SEMICONDUCTORS | AI | DESIGN | PACKAGING | MEMS | DISPLAYS

SEMICONDUCTOR DIGEST

NEWS AND INDUSTRY TRENDS

JULY 2020

Fab Models Built on AI

Collaborative Knowledge Sharing PAGE 20

Delivering Real-world Solutions

Today's lithography systems provide overlay accuracy and throughput capabilities far beyond what was previously thought possible. However, it is real-world, on-product performance that is vital to device makers.

Nikon combines a strong history of engineering expertise with superior scanner technology and industry-leading solutions to deliver exceptional manufacturing performance and productivity—now and for the future.

Nikon. Delivering Real-world Solutions.

See Nikon at:

Virtual SEMICON West 2020

and

Nikon is hiring:

Explore a career at:
nikonprecision.com/careers

www.nikonprecision.com

SEMICONDUCTOR DIGEST

NEWS AND INDUSTRY TRENDS

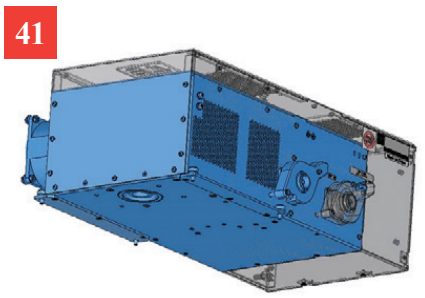
July 2020 | Volume 2 Number 5

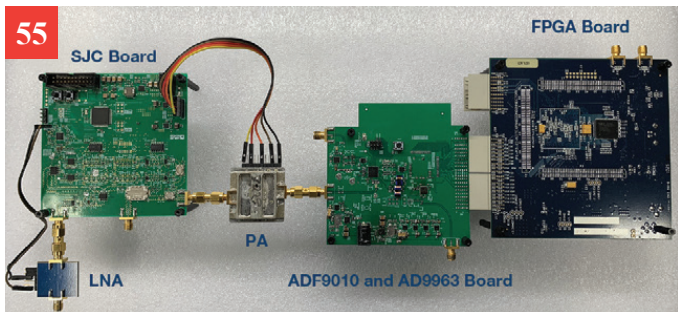
COVER: Photo courtesy of Eddie Shvartzman, esphotographer@gmail.com; www.esphotographer.com; Phone: 719-201-7006

Columns

Departments

FEATURES





MANUFACTURING A.I. BISTel

BISTelligence 2020

for NORTH AMERICA

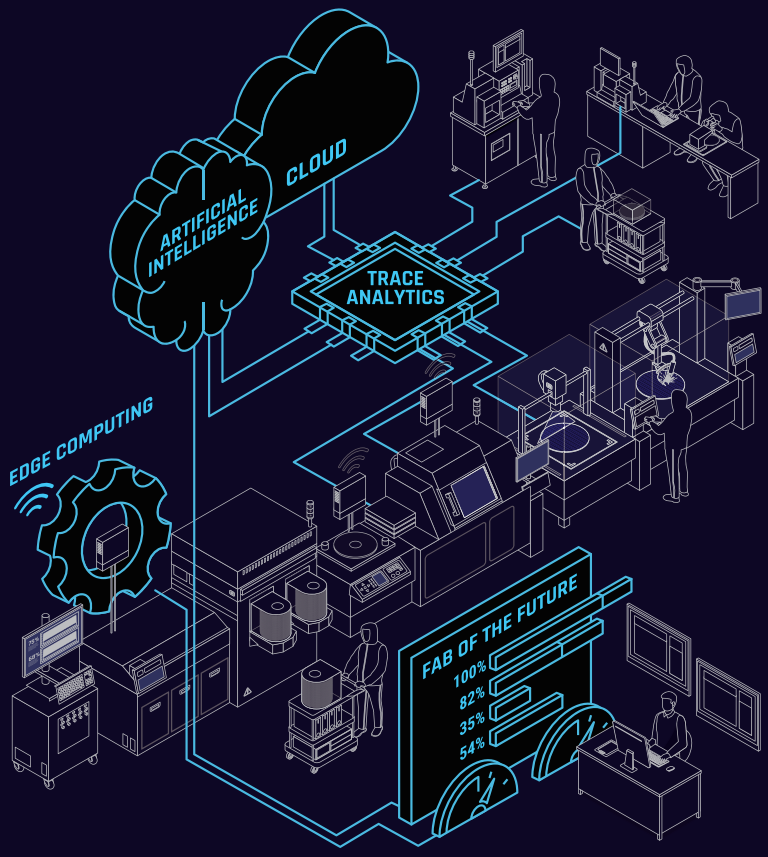
Come celebrate our 20th Anniversary and see
the latest A.I., cloud and edge computing technologies
powering the latest A.I. smart manufacturing applications.

Experience industry first technologies from the world leaders
in continuous improvement for manufacturing.

Register NOW
TO GET EXCLUSIVE CONTENT

AGENDA

- Topic 1 • TRACE ANALYTICS
- Topic 2 • ARTIFICIAL INTELLIGENCE
- Topic 3 • CLOUD
- Topic 4 • EDGE COMPUTING
- Topic 5 • FAB OF THE FUTURE



BISTelligence 2020
for NORTH AMERICA

Presents the fab of the future
July 20 - 31, 2020 | Online

BISTelligence is a virtual platform where we want to connect with you. It is a new digital experience that takes you to another level of smart manufacturing. It is where innovation meets the technology, and where digital transformation begins.

Register now at www.bistel.com/eventregistration

Editorial

Adapting to Change

"IT WAS THE BEST OF TIMES, IT WAS THE WORST OF TIMES..." SO BEGINS A Tale of Two Cities by Charles Dickens. He was talking about London and Paris in 1859 before the French Revolution, but the phrase seems apt today. It should be a time of excessive jubilation for the semiconductor industry – so many long-awaited advances are finally happening, including 5G, automated driving, artificial intelligence/machine learning, quantum computing, the Internet of Things (IoT), cloud computing, and smart just about everything.

Yet we find ourselves in the middle of a worsening pandemic. We miss our friends and family. Our hearts ache for those affected. We seemingly risk our health just going to the grocery store or to get our hair cut. The future is uncertain.

But we adapt. That's what people do. We take some time for introspection. We gain new respect for front-line healthcare workers and so many others. We take precautions. We find new routines. We find new ways to communicate. We carry on.

Manufacturing industries are adapting to this new way of life too. Fortunately for the semiconductor industry – which is in many ways the role model for quickly innovating and adapting to overcome new obstacles – the impact has been minimal.

In fact, although chip demand in some areas is down, it's up in other areas. Way up. As noted in this issue's article "How COVID-19 is Impacting the Memory Industry" starting on pg. 29, Netflix saw a surge in demand, as did online gaming platforms during the lockdown. Online shopping and video conferencing services such as Zoom and WebEx also saw huge increases. These all drive datacenter demand and the need for faster, more powerful semiconductors.

As essential businesses, semiconductor manufacturers and their many suppliers have continued operations. The industry has even moved to quickly adopt remote fab operations, e-diagnostics, cloud computing and other tools defined by the Industry 4.0 or "Smart Manufacturing" movement. This move is long overdue, having been delayed by mostly unfounded security concerns. In this month's Industry Observation guest editorial column (pg. 68), Dave Anderson, president of SEMI Americas, notes that "suppliers of software that is used to diagnose and manage semiconductor manufacturing tools remotely saw its usage more than double between February and April. It has remained at record-high usage levels during May and into June. A lot of folks have continued to handle vital diagnostics and control, but from home."

Additional good news is that manufacturers are also working hard to now adopt AI tools in their manufacturing operations. In this issue's cover story "Fab Models Built on AI Collaborative Knowledge Sharing," on pg. 20 the authors from BISTel note that, today, we have the ability to gather "more information than ever helping the system to digest and turn data into knowledge and insights. These new technologies are enabling the next generation of engineering empowerment and the first AI knowledge sharing fab models."

Our friends at SEMI have also adapted. As with many events this year, SEMICON West 2020 will be virtual, taking place July 20-23. Stop by the *Semiconductor Digest* booth for a chat and pick up a copy of our SEMICON West Show Daily.. virtually, of course!

—Pete Singer, Editor-in-Chief

SEMICONDUCTOR DIGEST
NEWS AND INDUSTRY TRENDS

Kerry Hoffman, Publisher
kerryh@semiconductordigest.com
978-580-4205

Pete Singer, Editor-in-Chief
psinger@semiconductordigest.com
978-470-1806

Shannon Davis, Web and News Editor
sdavis@semiconductordigest.com

Ed Korczynski, Senior Technical Editor
edk@semiconductordigest.com

Dave Lammers, Contributing Editor
dplammers@att.net

John Blyler, Contributing Editor
jblyler@semiconductordigest.com

Dick James, Contributing Editor
dickjames@siliconics.ca

Cindy Chamberlin, Art Director

Rich Mehta, Website Design

July 2020 VOL. 2 NO. 5•
Semiconductor Digest ©2020
ISSN 2643-7058 (print) ISSN 2643-7074 (online).
Semiconductor Digest is published eight times a year by Gold Flag Media LLC, 58 Summer St., Andover, MA 01810.

For subscription information, please visit www.semiconductordigest.com/subscribe.

Chroma

Driving Innovation to Success

Safeguarding Your Chipsets

Chroma's innovative test & automation turnkey solutions ensure the performance, quality, and success of your devices

VLSI Test Systems | **SoC/Analog Test Systems** | **Temperature Forcing Systems** | **Tri-temp ATC Test Handlers** | **Full Range ATC FT Handlers**

U.S.A.
Chroma ATE Inc. (U.S.A.)
7 Chrysler, Irvine,
CA 92618
T +1-949-421-0355
www.chromaus.com
info@chromaus.com

Taiwan
Chroma ATE Inc.
T +886-3-327-9999
www.chromaate.com
info@chromaate.com

The Netherlands
Chroma ATE Europe B.V.
T +31-318-648282
www.chromaeu.com
salesnl@chromaeu.com

Germany
Chroma Germany GmbH
T +49-821-790967-0
www.chromaeu.com
salesde@chromaeu.com

USA | Germany | Netherlands | Taiwan | China | Japan | Korea | SE Asia

Watch webcast

worldnews

U.S. — GLOBALFOUNDRIES secured a purchase option agreement for approximately 66 acres of undeveloped land adjacent to its most advanced manufacturing facility, Fab 8, in Malta, N.Y., near the Luther Forest Technology Campus (LFTC).

U.S. — Kioxia Holdings Corporation, a leader in memory solutions, announced the appointment of Michael R. Splinter as an independent director.

U.S. — Heidenhain Corp. announced the opening of its newly completed West Coast headquarters. This includes the expansion of its executive, sales and technical support offices, as well as demo facilities in San Jose, CA.

U.S. – National Instruments closed the acquisition of OptimalPlus, a supplier of data analytics software for the semiconductor, automotive and electronics industries.

U.S. and Asia — CyberOptics was awarded a 2020 EM Asia Innovation Award in the category of Test Equipment-Testing Software Suite for its CyberCMM. The award was presented to the company during a ceremony that took place July 3, 2020 at the National Exhibition and Convention Center in Shanghai.

U.S. — SK hynix Inc. started the full-scale mass-production of high-speed DRAM, 'HBM2E', only ten months after announcing the development of the new product.

news

Top-10 Semiconductor Suppliers Defy Weak Market Conditions in Q1

Defying a decline in market revenue, the world's top-10 semiconductor suppliers managed to generate revenue growth of 2.1 percent in the first quarter, as the companies benefitted from a COVID-19-driven increase in PCs and server sales.

The top-10 chipmakers collectively generated revenue of $63.6 billion in the first quarter, up from $62.2 billion in the fourth quarter of 2019, according to the Omdia Competitive Landscaping Tool (CLT) service. This contrasts with a 2 percent sequential decline in the overall global chip market in the first quarter, with revenue falling to $110.1 billion, down from $112.3 billion in the fourth quarter of 2019.

While Omdia typically employs a year-over-year comparison for market-share growth figures, the unprecedented downturn for the semiconductor market in 2019 and the pandemic in 2020 have made annual comparisons less meaningful, prompting a focus on the sequential aspect on the semiconductor market.

"Many of the leading semiconductor suppliers benefitted from strong demand for client PCs, enterprise PCs and servers in the first quarter," said Ron Ellwanger, senior analyst, semiconductor manufacturing, at Omdia. "COVID-19-driven stay-at-home orders spurred rising demand for PCs as locked-down consumers tried to remain connected with friends, family, clients and business associates. In turn, this phenomenon has stimulated demand for corporate PCs and servers, with organizations striving to accommodate increasing consumer demand for cloud services. All this demand translated into increased demand for the kinds of chips offered by the world's top semiconductor suppliers."

As a result of the increasing demand for computer platforms, the data-processing category was the only application market for semiconductors to attain growth during the first quarter. Data processing posted a 0.9 percent increase in revenue compared to the fourth quarter of 2019.

Top-10 semiconductor supplier growth fueled by memory

Memory-oriented semiconductor suppliers drove much of the growth among the top-10 suppliers, with the three leading companies in this area— Samsung, SK Hynix, and Micron—collectively expanding their revenue by 1.1 percent during the first quarter.

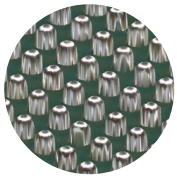
Big Performance at a Small Scale.

Unparalleled combination of high accuracy, high resolution and speed.

Copper Pillar

Flip Chip (c4)

Micro Bump



BGA Ball

BEST OF WEST
2019 WINNER
Presented by SEMICONDUCTOR DIGEST and SEMI®

2-3X FASTER

NanoResolution MRS™ Sensor

For Wafer-Level and Advanced Packaging Inspection and Metrology

The revolutionary NanoResolution Multi-Reflection Suppression™ (MRS) sensor technology inspects shiny or mirror-like surfaces with sub-micrometer accuracy for features as small as 25µm.

The NanoResolution MRS sensor enables metrology grade accuracy with superior 100% 3D and 2D inspection performance for attaining repeatable and reproducible measurements faster than alternative technologies – with throughput greater than 25 wafers (300mm) per hour, as compared to current sampling methods.

Visit us at
SEMICON WEST
July 20-23, 2020 | Virtual Event

CYBEROPTICS®
www.cyberoptics.com

Top-10 semiconductor supplier ranking in Q1, 2020 (ranking by revenue in millions of US Dollars)

Q4-19 Rank	Q1-20 Rank	Company Name	Q4-19 Revenue($)	Q1-20 Revenue($)	Revenue Percent Change	Revenue Percent of Total	Revenue Cumulative Percent
1	1	Intel	19,864	19,508	-1.8%	17.7%	17.7%
2	2	Samsung Electr	13,559	13,751	1.4%	12.5%	30.2%
3	3	SK Hynix	5,828	5,869	0.7%	5.3%	35.5%
4	4	Micron Technol	4,862	4,895	0.7%	4.4%	40.0%
5	5	Broadcom Limit	4,263	4,172	-2.1%	3.8%	43.8%
6	6	Qualcomm	3,534	4,050	14.6%	3.7%	47.5%
7	7	Texas Instrume	3,290	3,272	-0.5%	3.0%	50.4%
16	8	HiSilicon Techn	2,012	2,822	40.3%	2.6%	53.0%
9	9	nVidia	2,698	2,673	-0.9%	2.4%	55.4%
11	10	KIOXIA Corpora	2,331	2,564	10.0%	2.3%	57.8%
		Top 10 Companies	62,241	63,576	2.1%	57.8%	
		All Others	50,106	46,489	-7.2%	42.2%	
		Total Semiconductor	**112,347**	**110,065**	-2.0%	100.0%	

Top 10 semiconductor supplier ranking in Q1 2020 ranking by revenue in millions of US Dollars.

