100GHz(QFN, LGA)
Pitch : 0.25mm~
Crosstalk : -60dB
Impedance : 50Ω±10%



LEENO INDUSTRIAL INC.

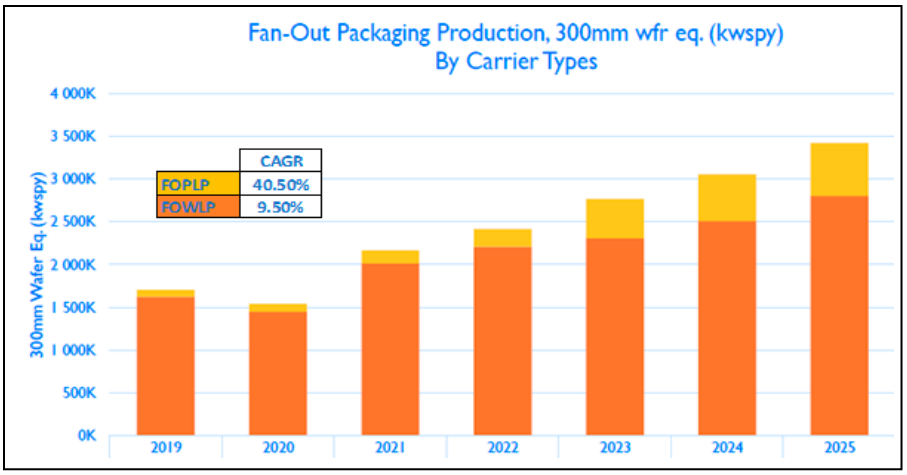


Figure 3: Production volume FOWLP vs. FOPLP. SOURCE: YDR20078 Fan-out Packaging 2020 Yole Report, Yole Développement

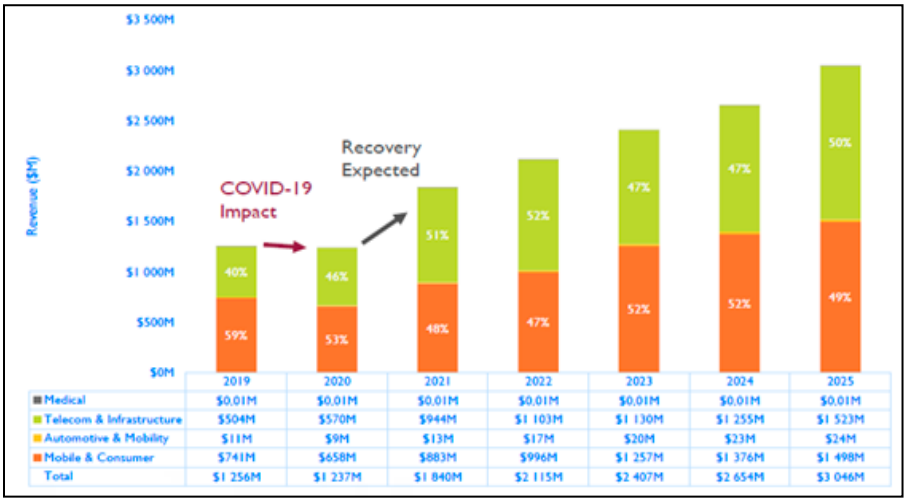


Figure 4: Fan-out packaging market revenue by end market. SOURCE: YDR20078 Fan-out Packaging 2020 Yole Report, Yole Développement

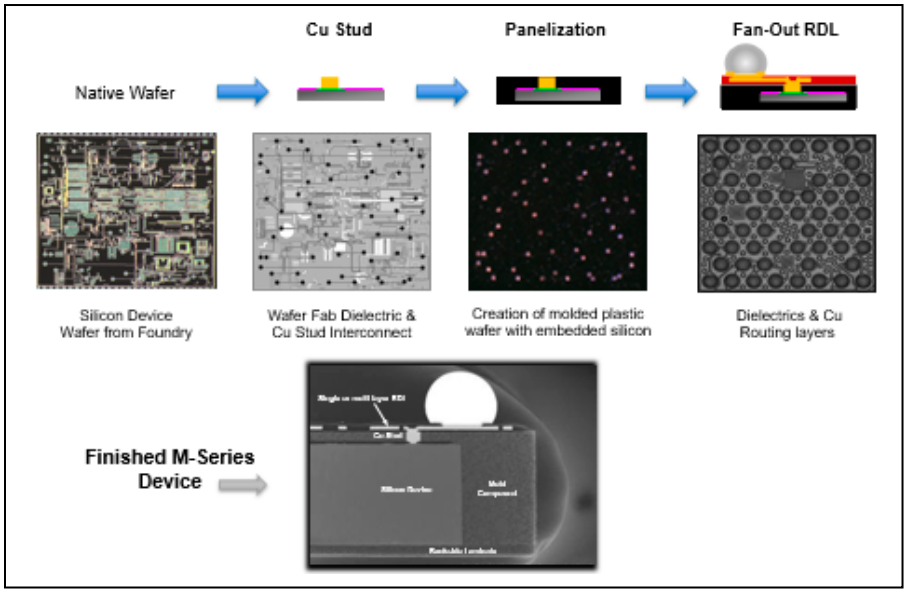


Figure 5: M-Series™ key process steps. SOURCE: DECA

telecom and infrastructure is improving moderately in 2020 and expected to recover in 2021 [2] (**Figure 4**).

Based on the positive outlook for FOWLP in general, outsourced semiconductor assembly and test suppliers (OSATS) have been gearing up to meet the customer demands for 2021 and beyond. One of the key areas in FOWLP development is expanding from the traditional 300mm round carrier to a 600mm x 600mm panel-level processing for fan-out. Specifically, we will be focusing on 6-sided die protection with mold compound material from a 300mm round carrier to 600mm x 600mm 6-sided die protection with mold compound material. We will further discuss some of the key challenges with managing die shift, panel warpage and panel design.

6-sided die protection with mold compound

The process for a molded M-Series™ device is illustrated in **Figure 5**. We begin with the incoming silicon wafer, either in 200mm or 300mm format. The wafer first undergoes a copper stud buildup process for connecting the silicon to the outside of the molded package in a subsequent process. The wafer is then singulated to prepare for the next step of panelization.

During the panelization process, the singulated die is picked and placed onto a temporary carrier face up, with the copper studs up. Once the entire panel is populated with known good die (KGD), the carrier undergoes a mold compression process. Subsequent processes such as Dielectric 1, RDL, Dielectric 2, UBM and solder ball placement are now similar to those done in WLCSP processing.

The unique portion of the photolithography process is the use of laser direct imaging, which is mask-less. In this instance, traditional reticles and masks are eliminated. The process works by dynamically adjusting a portion of the interconnect structure to accurately connect to the copper studs that are protruding through the mold compound for each individual die in the molded panel. A proprietary design tool adjusts the fan-out unit design for each package on the panel so that the first via layer and the fan-out RDL pattern are properly aligned to the pillars of the die. The design files for each panel are imported to a lithography machine that uses the design data to dynamically apply a

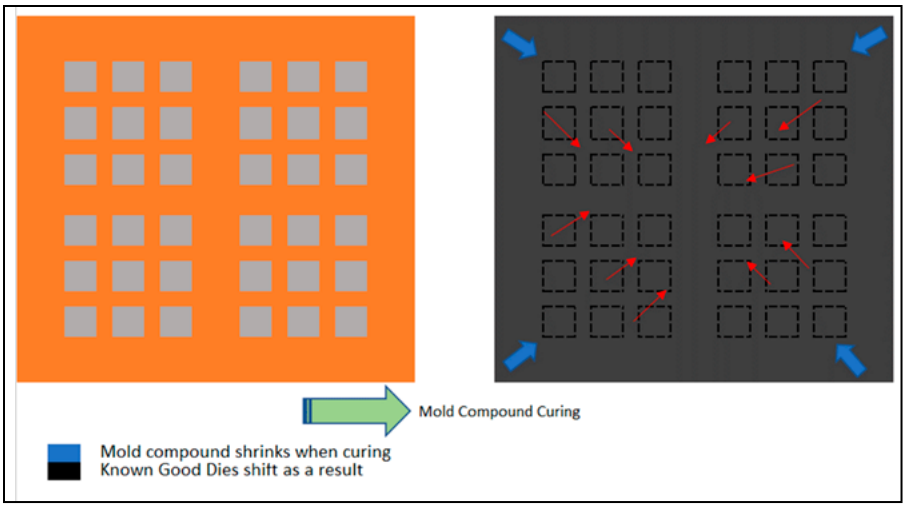


Figure 6: Die shift illustration from chip attach to post-mold cure.

custom adaptive pattern to each panel [1]. Utilizing this photolithography process helps with mitigating issues encountered during die shift (X, Y, Θ).

Key challenges in 600mm panel processing

Some of the key challenges in 600mm x 600mm PLP are die shift during the panelization process, warpage control, mold thickness variation, C-mode scanning acoustic microscopy (CSAM) for defect detection, panel segmentation and panel design to satisfy a >99.5% yield requirement. Some of these challenges are addressed in the sections below.

Panelization. In Figure 6, an example of why die shift occurs is illustrated. There are two major areas for die shift. The first will occur at the chip attach process and the second will occur during the mold compression process. During the chip attach process, singulated die may experience a shift in the x, y or Θ direction. This shift can be attributed to an incorrect recipe set-up or pick and place bond heads that are not aligned to the center of the die, or to optical misalignment.

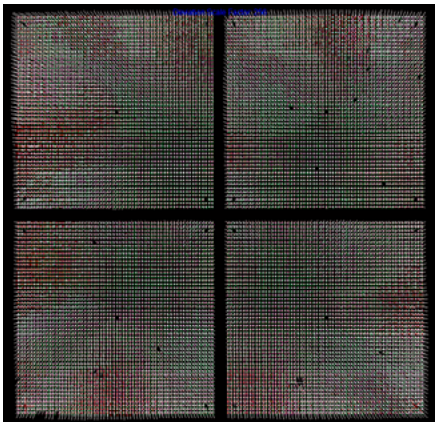


Figure 7: Die position compensation and mold dispense need to be optimized.

During the mold compression process, mold material tends to shrink and therefore tends to gravitate towards the center of the panel. Consequently, this shrinkage also affects die placement from the previous process. An example of a 600mm X 600mm panel that will require further improvements in mold dispense is shown in Figure 7. Referencing Figure 7, die shift during mold compression can be mitigated by applying a die position compensation during the chip attach process. In addition, the mold dispense pattern can be optimized to ensure better coverage and the minimization of total thickness variation (TTV) within the panel. The use of laser directed imaging to compensate for minor die shifts during RDL patterning will also contribute to overall panel yield.

CSAM inspection: mold voids detection. Mold voids can occur because of short gel time, short spiral flow length of the mold compound, contamination on the interface between silicon and mold compound material, or an un-optimized mold dispense profile and parameters, to name a few causes. In the example from Figure 8b, mold voids are normally detected with a CSAM metrology tool because a standard microscope will not be able to detect voids at the mold compound to the Si interface layer (Figure 8a). Figure 8c confirms post-Cu stud top grind that mold voids were present from CSAM inspection. In this example, the CSAM is performed prior to the Dielectric 1 process.

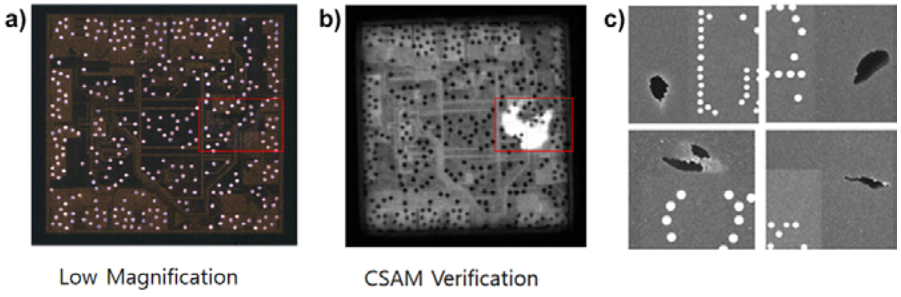


Figure 8: a) Example of CSAM defects: a) low magnification microscope; b) CSAM inspection; c) (four panels) Mold void after panel top grind post-CSAM.

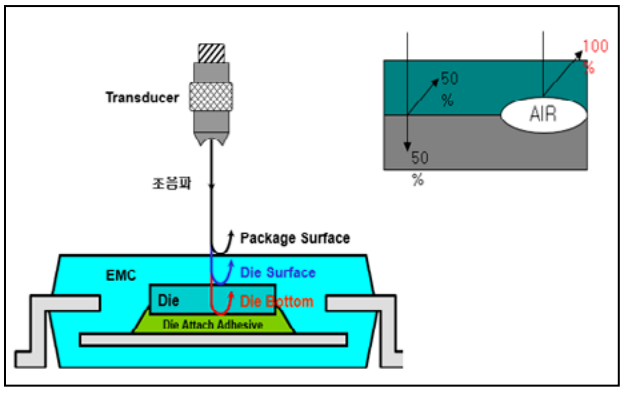


Figure 9: CSAM mechanism for checking mold voids.

In the example shown in Figure 9, the entire panel is submerged in water. Ultrasonic waves cannot pass through air pockets and therefore, the waves reflects at different interfaces or defects. For example, any air pockets or areas where the ultrasonic wave cannot pass through will be reflected 100%. This, in turn, provides us with a mapping of what the

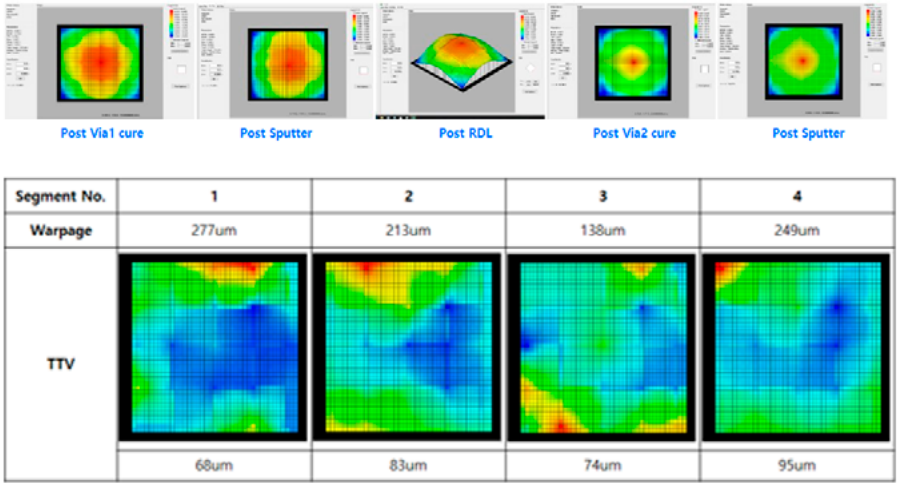


Figure 10: TTV and warpage profile results from evaluation.

Silicon BG Thickness	280um – 700um
Mold Thickness	405um – 815um
Fan Out Ratio	<1.1
TTV Spec Requirement	<50um
Warpage Spec Requirement	<1mm

Table 2: Evaluation plan for warpage and TTV for 600mm PLP.

defect area looks like and its location on the molded panel.

Panel warpage. In general, panel warpage is induced by a coefficient of thermal expansion (CTE) mismatch between the die and mold compound material. Other factors such as die size, fan-out ratio, die thickness, and overall mold thickness also play an important role in minimizing panel warpage. **Figure 10** shows a review of panel warpage and total thickness variation (TTV) for four panels that were used in an initial evaluation with parameters shown in **Table 2**. The specification requirement for this initial look-ahead build was for TTV to be less than 50µm and panel warpage to be <1mm. In the samples ran, all panels showed warpage to be <1mm with the average warpage across panels being 219.25µm. However, TTV tends to be thinner on the panel edges with an average of 80µm across all panels. In the evaluation described above, the next steps would be to optimize the mold pattern dispense to improve TTV.

Panel segmentation strategy. Panel segmentation for a 600mm x 600mm panel is another factor to consider when maximizing gross die per panel (GDPP). Some items for consideration are the fungibility of existing backend equipment, such as solder ball mount, backgrind equipment or testers/handlers. For example, one could continue utilizing 200mm or 300mm tooling for the ball mount process post-panel singulation to minimize capital expenditure. A slight modification to the panel chuck to fit the singulated panel geometry would be required, along with the ball dispense process. However, some considerations need to be taken into account for maximizing the GDPP. There is a tradeoff from segmenting the 600mm x 600mm panel into nine 200mm segments versus four 300mm segments.

Referencing **Table 3**, a 5mm x 5mm single-die package was used as a test vehicle. The total usable area loss was calculated by subtracting the sum of the package placement zone fiducial dies and panel ID from the total panel area. Option B comprises four 300mm panel segments and has a calculated usable area loss of 1X, in contrast with Option C, which comprises nine 200mm panel segments with a calculated usable area loss of 2X. As we get into larger package sizes or multi-die packaging for panel-level processing (PLP) for 6-sided die protection, the positive impact of 4 versus 9 segments on a 600mm x 600mm panel would be larger because of keep out zones and placement of the die to maximize usable panel space (**Figure 11**).

