Future Fab

INTERNATIONAL

ISSUE 43 – October 2012 – www.future-fab.com

In Association with:

leti · ITRS · College of Nanoscale Science & Engineering, University at Albany · GSA · SIA Semiconductor Industry Association · imec · Si Silicon Integration Initiative, Inc. · SEMATECH Accelerating the next technology revolution. · accellera Systems Initiative · Fraunhofer Mikroelektronik · IMECAS · SMTA Surface Mount Technology Association · A*STAR Institute of Microelectronics

Welcome ...

It's now been four years since the bottom fell out of the world's economic base and we suffered the worst economic downturn since the Great Depression. It's hard to believe it's been that long, but in the complex world of semiconductors, that time has seen the progression of two more technology cycles; Facebook passing 1 billion users; the rise of the tablet alongside the proliferation of the smartphone; and a new tech war between global technology superpowers that are suing each other as they encroach further and further into each other's sphere of influence.

Oddly enough, all of the aforementioned was driven by you – yes you, reading this right now. If you are a regular reader of this journal, then you are engaged or affiliated with the semiconductor industry in some form, and we just wanted to take a moment to say, "Well done; keep it up." It's been our honor to help you in whatever manner that is: to be effective, to think and do what you do. We know we're a small part of the ecosystem – but we're proud of that and hope you are too.

Enjoy!
The Future Fab Team

FUTURE VISIONS & CURRENT CONCERNS

DESIGN IMPLEMENTATION & PROCESS INTEGRATION

MANUFACTURING: FAB, SYSTEMS & SOFTWARE

LITHOGRAPHY LANDSCAPE

FRONT END OF LINE

METROLOGY, INSPECTION & FAILURE ANALYSIS

WAFER FAB & PACKAGING INTEGRATION

ASSEMBLY, TEST & PACKAGING TECHNOLOGIES

Biographies of Future Fab's Panel Members
For the full versions of the following biographies, please click here.

Welcome to our new Panel Member – Surya Bhattacharya

Surya Bhattacharya
Director, Industry Development; IME

Dr. Surya Bhattacharya is director, Industry Development, for IME's Through Si Interposer (TSI) program. He has over 18 years of experience ranging from 0.8 micron to 28 nm CMOS while working in the U.S. semiconductor industry at both fabless companies and integrated device manufacturers. Surya obtained his Bachelor of Technology degree in electrical eng. from the Indian Institute of Technology Madras, India in 1987, and his M.S. and Ph.D. in microelectronics from the University of Texas at Austin in 1993.

Paolo A. Gargini
Director of Technology Strategy for Intel Corporation

Dr. Paolo Gargini is also responsible for worldwide research activities conducted outside Intel for the Technology and Manufacturing Group by consortia, institutes and universities. He received doctorates in electrical engineering and physics from the Universita di Bologna, Italy.

Michel Brillouët
Senior Adviser, CEA-Leti

Michel Brillouët joined CEA-Leti in 1999, where he managed the silicon R&D. Prior to joining CEA-Leti, he worked for 23 years in the Centre National des Télécommunications (France Telecom R&D Center), where he held different positions in micro-electronics research. He graduated from École Polytechnique.

Alain E. Kaloyeros
Senior Vice President, CEO and Professor, College of Nanoscale Science and Engineering; University at Albany

Alain E. Kaloyeros has authored/co-authored over 150 articles and contributed to eight books on nanoscience, holds 13 U.S. patents, and has won numerous academic awards. He received his Ph.D. in experimental condensed matter physics from the University of Illinois, Urbana-Champaign, in 1987.

Gilbert J. Declerck
Executive Officer, imec; Member of the Board of Directors, imec International

Gilbert J. Declerck received his Ph.D. in electrical engineering from the University of Leuven in 1972. He has authored/co-authored over 200 papers and conference contributions. In 1993, he was elected fellow of the IEEE. Since July 1, 2009, Dr. Declerck has been executive officer imec and a member of the board of directors of imec International.

Didier Louis
Corporate and International Communication Manager, CEA-Leti

Louis joined CEA-Leti (France) in 1985, where he received a Ph.D. in metallurgy/electrochemistry from the University of Grenoble. He has written more than 30 papers related to etching and stripping processing and has co-authored more than 60 scientific papers and eight patents.

Pushkar P. Apte
President, Pravishyati Inc.

Dr. Pushkar P. Apte's strategy consulting firm focuses on the high-tech industry. He received his master's and Ph.D. from Stanford University in materials science and electrical engineering, and his bachelor's degree in ceramic engineering from the Institute of Technology, Varanasi, India.

Christian Boit
Head, Semiconductor Devices at Berlin University of Technology, Germany

The Berlin University of Technology is an institution for research and development in the areas of device simulation, technology, characterization and reliability. Christian Boit received a diploma in physics and a Ph.D. in electrical engineering on power devices, then joined Siemens AG's Research Laboratories for Semiconductor Electronics in Munich and has been a pioneer on photoemission.

Jiang Yan
Professor, IMECAS

Dr. Jiang Yan has authored and co-authored over 40 papers, holds 17 U.S. patents and 10 China patents. He received his Ph.D. in electrical engineering from the University of Texas at Austin in 1999.

Daniel C. Edelstein
IBM Fellow; Manager, BEOL Technology Strategy, IBM T.J. Watson Research Center

Dr. Edelstein played a leadership role in IBM's industry-first "Cu Chip" technology in 1997, in the introduction to manufacturing of Cu/Low-k insulation in 2004. He received his B.S., M.S., and Ph.D. degrees in applied physics from Cornell University.

Sitaram R. Arkalgud
Director of Interconnect, SEMATECH

Sitaram R. Arkalgud has over 20 years of R&D and manufacturing experience within the chip industry. He has a Ph.D. and a master's degree in materials engineering from Rensselaer Polytechnic Institute, and a B.S. in metallurgical engineering from Karnataka Regional Engineering College, Surathkal, India.

Peter Rabkin
Director of Device & Process Technology, SanDisk Corp.

Dr. Peter Rabkin focuses on development of novel 3D memory technologies and products. He holds a master's degree in physics from Tartu University and a Ph.D. in physics of semiconductors from the St. Petersburg Institute of Physics and Technology.

Giuseppe Fazio
Advanced Process & Equipment Control Sr. Engineer; Micron Semiconductors Italy

With a laurea degree in applied physics from Milan University, Giuseppe has working experience in several sectors, from research to industry, and vast experience in industrial and scientific instrumentation. He has authored/co-authored many articles, is an avid contributor at conferences and holds several patents in the semiconductor field.

Janice M. Golda
Director, Lithography Capital Equipment Development; Intel Corp.

Janice Golda manages an organization responsible for creating strategies and working with Intel's lithography, mask and metrology suppliers and subsuppliers to deliver equipment meeting Intel's roadmap technology, capacity and cost requirements. She is a member of the Berkeley CXRO Advisory committee, is Chairman of the Board for the EUV LLC and holds one U.S. patent.

Warren Savage
Chief Executive Officer, IPextreme

Warren Savage has spent his entire career in Silicon Valley, working with leading companies, where he focused on building a global scalable semiconductor IP business. In 2004, he founded, and still leads IPextreme in the mission of unlocking and monetizing captive intellectual property held within semiconductor companies and making it available to customers all over the world. He holds a B.S. in computer engineering from Santa Clara University and an MBA from Pepperdine University.

Klaus-Dieter Rinnen
Director/Chief Analyst, Dataquest

Klaus-Dieter Rinnen is director for Dataquest's semiconductor and electronics manufacturing group. He received a diploma degree in physics with minors in physical chemistry and mechanical engineering in Germany, and a Ph.D. in applied physics from Stanford University.

Luigi Colombo
TI Fellow

Dr. Luigi Colombo is a TI Fellow working on the Nanoelectronic Research Initiative (NRI). He is author and co-author of over 130 publications, three book chapters, and holds over 60 U.S. patents. Dr. Colombo received his Ph.D. in materials science from the University of Rochester.

John Schmitz
SVP & General Manager, Intellectual Property and Licensing; NXP

John Schmitz holds a master's degree in chemistry from Radboud University of Nijmegen, Netherlands, and a doctorate in physical chemistry from Radboud University Nijmegen. He has authored more than 45 papers in various scientific journals and has written books on IC technology and on thermodynamics.

Lode Lauwers
Director Strategic Program Partnerships for Silicon Process & Device Technology, imec

Lode Lauwers has an M.S. in electronics engineering and a Ph.D. in applied sciences. He joined imec in 1985 as a researcher. Lode manages imec's core partner research program on sub-32nm CMOS technologies.

Yayi Wei
Principal Member of Technical Staff; GLOBALFOUNDRIES

Dr. Wei investigates advanced lithography processes and materials. He has over 16 years of lithography experience, including DUV, 193 nm, 157 nm, 193 nm immersion, EUV and E-beam lithography. Dr. Wei has numerous publications and holds several patents in the field of lithography.

Click here to return to Table of Contents

Future Fab International

Publisher and CEO . . . . . | Nikki Wood

Editorial & Digital Projects Director . . . . . | Matt Grimshaw

Art Director . . . . . | David Witcomb

Editor . . . . . | Lisa Barr

Marketing Manager . . . . . | Lindsay Poole

Web . . . . . | Sara Smith

Production Manager . . . . | Jane Ellis

Published by:
Mazik Media Inc.
38 Miller Ave., Suite 9
Mill Valley, California 94941
USA

Tel: 415 381 6800
Fax: 415 381 6803

Email: converse@mazikmedia.com

ISSN: 1363-5182

Click here to return to Table of Contents

Pushkar P. Apte
President, Pravishyati Inc.

The semiconductor industry has been an amazing source of industrial innovation in recent history. Starting with the discovery of the transistor in Bell Labs in the 1940s, progressing through the development of the integrated circuit in the 1960s, the semiconductor industry has grown and flowered into a $300 billion juggernaut. In the process, it has driven incredible innovations that make the mind boggle – such as cramming a billion or more electronic transistors on a thumbnail, decoding the human genome and enabling people to communicate verbally, in pictures, and in video almost anytime, anywhere. For the first few decades, the industry relied primarily on geometric scaling – making stuff smaller. The new millennium brought a millennial shift, requiring many more innovations such as the shift to high-k dielectric materials and the move from aluminum to copper for on-chip interconnects. But this is just the beginning – future innovations in materials and device structures will be even more exotic, involving fundamental shifts like using photons to exchange information instead of electrons, and perhaps new substrates like graphene, instead of the old warhorse silicon.

While no one knows the exact nature and speed of these advances, it is abundantly clear that they won't be cheap. The semiconductor industry spends more on R&D than almost any other industry, and the spend is increasing rapidly, outpacing revenue growth. This is simply not sustainable – as the memorable line goes, "It's economics, stupid." Sustaining this blistering pace of innovation will require new models of pre-competitive collaboration across the entire supply chain in industry, and with academia, research institutes and the government. New innovation platforms and processes must develop – they may entail an unprecedented level of trust, information exchange and "co-opetition."

The two papers in this section describe different elements of this challenge – the first describes the global research institute imec's perspective on developing a rational intellectual property environment that rewards fairly but does not stifle the needed flow of ideas; while the second, from Lam Research, describes the breadth of innovations needed and exciting research programs under way at leading research universities like UC Berkeley, Stanford and MIT.

Optimizing an IP Portfolio Over an R&D Life Cycle

Kathleen De Belder, Sigrid Gilis, Vincent Ryckaert
Imec

As an independent research institute, imec performs world-leading research in nanoelectronics and nanotechnology, thereby creating new technologies and inventions all over the R&D life cycle. The R&D work covers both fundamental as well as late stage research, generating a diverse patent portfolio. Imec's intellectual property (IP) business model tries to leverage and increase the value of these IP assets.

Introduction to imec and Its R&D Life Cycle

Imec was founded in 1984 as a non-profit organization with the goal of strengthening the microelectronics industry in Flanders. The decision of the Flemish government to set up an R&D institute in the field of nanoelectronics was inspired by the strategic importance of microelectronics for the industry and by the major investments required to keep up with developments in this field. Nowadays, imec performs world-leading research in nanoelectronics. It leverages its scientific knowledge with the innovative power of its global partnerships in ICT, health care and energy. Imec delivers industry-relevant technology solutions. In a unique high-tech environment, its international top talent is committed to providing the building blocks for a better life in a sustainable society.

As an R&D center, imec creates many new technologies and inventions all over its R&D life cycle. The R&D work and efforts cover both early-stage research (fundamental research) as well as developing designs on demands or even performing low-volume manufacturing.

Figure 1 shows a global R&D platform. Early-stage technologies are typically developed in or with academia. These institutes are located on the high left of this platform where R&D costs are low and where most of the time, funding comes from the governmental. There is more time for doing research, and many options are left open. On the other hand, industrial players are located on the lower right side, where, typically, R&D costs are much higher. Next to this, there is only a short time period for doing development, due to the fact that the product has to enter the market within a fixed time frame.

Imec is situated in between the aforementioned areas, a position sometimes called "the valley of death." At this stage, governmental funding is exhausted, and the industry is still reluctant to invest money in too-immature technology. Therefore, imec has created a tunnel to bridge this gap by creating an R&D platform. This platform offers the opportunity to cooperate with industrial partners, but

to simultaneously continue the research activities of academia to prepare the technology for an industrial framework.

In this position, imec has the opportunity and potential to build up an IP portfolio across all the different phases of its R&D life cycle, covering several different IP stacks. This has resulted in a patent portfolio of about 1,025 families (translated into 2,850 patents and patent applications in Europe, the U.S., Japan, etc.). Almost half of the intellectual property rights (IPR) in this portfolio are co-owned by an industrial or academic partner. The IP stack covers materials, equipment, process steps, process results, devices, circuitry, systems and software.

Similar to its R&D activities, imec tries to gain value out of the IP portfolio in every stage of the R&D life cycle.

Imec's Business Model

Before going into detail in the monetization process, imec's business model needs some introduction.

Imec's business model is an open, collaborative model primarily based on non-exclusive licenses. The key features of the business model are: 1) avoiding IP blocking for our partners; 2) securing IP rights for our partners; and 3) enabling publications. The tools to facilitate this model consist of a selective patenting model and a carefully designed IP model. In every stage of the R&D life cycle, an open multi-partner model is promoted wherein the costs and IP are shared among the partners.

Figure 2 lists the R&D life cycle, going from funded research (early stage), to applied research projects, bilateral agreements toward development on demand, low-volume manufacturing, IP licensing and even selling IP. Arrows A and B illustrate the possible routes for handling the IP.

When early-stage IP has been created, typically, two routes are to be considered in the handling of IP. Route A is selected to use early-stage IP as background for (follow-up) projects, in bilateral collaborations (where new IP can be generated based on this background) and later on as background (with or without IP created halfway in the life cycle) in late life-cycle activities. Typically, imec grants non-exclusive use and exploitation rights to make, sell and offer to sell semiconductor products. In other words, a technology transfer is carried out via the residents of the partners who are collaborating jointly in the research programs. There is the possibility creating joint IP that will be on a without-accounting basis (i.e., IP can be used by each contributing party without financial accounting to another involved party or without asking consent).

The benefits of this model are twofold: The cost of R&D is being shared (costs for infrastructure, people, experience); and the available funding mechanism can be used for long- or midterm research (not stopping after academic R&D, no end in the valley of death). Furthermore, there is a well built-up IP portfolio when a product enters the market. There might be some arguments against this ecosystem, being: the need to have an agreement on the prosecution of IP; some loss of value due to the

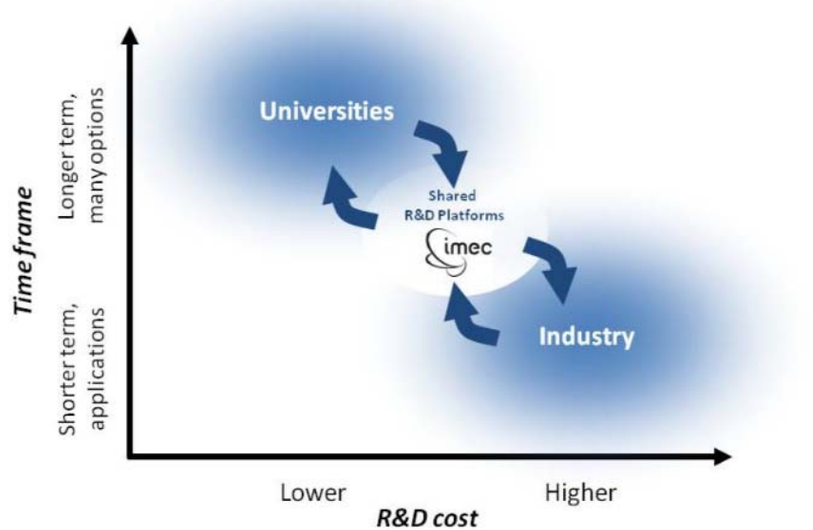


Figure 1. Representation of R&D Life Cycle: Time Frame for Doing Research vs. the R&D Cost

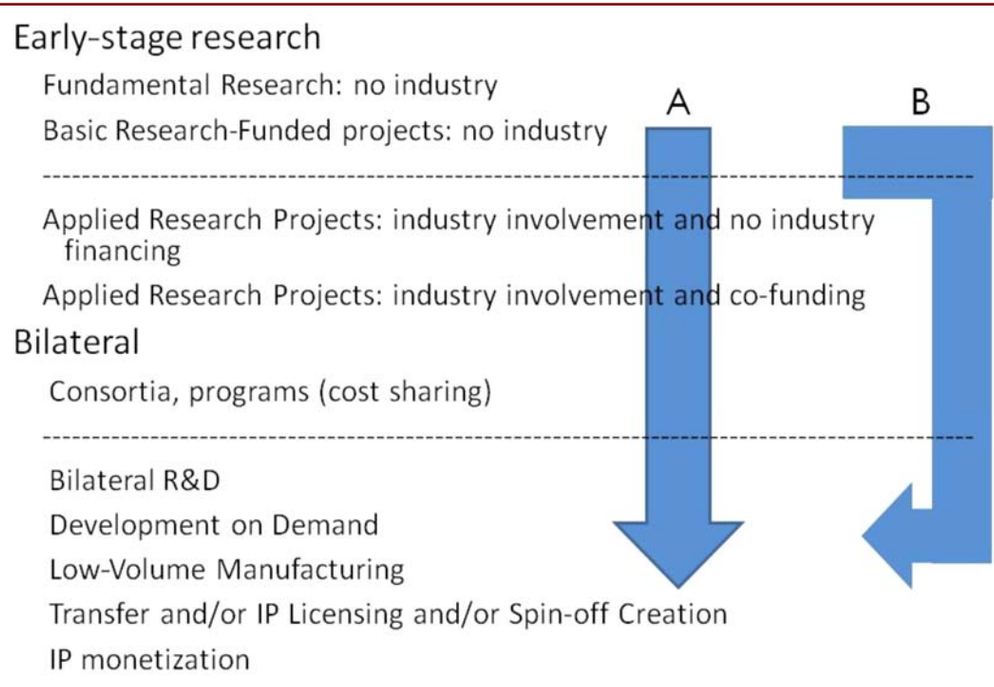


Figure 2. Possible Routes for Handling of IP Over the R&D Life Cycle

non-exclusivity of the system; and/or no rights to sue via a licensing system.

Imec manages its IP via one-to-one contracts (it is not a joint venture). The use of co-ownership without accounting avoids co-management. By following this path, the IP is fully exploited.

In some cases, it might be preferable to reserve the early-stage IP in order to be able to license or transfer it on an exclusive basis later. This is illustrated by route B in Figure 2.

Selecting this path also has its consequences; namely, a) financial means and critical mass are low; b) the entry into the market of the technology is risky; c) IP has to be reserved; d) the upturn for exclusive post-research exploitation is higher; and e) mutual expectations are to be matched by avoiding sitting on IP, by evaluating the situation frequently and by being prepared to switch to route A.

Academia should not only look for direct valorization of its IPR (via transfers or exclusive licenses), but should consider setting up a business framework wherein IP that is being created or acquired may be supported by licenses from other academia. This model of participation and collaboration between academia and/or other interested parties might be a way to unlock value for early life-cycle IP.

In both cases or routes, selling the IP to the secondary market should be the last option/resort for valorizing IP. However, before doing this, some aspects must be taken into consideration:

- Can the ecosystem be secured? Which parties can be approached? What are the possible risks or liabilities?
- Regarding business policy: Is this approach fair toward current partners? Can a submarine approach be avoided? Should the IP be offered to partners first?

Creating and Optimizing Additional IP Value Throughout Monetization

Besides taking one of the above-mentioned routes, why not jump on the train of IP monetization? As current success stories show, monetization of IP is a booming and upcoming market. It involves the transaction of a (granted) patent or patent family, without know-how, especially to non-practicing entities, hence for use of the IP right embedded in the patent as such. This rising trend of gaining additional revenues and more value out of IP through monetization is due to the failure of the early-stage IP market. In Europe, this market hardly exists. Besides that, this market failure is also caused by the lack of trust between the different R&D players and academia and/or by a mismatch of their business models.

The first experiment with IP monetization for imec was with an older patent family. A first choice was made by working with a defensive patent aggregator. This is a party that tries to keep the patents out of the hands of non-practicing entities (NPEs) that would assert them against companies that are likely to infringe these patents. Instead of fostering patent litigation, defensive patent aggregators operate as clearinghouses and try to set up a licensing deal with third companies.

Imec learned the following from this IP monetization experiment. Firstly, an efficient internal preparation is needed and a selection needs to be made of patent families that are without or have only minor encumbrances, e.g., having all assignments in place in case of co-owned patents.

A second important issue is to find out how one can evidence product infringement.

Thirdly, one needs to value the patent. This is a difficult exercise for early-stage IP, because there is no existing market yet. However, looking at the following parameters might help:

- Does the patent include know-how?
- Does the patent exist independently of the know-how?
- Is there a market yet?
- What is the stage/status of the patent? Of the know-how?
- Is the patent part of a patent cluster?
- What is the scope of the patent and is it technology independent?
- Is the patent linked to a standard?
- Is the patent blocking other technologies?
- Should one simply go for a cost-based approach (plus overhead)?
- Is there a time-to-market advantage?

Furthermore, prepare an offering according to your own business or sales model. The buyer will also perform a due diligence of the offered patent, so expect a lot of questions. Finally, there is the closing of the deal. All documents need to be in place (patent purchase agreement, assignments, documents, translations, etc.).

Currently, the imec IP model is more dynamic than ever before:

1. imec is recognizing and starting to use the IP market evolutions.
2. imec believes it is time to discuss IP in- and out-licensing with other EU research institutes.
   a. To overcome failure of early-stage IP market
   b. To overcome adverse effects of IP monetization
   c. To create true EU collaboration
3. imec is willing to bring its experience with academia to the table to achieve concrete solutions.

About the Authors

Kathleen De Belder

After some years of research at the Vrije Universiteit Brussel, Kathleen De Belder joined imec's patent group in October 2006 as a patent portfolio officer. She has followed several courses in European patent law, due diligence and licensing technology. In 2011, Kathleen moved to the IP Business group. Her work is directed toward giving IP advice in business concepts and contracts, and developing IP models.

Sigrid Gilis

After having obtained her master's degree in law at the Catholic University of Leuven (KUL), Sigrid Gilis became a registered member at the bar of Hasselt in September 2009. She specializes in IP, TMT and ICT law, but is also active in the sector of commercial and economic law and contract law. In June 2012, she completed a complementary post-graduate study in ICT, IP and media law at the KUL, organized by the ICRI. Currently, Sigrid is doing an internship with imec's IP Business group as an independent legal consultant.

Vincent Ryckaert

Vincent Ryckaert is a qualified European patent attorney, receiving further qualification on litigation of European patents. He has also undergone training in U.S. patent, licensing law and patent opinions and has an in-depth expertise in the field of software-enabled patents. Vincent leads imec's IP Business team. His efforts are directed toward IP business and intelligence work. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Click here to return to Table of Contents

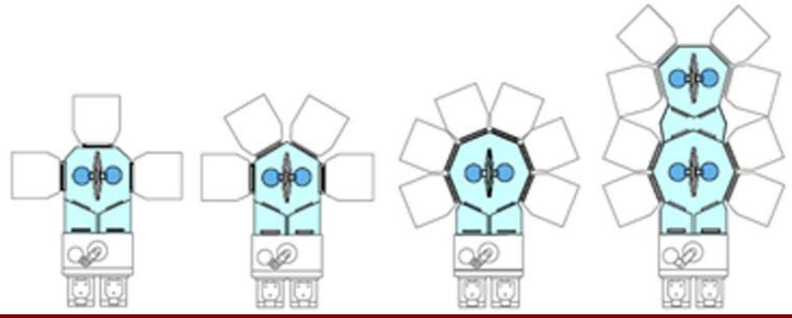
Thoughts on 450 mm Process Integration

450 mm is slowly but surely moving away from the "maybe" column toward the "inevitable" due to the largest four chip manufacturers building a roadmap through the much-publicized Global 450 Consortium (G450C) and other announcements such as the imec 450 clean room in Belgium. All are aiming to assist the rollout of what could be the final wafer transition before CMOS architecture finally runs out of ways to cheat physics and the industry must find new platforms to continue the never-ending consumer thirst for technology-related products.

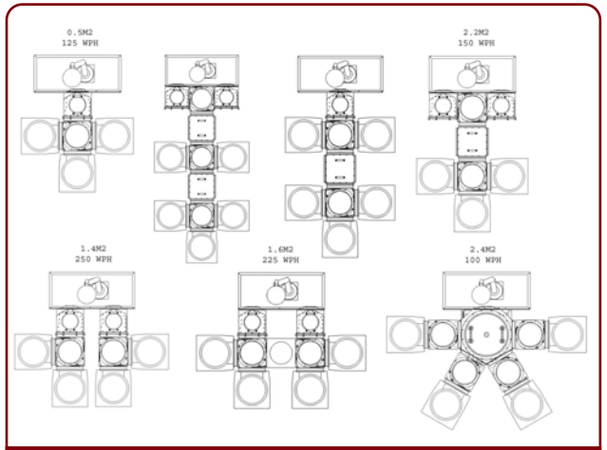
Many of the throughput issues that challenged 300 mm high-volume fabs will be exacerbated by the switch to the 450 mm wafer size; however, there is a plus side to such disruption. Many ideas that had been proposed as solutions in 300 mm have been rejected or remain in limbo due to numerous factors including risk, lack of support, and pushback from a conservative-leaning industry. 450 mm has the opportunity to breathe new life into numerous conceptual technologies due to its near-complete clean-sheet, and therefore have a far greater chance for adoption than they did in the 300 mm space. This is mainly due to issues such as economic scale and complexity, which will force fab designers, OEMs and process integrators to investigate all open avenues in the search for solutions to the huge challenges 450 mm entails.