These companies benefitted from the growth in the data-processing market, which is generating increasing demand for NAND flash used in enterprise solid-state drives (SSDs). Sales of NAND flash memory grew 6.9 percent sequentially in the first quarter, representing the highest growth rate of any device in the semiconductor market during the time period.

Qualcomm and HiSilicon lead in growth among top-10

While the memory area was the leading growth driver for the top-10, the best individual performances were posted by U.S. wireless semiconductor supplier Qualcomm and Chinese fabless system-on-chip (SoC) firm HiSilicon. Sixth-ranked Qualcomm attained robust 14.6 percent sequential growth in the first quarter, while HiSilicon surged by a staggering 40.3 percent—the highest rate of expansion of any top-10 supplier.

"Qualcomm is benefitting from the Chinese government's move to kickstart the economy by emphasizing the deployment of 5G infrastructure," Ellwanger said. "The company is taking advantage of China's subsidized 5G handset market, along with moves to accelerate the building of 5G infrastructure in the country."

HiSilicon is the chip division of China's Huawei. Despite the restrictions placed on Huawei by the US government, the company has managed to protect itself from the effects of the US/China trade war. The company is doing this by building up sufficient inventory to ride out the impact of the revised US trade restrictions, which are planned to go into effect in September.

Top-10 winners and losers

Although demand increased for PC-oriented chips in the first quarter, microprocessor chip leader Intel suffered a 1.8 percent drop in revenue during the period. The surge in computer demand in the first quarter was focused on lower-end systems, increasing demand for lower cost Intel microprocessors and chip sets, thus trimming its revenue.

KIOXIA, formerly known as Toshiba Memory, rose one rank to take the no.-10 spot, supplanting Sony Semiconductor. The company's semiconductor revenue increased by an impressive 10 percent. sd

NEWS & BLOGS

Taking Some Heat: Thermal Imaging for Fever Detection in 2020

As government and business leaders start to talk about "returning to normal," and looking to thermal cameras to help, questions remain about how and whether the latest technology can help prevent the spread of COVID-19. In this entry from the SEMI MEMS & Sensors Group Blog, Jean Brunelle, product manager for infrared imaging at Teledyne DALSA, gives an in-depth perspective. *https://bit.ly/3iD1Jcd*

Redefining the Semiconductor SubFab

In his blog, Alan Ifould, Head of Marketing, Operational Excellence, Edwards Vacuum notes that smart manufacturing and Industrie 4.0 are hot topics right now. "Most agree they are the future and they are everywhere ... but what do they really mean and what in the world is everyone so excited about?!," he asks (and answers). *https://bit.ly/2W2WetI*

American Foundries Act Would Provide Needed Investments in U.S. Semiconductor Manufacturing, Research

The Semiconductor Industry Association (SIA) applauded introduction in the Senate of the American Foundries Act of 2020, legislation that would provide federal investments totaling tens of billions of dollars for semiconductor manufacturing and research to help ensure America's continued leadership in chip technology, which is fundamental to our country's economy and national security. *https://bit.ly/2Zdvn01*

THE FLEXIBLE SOLUTION

COBRA Semiconductor screw vacuum pumps

SETTING NEW STANDARDS OF EFFICIENCY

THE BEST CHOICE
FOR SEMICONDUCTOR APPLICATIONS

COBRA Semicon state-of-the-art dry screw vacuum pumps are the ideal choice for all demanding processes in the production of semiconductors, photovoltaic cells, flat screens, and in numerous industrial coating applications. Low maintenance requirements, low operating costs, a very high uptime and a long life cycle, provide a very low total cost-of-ownership.

Rely on the trusted name for vacuum. Rely on Busch.

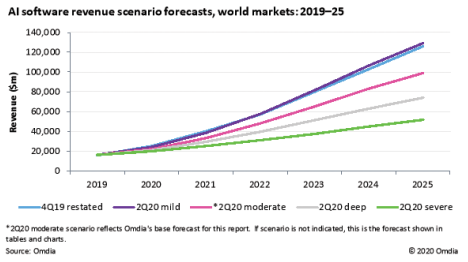
info@buschusa.com | www.buschusa.com

BUSCH

Global AI Software Market to Reach Nearly $100 billion in 2025

The worldwide market for artificial intelligence software will expand to $98.8 billion by 2025, rising by a factor of six from $16.4 billion in 2019—despite the varying effects of the COVID-19 pandemic across different industries, according to Omdia.

The artificial intelligence (AI) software market has experienced tremendous growth over the past several years. While the COVID-19 pandemic has dampened growth forecasts for the market, its impact is uneven across industry sectors. While some industries are retrenching—such as energy, oil, gas, & mining—a few markets, like healthcare, are accelerating AI adoption. Omdia has developed four market scenarios based on the duration of the pandemic and the severity of its economic effect. Omdia's moderate scenario presented in this press release foresees continued double-digit growth ahead, although the cumulative market size for 2019-2025 will be reduced by 22 percent compared to the pre-COVID forecasts.



End users focus on what AI software can do. As a result, Omdia has built a taxonomy of 340 AI use cases across 23 industry sectors that captures the software opportunity related to AI. The use cases explain how and why customers are deploying AI. They help distinguish market reality, i.e., actual deployment and usage, from market hype. Use cases driving the AI software market include voice/speech recognition, video surveillance, customer service & marketing virtual digital assistants (VDAs), network/IT operations monitoring & management and supply chain & inventory management.

"Economic effects from the COVID-19 pandemic have widened the dichotomy between early AI adopters—the 'AI haves'—and the trailing followers—the "AI have nots,'" said Omdia senior analyst, Neil Dunay. "Industries that have pioneered AI deployments and have the largest AI investments are likely to continue to invest in what they view as proven, indispensable technology for cost cutting, revenue generation, and enhancing customer experience."SD

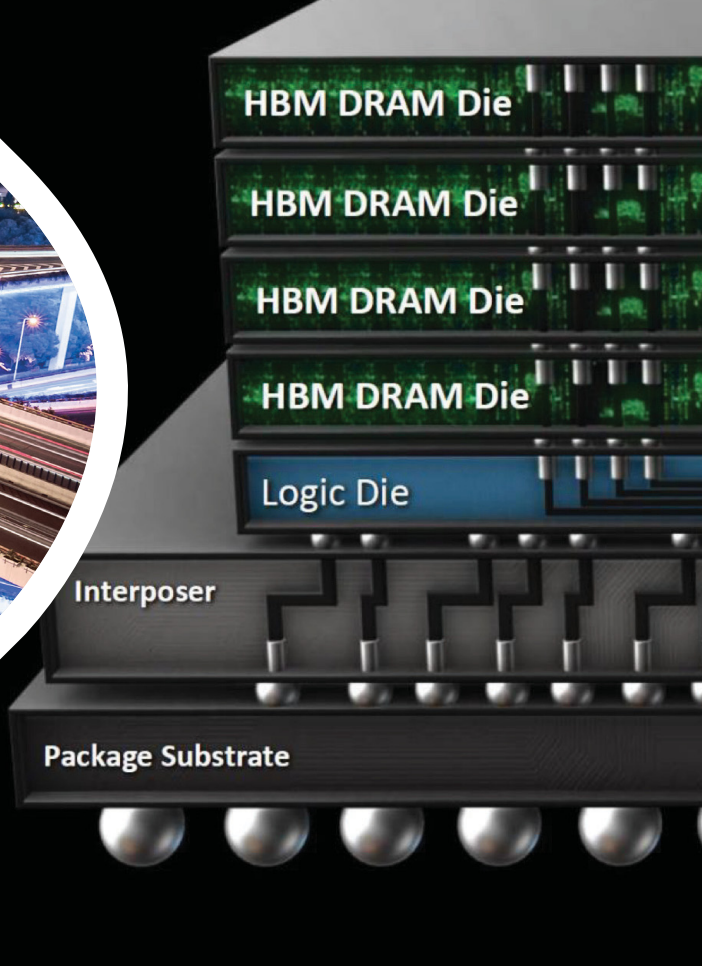
Taiwan Edges South Korea as Largest Base for IC Wafer Capacity

In its Global Wafer Capacity 2020-2024 report, IC Insights breaks down the world's installed monthly wafer capacity by geographic region (or country). Figure 1 shows the installed capacity by region as of December of 2019.

To clarify what the data represents, each regional number is the total installed monthly capacity of fabs located in that region regardless of the headquarters location for the companies that own the fabs. For example, the wafer capacity that South Korea-based Samsung has installed in the U.S. is counted in the North America capacity total, not in the South Korea capacity total. The ROW "region" consists primarily of Singapore, Israel, and Malaysia, but also includes countries/regions such as Russia, Belarus, and Australia.

Some observations contained in the Global Wafer Capacity Report 2020-2024 regarding IC capacity trends by region include:

- As of Dec-2019, Taiwan led the world in wafer capacity with about 22% of worldwide capacity installed in the country. Taiwan surpassed South Korea in 2015 to become the largest capacity holder after having passed Japan in 2011. China became a larger wafer capacity holder than Europe for the first time in 2010 and



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

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

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

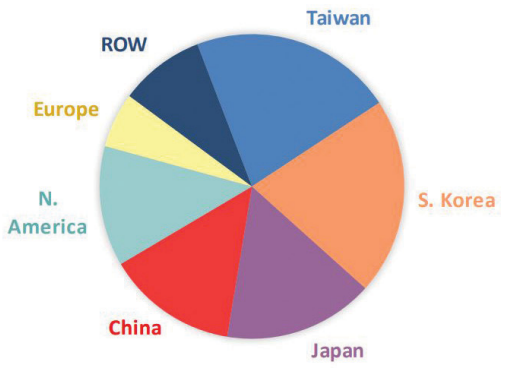
And wafer-level test and measurement becomes nearly essential to guarantee cost-effective fabrication and packaging. Successful verification at this level requires probing and measuring with extraordinary precision, optically, electrically and mechanically.

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

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

FormFactor™
Test insight from lab to fab.

Wafer Capacity at Dec-2019 – by Geographic Region
(Monthly Installed Capacity in 200mm-equivalents)



Region	Installed Capacity (K w/m)	% of Worldwide Total
Taiwan	4,208	21.6%
S. Korea	4,079	20.9%
Japan	3,114	16.0%
China	2,709	13.9%
N. America	2,492	12.8%
Europe	1,140	5.8%
ROW	1,765	9.0%
TOTAL	19,507	100%

Source: IC Insights

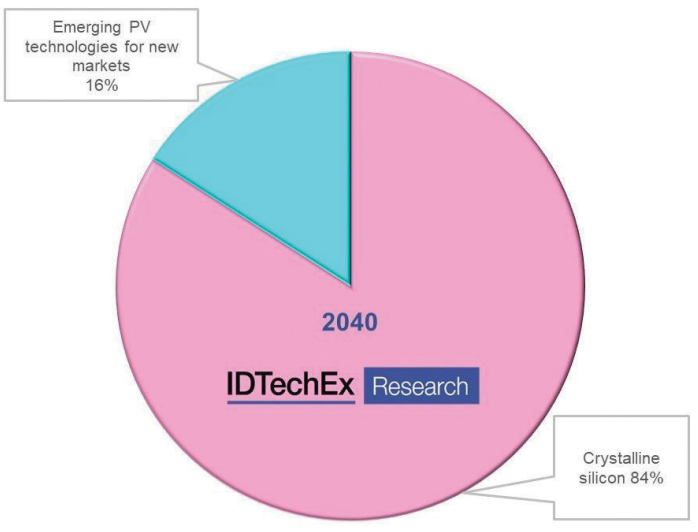
then surpassed North America in 2019. China held 14% of the world's capacity at the end of 2019.

- Taiwan is expected to hold on to the number one position throughout the forecast period. The country is forecast to add nearly 1.3 billion wafers (200mm-equivalent) in monthly fab capacity between 2019 and 2024.
- It is expected that China will surpass Japan in 2020 in terms of the amount of installed capacity. Two years later China is predicted to take over the number two spot from South Korea. The available capacity in China exceeded that of the ROW region for the first time in 2016 and then North America in 2019.
- China is expected to gain the most percentage points of capacity share over the 2019 to 2024 timeframe. While expectations have been tempered somewhat for the roll out of the large Chinese-led DRAM and NAND fabs now under construction, there is also a substantial amount of wafer capacity coming to China over the next few years from memory manufacturers headquartered in other countries and from local manufacturers of other device types.
- The share of capacity in North America is projected to decline over the forecast period, with the region's large fabless supplier industry continuing to rely on foundries, primarily those based in Taiwan. Europe's share of capacity is also expected to continue shrinking.

IC Insights' Global Wafer Capacity 2020-2024—Detailed Analysis and Forecast of the IC Industry's Wafer Fab Capacity report assesses the IC industry's capacity by wafer size, minimum process geometry, technology type, geographic region, and device type through 2024.

Emerging Photovoltaics: Materials Opportunity in New $38 Billion Market

The new IDTechEx report, "Materials Opportunities in Emerging Photovoltaics 2020-2040," is based on interviews by multi-lingual, PhD level IDTechEx analysts across the world and 20 years tracking the research and applications. They predict $38 billion dollars in 2040 without colliding



with the silicon-in-glass "power station" business. There will be many opportunities for premium pricing of its new specialist materials.

For example, over $10,000/W is currently paid for record 30% efficient lll-V compound PV in a designer watch, as an array on a satellite or surface of a high-altitude drone. Another emerging PV technology

copper-indium-gallium-diselenide PV has recently shot to over $2 billion sales in only ten years. Most of the emerging PV is thin film, flexible and some will be stretchable, creating many new markets. Tightly-rollable PV in your mobile phone, regular aircraft skin, billions of Internet of Things nodes?

Hundreds of millions more building facades need this lightweight PV. Tripled-efficiency indoor "lll-V" PV is newly on sale to be followed by underwater. Researchers even target three technologies for PV paint. Vehicle retrofit from boats to buses... this list goes on and on.

Excitingly, the PV research pipeline guarantees robust further improvements to the already dramatic leap in life, efficiency, cost and so on in recent years. CIGS will join organic photovoltaics OPV in being totally free of heavy metals that could be emitted during abuse or wrongful disposal. Rare materials subject to price hikes are being bypassed. Transparency, biodegradability, availability of wide area film, layering different technologies on top of each other - all open up more sales. SD

GaN and SiC Power Semiconductor Markets Set to Pass $1B Mark in 2021

The emerging market for silicon carbide (SiC) and gallium nitride (GaN) power semiconductors is forecast to pass $1 billion in 2021, energized by demand from hybrid & electric vehicles, power supplies, and photovoltaic (PV) inverters.

Worldwide revenue from sales of SiC and GaN power semiconductors is projected to rise to $854 million by the end of 2020, up from just $571 million in 2018, according to Omdia's SiC & GaN Power Semiconductors Report – 2020. Market revenue is expected to increase at a double-digit annual rate for the next decade, passing $5 billion by 2029.

These long-term market projection totals are about $1 billion lower than those in last year's edition of this report. This is because demand for almost all applications has slowed since 2018. Moreover, device average prices fell in 2019. A note a caution: The equipment forecasts used to create this year's forecast all date from 2019, and do not take account of the impact of the COVID-19 pandemic.

SiC Schottky diodes have been on the market for more than a decade, with SiC metal-oxide semiconductor field-effect transistors (SiC MOSFETs) and junction-gate field-effect transistors (SiC JFETs) appearing in recent years. SiC power modules are also becoming increasingly available, including hybrid SiC modules, containing SiC diodes with Si insulated-gate bipolar transistors (IGBTs), and full SiC modules containing SiC MOSFETs with or without SiC diodes.

SiC MOSFETs are proving very popular among manufacturers, with several companies already offering them. Several factors caused average pricing to fall in 2019, including the introduction of 650, 700 and 900 volt (V) SiC MOSFETs priced to compete with silicon superjunction MOSFETs, as well as increasing competition among suppliers.