Panel design: keep out zones. Keep out zones (KOZ), often referred to in FOWLP and FOPLP, are areas on the panel where active die placement is discouraged for process manufacturability and reliability purposes (**Figure 12**). For 600mm x 600mm PLP, dummy bump

			Package placement zone (per segment)			Package placement zone (per panel)	Panel ID (per segment)	Fiducial die package area (one pkg) *	Total Fiducial die package area (per segment) **	Total useable area (per segment)	Difference of useable area compared to baseline		Useable area loss as compared to baseline
	# segments per panel	Total panel area, sq. mm	X, mm	Y, mm	Area, Sq. mm	Area, Sq. mm	Area, Sq. mm	Area, Sq. mm	Area, Sq. mm	Area, Sq. mm	Area, Sq. mm	%	%
Option A (BL)	1	360,000	589.00	589.00	346,921.00	346,921.00	155	25	125	346,641.00	NA	NA	NA
Option B	4	360,000	288.85	288.85	83,434.32	333,737.29	620	25	500	332,617.29	14,023.71	95.95%	4.50%
Option C	9	360,000	188.96	188.96	35,705.88	321,352.93	1395	25	1125	318,832.93	27,808.07	91.98%	8.05%

* Assume 5x5 mm fiducial die package size

** Assume 5 fiducial die packages per segment

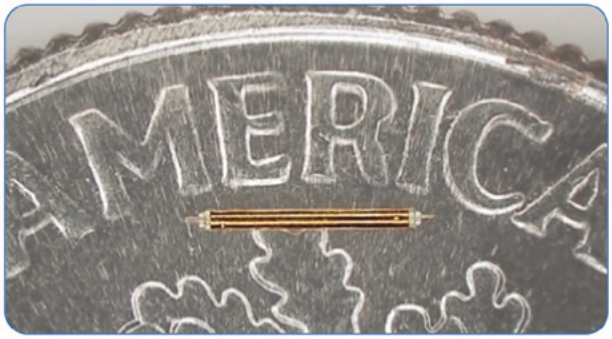
Table 3: 9-200mm segment vs. 4-300mm segment Usable Area Comparison. SOURCE: DECA

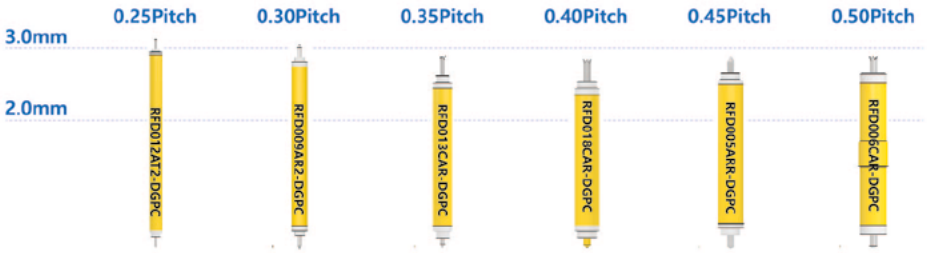
LEENO

Coaxial Probe for 100GHz

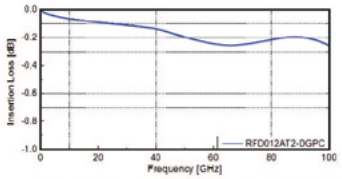
Specification

Frequency : 80GHz(BGA),
100GHz(QFN, LGA)
Pitch : 0.25mm ~
Crosstalk : -60dB
Impedance : 50Ω±10%

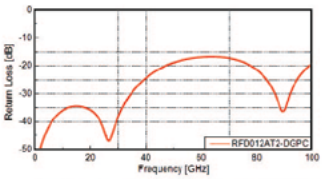




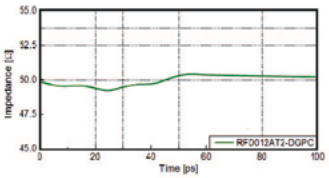
RFD012AT-DGPC(0.25P) S-Parameter & Impedance



Insertion Loss(-1dB): Up to 100GHz

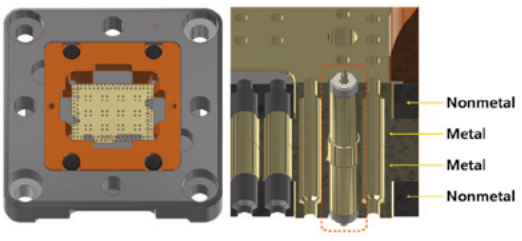


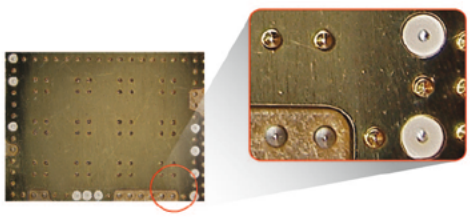
Return Loss(-10dB): Up to 100GHz



Impedance:49.2Ω

Coaxial Probe Socket Structure





For More Details, Please Contact Us.
We are Always Ready to Visit You.

GLOBAL LEADER LEENO

HEAD OFFICE
10 105beon-gil MieumSandan-ro Gangseo-gu, Busan, Korea

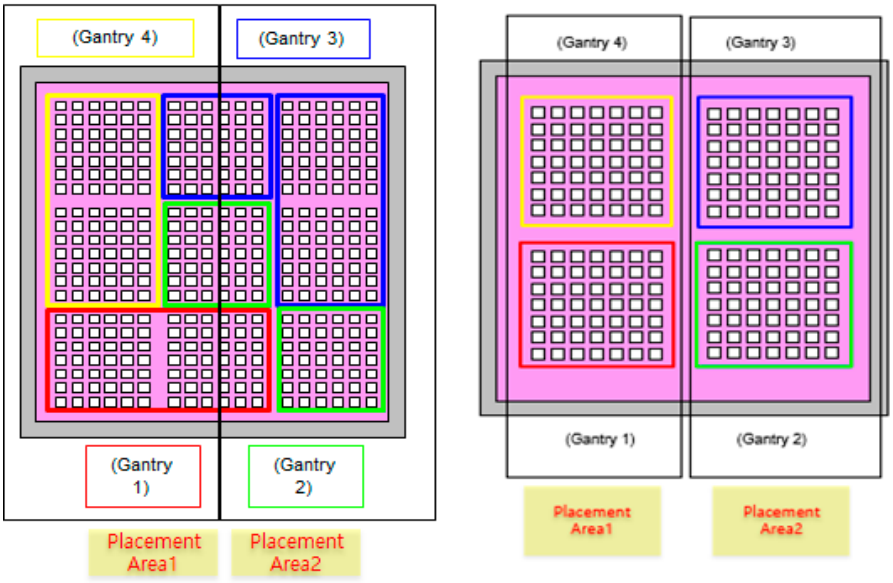


Figure 11: 9 segments vs. 4 segments for chip attach on a 600mm x 600mm panel for 6-sided die protection. SOURCE: DECA

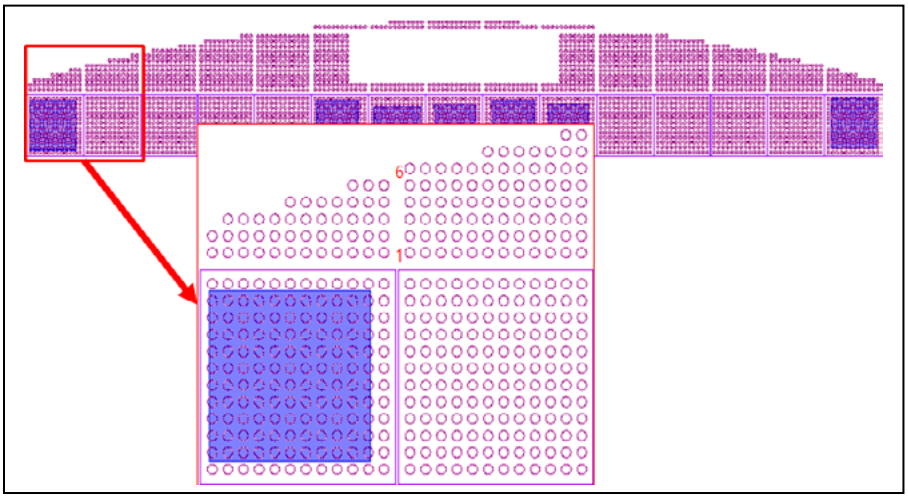


Figure 12: Example of a keep-out zone on a 300mm round pan, for M-Series™. SOURCES: DECA, nepes.

placement will need to be evaluated during panel design to ensure that panel processing in downstream processes like backgrind and singulation are robust. A decrease in usable area would negatively impact GDPP.

Summary

FOWLP revenue continues to experience a decline in 2020, reflecting COVID-19 impact for mobile and consumer applications. However, data-driven end systems continue to grow in delivering more data functionalities, thereby paving the way into 5G adoption. Pent-up demand is expected to return in 2021 as more technology-related business leaders are increasingly optimistic that businesses and consumers will return to a new normal. FOPLP is projected to gain 13% in volume from 5% in 2019, to 18% in 2025. At a 15% compound annual growth rate (CAGR), fan-out packaging is expected to be valued at $3.05B by 2025, up from $1.25B in 2019 [2,5].

Some of the challenges encountered on 600mm x 600mm FOPLP were discussed. We reviewed the impact of die shift in the x, y, and Θ directions in the chip attach process, and the use of laser-directed imaging to eliminate routing issues caused by die shift for RDL. We also reviewed CSAM metrology used for detecting mold defects on panel and die level process optimizations to meet warpage and TTV specifications, panel segmentation strategy, and keep out zones to maximize usable area and to ensure panel robustness at downstream processes.

FOPLP will continue to gain traction and appeals to high-volume applications where a key application will be the enabling of large die partitioning for HPC applications.

References

1. B. Rogers, “Implementation of a fully-molded fan-out packaging technology,” Deca Technologies, IWLPC, Nov. 2013.
2. F. Shoo, “2020 Fan-Out Packaging Market and Trends,” Yole Développement.
3. J. Aman Lim, “eWLB fan-out as one solution for 2.5D SiP,” IWLPC, Oct. 2017.
4. F. Wang, F. An, “Advanced eWLB solutions for mmWave applications,” IWLPC, Oct. 2018.
5. L. Wood, “Impact of COVID-19 on the market,” Business Wire, Aug. 12, 2020.

Biographies

Jacinta Aman Lim is the Technical Director of Business Development for Wafer Level Packaging at nepes, San Diego, CA USA. She has over 18 years of experience in semiconductor manufacturing, WLP, NPI and R&D. She graduated with MSE (Material Science), MBA and BSEE (Electrical Engineering) degrees from Arizona State U., and Business Analytics from The Wharton School of the U. of Pennsylvania. Email Jacinta.AmanLim@nepes.co.kr

Yun Mook Park is the Process Engineering Manager for the 600mm panel-level processing line at nepes, South Korea. He has over 18 years of experience in semiconductor manufacturing in wafer-level packaging, process engineering and R&D. He graduated from Young Nam U. in South Korea with a Bachelor of Science in Electronics Engineering.

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

Extreme Si thinning and nano-TSVs to advance 3D heterogeneous integration

By Dave Thomas [SPTS Technologies] Anne Jourdain [imec]

It is anticipated that most of the 3D system on chip (SoC) integration schemes for the future will require wafer-to-wafer (W2W) bonding, which directly joins the back-end-of-line (BEOL) layers of two wafers, in combination with via-last through-silicon via (TSV) connections. This article presents a new approach that combines extreme thinning of silicon to a final thickness of 500nm, together with subsequent etching of nano-scale TSVs at sub-500nm pitch. This allows for very high-density electrical connections between the back side and front side of a device wafer as part of the back-side power delivery network (BSPDN) integration.

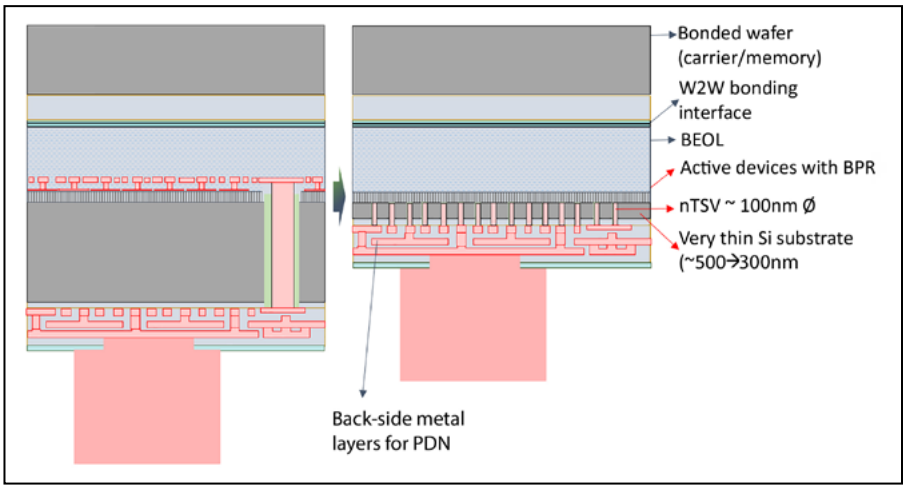


Figure 1: Typical device structures with micro- and nano-TSVs. The latter requires extreme thinning to a final Si thickness of ~500nm.

Introduction

Via-last TSVs are typically used to connect the back side to the front side of a device wafer through several microns of thinned Si. In most cases the micro-TSV dimensions are ~1µm diameter and 5µm deep [1]. For sub-micron interconnect pitches the thickness of the Si needs to be correspondingly reduced in order to preserve the TSV aspect ratio below 10:1—a typical maximum for subsequent oxide and metal deposition. Ideally, nano-TSVs with 180 x 250nm critical dimensions (CD) require the Si to be thinner than 1µm.

Figure 1 shows the device structures for micro- and nano-TSVs. The latter allows the movement of all power delivery lines to the back side of the wafer and the thinner Si also improves heat dissipation.

Silicon thinning

Thinning to 5µm can be achieved through a combination of grinding, chemical mechanical polishing (CMP), and dry etching to reach the target average thickness. Because of the non-uniformities associated with these processes the final total thickness variation (TTV) is typically 2µm [2]. The dry etching step, carried out in an SPTS Rapier™ XE system, removes 44µm of Si (to leave 5µm). However, even with exceptional control over the etching uniformity at (say) ±2%, this still leads to an induced TTV of ~1.8µm. These process combinations, therefore, cannot address the challenge of thinning consistently to ~0.5µm.

In order to achieve the desired final thickness, a combination of dry and wet etching is used along with an embedded etch-stop layer [3]. A 50nm-thick, high-quality, epitaxial layer of SiGe25% is first

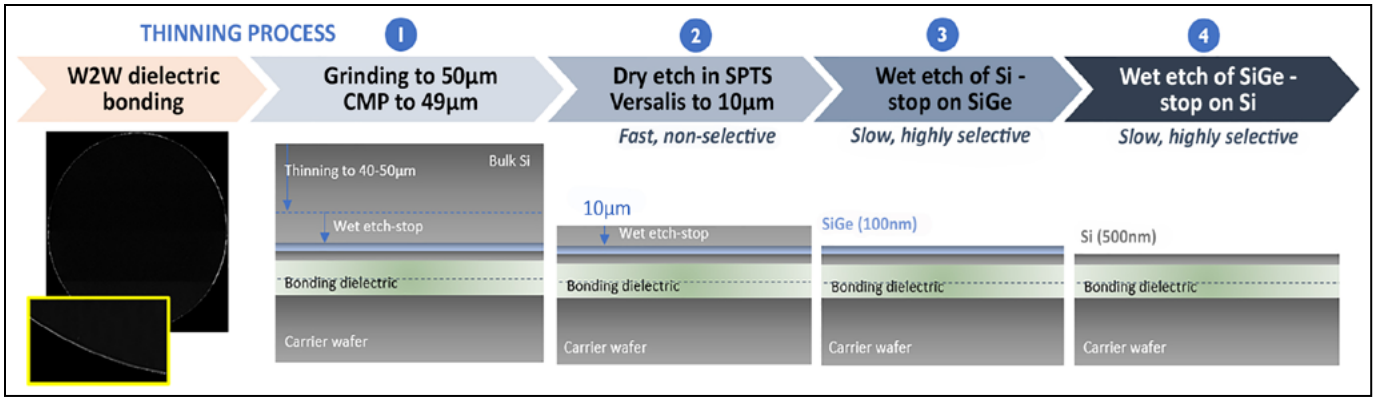


Figure 2: Post-W2W bonding process flow for extreme thinning to 500nm.

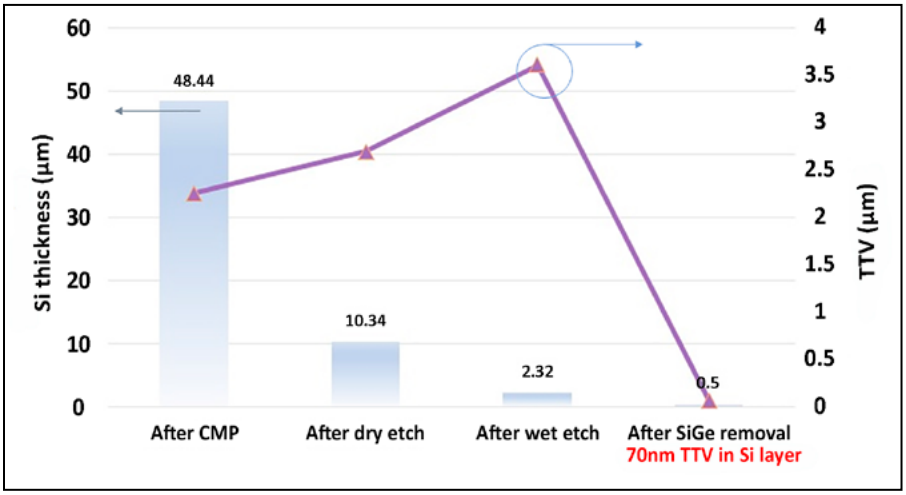


Figure 3: Si thickness and TTV of the device wafer across the thinning process steps.

grown on a 300mm Si wafer and a 500nm capping layer of epitaxial Si is grown on top. This Si capping layer is the starting point for device fabrication. The SiGe content, being SiGe25%, is selected as a compromise between wet etch selectivity to Si (being higher for higher Ge concentration) and film defectivity (being lower for lower Ge concentration because of lattice mismatch). Following the above substrate preparation, the next step is the fabrication of front-end-of-line (FEOL) devices in the 500nm thick Si capping layer. The top (device) wafer and a carrier wafer are then face-to-face dielectric bonded using a combination of 150nm SiO and 50nm of SiCN on both faces.