Many ideas from the last five years – such as platform standardization, whereby traditional single-vendor cluster tools (Figure 1) evolve into standardized/modular multi-purpose cluster tools (Figure 2) – offer solutions to the complexities involved with scaling up the wafer size and tool footprint while maintaining throughput and efficiency. This approach morphs a process line into a single tool or a few tools capable of an entire single layer. Instead of process lines, the process steps become layer clusters featuring multi-process tools with numerous chambers able to complete entire levels of the device.

This is no easy switch. The transition faces numerous challenges due to the nature of the business: Competing vendors will no longer be working with solely the manufacturer, but with each other to settle on a standard platform – an approach many consider unworkable. There are similar examples common in fabs already; the litho cell is just one example. Obviously, there are differences. The litho cell generally doesn't have competing vendors integrated onto a common platform, but the concept is sound, even if the political and infrastructure challenges faced will not be easy.



Today's Common Tools With No Commonality



Sample Platform Layouts

20/20 VISIONS
FUTURE VISIONS & CURRENT CONCERNS

More Than Technical Innovation Required for CMOS Extension

David Hemker
Lam Research Corp.

Complementary metal oxide semiconductor (CMOS) process technology has been the standard in the semiconductor manufacturing industry for more than 40 years. CMOS has advanced at a rapid pace guided by Moore's Law. Approximately every two years, circuit line widths are driven smaller, device performance is improved and cost is reduced, putting more processing power into a handheld smartphone than all of the Apollo space missions combined.

But some are predicting that CMOS will soon reach the proverbial "brick wall" of the laws of physics, where conventional scaling becomes impossible (or at least impractical). If this does happen, the results will not only be felt in the fab, but will also negatively impact global economies that are influenced by demand for the latest and greatest technologies.

So what can be done?

First, consider that the end of CMOS has been predicted many times over the past two decades. People used to think that optical lithography was going to be the end of CMOS scaling – that you couldn't produce patterns smaller than the wavelength of the source illumination. Additionally, there were skeptics who didn't believe it was possible to reliably use deep UV sources in production, let alone 193 nm immersion lithography, double patterning, quadruple patterning or even EUV lithography. But innovation has proven them wrong. Indeed, the industry has continued to invent technology "extenders" that have kept CMOS alive and well.

When device speed improvements could no longer be achieved with traditional methods, the industry invented materials with higher dielectric constants to continue gate oxide scaling, metal gates to reduce resistance, and hybrid materials like SiGe and strain engineering to enhance mobility. To achieve voltage scaling, reduced voltage operation and lower leakage in logic devices, we now have 3D FinFET structures replacing planar structures. In memory devices, planar NAND is being replaced by 3D NAND to improve device densities, and DRAMs are migrating to buried bit lines and buried word lines. To emphasize the amount of change that is taking place in the industry, consider this point: Before 2000, there were 19 periodic table elements being used in semiconductor manufacturing. Today more than 60 materials are either in use or under evaluation.

Based on our customers' roadmaps, virtually all semiconductor manufacturing will still use CMOS at the end of this decade. But this does not mean that CMOS will be around forever. Despite the industry's inventiveness, there are things that could ultimately spell the end of the road for CMOS. The first of these is the transistor itself. Preventing transistor leakage on ever-smaller devices is a significant challenge that will need to be addressed. Failing to do so will result in larger end-products or shorter battery life. Another possible disruptor to CMOS is the copper interconnect. Shrinking interconnects increases the electrical resistance in copper wiring, which in turn reduces performance. The industry is looking at alternative metals, new deposition processes, and graphene and carbon nanotubes as possible solutions to this issue. Perhaps the most likely factor to drive CMOS to an end is cost. Even if technology challenges are overcome, the process may be too expensive for scaling to remain economically feasible.

So, what might a successor to today's CMOS technology look like? Academia and

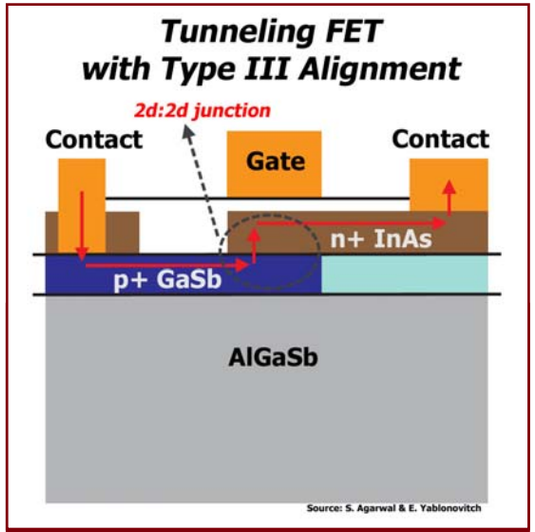


Figure 1. Nanoelectronics tunnels electric current through a barrier.[1]

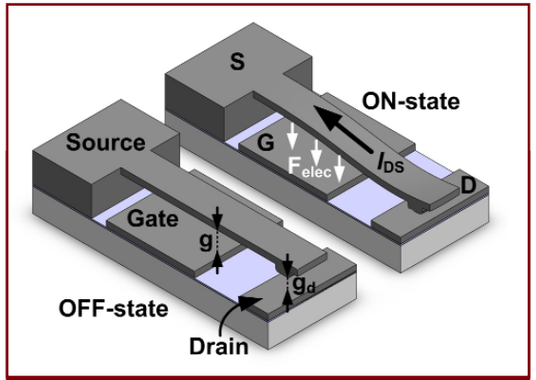


Figure 2. Nanomechanics replaces transistors with tiny switches.[1]

industry around the world are exploring several options. The first is nanoelectronics (Figure 1). This approach proposes replacing transistors with a new switching mechanism that tunnels electric current through a barrier. It would significantly reduce device power consumption, but faces significant challenges around band-edge sharpness and alignment.

The second approach is nanomechanics (Figure 2). This replaces transistors with tiny switches that open and close mechanically and have no "off-state" current. This is great for power consumption, but it also has significant challenges related to reliability, manufacturability and device design.

A third approach is nanophotonics (Figure 3) – a process that uses light, rather than electrons, to function. The good news about nanophotonics is that it uses very little energy and has significant room for growth before reaching its theoretical limits. Challenges to this method include processing hybrid materials (III-V) that require highly toxic raw materials, as well as processing silicon and III-V materials on the same substrate.

The fourth approach is nanomagnetics (Figure 4). This energy-efficient method leverages the natural spinning movement of electrons, rather than their charge, to store and process data. IBM recently demonstrated that it is possible to control and manipulate the spin of electrons using nanomagnetics (e.g., spintronics). Key challenges to this approach include processing mostly nonvolatile novel compounds, reliability and compatibility with silicon wafer fab processes.

While significant innovation is required to overcome the challenges with each of these technologies, the good news is that academia and industry are already working together well in advance of when they may be commercialized. The Center for Energy Efficient Electronics Science (E3S) at the University of California at Berkeley is an example of a public-private consortium that is focused on evaluating each of these approaches. The group, funded by the National Science Foundation, is collaborating with other academic institutions including the Massachusetts Institute of Technology, Stanford University and others, as well as corporations, including Lam Research.

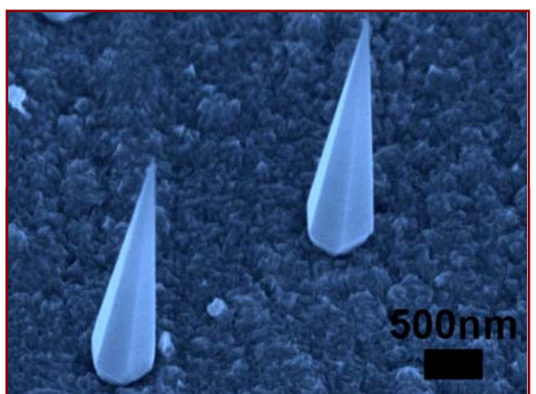


Figure 3. Nanophotonics uses light rather than electrons to power devices.[1]

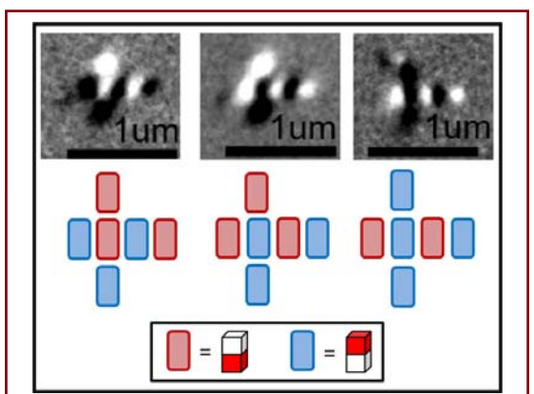


Figure 4. Nanomagnetics uses the natural spinning movement of electrons to store and process data.[1]

While it is not clear which technology, if any of these, will win out, five of the world's leading semiconductor industry associations have jointly outlined eight criteria they consider essential to a CMOS successor. Published in the "International Technology Roadmap for Semiconductors" (ITRS) in 2009, these criteria include scalability, incremental performance improvements, energy efficiency, an OFF/ON ratio for memory devices to minimize power dissipation, gain (for logic devices), operational reliability, room-temperature operation with tolerance for higher internal-device temperatures, and, perhaps most importantly, both technological and architectural compatibility with CMOS.[2]

Between now and the end of CMOS – whenever that is – we can expect that the industry will continue on its rapid pace of innovation. By the end of this decade, chipmakers will be manufacturing at least some capacity in volume on 450 mm wafers. The leading-edge manufacturing node will be 7 nm, and 10 nm will be in full production. Extreme ultraviolet (EUV) lithography will be in play, and double- and quad-patterning of EUV-treated wafers will follow. New processes will be explored, including directed self-assembly, which uses phase separation of polymers to create patterned wafers as a potential alternative to lithography, or more likely to augment optical lithography.

Wafer fabrication equipment will continue to evolve and become even more sophisticated in 2020, with more sensors and "intelligence." Some tools may even be self-diagnosing and able to restock their own consumables. New approaches to deposit and remove many of the new materials may be required, and even new ways of building structures may emerge.

No matter how we get there, one thing is certain: The extension of CMOS is getting harder and more expensive. Innovation – not just in technology, but in how we partner and collaborate – is essential to overcome the financial burden of developing next-generation technologies. Early collaboration with semiconductor manufacturers, fabless companies, peers, suppliers, research consortia and universities is essential not only to reducing risk but also to efficiently achieving the level of technical innovation that is required in the future. Lam Research is actively pursuing new innovation models to help our customers extend CMOS for as long as possible, and we are investigating the successors to CMOS so these technologies are ready when our customers need them.

Endnotes

1. K. Ganapathi et al. “Center for Energy Efficient Electronics Science” website, University of California at Berkeley, http://www.e3s-center.org/research/index.htm#theme1-sec, Aug. 2012.
2. “Emerging Research Devices,” International Technology Roadmap for Semiconductors, http://www.itrs.net/Links/2009ITRS/2009Chapters_2009Tables/2009_ERD.pdf, 2009.

References

- Kartik Ganapathi et al. “Recent Simulation Results on TFET Structures: Transport, Limitations on ON Current, Ballistic Leakage, and Structural Optimization,” Talk or presentation, Center for Energy Efficient Electronics Science, Jan. 27, 2011.
- Youngki Yoon et al. “Scaling Study of Graphene Transistors,” 2011 11th IEEE International Conference on Nanotechnology, August 2011.
- Giovanni De Micheli, “Are We Reaching the End of CMOS Scaling Limits?” EE Times, May 22, 2000.
- Matt Grimshaw, “Chip Watch: The End of CMOS and Its Effect on the World,” Silicon Valley Watcher, Aug. 28, 2009.
- “IBM Scientists Waltz Closer to Using Spintronics in Computing,” IBM Research and the Solid State Physics Laboratory at ETH Zurich, http://www-03.ibm.com/press/us/en/pressrelease/38566.wss, Aug. 12, 2012.
- Tsu-Jae King Liu et al. “The Relay Reborn,” IEEE Spectrum, April 2012.
- A. Barman et al. “Size Dependent Damping in Picosecond Dynamics of Single Nanomagnets,” Applied Physics Letters, vol. 90, Issue 20, May 17, 2007.

About the Author

David Hemker

Dr. David Hemker is chief technology officer at Lam Research Corp., in Fremont, Calif., where he focuses on corporate technology development. He joined Lam Research in 1998 and was named CTO in 2012. Previously, Dr. Hemker served as vice president of New Product Development. Prior to joining Lam, he was vice president of technology for PMT/Trikon, where he was responsible for research and development, technology and engineering. Dr. Hemker also spent five years in various technology roles at Applied Materials. He received his doctorate and master’s degrees in chemical engineering from Stanford University, and his bachelor’s degree in chemical engineering from the University of Wisconsin. Dr. Hemker holds over 35 patents and has authored numerous technical publications in the areas of semiconductor processing and thin film applications. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Steven E. Schulz
President and CEO; Silicon Integration Initiative, Inc.

It seems that nearly the entire semiconductor landscape is becoming limited by power and energy management, where this topic has overtaken timing performance to satisfy battery life, thermal constraints, reliability or global "green" resource management.

At one time, a majority focus was placed on the main culprit, dynamic (sometimes called switching) power. However, at smaller process nodes, leakage competes with dynamic power and can even dominate total power consumption. While every new process node brought with it relative increases in leakage, it could be ignored as a small portion of total power consumption, so emphasis was placed on design techniques to provide large dynamic power savings. Now, however, leakage has emerged as a serious limiter to continued semiconductor device integration.

In the following paper from ARM Fellow Jean-Luc Pelloie, the author makes the modern-day case for fully depleted SOI (FDSOI) devices. Mr. Pelloie takes a designer's perspective, assessing the "pros and cons" of FDSOI along with bulk CMOS and other alternatives such as FinFETs. The author makes several excellent points regarding the dependence upon supporting design flow infrastructure necessary to take advantage of FinFETs and FDSOI. As appealing as a new transistor device may seem as a solution to the leakage problem, it becomes clear that it may require much more work and cost in infrastructure changes before it can be effectively inserted and used in new designs.

Novice designers who see their role primarily in terms of “programming and verifying” RTL code to deliver digital functionality into silicon might be initially tempted to consider the details of FDSOI and FinFET outside their immediate scope. That would be a huge mistake. The ability to design in the future will increasingly depend upon optimizing all aspects of an interconnected technology chain (and supply chain). For almost every category of IC in the years ahead, managing power – specifically leakage power – will be a critical success factor.

DESIGN IMPLEMENTATION & PROCESS INTEGRATION

Designing With FDSOI vs. Bulk

Jean-Luc Pelloie
ARM

With gate length scaling down to follow Moore's Law, it is more and more difficult to achieve a good leakage current compatible with low-power applications requirements. This is due to the loss of control of the vertical electric field through the channel of the MOS transistor, which triggers the short channel effects. Until now, in bulk CMOS, the control was maintained by shrinking the gate insulator thickness (gate oxide has now been replaced by high-k insulator) and increasing the doping in the channel region close to the source drain junctions (halo or pocket implants). This methodology is reaching its limits, as increasing the channel doping results in an increase of junction leakage and gate-induced drain leakage (GIDL) currents.

Fully depleted SOI (FDSOI) transistor is not a new device, as it started to be used more than 20 years ago. At that time, it was considered to be the ideal MOS transistor, being able to switch with a 60 mV/decade subthreshold swing at room temperature. FDSOI has not been popular on the manufacturing side in the past, and was mainly used by OKI for production at 0.25 µm and 0.15 µm nodes for ultra-low-power circuits used in watches.[1] People were wrongly afraid of the possible variability of the devices because of the threshold voltage dependence on the silicon film thickness, and were not inclined to move from bulk to SOI, as they did not meet major issues using bulk. That situation has changed with the coming 20 nm node and beyond.

The probability is low that bulk will be able to maintain the leakage current as low as it was for the previous generations. Many companies have started to work on alternative solutions, and FDSOI is emerging as the best short-term solution from a manufacturing point of view. Other solutions are based on 3D devices (FinFET, multi-gate, nanoribbons ...) which will demand more development time to reach the manufacturing maturity. Note that all these 3D devices are fully depleted devices and can be implemented on either SOI or BULK substrates.

Coming back to the physics, FDSOI has the advantage that the vertical electric field in the channel can be controlled by adjusting the film thickness. The smaller the gate length, the thinner the silicon film must be to achieve a low leakage current. A general rule of thumb is that the silicon film thickness must be less than one-fourth of the gate length; for instance, a 5 nm silicon film must be used for a 20 nm gate length. The threshold voltage then becomes fixed by the silicon thickness and the metal gate work function; there is no need for doping in the channel region. This is a tremendous advantage, as one can immediately determine that the variability due to random dopant fluctuation no longer exists.

The use of an ultra-thin silicon film may become a manufacturing obstacle at smaller nodes, being difficult to deal with a few nm of silicon. This constraint may be released by depleting the device from several gates controlling the silicon film between them. The next step will be to use the FinFET: a vertical dual-gate fully depleted device, where the required silicon thickness may roughly be multiplied by 2 with respect to planar FDSOI. The next question is, What are the implications on the design side?

For planar FDSOI, a simple and direct answer is: There is no change from bulk; all the electronic design automation (EDA) views used for circuit design are identical. Most of SOI circuit designs in the last 20 years have used partially depleted SOI (PDSOI), and designers had to deal with the history effect of the devices[2] to achieve a reliable timing closure of the circuit paths. This effect does not exist in FDSOI, and timing analysis is run exactly the same way than bulk.

The differences between planar FDSOI and bulk are only related to the device electrical features. In particular, the SPICE

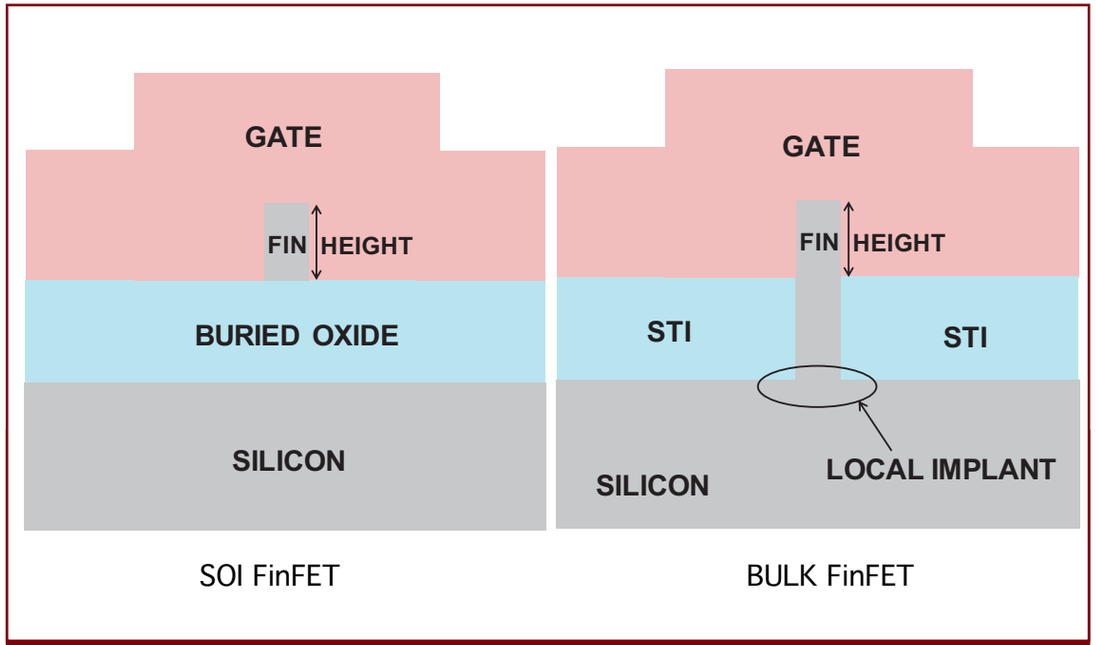


Figure 1. SOI and BULK FinFET Cross-section Comparison

model must be dedicated to FDSOI, a bulk SPICE model will not accurately simulate the FDSOI electrical features (current and charge models). Several SPICE models are already available, such as BSIM-IMG (University of California at Berkeley), HiSIM (Hiroshima University) or PSP-SOI (Arizona State University). It is also worth mentioning that all interconnect metal layers above the transistors are identical for SOI and bulk, and the corresponding wire loads are equivalent. A direct migration of existing bulk EDA views may be applied to FDSOI.

Though not necessary, slight changes may be achieved to get a better optimization due to the different FDSOI electrical features: The transistors used in the logic standard cells (optimum ratio between NMOS and PMOS is different between FDSOI and bulk), memory periphery and input/output (IO) may be resized for optimal performance. The SRAM bit cell may be identical or also slightly modified to better balance the pull-up, pull-down and pass-gate transistors. A much better margin is expected due to the reduced variability, and low-voltage memory operation will be easier to achieve,[3] which is a significant advantage for low-power applications. When moving from bulk to FDSOI, the only EDA views that need to be modified are those containing the electrical information: timing, power and noise, obtained from SPICE simulations (data contained in liberty or .lib files). This is identical to the recharacterization required when a bulk process modification results in a change of the SPICE models.

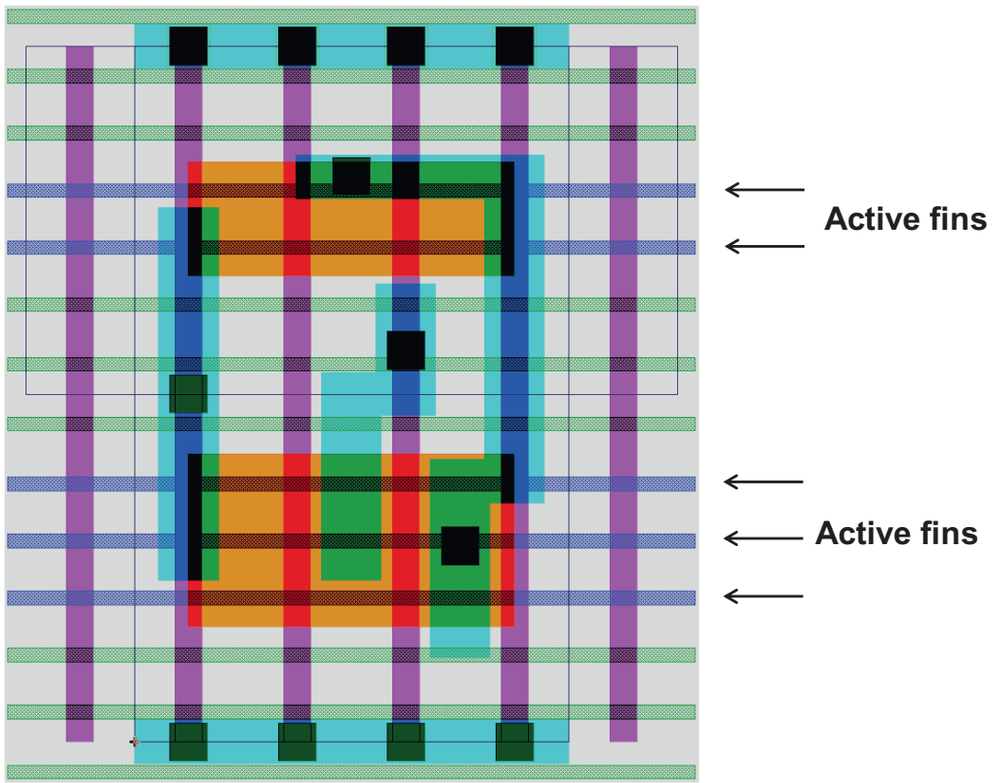


Figure 2. FinFET-based NAND2 Cell

When implemented on a thin buried oxide layer (typically 10 nm thick for 20 nm node), planar FDSOI offers another strong advantage: The threshold voltage of the transistors can be modulated by the back-gate bias applied to the substrate underneath the buried oxide.[3] The modulation is much more efficient than bulk; typically, 70 mV/V can be achieved, which translates in a 10x reduction of the leakage for a 1V change of the back-gate bias. The well bias has been extensively used in bulk designs to reduce the leakage (back bias) or improve the performance (forward bias), but has become fairly inefficient at advanced technology nodes because of the reduced threshold voltage modulation. FDSOI enables the continuation of implementing the reverse/forward biasing to mitigate power and performance.

FinFET may be seen as a 90° rotated planar FDSOI with the front and back gates connected together. It becomes a vertical device, and the silicon thickness between the gates is controlled through lithography and etching process steps. The bulk version adds more process complexity: The bottom part of the fin is not isolated from the silicon substrate, requiring an additional local doping to avoid drain-source punch-through leakage, and the fin height results from a combination of silicon etch and insulator planarization and etch-back.

In the SOI version, the fin height corresponds to the silicon film thickness after etching (Figure 1). The electrical features of the bulk and SOI versions are identical, and there is no difference when designing with either version. Contrary to planar FDSOI, FinFET has a significant impact on the design because of its discrete width. The width of the transistor is no longer the length of the drain-source diffusion; it now results from the fin height (roughly twice). The total width and then the delivered current of an active device correspond to the sum of fins connected together. It was usual in the past to compare the current per unit width between two technologies to indicate their performance. That no longer makes sense when comparing FinFET to planar. What is important is the current per unit area, which relates to the number of active fins that may fit in a given area. A real comparison may only be achieved at circuit level. Cells using FinFET have a similar layout to planar with the addition of the fins (Figure 2). A direct porting of a planar design (bulk or FDSOI) to a FinFET-based design may prove challenging because of the width mismatch. In addition to the EDA views modified because of different SPICE models mentioned for FDSOI, the layout views are also different. The FinFET discrete width is not an obstacle to analog design as long as a significant total device width is required so that many fins are used in parallel.

For both planar FDSOI and SOI FinFET, the protection against electrostatic discharge may be more area-consuming than its bulk counterpart. This is due to the limited silicon volume used by the active devices. This area increase is generally compensated by the additional area required for latch-up protection in bulk.