"Declining prices will eventually spur faster adoption of SiC MOSFET technology," said Richard Eden, senior principal analyst for power semiconductors at Omdia. "In contrast, GaN power transistors and GaN system ICs have only appeared on the market quite recently. GaN is a wide-bandgap material offering similar performance benefits as SiC, but with a higher cost-reduction potential. These price and performance advantages are possible because GaN power devices can be grown on either silicon or sapphire substrates, which are less expensive than SiC. Although GaN transistors are now available, sales of GaN system integrated circuits (ICs), from companies such as Power Integrations, Texas Instruments and Navitas Semiconductor are forecast to rise at a faster rate."

SiC and GaN power semiconductor market trends

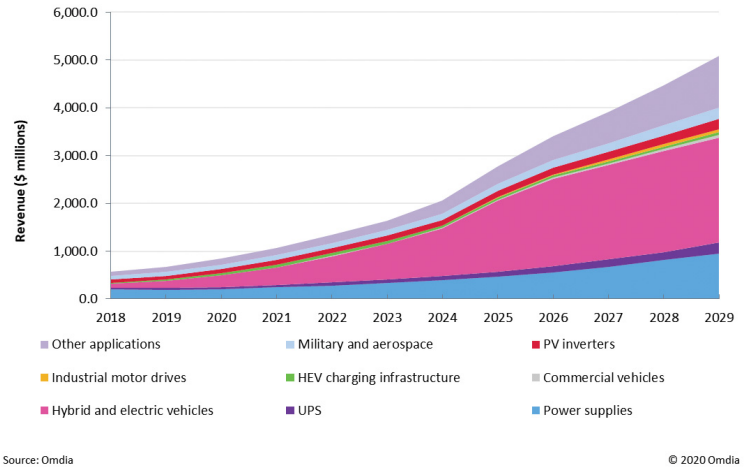
By the end of 2020, SiC MOSFETs are forecast to generate revenue of approximately

$320 million to match those of Schottky diodes. From 2021 onwards, SiC MOSFETs will grow at a slightly faster rate to become the best-selling discrete SiC power device. Meanwhile, SiC JFETs are each forecast to generate much smaller revenues than those of SiC MOSFETs, despite achieving good reliability, price and performance.

"End users strongly prefer normally-off SiC MOSFETs, so SiC JFETs appear likely to remain specialized, niche products," Eden said. "However, sales of SiC JFETs are forecast to rise at an impressive rate, despite having very few active suppliers."

Hybrid SiC power modules, combining Si IGBTs and SIC diodes, are estimated to have generated approximately $72 million in sales in 2019, with full SiC power modules estimated to have generated approximately $50 million in 2019. Full SiC power modules are forecast to achieve over $850 million in revenue by 2029, as they will be preferred for use in hybrid and electric vehicle powertrain inverters. In contrast, hybrid SiC power modules will be used in photovoltaic (PV) inverters, uninterruptible power supply systems and other industrial applications, mainly, delivering a much slower growth rate.

Figure ES.2: SiC and GaN power semiconductors by application



Source: Omdia

Global market revenue forecast for GaN and SiC power semiconductors (millions of US dollars).

What has changed since 2019?

There are now trillions of hours of device field experience available for both SiC and GaN power devices. Suppliers, even new market entrants, are demonstrating this by obtaining JEDEC and AEC-Q101 approvals. There do not appear to be any unexpected reliability problems with SiC and GaN devices; in fact, they usually appear better than silicon.

SiC MOSFETs and SiC JFETs are available at lower operating voltages, such as 650V, 800V and 900V, allowing SiC to compete with Si Superjunction MOSFETs on both performance and price.

End-products with GaN transistors and GaN system ICs inside are in mass production, particularly USB type C power adaptors and chargers for fast charging of mobile phones and notebook PCs. Also, many GaN devices are being made by foundry service providers, offering in-house GaN epitaxial crystal growth on standard silicon wafers, and potentially unlimited production capacity expansion as volumes ramp. SD

ASMC 2020 to Go Virtual

Industry leaders and visionaries will gather in virtual format August 24-26, 2020, for the SEMI Advanced Semiconductor Manufacturing Conference (ASMC) to provide the latest insights and developments across a wide range of critical industry topics from yield management to metrology in the era of artificial intelligence (AI). The conference will feature more than 35 hours of technical content – presented live and available on-demand – on leading-edge semiconductor manufacturing strategies and methodologies. Registration is now open.

The live panel discussion Exascalers, Hyperscalers: Computing at the Edge – Different Modes, Different Nodes will highlight the conference. Panelists will include experts from industry leaders including Applied Materials, Binghamton University, Corning, IBM Global Think Labs and Tercero Technologies. In addition, a tutorial presented by Intel and Entegris will focus on Contamination Control: Enabling High-Yield Manufacturing.

Now entering its fourth

decade, ASMC is the premier event for industry professionals to network and share best practices for semiconductor manufacturing. The conference is co-chaired by Fred Bouchard of SpareTech and Armando Anaya of Northrop Grumman.

ASMC 2020 Keynotes and tech tracks

- Thomas Sonderman, President, SkyWater Technology Sonderman will highlight requirements for manufacturing leadership in the AI era.
- Jesus del Alamo, PhD, Donner Professor, Professor of Electrical Engineering, MITdel Alamo will discuss nanoscale III-V for sustained CMOS innovation.
- Robert Maire, Semiconductor Advisors Maire will explore future market trends.

ASMC 2020 technical tracks include:

- Advanced Equipment Processes and Materials
- Advanced Metal Structures
- Advanced Process Control
- Contamination-Free Manufacturing
- Defect Inspection
- Factory Optimization
- Novel Metrology Techniques
- Optical Metrology and Machine Learning
- Patterning and 3D Metrology
- Smart Manufacturing
- Yield Enhancement/Yield Methodologies

For more conference details, contact Margaret Kindling at mkindling@semi.org.

Imec and GF Announce Breakthrough in AI Chip

Imec and GLOBALFOUNDRIES announced a hardware demonstration of a new artificial intelligence chip. Based on imec's Analog in Memory Computing (AiMC) architecture utilizing GF's 22FDX® solution, the new chip is optimized to perform deep neural network calculations on in-memory computing hardware in the analog domain. Achieving record-high energy efficiency up to 2,900 TOPS/W, the accelerator is a key enabler for inference-on-the-edge for low-power devices. The privacy, security and latency benefits of this new technology will have an impact on AI applications in a wide range of edge devices, from smart speakers to self-driving vehicles.

Since the early days of the digital computer age, the processor has been separated from the memory. Operations performed using a large amount of data require a similarly large number of data elements to be retrieved from the memory storage. This limitation, known as the von Neumann bottleneck, can overshadow the actual computing time, especially in neural networks – which depend on large vector matrix multiplications. These computations are performed with the precision of a digital computer and require a significant amount of energy. However, neural networks can also achieve accurate results if the vector-matrix multiplications are performed with a lower precision on analog technology.

To address this challenge, imec and its industrial partners in imec's industrial affiliation machine learning program, including GF, developed a new architecture which eliminates the von Neumann bottleneck by performing analog computation in SRAM cells. The resulting Analog Inference Accelerator (AnIA), built on GF's 22FDX semiconductor platform, has exceptional energy efficiency. Characterization tests demonstrate power efficiency peaking at 2,900 tera operations per second per watt (TOPS/W). Pattern recognition in tiny sensors and low-power edge devices, which is typically powered by machine learning in data centers, can now be performed locally on this power-efficient accelerator.

"The successful tape-out of AnIA marks an important step forward toward validation of Analog in Memory Computing (AiMC)," said Diederik Verkest, program director for machine learning at imec. "The reference implementation not only shows that analog in-memory calculations are possible in practice, but also that they achieve an energy efficiency ten to hundred times better than digital accelerators. In imec's machine learning program, we tune existing and emerging memory devices to optimize them for analog in-memory computation. These promising results encourage us to further develop this technology, with the ambition to evolve towards 10,000

TOPS/W".

"GlobalFoundries collaborated closely with imec to implement the new AnIA chip using our low-power, high-performance 22FDX platform," said Hiren Majmudar, vice president of product management for computing and wired infrastructure at GF. "This test chip is a critical step forward in demonstrating to the industry how 22FDX can significantly reduce the power consumption of energy-intensive AI and machine learning applications."

Looking ahead, GF will include AiMC as a feature able to be implemented on the 22FDX platform for a differentiated solution in the AI market space. GF's 22FDX employs 22nm FD-SOI technology to deliver outstanding performance at extremely low power, with the ability to operate at 0.5 Volt ultralow power and at 1 pico amp per micron for ultralow standby leakage. 22FDX with the new AiMC feature is in development at GF's state-of-the-art 300mm production line at Fab 1 in Dresden, Germany.

World's Widest Graphene Nanoribbon Shows Promise

With literally the thickness of one carbon atom and electrical properties that can surpass those of standard semiconductor technologies, graphene nanoribbons promise a new generation of miniaturized electronic devices. The theory, however, remains far ahead of reality, with current graphene nanoribbons falling short of their potential. A new collaborative study seen in Communications Materials by a project of CREST, JST Japan including Nara Institute of Science and Technology (NAIST), Fujitsu Laboratories Ltd. and Fujitsu Ltd., and the University of Tokyo reports the first ever 17-carbon wide graphene nanoribbon and confirms it has the smallest bandgap seen to date among known graphene nanoribbons prepared by a bottom-up manner.

Large-scale integrated circuits (LSIs) that use silicon semiconductors are used in a wide range of electronic devices, anywhere from computers to smartphones. They are actually supporting our lives and almost everything else these days. However, although LSIs have improved device performance by reducing the size of the devices, LSI miniaturization is approaching its limit. At the same time, commercial demand continues to put pressure on companies to make higher performing smartphones at smaller sizes, while industry pressure is demanding large-scale manufacturing with smaller equipment.

Other methods and/or materials are definitely needed to solve these problems, says the group leader Dr. Shintaro Sato, Fujitsu Ltd.

"Silicon semiconductors are giving us better performance at smaller sizes. However, we are reaching the limit in how small we can make devices. Thus, we have high expectations for the performance of graphene nanoribbons, which have semi-conducting properties that are only one atom thick – a 2D material," he notes.

Graphene nanoribbons are honeycomb-like structures and, compared to graphene and carbon nanotubes, are the lesser known carbon-based semiconductor family member. Graphene nanoribbons exhibit unique electronic and magnetic properties that do not appear in two-dimensional graphene.

"Interestingly, the electronic and magnetic properties of graphene nanoribbons are widely tuned as a function of the width and edge structure." says Prof. Hiroko Yamada at NAIST.

Conventional graphene nanoribbons, which are promising type of nanoribbon for device application, display width-dependent band gap. They can be classified into three subfamilies (3p, 3p + 1, 3p + 2), their band gaps being inversely proportional to the width of those families. Basically, wider armchair-edge graphene nanoribbons belonging to the 3p + 2 subfamily have the smallest bandgaps among different graphene nanoribbons, having considerable potential to be exploited in GNR-based devices.

So far, 13-armchair graphene nanoribbons belonging to the 3p + 1 subfamily with a band gap of more than 1 eV have been reported, but Sato, Yamada and colleagues show the synthesis

SMART PUMPS FOR SMART MANUFACTURING

LEVITRONIX® PUMP SYSTEMS

The Only Pump with High Performance Data Acquisition, Ideal for:

// Predictive Maintenance
// Continuous Monitoring of Equipment Performance
// Increased System Uptime

LEVITRONIX®

sales@levitronix.com · www.levitronix.com

of a 17-graphene nanoribbon belonging to the 3p + 2 subfamily, which have even smaller bandgaps.

The graphene nanoribbon synthesis was based on the bottom-up approach, called "on-surface synthesis," and a dibromobenzene-based molecule was used as a precursor for on-surface graphene nanoribbon synthesis.

"There are many methods to synthesize graphene nanoribbons, but to produce atomically precise graphene nanoribbons, we decided to use the bottom-up approach. The important point is that the structure of the precursor can define the ultimate structure of graphene nanoribbons if we use the bottom-up approach," explains NAIST's Dr. Hironobu Hayashi, who also contributed to the study.

Scanning tunnel microscopy and spectroscopy by Dr. Junichi Yamaguchi at Fujitsu. Ltd. and non-contact atomic force microscopy by Dr. Akitoshi Shiotari and Prof. Yoshiaki Sugimoto at The University of Tokyo confirmed the atomic and electronic structure of the acquired 17-armchair graphene nanoribbons. Additionally, the experimentally obtained bandgap of 17-armchair graphene nanoribbons was found to be 0.6 eV, and this is the first demonstration of the synthesis of graphene nanoribbons having a band gap smaller than 1 eV in a controlled manner.

"We expect these 17-carbon wide graphene nanoribbons to pave the way for new GNR-based electronic devices," says Sato.

Value of Quantum Computing Remains Uncertain for at Least 10 Years

As innovation continues to accelerate, quantum computing has become an increasingly important technology to monitor as part of the broader wave of digital transformation. Quantum computing aims to solve complex problems that are impossible to address with today's supercomputers and has strong potential across multiple industry sectors, including pharma, energy, finance, logistics, manufacturing, and materials. However, there are significant obstacles in developing the technology that are currently limiting, and these obstacles will continue to challenge developers over the coming years. Over the next 10 years, it is uncertain if quantum computing will consistently outperform today's supercomputers for useful business-related problems, if at all. In the new report "Preparing for Quantum Computing," Lux Research addresses what businesses need to know about quantum computing, including why it's better, when it will become available, and how a company should engage with it.

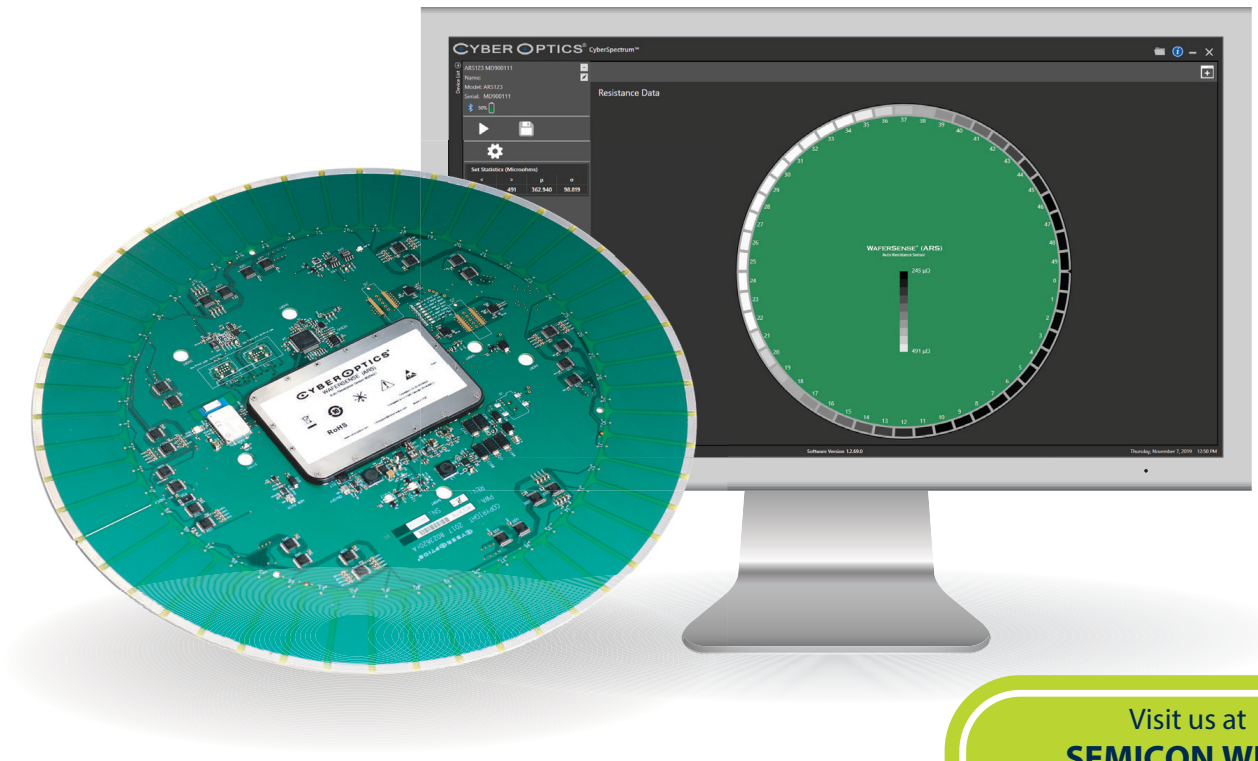
"Today's supercomputers tackle difficult problems including weather modeling, genomic analysis, and computational fluid dynamics, but even the best supercomputers will always be limited in specific areas. They're still unable to handle some important problems in areas like chemical product design, protein folding, or supply chain optimization." says Lux Research lead report author Lewie Roberts. "It's our belief that quantum computing will one day enable multiple industries to address some of these key problems, moving past today's barriers and enabling further innovation." Today, the main problems being targeted by quantum computing are the simulation of quantum systems, machine learning, and optimization.