After wafer-to-wafer (W2W) bonding, the next part of the process flow is shown in **Figure 2**. The top wafer is thinned by grinding to 50µm. Grinding is mechanically damaging to the Si surface and sub-surface, therefore the remaining Si thickness should be larger than the affected Si thickness, as cracks and disclocations must be prevented from reaching the active device areas, because they would impact device performance and reliability. A further 1µm is then polished from the Si by CMP to smooth the surface. Then a dry etch step removes a further 39µm of Si to leave 10µm above the SiGe layer. Dry etching benefits from being a high-rate step that avoids any mechanical contact to the wafer. It also allows for in situ thickness monitoring by near-infra-red (NIR) interferometry. After dry etching, the remaining Si has a TTV around 2µm. Dry etching is fast (~9µm/min) but is not selective to SiGe. Therefore, a wet etch is needed to remove the remaining 10µm of Si to the SiGe. The wet etching compensates for the 2µm TTV leaving a SiGe TTV at ~20nm. The wet etch selectivity of Si to the SiGe is not high enough that wet etching could be used for the entire Si removal. The aim of the dry etching is to get close enough to the SiGe layer to allow the wet etch to stop within the SiGe and not breach it. The SiGe layer thickness also has to be at least 50nm to prevent that. Finally, the SiGe layer itself is wet etched, using a highly SiGe-to-Si selective chemical etchant, exposing the epitaxial 500nm-thick Si layer. **Figure 3** shows the Si thickness and TTV of the device wafer during the thinning processes. The final Si has a TTV around 70nm—identical to the thickness variation of the epitaxially-grown Si layer. The wafers are then ready for nano-TSV processing.

Nano-TSV formation

The TSV process flow is shown in **Figure 4**. The TSV patterning is done by through-Si alignment and needs to ensure that the nano-TSVs are aligned to the bottom metal 1 (M1) layer. This requires <20nm of overlay tolerance. The TSVs have a 180 x 250nm oblong top CD and are 500nm deep (the final thickness after the thinning steps). TSV etching is also carried out on the SPTS Rapier™ XE system, this time using a Bosch process with short cycle times to minimize the sidewall scalloping and assist with subsequent depositions. The TSV etch needs to stop on a thin dielectric layer that covers M1 as this avoids metal re-sputtering during TSV over-etching, which could otherwise lead to device reliability issues [4]. After TSV etching, 10nm of oxide is deposited conformally by plasma-enhanced atomic layer deposition (PEALD) throughout the TSV. Then the TSV and M1 liners are etched in a

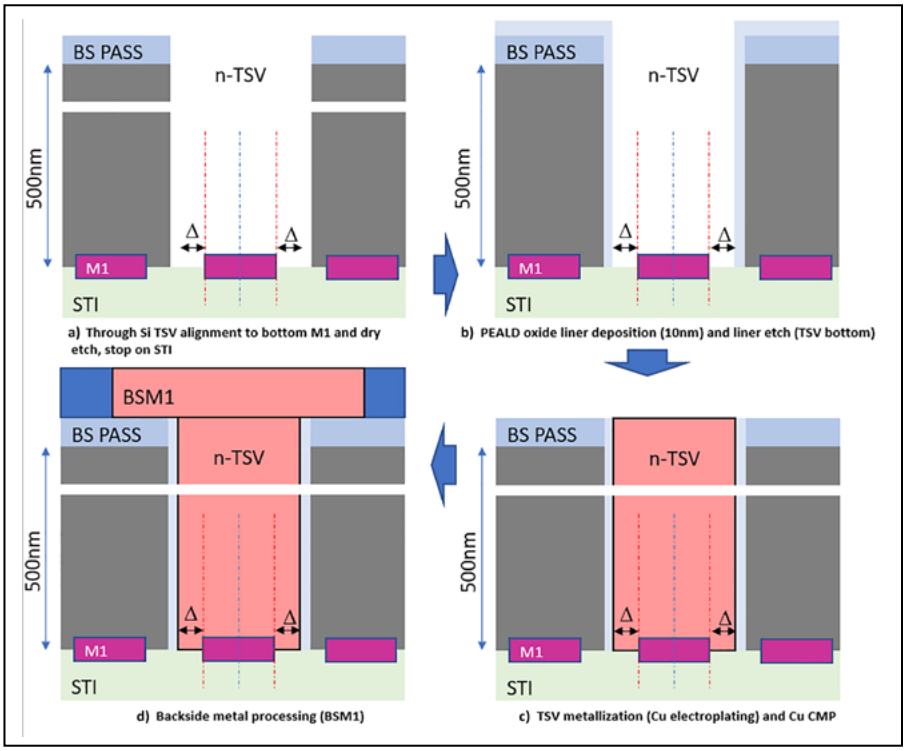


Figure 4: TSV patterning process flow: a) Through Si-TSV alignment to bottom M1 and dry etch, stop on STI; b) PEALD oxide liner deposition (10nm) and liner etch (TSV) bottom; c) TSV metallization (Cu electroplating) and Cu CMP; d) Back-side metal processing (BSM1).

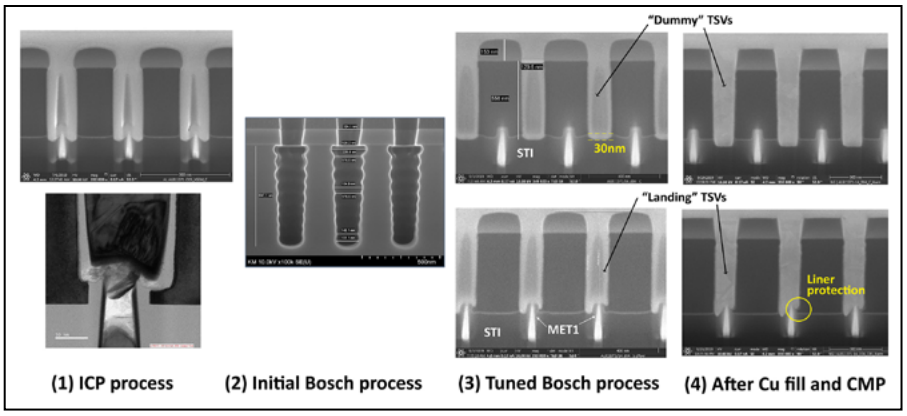


Figure 5: Nano-TSV etch process tuning and example after Cu fill and CMP.

single process to expose M1. The TSVs are then lined with Ta/TaN barrier and seed metals, plated with Cu and planarized by CMP. Finally, a Cu single damascene step completes the back-side metal layer (referred to as BSM1). There are two types of TSVs: "dummy" TSVs that land on the shallow trench isolation (STI) dielectric and "landing" TSVs that eventually connect to M1.

Figure 5 shows the results of the TSV etch process tuning. Initially, an inductively-coupled plasma etching (ICP; non-Bosch) process was attempted, but this lacks the selectivity to oxide that is required and exposes M1 because its thin liner is also etched away. Then taking a Bosch etch process that was originally developed for 1 x 5µm TSVs, the scalloping is too large within the nano-TSV dimensions and would dramatically impede deposition and filling. The Bosch process was therefore tuned by reducing the step times to give a smaller scalloping level consistent with the nano-TSV dimensions.

Electrical assessment

An electrical test vehicle was designed to demonstrate the very high-density electrical connections that can be achieved between the front side and back side of an extremely thin device wafer. The electrical characterization consisted of single measurements of Kelvin resistance (for a single TSV) and lengthy daisy chains connecting front-side M1 to backside BSM1 through multiple TSVs. The study also included the impact of overlay between the TSVs and M1. The detailed overlay data is beyond the scope of this article but, simply put, the Kelvin resistance is minimized when there is alignment between the TSV and M1, but also when the TSV etches beyond the M1 tips such that the contacting area is also maximized. The measurements indicate that the actual front-side to back-side overlay that has been achieved is <15nm.

The daisy chain resistance was evaluated as a function of the number of nano-TSVs from 40 up to 800. Around 99% electrical yield was achieved for all structures. **Figure 6** shows the daisy chain resistance data. Total chain resistance increases with the number of TSVs as expected. The data for resistance per chain link shows that the electrical yield of 99% for the Bosch etched TSVs reduces dramatically to ~70% for the ICP (non-Bosch) TSV etching. This is consistent with the data in **Figure 5**

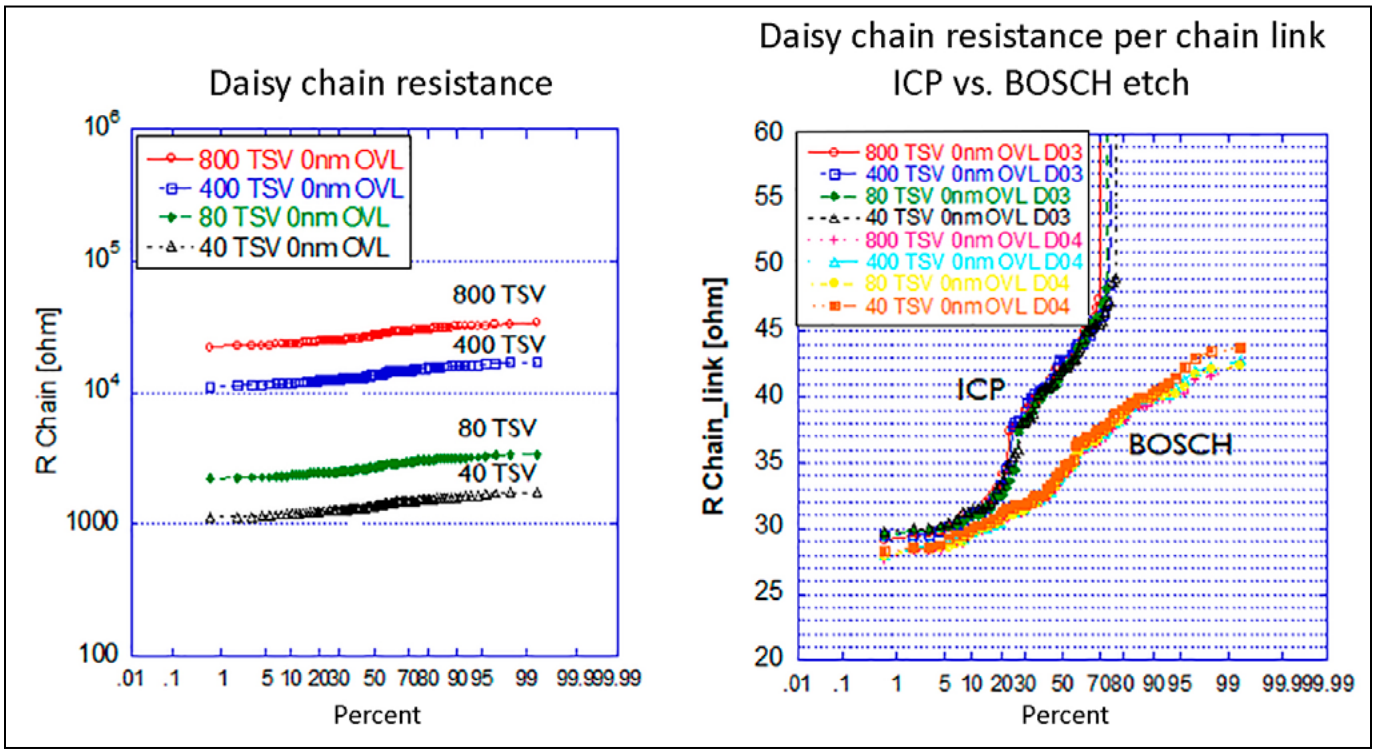
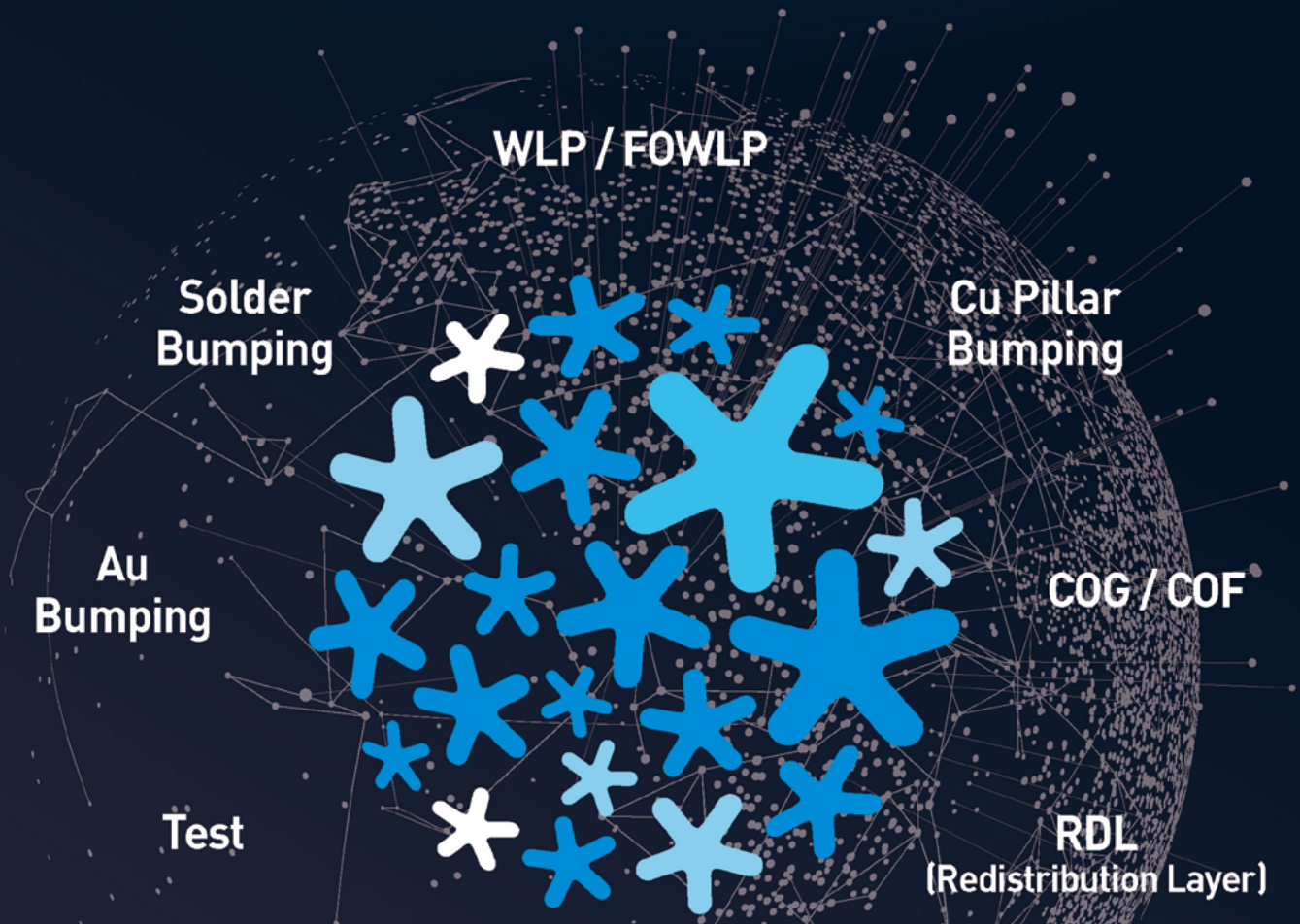


Figure 6: a) Daisy chain resistance as a function of the number of TSVs; and b) Daisy chain resistance per chain link, comparing ICP and Bosch etching.

WORLD NO.1

Specialized IC Packaging Company



Look Beyond Your Current Needs
to Make your Future Business Successful

LB Semicon

138, Cheongbuksandan-ro, Cheongbuk-eup, Pyeongtaek-si, Gyeonggi-do Korea
TEL +82-31-680-1600 **FAX** +82-31-680-1798
www.facebook.com/lbsemicon www.lbsemicon.com

whereby the ICP etching sputters some of M1 because of the low dielectric selectivity. This sputtered metal changes the TSV-to-M1 contact resistance. From this data it is clear that Bosch etching of nano-TSVs appears more stable and reliable than ICP etching.

Summary

Sub-500nm pitch interconnects that electrically connect the back side and front side of a device wafer have been realized by combining extreme wafer thinning to 500nm and nano-scale via-last formation. A combination of grinding, CMP, dry etching and wet etching can be used for the thinning of wafers so long as they contain a grown-in SiGe etch stop layer. Tight control over the final Si thickness has been achieved with TTV <70nm. Nano-TSVs with 180nm minimum CD have been etched through the remaining Si using a Bosch process that has the required selectivity to the M1 dielectric and the desired sidewall quality for subsequent deposition and plating. Functional electrical structures have shown 99% yield of the front-side to back-side connections and the data confirms the overlay between front side and back side to be <15nm.

References

1. J. De Vos, et al., "Hole-in-one TSV, a new via-last concept for high-density 3DSOC interconnects," ECTC, 2018.
2. A.Jourdain, et al., "Extreme thinning of Si wafers for via-last and multi-wafer stacking applications," IEEE ECTC, May 2018.
3. A.Jourdain, et al., "Extreme wafer thinning and nano-TSV processing for 3D heterogeneous integration," in IEEE ECTC, June 2020.
4. S. Van Huylenbroeck, et al., "A highly reliable 1x5µm via-last TSV module," IITC, 2018.