Endnotes

1. J.L. Pelloie, "FDSOI Design Portability from BULK at 20nm Node," IEEE SOI conference proceedings, 2011
2. J.L. Pelloie et al. "Timing Verification of a 45nm SOI Standard Cell Library," IEEE SOI conference proceedings, 2010
3. T. Skotnicki, "Competitive SOC with UTBB SOI," IEEE SOI conference proceedings, 2011

About the Author

Jean-Luc Pelloie

Jean-Luc Pelloie, ARM Fellow and director of SOI Technology, is an SOI expert. He co-founded SOISIC in April 2001, the first independent company offering SOI physical IP blocks, after having spent more than 10 years at LETI (Grenoble, France) on SOI technology development, including collaborations with IBM and TI. Jean-Luc joined ARM at the end of 2006 as a result of SOISIC's acquisition. Since then, he has driven the development of three SOI physical IP generations at 45 nm, 32 nm and 22 nm, and is currently working on 14 nm FinFET-based physical IP development. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Alan Weber
President, Alan Weber & Associates

The latter part of 2012 finds most industry veterans older yet wiser after the lessons of 2009, while still bracing for a rough couple of quarters. Nevertheless, these are the times when well-prepared companies focus on technology improvements that will better position themselves for the inevitable market improvements down the road, and the two papers in this section discuss such technologies in depth. Interestingly, both papers deal with complex system integration issues at the factory level, but at two very different ends of the application spectrum.

The first paper in this section provides an excellent overview of the connection points between a virtual metrology system and the various applications that must support it or are affected by its results. This is a refreshing departure from most of the literature that has been published on virtual metrology (VM) to date, which tends to focus on the analysis, modeling and algorithmic aspects of the technology (which are often very process- or solution-specific) rather than the practical issues that must be dealt with to realize VM in a production environment. In particular, the authors point out the multiple data sources that must be integrated beyond the trace data from the process tools; namely, the wafer or lot start/stop events, the logistic/context information (which may come from the MES), and the metrology results from prior runs. Moreover, they identify the related applications that must comprehend the insertion of VM into the application suite; namely, (dynamic) lot sampling, R2R control, and equipment health monitoring, among others. All in all, this article is a "must read" treatment of the topic for people considering a virtual metrology implementation.

The second paper deals with migration of an entire manufacturing execution system (MES) from a legacy level of capability to a new set of production functionality over some period of time. Migration of this scale can be thought of as a combination of integration and "dis-integration" processes of multiple types, including data sources and repositories, applications, user interfaces and especially the underlying business processes ... a daunting prospect for any company. The authors rightly point out that undertaking an MES migration requires clear justification, thorough requirements gathering, careful technology selection and planning, and above all, an experienced team that has done this before. Even though the audience for this paper may at first glance appear to be limited, the insights offered apply to a broad range of manufacturing system development/support organizations.

MANUFACTURING: FABS, SYSTEMS & SOFTWARE

Integrating Virtual Metrology in a Manufacturing System

Cristina De Luca,[1] Andrea Schirru[2]
[1]Infineon Technologies Austria [2]University of Pavia

Abstract

In semiconductor manufacturing, metrology operations are expensive and time-consuming; for this reason, only a small sample of produced wafers are evaluated. Virtual sensors are used to estimate the result of an operation measurement relying on processing data parameters. This article describes how virtual metrology should be integrated in a manufacturing system.

Introduction

Manufacturing science is a primary enabler for semiconductor companies to remain competitive for the next 10 years of technology evolution. Competitiveness means to be able to produce high-quality nanoscale devices at a reasonable cost, quickly and effectively integrating rapid changes in process technologies. Manage complexity, enable flexibility and agility, rapidly adapt to technology changes are key phrases used to describe most of the semiconductor challenges of the future.

During the last 20 years, semiconductor manufacturers have focused on advanced process control (APC) topics, including fault detection and classification (FDC) and run-to-run (R2R) controllers. After early deployments on specific processes such as chemical vapor deposition (CVD), lithography, etch, epitaxy, etc., the efforts were concentrated in the realization of a fab-wide solution (Moyne, 2004). The driving factor, especially in the last 10 years, has been the need to improve yield (Tobin and Neiberg, 2001) and quality without impacting productive throughput; such topics are even more important for the industries facing high product mix and rapid product development phases.

Some research has focused on sampling techniques. While on one side, this would allow for reducing costs by avoiding unnecessary measurement operations, this posed sensitive problems for R2R controllers, which need a constant flow of real measurements to work. Constant challenges include the costs and reliability of equipment integration for data collection and interaction with existing software (such as R2R controllers, FDC, *in-situ* sensors, yield management systems). The theory of a fab-wide solution that goes from APC to yield and integrates components such as scheduling and dispatching represents the next step in the semiconductor industrial roadmaps.

Actually, the need (ITRS 2011, Factory Integration) is to move factory operations from a reactive mode to a predictive mode: *"Reactive practices such as FDC (fault detection and classification), preventive and unscheduled maintenance and real-time scheduling and dispatching have been used in the past and also today but unfortunately they cannot eliminate waste on product, downtime, cycle time, yield loss etc. ... this is due to the fact that there is a reaction to an incoming problem."*

This means to migrate from fault detection to fault prediction and in this way to reduce control events and fault occurrences, better scheduling maintenance events in order to reduce unscheduled downtime and improve efficiency in daily maintenance work. On the other side, one of the important components is virtual metrology, which will ably provide metrology values in real time at different product levels (lot, wafer, sites) at different equipment levels (equipment type, chamber, sub-chamber). Data such as sensors or APC key values, recipes, logistic data of lots – wafers and equipment, and also, if possible, maintenance information – are merged in a sequential data flow available for a virtual metrology algorithm or software.

A key challenge to reach the "prediction" goal is the development of robust, configurable and real-time models and algorithms. After data filtering and quality data check, the virtual metrology data are available for all the "customers" that need to use this information. One customer is the R2R controller that will benefit from the good wafer/site information, solving the conflict between having no measurement for costs and having measurement for better control. Another example is the "skip lot sampling," which, based on a so-called wafer at risk, could decide when and what to skip. An equipment/process health factor system that can trigger maintenance or other events can be also a customer of VM information.

These are the main "systems" that could benefit from virtual metrology and the reliability of the calculations (see Figure 1). A great challenge in doing this is certainly the high-mix factory environment and the complex process control (reaction to frequent recipes and process tools changes). The expectation's target is the improvement of efficiency: productivity waste reduction, better product quality, higher process monitoring and controllability, yield and cycle time improvement, and last

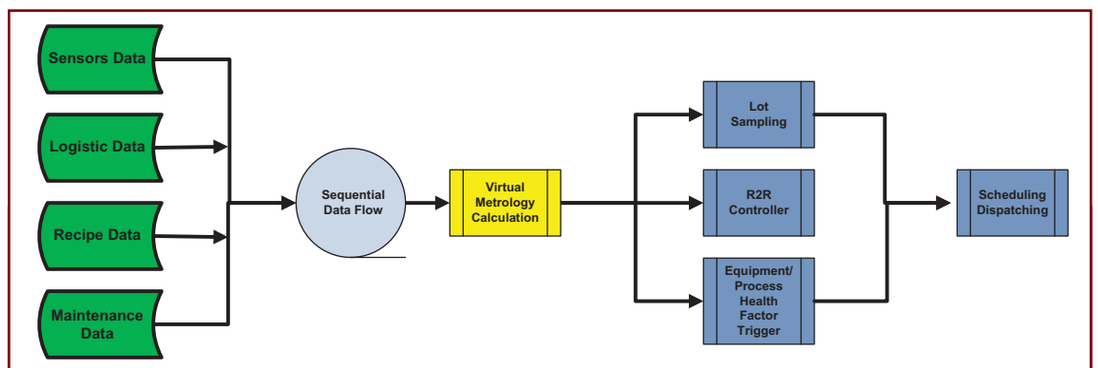


Figure 1. General Virtual Metrology Data Flow: Input Data & Output & Possible Software Customers

but not least, reduction in utilities and power consumption. Following is a view on how to integrate virtual metrology in a manufacturing system.

Data Collection

To implement an automated virtual metrology system, it is necessary to develop a robust online data import library that will be responsible for merging information from heterogeneous sources to a common format. Specifically, the following sources of information should be available:

- *Raw equipment data dispatcher:* This interface sends, on a fixed-time basis, the sensor information coming from a piece of equipment, as well as the set parameters for the recipe currently being processed.
- *Synchronization component:* This service, which can be either independent or included in the previous point, has the task of sending synchronization tokens about processed wafers (e.g., when the processing of a certain wafer begins or ends), along with minimal logistic information.
- *Logistic data provider:* This component is able to provide the full logistic scope of a certain wafer on a request-reply mode, receiving the minimal logistic information discussed above as input.
- *Metrology data provider:* This allows the integration of the full process data (collected by means of merging the information of the first three components) with metrology information, whenever available.

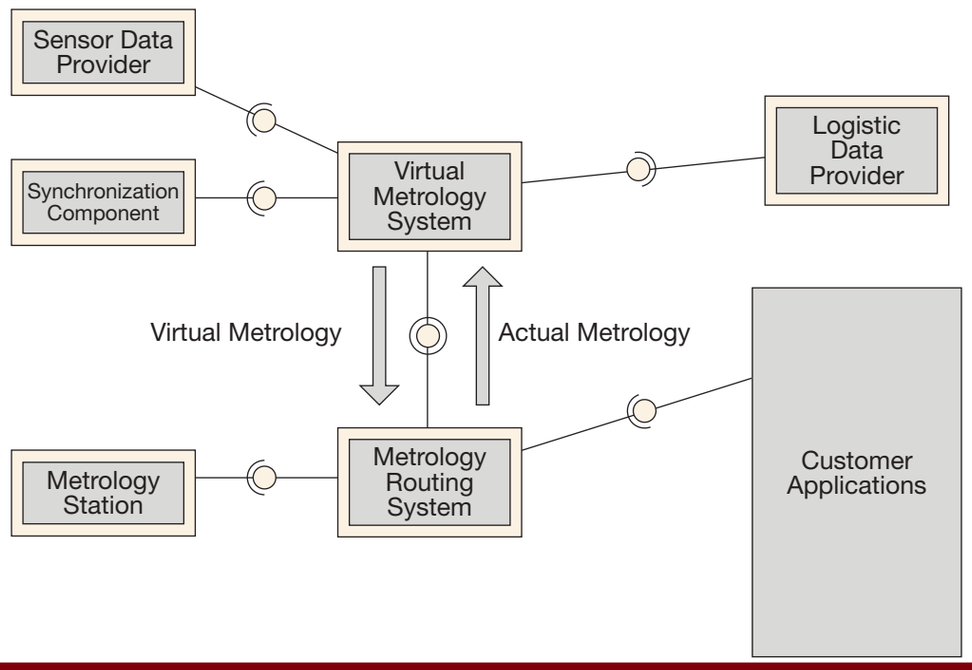


Figure 2. Architecture and Dependencies of the Virtual Metrology System

Using a wafer-specific key related to the above systems (e.g., an RFID-based identity of the silicon slice, combined with minimal process-related key to identify the correct step), enables building a homogeneous learning dataset (Figure 2) with the following:

- *Unique wafer identifier (UWI):* an internal key that uniquely represents a silicon slice undergoing a specific process step in a specific piece of equipment
- *Complete logistic infomation*: all the logistic data (e.g., chamber position, recipe name, etc.) associated to the UWI
- *Complete process data:* all the sensor readings collected during processing, as well as equipment set points
- *Metrology data:* complete measurement type (categorization) and (possibly multivariate) measurement results

Following such dataset structure, it is possible to configure a dataset extraction system for a specific logistic configuration (e.g., a certain recipe or process type) and associate it to a specific measurement (e.g., a thickness measurement). After this step, the dataset is ready to be processed by a mathematical algorithm.

Model Creation and Prediction

Whenever it is possible to update a VM model (that is, every time a complete wafer is available – including metrology information), the above described dataset is elaborated by a mathematical procedure in order to provide a VM model.

While the literature in terms of VM models is quite wide, including hierarchical approaches,[3] kernel-based methodolo-

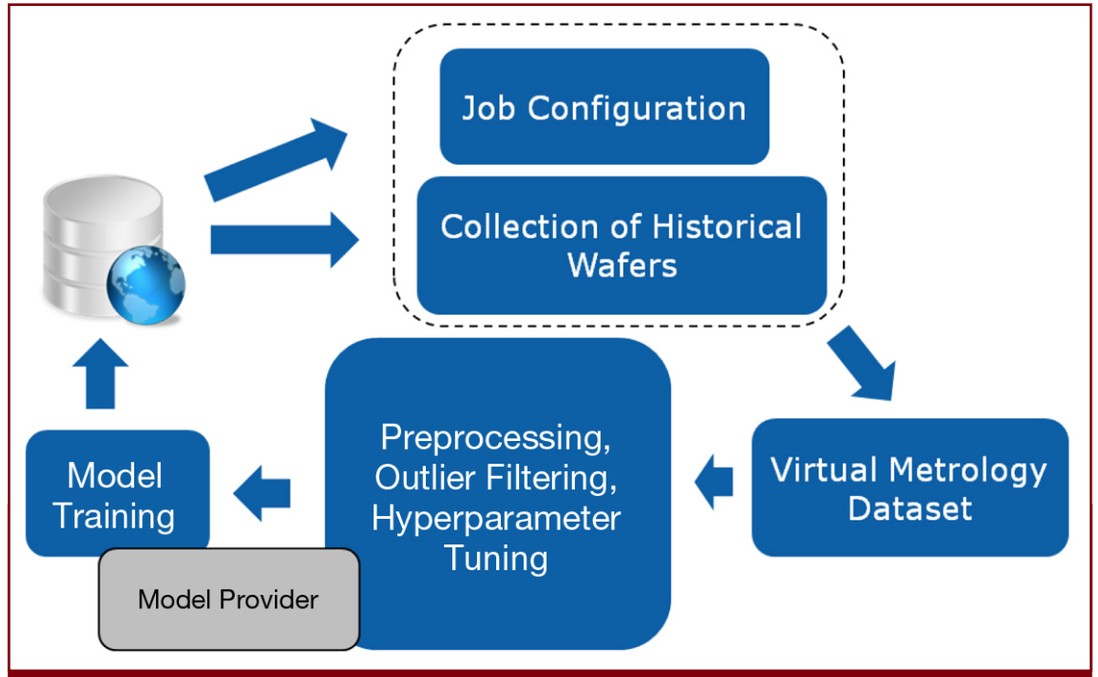


Figure 3. Logical Flow for Model Computation

gies[4] as well as models profoundly impacted by process expertise,[5] the main focus from the software architecture point of view must be set on the generic interface between historical dataset and mathematical library. Specifically, a system of interfaces must be set in a way that any algorithm can be implemented as a black-box procedure, where only the structures of input (historical dataset) and output (the model) are predetermined. From the point of view of architectural development, this allows for the decoupling of the system from the employed algorithms, ensuring forward compatibility for new breakthroughs in the VM learning field.

Special emphasis must be set on the parameter selection for the involved algorithm; in fact, it is well-known that most learning algorithms depend on an additional set of parameters (called "hyperparameters") that must be tuned to obtain optimal predictive power. Given that the procedure used for such tuning operations is usually the same for very different algorithms, it is advisable to implement a wrapper procedure to deal with them (e.g., by means of a grid search-based approach). In this way, a significant overhead in algorithm programming will be avoided, and suboptimal models will be ruled out by a centralized procedure (Figure 3).

As soon as a model is calculated, it is possible to use it to make predictions for new wafers; specifically, every time a wafer that matches the model configuration is available (even without metrology data), a virtual measurement will result from the input-output structure of the most recently calculated model. Depending on the capabilities of the model, additional indicators might be available, such as confidence intervals or virtual goodness-of-fit (GOF).

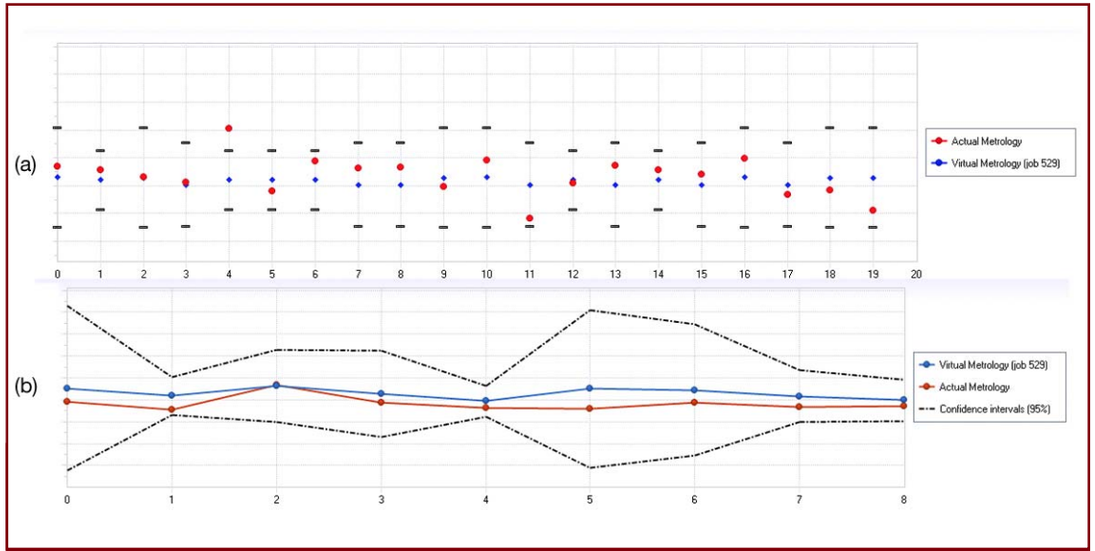


Figure 4. Comparison of real and wafer-level virtual measurements with 95% confidence interval for CVD thickness average (42 wafers from different technologies, same target) and site-level thickness (9 points) for 1 wafer. All the real and virtual measurements are between the control limits not included in the pictures.

An example of the results of a VM algorithm is given in Figure 4, comparing in 4(a) real average thickness productive data and virtual ones for 42 wafers (different technologies, same target) and in f(b) the nine virtual sites as well as real measurement for one wafer. Upper and lower control limits are not included in the figure as they are well outside the considered window. In the best case, such as this, all the measurements are well inside the specified control limits. If some of them are outside the 95 percent confidence limit, or the upper/lower control limits, the VM values for that wafer are discarded.

Conclusions

This article has briefly explained how to implement a virtual metrology system in a real semiconductor manufacturing environment. Among the main challenges that such an endeavor presents is having a robust data collection system for online VM calculation and wafer/equipment monitoring purposes. The example shows the results of a VM-algorithm application with real online data. Furthermore, it is a key aspect to provide a flexible interface for model calculation in order to lower the implementation effort for new methodologies.

References

1. Moyne, J., "Making the Move to Fab-wide APC," in Solid State Technology, vol. 47, No. 9: pp. 47-52 (2004)
2. Tobin Jr., K. and Neiberg, L., "Metrology Data Management and Information Systems," in Handbook of Silicon Semiconductor Metrology, SEMATECH
3. Schirru, A. et al. "Multilevel Kernel methods for virtual metrology in semiconductor manufacturing," IFAC World Congress 2011, 11614-11621
4. Schirru, A. et al. "Virtual Sensors for Semiconductor Manufacturing: A Nonparametric Approach - Exploiting Information Theoretic Learning and Kernel Machines," Informatics in Control, Automation and Robotics, (2013) 175-193
5. Purwins, H. et al. "Regression Methods for Prediction of PECVD Silicon Nitride Layer Thickness," Automation Science and Engineering (CASE), 2011 IEEE Conference, 387-392

About the Authors

Cristina De Luca

Cristina De Luca received her laurea degree in statistical and economical science, University of Padova (Italy) and her Ph.D. degree in mathematics, University of Klagenfurt (Austria). She joined Infineon Technologies Austria in 2002, where she is currently innovation project manager and staff engineer on the Innovation Management team. Dr. De Luca has worked on a wide range of R2R control applications and has significant experience in advanced process control. She was project manager of the "IMPROVE" ENIAC Project for Work Package "Predictive Equipment Behavior."

Andrea Schirru

Andrea Schirru received his master's degree in computer science engineering and his Ph.D. degree in model identification from University of Pavia (Italy). He participated in the EU-ENIAC project IMPROVE. Dr. Schirru's main research interests include virtual metrology, fault detection and classification, and machine learning models. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

MANUFACTURING: FABS, SYSTEMS & SOFTWARE

MES Migration: An Absolute Need or a Utopia?

Francisco Almada-Lobo, Jeff Peabody
Critical Manufacturing

Abstract

Semiconductor manufacturers are among the earlier adopters of MES systems. Once vital to maintain the competitiveness of their operations, these applications are becoming obsolete, and the need to circumvent their limitations causes significant cost, time and risk and will sooner or later impact these companies' business. This article describes the limitations of legacy MES and the main aspects to consider in a migration project.

The Need to Migrate

Introduction

Manufacturers operating in high-tech sectors such as the semiconductor industry must continuously embrace change or find themselves overtaken by more agile competitors. Forward-looking companies can point to a history of innovation and continuous improvement – and know that further change will be required to remain successful. These changes encompass everything from product conception, design and manufacturing to the sale, support and retirement processes.

Manufacturing execution systems (MES) play a vital role in this perpetual evolution. Almost no change can be done without the appropriate enforcement, recording and process automation that MES systems provide.

The MES is often considered a very mature market. However, many of these systems have evolved from homegrown applications, and many commercial systems have been developed to meet the narrow requirements of certain industries.

Given the effort to build a new MES and the need for a return on such investments, the majority of existing solutions were developed 10, 20 or even more years ago, prior to the major technology advancements in recent years. Early suppliers accumulated a significant customer base that could not be abandoned or disrupted even when new technologies became available. Thus, improvements in legacy systems tended to be small and incremental. However, these systems needed to be maintained because, as we'll see later, manufacturers were also reluctant to change, providing a unique environment for extending the life of sometimes extremely obsolete technologies.

Without fast-evolving MES products, the burden of enabling the manufacturing process evolution was passed on to manufacturers. Bigger semiconductor corporations had dozens or even hundreds of persons building extensions and work-arounds to overcome system limitations. Very-low-level technical framework features had to be used due to the legacy technologies involved. Every new feature had to be patched in a complex network of functional blocks that had not been designed for that purpose, requiring considerable time, effort and risk, and an exponential future maintenance cost (see Figure 1).

Is this history? Not quite, as a majority of the market remains in this state.

Main System Limitations

Given the obsolescence of the technologies involved, many aspects limit the evolution and scope of the manufacturing processes. However, we will focus on factors causing the greatest constraints.

Modeling

The modeling capability is perhaps the most important, because it is one of the most difficult to be circumvented. Semiconductor processes are complex. There are cascading flows, process loops, optional steps, rework paths, alternate flows and non-sequential steps.

Material flows are frequently not discrete in nature, forcing the use of hierarchical logical material structures. Materials need to be associated with position-controlled container models, and tracked to sublevels of clustered equipment. Dispatching or sorting rules need to be dynamic and consider both cell and line optimization criteria.

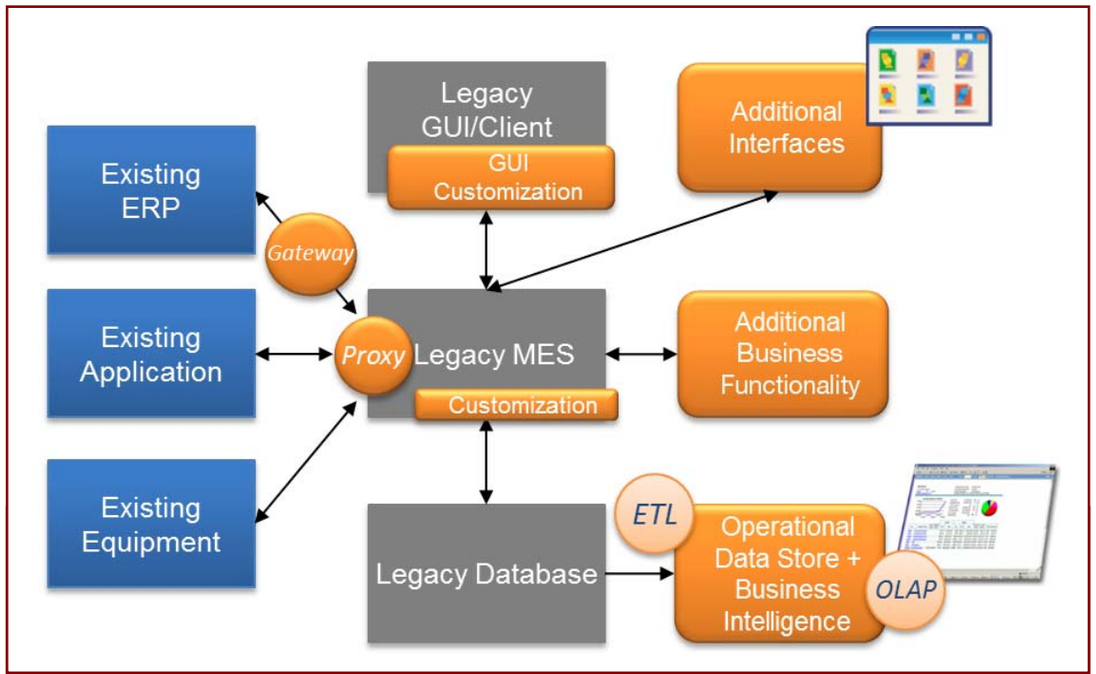


Figure 1. Most Common Legacy System Enhancement Strategies

Engineering data collections need to be recorded in real time, some manually collected, while others are collected automatically from measurement or production equipment. These often need complex calculations, dynamic sampling and to be checked against limits.

Without realistic modeling of the operations, the result is inevitably a compromise in visibility, traceability and control. That can be fatal.

Rule Enforcement

Rule enforcement is the second category. It includes all aspects of misprocess prevention or of immediate reaction upon anomaly situations. Prevention includes verification of limits, measurements, elapsed times, checklists, operations performed and other preconditions. Reaction includes automatic material dispatching, equipment actions, remeasurements, triggering exception protocols and other possibilities.

A system that lacks rule enforcement compromises product quality, yield and cycle time, and ultimately profitability.

Functional Coverage, Integration and Usability

While all MES systems deal with material and resource tracking, the full scope of the manufacturing operations is much broader. Beyond tracking, a fully integrated system comprises monitoring, controlling, dispatching, quality and recipe management, equipment integration and automation, documentation management, maintenance management, data analysis and business intelligence.