Currently, quantum computing faces many barriers that limit its near-term development. There are major challenges in hardware development, which severely limit further software development. Quantum bits, or qubits, are inherently unstable, thus reducing the accuracy of any computation that relies on them; this is the first major obstacle to commercialization. For this reason, problems that lack clearly defined answers (like machine learning) but still benefit from improved solutions are the best problems to target with quantum computing. Hardware developers hope to increase the stability of qubits but

Continued on page 67

No Down Time.

Real-time resistance measurements of plating cell contacts in semiconductor ECD applications.



Visit us at
SEMICON WEST
July 20-23, 2020 | Virtual Event

WaferSense® Auto Resistance Sensor™ (ARS) with CyberSpectrum™ Software

Quickly identifies and monitors resistance measurements to detect residue affecting plating pins.

Process and equipment engineers in semiconductor fabs can predict when a tool needs maintenance with quantitative analysis of measured mean resistance over time, shorten equipment maintenance cycles, and improve cell-to-cell uniformity with the wafer-like, 4-wire resistance sensor and CyberSpectrum software's objective and repeatable data. Semiconductor fabs and OEMs worldwide value the accuracy, precision and versatility of the WaferSense Auto Resistance Sensor (ARS). The most efficient and effective wireless measurement device for resistance.

CYBEROPTICS®
www.cyberoptics.com

Fab Models Built on AI Collaborative Knowledge Sharing

STEWART CHALMERS and **TOM HO**, BISTel, Santa Clara, CA

New AI knowledge sharing fab models will play a leading role in shaping the fab of the future.

ADVANCES IN Artificial intelligence (AI), cloud, big data, edge computing and internet of things (IoT) are pioneering the first AI knowledge sharing fab models. An array of AI smart manufacturing applications is fueling a resurgent semiconductor manufacturing industry. Smart applications are coming together to form the first knowledge sharing fab models. Understanding the power of the first AI fab knowledge sharing models is not enough, one must also understand the specific challenges.

Sharing the knowledge, means understanding the problems

How do you share the knowledge among all of the engineers in the factory? One of the challenges to knowledge sharing across the fab is complacent disengagement (FIGURE 1). This behavior underscores the difference between experienced engineers and inexperienced engineers. Experienced engineers bring the advantage of industry knowledge and "boots on ground" skills. It is advantageous for companies to share that knowledge across all the engineers in the fab. The fab of the future seeks to avoid siloed knowledge. Engagement and collaboration across the fab should be encouraged. This is ever more important given the aging of the semiconductor workforce. As the population ages, we have seen the difficulties our industry has had in attracting new talent. A dilemma is created. How do you shorten the gap between experience and inexperience? Knowledge is needed to accelerate response times solving problems quickly and efficiently.

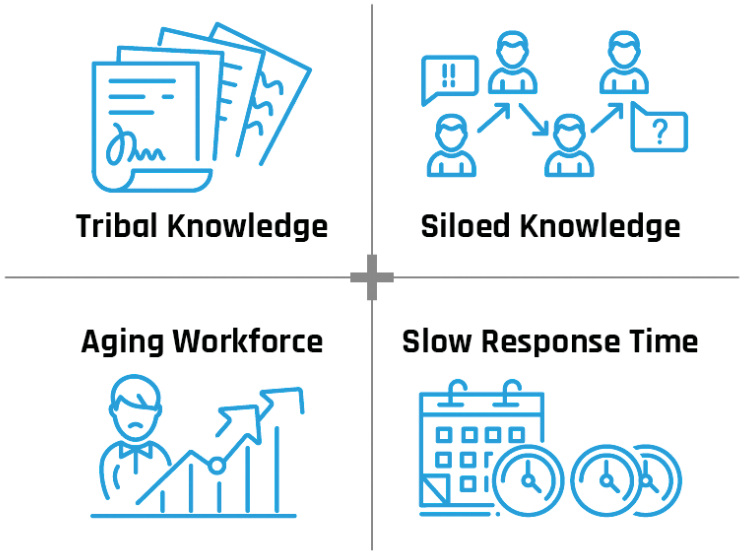


Figure 1. One of the challenges to knowledge sharing across the fab is complacent disengagement.

Capturing knowledge and sharing it quickly among engineers

When engineers are empowered, smarter decisions are made, redundant work is eliminated, and problems can be targeted with efficiency (FIGURE 2). It is critical to achieving this objective that engineers have the tools to share knowledge across the organization.

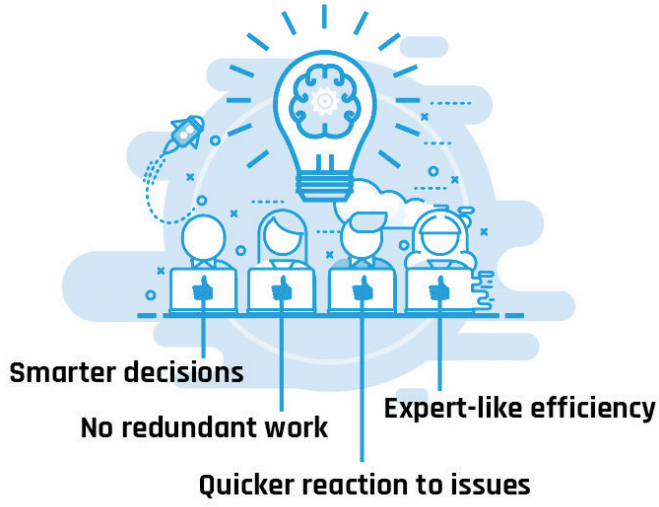


Figure 2. When engineers are empowered, smarter decisions are made, redundant work is eliminated, and problems can be targeted with efficiency.

What decisions are being made to solve certain problems?

In FIGURE 3, everyone has data and when we connect the data it becomes information and then knowledge. As we increase our knowledge, engineers began to enjoy greater insights. Engineers can then solve problems. Solving problems requires knowledge, experience and wisdom. It is also important that there is a path to getting there – the data evaluation process.

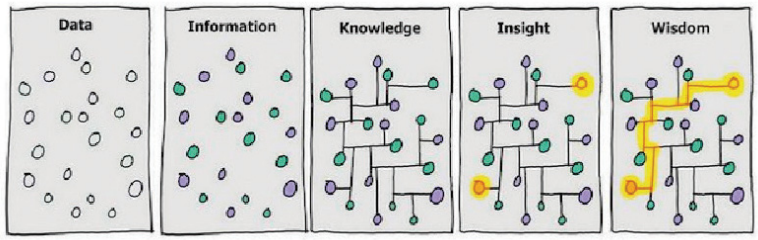


Figure 3. Solving problems requires knowledge, experience and wisdom. It is also important that there is a path to getting there – the data evaluation process. *Cartoon by David Somerville, based on a two pane version by Hugh McLeod.*

Knowledge sharing fab models empower engineers

Recent advances in AI smart manufacturing applications have created enabling technologies that have moved the knowledge sharing ball forward. AI, cloud, big data, edge computing and IoT technologies help capture and retain enormous data. These technologies empower engineers to make quicker and more insightful decisions. This is the bridge to efficiency. Today, we have the ability to gather much more information than ever, helping the modern fab to digest and transform data into knowledge and actionable insights.

Current smart manufacturing operates under the premise that decisions follow decision independent processes of detection, analysis and prediction. Detection identifies problems and alarms engineers that something is wrong. The second aspect of the process is analysis to pinpoint the root cause of the problem. Prediction involves identifying issues before they occur. It may also involves predicting the remaining useful life of critical equipment. These types of applications are now starting to incorporate AI elements that make the process quicker and smarter (FIGURE 4). Smart applications make analysis accurate and more exhaustive. By integrating AI into BISTel's smart applications, BISTel enables engineers to be more efficient in identifying the problems. Understanding these issues helps us to create solutions that can predict the behavior of certain tools and processes. Many of today's AI applications are mostly point of use. This is an evolution and, in the future, fabs will adopt a truly comprehensive knowledge sharing fab model where all data and insights are shared and understood.

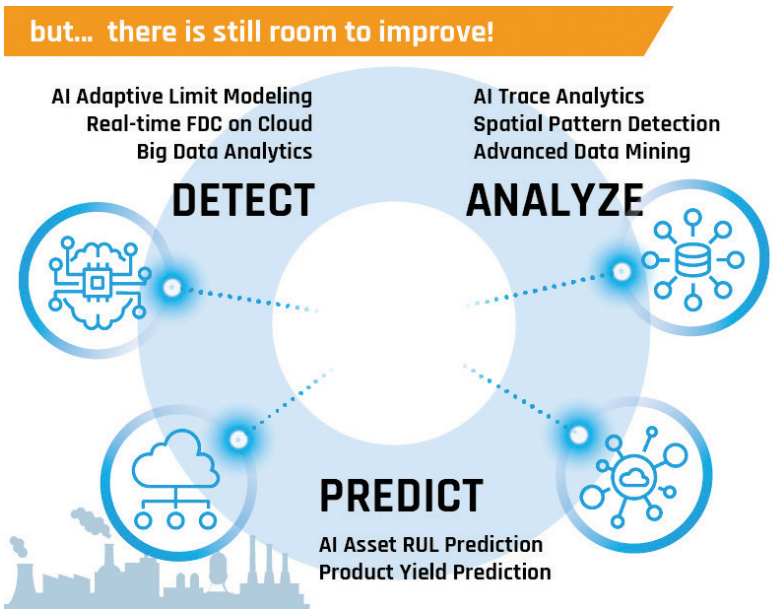


Figure 4. Smart applications make analysis accurate and more exhaustive.

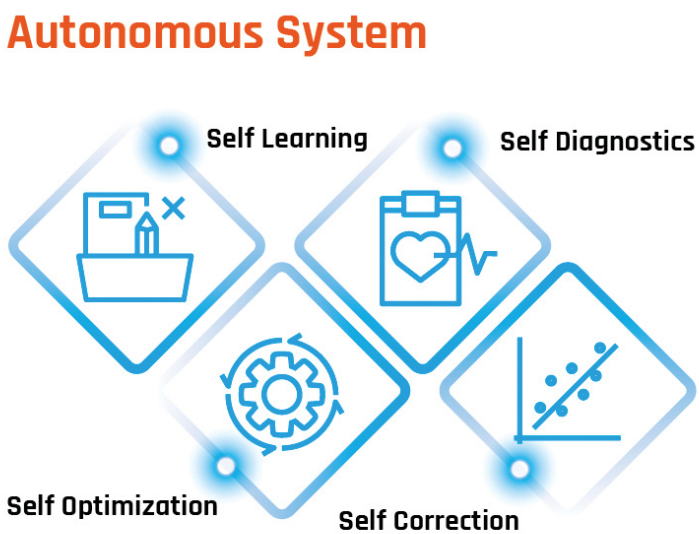


Figure 5. An autonomous system discovers new things and learns the relationship between cause and effect, and it uses that knowledge and acts on it independently.

How can we create a system where the system learns by itself?

How can we create a system where the system learns by itself? When it learns, the new Decision Support System (DSS) it offers diagnoses and then self-optimizes and, if warranted, self -corrects. In 2020, BISTel will introduce the industry's first AI central

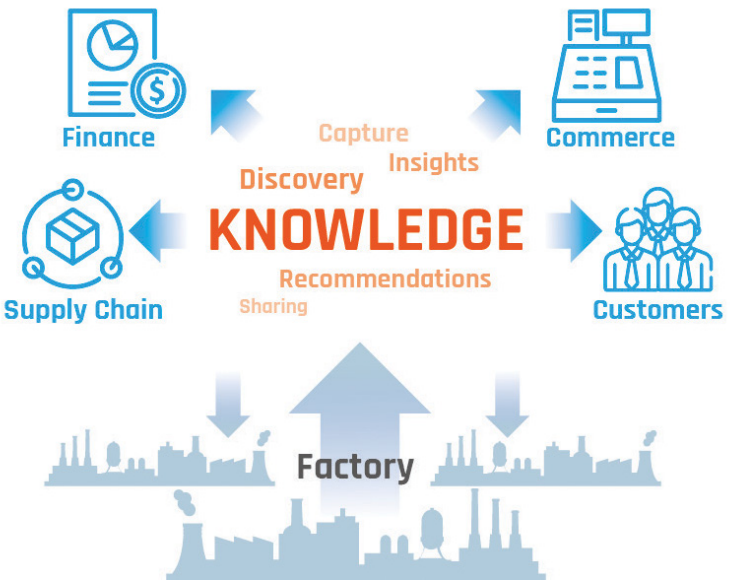


Figure 6. The technology that provides for the decision support system could very easily reach beyond the factory.

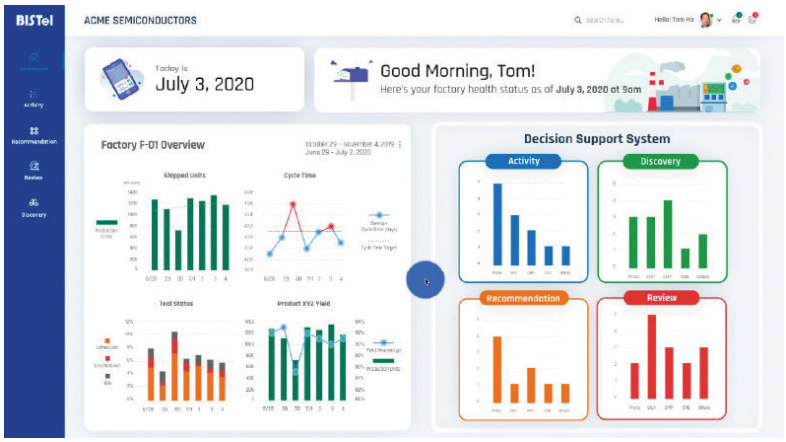


Figure 7. Key activities being performed by the DSS are Activity, Recommendations, Reviews and Discovery.