Biographies

Dave Thomas is VP, Product Management at SPTS Technologies, Newport, UK. With previous positions at Philips Components and Nortel, he has been with SPTS since 1994, working with both PVD and Etch product lines. He holds a BSc in Chemistry from Leeds U. (UK), and an MSc in Surface Chemistry and PhD Plasma Etching & Deposition from U. of Bristol (UK). Email Dave.Thomas@spts.com

Anne Jourdain is a Principal Member of Technical Staff at imec, Leuven, Belgium. She joined imec in 1999, and is currently responsible for process integration activities and wafer thinning in various projects within the imec's 3D System Integration Program. She's also part of their STCO (System-Technology Co-Integration) program. She holds an MSc degree in Opto-Microelectronics from Ecole Nationale Supérieure d'Ingénieurs de Caen (France) and a PhD from U. Joseph Fourier of Grenoble (France).

Revealing invisible defects on large 600mm panels

By Woo Young Han [Onto Innovation]

In their continuing drive to pack more computing power and speed into less space, semiconductor manufacturers are using advanced packaging (AP) processes to integrate multiple die of different types within a single package and to increase input/output (I/O) connectivity for large, complex chips. The use of front-end-like processes to create ever smaller features on ever larger substrates is increasing the need for process control and inspection in AP processes. Novel materials like organic polymer dielectrics pose special challenges to conventional front-end optical technologies. Our new illumination technology, Clearfind®, specifically addresses these issues to provide high-sensitivity defect detection.

Packaging evolution

Packaging processes have evolved from relatively simple, inexpensive technologies to costly, complex processes that have adopted and adapted process technologies developed for wafer fabrication. Some of these processes are discussed in the sections below.

Wire bonding. Traditionally, packaging uses thin wires bonded between I/O pads at the edge of the chip and a wire frame that includes pins for connection to a printed circuit board. The chip and frame are encapsulated for protection from the external environment, resulting in a final package that is much larger than the chip.

Flip chip. Chip-scale processes, like flip chip, form contact pads on the top surface of the die, which, when the separated die are flipped over, mate with solder balls on a connecting package substrate. Flip-chip packages allow many more I/O connections because the entire surface of the chip, not just the edges, can be used for contacts. The resulting package is smaller than wire bonding, but usually larger than the chip.

Wafer-level processing. Wafer-level processing (WLP) uses front-end-like processes to form packaging structures on chips while they are still part of the wafer on which they are fabricated. WLP has the benefit of creating small packages – the same size as the chip – but that small size ultimately limits the space available for I/O connections.

Fan-out wafer-level processing. Fan-out wafer-level processing (FOWLP) offers increased I/O capability. Separated chips are embedded in a round substrate with space added between the chips. Overlying redistribution layers (RDL) route signals from contacts on the top surface of the chip to contacts on the top surface of the larger substrate extending beyond the area of the chip itself. The round, wafer-like form factor of the reconstituted FOWLP substrates permits the use of process equipment and handlers designed for wafer processing with minimal modification. But it also limits the size of the substrate (and therefore the number of die that it can contain), and it wastes space near the curved edges of the substrate where rectangular die/packages do not fit efficiently.

Fan-out panel-level processing. Fan-out panel-level processing (FOPLP) is similar to FOWLP except the panel substrates are not limited to wafer-like form factors. They can be rectangular, to increase spatial efficiency, and larger, to process more die per panel at lower cost per die.

Challenges of FOPLP

As with most things electronic, there is ever-present pressure to reduce feature dimensions and spacing and to increase substrate size. Panel sizes have already grown as large as 730mm X 920mm and larger sizes are possible. The large number of expensive known-good-die contained on a panel makes process failures especially costly. Panels require handlers designed to accommodate not only their rectangular shape, but also their increased size and weight. Challenges associated with accurately positioning die, measuring shifts that unavoidably occur during the process [1], and maintaining the flatness of the repositioned die and the overall panel are also nontrivial.

This article focuses on inspection challenges for FOPLP processes. RDL line widths are now typically in the few tens of micrometers range, but line widths down to 1µm-2µm are on many roadmaps. This implies the need for inspection sensitivity to defects as small as 0.5µm. Although defects of this size are well within the detection range of optical inspection technologies, the materials used in FOPLP processes raise

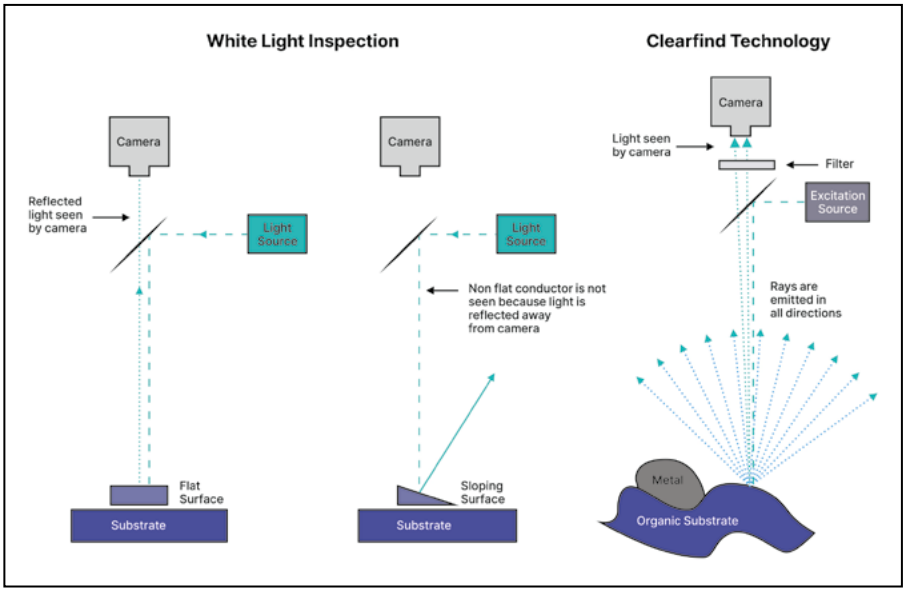


Figure 1: Simplified illustrations of the differences between traditional white light BF and DF illumination and CF technology.

specific issues. The organic polymer used as an insulator in RDL is essentially transparent, making any residues left after critical cleaning steps nearly invisible with conventional illumination techniques. Residues on metal contact surfaces can increase contact resistance or prevent electrical continuity completely. Contact residues can also degrade reliability, contributing to costly field failures. The graininess of metals used as conductors interferes with defect detection on metal surfaces, and the transparency of polymers, which allow the graininess of underlying metals to show through, extends that difficulty to insulators as well. Moreover, the automated detection routines used in optical inspection usually rely on a comparison of the sample to a "golden" standard. Any differences register as defects. Random differences in the grain pattern can therefore create thousands of false-positive nuisance defects, overwhelming the detection algorithm.

Illumination technologies

Conventional illumination approaches include bright-field (BF) and dark-field (DF) techniques. The work presented here uses a new technique, known as Clearfind® (CF), as implemented in our Firefly® macro defect inspection system. Figure 1 illustrates the essential characteristics of the three techniques. BF and DF systems typically use a broadband white light source. To simplify the discussion, assume that the sample surface is essentially flat and the features of interest – defects – are small and irregularly shaped. In bright field illumination the camera objective and illumination source are positioned on a common axis perpendicular to the substrate surface such that the camera sees the specular reflection of the illumination. The entire field of view appears to be uniformly illuminated—both the background surface (the field) and any features on it are bright. In dark-field illumination, the camera is positioned away from the direction of the specular reflection of the illumination source. On a perfectly flat, mirror-like surface, the specular reflection from the substrate is directed away from the camera and the field is dark. But any defect or surface irregularity that scatters light out of the specular beam will be bright. It is this characteristic that makes dark-field illumination particularly good at seeing small particles and defects on a flat specular surface, like that of a bare silicon wafer.

The light source for CF illumination is laser based. The light is monochromatic with stable wavelength and output power. The laser beam is collimated and expanded into a horizontal line at the sample and then scanned over the surface. Stimulated by the illumination, the sample emits light at a different wavelength and a wavelength filter in the optic path prevents reflected laser light from reaching the imaging camera. The intensity of the light emitted by the sample depends on the type of material illuminated. A high-speed, near-infrared laser-triangulation autofocus system maintains a constant distance between the imaging optics and the area being scanned. Imaging is accomplished using a high-resolution line scan camera. The image pixel size corresponds to 1.4µm on the sample surface at 4X and 0.7µm at 10X. CF technology is most powerful as part of a comprehensive inspection regime that may also include bright-field and dark-field inspection. The BF inspection results shown here were acquired on the same automated optical inspection (AOI) platform that acquired the CF images.

Results

The following sections discuss the results obtained by the study.

RDL sample. The sample shown in Figure 2 is a large molding compound panel. Figure 2a shows a 10X CF image and Figure 2b

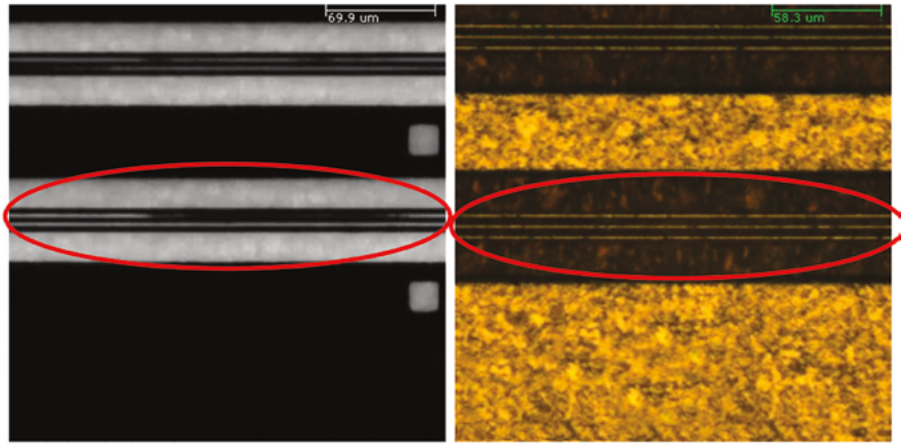


Figure 2: a) (left) CF and b) (right) BF images of 2µm horizontal RDL captured with a 10X objective (0.7µm/pixel)

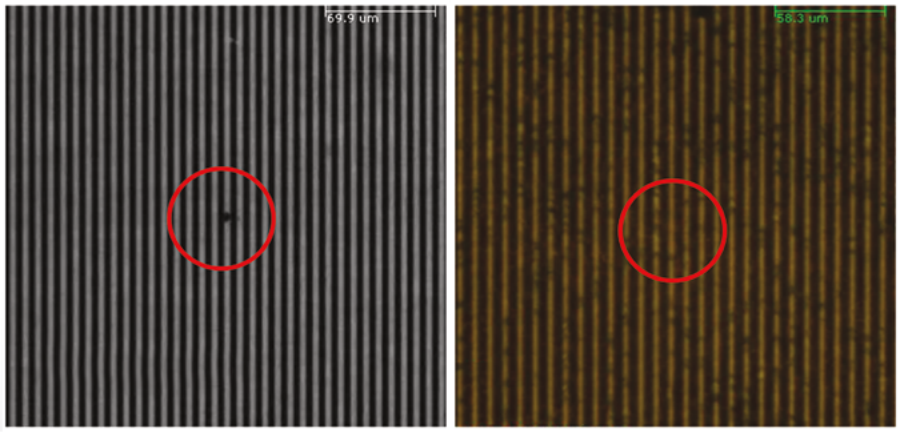


Figure 3: a) (left) CF and b) (right) BF images of vertically oriented 3.5µm RDL lines captured with the 10X objective (0.7µm/pixel)

shows a bright field image of the same area. In the areas circled in red, three 2µm RDL lines separated by 2µm spaces run horizontally across the image. A wider band of organic polymer runs beneath the RDL lines, isolating them from an underlying layer of metal. Contrast is reversed between the CF and BF images – metal is dark and polymer bright in the CF image, while metal is bright, and polymer is semi-transparent and darker in the BF image.

In the BF image (**Figure 2b**), the underlying metal surface showing through the polymer film looks very much like the RDL lines, making it difficult to distinguish the upper layer metal features from the material below. The graininess of the metal also obscures real defects and interferes with the automatic detection algorithm. Inspection of this sample with BF illumination resulted in high nuisance defect counts without finding real process issues on the wafer.

The CF image (**Figure 2a**) shows a clear contrast difference between the metal RDL lines and the underlying polymer. The absence of texture and graininess in the metal RDL lines and underlying metal layer permits extremely sensitive inspection with detection of single-pixel defects. A short between the upper two RDL lines that was not seen in the BF image is clearly visible in the CF image.

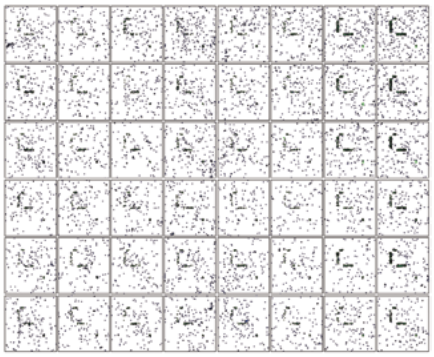


Figure 4: CF inspection results shown in a composite panel map.

Figure 3 shows 10X CF (**Figure 3a**) and BF (**Figure 3b**) images of the same area of vertically oriented 3.5µm RDL lines on a panel. The grainy metal surface and the noisy background in the BF image prevent clear visibility of the edges and continuity of the RDL lines. Unlike the BF image, the boundaries and the continuity of the RDLs can be clearly seen in the CF image. A short can be seen between the RDL lines in the CF image. The same short is not visible in the BF image.

Figure 4 shows the CF inspection results of RDL lines on a panel. The dots on the panel map represent defect locations. More defects were found on the right side of the panel and there appears to be a repeating pattern within each die/package.

Through-silicon via sample. The next is a large panel with through-silicon vias

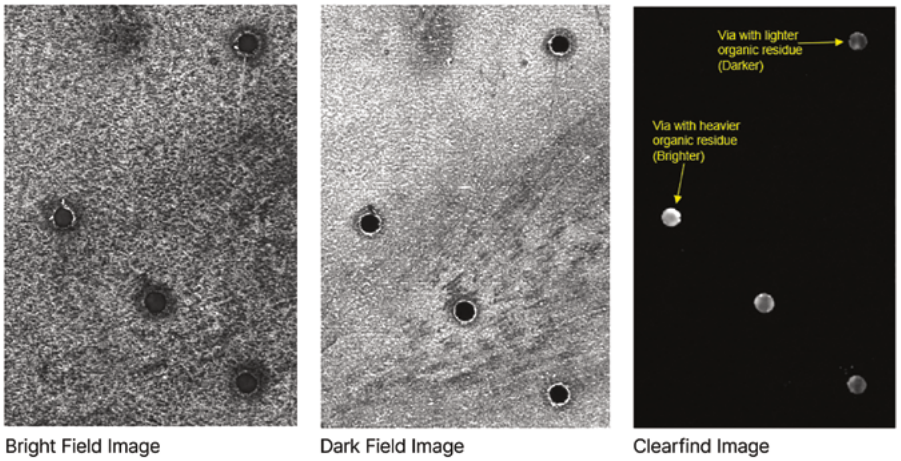


Figure 5: TSV images captured with a) (left) BF, b) (middle) DF, and c) (right) CF illumination using the 10X objective (0.7µm/pixel)

(TSVs) drilled for a 3D-IC process. The TSVs are about 30µm wide and 20µm deep and the pitch between TSVs is about 120µm. The panel is covered with a metal layer and has been through a "de-smear" cleaning process to remove organic residue from the vias. Inspections with BF and DF illumination have difficulty detecting the organic residue because of its transparency. Figure 5 shows 10X images of the same TSVs under BF (Figure 5a), DF (Figure 5b), and CF illumination (Figure 5c). All four vias have organic residue in them but it is difficult to see in the BF and DF images. The residue is clearly visible in the CF images and the brightness of the residue indicates the amount of residue present (brighter is more). Background noise from the metal surface variations can be seen in the BF and DF images, but is not visible in the CF image.

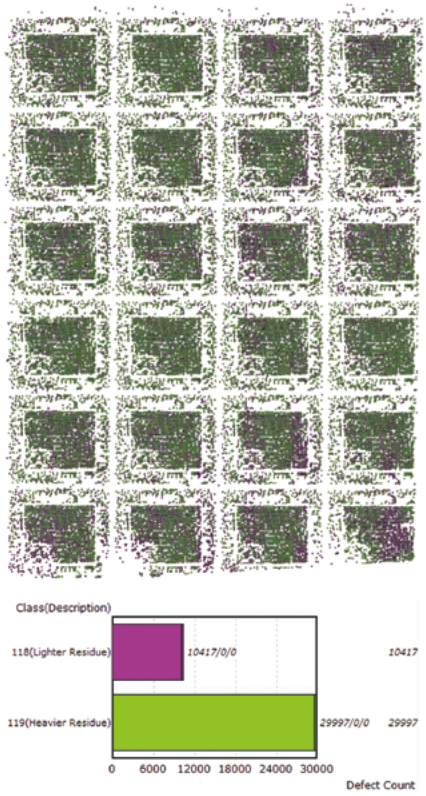


Figure 6: CF inspection results shown as a a) (top) defect map, and b) (bottom) as a histogram distinguishing lighter and heavier deposits.

Figure 6 shows the CF inspection result of the TSV panel. The dots on the panel map represent defect locations and show organic residues present in TSVs across the entire panel, indicating that the "de-smear" cleaning process did not work as expected. The TSVs will be filled with metal and organic residue may cause deplanarization and lead to connectivity issues. Organic residue can also add electrical resistance in TSVs.