Can't this functionality be provided by other systems? It can, and depending on the specialization requirements, sometimes it should. However, there must be a balance between what's in scope of the MES, considering that external systems may mean a duplication of master data and a significant increase of integration and maintenance effort. This becomes overwhelming with legacy systems and, every subsequent addition is likely to exponentially worsen the problem.

The user interface plays an important role in terms of efficiency and usability. It must be ergonomic, integrated, oriented to the operator flow and easily adaptable to changing scenarios. Some traditional systems still rely on character-based terminals, while others come with Visual Basic or Web-based interfaces that provide a less-than-ideal user experience and prove cumbersome to extend and maintain.

So Why Don't Companies Change Their MES Systems?

They certainly can, and it is inevitable that they will do it sooner or later. Paradoxically, as companies are afraid of doing it, they make the situation worse by investing more effort and cost in adding workarounds and patches to keep up with their dynamic environments.

Because of the dependency production has on the MES, fab managers have been heard comparing MES migration to a heart transplant. Getting a newer and leaner system is an appealing concept, but in the analogy of the heart transplant, the new heart must be fully compatible with all other organs and the surgery must be carefully planned and executed.

Given the risks involved (in the MES migration), the obvious question is, What are the benefits?

There are four primary drivers:

- End-of-life support
- High maintenance and operational cost
- Effort to modify, extend and integrate
- High turnaround times

Perhaps the only lethal argument to an immediate change would be the end-of-life support, but legacy system suppliers contribute to the postponement of the decision by continuously announcing support extensions.

So, the justification must allow for a midterm gain, considering risk, cost and missed revenue (opportunity costs) arguments. The short-term argument may be difficult to see, but the long-term peril is simple and easily understood: Coping with the required process agility will slowly become prohibitively expensive until it becomes impossible. Every time the decision is postponed, it makes the future migration more expensive and riskier.

The Migration

Three aspects must be considered when planning the migration: the selection of the MES; the migration scenario; and the project team.

	"Big Bang"	Parallel System	Phased Introduction
Risk	Very difficult to roll back Very difficult to test in production	Easy to roll back	Somewhat easier to roll back
Cost	$ Requires the least effort and time	$$$ Requires the greatest effort to maintain & synchronize the two systems	$$ Effort is in between the other two scenarios
Time	The fastest migration process, but requires the most preparation	Somewhat fast migration process, but requires a lot of preparation	Slowest migration process, but requires the least preparation
Remarks	• The most disruptive and risky strategy • Allows taking immediate advantage of the new system	• The most costly migration strategy • May postpone new functionality until switch-off of old system	• The most balanced migration strategy • Requires two reporting/ analysis systems for the duration of migration

Figure 2. Assessment of MES Migration Strategies

Selection of the Right MES

The selection of the right MES system is key for the success of the migration project and the achievement of its objectives.

First, the new MES needs to address the majority, if not all, of the shortcomings of the legacy system. Vendor guides from analyst companies like Gartner[1] or Logica[2] are excellent starting points. The system must have full modeling possibilities for materials, equipment, containers, product structures, flows, steps, data collections, but with high flexibility. Remember that it will support the current and future business needs. It must allow the enforcement of rules and processes both preventively and reactively. And it must contain all necessary functionality to cover as completely as possible the manufacturing system's domain in a modular, yet service-oriented fashion. This facilitates minimal integration effort within the modules, but still provides room for specialized functionality to be performed by external systems with low integration efforts.

Second, it must eliminate the need for another migration for many years. It makes little sense to replace a 15-year-old system with a 10-year-old system. The goal is to use the new system for 10 or more years. Beyond the model flexibility and functional modular coverage, it should allow an easy evolution in terms of configurability, extensibility and interoperability.

Third, and often ignored, the new system should be migration compatible. This refers to the ability of the new MES to "mimic" the legacy MES to allow migration. This is ultimately related to the migration strategies described later, which include the "big bang" approach, parallel systems or phased introduction. The "big bang" approach is limited to master data migration, but the need for compatibility is considerably higher in the phased introduction and extremely high in a parallel system approach.

The Migration Strategy

Exploring the different migration strategies would provide sufficient content for a much longer article. A migration project is complex and needs to include:

- Migration of master data
- Migration of manufacturing software processes
- Reporting systems
- Historical information
- Integration with other applications
- Deployment plans and procedures, including software, hardware and people
- Risk mitigation and fallback solutions
- Hand-over to operations & support

Figure 2 summarizes the most common options. There are, of course, possible variations within each scenario.

The main point is that there's no right or wrong strategy, but one may better fit your needs. To determine the best strategy, the company must answer the following questions:

- What is the level of risk the company can cope with?
- What is the amount of effort (cost and time) that can be invested in the project?

The migration strategy can now be part of the MES selection criteria.

With MES capabilities and strategy determined, thorough planning with involvement from representative stakeholders is essential. Answering the following additional questions helps to better determine the project complexity:

- How much production downtime is acceptable (from zero to *n* days)?
- What historical data needs to be used within the new system (versus only in analytical or business intelligence databases)?
- How complex are the interfaces to applications that require MES interfacing?

The Team

While there are many MES professionals with significant manufacturing and MES knowledge, resources specifically with migration experience are few. Even if a company decides to use a supplier with a history of successful deployments, it is essential for the team to include someone with migration experience. Although each migration is unique and highly dependent on the environment and selected applications, as well as on the strategy, persons who have been exposed to migration projects know what needs to be considered and how to react to specific migration challenges. Leveraging this experience is invaluable and can often determine the success of such a challenging project.

Conclusions

In a Nutshell

Production and site managers constantly face a difficult balancing act. On one hand, they are challenged to excel on day-to-day operations, improving quality and throughput, while at the same time reducing cost. On the other hand, they need to ensure the future long-term viability and competitiveness of their company.

Migrating an obsolete MES is an absolute must. If not done, it will sooner or later have a negative impact on the company's business. The longer it takes, the higher the risk and the higher the effort.

It is definitely not a utopia. It has been successfully done in very complex and high-volume manufacturing environments. It requires the right migration strategy, the right MES, proper planning and of course, experienced people who have done it before.

Endnotes

1. Vendor Guide for Manufacturing Execution Systems, 2012, Gartner
2. MES Product Survey, 2012, Logica

About the Authors

Francisco Almada-Lobo

Francisco Almada-Lobo worked for 12 years in semiconductor IT for Siemens, Infineon and Qimonda. In 2004, he led Infineon's first migration of an MES system in a running high-volume facility, and managed until 2009 a development center implementing MES projects worldwide. Francisco co-founded Critical Manufacturing in 2009, where he has been the CEO since 2010.

Jeff Peabody

Jeff Peabody served as director of Sales and Marketing at itemic Corp. and as product sales manager for KLA-Tencor's Process Control group. From 2006 till 2009, he held a senior sales executive position in the Global Services group of Applied Materials. Jeff has been Critical Manufacturing's VP for Sales and Business Development since 2010. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Janice M. Golda
Director, Lithography Capital Equipment Development; Intel Corp.

Building the Perfect Mask

In this section, Vibhu Jindal and his colleagues at SEMATECH describe their comprehensive search-and-destroy mission to find defects on EUV mask blanks, eradicate their sources and help enable EUV mask equipment and materials suppliers to do the same. The targets of this mission are to meet the memory requirement of zero defects larger than 100 nm, and less than 22 defects larger than 35 nm per blank, and to meet the logic requirement of zero defects larger than 75 nm and less than three defects larger than 25 nm per blank. The numbers are daunting.

This mission has required seeking out and destroying defect sources in mask substrate preparation as well as in the subsequent multilayer deposition process, leaving no rogue defect source to survive. Substrate preparation involves removal of material via various polishing techniques to deliver an extremely flat surface, followed by cleaning to remove defects generated during polishing. Multilayer deposition requires depositing more than 80, yes 80, layers of film to build the EUV reflective film stack without adding defects. Destroying defect sources in the multilayer deposition process has required SEMATECH to seek out the defect sources in the deposition equipment and address them by working with the supply chain to improve equipment design and materials specifications.

As a result of current progress, SEMATECH has demonstrated production of EUV blanks that can meet defect density as required for memory applications; however, further improvements are required to meet the logic requirements, as well as to enable the supply chain to deliver consistent yield to enable high-volume manufacturing. The mission will continue ...

LITHOGRAPHY LANDSCAPE

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Proven Solutions Through Evolution

Nikon Corporation has been one of the world's leading optical companies for more than 90 years. Nikon developed the world's first production-worthy step-and-repeat photolithography tool in 1980. Since then, over half of all integrated circuits printed have been manufactured on Nikon steppers and scanners.

Nikon has a long-established precedent of leading the industry through innovation and the continuous evolution of our proven lithography solutions, providing photolithography systems spanning the range of resolutions required by today's IC manufacturers. From high-throughput i-line steppers to advanced immersion ArF scanners for 22 nm half-pitch applications and beyond, Nikon delivers exceptional performance with the lowest cost of ownership, and the most comprehensive customer support of any manufacturer.

In 2007, Nikon shipped the industry's first 1.30 NA immersion scanner, the NSR-S610C, for 45 nm half-pitch production. Later that year, Nikon also introduced the NSR-S310F and NSR-S210D non-immersion scanners. These systems were evolutions of the S610C immersion platform that incorporated Tandem Stage technology to provide optimal performance and cost of ownership for dry lithography applications.

Then, in 2009, to meet the stringent requirements for 32 nm double patterning and provide extendibility to next generation applications, Nikon introduced the NSR-S620D immersion scanner, which is based on the innovative *Streamlign* platform. This was followed in 2011 by the NSR-S320F, an evolutionary *Streamlign* platform-based system designed to deliver exceptional performance and productivity for the most critical dry ArF applications. The most recent release from Nikon is the NSR-S621D immersion scanner. This latest evolution of the proven *Streamlign* platform fully satisfies the aggressive overlay and throughput requirements of high-volume immersion double patterning applications at 22 nm half-pitch and beyond.

Next generation lithography techniques continue to evolve, but IC makers need solutions today that will keep them on their aggressive technology roadmaps. With a history of innovation and evolutionary lithography solutions, Nikon will be there to ensure you maintain your production timelines.

LITHOGRAPHY LANDSCAPE

Reducing Defects in EUV Mask Blanks to Enable High-Volume Manufacturing

Vibhu Jindal, Patrick Kearney, Arun John, Frank Goodwin
SEMATECH

Extreme ultraviolet (EUV) lithography is the leading next-generation technology to succeed optical lithography beyond the 22 nm node,[1] and the availability of defect-free masks is one of the top two most critical technology gaps hindering its commercialization.[2,3] The masks for EUV lithography are reflective, composed of a multilayer structure consisting of molybdenum and silicon bilayer film, a capping layer, and a patterned absorber layer formed on a 6-inch glass substrate made of low thermal expansion material. EUV mask blanks are defined as multilayer structures with a capping layer but without an absorber mask pattern. Defects can be incorporated at different locations in the blanks depending on their various sources. Figure 1 shows all the possible defect locations. Substrate pits and embedded particles on the substrate, due to chemical mechanical polishing (CMP) or the cleaning process, are major concerns, as they are small and often difficult to detect. Particle residues left on the substrate due to cleaning, storage or handling are another major contributor of mask blank defects. Other significant sources are deposition particles within or on top of the multilayer due to material handing within the deposition tool and from the deposition process and pits or particles added post-deposition from storage, cleaning or handling. The insufficient progress across the industry in reducing mask blank defects has prompted SEMATECH to concentrate efforts on identifying their major sources, implementing mitigation techniques, and demonstrating a yielding EUV mask blank multilayer deposition process with low defect density. Ongoing research at SEMATECH has provided opportunities for blank suppliers to acquire early learning on the deposition, cleaning and inspection tools, providing the impetus to improve EUV mask blank quality for the industry. This work has enabled mask blank and tool suppliers to improve processes for multilayer deposition,[2,3] develop novel cleaning techniques for EUV substrates and blanks,[4] and evaluate the metrology infrastructure for mask blank defects to match the industry's needs.[5]

The current industry requirement for memory applications specifies that EUV blanks contain 0 defects >100 nm, considered killer defects and less than 22 defects >35 nm in the quality area. The defect requirements for logic applications are more stringent, requiring 0 defects >75 nm and less than 3 defects >25 nm. SEMATECH has been able to determine the critical components and sources of defects that are impeding progress in mitigating EUV mask blank defectivity. Defects from the shields and targets are large killer defects that must be completely eliminated. Characterization of the target surface after sputtering shows a tendency for nodule formation and roughness near the edges, primarily due to divergence in the ion beam. Based on the analysis of mask blank defects and the tool, the two main contributors to target defects were determined to be from scattered ion-beam sputtering on the edges of the target and the material properties (dopant properties, void density, surface finish, etc.) of the bulk target materials. A number of defects can also originate in the shield areas when (a) installed shields are not cleaned properly; (b) heavy depositions cause cracking or flaking on shield surface; and (c) the shield surface is etched by the scattered ion beam. The collective improvements to tool components and processes were able to significantly drive down the number of large defects. The recent improvements in cleaning processes at SEMATECH were also able to reduce total defectivity and improve the yield of high-quality mask blanks. With the implementation of these improvements, SEMATECH was able to report a new champion mask blank three times in the last four quarters. A champion mask blank of 28 defects >45 nm was reported in 2011,[6] with a further reduction to 19 defects >45 nm by the end of Q1 2012. The latest champion mask blank at the end of Q2 2012 shows 12 defects >45 nm, with only 8 defects >50 nm within the quality area (132

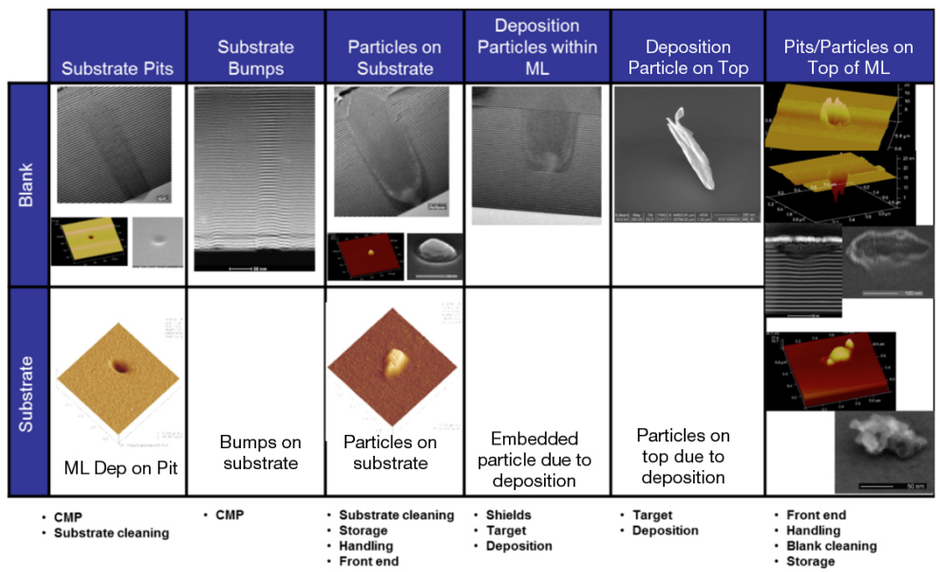


Figure 1. EUV masks blank defects categorized based on location and possible defect sources.

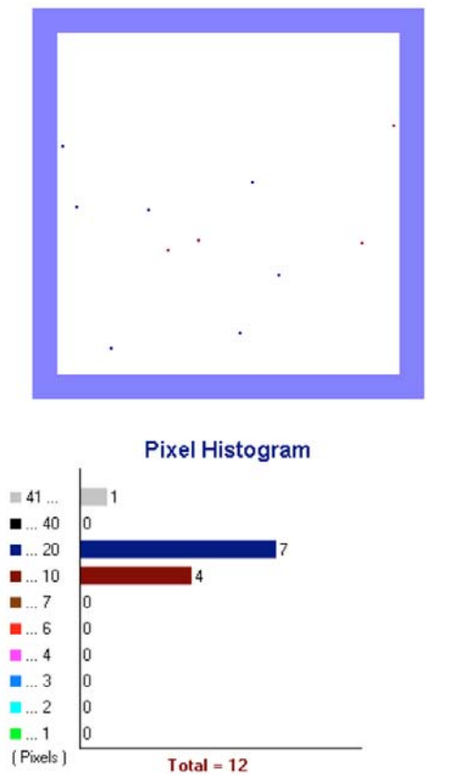


Figure 2. Champion EUV Mask Blank With 12 Defects >45 nm and 8 Defects >50 nm as Inspected by the Lasertec 7360 in Dense Scan Mode

mm by 132 mm) as measured by the Lasertec M7360 (Figure 2) in a dense scan mode.

As a result of our current progress, SEMATECH has demonstrated production of EUV blanks that can meet defect density as required for memory applications; however, further improvements are required to meet logic requirements. The yield of the mask blank deposition tool, which has been an increasing concern, was also improved in the last year. The yield of high-quality mask blanks (less than 20 defects >45 nm) in the clean phase was found to be just over 20 percent (Figure 3(a)). SEMATECH has a clear path to drive further improvement of EUV mask blank champion defectivity and to address the lack of corresponding yield through the design of next-generation deposition tools equipped with improved handling capability. Improvements in some of the critical components within the ion beam deposition systems – such as the ion beam source, substrate fixture and substrate chucking method – are also required. The next-generation deposition tool should have ion sources in which divergence will be tightly controlled and scattered high-energy ions and neutrals minimized. As defect data have shown, the dual-pod reticle carrier has superior defect protection when compared with a standard reticle standard pod. The next-generation tool should be configured to support and integrate the dual-pod front-end handler. Furthermore, tool suppliers should proactively be involved in these advancements and improvements.

The demands for EUV mask blanks will increase over the next few years; however, the industry is currently not equipped to meet these demands. The need for high-quality EUV mask blanks is expected to reach up to 5,500 per year in 2015. The high demand will generate capacity issues for ion beam deposition systems and unavailability issues for high-quality substrate yields. The supply of high-quality EUV substrates has often been overlooked. The current yield on such quartz substrates (less than 5 defects >45 nm), as procured by SEMATECH, is insufficient, (Figure 3(b)). The yield on low thermal expansion material substrates is even less with the current technology. As can be seen in Figure 4, more than 18,000 high-quality substrates are needed to yield 5,500 high-quality mask blanks per year at a 30 percent yield in the ion beam deposition system. Meeting such demands involves addressing numerous substrate requirements, especially considering the current issues with substrate quality and cleaning. Achieving these targets will require a new multilayer deposition tool to improve the deposition yield on high-quality blanks, extensive innovations to improve substrate quality and cleaning to enable high-volume manufacturing of low-defect density blanks.

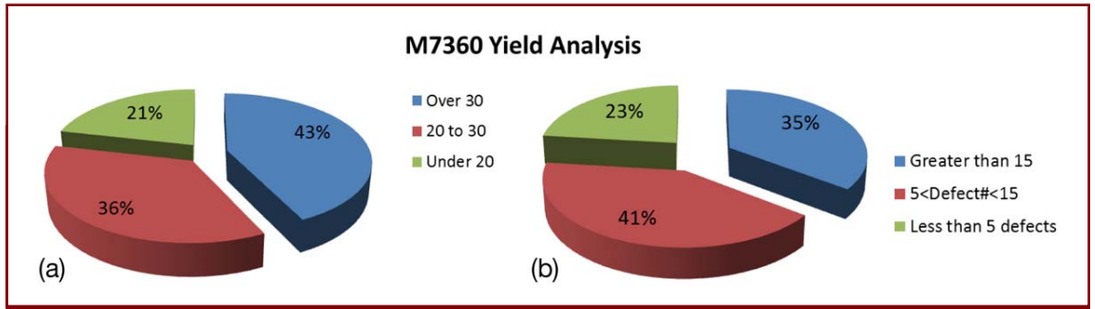


Figure 3. (a) Yield of EUV mask blanks with under 20, 20-30 and over 30 defects >45 nm after EUV mask blank deposition; (b) yield of EUV substrates under 5, 5-15 and over 15 defects >45 nm after substrate preparation and cleaning

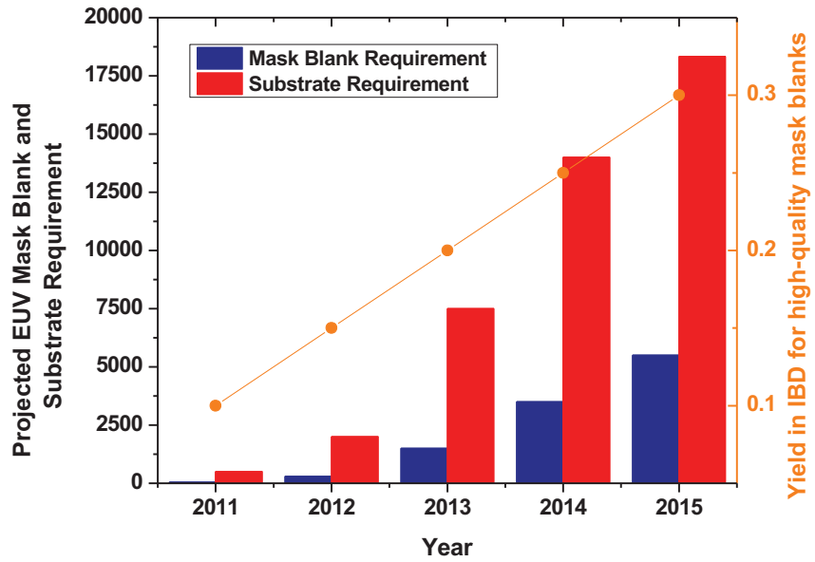


Figure 4. Projected EUV Mask Blank and High-Quality Substrate Requirements Based on Yield in an Ion Beam Deposition System

References

1. S. Wurm et al. "SEMATECH's EUV program: a key enabler for EUVL introduction," Proc. SPIE 6517, Emerging Lithographic Technologies XI, 651705 (March 23, 2007)
2. V. Jindal et al. "Ion beam deposition system for depositing low defect density extreme ultraviolet mask blanks," Proc. SPIE 8322, Extreme Ultraviolet (EUV) Lithography III, 83221W (March 29, 2012)
3. V. Jindal et al. "Modeling the EUV multilayer deposition process on EUV blanks," Proc. SPIE 7969, Extreme Ultraviolet (EUV) Lithography II, 79691A (April 6, 2011)
4. Abbas Rastegar and Vibhu Jindal, "EUV mask defects and their removal," Proc. SPIE 8352, 28th European Mask and Lithography Conference, 83520W (April 16, 2012)
5. V. Jindal et al. "SEMATECH's infrastructure for defect metrology and failure analysis to support its EUV mask defect reduction program," Proc. SPIE 7969, Extreme Ultraviolet (EUV) Lithography II, 79690I (March 29, 2011)
6. V. Jindal and F. Goodwin, "Roadmap for Defect-Free Extreme Ultraviolet Mask Blanks," Future Fab International, Issue 39 (October 2011)

About the Authors

Vibhu Jindal

Vibhu Jindal is the group leader in SEMATECH's EUV Mask Blank Defect Reduction program. He received his B.S. in material science from the Indian Institute of Technology, Bombay, and his Ph.D. in nanoscale science and engineering from SUNY-Albany.

Patrick Kearney

Patrick A. Kearney earned B.S. degrees in physics and mathematics with minors in optics and electronics from the Rose-Hulman Institute of Technology in 1987, and a Ph.D. in optical sciences from the University of Arizona in 1994. From there he moved to Lawrence Livermore National Laboratory until 2004, developing and refining several techniques for depositing multilayer mirrors. In 2004, Patrick moved to SEMATECH's EUVL Mask Blank Development Center to aid in the development of EUVL mask blank infrastructure.

Arun John

Arun John Kadaksham has been a project engineer at SEMATECH for six years. His primary focus is reducing nanoparticle defects on substrates for EUV lithography. Dr. Kadaksham previously held positions as a post-doctoral research associate at Clarkson University, where his research was mainly focused on nanoparticle adhesion and removal.

Francis Goodwin

Francis Goodwin is the program manager for SEMATECH's EUV Mask Blank Defect Reduction program and Mask Blank Development Center. He has more than 20 years of experience in lithography, has held management and engineering positions throughout his career, and has supported ASIC, logic and advanced DRAM operations. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

FRONT END OF LINE

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Jiang Yan
Professor, IMECAS

In the semiconductor industry, high yield and low cost are two very important issues that are closely related to the manufacturing profit. There are many factors and issues that could impact the performance of the yield and the cost. In this section, there are three excellent papers that discuss ways to enhance yield and to reduce the process cost.

The paper from the University of Milan and Technofittings S.r.l. introduces a novel optical method to measure the particles in fabs. The measurements at different situations and the reliability of the method are presented in the paper.

In terms of yield enhancement, management improvement for airborne molecular contamination (AMC) is proposed in the paper from Pfeiffer Vacuum and GLOBALFOUNDRIES. As AMC is a key contributor to yield loss, control of AMCs is very important to yield enhancement. The concepts of DOSE and $DOSE_{lim}$ and evaluation of the various solutions are discussed in the paper.

In the paper from Texas Instruments, several effective ways to reduce process costs by chemical savings are introduced – a key topic, as usage of chemical materials in semiconductor manufacturing fabs is huge and costly. Good methods for chemical saving include chemical reuse, increased bath lifetime and equipment reconfigurations.

Click here to return to Table of Contents

Hitachi High Technologies Introduces the New M-9000XT Lynx Etch System

PRINT this article E-MAIL this article

Hitachi High Technologies, Inc. (HHT) is proud to introduce the semiconductor industry's highest-volume production etch tool for critical and non-critical etch layers.

The HHT M-9000XT Lynx is a linear production platform that offers best-in-class throughput per footprint of any etch cluster available to the market. The Lynx platform offers a significant advancement in extendibility by integrating from one to nine process chambers, thus allowing the M-9000XT to be configured to meet the exact needs of SC manufacturers. The ability to add process chambers allows the M-9000XT to grow with production capacity requirements. Manufacturers can take advantage of a small R&D investment by installing a single-transport unit module to support one to three process chambers. As the development moves toward production, the SC manufacturer can add up to two additional transport unit modules, allowing a maximum of nine process chambers to support high-volume manufacturing (HVM). Flexibility and extendibility are the advantages to the new Hitachi High Technologies M-9000XT.

The HHT M-9000XT is capable of integrating different chamber types but the HHT Microwave ECR or M-XT chamber is the primary chamber for today's state-of-the-art etch processes. The M-XT plasma etch chamber delivers the best etch uniformity, tightest process control and highest productivity of any etch system available to the SC market.