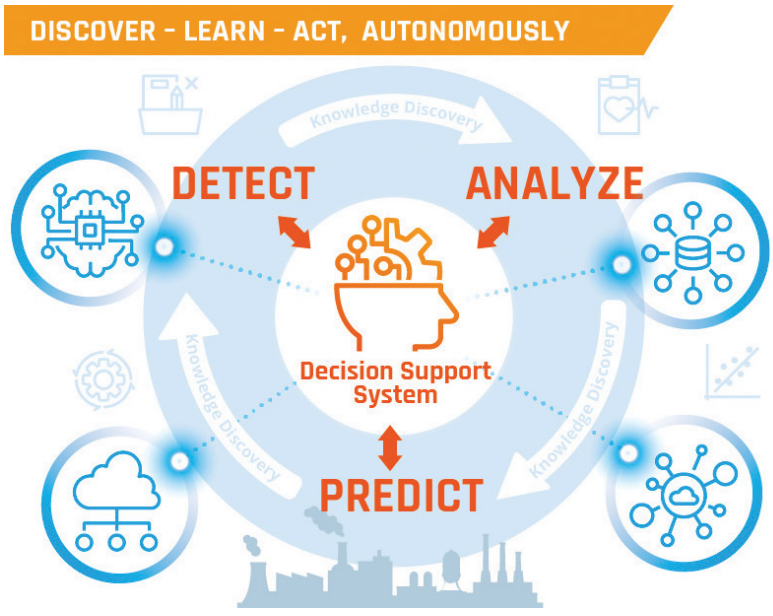


Figure 8. DSS supports quicker decision making.

platform for knowledge sharing across the fab and across the manufacturing ecosystem (FIGURE 5). The new Decision Support System (DSS) is smart enough to capture the learning but also act on the learning. This is what we refer to as the "knowledge sharing fab model". Detection, analyses, and prediction integrated into one AI system that continues to build and share knowledge for the modern fab. The Decision Support System (DSS), conceptualized as a brain of sorts, discovers and captures all of the constant fragments of data extrapolated from the fab environment. This is a hallmark of the fab of the future.

AI DSS goes beyond the factory

The technology that provides for the decision support system could very easily reach beyond the factory (FIGURE 6). All have the potential to be revolutionized by such an autonomous system. It discovers new things and learns the relationship between cause and effect, and it uses that knowledge and acts on it independently. DSS has the power to connect across factories, organizations and industries to include supply chains, finance, IT, logistics and others. This is not just an application that resides within fab or factory. DSS can share knowledge across many different types of manufacturing organizations.

So, what do the first knowledge sharing fab models look like?

In the first knowledge sharing fab model, engineers are presented each morning with a personalized dashboard at their workspace. The dashboard may offer an update on the current status of the fab. Examples of updates may include illustrations representing number of units shipped, cycle time and yield status. Key activities of DSS may be depicted on an engineer's dashboard. In FIGURE 7, key activities being performed by the DSS are presented – Activity, Recommendations, Reviews and Discovery.

The ACTIVITY feature provides information to the engineer on the health status of equipment and processes. In predictive based maintenance, knowledge sharing allows for an immediate influence on the fab and sub fab operations. For instance, in Fig. 7, if the DSS identifies a malfunction in a specific pump. The problem is highlighted on the engineer's dashboard. This problem could be affecting the tool performance and yield in the fab. The system predicts the remaining useful life (RUL) of the asset and then creates a work order flow back to the enterprise resource planning system (ERP) where the system can order

parts and/or schedule maintenance. This prediction also includes actions and recommendations for the engineer. The ability to predict when an asset will fail is having a huge impact on pharmaceutical, semiconductor and oil and gas manufacturing operations, reducing system downtime and saving considerable maintenance costs. The system logs the activity, as well as the correction, displays the incident and outcome and learns from it.

The RECOMMENDATIONS feature identifies problems and generates a set of solutions that is then presented to the engineer. Some solutions may call for approval. All of the work is done by the system and the human makes the decisions based on this knowledge capture.

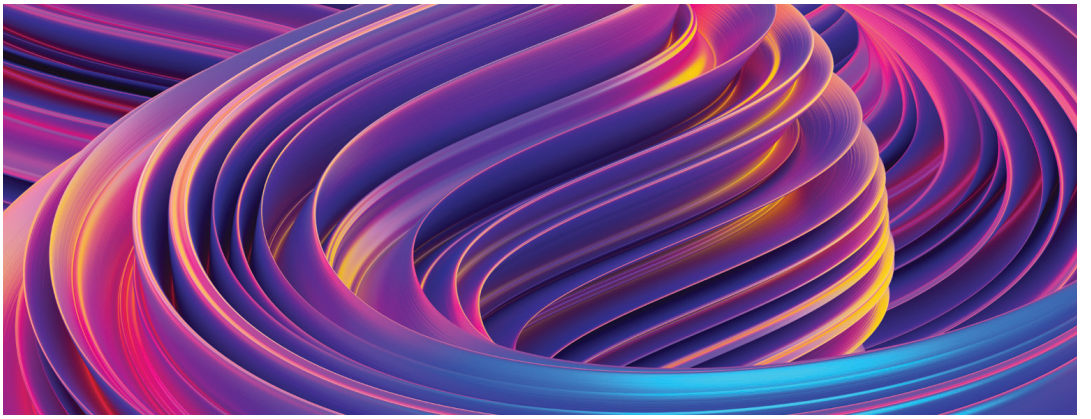
A third feature Is REVIEW. If a novel sequence of events is presented, The DSS flags the engineer or operator for specific input to the system. If, for example, the system identifies an identical alarm repeat (IAR). The system will recognize the novel pattern, analysize the data and then, if needed, advises a review. The engineer may provide domain input to the system. Any subsequent input may be incorporated into a new protocol. In future, DSS would then identify and act autonomously. The knowledge has been captured and shared.

Feature four is DISCOVERY. This is an example of knowledge empowerment. If, in practice, the DSS, notes a trace alarm, records when it happened and on what lot, wafer and recipe and tool, DSS will then enrich the knowledge base. The system learns by itself. It transforms observed behavior from humans and processes, "connects the dots" and created a solution (FIGURE 8). DSS helps immensely by supporting quicker decision making and/or support.

AI based decision system support systems will proliferate over the next several years. New AI knowledge sharing fab models will play a leading role in shaping the fab of the future. Autonomous thinking accelerates our path to the fab of the future. sd

About the authors

Stewart Chalmers is VP Business Development and Marketing BISTel. He can be reached at stewrt.chalmers@bistel.com Tom Ho is President & GM, BISTel America.



Metrology Solutions for Gate-All-Around Transistors in High Volume Manufacturing

NICK KELLER and **ZHUAN LIU**, Onto Innovation, Milpitas, CA

High-speed, non-destructive OCD metrology has the capability needed to support process control throughout the GAA process flow.

GATE-ALL-AROUND (GAA) TRANSISTORS offer significant performance advantages at advanced nodes, but only at the cost of significant increases in process complexity. Complicated three-dimensional structures and shrinking critical dimensions make precise, accurate metrology in GAA manufacturing processes both more important and more challenging. Scatterometry-based optical critical dimension (OCD) metrology has become mainstream in the last several generations of semiconductor development, in part because of its ability to measure three dimensional shapes and subsurface/re-entrant features. The latest generation of OCD systems combines improvements in signal-to-noise ratios, signal fidelity and advanced machine learning capabilities that allow it to support the most challenging GAA process steps with repeatable measurements and production worthy throughput.

GAA vs planar and finFETs

The revolution in electronics that became the semiconductor industry was built on the discovery that the flow of electric current could be controlled by transistors fabricated from semiconductor materials, and the subsequent developments that found ways to continually decrease the size of those transistors. For most of its history, the industry has focused on planar transistor architectures in which a gate positioned over a channel controls the flow of current through the channel between a source and a drain. A voltage applied to the gate creates an electric field (FET - field effect transistor) that excludes or permits carriers in the channel thus turning the current on or off. The source, channel, and drain are coplanar, created at the surface of a semiconductor wafer, with the gate positioned over the channel (**FIGURE 1**). Increasing the computing power of an integrated circuit is essentially an exercise in reducing its aerial (X and Y) dimensions.

As nominal gate lengths (nodes) approached 20nm, planar devices encountered short channel effects, such as increasing leakage currents, that degraded their performance. To combat these effects manufacturers moved to finFETs, in which the channel has the shape of a fin, surrounded on three sides by the gate (Fig. 1). This increased the effective area of the gate in proximity to the channel. FinFET devices have allowed continuing increases in computing power through the current leading-edge nodes – around 5nm. (The node name no longer accurately reflects the gate length but is rather a convention to reflect successive

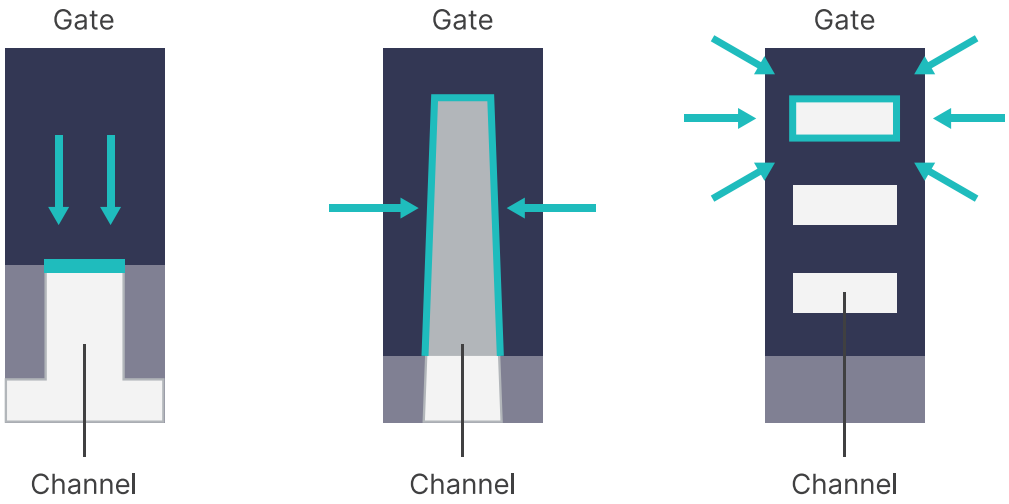


Figure 1. Planar, fin and gate-all-around field effect transistor architectures [1].

generations of increasing device density and computing power.) Beyond this node, manufacturers have encountered limitations for finFETs and have had to consider other architectures. Several have chosen the gate-all-around (GAA) design in which, as the name suggests, the gate completely surrounds the channel (Fig. 1). GAA devices promise continuing improvement in performance but include three dimensional features that greatly increase the complexity of the manufacturing process.

Process control and optical metrology

Process control, ensuring that the process reliably creates functional devices with physical and electrical characteristics that fall within established process windows, is an essential part of every semiconductor manufacturing operation. Metrology, the science and practice of measuring process performance, provides the basis for process control. Just as manufacturing processes have evolved to create smaller, more complex devices, measurement technologies have had to change to monitor the new processes. Most image-based optical metrology became obsolete as critical dimensions (CD) decreased into the sub-micrometer range decades ago. The mainstay of the industry since that time has been scanning electron microscopy – specially designed for CD measurements (CD-SEM). CD-SEM is non-destructive and provides top-down, two-dimensional measurements. Cross sectional SEM (XSEM) can provide three-dimensional information, but at the cost of additional, destructive sample preparation. As device sizes have continued to decrease, dimensions have exceeded the resolution of SEM and manufacturers have adopted transmission electron microscopy (TEM). TEM can resolve individual atoms, but only at the cost of time-consuming, destructive sample preparation. It remains the gold-standard for accuracy and reference measurements, but its slow turnaround and low throughput are not well-suited to process control applications.

Although critical dimensions long ago passed beyond the resolution capability of optical imaging, other optical techniques can deliver fast, repeatable, non-destructive measurements. Optical critical dimension (OCD) measurements based on scatterometry derive shape, dimension and composition information from the scattering patterns observed in light that has interacted with the sample. It requires a regular array of similar features, but these are common in integrated circuits. The target may be an in-circuit feature, such as a line array, or a specially designed measurement target, typically located in the area between die on the wafer.

The simplest illustration of an OCD a measurement is the interference pattern created when light falls on a regularly spaced array of lines and spaces. The spacing of the interference fringes is a function of the wavelength of the light, the configuration of the optical path, and the spacing of the lines. Because the information is carried in the phase relationships of the light waves, the technique is not constrained by wavelength-related diffraction limits on image resolution. As manufactures moved beyond the 20nm node and adopted finFET architectures with critical three-dimensional features, scatterometry entered the mainstream of process control metrology.

Scatterometry for semiconductor manufacturing process control is based on ellipsometry. An ellipsometer measures the effects of reflection (or transmission) on polarized light. Ellipsometry has long been used in semiconductor metrology to characterize thin films in multilayer stacks. It is exquisitely sensitive and accurate, capable of measuring films as thin as a single atomic layer. Ellipsometers measure a material's complex refractive index or dielectric tensor to determine fundamental physical properties. They can be used to characterize film thickness, composition, roughness, crystalline nature, doping concentration, electrical conductivity, and more.

Conventional ellipsometers look at polarized light reflected from the sample and compare it to the known polarization state of the incident light to measure the complex reflectance

$$MM = \begin{bmatrix} MM_{11} & MM_{12} & MM_{13} & MM_{14} \\ MM_{21} & MM_{22} & MM_{23} & MM_{24} \\ MM_{31} & MM_{32} & MM_{33} & MM_{34} \\ MM_{41} & MM_{42} & MM_{43} & MM_{44} \end{bmatrix}$$

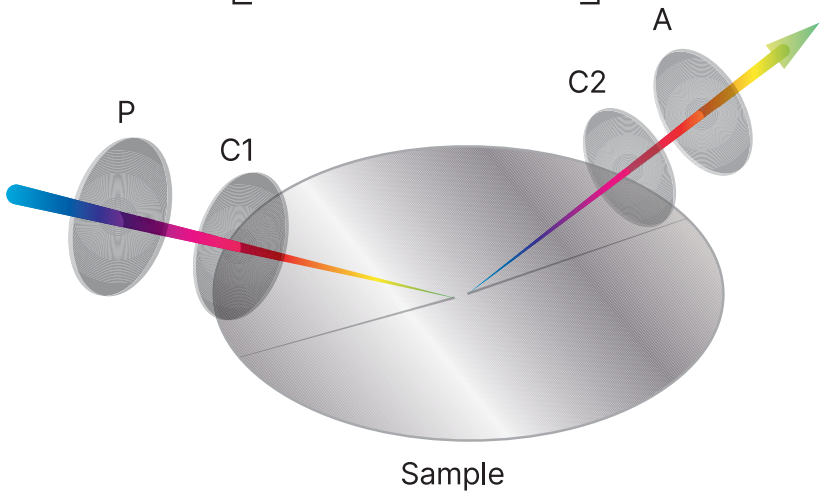


Figure 2. An ellipsometer compares incident polarized light and reflected polarized light to determine structural and material properties. With two compensators, it can acquire all 16 elements of the Mueller matrix that completely describes the reflection.

ratio, composed of two parameters, an amplitude component (ψ) and a phase shift difference (Δ). Spectroscopic ellipsometers use a broadband light source and measure these parameters as a function of wavelength. When used to measure thin, unpatterned films, the analysis typically assumes the sample is composed of a small number of discrete, well-defined layers that are optically homogeneous and isotropic. These assumptions are valid, and the two parameters, ψ and Δ, are sufficient, for most thin film applications.

The assumptions are not valid for scatterometry measurements of complex three-dimensional features. Conventional spectroscopic ellipsometry measures only the amount of incident p-polarized (electric field parallel to the plane of incidence) light that is reflected as p-polarized light and the amount of incident s-polarized (electric filed perpendicular to the plane of incidence) light that is reflected as s-polarized light. However, there may also be cross polarized scattering, p to s and s to p. Mueller matrix spectroscopic ellipsometry (MMSE) captures a complete description of the polarized reflection, including cross-polarization and circular polarization, in a matrix of 16 elements at each wavelength. Cross polarization carries important information about characteristics such as symmetry, edge roughness and anisotropic optical properties. It is essential for characterizing 3D structures. The full Mueller matrix can be measured using a spectroscopic ellipsometer with dual rotating compensators (**FIGURE 2**), one between the polarizer and the sample and one between the sample and the analyzer.

The utility and value of full Mueller matrix ellipsometry varies with the application. In some cases, it is essential, such as the measurement of structural anisotropy like tilt and overlay shift. In other cases, it is not necessary, but still valuable, such measurements of complex structures where the additional information can help in parameter decorrelation. Finally, the additional information may be only potentially valuable but is essentially free, as when the dual compensators allow full Mueller matrix collection without using different analyzer angles.