FOPLP sample. The next example is a 450mm molding compound panel. Figure 7 shows 4X BF (Figure 7a) and CF (Figure 7b) images of RDL and under bump metallization (UBM) pad on the panel. The background metal surface underneath the transparent film looks just like the RDL and the UBM pad in the BF image, making it difficult to distinguish the upper layer metal patterns from the underlying metal layer. The metal texture and graininess also add noise to the image, making it difficult to detect defects and interfering with the detection algorithm. Inspection with BF illumination resulted in high

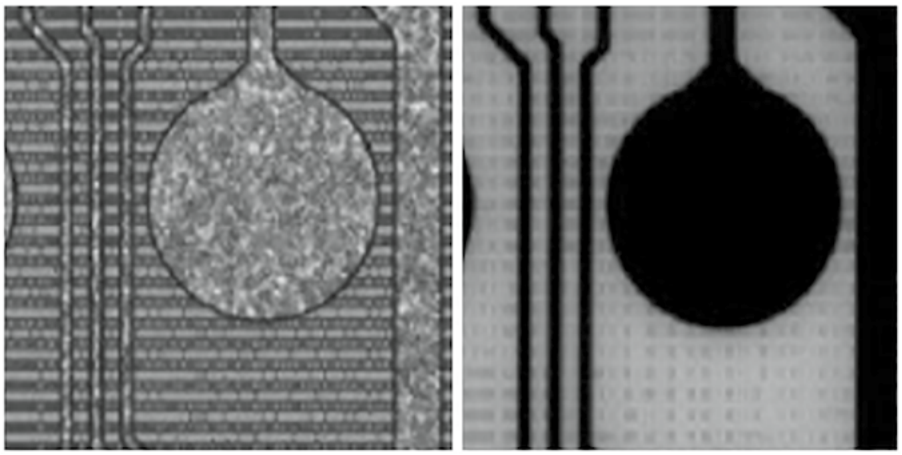


Figure 7: a) (left) BF and b) (right) CF images of RDL and UBM pads captured with 4X objective (1.75µm/pixel)

nuisance defect counts without finding real process issues on the wafer. The CF image (**Figure 7b**) shows a clear contrast difference between the previous metal layer in the background and the overlying RDL and the UBM structures. The texture and graininess of the metal are not visible, permitting high-sensitivity inspection with detection of single-pixel defects.

Figure 8 shows a composite map from CF inspection of the RDL and UBM panel. A repeating rectangular pattern matches the reticle layout of the lithography process. The defects comprising the pattern were determined to be undersized RDL lines, which were ultimately traced to a problem with the condenser lens of the lithography tool that produced thinner RDL lines on the die at the lower left corner of the reticle. **Figure 9a** shows CF images of thinner (8µm) RDL lines and **Figure 9b** shows images of normal (14µm) RDL lines. Five panels were in the lot and all of them had the same thin RDL problems caused by the lithography tool. All were inspected with both BF and CF illumination. The BF inspections found none of the thin RDL defects while the CF inspections found the reticle pattern on all five panels.

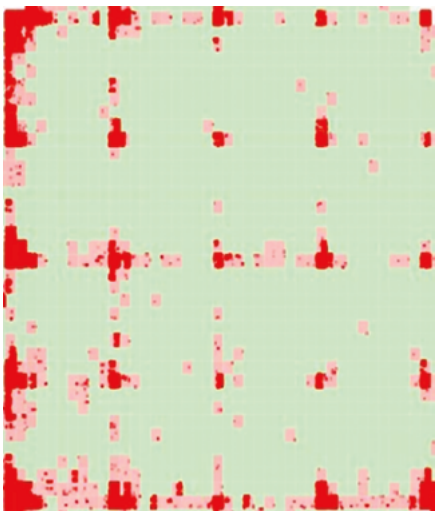


Figure 8: CF panel map of defects on the RDL/UBM sample reveals a repeating pattern that matches the lithography reticle layout.

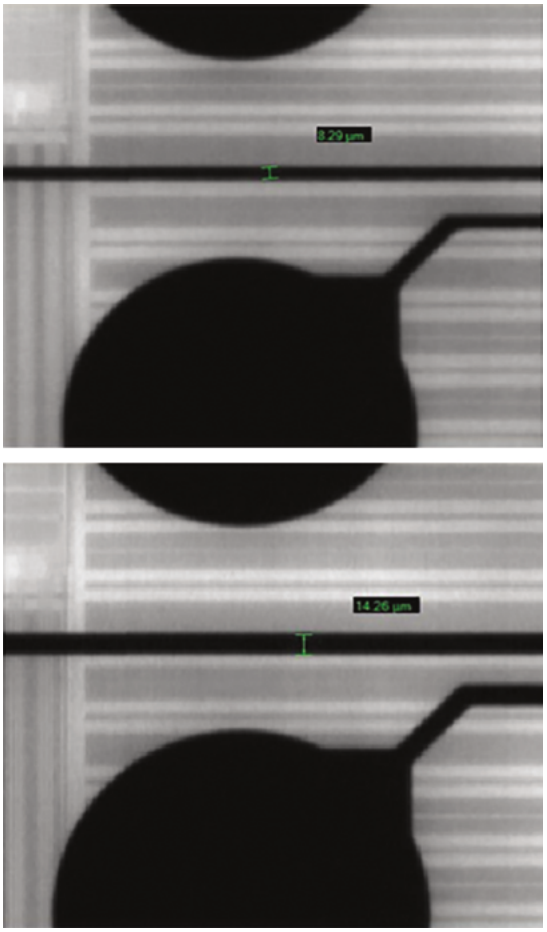


Figure 9: CF images of thin (top) and normal (bottom) RDL lines.

Summary

CF illumination technology provides clear advantages in detecting defects that are characteristic of FOPLP processes on large panels. It readily detects organic residues that are transparent and essentially invisible under BF and DF illumination. Its ability to eliminate the metal graininess and texture that obscure real defects and generate false-positive nuisance defects in BF/DF inspections allows CF technology to deliver high-sensitivity single-pixel defect detection.

Reference

1. G. Singh, K. Best, R. McCleary "Advanced packaging lithography and inspection solution for next-generation FOWLP-FOPLP processing," IEEE Xplore, Oct. 2016.

Biography

Woo Young Han is an applications engineering manager at Onto Innovation, Richardson, Texas. He joined the company in 2000 and is currently managing the applications engineering group for the inspection business. He holds an Electrical Engineering degree from U. of Toronto. Email wooyoung.han@ontoinnovation.com

Going beyond traditional temporary bonding materials

By Taro Shiojima, Munehiro Hatai, Minoru Inoue, Ryoichi Watanabe, Toshio Enami, Daihei Sugita
[SEKISUI CHEMICAL CO., LTD.]

High-performance computing (HPC) applications require not only acceleration of processing performance speed, but also reduction of power consumption and a lower cost structure. To meet those requirements, 3D packaging has recently been developed for multi-chip heterogeneous applications using through-silicon via (TSV), wafer on wafer (WoW), chip on wafer (CoW), and other technologies [1-5].

It is required that the technologies noted above stack multi-chips in a package and bond the chips to chips, or to wafers, directly [6,7]. For these technologies, several temporary bonding materials are used to support the handling of wafers. The wafer is temporarily bonded on the carrier wafer and then bonded to the other wafer after several treatments. After the previous steps, the carrier wafer is released to fabricate the 3D packaging. These temporary bonding materials are needed to combine two conflicting requirements: 1) securely holding the wafer during processing, and 2) easily debonding the hard carrier during the peeling off process. In recent years, these required performances have been getting more difficult to achieve because of the increased complexity of the fabrication process. The chemical and thermal stresses that temporary bonding materials suffer are also getting larger. Additionally, these packages are becoming thinner and thinner in order to stack more chips, so the wafer thickness after back-grinding is becoming thinner. For example, achieving a thickness <30µm and completing the debonding process are getting more difficult. These trends mean the difficulties of using temporary bonding materials are becoming more challenging and the requirements for new temporary bonding materials are growing [8].

Several kinds of resins and tapes are used as a temporary bonding material. The comparison of the carrier debonding methods of these temporary bonding materials is shown in Figure 1. The mainstream applications of these carrier debonding methods are mechanical, thermal and laser ablation. These debonding methods, however, have difficulties such as, risk of device damage while debonding, thermal resistance, and high process cost—including the cost of the debonding equipment. In response to these difficulties, more thermal-resistant and easy-to-debond temporary bonding materials have been developed.

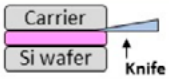
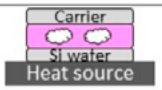
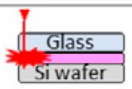
Debonding Method		Thermal resistance	Process Cost	Device damage
Mechanical		Good	Fair	Big risks
Thermal Debond		Poor	Good (Rinse-free)	Good
Laser ablation		Good	Expensive	Good
UV Gas Debond (SELFA)		Good	Good (Rinse-free)	Good

Figure 1: Comparison between SELFA and the other temporary bonding materials.

In this article "self-releasing adhesive tape (SELFA)" as a temporary bonding material is introduced. The new tape has a unique releasing system: ultraviolet (UV)-triggered gas generation from the tape. By using UV irradiation, gas is generated from the adhesive layer and the glass carrier is released automatically. This self-releasing technology enables the debonding system to be very simple and suitable for thinner wafers. By combining this gas generation technology and our adhesive design technologies, SELFA has chemical and thermal resistance up to 260°C. This means the new tape overcomes errors in the processes that traditional temporary bonding materials cannot even survive. In this paper, several advantages of the new technology are discussed, such as low total thickness variation (TTV), thermal resistance, no-residue, stress-free releasing technology, and so on.

Design concept and technologies

SELFA is a double-sided adhesive tape that has two different acrylic adhesive layers. One side is a self-releasing adhesive layer attached to a transparent hard carrier, such as a glass wafer, and the other side is an easy-to-peel adhesive layer that is attached to a device wafer. The tape structure is shown in Figure 2. The self-releasing adhesive layer is designed for carrier debonding assisted by gas generation. Gas is generated from the adhesive by UV irradiation and raises the glass carrier up so that the contacting area between the glass and the adhesive layer dramatically decreases. By decreasing the contacting area, adhesion strength becomes lower and the glass carrier can be detached free of stress. After glass carrier debonding, the adhesive tape can be smoothly removed from the wafer because the easy-to-peel adhesive was specifically designed for controlling adhesion strength under processing.

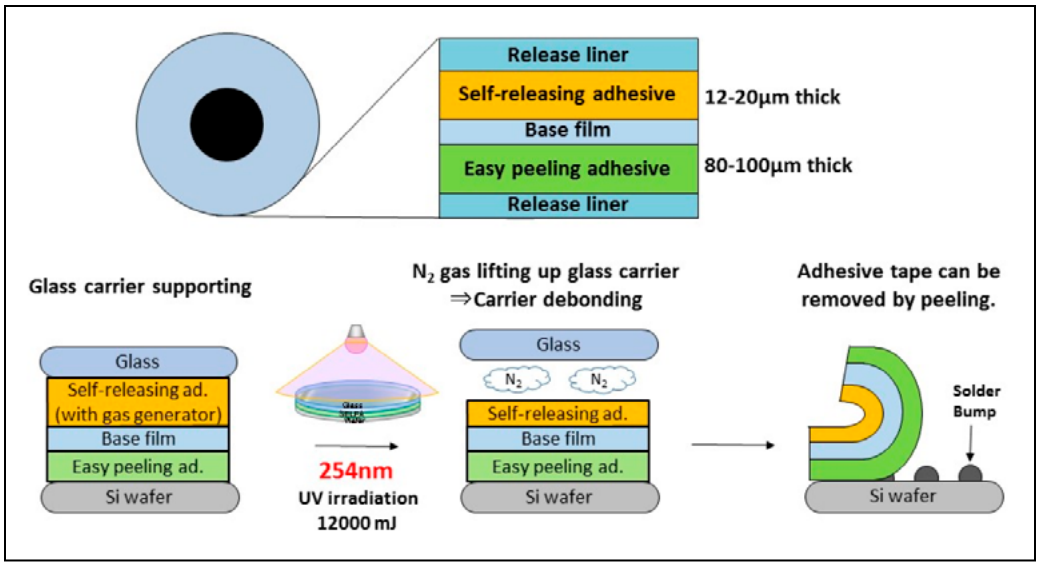


Figure 2: a) (top) Structure of SELFA; and b) (bottom) Gas debonding scheme.

The gas generator designed for gas debonding is one of the key technologies of SELFA. From the perspective of safety and to preclude contamination of the wafer, the gas used for carrier debonding should be inactive. In self-releasing adhesive, nitrogen compounds are dispersed and serve as the gas generator. N_2 gas is generated by the use of UV radiation at a wavelength of 254nm. In addition to the UV reactivity, the gas generator used in SELFA must have a thermal resistance of over 250°C, otherwise the tape can be delaminated because of the decomposition of gas generators during thermal processes such as chemical vapor deposition (CVD) or reflow. To improve the thermal resistance of the gas generator, the chemical structure of the nitrogen compound was optimized. By tuning the structure of the nitrogen compound, the gas generator has high thermal stability at temperatures over 250°C. In our tape, heterogeneous compounds, including nitrogen as a gas generator, disperse in an acrylic adhesive layer and react by way of the 254nm wavelength UV irradiation.

Another key SELFA technology is designing the adhesive to enable controlling adhesion strength during the process. As a temporary bonding material used in wafer fabrication processes, the adhesive tape must have chemical and thermal resistance so as to inhibit delamination, which can occur in the event of the tape dissolving or decomposing during the chemical or thermal processes, respectively. Another difficulty of using a temporary bonding material is that the adhesion strength needed to bond to the device wafer increases during the high-temperature process. Figure 3 shows how the adhesion strength changes between the adhesive and the adherend during the wafer fabrication process. For conventional adhesive tape, the adhesion layer becomes soft and its wettability to the adherend becomes greater during the thermal process so that the adhesion strength rises after the high-temperature process. This increase in the adhesion strength causes difficulty during the carrier debonding and also causes residue on the wafer after removal of the adhesive tape. To prevent the increase in adhesion strength, two approaches can be considered: 1) hardening the adhesive layer to prevent softening during thermal processing; and 2) decreasing the affinity for the adherend to reduce wettability.

For the first approach – hardening the adhesive layer – a chemical crosslinking structure is introduced into the adhesive resin. The adhesive resin contains unsaturated double bonds in the main chains and a UV initiator, which reacts to 405nm UV irradiation. By using UV radiation, unsaturated double bonds react

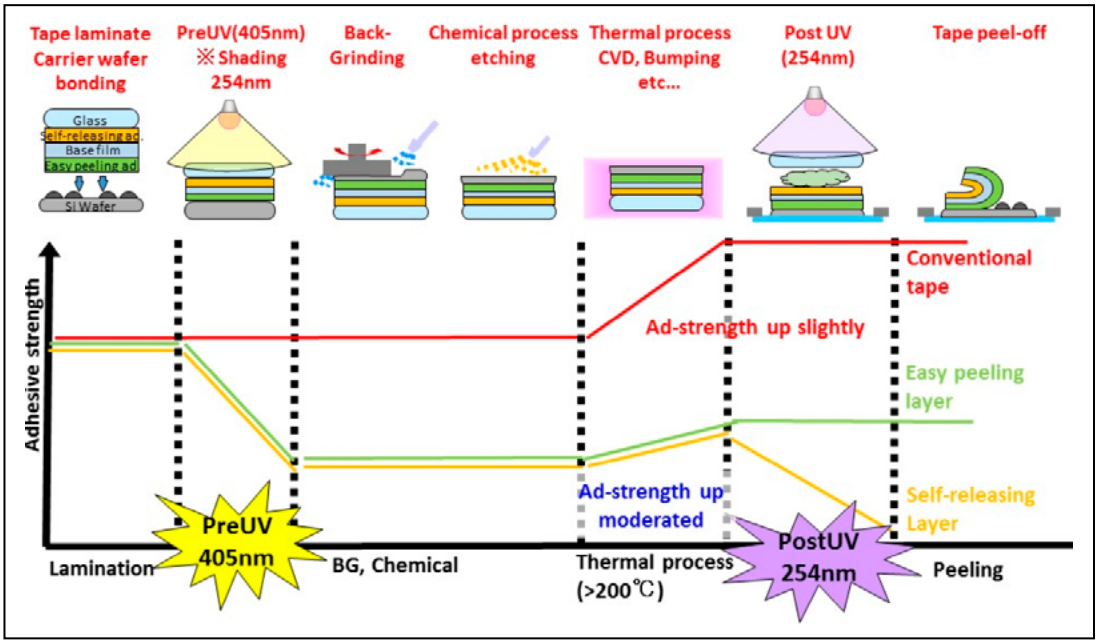


Figure 3: Image of adhesion strength change during the wafer fabricating process.