Etch uniformity is the key design concept for the M-XT. HHT retained the Microwave ECR plasma source and all its advantages while redesigning the gas delivery and evacuation of the chamber to provide superior uniformity of reactive gases reaching the wafer surface and byproducts being removed from the process chamber.

Combining the highest-capability etch chamber with the new ultra-high-throughput platform allows HHT to offer the most significant new etch tool to the SC market. When your fab is ready to shift into high gear, contact Hitachi High Technologies, Inc. to learn more about the possibilities to improve your process and increase your productivity with the HHT M-9000XT.

HITACHI
Inspire the Next

Hitachi High Technololgies supplies plasma etch systems to the world's leading SC and HDD manufacturers. HHT customers rely on Hitachi's technology, innovation and reliability to help them succeed in creating today's most advanced microprocessors, DSPs, memory devices and HDDs. Hitachi etch systems are renowned for their superior technology and production-proven reliability in the most demanding SC and HDD manufacturing environments.

Hitachi High Technologies America
1375 N. 28th Ave.
Dallas, Texas 75261-2208
1.877.ECR.ETCH
Etch.Sales@Hitachi-HTA.com
www.Hitachi-hta.com

This Future Fab chapter is sponsored by Hitachi

FRONT END OF LINE

Monitoring Fluid Suspensions: Current Results & Perspectives of a Novel Method Based on Light Scattering

M.A.C. Potenza,[1] T. Sanvito,[1] A. Pullia,[1] D. Di Cola,[2] D. Lavalle[2] G. Fazio,[3] A. Filippini,[3] L. Caudo,[3] G. Spinolo[3]
[1]Physics Department, University of Milan,
[2]Technofittings S.r.l., Rome, [3]Micron S.p.A., Italy

Abstract

We present the results of a feasibility study for the exploitation of a novel light-scattering method for monitoring water suspensions of submicron particles of interest for nanoelectronics fabs. In a collaboration between the University of Milan and Techno Fittings S.r.l., we realized a prototype of a device based upon a recent optical method for single particle sizing that is calibration free and provides results without any free parameter. We show results obtained with calibrated, monodispersed spherical particles, as well as samples with broad size distributions of grains made of different materials. The suitability of the method for specific nanoelectronics applications is discussed.

We recently proposed[1] a novel light-scattering method for measuring the size of single submicron particles, claiming a potential interest for performing measurements of interest for nanoelectronics fab processes. This method relies on the illumination of a single particle brought into the scattering volume and the consequential interference of the scattered wave front with the transmitted beam. Exploiting the fundamentals of optics, more information than the basic cross-section measurement can be recovered from the time-dependent intensity distributions at a given distance from the scattering volume. This method is capable of overcoming the limitations of the current optical methods measuring single particles, typically due to the presence of stray light, fake events, multiple events, determination of the traversing position of the particle within the light beam, calibration problems and rejection of air bubbles. For a complete overview of the traditional methods, one can refer to Terrell, and Mitchell and Bast.[2,3] On the basis of some typical issues at the nanoelectronics fab processes – such as, e.g., the quality of the slurries used for chemical mechanical polishing (CMP) processes – we designed and realized a prototype especially dedicated to the aim.

The method is the result of the activity performed during the last few years at the University of Milan, and the recent participation of all the authors to the European project IMPROVE, within the ENIAC platform. During the last 10 years, several optical methods have been conceived and realized based on the measure of the transmitted light just downstream a sample.[4-6] These schemes are as simple as possible, with almost no instrumentation needed and free from any complex alignment procedure. The measure of the transmitted light provides determination of the overall power reduction of the beam, which is ultimately a perfect measure of the particle cross section. At the same time, this geometry also promotes a self-reference condition, where the intensity distribution at a distance from the scattering volume is given by the interference of the scattered wave front and the transmitted beam. The self-reference condition naturally gives an intrinsic calibration of the signals, and allows the development of a model describing the instrument without any free parameter.

Particle suspension is brought into the device after a very high dilution, such that a single particle at a time is statistically present within the scattering volume. The flux is forced along a given, known direction at a definite speed (0.1 m/s, in our case), and passes through a focused laser beam. The intensity distribution due to the interference between the wave front scattered by the particle and the main transmitted beam changes with space and time depending on: 1) the size of the particle; 2) the refractive index of the particle; 3) the position of the particle across the light beam. Once the velocity (direction and speed) is known, one can cope with just two parameters the particle is endowed with (see below for further details). Therefore, the fast acquisition of intensity signals is a crucial point here, and resulted in the development of a dedicated front-end electronics (FE). Indeed, the ultimate limit of this method is ideally represented by the capability to tell apart the true signals generated by particles passing through the light beam from the very intense transmitted beam. This is an issue here, since the signals to be detected are as small as several 10^{-4}, corresponding to the limitations imposed by the current laser sources' stability (relatively cheap sources are considered here, as imposed by the fab environments). The developed FE is an innovative preamplifier capable of an automatic, real-time rejection of the DC signal, based on a double-ring negative feedback architecture with good stability and linearity. It physically separates the DC and the AC signals, bringing the small fluctuations to zero average.[7]

Under these conditions, a sensibility close to the limits of current optical devices has been achieved. Depending on the particle refractive index, the current sensibility is approximately 150-200 nm in diameter. Notice that this is not the ultimate sensibility for the present method, which can be capable of detecting and measuring the size of much smaller par-ticles if a more tightly focused beam is used. On the other hand, a decrease of the scattering volume will appreciably reduce the counting rate. The current sensibility represents a trade-off that permits one to obtain relatively fast measurements of a suspension, with a resolution down to approximately 20 nm in the lower size range and an accuracy of several percent for each bin of the distribution. We care-fully analyzed the laser stability issue, and compared several sources as detailed in Sanvito.[8]

After the acquisition of intensity signals with the FE, the data reduction scheme actually represents a fundamental feature of the method. A procedure of pulse shape analysis (PSA) permits: 1) a strict classification of the measured signals, telling the position of the particles within the scattering volume; 2) access to information beyond the sole cross section on the basis of the peculiar signatures that are present in the intensity distributions. Briefly, the PSA operates as follows: 1) the time series analysis of two parameters, *a(t)* and *b(t)*, are determined; 2) any fake, multiple or bubble event can be eliminated; 3) the position of the particle passing though the light beam, and the corresponding intensity profile of the illuminating beam at that position are determined; 3) two physical parameters, *A* and *B*, are determined from the time series *a(t)* and *b(t)*. Parameter *A* is related to the particle cross section (that slightly depends on the particle's refractive index); parameter *B* depends on the optical thickness of the particle; i.e., the product of the diameter and the refractive index. Note that just due to the self-referencing conditions, each time series can be normalized so that *a(t)*, *b(t)*, *A*, *B* are pure numbers.

In Figure 1, we show the results obtained with water suspensions of calibrated, monodispersed polystyrene spheres (200, 240, 290, 430 nm in diameter). The size distributions are shown, recovered on the basis of accurate Mie computations for the optical cross sections. As can be seen, a very high resolution is obtained. While the sensibility of the device is given by the laser stability as discussed above, the resolution is mainly given by the PSA used to obtain the intensity profile encountered by the particle when traversing the beam. This depends on different experimental parameters, such as the beam shape quality and the laser noise. Note that although decreasing the particle diameter, the raw signals are more spread out due to unavoidable noise; the actual size resolution is higher just for the smallest particles, thanks to the stronger dependence of the signal amplitudes on the size itself.

Figure 2 shows the size resolution obtained for a suspension of ground powder. These tests have been performed to check for the reliability of the method with more realistic, non-ideal samples (as the polystyrene spherical particles are). In this case, the results show that the method

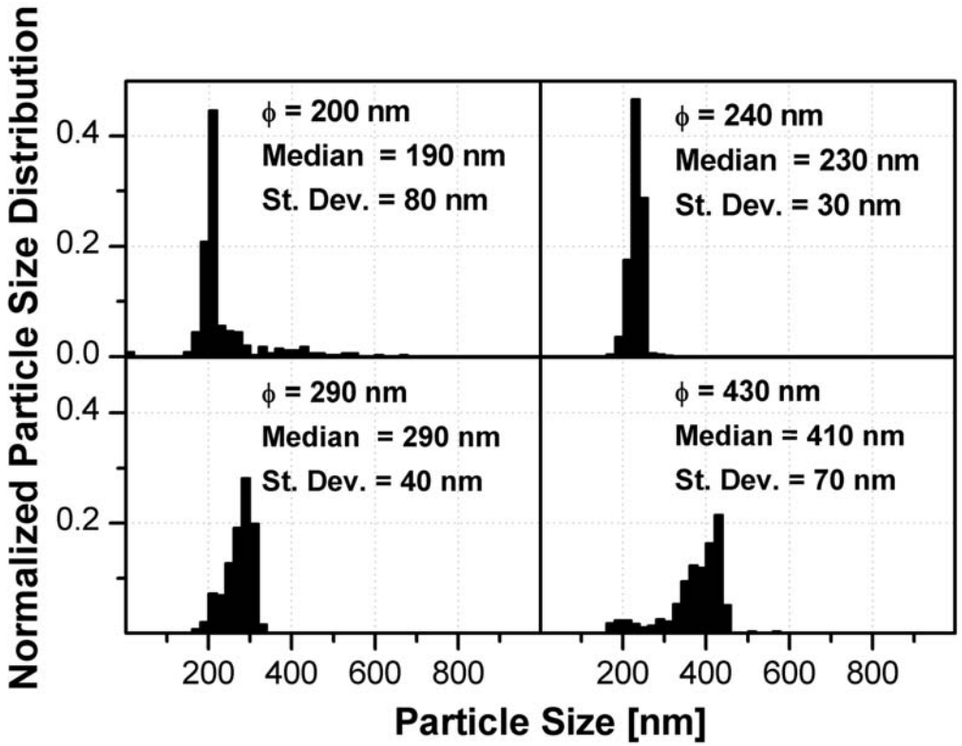


Figure 1. Results obtained with water suspensions of calibrated, monodispersed polystyrene spheres (200, 240, 290, 430 nm in diameter). Histograms indicate the fractional content of each bin.

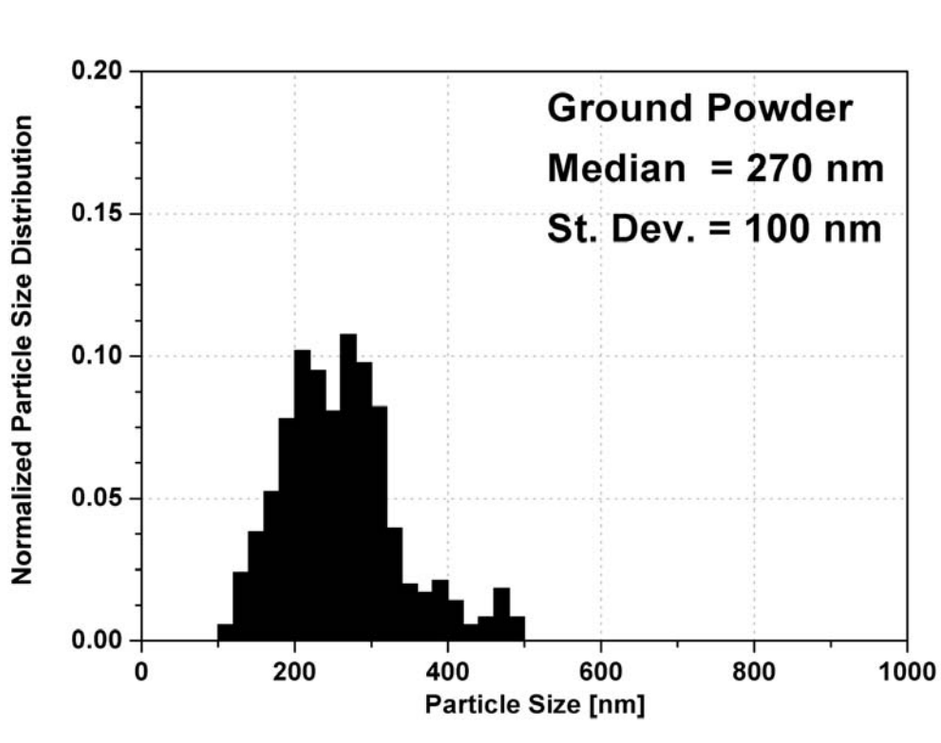


Figure 2. The Size Distribution Obtained With a Suspension of a Ground Powder Endowed With a Large Polydispersity and Nonspherical Shape of the Particles

maintains the performances. The only small limitations are: 1) the knowledge of the refractive index; and 2) the effect of the nonspherical shape of the particles. Both slightly reduce the resolution, as a given refractive index and the spherical approximation are assumed (as is usually done) for recovering the particle diameter.

In Figure 3, we show the size distribution obtained from a measurement performed with a suspension of ceria particles commonly used to polish optical surfaces. In this case, the refractive index is well-known (2.1), and can be used to determine the cross sections (under the spherical approximation). The binning is compatible to the size resolution, while the error bars are obtained from the Poisson statistics for each bin. As can be seen, the method can give insight into the details of a size distribution, thanks to the superior resolution and reliability of assigning the size. This represents a feature that suggests potential applications in measuring sizes of submicron particles. This collaboration aims to assess the feasibility of such measurements for in-line monitoring and analysis of liquids for nanoelectronics fabs such as CMP slurries (similar to the ceria samples shown in Figure 3). A possible integration of the device within the advanced process control could result in a continuous monitoring of the slurry, with the possibility of automatic feedback on the process.

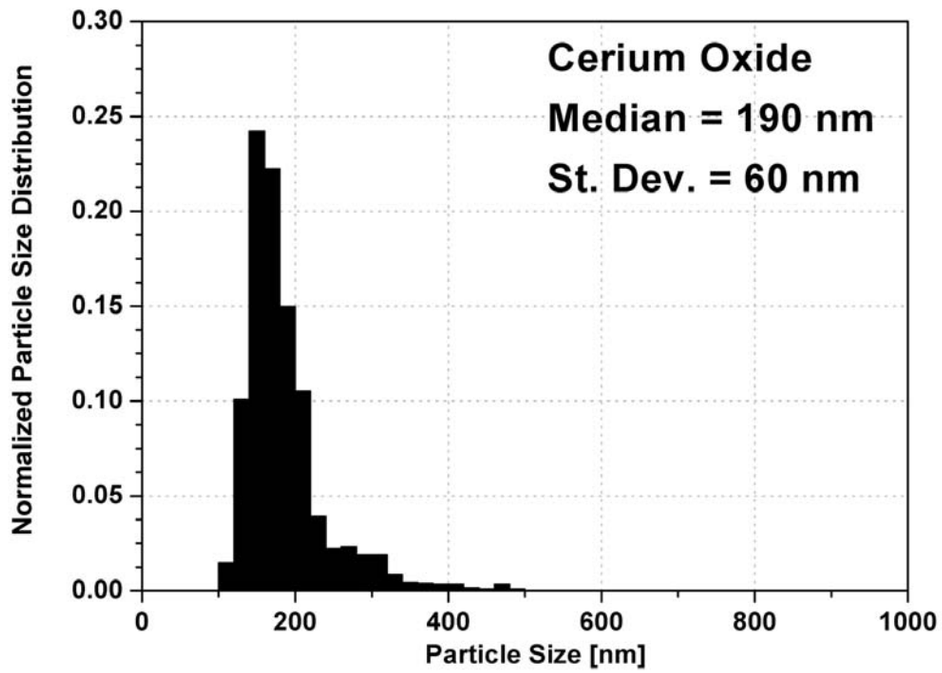


Figure 3. The Size Distribution Obtained From a Measurement Performed With a Suspension of Cerium Oxide, Obtained by Strongly Diluting a Slurry Commonly Used to Polish Optical Surfaces

Acknowledgments

The authors gratefully acknowledge the following: The apparatus has been realized by D. Viganò (Workshop of the Department of Physics, University of Milan). This work has been developed within the framework of the European project IMPROVE, with financial support from University of Milan and Techno Fittings S.r.l.

Endnotes

1. M.A.C. Potenza et al., "A Novel Method for Monitoring Fluid Suspensions: Current Results and Perspectives," Future Fab International, vol. 39 (2011)
2. E. Terrell, et al. "Understanding Liquid Particle Counters," PMS Inc. Boulder, Colorado (2006)
3. J. Mitchell and B. Bast, "Comparing Particle Counter Performance Characteristics for Analyzing High-Purity Water," Ultrapure Water, 48 (2000)
4. M.D. Alaimo et al. "Probing the Transverse Coherence of an Undulator X-Ray Beam Using Brownian Particles," Phys. Rev. Lett. vol. 103 194805 (2009)
5. M.A.C. Potenza et al. "Confocal zero-angle dynamic depolarized light scattering," Eur. Phys. J. E 31 69-72 (2010)
6. M.A.C. Potenza et al. "How to Measure the Optical Thickness of Scattering Particles from the Phase Delay of Scattered Waves: Application to Turbid Samples," Phys. Rev. Lett. 105 193901 (2010)
7. A. Pullia et al. "A Low-Noise Large Dynamic-Range Readout Suitable for Laser Spectroscopy with PIN Diodes," in press, in Rev. Sci. Instrum.
8. T. Sanvito, et al. "A method for characterizing the stability of light sources," in preparation.

About the Authors

University of Milan, Physics Department
M.A.C. Potenza – Assistant professor
Tiziano Sanvito – Ph.D. student in physics
A. Pullia – Associate professor

Technofittings S.r.l.
Domenico Di Cola – General manager
Daniele Lavalle – R&D manager

Micron S.p.A., Italy
Giuseppe Fazio – Senior Engineer, APC/AEC
A. Filippini – Senior Engineer, Contamination Control Lead
Leonardo Caudo – Senior Engineer, CMP Lead
Giulia Spinolo – Manager, CMP ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

FRONT END OF LINE

Yield Enhancement Through Airborne Molecular Contamination Management

Arnaud Favre,[1] Julien Bounouar,[1] Emmanuelle Véran,[1]
Astrid Gettel,[2] Marco Ackermann[2]
[1]adixen Vacuum Products by Pfeiffer Vacuum, Annecy, France
[2]GLOBALFOUNDRIES, Dresden, Germany

Introduction

With chip dimension shrinking, airborne molecular contamination (AMC) is well-known as a key contributor in avoiding yield loss and quality degradation.[1-2] And as for everything related to yield, significant investments to fight against this invisible enemy are made. As a consequence, there are as many strategies as there are semiconductor fabrication plants; bay and EFEM filtration; wafer and reticule stocker with AMC filtration or purge capabilities; new substrate carrier materials; stand-alone, EFEM or stocker purge; Qtime management – it is quite the mixed bag.

To solve this problem in a cost-effective manner, we need to fully understand the causes and their effects, and finally manage a global solution addressing each cause. In this paper, the importance of AMC-behavior understanding will be highlighted and the links between AMC and yield will be exposed. Finally, current solutions will be described and evaluated. An example of AMC management results in a 300 mm advanced fabrication plant will be then discussed.

AMCs

Species to Be Controlled

One of the Yield Enhancement ITRS group's topics is Wafer Environment Contamination Control (WECC). This subgroup is involved in defining the type and thresholds of AMCs that may lead to wafer defects. In 2011, a restructuration of the YE WECC table was published; one of the goals was to set the difference between cleanroom and FOUP contamination species and thresholds. Please refer to the YE3 table in the Yield Enhancement chapter: (http://www.itrs.net/Links/2011ITRS/Home2011.htm).[3]

From this YE3 table, we can extract the list of AMCs that can lead to wafer defects, in order to define the species that needs to be controlled inside FOUPs:

- Total inorganic acids
- HCl
- HF
- HBr
- HNO_x
- Total organic acids
- Total bases
- Total other corrosive species
- H_2S
- Total sulphur compounds
- Volatile organics (with GCMS retention times ≥ benzene, calibrated to hexadecane)

AMC Sources Inside FOUP

There are three AMC sources in FOUP environment (Figure 1).[4-5]

- *Cleanroom contamination*, through the FOUP leakage or filters – As cleanrooms are well-controlled, this contamination can be considered "under control" in comparison with the two other sources of contamination.
- *Wafer outgassing* – After each process step, wafers are stored inside FOUPs, waiting for the next process step. Wafers will then outgas a significant amount of last process byproduct.
- *FOUP outgassing*, coming either from the polymer material itself, or from the outgassing of previously adsorbed molecules on the polymer (memory effect).

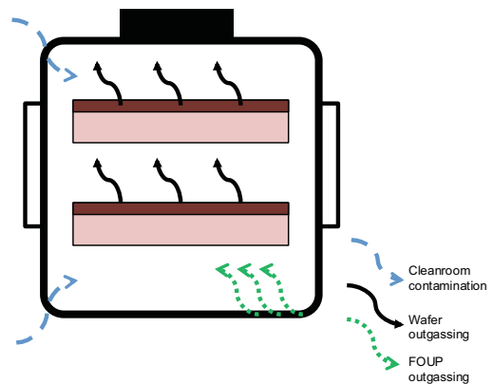


Figure 1. AMC Contamination Sources in FOUP

Mechanism of Wafer Degradation

From the two AMC main sources inside FOUPs, mechanisms of wafer defects due to AMC are illustrated in Figure 2.

Shrinking dimensions, together with process gas quantities, which increase in tandem with throughput ramp-up (less purging time/wafer), are the root causes for wafer AMC outgassing. As a consequence, and as FOUP is a closed environment, these wafers will outgas during the storage between two process steps in the FOUP environment, and will lead to an increase in AMC concentration. Chemical reaction may occur as well as defect growth, leading to yield loss (Figure 3).

Moreover, the ability of polymer FOUP to absorb molecules can be relatively high[6] and degassing of absorbed molecules could last for days or even weeks. FOUP wet cleaning is not efficient enough to remove AMC from the FOUP. As a consequence, cross-contamination can occur with new wafer batches.

The Concept of Dose

As explained above, we can summarize wafer damage due to AMC with the schematic found in Figure 4.

Interaction of AMC, substrate materials and moisture during wait time between two process steps may lead to particle growth and yield decrease. AMC concentration [AMC] and storage time (St) are key parameters to control; that is the reason we introduce the "DOSE" concept:

$$DOSE = [AMC]*St$$

Moisture cannot be considered completely as a contaminant by itself, but it can act as a catalyst that increases the impact of some AMCs. As an example, a

higher moisture level can increase the corrosion effect of acids.

DOSE is thus a "quantity of contamination," and defects will appear when DOSE > $DOSE_{lim}$, where $DOSE_{lim}$ is a function of material surface and contaminant type. The current trend is a decrease of $DOSE_{lim}$, as new materials such as ultra-low-K are introduced in production. This is the reason AMC-related issues are more and more critical to FOUP-related contamination.

AMC Concentration During Storage

There are different concentration behaviors inside FOUPs, and it is important to understand that they are related to the last wafer process step (Figure 5).

- Right after a process, the FOUP will be first open on the EFEM, and AMC concentration can be considered almost zero (depending on EFEM filtration or, without EFEM filtration, on cleanroom quality).
- Then, wafer (and/or FOUP) outgassing takes place. We observe a competition between two phenomena. Wafer outgassing leads to an increase of AMC concentration inside the FOUP, while FOUP leakage and AMC adsorption on FOUP inner surfaces lead to a decrease of concentration. This competition is represented by a second-order function.
- The maximum in AMC concentration can occur right after the FOUP closure, after the last wafer has been inserted, or later during the storage time.

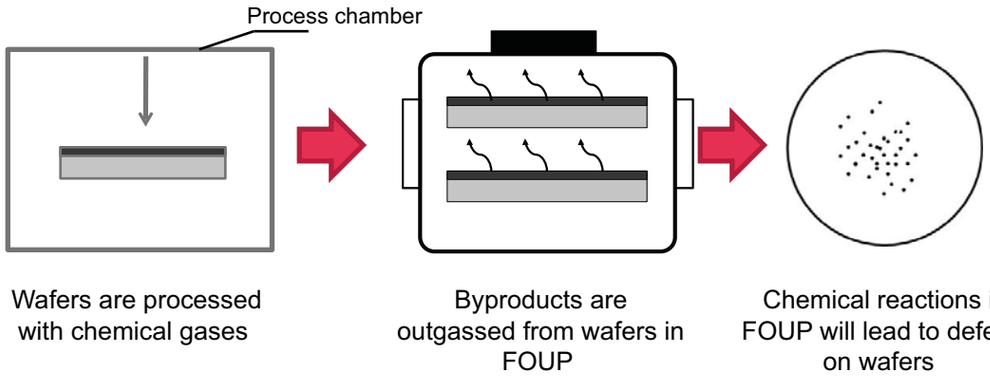


Figure 2. Defects Due to Wafer Outgassing

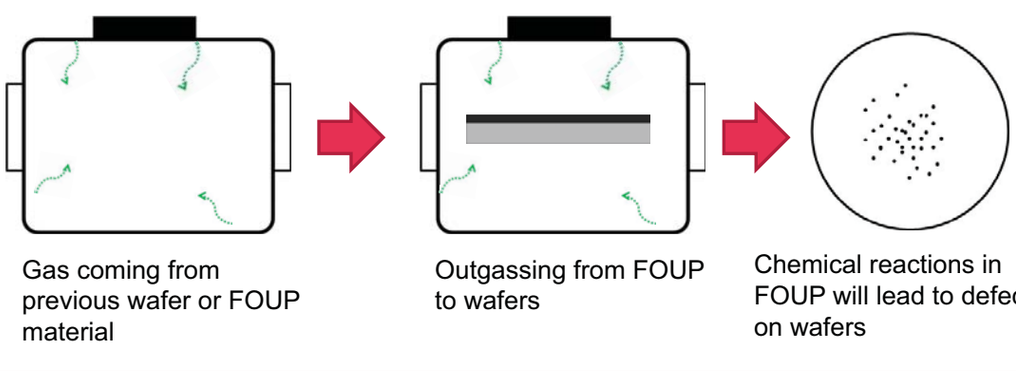


Figure 3. Defects Due to FOUP Outgassing

As a conclusion, DOSE can be expressed as the integration of these curves, i.e., integration of concentration with time. As DOSE is a quantity, its unit is [mol] (Figure 6).

In this example, $DOSE_{lim}$ cannot be exceeded after process 1. Storage after this process is not critical. However, after processes 2 and 3, defects will appear respectively after t_{max2} and t_{max3}. Actions need to be undertaken for these two critical processes.