Data analysis – RCWA and machine learning

OCD is an indirect measurement. No analytical solution exists to derive the desired physical or material properties directly from the measured parameters. Rather, the process relies on the development of models and their comparison to acquired data. The classic process relies on rigorous coupled wave analysis (RCWA) to generate a set of expected Mueller matrix elements based on theoretical interactions of light with a virtual model of the structure that includes shape, dimensions, material and optical properties, and more. Parameters of the model are varied, and the resulting matrix elements recorded. Regression analysis seeks to identify key features of the matrix element spectra that vary predictably and uniquely with the parameter of interest and can therefore serve as reliable proxies. The modeling process can be time-consuming, computationally intensive, and expensive. In use, measured data are compared to the modeled data to infer the desired value, and measurements are quite fast.

Recent developments in artificial intelligence (AI) and machine learning (ML) offer significant improvement in the cost and time to solution. Machine learning essentially automates the process. Given an appropriate dataset of measured MMSE spectra and reference values, machine learning can often find the salient spectral features and quantify their relationships to the parameters of interest without physical modeling or structured regression analysis. ML-based solutions are unlikely to completely replace model-based solutions. Rather, they will provide a complementary capability for situations where modeling is especially challenging. The ideal space for ML solutions will be situations where modeling costs are high because of the complexity of the structure, the key parameter of interest has dominating or unique sensitivity in the signal, and reference data is abundantly available.

GAA process

One design for GAA transistors incorporates multiple vertically stacked

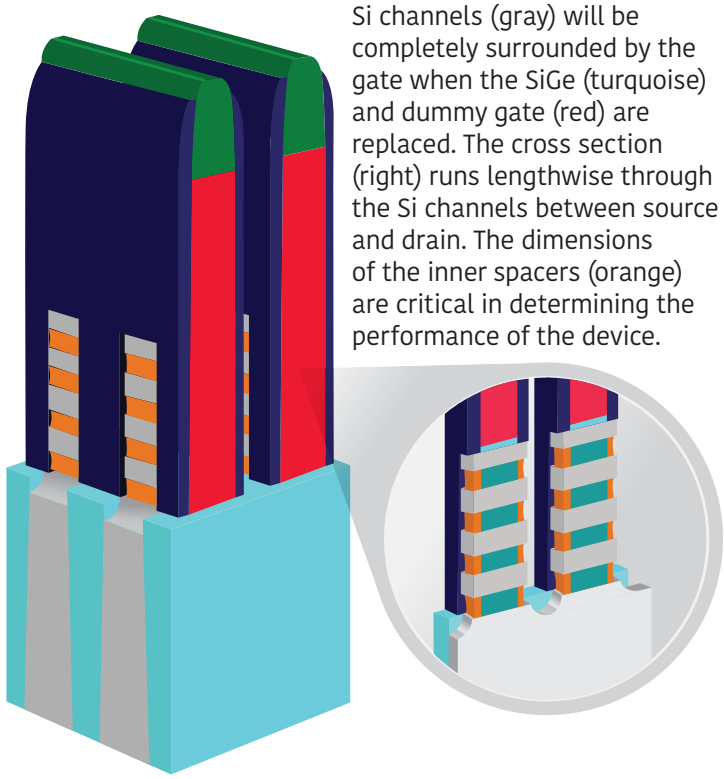


Figure 3. GAA transistor - the Si channels (gray) will be completely surrounded by the gate when the SiGe (turquoise) and dummy gate (red) are replaced. The cross section (right) runs lengthwise through the Si channels between source and drain. The dimensions of the inner spacers (orange) are critical in determining the performance of the device.

nanosheet channels passing through a single gate. The GAA process flow bears some similarity to finFET processes. The process begins with the creation of a superlattice, a stack of alternating, epitaxially deposited silicon and silicon germanium layers. Trenches etched through the lattice create fin-like structures, with each fin containing 3-4 silicon nanosheet layers that will become the transistor channels. The silicon layers alternate with SiGe layers that will eventually be replaced by gate materials. Dummy polysilicon gates are deposited across the nanosheet-fins and spacer material conformally deposited over all. Source and drain are etched either side of the gate, cutting through and exposing the ends of the Si channels. In a series of critical steps, the exposed SiGe between the ends of the Si channels is selectively etched to create cavities for the inner spacers, and inner spacers are deposited in the cavities (**FIGURE 3**). These features are extremely small and yet their dimensions are critical for several reasons. The depth of the cavity/inner spacer determines the length of the gate, the inner spacer protects the subsequently deposited source and drain during layer release when the dummy gate is etched away and replaced with gate materials, and the spacer suppresses parasitic capacitance between the source/drain and gate.

Each successive finFET node has required about a 30% improvement in the signal to noise ratio (S/N) in OCD systems (**Figure 4**). The transition to GAA architectures with exceptionally small, complex features like inner spacer cavities and inner spacers puts additional pressure on S/N. Still, OCD has the core capability needed for GAA processes. Onto Innovation has been developing OCD solutions for the GAA device with its Atlas OCD system. It was found that the earlier generation Atlas III system has a noise level close to the signal level for a 0.1nm increase in SiGe etch back, which limits the capability to measure the individual SiGe etch back. Improvements in the subsequently introduced Atlas III+ system provided a better signal for that measurement, and the recently introduced Atlas V system improves even more in signal to noise ratio, which provides the possibility to measure the individual dimensions of the SiGe etch back.

An overall (average) value for SiGe cavity etch is helpful, but what is really needed is a separate measurement for each spacer because the transistor will be defined by the "worst" nanosheet. The Atlas V metrology system, using the newly released AI-guided engine in the OCD solution software that includes both classical modeling and advanced

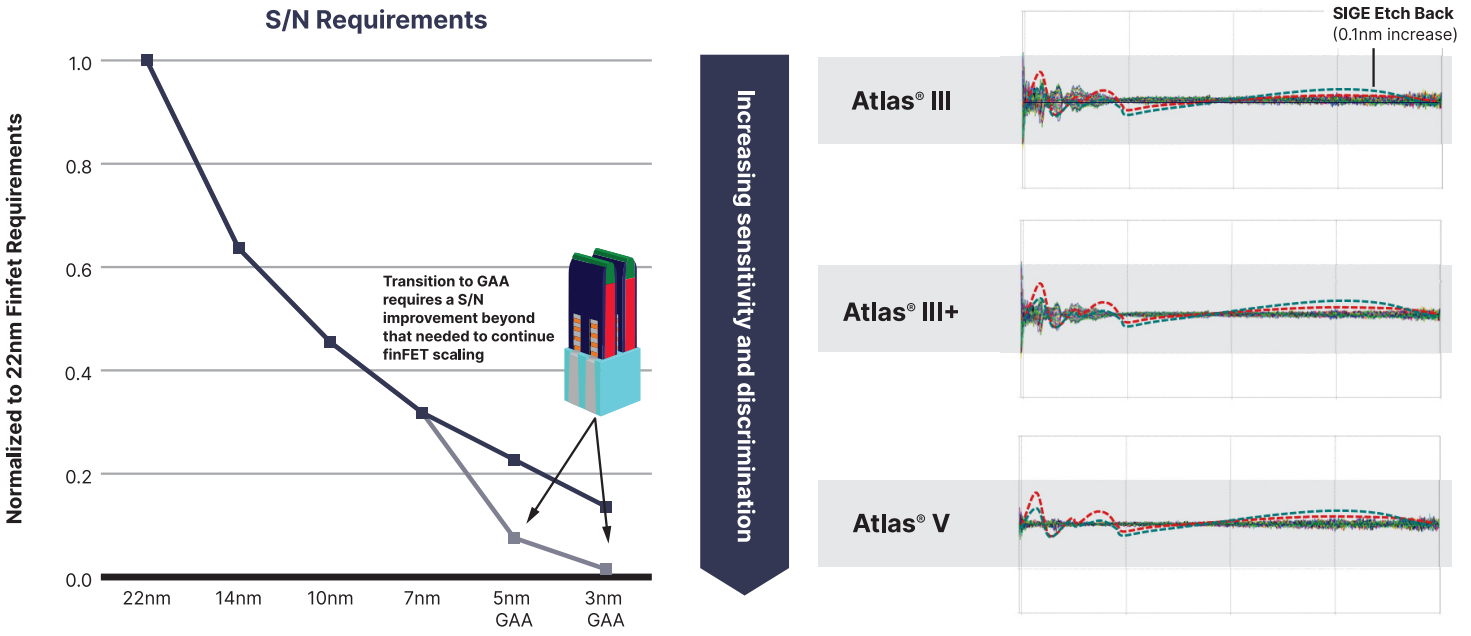


Figure 4. (Left) GAA transistors offer better performance than finFETs but are more complex and require a significant improvement in S/N of OCD measurement systems. The spectra (right) the steady improvement in S/N over three generations of OCD metrology systems.

machine learning capabilities, demonstrated individual inner spacer measurements with significant improvement in P/T (precision/tolerance) compared to average inner spacer measurements made with the Atlas III+ system (**FIGURE 5**).

Manufacturers are exploring other measurement technologies for some GAA process steps, most are X-ray-based. X-ray reflectivity (XRR) can measure the thin films in the initial superlattice, though this is an easy measurement for ellipsometry. High resolution X-ray diffraction (HRXRD) is useful for single crystal thin films. Onto Innovation has also done some work with critical dimension small-angle X-ray scattering (CD-SAXS), a scatterometry-like technique, but it is currently slow and expensive. X-ray photoelectron spectroscopy (XPS) is a surface analysis technique that can measure the composition, chemical state, and electronic state of elements in thin films.

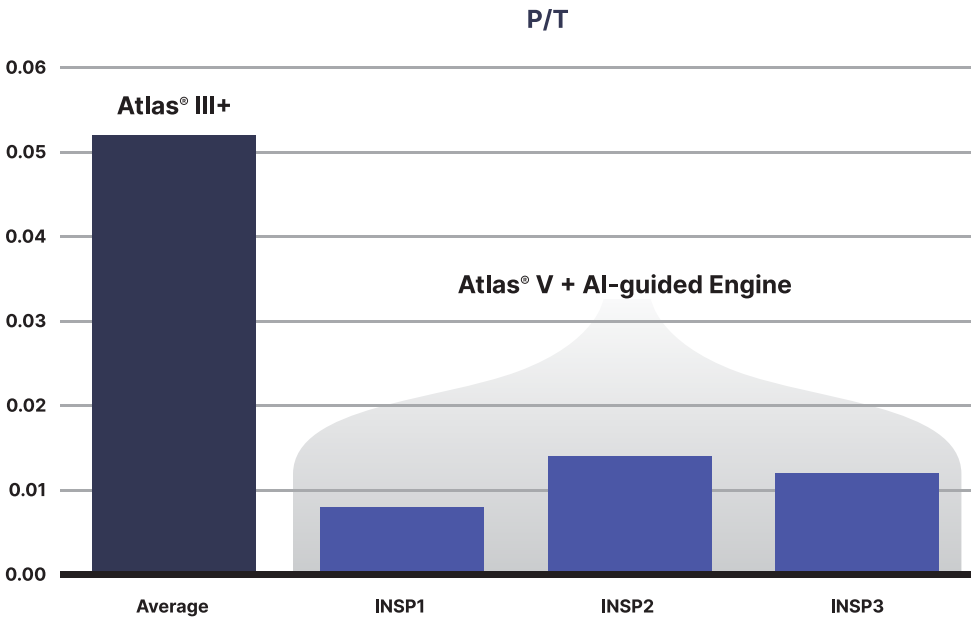


Figure 5. One of the greatest metrology challenges in the GAA process flow is the measurement of inner spacer cavity etch and the inner spacer itself. The latest generation systems can provide individual spacer measurements with several times better precision/tolerance than the average measurements available from previous generation systems.

Conclusion

GAA transistors are the most likely candidate to replace finFETs, putting more computing power in less space. They offer improved performance but require more complex processes to create smaller, three-dimensional structures. Controlling these processes requires improved metrology. OCD metrology has the fundamental capability. Current generation systems have demonstrated the ability to measure inner spacers, among the most challenging features. High-speed, non-destructive OCD metrology has the fundamental capability needed to support process control throughout the GAA process flow. SD

How COVID-19 is Impacting the Memory Industry

WALT COON and **MIKE HOWARD**, Yole Développement, France

Heading into 2020 both the DRAM and NAND industries were projected to have turn-around years after suffering through much of 2018 and 2019. However, with the global lockdown triggered by the COVID-19 pandemic the outlook for the industry has changed considerably, and not all the change is for the worse. In this article we explore the impact on memory demand, how suppliers are expected to react to the pandemic, and the likely impact to pricing over the near-to-midterm.

IT IS EASY TO IMAGINE HOW A GLOBAL PANdemic that results in widespread lockdown might negatively impact demand. Perhaps a more illuminating question is if there are any demand segments that have been positively impacted by COVID-19. The short answer is yes. One of the big winners during this crisis has been datacenter, and we thought we would share a few anecdotes that illustrate this.

The first example is Netflix. Over the last eight years Netflix has averaged about four and a half million new customers per quarter. In Q1-20 much of the world went on lockdown and we saw a huge surge in media streaming. The result was that in Q1-20 Netflix had nearly 16 million new users - more than three times their average and 60% more than their prior record - and Q1 only had a few weeks of lockdown for much of North America and Europe. Clearly this has been a boon for streaming media and the datacenters that host streaming media.

Another area where we have seen surging demand is gaming. Steam, the online gaming platform, averaged about 16 million users per day over the six months prior to lockdown. During lockdown it has averaged 23 million users per day, a nearly 50% jump in demand.

Other categories that have been driving datacenter demand include online shopping and video conferencing services such as Zoom and WebEx, which are taking the place of face-to-face interaction as to the world adjusts to "work from home".

Before COVID we expected the Datacenter to generate about $38 billion in memory revenue in 2020, with revenue weighted towards the second half of the year. Today we anticipate slightly higher revenue but split evenly between the first and second half of the year. Essentially, the first-half surge in demand has changed the half-on-half dynamics for 2020. Cloud Service Providers have been big buyers thus far this year

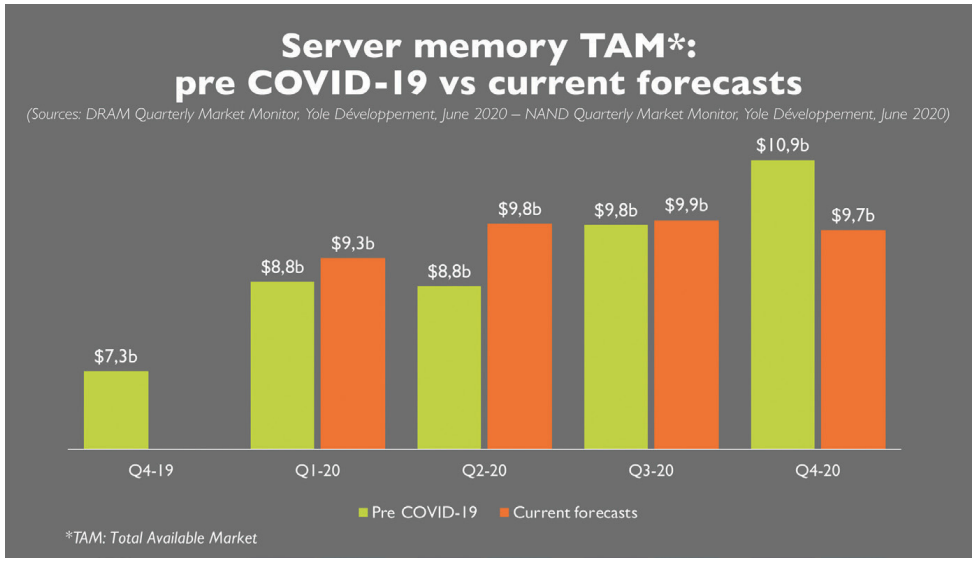


Figure 1

as they have seen significant upticks in demand and have grown inventory to hedge against any potential supply chain disruptions that might arise due to COVID. Further, traditional enterprise has been scrambling to assess "work from home" and budgets have not been cut at this stage (**FIGURE 1**).