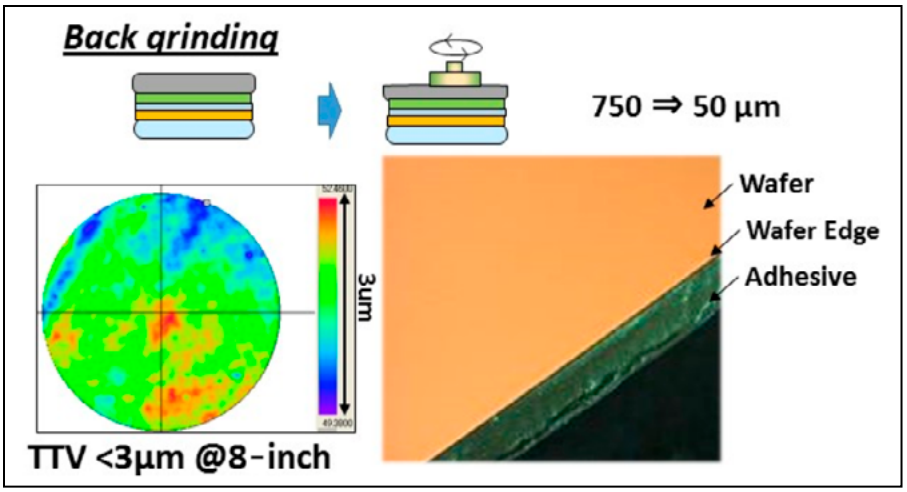


Figure 4: Wafer thickness variation in an 8-inch wafer and edge appearance after backgrinding supported with a glass carrier and SELFA.

by radical polymerization and the adhesion layer becomes rigid. The tensile modulus of elasticity of the adhesion layer increases from about 1~10 x 10^4Pa before UV irradiation, to 1~10 x 10^6Pa after UV irradiation. By using low-energy UV irradiation, polymerization in the adhesive layer can be achieved without N_2 gas generation occurring if high-energy UV radiation were used. Figure 3 shows some suggested applications, such as laminating the self-releasing adhesive tape to the wafer and irradiating using 405nm UV light before loading into processes like backgrinding (BG), chemical vapor deposition (CVD), or other chemical processes. By hardening the adhesive layer before loading into several processes, the degradation and softening of the adhesive layer during these processes can be prevented. To distinguish this UV irradiation at 405nm for polymerization of the resin before chemical or thermal processing from the use of UV irradiation at 254nm for N_2 gas generation, we refer to the 405 nm-wavelength UV irradiation as "pre-UV." By using pre-UV irradiation, the adhesive got rigid and the adhesion strength gets lower. This crosslinking of the adhesive resin also improves chemical and thermal resistances because the resin is stiffly bonded and it prevents components in the adhesive from dissolving or decomposing in chemical or thermal processes, respectively.

For the second approach – to decrease affinity to the adherend – the surface polarity of the adhesive resin is controlled by adding low polarity components into the adhesive layers. By adding low polarity components to the adhesive layers, the surface polarity is lowered and the rise in the adhesion strength because of heat becomes moderated, as shown in Figure 3. This means the tape can be removed easily after the thermal processes. Combining these two methodologies—pre-UV and adding a low polarity component—enables easy removal of SELFA from the wafer after processing.

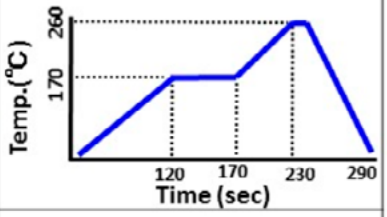
Process applicability

To determine the applicability of SELFA as a temporary bonding material, the following points were evaluated: 1) total thickness variation (TTV) after BG; 2) chemical resistance; 3) thermal resistance; 4) gas debonding; and 5) residue after tape removal. The evaluation followed the process flow shown in Figure 3.

Figure 4 shows the wafer thickness variation after BG supported by a glass carrier with SELFA. The wafer was ground down smoothly from 750µm to 50µm without wafer cracking and the TTV of the thin wafer was less than 3µm. It means that smooth BG can be realized by having the hard carrier supported with SELFA without

Solvent	Temp. (℃)	Process time (sec)	Weight loss (wt%)
Tetrahydrofurfuryl Alcohol 75-100%	60	420	<2
N-Methyl-2-pyrrolidione	50	3000	<2
Tetra Methyl Ammonium Hydroxide (TMAH) 3%	22±3	200	<2
Potassium Hydroxide (KOH) 10%	22±3	90	<2
Propylene Glycol Monomethyl Ether (PGME) 79-81%	22±3	60	<2
Copper Sulfate ($CuSO_4$) 18%	22±3	4200	<2
Sulfuric Acid (H_2SO_4) 5-10%	22±3	4200	<2
Hydrochoric acid (HCl) 0.006%	22±3	4200	<2
Acetic Achid 10%	22±3	60	<2

Hot plate test	Reflow test
220℃ x 2hr x 3passes	Max. 260℃ x 10passes
Result: No delamination	Result: No delamination



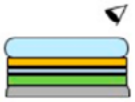


Figure 5: a) (left) Chemical resistance; and b) (right) Thermal resistance of SELFA.

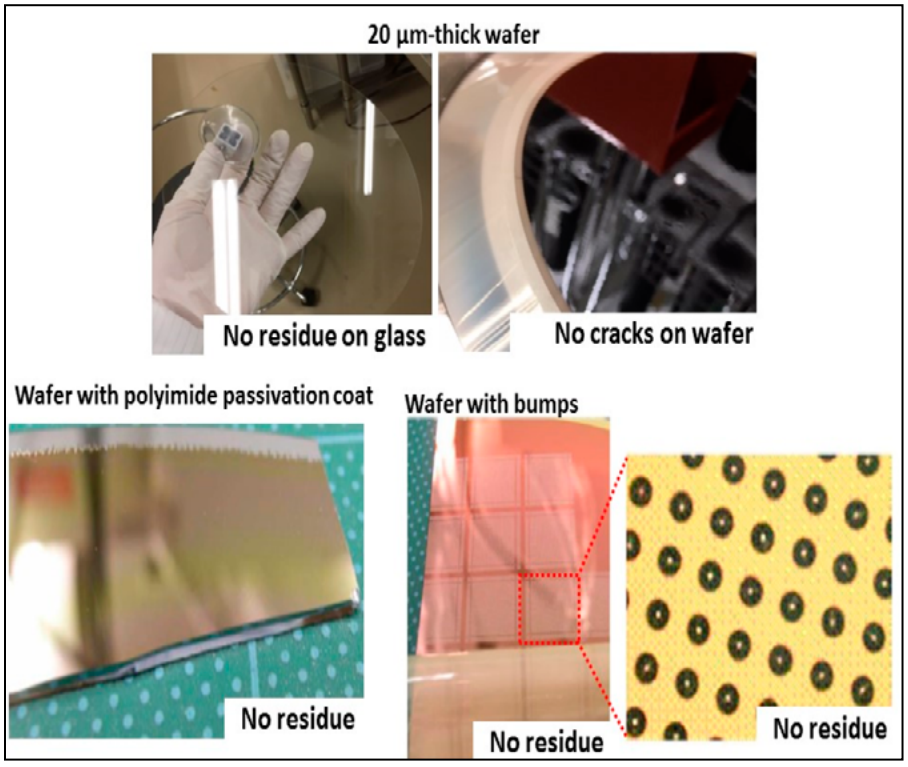


Figure 6: a) (top) Appearance of a 20µm-thick wafer; and b) (bottom) Several types of wafers after glass debonding.

an additional planarization process. This TTV value is the same as for the other temporary bonding materials [9]. No edge crack was observed from the wafer edge.

Chemical resistance was evaluated by measuring the weight loss of SELFA in several solvents. **Figure 5a** shows the chemical resistance of the new tape against solvents and conditions used in wet processing. The weight loss values of SELFA were <2% in any of the conditions we evaluated. This means that the new tape is stable in these solvents and can safely support wafers in wet processes. This chemical resistance is enabled by chemical crosslinking of the adhesion layer by using pre-UV. The resin in the adhesive is chemically bonded and makes a three-dimensional network by use of the 405nm UV irradiation. This chemical networking improves the stability of the adhesive and prevents delamination.

Thermal resistance was evaluated by observing the delamination behavior of the wafer/SELFA/glass structure during and after heating. **Figure 5b** shows the evaluation scheme of the thermal resistance and the appearance of the wafer/SELFA/glass structure after evaluation. Thermal resistance was evaluated after pre-UV irradiation. Two thermal conditions were prepared as follows: 1) hot plate, and 2) reflow, which were set up as simulations of thermal processes, such as redistribution layer (RDL) curing and soldering, respectively. Delamination was not observed after 3 cycles at 220°C x 2hr treatment on the hot plate in atmosphere. Likewise, delimitation was not observed after 10 cycles at 260°C x 5min treatment in the reflow process. Delamination was prevented because outgassing from SELFA was decreased by using thermal resistant material for the acrylic resin, base film, low polarity component, and the gas generator, which are stable in high temperature, such as <260°C. The combination test of hot plate and reflow was also evaluated. Delamination was not observed after continuous treatment of 220°C x 2hr on a hot plate

and 260°C x 5min reflow. This result shows that SELFA can be applied to higher temperature treatments such as RDL processing and soldering. It means SELFA is applicable to both fan-out wafer-level package (FOWLP) processing and to fan-out panel-level package (FOPLP) processing.

Glass carrier debonding from the wafer/SELFA/glass structure was demonstrated with a wafer thickness of <100µm. Before carrier debonding, the wafer supported with glass was treated for 1hr at 200°C in an oven and then transferred onto a dicing (DC) tape. The 200°C thermal treatment was a simulation of a high-temperature treatment like CVD. After UV irradiation (254nm) the glass carrier was manually picked up with a suction stage (**Figure 6a**). The glass carrier was easily debonded from SELFA without cracking the wafer. The gas that was generated from the adhesive layer attached to the glass carrier enabled the glass carrier to be lifted up from the adhesive; the contact area between the glass and the adhesive thereby became approximately zero. So the glass was able to be debonded from the thin wafer without cracking it. We also confirmed that the tape can be peeled off from the thin wafer without cracking. This result shows that SELFA has a potential as a temporary bonding material applying to a very thin wafer such as one that is 20µm thick, which will be demanded in the near future for 3D package applications.

The surface of the wafer was also observed after removal of SELFA. Several adherends: a Si bare wafer, a wafer with a polyimide passivation coat, and a wafer with bumps were used. **Figure 6b** shows the wafer appearance after tape removal. No residue of adhesive was observed on the tested wafers.

Application idea

The various application ideas of SELFA have been proposed. In this article, two processes are introduced, CoW and FOWLP. **Figure 7a** shows process images of CoW using the new tape. In the CoW process, SELFA is laminated on the buffer wafer first. Then, BG and back-side fabrication is completed. After that, core chips are bonded by thermocompression bonding. In our thermocompression bonding test with a 300°C bonding head, 8 chips were successfully stacked without delamination between the buffer wafer and the adhesive tape. After that, the wafer-level molding can be applied.

The second application idea is for a FOWLP, chip-first, face-up process. **Figure 7b** shows process images of FOWLP using the new tape. In this process, chips are mounted on the SELFA and after that, wafer-level molding, BG and bumping were applied. During the FOWLP process, chip shift during the molding step is one of the major challenges. In our test, chip shift and subduction during molding were measured and controlled within 2µm, and 1µm, respectively.

In addition to the processes noted above, the new tape has been used in several other applications and many product line-ups are prepared to deal with each customer's process. Adhesive strength and layer thickness can be custom-made. Several single-sided types utilizing the easy-peeling adhesive design are also being readied.

Summary

In this paper, self-releasing adhesive tape (SELFA) as a temporary bonding material and a thin wafer supporting system with SELFA and glass carrier were introduced. This temporary bonding adhesive tape has a unique technology for debonding carriers, which enables self-releasing by gas generation. At the same time, the adhesive resin has been designed for easy peeling. By combining these two technologies, self-releasing adhesive tape can be applied to 3D integration processes including wet and high-temperature processes. It is anticipated that this new tape will drive further wafer thinning and support handling of devices in the upcoming 3D integration era.

References

1. N. Araki, S. Maetani, Y. S. Kim, S. Kodama, T. Ohba, "Development of resins for bumpless interconnects and wafer-on-wafer (WOW) integration," 2019 IEEE 69th Elec. Comp. and Tech.

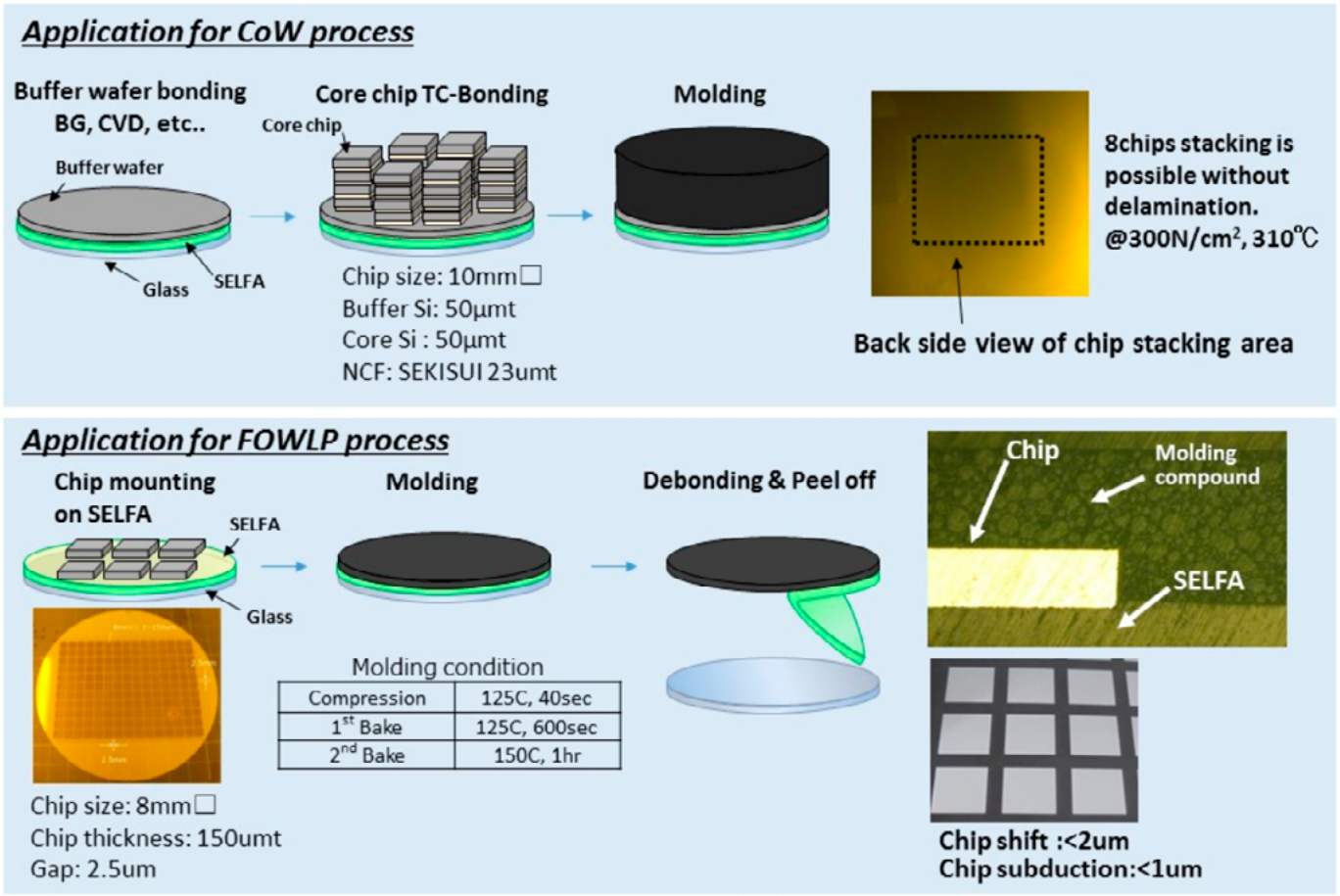


Figure 7: Application idea of SELFA to CoW and FOWLP processes.

Conference (ECTC), pp. 1002-1008.

2. L. England, D. Fisher, K. Rivera, B. Guthrie, P-J. Kuo, C.-C. Lee, et al., "Die-to-wafer (D2W) processing and reliability for 3D packaging of advanced node logic," 2019 IEEE 69th ECTC, pp. 600-606.
3. Y. S. Kim, S. Kodama, Y. Mizushima, N. Araki1, C. Hsiao, H. Chang, et al., "Optimization of via bottom cleaning for bumpless interconnects and wafer-on-wafer (WOW) integration," 2018 IEEE 68th ECTC, pp. 1962-1967.
4. A.-J. Su, T. Ku, C.-H. Tsai, K-C. Yee, D. Yu, "3D-MiM (MUST-in-MUST) technology for advanced system integration," 2019 IEEE 69th ECTC, pp. 1-6.
5. A. Jouve, L. Sanchez, C. Castan, M. Laugier, E. Rolland, B. Montmayeul, et al., "Self-assembly process for 3D die-to-wafer using direct bonding: A step forward toward process automatization," 2019 IEEE 69th ECTC, pp. 225-234.
6. F. C. Chen, M. F. Chen, W. C. Chiou, D. C. H. Yu, "System on integrated chips (SoIC™) for 3D heterogeneous integration," 2019 IEEE 69th ECTC, pp. 594-599.
7. F.-C. Hsu, J. Lin, S.-M. Chen, P.-Y. Lin, J. Fang, J.-H. Wang, S.-P. Jeng, "3D heterogeneous integration with multiple stacking fan-out package," 2018 IEEE 68th ECTC, pp. 337-342.
8. WonMyoung Ki, WonGeol Lee, IlBok Lee, InSu Mok, WonChul Do, Moh Kolbehdari, "Chip stackable, ultra-thin, high-flexibility 3D FOWLP (3D SWIFT® technology) for hetero-integrated advanced 3D WL-SiP," 2018 IEEE 68th ECTC, pp. 580-586.
9. M. Lueck, A. Huffman, M. Butler, D. Temple, P. Garrou, "Temporary wafer bonding materials and processes," Additional Conferences (Device Packaging, HiTEC, HiTEN, & CICMT): Jan. 2012, Vol. 2012, No. DPC, pp. 001452-001476.

Biographies

Taro Shiojima is a Material Engineer in the R&D Institute at SEKISUI CHEMICAL CO., LTD., Osaka, Japan. He engages in the development of materials for 3D packaging applications and holds a Master's degree in Material Engineering from the U. of Tokyo, Japan. Email taro.shiojima@sekisui.com

Munehiro Hatai is Chief Manager in the R&D institute at SEKISUI CHEMICAL CO., LTD., Osaka, Japan. He has been developing semiconductor materials for 20 years and established the basic technology of SELFA. He holds more than 40 Japanese patents.