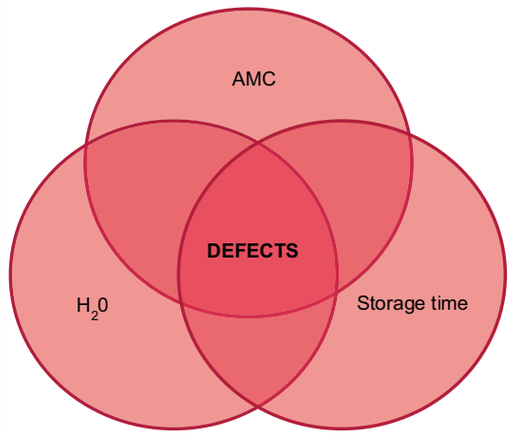


Figure 4. Schematics of Defect Apparition

Control of AMCs

FOUP Measurement

To control AMCs, it is crucial to monitor their concentration *[AMC]* inside FOUPs. There are two general methods: offline and online measurements. As the concentration of AMC is time dependent, AMC measurement must be carried out in the production environment to be able to evaluate the AMC concentration behavior, and to set up quickly the containment actions.

adixen Vacuum Products, by Pfeiffer Vacuum, has developed and patented an innovative equipment. This equipment measures *[AMC]* (total acids, total amines, total volatile organic compounds and moisture) at ppb_v level inside FOUPs (with or without wafers inside) within two minutes, and can be integrated into the production flow. This equipment – APA302 – is widely used in semiconductor fabrication plants for advanced nodes, such as at GLOBALFOUNDRIES, and enables rapid identification of AMC issues and solution qualification.

Current Solutions

Various industrial solutions are available, currently installed in semiconductor fabs. However, before listing and evaluating them, it is necessary to identify the parameters they are dealing with.

Coming back to *DOSE* expressions:

*DOSE = [AMC]*St* with apparition of defects when *DOSE* exceeds $DOSE_{lim}$

$DOSE_{lim}$ cannot be modified, as it depends on wafer materials and process gas. As a consequence, *AMC* concentration *[AMC]* and storage time *St* are the only two parameters to consider lowering the *DOSE*.

Many industrial solutions address these two parameters.

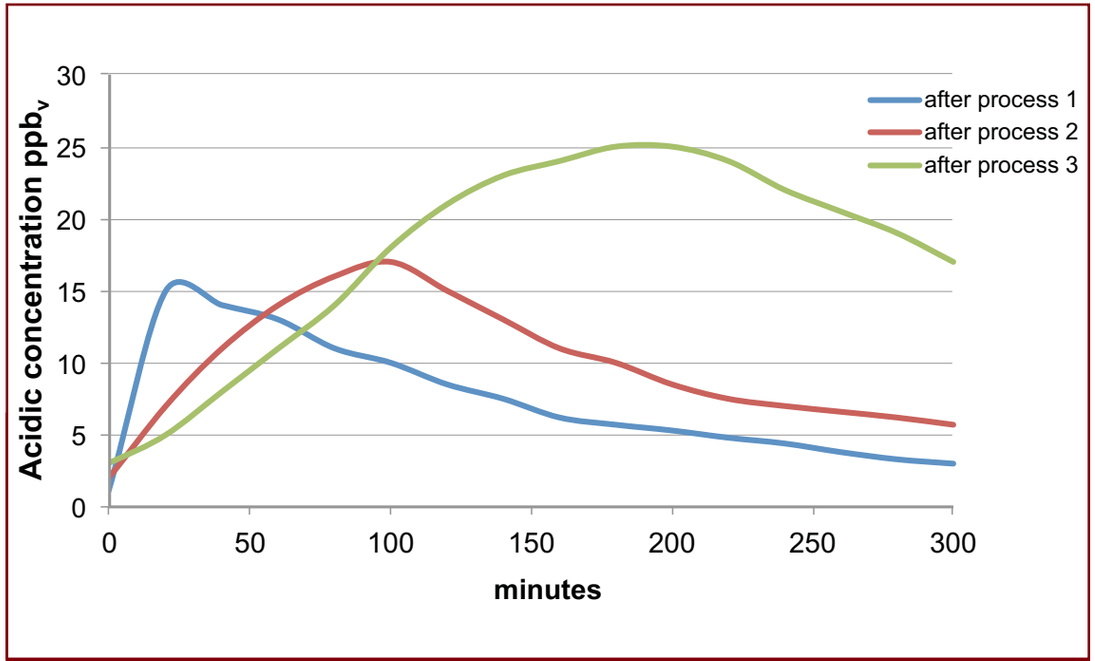


Figure 5. Example of Acidic Concentration Evolution During Storage After 3 Different Process Steps

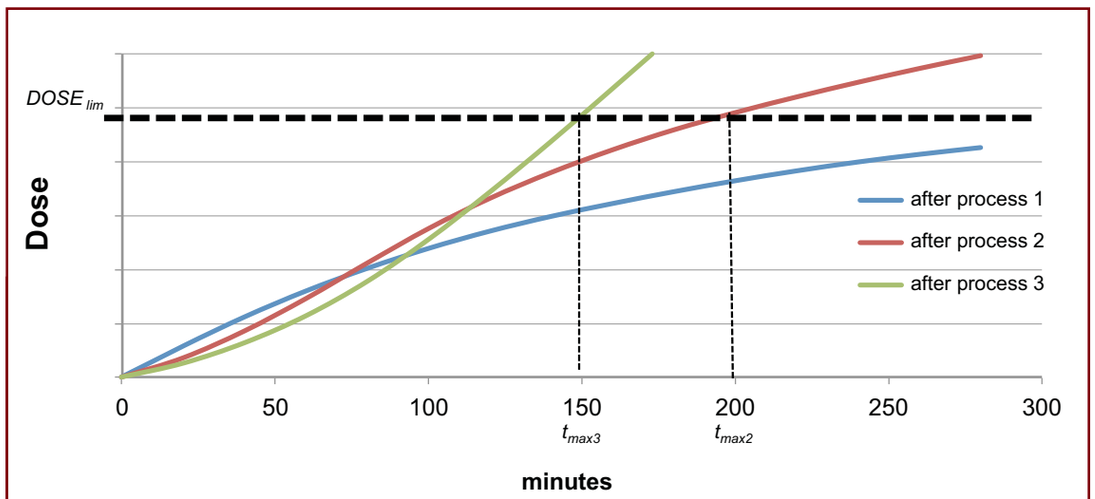


Figure 6. DOSE Accumulation After 3 Different Processes

Queue time (Qtime): setup of maximum duration between two process steps. With dimension shrinking, maximum duration is decreasing drastically (sometimes <2h) and it is more and more difficult to handle in a production environment without huge investment. If the Qtime is exceeded, wafers are either scrapped or reworked.

FOUP Change: After the critical step, wafers are removed from the initial FOUP and introduced in a clean FOUP through the wafer sorter.

FOUP N_2 purge: Specific FOUPs have the ability to be purged with clean N_2 either on a specific load port, or in a specific stocker.

EFEM N_2 Purge: During the wafer processing, the FOUP is purged with N_2 on a specific N_2 EFEM.

Lot Split: Before the critical step, 25 wafer lots are divided in different lots (fewer wafers in the FOUP, less outgassing).

Vacuum Purge: This innovative solution has been introduced and patented by adixen Vacuum Products, by Pfeiffer Vacuum, and has demonstrated yield enhancement. After loading the chamber with a FOUP containing wafers, the pressure in the chamber is reduced to < 0.1 mbar. Then the decontamination process is applied and contamination is removed. After this, the chamber is purged with clean nitrogen and returned to atmospheric pressure. The wafers and FOUP are now protected from contamination and Qtime can be extended to one day.

Table 1, based on various experiments, [7-10] describes impact on *[AMC]* and St for each solution:

Currently, almost all advanced fabs are using at least one of these solutions, and most of them are using several, depending on the process steps. GLOBALFOUNDRIES is one of the leaders in AMC management in the semiconductor industry.

Conclusion

The control of AMCs in FOUPs has become a major stake to ensure an optimum yield. ITRS and end-users are looking for better AMC control inside FOUP. This paper has shown the importance of con-

Solutions	[AMC]	St	Solutions Global Evaluation
Qtime	0	+	• Need process tool investments • Increase production complexity
FOUP change during process	+	0	• No action on wafer outgassing
FOUP N_2 purge	+	0	• Need continuous N_2 purging to be efficient
EFEM N_2 purge	+ -	0	• Need to be coupled with N_2 purge during storage to be efficent
Lot split + Qtime	+ -	+	• Need process tool investments • Increase production complexity
{FOUP + wafer} Vacuum purge	+ +	0	• One action to suppress both FOUP & wafer outgassing issues • Yield enhancement

Table 1. Evaluation of Solutions

cepts like *DOSE* and $DOSE_{lim}$ to define the right solution to AMC yield issues. Evaluation of the various solutions has been presented.

Moreover, as new materials will be used in production for the next-generation nodes, $DOSE_{lim}$ should become so restrictive that alternatives to atmospheric-pressure transport between critical processes would need to be investigated.

Endnotes

1. T.Q. Nguyen et al. "Identification and quantification of FOUP molecular contaminants inducing defects in integrated circuits manufactuing," Microelectronic Engineering (2012)
2. H. Fontaine et al. "Impact of the volatile acid contaminants on Cu interconnects electrical performances," ECS Transactions, vol. 25 (5), pp. 79-86 (2009)
3. International Technology Roadmap for Semiconductors (ITRS), Yield Enhancement, 2011, www.itrs.net
4. H. Fontaine et al. "Study of the volatile organic contaminants absorption and their reversible outgassing by FOUPs," Solid State Phenomena, vols. 145-146, pp. 143-146 (2009)
5. H. Fontaine and M. Veillerot, "Plastic containers contamination by volatile acids: accumulation, release and transfer to Cu-surfaces during wafers storage," Solid State Phenomena, vol. 134, pp. 251-254 (2008)
6. P. Gonzáles et al. "FOUPs Polymers Against AMCs: The HF Case," Future Fab international, issue 42
7. T. Kamoshima et al. "Ambient Gas Control in Slot-to-Slot Space inside FOUP to Suppress Cu-Loss after Dual Damascene Patterning," ISSM 2007
8. A. Favre et al. "Yield enhancement through new solutions for queue time containment," ISSM 2008
9. S. André et al. "Yield Enhancement Using New Solutions for Queue Time Containment," ARCSIS 2008
10. C. Martin et al. "Metal corrosion, what is the solution?," ARCSIS 2007

About the Authors

Arnaud Favre – Head of the R&D department of Advanced Systems, adixen Vacuum Products

Julien Bounouar – Process Manager, adixen Vacuum Products

Emmanuelle Véran – Semiconductor Market Manager, adixen Vacuum Products

Astrid Gettel – Project Leader, Factory Automation Tech Development department, GLOBALFOUNDRIES

Marco Ackermann – Principal ETCH Process Engineer, GLOBALFOUNDRIES ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Point-of-Use Chemical Reuse

Mark Simpson and Allen Page
Texas Instruments

PRINT this article E-MAIL this article

Wafers susceptible to particle and residue contamination necessitate specialized chemical mixtures to clean and etch them during the manufacturing process. Organic amine-based solvent strippers are required for removal of etch residue, often followed by chemicals such as isopropyl alcohol (IPA) and/or deionized water to clean off the original chemical. "Wet" processing equipment contains multiple tanks in sequence filled with these chemicals. They must be frequently dumped to waste and replenished to keep contaminates from being redeposited onto wafers. Chemical use is therefore one of the higher costs in wafer production due to the quantity used and a stringent need for cleanliness.

This article reviews several methods for chemical cost savings in post-metal wafer processing from bath life extension to chemical point-of-use reuse with emphasis on qualifying process change through inline controls and monitoring. Even with high customer requirements to produce wafers, reduced costs can be achieved through creative and effective techniques.

High Cost of Chemicals in a Fab

Semiconductor manufacturing commands many processing steps that use selective chemicals as part of its cleaning or etching phases. Aqueous and solvent-based strippers formulated to remove organic residues, such as photoresist or post-dry etch material, are usually found under the supplier's specialty chemical section. These solvents are often difficult and expensive to produce and a challenge to dispose of in an environmentally conscious way without amplifying significant wafer costs. In addition, wafers that go through a specialty solvent immersion or spray are regularly followed up with a rinse in IPA. Although not as expensive as the solvent strippers in terms of liters-for-liters, chipmakers must use higher quantities of semiconductor-grade IPA and pay a premium for its high purity and cleanliness. And unlike ethanol or methanol, IPA is generated from fossil fuel, which adds to the environmental concerns.

The demand for state-of-the-art integrated circuits fuels the need for an increasingly complex set of manufacturing rules. How do fabs provide top-quality chips to customers without compromising performance or cost? The solution may be as simple as extending a chemical bath life or sending a waste stream somewhere for reuse. This "low hanging fruit" gives the lowest return on investment (ROI); however, it is a great segue into higher-ROI projects.

Keep in mind that even simple and equally effective process changes require

review by a change approval committee. This will be discussed in detail later.

Extending Chemical Bath Life

The most efficient way to reduce chemical usage is to simply extend the bath life. Imagine it as reprogramming a dishwasher at home to prolong the water used in the wash cycle by reusing it for the rinse cycle. Water usage would be cut in half, but the dishes may not be as effectively cleaned. By the same token, it is easy to calculate that for every x time added to a bath, y amount of chemicals will be saved, but minimum specifications of removability and cleanliness may be affected. Figure 1 shows a typical process sequence using recirculated tanks. Pumps move solvent from outer to inner weir of each tank through a filter. Solvents are frequently dumped to drain and then refreshed to continue in its strength and to keep the filter from overloading with contaminates.

Bath life is determined by many factors: how quickly loaded the filter becomes with each run, concentration changes over time and/or loss of chemical integrity in effective cleaning and particle removal. Baths are automatically dumped as programmed by the engineer at either time-based or-run based intervals. Extending the initial IPA rinse and doubling the amount of wafers processed is one low-cost solution to stretch the amount of runs between chemical changes. No associated hardware modifications or re-plumbing and rerouting of piping are necessary. Only a recipe or parameter change is needed.

Waste Stream Recycle

Diverting spent water or chemicals for reuse elsewhere is not a new concept. Manufacturers routinely plumb drain water from wet hoods to the industrial wastewater system to aid in dilution of acid waste. Reuse of water is utilized in non-critical applications like HVAC cooling towers. Using the dishwasher analogy, imagine rerouting its drain to water the grass outside. Solvent waste comprised mostly IPA is usually collected and periodically emptied into a tanker truck for disposal. If water content is low enough, the waste can be recycled as fuel or for other applications by extracting the specific compounds needed for reuse. In certain cases, the waste solvent is actually purchased and the company avoids disposal costs. Special internal equipment programming is needed to divert drains to different outputs based on liquid type, cleanliness or concentration. Re-plumbing of drain lines outside the tool is necessary, but the change is downstream and has no effect on the process itself.

Concentration Replenishment

There are numerous chemistries that lose concentration over time due to evaporation, process drag-out, tank level replenishment or chemical reaction. Often, equipment is set up to bypass this problem by replenishing extra chemicals at a certain rate in order to extend bath life. It is common practice to add hydrogen peroxide into a hot sulfuric-peroxide mixture prior to wafer entry in a post-ash clean bath. Similarly, water-based solvent chemicals used in post-metal processes, which also operate in heated tanks, can have water replenished to remove polymers from wafers effectively. Concentration of water within the solvent is critical, and the evaporation rate of the water due to elevated temperatures limits its useful life. One solution is to determine water evaporation rate of a solvent over time. Then, deploy an integrated flow controller-metering valve to add de-ionized water back into the bath at the same rate of evaporation. An in-line chemical analyzer, like the WetSpec 200™ (manufactured by CI-Semi Inc.), measures water concentration every two to three minutes and can provide feedback to the metering valve.[1] This method allows for a 600 percent increase in the life of the bath, which transformed into an enormous reduction in cost and solvent supply

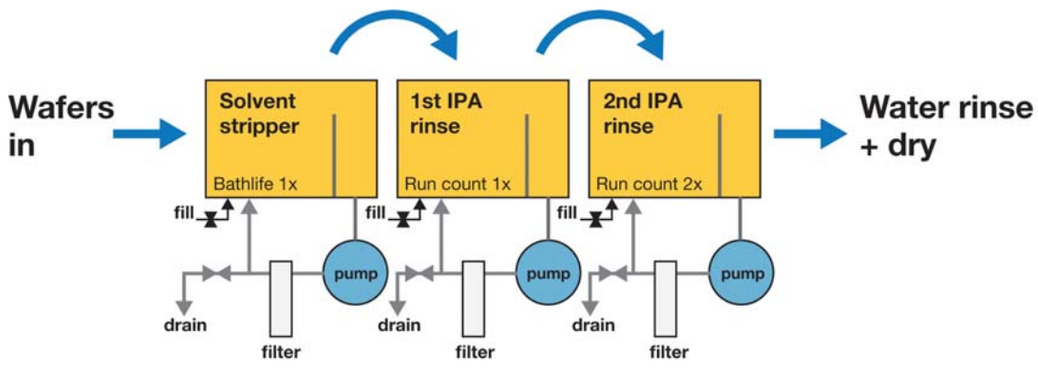


Figure 1. Typical Post-Metal Process Sequence Using Recirculated Tanks

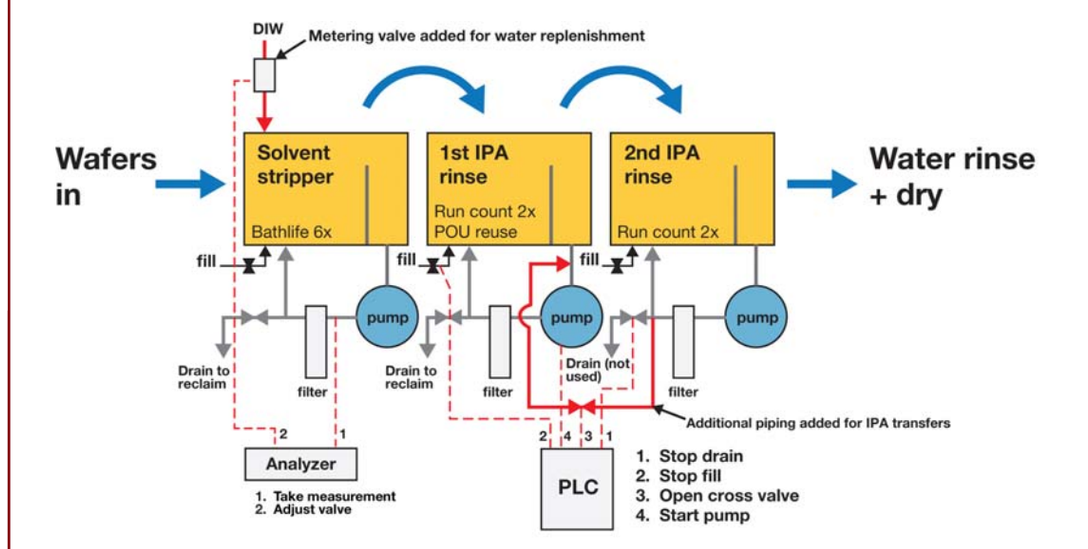


Figure 2. Bath Life Extension, Stripper Replenishment, Drain Reclaim and IPA Reuse

reduction.[2] Chemical concentration control can be challenging. There are risks associated with building a replenishment system into a process where solvent-to-water ratio is key: Too much water can cause corrosion of metal, and too little can cause ineffective polymer removal. In-line controls are crucial to ensure reliability and repeatability.

Point-of-Use Chemical Reuse

Each of the chemical cost-saving methods discussed thus far is relatively low risk and industry proven. In spite of extending bath life, much IPA is still consumed in post-metal processes. Because the sequence of carry-over from the solvent stripper bath makes the second in-line IPA bath cleaner than the first, connecting piping from the second bath to the first allows reuse of IPA at the point-of-use. By attaching I/O signals from the process tool to a programmable logic controller (PLC), engineers can control valves and pumps so that the IPA transfers from the second bath to the first at the start of its fill cycle.

Figure 3. Engineers monitor piping and valve status, I/O signals and alarm conditions through touch screens.

Revisiting the dishwasher example, cleaner water used in the rinse cycle can be saved for reuse in the wash cycle of the *next* load of dishes. This passive system only performs the transfer when the signal of the main tool starts to drain the first bath. This would reduce the amount of IPA chemical changes by half for a two- tank process – only one tank changes out. Even more savings can be yielded than the extended bath life scenario where eventually both tanks still dump. Reusing IPA at its point-of-use eliminates the need for any special storage system. Figure 2 shows the overall modifications that can be utilized for chemical cost savings on a single piece of equipment and the sequence of events during IPA transfers. The PLC utilizes a touch screen that features different areas to look at, such as overall piping and valve status, I/O signals and alarm conditions (Figure 3).

Qualifying the Process Change

Even simple process changes, such as chemical reduction methods, require qualification. Knowing the five steps to proposing change is essential:

1. *Proof of Concept* – Prove the concept will work for any process change through initial testing. This can include analysis of liquid samples for process of record (POR) versus new processes. Take Texas Instruments' life extension/replenishment project (Figure 4). Composition of the bath remains stable after 12 hours and up to 78 hours. Particle checks may also be utilized to prove the new process is as clean as POR.
2. *Initial Presentation for Change* – Present data outlining the concept to a change review board for initial approval. Include the number and variation of split wafer lots.
3. *Split Lot Results* – Present the results for the new process versus baseline, or POR. This often includes in-line defect, parametric and probe data. These tests need to be statistically equivalent and in specification. A limited release may then be granted.
4. *Limited Release* – Run a predetermined number of lots using the new process. Once these are completed, the configuration is returned to POR, awaiting results. Compile in-line defect, parametric and probe data again, and compare the limited release lots to POR lots for verification of sameness.
5. *Full Final Release* – Show all data to the final change review board. If the change is considered equivalent or improved over POR, then it will be granted full final release, and equipment may then be modified for the change. Specifications may have to be updated and personnel trained. Once the change is fully implemented, it is tracked for cost savings, improved probe data and improved throughput.

Summary

Fabs are conscious of the need to provide top-quality chips to customers who impose rigorous standards of performance while holding down costs in order to turn a profit. The ongoing price increase for depleting raw materials demands that engineers change the way they think. From simple changes in chemical bath life to

Bathlife, hr	Composition				
	Water	Amine	Inhibitor A	Inhibitor B	Fluorine
0.00	18.76	79.25	0.90	1.09	ND
12.00	15.49	82.31	0.93	1.27	ND
24.00	15.95	81.87	0.89	1.29	ND
36.00	15.97	81.79	0.89	1.35	ND
48.00	16.59	81.24	0.81	1.35	ND
54.00	16.53	81.24	0.83	1.40	ND
60.00	16.56	81.20	0.79	1.44	ND
66.00	16.46	81.29	0.79	1.45	ND
72.00	16.43	81.26	0.75	1.55	ND
78.00	16.35	81.38	0.75	1.51	ND

ND means < 1ppm

Analytical Methods

Stripper compositions were determined by gas and high pressure liquid chromatography. Fluorine level was determined by ion chromatography.

Figure 4. Initial Data Collection for a Change in Process Proposal

more involved equipment reconfigurations to recycle its resources at its point-of-use, manufacturers must take a discerning look at its processes and develop creative methods for cost reduction while maintaining process integrity.

References

1. M. Simpson et al. "Monitoring recirculated baths using NIR inline chemical analysis," Solid State Technology, pp. 44-46, October 2007
2. M. Simpson et al. "Achieving closed-loop process control of DI-Water replenishment," pp. 77-84, Micro Magazine, July 2004

About the Authors

Mark Simpson
Mark Simpson, TMG member of the Technical Staff, joined Texas Instruments in 2000 and currently holds a position as senior engineer in the Plasma Etch group at DMOS5 in Dallas. He received his B.S. in EET from Texas A&M University in 1987.

Allen Page
Allen Page joined Texas Instruments in 2007 and currently holds a position as Senior Engineer in the Plasma Etch group in DMOS5 in Dallas. He received his B.S. in math and physics from East Central University in 1977 and his M.S. in instrumentation from the University of Arkansas in 1981. ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

David G. Seiler
Chief, Semiconductor and Dimensional Metrology Division, NIST

Cost-effective manufacturing is critically dependent upon the development of defect detection, defect review and classification technologies. The 2011 ITRS Roadmap Yield Enhancement chapter points out that it is a challenge to detect multiple killer defects and to differentiate them simultaneously at high capture rates, low cost of ownership and high throughput. Multi-component architectures such as in high-end microprocessors and graphic processors are becoming more common, and their corresponding kill ratios thus become more difficult to estimate.

Garry Tuohy of GLOBALFOUNDRIES has put together an elegant method of calculating kill ratios and loss estimates for manufacturing defects that utilizes the discrete component results. The methods should generally be applicable when quantifying the defect-generated yield loss per wafer inspection step.

METROLOGY, INSPECTION & FAILURE ANALYSIS

Calculating Kill Ratios on Multi-Component Devices

Garry Tuohy
GLOBALFOUNDRIES Inc.

Abstract

The advent of multi-component devices has served to slow the reduction in die size. Relying on die-level contingency analysis to determine kill ratios for ever-shrinking design rules is no longer adequate. A means of utilizing the data from die components is necessary in order to preserve kill ratio accuracy. The application of a yield model is presented as a means of utilizing component-level data for kill ratio calculation.

Introduction

The progress to multi-component architectures in high-end microprocessors and graphic processors has been well under way for the past decade. Leading-edge microprocessors are now very heterogeneous in nature, containing a wide range of discrete component types (e.g., NorthBridge, PCIe, Display Port).

More recently, this trend has also appeared in general purpose microprocessors and even microcontrollers. This indicates that a method of calculating kill ratios and loss estimates for manufacturing defects that utilizes the discrete component results should now be more generally applicable when attempting to quantify the defect-generated yield loss per wafer inspection step.

The Challenge

The primary difficulty with utilizing component-level SORT data when calculating kill ratios are those die affected by a gross defect or a peripheral defect, both of which can result in the loss of component-level data.

In the case of the examples shown in Figure 1, the components of the middle die cannot be treated independently, but rather need to be considered as a single die-level failure.