There is, however, ample uncertainty about the second half of 2020. We expect traditional enterprise to reduce spending as the full economic impact of COVID is better understood. Additionally, although CSPs have been big drivers of first half demand, there is concern that shrinking ad revenues, overly robust memory inventories, and general economic malaise will prompt a spending cut in the second half.

PCs

Another category which we believe will fare relatively well during the COVID pandemic is PCs. Our pre-COVID forecast was for ~$24.5 billion to be spent on memory for PCs and our current outlook is only down about $1B from this level. While Q1 PC unit sales were much lower than anticipated, much of that was due to supply chain constraints rather than a lack of demand. Those constraints have now been resolved and we expect Q2 sales of PCs to be up quite dramatically.

Clearly "work from home" is causing the surge in PC demand as people realize that getting real work done – especially for extended durations - requires the right equipment. This surge in PC demand is likely a one-time uplift in demand and not a systemic change to the PC market. (**FIGURE 2**).

For the year, we have reduced our PC unit forecast about 1%. Even though we are seeing strong demand today as people attempt to work from home, we expect PC sales in the second half of 2020 to suffer as economic headwinds take hold, people working from home have made the needed upgrades, and enterprise looks to tighten budgets.

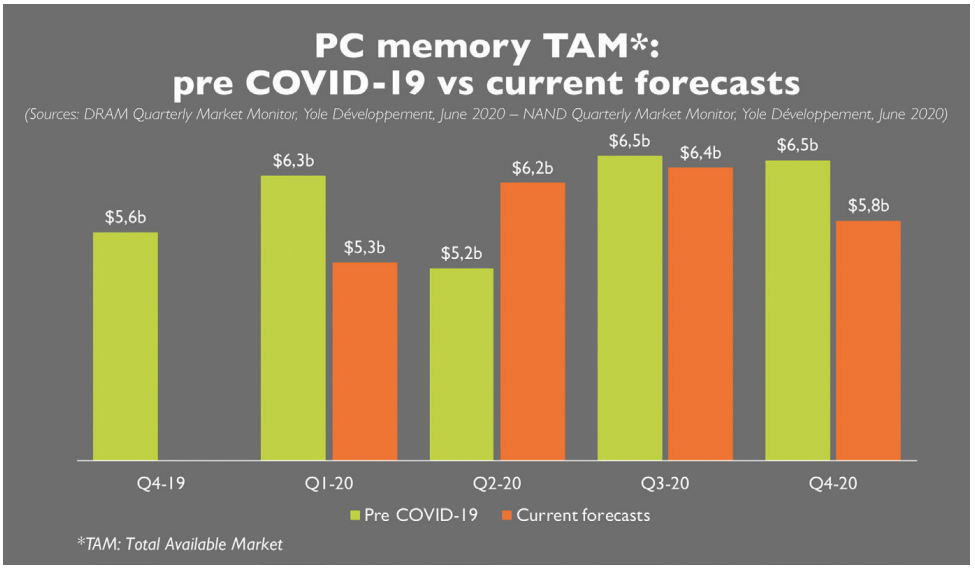


Figure 2

Smartphones

Smartphones is the category that has been the most adversely impacted by COVID. At the beginning of the year we expected 2020 to be a strong year for smartphones. With enthusiasm around 5G driving an upgrade cycle, we expected new smartphone shipments to be around 1.4 billion units and memory revenue to reach nearly $40 billion growing to $50 billion next year. Today we expect smartphone shipments to come in around 1.1 billion and memory revenue to be about 20% lower at $32 billion in 2020 before recovering in 2021 (**FIGURE 3**).

People's inability to get out and shop for a new phone during lockdown, coupled with economic uncertainty and hardship, has resulted in people holding onto old phones for longer and postponing upgrades. However, we believe that long-term the segment will bounce back. Phones do not last forever and there is not currently a substitute device for smartphones (as the smartphone was the replacement device for the PC ten years ago). Therefore, eventually old phones will need to be replaced. This will result in "catch up" demand which we expect to start seeing in 2021. We anticipate the smartphone market will perform very well next year.

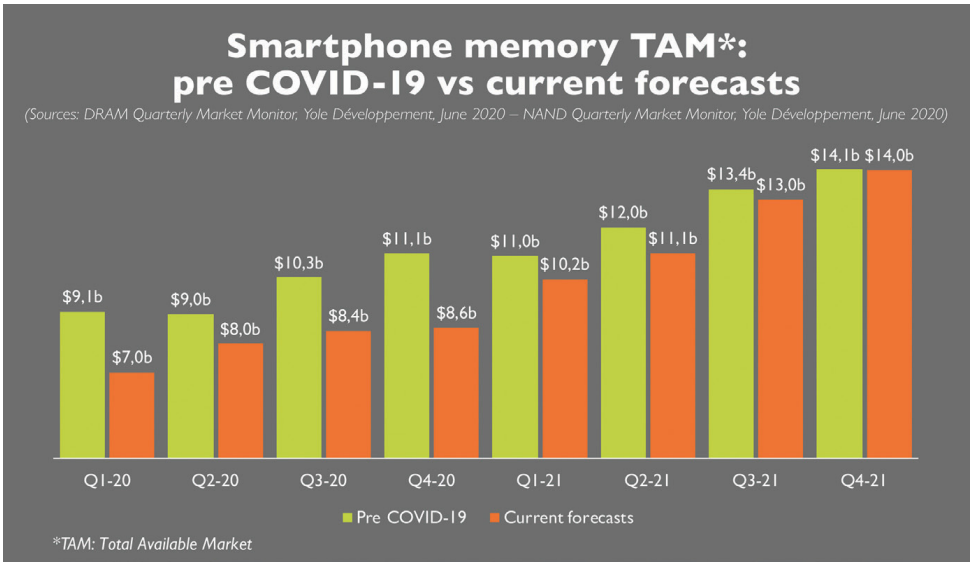


Figure 3

Memory supplier reactions

Capital expenditures Prior to the COVID-19 outbreak expectations were for combined DRAM and NAND capital expenditures (capex) of $38.8 billion in 2020, down 15% year-over-year, including both WFE and infrastructure spend. This decrease was a result of the memory market downturn that plagued 2019—and much of 2018 for NAND—as well as the timing of infrastructure build outs and technology transitions.

The current forecast assumes a much steeper 2020 capex drop-off, with combined DRAM and NAND capex of $33.8 billion, down 26% from 2019 and 13% lower than the prior outlook. It is expected that the memory suppliers will be more cautious with investments this year due to the uncertainty around second half demand and longer-term economic ramifications. The memory suppliers are likely to err on the side of caution and push spend into 2021 or potentially further depending on market conditions (**FIGURE 4**).

The ramifications of reduced capex are significant. Lower capex equates to slower technology transitions, fewer new wafer additions, diminished bit growth, and reduced cost per bit declines. Smaller cost declines may lead to lower profit margins and lower revenue TAM, depending on the impacts to memory average selling prices (ASPs). Given the uncertainty in the market, reducing capex in the near-term appears to be a prudent decision. The suppliers can hope that lower capex (and therefore lower bit shipments) will result in higher ASPs, offsetting smaller cost per bit declines.

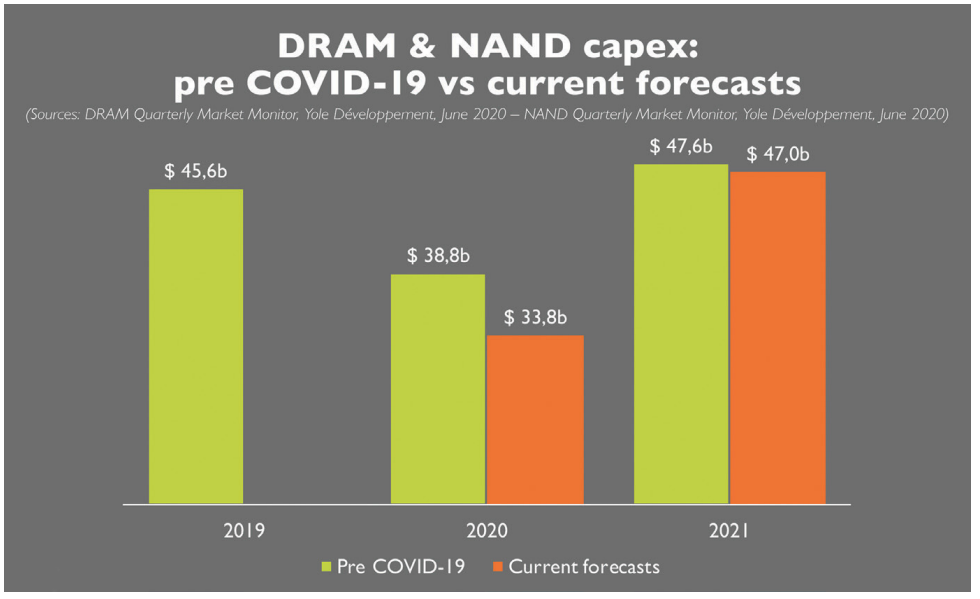


Figure 4

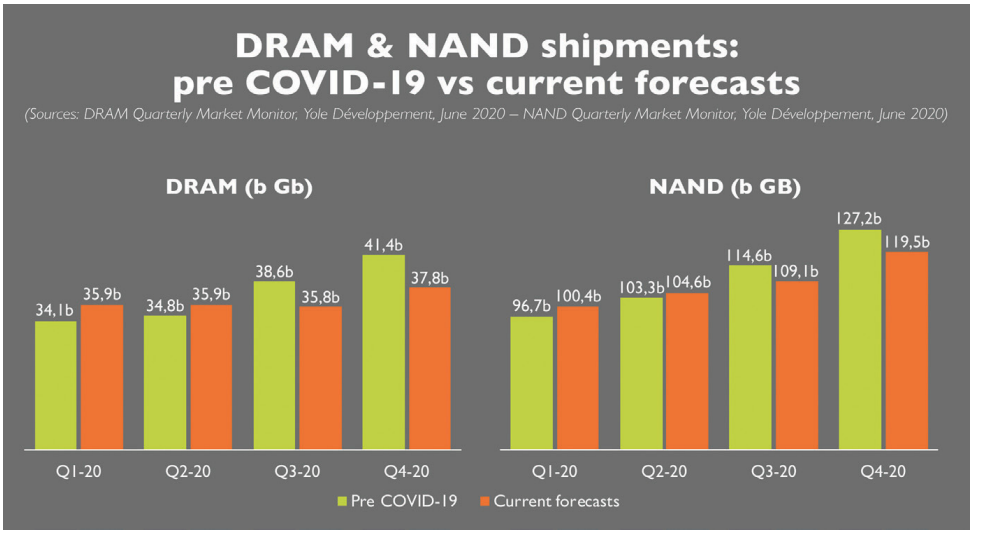


Figure 5

Memory supply The impacts from the pandemic were felt by the memory suppliers almost immediately and have been reflected in first half bit shipments for both DRAM and NAND. The widespread work- and learn-from-home transition has provided a near-term boost to the memory suppliers, leading to first half memory bit shipments higher than previously anticipated. Shipment growth has been led by strong server and PC demand and customer buy-aheads due to supply chain concerns, which have helped offset initial COVID-related weakness in the smartphone and consumer markets.

Looking ahead to the rest of the year, although datacenter demand is expected to remain resilient, we anticipate continued weakness in the smartphone and consumer markets and softening PC demand after the initial surge from the first half wanes. Demand for traditional enterprise servers is also at risk, as economic uncertainty may lead to more conservative IT spend. As a result, second half 2020 bit shipment expectations have been lowered for both DRAM and NAND (**FIGURE 5**).

The impact to NAND bit shipments is not expected to be as significant as DRAM for a several reasons. Prior to the outbreak, 2020 NAND bit growth was already expected to be constrained, with the market just emerging from a major downturn and supply impacts from previous capex cuts taking hold. Additionally, NAND has the continued benefit of the HDD-to-SSD replacement cycle in PC's, with the current PC demand surge coming from corporate buyers who overwhelmingly use SSD-based storage along with Chromebooks on the education side using NAND-based storage. Finally, the introduction

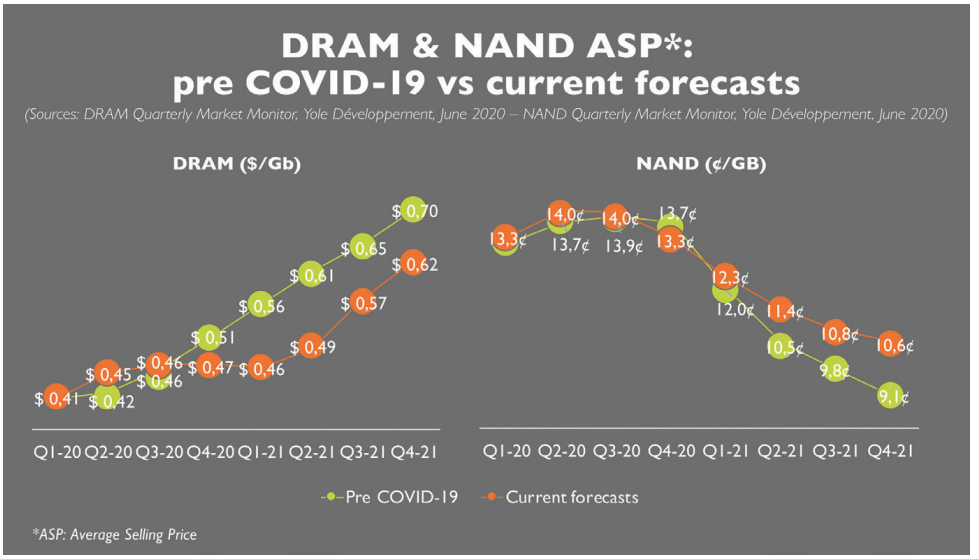


Figure 6

of new gaming consoles later this year, which are shifting from HDD to high density SSD-based storage solutions, will provide a significant boost to bit demand in the second half.

For the full year, DRAM bit growth (2020 vs. 2019) has been lowered from 17% in the prior forecast to 15%, while NAND bit growth has been lowered from 30% to 29%.

Memory pricing Strength in datacenter and PC demand has led to higher-than-anticipated pricing for both DRAM and NAND in the first half of 2020. However, weaker demand in the second half is likely to reduce pricing relative to prior expectations. The reality is that second half pricing will be largely dictated by the response of the suppliers to this pandemic. Will they adjust wafer output and technology transitions in the face of demand uncertainty, or continue along the path of their prior plans? There are several actions the supplier can take to bolster pricing in the face of demand uncertainty, including reducing capex, lowering fab utilization, and holding strategic inventory.

Strong datacenter demand has driven server DRAM pricing up ~40% since the end of 2019. Server DRAM prices, however, are not likely to rise much more in 2020 as large CSPs have ample inventory and the availability of server DRAM should be greater in the second half of 2020 as suppliers have shifted wafers from mobile DRAM to compute. The surge in H1 2020 pricing will result in overall pricing for the year being slightly better than previously expected but 2021 prices are not expected to climb as high due to the drag on demand from the anticipated economic fallout (**FIGURE 6**).

The NAND downturn that spanned most of past few years resulted in steep losses for the industry in 2019. Although industry margins moved positive in Q1 2020, there is little room for large price declines in the coming quarters. Additionally, capex reductions will impact the ability of the NAND suppliers to lower costs both this year and beyond. As a result, the full year outlook for blended NAND pricing in 2020 is mostly unchanged, with pricing expected to be up 6% year-over-year. Based on the current forecast, long term NAND industry margins are not sustainable, and the industry likely needs to see consolidation or some other structural shift to generate sufficient returns.