High-throughput flexible direct imaging for packaging/MEMS fabrication

By Shota Majima, David Hyde [SCREEN Semiconductor Solutions Co., Ltd.]

There are a number of direct imaging systems for advanced packaging and microelectromechanical systems (MEMS) applications. Several are discussed in the sections below, and this article will propose using direct imaging for the die-first fan-out packaging of heterogeneous devices that use high-density organic interconnects. This last application will be key for future high-performance computing (HPC).

Mask-less direct imaging lithography

The most common patterning method of semiconductor microfabrication is to project a mask pattern onto the surface to be exposed. In R&D departments, maskless direct imaging systems are more common as the need for a flexible lithography process has taken priority over the higher throughput of a mask-based system. However, as the pattern designs in mass production become smaller, and of increasing complexity, the cost of manufacturing the device is increasing, partially driven by the mask production and the complexity in resulting processes. The combination of these factors has become a financial burden and a design constraint to manufacturers. In fact, MEMS manufacturing and some advanced packaging techniques, like heterogeneous or chiplet integration, are now looking toward maskless systems to resolve these issues. Especially in die-first fan-out packaging technology, the flexible compensation exposure of a maskless system is necessary to achieve production level throughput with the high-overlay accuracy needed to maintain yield. The stepper/aligner approach of mask-based lithography enjoys high overlay accuracy, but often does so at the expense of throughput.

Direct imaging with spatial light modulator

The Grating Light Valve (GLV™) is a high-performance spatial light modulator composed of thousands of free-standing silicon-nitride micro-ribbons anchored on the surface of a silicon die. By electronically controlling the deflection of the ribbons, the GLV™ functions as a programmable diffraction grating, enabling attenuation, modulation and switching of laser light with unparalleled resolution, speed, and precision. Using this new technology, we have developed a flexible direct imaging system for advanced packaging and MEMS industries with support for resolutions down to 2/2µm. When configured with multiple exposure heads, this system can process a wide range of substrate types, from 200mm/300mm wafers up to 600mm square panels. The ability to control the reflected light level by changing the electrical bias at each ribbon allows the use of multiple exposure dose levels (grayscale exposure) in the same exposure operation. Figure 1 illustrates the light valve's structure (Figure 1a), including grayscale exposure samples. With grayscale exposure, the resist side wall angles can be controlled to generate the tapered shape needed for Cu sputtering (Figure 1b) and avoid the reverse taper that leads to sputtering defects and plating continuity failures. This technique can also be applied to produce dual-damascene

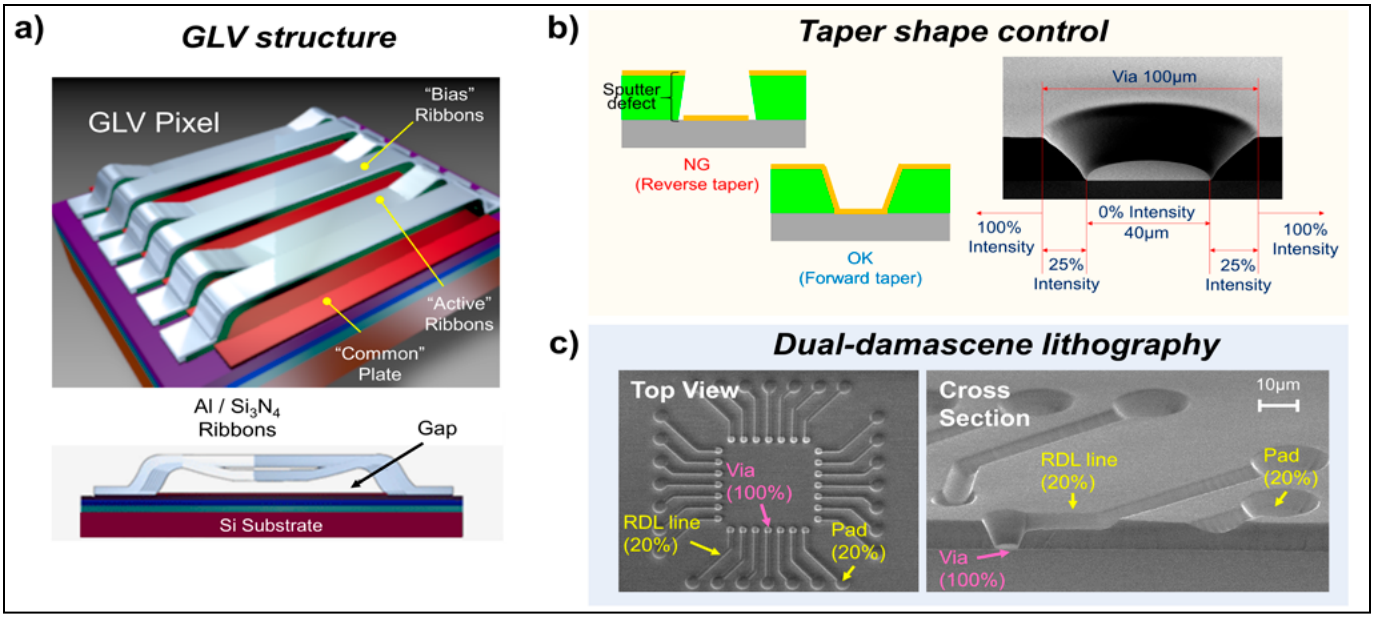


Figure 1: Illustration of GLV™ a) structure, b) taper shape control, and c) application to dual-damascene lithography.

structures (Figure 1c), by applying a full dose at the via and a partial dose for the redistribution layer (RDL) lines, a single-exposure process is used without impacting throughput. The inherent flexibility of maskless exposure, in conjunction with the light valve's ability to modulate the exposure dose can reduce the number of process steps required and increase yield while maintaining throughput.

Dynamic data generation

Direct imaging systems have a wide range of functions to increase flexibility for advanced packaging and MEMS fabrication processes, many of which utilize dynamic data generation. Consider three different alignment methods. The first and simplest of these is global alignment, which measures two or four alignment marks on the substrate and then adjusts offsets in the X, Y, and θ directions. Global alignment uses dynamic data generation to compensate for linear expansion/contraction and is particularly well suited for rigid processes, such as silicon or glass-based substrates. Next is local alignment, which expands on the global alignment capabilities by compensating for nonlinear distortion of the substrate. Suitable for many processes – including panel circuit board (PCB) – local alignment measures up to 400 points on a substrate, calculates the difference between design position and actual position, and dynamically adjusts the exposure data for the entire substrate. The third method – die-by-die alignment – is crucial for die-first panel/wafer-level packaging processes. In these die-first processes, reconstituted substrates exhibit random die offsets that require compensated exposure to achieve acceptable overlay performance. Steppers use a chip-by-chip approach, measuring and exposing each die individually, but suffer from significant decreases in throughput as a result. On the other hand, direct imaging tools can expose all the die with one scanned exposure, using

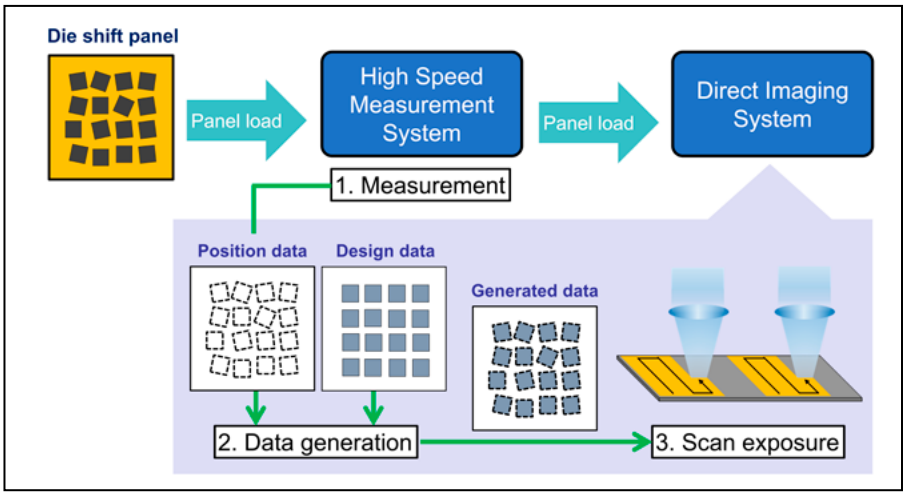


Figure 2: High-throughput compensated exposure process flow.

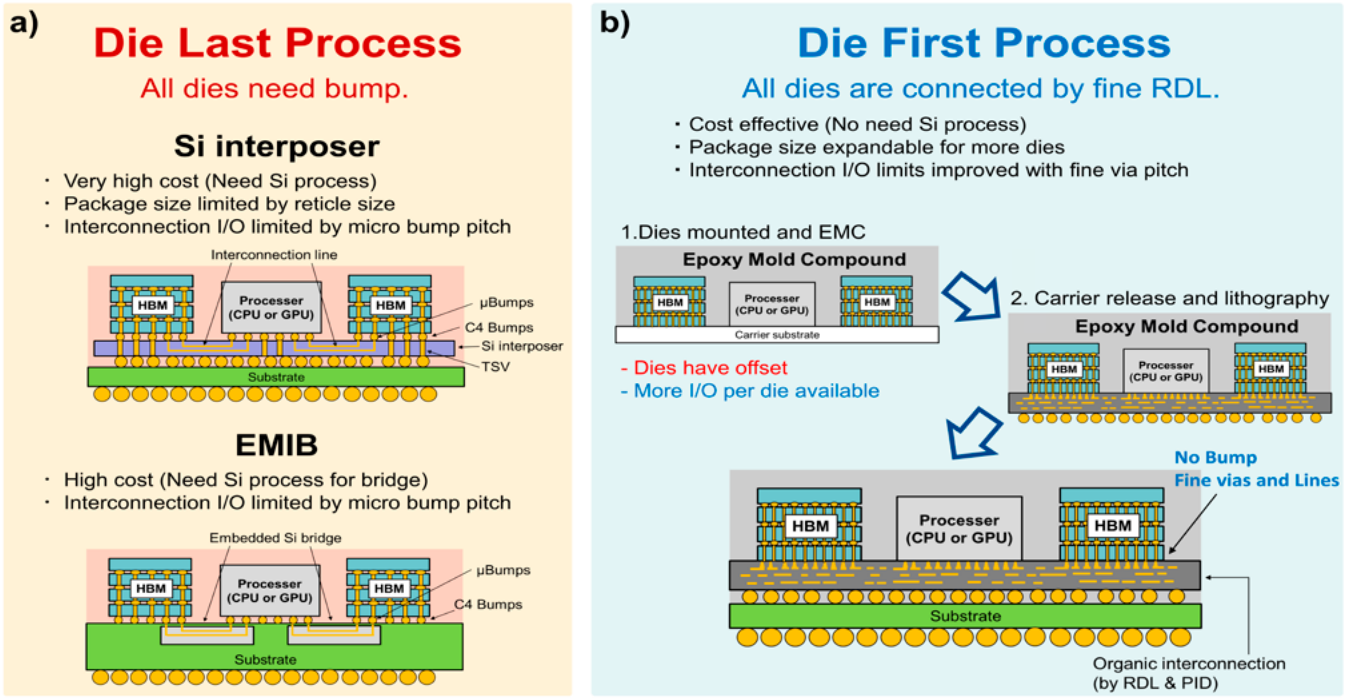


Figure 3: Direct imaging for heterogeneous devices: a) Approaches utilizing silicon interposer/EMIB where all dies need a bump; and b) a die-first approach where all dies are connected by fine RDL.

the pre-measured position data of each die to dynamically generate compensated exposure data without impacting throughput. Using pre-measured data and dynamic exposure pattern generation also enables selective patterning, such as selecting pad openings in arbitrary areas, or not exposing die whose displacement is too large.

Dynamic pattern generation supports unique functions for high-volume manufacturing as well. One of these is an auto-numbering function to provide traceability as required for critical devices (automotive, healthcare, etc.). Auto-numbering functions can simultaneously add unique ID numbers for each die during the normal exposure process. Using simple rules, unique ID numbers are automatically generated and applied to individual dies. Auto-wiring is another example, beneficial to advanced packages requiring electrical interconnects between multiple dies, each with independent displacement. In these packages, die displacement compensation alone is not enough—the exposure data for the interconnect lines will also need to be adjusted, and dynamic pattern generation with a direct imaging system can do so without impacting throughput. In both the advanced packaging and MEMS industry segments, direct imaging's unique capabilities are improving yield and enabling traceability of critical devices.

High-speed compensation exposure system

One challenge to date for the mass production of heterogeneous integration has been the throughput of the lithography system. Dies on the reconstituted substrate have random and independent offsets from mounter accuracy errors and/or movement (expansion/contraction) from the epoxy mold compound (EMC). These offsets must be measured before exposure and the pattern modified to compensate. This measurement time increases with the number of dies on the substrate, further reducing the throughput of the exposure system. However, we have shown it is possible to pre-measure the die with a stand-alone tool to realize the full advantages of the direct imaging approach (Figure 2). In the measurement tool, the exact positions of all dies are recorded. Twin time delay integration (TDI) cameras are used to scan the substrate and create an image of the entire surface. Our algorithm then efficiently and rapidly extracts localized images, wherein alignment mark matching is performed on each die. With this system, the measurement time for 5,000 mounted dies on a 515x510 substrate is as low as 120s. After measurement, the substrate design and measurement data are loaded to the direct imaging tool. The direct imaging tool then evaluates the measured position against the original design data, dynamically generates the compensated data for the entire substrate, and executes the scanned exposure. While exposing, subsequent substrates are measured with the stand-alone tool and compensated exposure data is produced. By utilizing measurement and exposure processes in parallel, high throughput, including exposure data compensation, is realized. The key aspects for this approach are high-speed measurement and high overlay accuracy.

The method described above is proven to be very effective at compensating for displacement. For example, we mounted 600 dies onto a 515x510mm substrate. The stand-alone measurement system was used to measure the displacement of each die, with results showing displacements of |Average|+3σ = 81.5µm and 162.0µm in the X and Y directions, respectively. After exposing with the dynamically-generated compensation data, overlay results were |Average|+3σ = 1.9µm and 1.4µm, clearly showing our system's ability to flexibly compensate for die displacement across the entire panel.

Summary

In this article, we have discussed the flexibility of direct imaging systems for advanced packaging and MEMS industries. Now, we propose using direct

imaging for die-first fan-out packaging of heterogeneous devices utilizing high-density organic interconnects for future high-performance computing (HPC) devices. In Figure 3a, two popular packaging technologies for HPC that utilize a silicon interposer or embedded multi-die interconnect bridge (EMIB) techniques are shown. These processes are referred to as "die-last processes" as the fan-out RDL patterns and interconnect lines are built on the substrate before the dies are mounted. To connect the substrate and dies, C4 bumps and µbumps are used. Figure 3b shows our die-first proposal that eliminates the bumps required in the die-last process. With this method, all of the dies are fixed to a temporary carrier with epoxy mold compound (EMC). After releasing the temporary carrier, the RDL layer is built up, utilizing the compensated exposure and auto-wiring abilities of the direct imaging system to connect the dies and compensate for random die placement offsets. The silicon interposer or embedded silicon bridge interposers are no longer needed, leading to an overall reduction in process cost. In addition, package size is not limited when using maskless exposure, allowing for additional dies per package when larger substrates are selected. This type of die-first process also removes the need for C4 and µbumps, which in turn supports the use of smaller contact vias and reduced line pitches. With the increased area available after reducing the vias and lines, additional I/Os can also be added. These changes are possible with the direct imaging system's auto-wiring and high-resolution capabilities. The proposed process enables advances in future high-performance packaging that are expandable and enable higher I/O densities while remaining cost effective. We believe these benefits of the maskless exposure system will expand packaging design beyond current limitations, enabling the next generation of dynamic device innovation.

Biographies

Shota Majima is a Process Engineer for semiconductor equipment working at SCREEN Semiconductor Solutions, Kyoto, Japan. He engaged in the development of dry etching process for FEOL for 3 years as his first career. After that, he joined lithography process development for the direct imaging tool and has engaged in the development of various kinds of packaging technologies for 3 years. He holds a Master's degree in Electronic Engineering from Nara Institute of Science and Technology (NAIST). Email majima@screen.co.jp

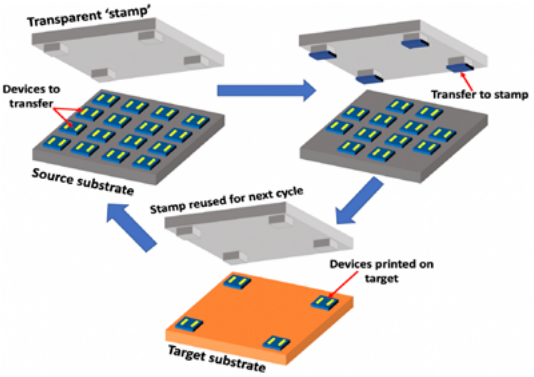
David Hyde is a Product Engineer at SCREEN-SPE USA, Sunnyvale, CA. His early career focused on prototyping solutions for power generation control systems before transitioning to semiconductors. He now supports tool customization for existing product lines in addition to engaging in new product design and evaluation.

Adhesive bonding for flexible microLED display assembly

By Goutham Ezhilarasu, Subramanian Iyer [UCLA] Ajit Paranjpe [Veeco Instruments, Inc.] Jay Lee [DISCO Corporation] Frank Wei [DISCO Hi-Tec America, Inc.]