To illustrate the issues associated with die-level fails, consider the synthesized wafer maps in Figure 2. The yield difference between the defective die (*Yd*) and non-defective die (*Yc*) is traditionally used to calculate the kill ratio according to equation (1). This results in a kill ratio estimate of 9.2 percent. Performing the same calculation at the component level but treating die-level fails as single elements results in a kill ratio estimate of 17.4 percent.

$$KR = 1 - (Yd/Yc) \qquad (1)$$

The accuracy and selectivity of the die-level correlation is clearly under suspicion because of the relatively large die size. However, the component-level calculation is also questionable because the defective components include a higher proportion of die-level fails (Figure 2c). The larger area of the die-level fails naturally have a higher probability of being defective than a true component and hence adversely influencing the kill ratio. This higher probability needs to be factored into the component-level yield values prior to calculating a component-level kill ratio.

The Impact of Component Area on Defectivity

Examining the difference in the probability of die elements being defective for typical die and component sizes over the range of wafer defectivity values produces the response shown in Figure 3. Unsurprisingly, at low defectivity levels (i.e., <20 defects/wafer), the difference in the probability of being defective between

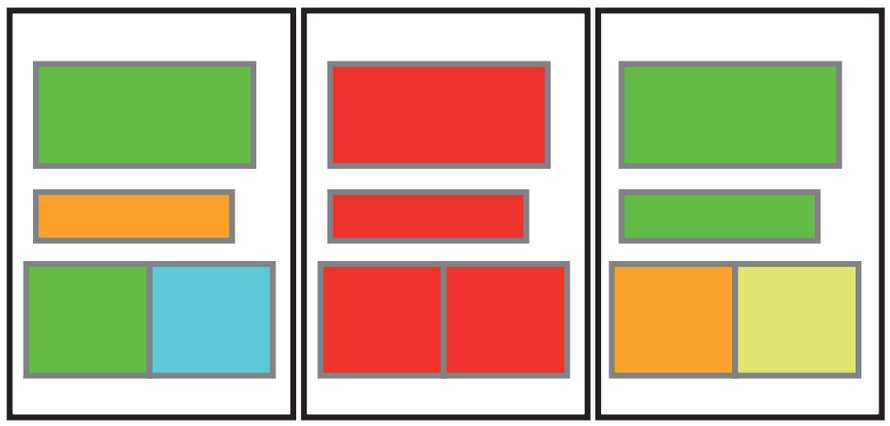


Figure 1. Example of 3 Die With a Defect-Generated Die-Level Fail on the Middle Die

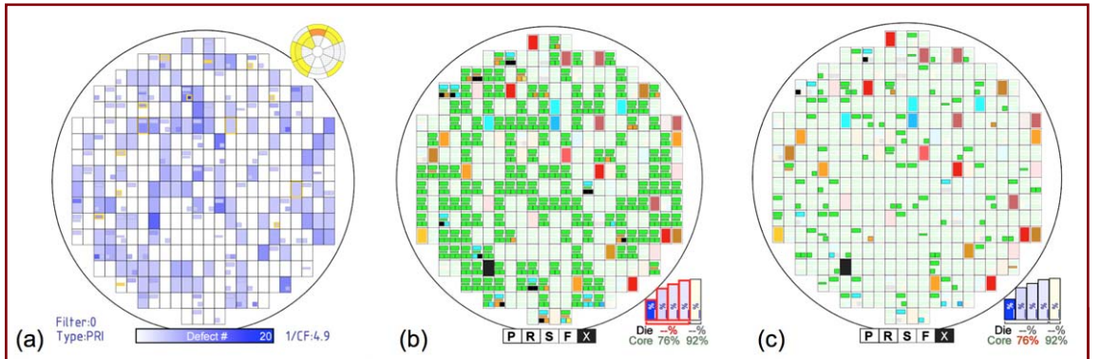


Figure 2. Artificially generated wafer maps showing (a) a defect density map and SORT maps highlighting (b) the defective die and (c) defective components plus defective die-level fails.

a die and component is low and would only marginally affect the kill ratios. Beyond this defectivity level, the difference in the probability becomes unacceptable, as it will increasingly put any die-level fails into the defective category, thus erroneously increasing the kill ratio.

A Solution: Area Correction

One method of resolving this issue is to use a yield model to translate the yield values for the defective and non-defective components to the equivalent yield for die elements of the size of a typical component. This yield conversion is performed according to equation (2), which is derived from the negative binomial equation as first proposed by Okabe, et al[1], Stapper[2] and later comprehensively validated by Stapper et al.[3]

These calculations are performed for the defective and non-defective die elements at every wafer step, where the clustering factor (α) as defined by equation (3) is obtained from the inspection scan where $\bar{\lambda}$ and σ^2 are the mean and variance for the number of defects per die, as described by Cunningham.[4] The actual component area is represented by *Ac* and the average area and yield of the wafer elements under investigation are represented by *Ai* and *Yi*, respectively.

$$Yo = 1/\left(1 + \left(\frac{Ac}{Ai}\right) * \left(Yi^{-1/\alpha} - 1\right)\right)^{\alpha} \quad (2)$$

$$\alpha = \bar{\lambda}/(\sigma^2 - \bar{\lambda}) \quad (3)$$

Applying equation (2) to the example wafer results in the yield conversions as shown in Table 1. The area-corrected yields result in a kill ratio of 4.6 percent. This is in line with the expected value for the primary defect mechanism at the inspection step in question, and indicates that the area-correction method does allow the component-level data to be utilized as a more precise assessment of kill ratios, for individual wafer inspection steps.

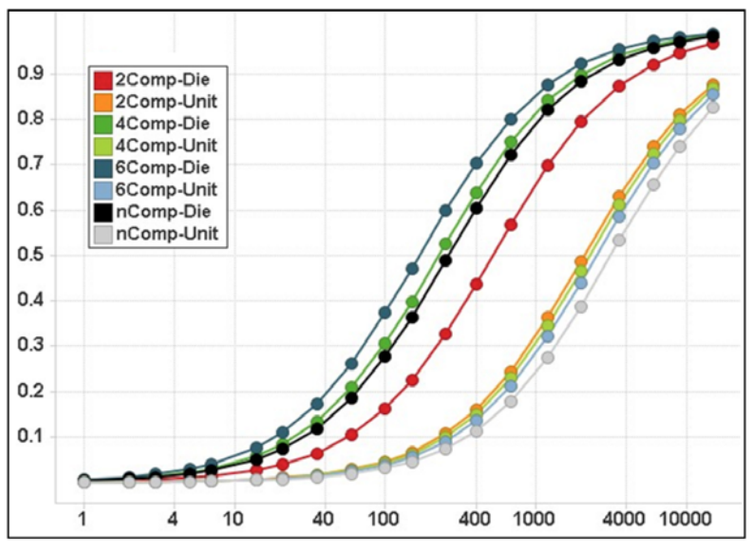


Figure 3. Probability of Die or Components Being Defective vs. Defects/Wafer

Calculating Loss From Component-Level Kill Ratios

Estimating the number of die lost from the component-level kill ratio requires a different treatment than that normally used by the die-level calculation. The method utilized is the same as that to estimate loss for die with a given number of defects where the kill potential per defect is known. In this case, instead of defects, the number of defective components per die is used. The number of lost die is given by equation (4).

	Yi [%]	*Yo* [%]
Defective Elements	76	87
Non-Defective Elements	92	94

Table 1. Defective and Non-defective Yields Before (*Yi*) and After (*Yo*) Area Correction

$$Die\ Loss = \sum_{i=1}^{n} D_i - (D_i * (1 - K_c)^i) \quad (4)$$

K_c = Component – Level kill ratio
i = number of defective components per die
D_i = number of die with i defective components

Special consideration needs to be given for defect signatures that show a high degree of selectivity for complete die-level fails (i.e., high defect kill potential, *Kc* >80 percent) and that have no clustering (i.e., clustering factor >10). In such cases, the average area of the defective units approaches that of the die area, and the absence of any clustering means that no kill-ratio scaling occurs because there are relatively few defective components per die. This results in the component-level loss estimate that significantly underestimates the loss that is accurately estimated by the die-level calculation. The die-level loss estimate should be used when these characteristics are observed.

Component-Type Specific Kill Ratios

Naturally, once the data for all components have been combined to produce an

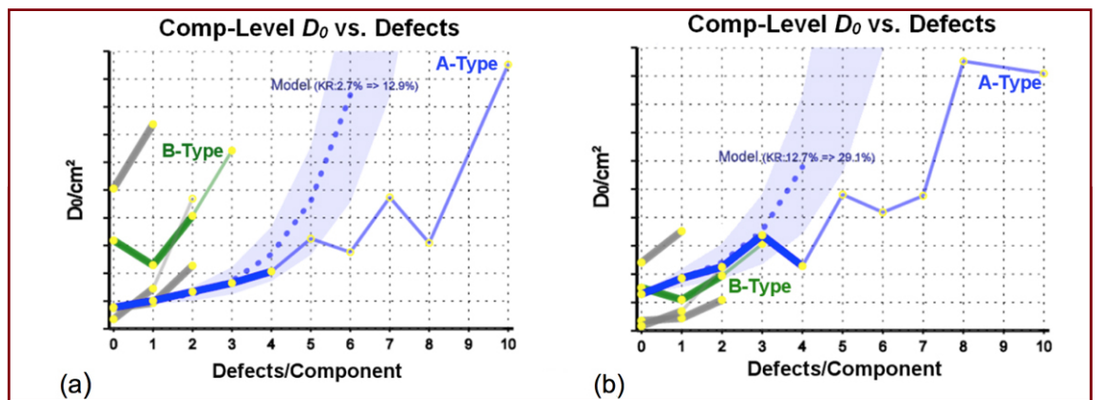


Figure 4. Visualization of specific component-type's yield response vs. the number of defects per component (a) including A-type component redundancy repair [2.7% per defect, rising to 12.9% for all A-types]; and (b) excluding redundancy [12.7% per defect, rising to 29.1% for all A-types]

overall kill ratio for a wafer inspection step, the following questions arise: What are the kill ratios for the individual component types? Do they display differing yield responses? Displaying the yield response versus the number of defects per component is best visualized by translating the yields into D_0 values.

The shaded blue areas in Figure 4 depict the expected response of the largest component type (i.e., the A-type component) if the yield relationship between the non-defective A-type components and those with a single defect is maintained over an extended range of defect density.

A monotonic response indicates a clear yield impact attributable to the relevant defects and lends certainty to the modeled kill ratio. The average component-level kill ratio for all component types should normally be within a range defined by the largest component type, from the kill ratio for components with a single defect to the average value for all A-type components. Any unexplained divergence would indicate the presence of an atypical failure mechanism.

Conclusion

The application of the negative binomial model has been demonstrated to allow component-level SORT data to be utilized for more precise kill-ratio calculation and subsequently yield loss assessment when quantifying the defect-assignable loss on a wafer inspection step basis.

Plotting of D_0 for each component type versus the number of defects per component can illustrate differences in the response of different component types to increasing defectivity for specific wafer inspections scans and serve to increase the confidence in the aggregated component-level kill ratios.

References

1. Okabe et al. "Analysis on Yield of Integrated Circuits and a New Expression for the Yield," Elec. Eng., Japan, vol. 92, pp 135-141, December 1972.
2. Stapper, C.H., "Defect Density Distribution for LSI Yield Calculations," IEEE Trans. on Electron Devices," vol. 20, No. 7: pp 655-657, July 1973.
3. Stapper, C.H. et al. "Integrated Circuit Yield Statistics," Proc. of the IEEE, vol. 71, No. 4: pp 453-470, April 1983.
4. Cunningham, J.A. "The Use and Evaluation of Yield Models in Integrated Circuit Manufacturing," IEEE Trans. on Semiconductor Manufacturing, vol. 3, No. 2: pp 60-71, May 1990.

About the Author

Garry Tuohy

Garry Tuohy received his B.Sc. degree in applied physics from the University of Salford, Greater Manchester, U.K., in 1994. He has worked in the semiconductor industry since 1996, mostly in the U.K. and Germany. Since 2003, he has been working in yield engineering in Dresden, Germany, for AMD and now GLOBALFOUNDRIES, where he is a member of the technical staff. His work primarily involves developing systems for kill ratio and defect loss analysis, and web-reporting. ■

Click here to return to Table of Contents

Click here to return to Table of Contents

Christo Bojkov

Senior Package Development Engineer, TriQuint Semiconductor

Cu-to-Cu Thermo-compression Bonding: The Reliability Challenge of TSV Integration

Higher IC density, reduced cost, improved electrical performance and high-reliability expectations have imposed significant challenges on the integration flows of TSV. As in most cases, the simplest way of forming a reliable interconnect brings the highest level of uncertainty and the need for profound fundamental studies.

This is the case with Cu-to-Cu thermo-compression bonding in TSV. From the very beginning of TSV development, the simple thermo-compression bonding of Cu has attracted numerous research groups, universities and companies. Data have been reported at different conferences from as low as room temperature bonding to as high as 600 °C. The bulk of data, though, comes from the 200-400 °C range, always accompanied with some form of a proprietary surface pre-treatment process of the Cu metal interface. These can be either wet (acidic rinse) or plasma (hydrogen) clean, pre-sputter Ar-etch or some form of organic protective layers. Usually a post-bonding annealing step is also required to further improve the crystalline structure at the interface. Preserving Cu-interface from oxidation and contaminations prior to bonding is a well-established process in the Cu-damascene flow, especially after the CMP step, but the need for physical bonding has necessitated applying pressure and the requirement for metallurgical connection between the two bulk Cu phases. In one example, engineers from EVG have shown that uniform crystalline structure must be formed between the two bulk phases, thus securing homogeneous and defect-free metallurgical connection. A similar conclusion has been reached by a team from imec (Belgium) developing the so-called "Insertion Cu-Cu bonding."

In this section, you will find another excellent example, from the Institute of Microelectronics (A*STAR) and Nanyang Technological University (Singapore) aimed at establishing high-reliability of the Cu-to-Cu bonding interface. This study is very attractive in providing a correlation of results from temperature cycling with contact resistance at the interface along with the thermo-mechanical stresses profiles developed at the TSV bottom section. Looking forward to their next article.

WAFER FAB & PACKAGING INTEGRATION

An Optimal Cu-to-Cu Thermo-compression Bonding Process Window Compatible With 3D Wafer Stacking and Stability

H.Y. Li,[1] L. Peng,[1,2] F.X. Che,[1] C.S. Tan,[2] G.Q. Lo[1]
[1]Institute of Microelectronics, A*STAR (Agency for Science, Technology and Research), Singapore
[2]Nanyang Technological University, Singapore

Abstract

The Cu-Cu thermo-compression bonding process being compatible with 3D integration is important since all aspects of the FEOL, TSV, BEOL, back-side metalization and assembly processes need to be considered. Bonding temperature and force are critical parameters during Cu-Cu bonding. A high bonding temperature induces high TSV stress in the Si substrate; meanwhile, high contact resistance and its instability are observed when the bonding temperature is <300 °C, according to our study. Therefore, bonding temperature of 300 °C is suggested for Cu-Cu bonding and applied for 3D-IC stacking.

Introduction

Cu-Cu thermo-compression bonding is a promising technology for 3D wafer stacking. Reports on Cu-Cu thermo-compression bonding have attracted significant attention.[1-3] Specifically, electrical and reliability results of 200 °C direct bonding were reported by Di Cioccio et al.[3] Cu-Cu thermo-compression bonding is critical technology to address the ever-scaling requirement of the bump pitch in order to increase the bump density. The important applications of Cu-Cu bonding include direct Cu-TSV bonding with Cu-pad to form 3D interconnect.[4-7] However, high-temperature Cu-Cu bonding during 3D wafer stacking can reduce negative impact on the process integration; for instance, with temporary bonding resulting in de-bonding during the back-side CVD, PVD and metalization.

Cu-Cu bonding conditions are critical for the devices with Cu-TSV in via-middle (VM) implementation. Cu-Cu bonding is performed after devices and Cu-TSV fabrication. Therefore, bonding temperature and pressure are applied on the Cu-TSV and active device as well. Thermal stress from TSV affects the active device in the nearby surrounding Si significantly. The bonding condition is known to affect the keep-out-zone (KOZ) between TSV and devices around.[7] Thus, the bonding temperature has been identified as the primary contributor to the buildup of TSV thermal stress compared with bonding pressure.[7] Therefore, Cu-Cu thermo-compression bonding temperature is an important parameter to be considered in the context of 3D process integration.

Materials	CTE (ppm/°C)	Young's Modulus (GPa)	Poisson's Ratio
Silicon	2.8	131	0.28
Cu	17	117	0.35

Table 1. Parameters Used in Thermal Stress Simulation

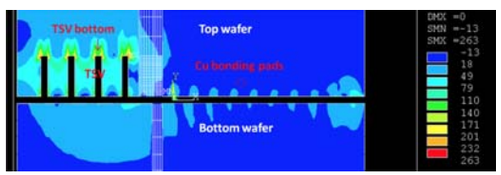


Figure 1. Thermal-stress Distribution When Heat and Pressure Are Applied and Hold at High Bonding Temperature

Bonding Conditions and Effects on Embedded TSV

High bonding force and temperature are required to ensure the bonding quality and site-to-site uniformity. However, the bonding temperature and force are loaded on Cu-TSV as well. As a result, TSV generates thermal stress in Si substrate during bonding process. The thermal stress pres-

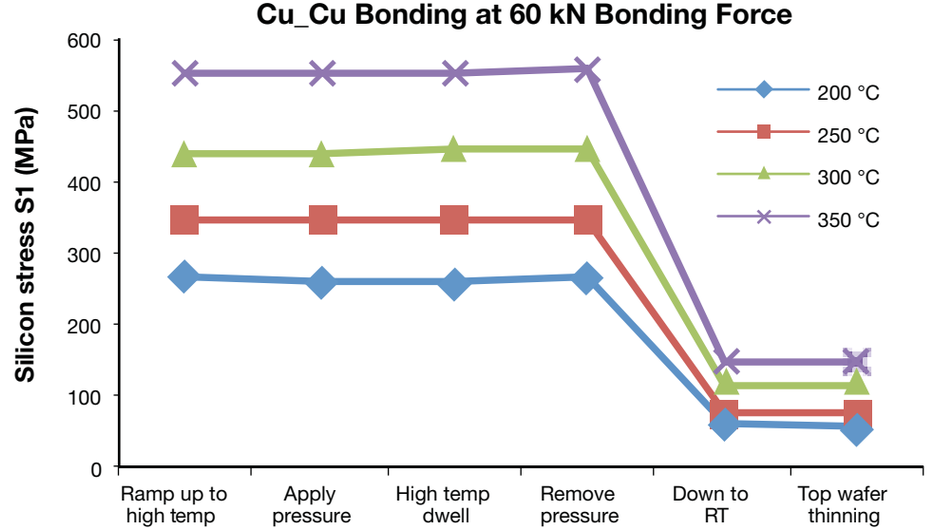


Figure 2. Simulation Results of Cu-Cu Bonding at 60 kN for 1 Hour Held at Different Temperatures

ents challenges during subsequent wafer processes such as back-grinding and TSV via reveal. Therefore, simulation of bonding force and temperature is necessary to guide the 3D process integration. The thermal stress in Cu-TSV (Φ: 5 µm, H: 30 µm) from Cu-Cu thermal-compression bonding is simulated by ANSYS. The simulation considers temperature ramp-up, applied pressure, high temperature dwell, pressure removal and cooling down to room temperature. The major parameters used are shown in Table1.

The maximum stress is concentrated around the TSV bottom corner area when heat and pressure are applied and held at high temperature. The result is illustrated in Figure 1. A second high stress is distributed in the Si substrate near TSV bottom. When the top bonded wafer was thinned down to this area by the back-grinding system, complicated mechanical stress easily induced wafer crack. The maximum stress near TSV in Figure 1 is not symmetric, which increases the via-revealing process challenge.

Extremely high thermo-mechanical stress could cause the Si substrate to crack. The stress weak point near the TSV bottom could affect wafer back-side thinning and via reveal. A previous study[8] showed that the bonding yield is improved with both the bonding force and the temperature. Therefore, it is essential to identify and select a Cu-Cu bonding process that results in the least-TSV-induced thermo-mechanical stress in the Si substrate. The maximum thermo-mechanical stress in the Si area near the TSV bottom is simulated for Cu-Cu bonding process at 200, 250, 300, and 350 °C with 60 kN bonding force. The maximum stress distribution during the bonding process is shown in Figure 2.

As shown in Figure 2, the Si-substrate is under high stress when the respective bonding temperature is applied, which is reduced during the cooling step. As expected, a higher bonding temperature results in higher stress. However, there is no significant variation in the stress value when the bonding pressure is applied, saturated at high temperature and removed within individual bonding process. Therefore, it is clear that the bonding temperature plays a significant role in the Si substrate stress buildup.

To further understand the role of bonding pressure, the simulation is repeated at 300 °C for 10, 30 and 60 kN of bonding force, and the results are shown in Figure 3. As can be seen, the maximum Si stress has no significant dependency on the bonding force. Again, based on an earlier study,[8] the bonding force merely increases the contact area to improve the overall bonding quality.

The conclusion drawn from the simulation results on the dependency of the maximum Si stress level in response to bonding temperature concurs with the earlier report.[7] One guideline is to reduce the bonding temperature while maintaining a reasonable level of bonding force in order to obtain optimum bonding quality. The next section investigates and identifies suitable Cu-Cu bonding temperature for 3D process integration.

Impact on Cu-Cu Contact Resistance and Its Stability

200 mm Si-(100) wafers with two layers of single damascene Cu are used for wafer-on-wafer face-to-face stacking. Self-assembled monolayer (SAM)[9] is coated on a Cu bonding pad after a top dielectric recess that reduces the Cu oxidation effect. SAM is desorbed before Cu-Cu thermo-compression bonding. A cross-bar Kelvin structure is used to characterize

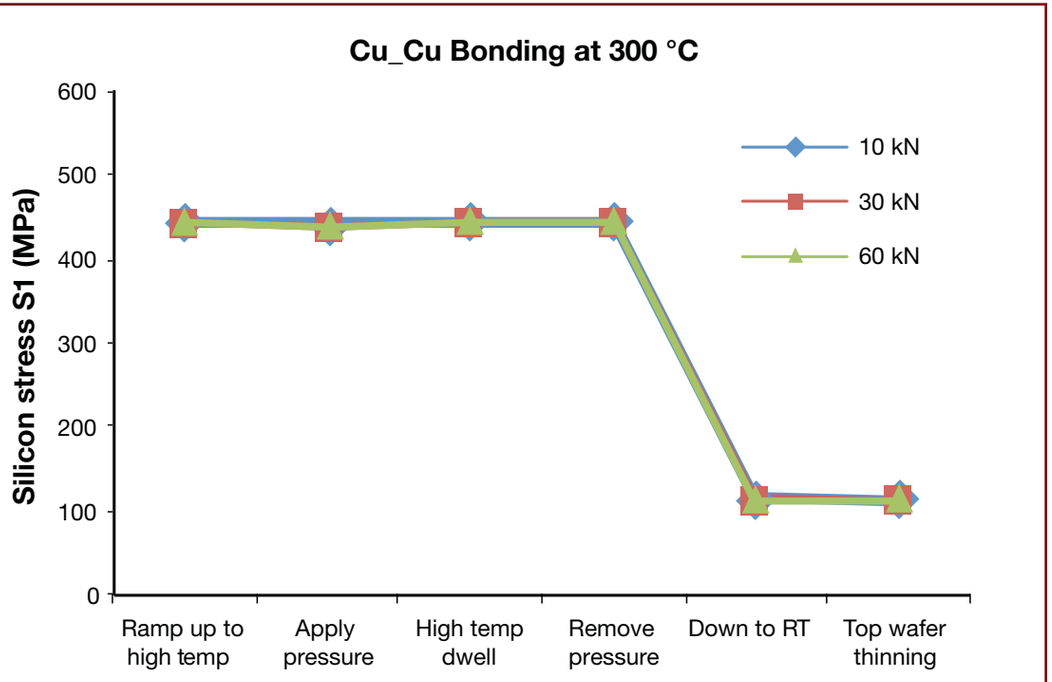


Figure 3. Simulation Results of Cu-Cu Bonding at 300 °C (1 hour) With Different Bonding Forces

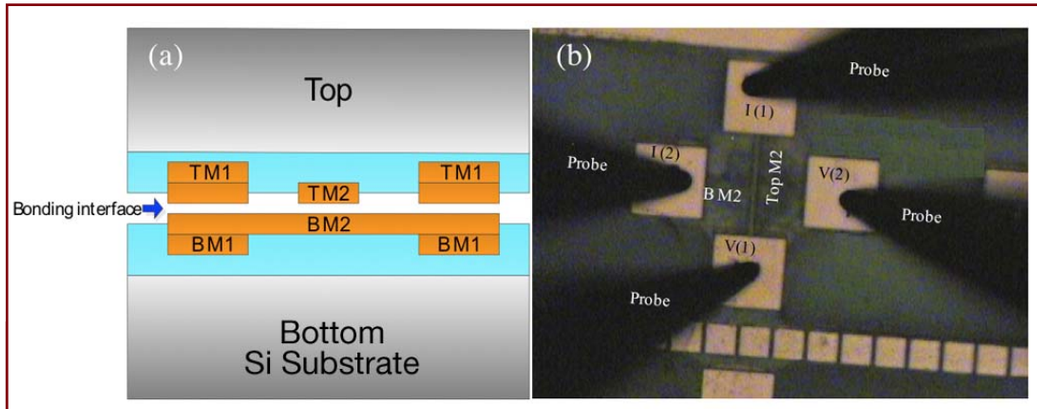


Figure 4. (a) Cross-bar Kelvin Structures Schematic; and (b) After Top Si Removal

the effect of bonding temperature on the contact resistance as shown in Figure 4. This structure can overcome wafer-to-wafer misalignment during contact resistance study. The top and bottom metal line width is 5 µm.

The bonding process is performed at 225, 275 and 300 °C, respectively, with 60 kN bonding force for one hour. The Cu-Cu thermo-compression bonding contact resistance is characterized by a four-point measurement. The contact resistance is ~5, 3.5 and 2.3 mΩ for bonding temperature at 225, 275 and 300 °C, respectively. As clearly shown in Figure 5, lower contact resistance is achieved when higher bonding temperature is applied.

Cu-Cu contact resistance reliability is investigated by thermal cycle test (TCT). The sample is subjected to thermal cycling test with the temperature ranging from -40 °C to 125 °C. The ramp-up/down rate is ~15 °C/min and each thermal cycle is ~52 min. The Cu-Cu contact resistance is measured after 200 and 500 thermal cycles. The Cu-Cu contact resistance is consistent after 200 and 500 cycles of TCT when bonding is done at 300 °C. However, fluctuation in the contact resistance is observed for Cu-Cu contact bonded at 225 °C and 275 °C after TCT. Therefore, 300 °C is selected as the Cu-Cu thermo-compression bonding temperature in our bonding study. This bonding temperature can provide stable contact resistance.