In recent years, there has been growing interest in fabricating high-resolution information displays using micro-scale inorganic light emitting diodes (i.e., microLEDs) as an alternative to organic LED (OLED) or liquid crystal display (LCD) displays. The main reasons for this interest is the remarkable quality of displays made using microLEDs that possess several benefits [1]: 1) microLEDs are based on compound semiconductor material systems like GaN, GaAs or InP that have far superior emission properties like sharper line width, higher quantum efficiencies, and strong luminance exceeding 10^6cd/m^2; 2) Resistant to environmental conditions such as temperature and humidity due to their chemical stability leading to longer operational lifetimes (>100,000 hours); and 3) Ultra-fast response times typically in the nanosecond range because of the high mobility of carriers. Such high-quality microLED displays can have a wide variety of applications ranging from the commercial/defense sectors for augmented reality (AR) and virtual reality (VR) device displays, automobile and heads-up displays (HUDs), to the medical sector for visible light therapy and optogenetics.

In spite of the enormous market potential for microLED displays, the technology is, however, mostly in the research phase [1]. The main reason for this is the difficulty in mass manufacturing microLED displays at competitive costs because of the poor yields in assembly. Unlike OLED materials that can be directly deposited on a target substrate to fabricate the LED device, microLEDs contain a complex stack of compound semiconductor material layers that are grown either epitaxially, or using metal organic chemical vapor deposition (MOCVD) on lattice-matched growth substrates at high temperatures exceeding 900ºC. In order to assemble such devices on a target such as a Si CMOS backplane or a flexible substrate, the prefabricated devices on the growth substrate are to be released from it and then transferred onto the target. Because of the extremely small size and thickness of the individual LED (<100µm side, <7µm thickness), sequential pick and place for assembly is impractical. Therefore, a mass transfer process is generally used wherein a large block of devices from the growth substrate are transferred and assembled onto the target in a massively parallel fashion [2] (**Figure 1**). Before releasing from the growth substrate, the prefabricated microLEDs are attached to a temporary carrier by bonding (metallic or adhesive) or stiction (Van der Waals [VdW] in elastomers) for accurate registration. The release process of the devices from the growth substrate is done either chemically (for GaAs, InP) or optically (for III-N), depending on the material system. After release, a transfer printing process is utilized wherein a "stamp" is used to pick up selected patterns of microLEDs from the temporary carrier to the target substrate for final assembly. For a full color display assembly, this mass transfer has to be repeated three different times to assemble the three colors for each pixel.



Transfer technique	Developer	Schematic of technique
Electrostatic stamp	LuxVue	
Magnetic Stamp	ITRI	
Viscoelastic stamp	X-Celeprint	

Figure 1: Schematic of a typical mass transfer process (left); Summary of various mass transfer approaches in literature (right)

Different methods of transfer printing using different types of stamps have been explored and are summarized in the table in **Figure 1**. Although all the transfer approaches do work in principle, their application to product manufacturing have mostly been stifled [2]. Transfer approaches based on electrostatic forces as developed by LuxVue utilize the electrostatic gripper action of a voltage-driven stamp to attract and pickup microLEDs for transfer. This approach is very sensitive to substrate planarity and surface contamination, and involves a relatively complex process with specialized equipment such as an active stamp. Approaches based on electromagnetic forces also suffer from similar drawbacks of the electrostatic case in addition to limited scalability to handle very small microLEDs (<10µm pitch). Mass transfer using viscoelastic stamps is the most popular technique reported in literature, and is being explored for commercialization by the startup XDC.

The viscoelastic mass transfer technique relies on the relatively weak Van der Waals stiction between the stamp and the microLEDs for pickup. They are, therefore, not very reliable in holding the devices in place during the transfer printing process as a small percentage of devices could shift or even fall off during the transfer. The viscoelastic property of the elastomeric stamp also causes it to exhibit peel-rate dependent interfacial adhesion, a property that is exploited to allow the same polydimethylsiloxane (PDMS) stamp to both pick up devices from the source using a higher pull rate and release the devices to the target using a lower pull rate. The complex transfer physics involved makes

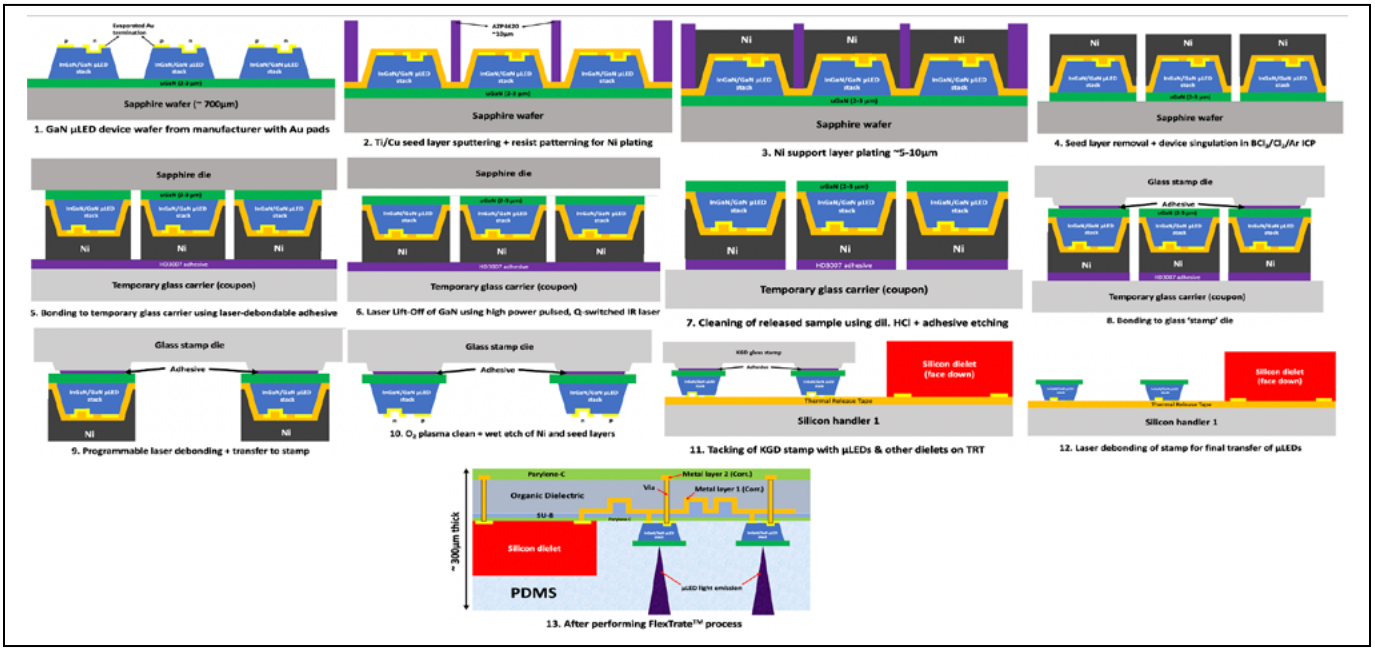


Figure 2: Full process and assembly flow for mass transfer using adhesive bonding.

this technique less attractive as careful fine tuning of process conditions is required to attain substantial yields. It also requires the use of specialized equipment, which may not be available in most outsourced semiconductor assembly and test suppliers (OSATS) or packaging facilities.

For a mass transfer process to be commercially viable, transfer yields of >99.9999% (six 9s of yield or <1ppm failure) are to be achieved as even a few dead pixels in a commercial display are unacceptable. The mass transfer techniques explored in this summary have demonstrated around 99.99% yield, which is impressive, however falls short of the stringent yield requirement of > six 9s for commercialization. Even though using post-assembly repairs and redundant pixels has been suggested to overcome this yield issue, such solutions are not economically viable given the high cost and time of repairs and material cost of the microLEDs.

To overcome the yield issues and other drawbacks, we are currently developing a transfer process using thermoplastic adhesive bonding and programmable-laser debonding to selectively pick up and mass transfer InGaN/GaN MQW microLEDs grown on c-plane sapphire substrates [2]. A thermoplastic, laser-debondable polyimide based adhesive (HD3007) is used to attach the microLEDs to a temporary glass carrier before the laser lift-off (LLO) process that releases the devices from the growth substrate. The same adhesive, deposited on a lithographically-patterned glass stamp, is used to transfer print selected arrays of programmably-debonded microLEDs from the temporary carrier to the target substrate, which in our case is an ultra-flexible PDMS substrate called FlexTrate™ [3]. Because simple adhesive bonding and laser debonding are used for the mass transfer, process complexity is significantly reduced, and the process can be easily performed at any packaging facility without the need for specialized equipment (only a substrate bonder and laser debonding system are required). The full process and assembly flow are given in **Figure 2**. As strong adhesive bonding, instead of weaker electrostatic/electromagnetic or Van der Waals (VdW) forces is used for mass transfer, potentially higher yields of transfer (<1ppm defect) and finer alignments can be achieved because devices will not fall off or shift during the transfer. Before the substrate release process, the microLEDs are also protected with a 5-10µm electroplated Ni stress buffer to prevent any damage during the LLO process. The LLO process itself is done using a novel DPSS laser system (DFL7560L) developed by DISCO Corporation that uses gentle (fluence per pulse <<1J/cm^2) overlapping gaussian beams (>>10s of Hz repetition rate) with a small spot size to achieve low-stress, gap-free device liftoff [4]. A combination of the diode-pumped solid state (DPSS)-LLO process and the Ni stress buffer layer allows us to attain nearly 100% LLO yield for the InGaN/GaN devices on sapphire.

Figure 3 gives experimental details of the aforementioned mass transfer and assembly process at different critical stages. The devices that were electrically measured while on the stamp – after pickup for final assembly – showed less than 5% degradation in forward current at operating forward voltage of 4V when compared to virgin devices on sapphire indicating that the devices were undamaged by the mass transfer process. The devices are finally printed onto a thermal release tape (TRT) laminated on a 4" carrier wafer. This tape also contains other components like Si dielets flip-chip bonded on it. A die-first fan-out wafer-level packaging (FOWLP) process called FlexTrate™ is then carried out, which involves compression molding of PDMS to reconstitute the packaging substrate with embedded microLEDs and dielets, followed by fabrication of metal interconnects using a back end of line (BEOL) Cu plating process and surface passivation using Parylene-C.

Summary

Commercialization of microLED display technology requires the development of mass transfer approaches that are cost effective, high yield, and easy to implement. Current mass transfer approaches in literature and

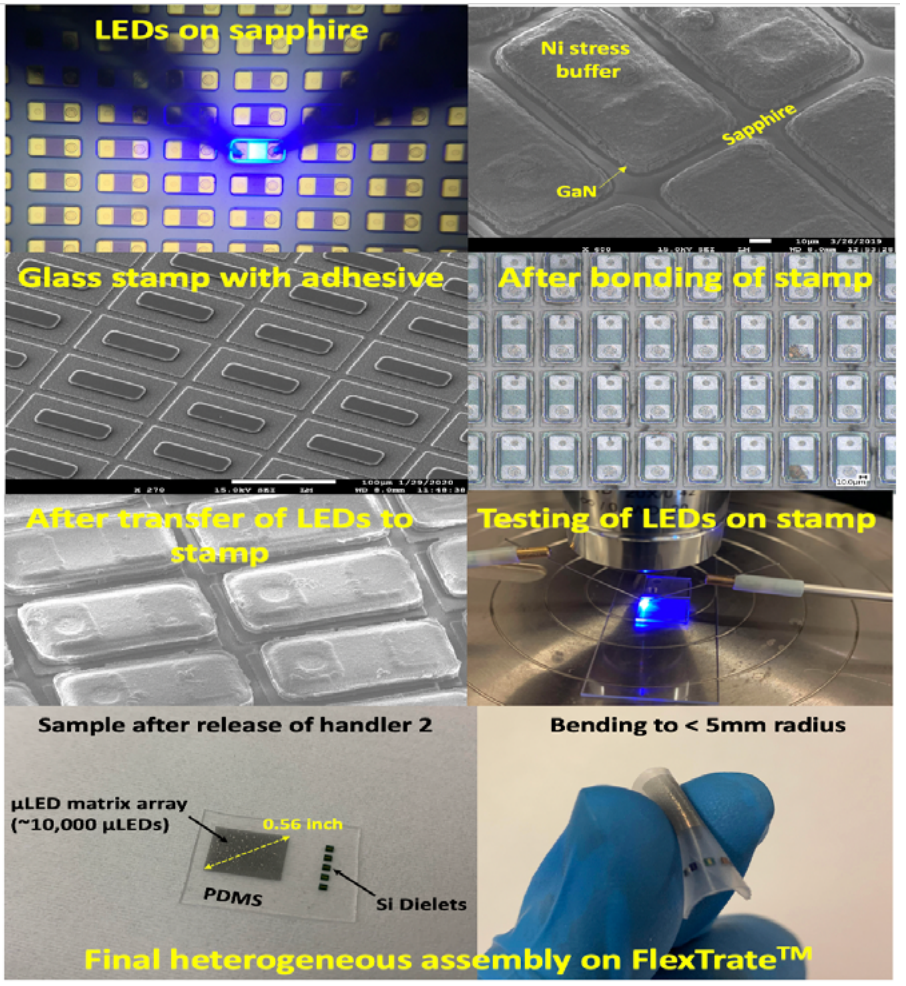


Figure 3: Experimental details of the process shown in Figure 2.

industry are relatively expensive, require specialized equipment, and have been demonstrated to attain yields of ~99.99%, which falls short of the commercially viable yields of six 9s. At the moment, we have demonstrated the assembly of blue InGaN/GaN microLEDs (50µm X 100µm) at >200PPI densities on our PDMS substrate without metallization. In the near future, we plan to demonstrate a fully functional passive matrix display on FlexTrate™ with heterogeneously integrated microLEDs and Si CMOS driver integrated circuits (ICs) with two levels of electroplated Cu metallization at <40µm interconnect pitch.

References

1. K. Ding, et al., "Micro-LEDs, a manufacturability perspective," Applied Sciences 2019, 9, 1206.
2. G. Ezhilarasu, et al., "High yield precision transfer and assembly of GaN µLEDs using laser-assisted micro transfer printing," 2019 IEEE 69th Elec. Comp. and Tech. Conf. (ECTC), Las Vegas, NV, USA.
3. G. Ezhilarasu, et al., "A heterogeneously integrated, high-resolution and flexible inorganic µLED display using fan-out wafer-level packaging," 2020 IEEE 70th ECTC, Orlando, FL, USA.
4. DISCO Technology Press Release, Dec. 2015. https://www.disco.co.jp/eg/news/press/20151207.html

Biographies

Goutham Ezhilarasu is a graduate student researcher at the Center for Heterogeneous Integration and Performance Scaling, Department of Electrical Engineering, U. of California, Los Angeles, CA USA. Email: goutham93@ucla.edu

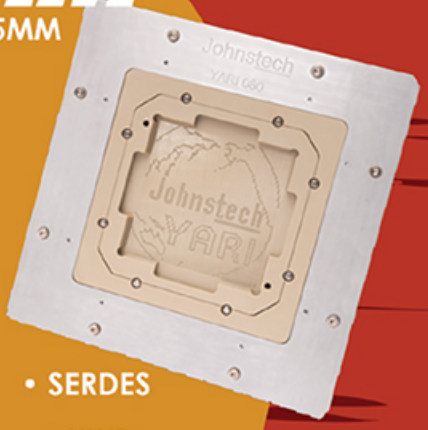
Johnstech®

CONTACTOR / SPRING PROBE APPLICATIONS

SHOTO™
3MM

- Wi-Fi
- BLUETOOTH
- LTE (5G)

YARI™
4.5MM



- SERDES
- UWB
- GPS

Crown Tip
BGA & WLCSP

EXPERIENCE A CUT ABOVE

- Rapid Delivery Enabling Faster Time to Market
- High-Quality Test Solutions
- Extensively Experienced Field Service and Support

Spear Tip
QFN & LGA

FOR MORE INFO, VISIT JOHNSTECH.COM
JOHNSTECH.COM/PRODUCT-SERIES/SPRING-PROBE/

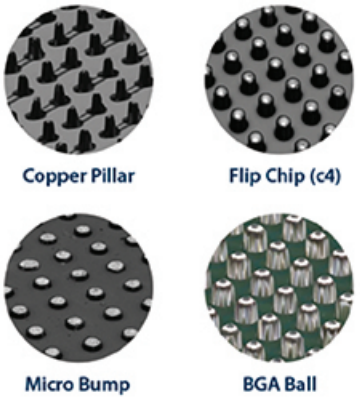
CERTIFIED ISO 9001:2015 CERTIFIED

Seriously Fast.

WX3000™ Metrology and Inspection Systems for Wafer-Level and Advanced Packaging

Powered by Multi-Reflection Suppression™ (MRS™) Technology

- **Improve Yields**
  High-Resolution & High Accuracy.
- **Increase Throughput**
  2-3X Faster Than Alternative Solutions.
- **Improve Processes**
  100% 3D and 2D Inspection in a Single Pass.



2-3X Faster with High Resolution and High Accuracy

WX3000 3D and 2D metrology and inspection system provides the ultimate combination of high speed, high resolution and high accuracy for wafer-level and advanced packaging applications to improve yields and processes.

Powered by Multi-Reflection Suppression (MRS) Sensor Technology
The 3-micron NanoResolution (X/Y resolution of 3 micron, Z resolution of 50 nanometer) MRS sensor enables metrology grade accuracy with superior 100% 3D and 2D measurement performance for features as small as 25-micron.

100% 3D and 2D metrology and inspection can be completed simultaneously at high speed (25 300mm wafers/hour and 55 200mm wafers/hour) as compared to a slow method that requires two separate scans for 2D and 3D, and only a sampling process.

CYBEROPTICS®

www.cyberoptics.com