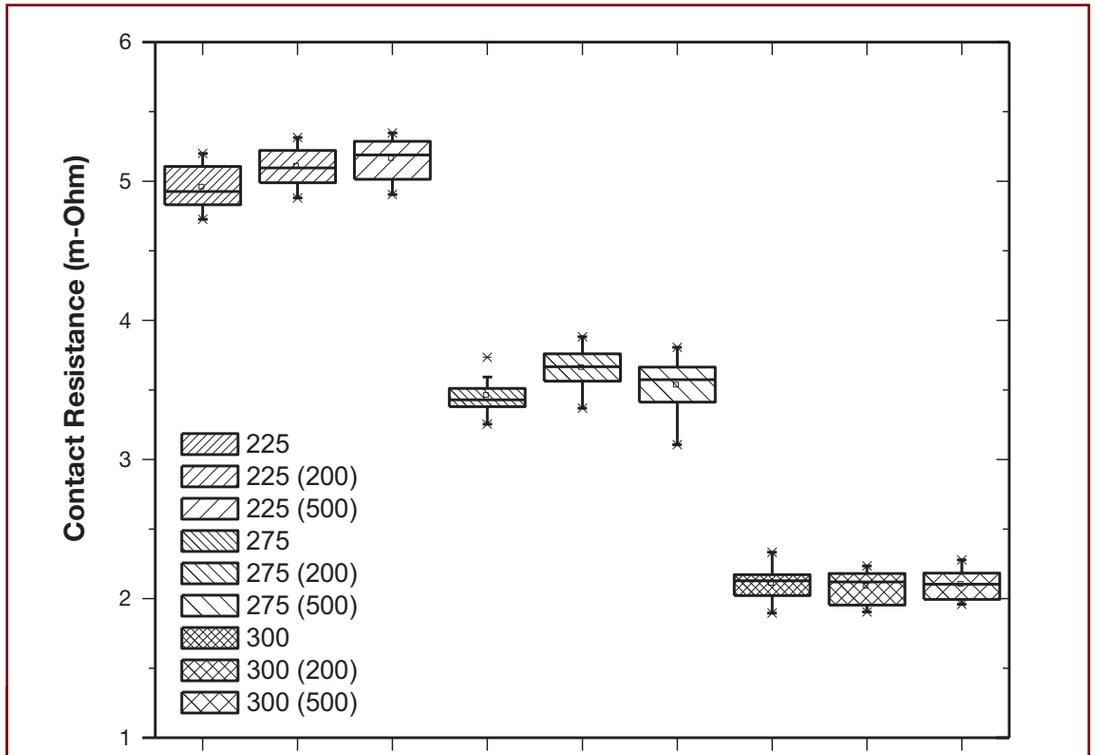


Figure 5. Cu-Cu Contact Resistance Due to Bonding Temperature and Its Reliability

Summary

In this study, it is noted that when the bonding temperature is <300 °C, higher resistance and unstable contact resistance are observed from the thermal cycling test. Meanwhile, increased bonding pressure can improve bonding yield and quality.[8] Therefore, Cu-Cu bonding temperature at 300 °C is chosen along with 60 kN of bonding force based on this study. In addition, surface pre-treatment with SAM is applied to achieve fine-pitch Cu-Cu interconnects bonding for wafer-on-wafer stacking.

References

1. R. Taibi et al. "Investigation of stress induced voiding and electromigration phenomena on direct copper bonding interconnects for 3D integration," in IEDM Tech. Dig., 2011, pp. 135-138.
2. Y.H. Hu et al. "Cu-Cu Hybrid Bonding as option for 3D IC stacking," IITC 2012, pp. 1-3.
3. L. Di Cioccio et al. "200°C direct bonding copper interconnects: electrical results and reliability," IEEE 3D Systems Integration Conference, 2011, pp. 1-4 (9-8)
4. Y.H. Hu et al. "3D Stacking Using Cu-Cu Direct Bonding," IEEE 3D IC 2011, pp 1-4 (1-3).
5. Cedric Huyghebaert et al. "Cu to Cu interconnect using 3D-TSV and Wafer to Wafer thermo-compression bonding," IITC 2010, pp. 1-3.
6. W.H. The et al. "Recent Advances in Submicron Alignment 300mm Copper-Copper thermocompressive Face-to-Face Wafer-to-Wafer Bonding and Integrated Infrared, High-Speed FIB Metrology," IITC 2010, pp. 1-3.
7. Chukwudi Okoro et al. "Analysis of the Induced Stresses in Silicon During Thermocompression Cu-Cu bonding of Cu-Through-Vias in 3D-SIC Architecture," IEEE ECTC 2007 pp. 249-255.
8. H.L. Leong et al. "Application of contact theory to metal-metal bonding of Silicon wafer," J. Applied Physics 2007 102 p. 103510.
9. D.F. Lim et al. "Achieving low temperature Cu to Cu diffusion bonding with self-assembled monolayer (SAM) passivation," IEEE 3DIC, 2009.

About the Authors

H.Y. Li – Scientist, Institute of Microelectronics

L. Peng – Ph.D. student, Nanyang Technological University and Institute of Microelectronics

F.X. Che – Scientist, Institute of Microelectronics

C.S. Tan – Assistant Professor, Nanyang Technological University

G.Q. Lo – Deputy Executive Director, Institute of Microelectronics ■

Click here to return to Table of Contents

PRINT this article E-MAIL this article

Click here to return to Table of Contents

Surya Bhattacharya
Director, Industry Development; IME

With the current 2.5D and 3D technologies, multi-functional die systems are confined in a single package with few input/output (IO) going to the external printed circuit board (PCB) motherboard. In particular, the die-to-die (e.g., logic to memory read and write operations) interfaces can't be probed with normal "intrusive" test probes.

It is then necessary to review the testing methodology to qualify 2.5D/3D systems during pilot test, and verify the proper functionality of the different blocks within the system prior to and during operations on a larger PCB motherboard.

A set of existing chip-embedded testing strategies can be used for such "closed" systems such as boundary scan testing and built-in self test (BIST). Yet more advanced "chip embedded instrumentation" intellectual properties (IPs), mimicking laboratory equipment test boxes (e.g., logic scope or bit error rate tester) inside the dies under test (DUT) need to be developed as well. Those "chip-embedded instrumentation" IPs could be implemented as hard IP macro or soft IP macro and accessed through a JTAG interface from the PCB motherboard.

Currently, IME researchers (with their industrial partners) are developing a field programmable gate array (FPGA) with 3D-DRAM 2.5D high-performance system that will use "chip-embedded instrumentation" soft IPs to test the FPGA to 3D-DRAM wide IO interface.

This FPGA-assisted test technique for wide IO logic to 3D-DRAM systems will enable the proper qualification process for future high-performance applications and could be migrated for an ASIC version of a wide IO-based system in a 2.5D or 3D configuration at very high speed rates.

Embedded System Access: Changing the Paradigm of Electrical Test

Thomas Wenzel, Heiko Ehrenberg
GÖPEL electronic GmbH

Looking for the Panacea of Test

The test of electronic circuits has been a key topic in the industry since the first transistor was developed, and today it is as relevant as ever. Test strategies are graded by how close they come to the ideal test solution that doesn't add any cost to the product under test, either during the design or during production. Most of us agree that product testing is absolutely necessary, as part of design validation, as a quality indicator for manufacturing process control or for the detection of defective products prior to shipping them to a customer. However, we do have certain requirements that should be met by our test solutions: test development and execution should be fully automated and should be done in essentially no time; we want the test equipment to be very inexpensive; and we want fault coverage of 100 percent. Industry trends give cause for concern, though, considering that the cost of test today can be a significant part of the overall development and manufacturing cost.

Responsible for this development are primarily the complexity, high-speed designs and the lack of available test access of many of today's printed circuit board assemblies (PCBAs), or boards, for short. The combined forces of these characteristics result in systematic changes in the balance of product design and product test. We start to see a correlation between problems seen in chip test and those seen in board test.

While boards look more like integrated circuits (IC) due to the loss of access to internal circuit nodes, the rapid development of three-dimensional (3D) ICs with multi-die integration results in structures that are similar to boards and systems. The 3D board with very little physical access seems to be looming on the horizon. At the same time, the combination of new packaging and integration technologies result in hitherto-unaccustomed complexity. While several years ago, multiple boards were necessary to create complete system designs, today some such systems can be realized in IC as system-on chip (SOC) or system-in package (SIP) designs. As a result, board size can be minimized and new possibilities are available to create super-complex systems. No matter how a design is arranged, however, from

the perspective of test engineering, the fundamental questions are: 1) How such highly complex systems can be tested appropriately and efficiently; and 2) How one can take advantage of synergies between chip test and board test approaches?

Non-invasive Test Access?

Partitioning circuit structures into testable elements is a prerequisite for a successful test strategy. This is one of the reasons why in-circuit test (ICT) became so successful for board-level tests. ICT approached circuit test structurally and tests components individually; however, the required bed-of-nail-based invasive test access is becoming a big dilemma with modern boards. Test access problems were predictable, which resulted in the creation of IEEE Std 1149.1 in 1990.

The brilliance of IEEE Std 1149.1 is the open expandability of its register architecture combined with the universal test bus interface (test access port, TAP) and its protocol definition. These properties allowed IEEE Std 1149.1 to become the base technology for new non-intrusive methodologies and standards for testing, debugging, programming and emulation. As a result, the portfolio of test access strategies at the board level has definitively changed.

Today we can differentiate three principle classes of access strategies (Figure 1):

- Native connector access (access through design-integrated I/O interfaces)
- Intrusive board access (access through physical test nails and probes)
- Embedded system access (access through design-integrated test bus)

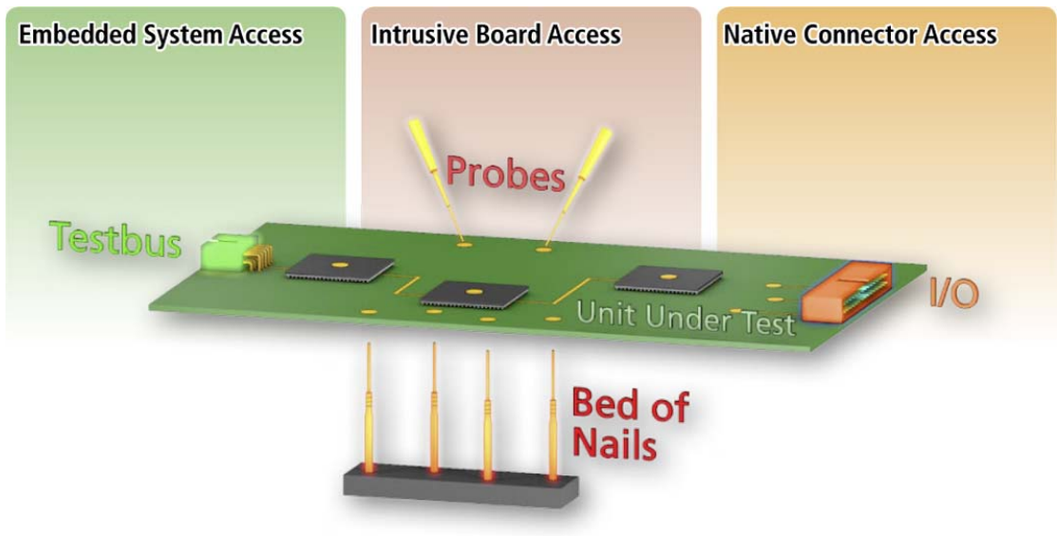


Figure 1. Classification of Electrical Test Access Strategies at Board Level

While these classes are not mutually exclusive in their practical utilization, the applicability of an actual combination of these access strategies depends on the individual capabilities of the chosen automated test equipment (ATE) platform.

So, how do these access strategies relate to each other, and what does embedded system access mean practically?

Paradigm Change: A New Era?

A look at the qualitative development of trends for the various access strategies reveals interesting facts, including a long adoption period of IEEE Std 1149.1 as the first representative for embedded system access. The accelerated adoption of embedded system access in the market is primarily due to the fact that it is now a class by itself, comprising a variety of non-invasive access technologies, including:

- Boundary-scan test (IEEE Std 1149.1/.4/.6/.7)
- Processor-emulation test
- Chip-embedded instrumentation (IJTAG, IEEE P1687)
- In-system programming
- Core-assisted programming
- FPGA-assisted test
- FPGA-assisted programming
- System JTAG (SJTAG)

The electrical access embedded in the target system allows embedded system access to work without invasive test nails and probes. In principle, every ESA technology utilizes a task-specific pin-electronic that is controlled by the test bus and, as a result, can directly execute test functions and programming routines in the target system. This target system can be an individual chip, a board or a complete system assembly; embedded system access can be utilized throughout the entire product life cycle.

Property	Boundary Scan Test	Processor Emulation Test	Chip-Embedded Instruments	FPGA-Assisted Test
Test type	structural	functional	open*	open*
Test speed	static	dynamic	open*	open*
Access through	Boundary-Scan IC	processor	IJTAG-IC	FPGA
Pin-electronics	Boundary-Scan-Register	system bus	IP-Interface	IP-Interface
Configurable IP**	no	no	open*	yes
Fault coverage	static	dynamic	open*	open*
Level of diagnostics	pin	net/pin	open*	open*
Related IEEE standard	IEEE Std 1149.x	IEEE Std 1149.7/ ISTO 5001	IEEE Std 1149.1, IEEE Std 1149.7, IEEE P1687	IEEE Std 1149.1, IEEE Std 1149.7, IEEE P1687

* Depending on the implementation ** Intellectual property

Table 1. Comparing ESA Technologies Relevant to Board-Level Test

Embedded System Access: A Portfolio of Complementary Technologies

A detailed analysis of key ESA technologies at the board level reveals considerable differences in operation and goals.

Table 1 reflects the complementary character of the various technologies and, as the following discussion will further explain, it becomes clear how important it is for ATE platforms to support all these ESA technologies alike.

Boundary Scan utilizes so-called boundary-scan cells, combined into a boundary-scan register, as primary access points for a target system's circuit nodes. The boundary-scan register is accessed and controlled through the test access port (TAP). All vectors are scanned serially.

However, since boundary-scan tests are static in nature, dynamic defects usually cannot be detected, let alone be diagnosed. In addition to IEEE Std 1149.1, various related standards have been created or are in development.

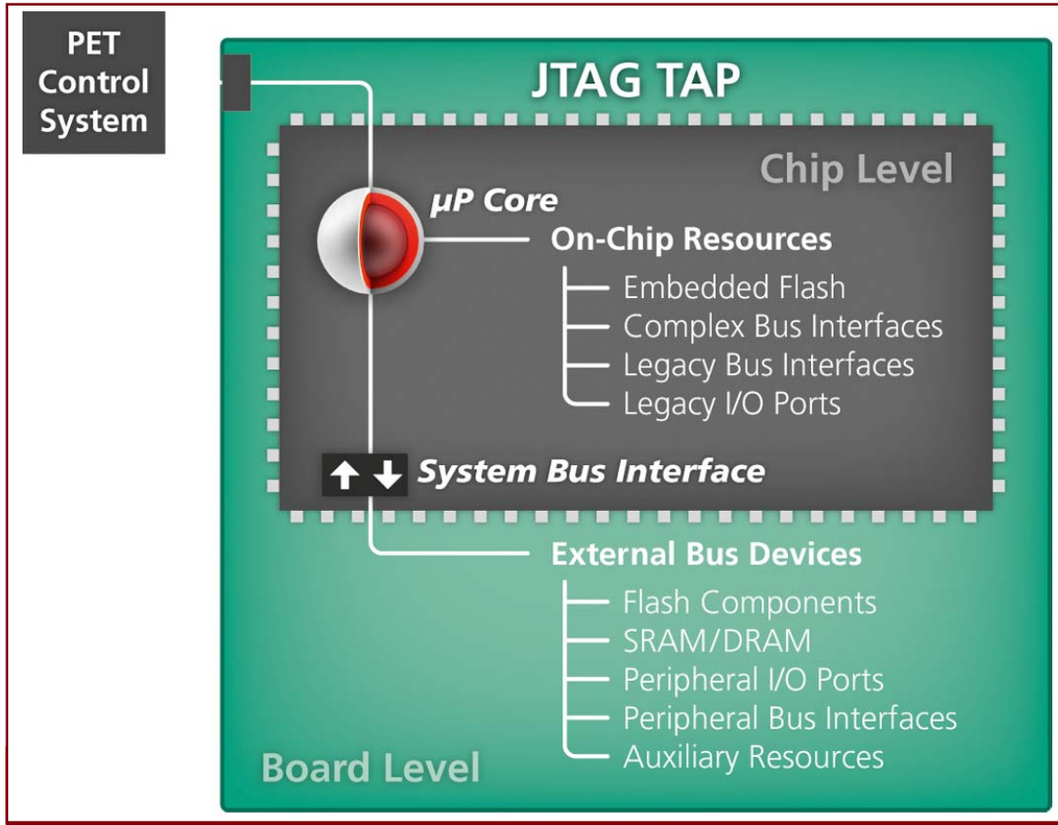


Figure 2. Principle of Processor Emulation Test (PET)

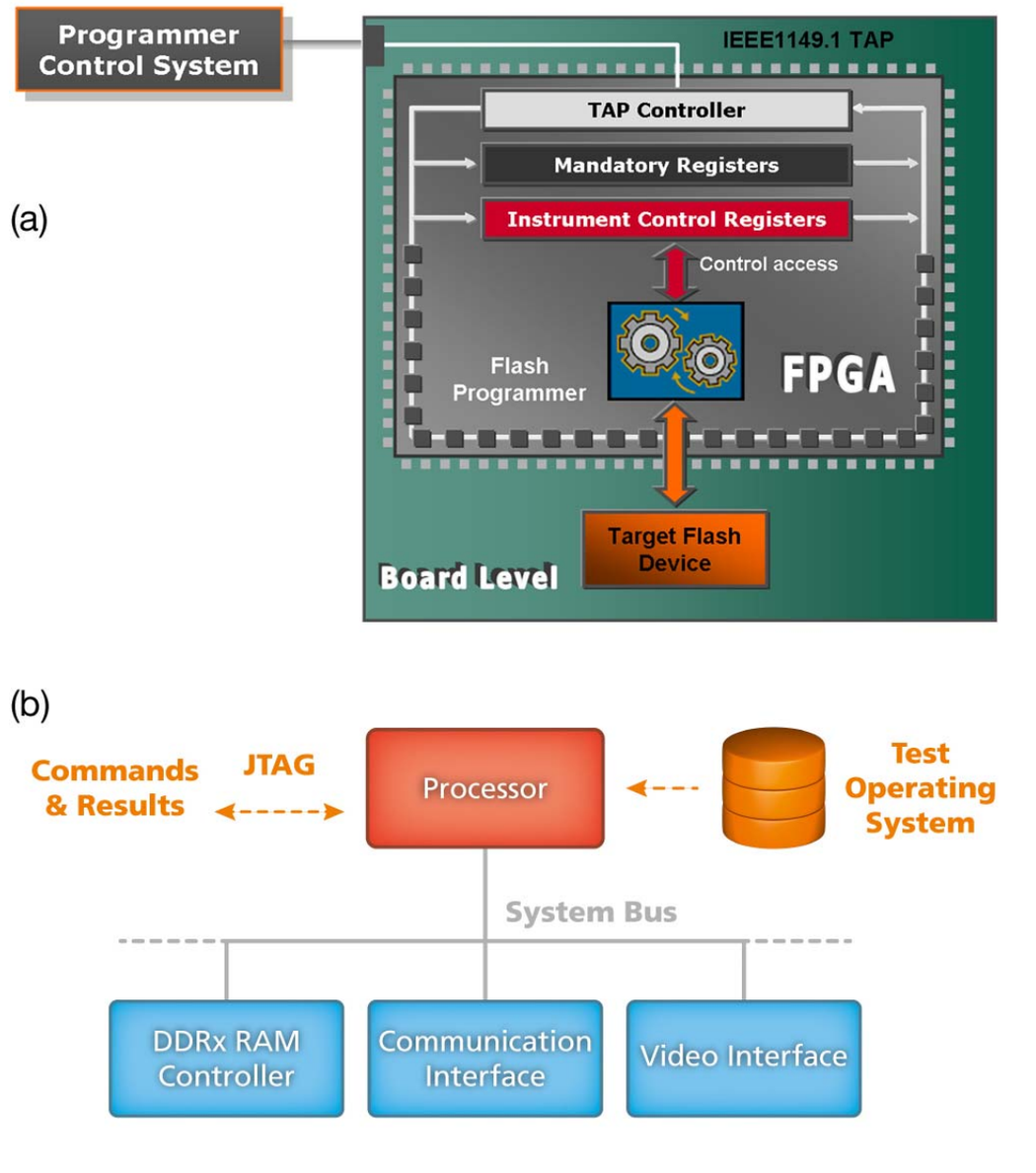


Figure 3. a) Programming flash memory per FPGA-assisted programming; b) Embedded test bus controller at board level

Processor Emulation Test (PET) utilizes the debug interface to transform the processor core temporarily into a native test controller (Figure 2). The processor and its system bus interface become the pin-electronics used as access points for the connected circuitry in the target systems. Remote-controlled through the JTAG interface or some other debug interface, the processor core utilizes write and read access to the system bus with respective test vectors in order to manipulate and test the connected internal and external resources and components.

Chip-embedded Instruments are test and measurement intellectual property (IP) blocks integrated into ICs, often accessible through the JTAG port. The functionality of chip-embedded instruments is completely open and ranges from simple sensors over complex signal processing and data collection all the way to complete analysis instruments and programming engines. The IP is either integrated permanently in the chip (hard-macro), or it can be temporarily instantiated and configured (soft-macro) in field-programmable gate arrays (FPGA). As a result, the pin-electronics are unrestricted in principle and can provide a wide variety of functionality, within the frame and scope of the respective technology of the host device, of course.

In particular, FPGA-embedded instruments have enjoyed strong interest recently.

Figure 4. Control of ESA Applications Through External Hardware and Software

By enabling strategies such as FPGA assisted test (FAT) and **FPGA-Assisted Programming (FAP)**, they provide enormous flexibility for the adaptation to individual test and measurement requirements.

Chip-embedded instruments have been utilized for years in chip test; for example, in the form of built-in self-test (BIST) IP. However, access to these instruments has not been standardized in the past; something that will be changed with the new IEEE P1687 (also known as IJTAG).

In-System Programming (ISP) is a collective term for the programming of Flash devices via boundary scan and for the programming of PLD/FPGA devices through their TAP and built-in programming registers, while the devices are mounted on the printed circuit board. For in-system programming of PLD/FGPA, special standards exist, such as IEEE Std 1532, JESD-71 and an industrial standard called serial vector format (SVF).

The premise of the **Core-Assisted Programming (CAP)** strategy is similar to processor emulation test. The processor is controlled through its native debug interface in a way that allows Flash or FPGA (design permitting) connected to the system bus to be erased, programmed and verified. In the case of Flash, it does not matter whether it is integrated in the processor/micro controller unit (on-chip Flash) or connected as external, discrete Flash device(s). Furthermore, it is possible to load only the Flash handler/programming engine via JTAG into the processor and to download the Flash data image through a high-speed communication interface on the processor CAP technology.

One of the most interesting technologies for Flash ISP, referred to as **FPGA-Assisted Programming (FAP)**, is based on FPGA-embedded instruments. The embedded instrument in this case is a programming engine (programmer) soft macro, typically provided by a tool vendor and temporarily downloaded into the FPGA. Depending on the architecture of the programmer IP and the performance of the external control system, drastic improvements in programming speed compared to boundary-scan based ISP are possible.

The last access technology in this discussion is referred to as **System-Level JTAG**. While remote control through an external controller is possible, this technique typically employs a central test control unit integrated directly into the system design. Test vectors are usually stored locally on the system and a separate IC is commonly used as the test bus controller (although there is also the possibility of integrating the test bus controller function in an IC that also performs other functions in the system design). As the name implies, this method can be employed not only for individual boards but also for systems comprising multiple boards and modules.

The Transformation to the System-integrated Tester

The transition from traditional invasive test access and techniques to embedded system access is not a marginal change in the handling of test and programming vectors, but rather a fundamental technological metamorphosis. Characteristics of these changes include:

- Integration of test electronics in the system under test
- Un-separable coupling of functional and test circuitry in the system design
- Forming of partitioned test centers with various features
- Significantly wider range of test and programming strategies

- Possible utilization throughout the entire product life cycle
- Flexibility of reconfigurable pin-electronics with FPGAs
- Availability of completely new instrumentation platforms

In practice, embedded system access represents in principle a transformation from a purely functional design into a functional design with integrated test capabilities, a combination of unit under test and tester, so to speak (Figure 3).

Depending on the actual implementation of embedded system access, a wide variety of applications is possible (Figure 4). Currently, FPGA-based test in particular is a technology driver for progressively more complex test and measurement functions. This includes applications such as:

- Voltage measurements
- Frequency measurements
- Temperature measurements
- Bit error rate tests (BERT) for high-speed signals
- Event counters
- Logic scopes, etc.

Multidimensional Requirements for Tester Instrumentation

So far, we have primarily discussed the JTAG interface as the test bus. However, there are also a number of proprietary bus interfaces used in the industry; in particular, for debug interfaces on processors, such as serial wire debug (SWD), spy-bi-wire (SBW) or background debug mode (BDM). For ATE vendors, this means their test bus controllers need to provide the required flexibility to support any of such interfaces; even a mix of different test bus interfaces in multi-processor applications should be supported. Furthermore, the various ESA technologies must be supported by powerful software tools and must be made available to the user in intuitive graphical user interfaces. In this context, we need to consider not only the independent use of individual ESA methods, but also the potentially interactive application of various ESA technologies in order to gain extra benefits.

About the Authors

Thomas Wenzel
Thomas Wenzel is co-founder and managing director of GÖPEL electronic GmbH in Jena, Germany. He heads the company's JTAG/Boundary Scan division. As graduate engineer, Thomas worked as a hardware designer with Carl-Zeiss Jena before founding GÖPEL electronic in 1991.

Heiko Ehrenberg
Heiko Ehrenberg is president of GÖPEL electronics LLC in Austin, Texas. He opened GÖPEL electronic's subsidiary in the U.S. Today Heiko works in several IEEE groups, dealing with future test and programming standards. ■

Click here to return to Table of Contents

ADVERTISERS INDEX