Conclusion

Effects from the COVID-19 pandemic on the memory markets have been immediate and dramatic and are expected to continue impacting the memory markets into the foreseeable future. Although there has been some demand upside in the near term due to changing work, education, and leisure habits, the long-term economic impacts are likely to be severe, and hamper demand in the mid- to long-term. It is imperative that the memory suppliers respond proactively and with caution given the uncertainty in the markets to ensure the long-term health of the memory industry. Yole will continue to closely monitor the entire supply chain and adjust our outlook as needed.

About the authors

Walt Coon joins Yole Développement's memory team as VP of NAND and Memory Research, part of the Semiconductor & Software division. Based in the US, Walt is leading the day-to-day production of both market updates and Market Monitors, with a focus on the NAND market and semiconductor industries.

Walt spent 16 years at Micron Technology, managing the team responsible for competitor benchmarking, and industry supply, demand, and cost modeling. His team also supported both corporate strategy and Mergers & Acquisitions analysis. Previously, he spent time in Information Systems, developing engineering applications to support memory process and yield enhancement.

Mike Howard is a member of the memory team at Yole Développement (Yole) as VP of DRAM and Memory Research. Mike's mission at Yole is to deliver a comprehensive understanding of the entire memory and semiconductor landscape (with special emphasis on DRAM) via market updates and Market Monitors. Mike is also deeply involved in the business development of all memory activities. Mike is based in the US.

For the decade prior to joining Yole, Mike was the Senior Director of DRAM and Memory Research at IHS. Before IHS, Mike worked at Micron Technology where he had roles in corporate development, marketing, and engineering.

Five Trends That Will Shape the Future Semiconductor Technology Landscape

SRI SAMAVEDAM, senior vice president of CMOS technologies at imec, Leuven, Belgium

Is Moore's Law still alive, and which applications will benefit from ultra-scaled technologies? How will data centers cope with the overwhelming amount of data? And will we be able to break the memory wall in traditional Von Neumann computing architectures?

In the past few decades, growth of the global semiconductor industry has been driven largely by the demand for cutting-edge electronic devices such as desktops, laptops and wireless communication products, and by the rise of cloud-based computing. Growth will continue with new application drivers for the high-performance compute market segment.

First, the amount of data keeps on growing exponentially, a trend that will be accelerated by the rollout of 5G networks. We need more and more servers where these data are processed and stored. Following a 2020 Yole report [1], a compound annual growth rate of 29% is expected for the high-end central processing units (CPUs) and graphical processing units (GPUs) that are at the heart of these servers. They will support a host of datacenter applications, such as supercomputing and high-performance-computing as a service. Faster growth rate is expected for GPUs – triggered by emerging applications such as cloud gaming and artificial intelligence. Recent corona-related remote work and education will also leave their mark on the internet traffic. In March 2020, for example, internet traffic increased by nearly 50% and commercial internet data exchange in Frankfurt set a new world record [2] for data throughput at more than 9.1 Terabits per second.

A second main driver is mobile systems-on-chips (SoCs) – the chips in our smart phones. This market segment is not growing as fast, but the demand for more functionality in these SoCs in form-factor constrained chip area will drive further technology innovations.

Beyond the traditional dimensional scaling of logic, memory and 3D interconnects, these emerging applications will need to leverage cross-domain innovations. There is a need for new modules, new materials and architecture changes at the device, block and SoC level to realize the benefit at a system level. Below, these innovations are summarized in five major semiconductor technology trends.

Trend 1: Moore's Law will continue for the next 8 to 10 years...

CMOS transistor density scaling will roughly continue to follow Moore's Law for the next eight to ten years. This will be enabled mainly by advances in EUV patterning and by the introduction of novel device architectures which will enable logic standard cell scaling.

Extreme ultraviolet (EUV) lithography was introduced in the 7nm technology node to pattern some of the most critical chip structures in one single exposure step. Beyond the 5nm technology node (i.e., when

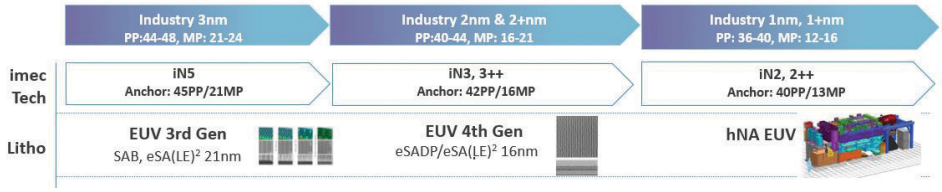


Figure 1. An imec view on the EUV lithography roadmap (PP=poly pitch; MP=metal pitch; SAB=self aligned blocks; eSALELE=innovative approach to self-aligned litho-etch litho-etch; SADP=self-aligned double patterning).

critical back-end-of-line (BEOL) metal pitches are below 28-30nm), multi-patterning EUV lithography becomes inevitable – adding significantly to the wafer cost. Eventually, we expect high-numerical-aperture (high-NA) EUV lithography to become available for patterning the most critical layers of the industry's 1nm node. This technique will push the multi-patterning of some of these layers back to single patterning, providing cost, yield and cycle-time relief (**FIGURE 1**).

Imec contributes to advancing EUV lithography for example by investigating stochastic defectivity. Stochastic printing failures are random, non-repeating, isolated defects such as microbridges, locally broken lines and missing or merged contacts. Improvement in stochastic defectivity could lead to the use of lower dose exposures and thus improve throughput and cost. We try to understand, detect and mitigate stochastic failures and could recently report an order of magnitude improvement in stochastic defectivity.

To accelerate the introduction of high-NA EUV, we are installing Attolab – allowing to test some of the critical materials for high-NA EUV (such as mask absorber layers and resists) before the high-NA tool will be available. The spectroscopic characterization tools in this lab will allow us to look at crucial EUV-photon reactions with resists at attosecond timeframes, which are also relevant to understand and mitigate stochastic defect formation. Today, we have successfully completed phase one of the Attolab installation and expect to have high-NA EUV exposures in the coming months.

Apart from advancements in EUV lithography, Moore's Law cannot continue without innovations in the front-end-of-line (FEOL) device architecture (**FIGURE 2**). Today, FinFET devices are the mainstream transistor architectures, with the most advanced nodes having 2 fins in a 6-track (6T) standard cell. However, scaling down FinFETs to 5T standard cells results in fin depopulation with only 1 fin per device in the standard cell, causing a dramatic drop in the device performance per unit area. Vertically stacked nanosheet devices are considered as the next generation device, being a more efficient use of the device footprint. Another critical scaling booster is the buried power rail (BPR). Buried in the chip's FEOL instead of in the BEOL, these BPRs will free up interconnect resources for routing.

Scaling nanosheets into the 2nm generation will be limited by n-to-p space constraint. Imec envisions the forksheet architecture as the next generation device. By defining the n-to-p space with a dielectric wall, the track height can be further scaled. Another standard cell architecture evolution that will help with routing efficiency is a vertical-horizontal-vertical (VHV) design for metal lines, as opposed to traditional HVH designs. Ultimate standard cell scaling down to 4T will be enabled by complementary FETs (CFETs), that fully exploit the third dimension at the cell level by folding n-FETs over p-FETs or vice-versa.

Trend 2: ... but logic performance improvement at fixed power will slow down

With the above-mentioned innovations, we expect transistor density to follow the path mapped out by Gordon Moore. But node-to-node performance improvements at fixed power – referred to as Dennard scaling – have slowed down due to the inability to scale supply voltage. Researchers worldwide are looking for ways to compensate for this slow-down and further improve the chip's performance. The aforementioned buried power rails are expected to offer a performance boost at system level due to improved power distribution. Besides, imec looks at incorporating stress into nanosheet and forksheet devices, and at improving the contact resistance in the middle-of-line (MOL). Further out, the sequential CFET device will provide the flexibility for incorporating high mobility materials since the n-device and p-device can be optimized independently.

2D materials such as tungsten disulfide (WS_2) in the channel promise performance improvements because they enable more aggressive gate length scaling than Si or SiGe. A promising 2D-based device architecture involves multiple stacked sheets each surrounded by a gate stack and contacted from the side.

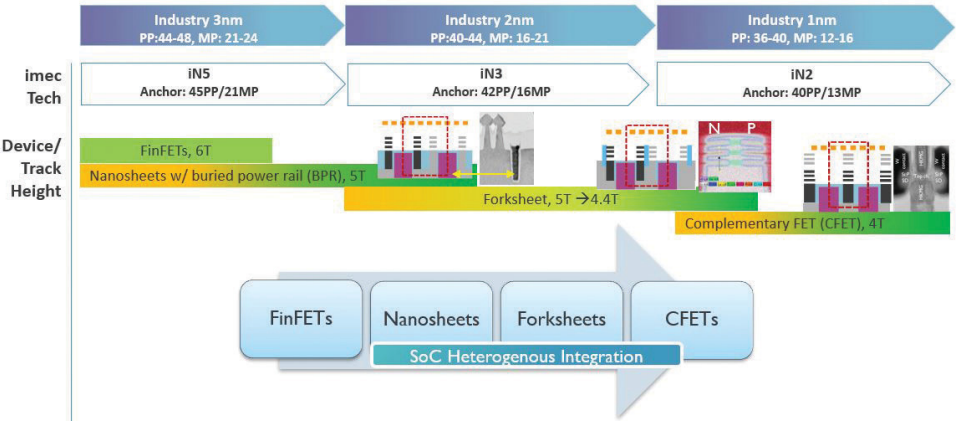


Figure 2. An imec view on the FEOL roadmap.

Simulations suggest these devices can out-perform nanosheets at scaled dimensions targeting 1nm node or beyond. At imec, dual-gate transistors with bilayer WS_2 on 300mm wafers have already been demonstrated, with gate lengths down to 17nm. To further improve the drive current of these devices, we strongly focus on improving the channel growth quality, incorporating dopants and improving contact resistance in these novel materials. We try to speed up the learning cycles for these devices by correlating physical properties (such as growth quality) with electrical properties.

Besides the FEOL, routing congestion and RC delay in the BEOL have become important bottlenecks for performance improvement (**FIGURE 3**). To improve the via resistance, we are looking at hybrid metallization using Ru or Mo. We expect semi-damascene metallization modules to simultaneously improve resistance and capacitance in the tightest pitch metal layers. Semi-damascene will allow us to increase the aspect ratio of the metal lines (to lower resistance) by direct patterning and use airgaps as a dielectric in between the lines (to control the capacitance increase). At the same time, we screen a variety of alternative conductors like binary alloys as a replacement for 'good old' Cu, to further reduce the line resistance.

Trend 3: More heterogeneous integration, enabled by 3D technologies

In industry, we see more and more examples of systems being built through heterogeneous integration leveraging 2.5D or 3D connectivity. These options help address the memory wall, add functionality in form-factor constrained systems or improve yields on large chip systems. With the slowing logic PPAC (performance-power-area-cost), smart functional partitioning of SoC (system on chip) can provide another knob for scaling. A typical example is high-bandwidth memory (HBM) stacks, consisting of stacked dynamic random access memory (DRAM) chips that connect directly through a short interposer link to processor chip, such as a GPU or CPU. More recent examples include die-on-die stacking in Intel's Lakefield CPU or chiplets on interposer in case of AMD's 7nm Epyc CPU. In the future, we expect to see many more of these heterogeneous SoCs – as an attractive way to improve system performance.

At imec, we bring about the benefits at SoC level by leveraging innovations that we are making across the different domains (such as logic, memory, 3D...). In order to connect the technology options to the performance at system level, we have set up a framework called S-EAT (System benchmarking for Enablement of Advanced Technologies). This framework allows us to evaluate the impact of specific technology choices on the system level performance. For example: can we benefit from 3D-partitioning the on-chip memory at lower levels of the cache hierarchy? And what happens at the system level if static random access memory (SRAM) is replaced by a magnetic RAM (MRAM) memory?

As an illustration, we have used this platform to find the most optimal partitioning of a high performance mobile SoC containing a CPU and L1, L2 and L3 caches. In a traditional design, the CPU would reside next to the caches in a planar configuration. We assessed the impact of moving the caches to another chip, stacked with 3D wafer bonding techniques to the CPU chip. As the signals between cache and CPU now travel shorter distances, an improvement in speed and latency can be expected. The simulation experiments concluded that it was most optimal to move L2 and L3 caches to the top tier instead of L1 only or all 3 caches simultaneously.

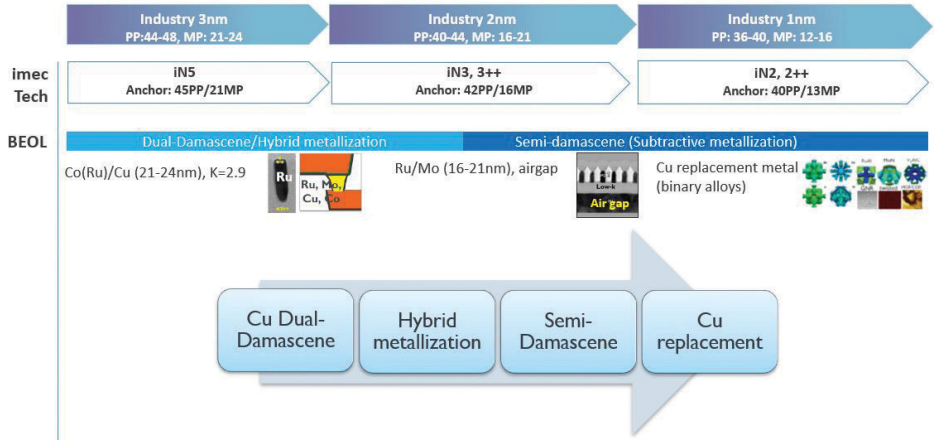


Figure 3. An imec view on the BEOL roadmap.

To enable partitioning at these deeper levels of the cache hierarchy, a high-density wafer-to-wafer stacking technology is required. We have demonstrated wafer-to-wafer hybrid bonding at 700nm interconnect pitch and believe the advancements in bonding technology will enable 500nm pitch interconnects in the near future.

Heterogeneous integration is enabled by 3D integration technologies such as die-to-die or die-to-Si-interposer stacking using Sn microbumps or die-to-silicon using hybrid Cu bonding. The state-of-the-art Sn microbump pitches in production

have saturated at about 30µm. At imec, we are pushing the boundaries of what is possible today. We have demonstrated a Sn-based microbump interconnect approach with interconnect pitch down to 7µm. Such high-density connections leverage the full potential of through-Si via technology and enable >16x higher 3D interconnect densities between die or between dies and a Si-interposer. This allows for a strongly reduced SoC area requirement for the HBM I/O interface (from 6 down to 1 mm^2) and potentially shortens the interconnect lengths to the HBM memory stack by up to 1 mm. Direct bonding of die to silicon is also possible using hybrid Cu bonding. We are developing die-to-wafer hybrid bonding down to 3µm pitches with high tolerance pick and place accuracy, leveraging the learning from wafer to wafer hybrid bonding.

As SoCs are becoming increasingly more heterogeneous, the different functions on a chip (logic, memory, I/O interfaces, analog, ...) need not come from a single CMOS technology. It may be more advantageous to use different process technologies for different sub-systems to optimize design costs and yield. This evolution can also answer the needs for more chip diversification and customization.

Trend 4: NAND and DRAM being pushed to their limits; emerging non-volatile memories on the rise

The total memory IC market forecast [3] suggests that 2020 will be a flat year for memory relative to 2019 – an evolution that can partly be related to the COVID-19 slowdown. Beyond 2021, this market is expected to start growing again. The emerging non-volatile memory market [4] is expected to grow at >50% compound annual growth rate – mainly driven by the demand for embedded magnetic random access memory (MRAM) and standalone phase change memory (PCM).

NAND storage will continue to scale incrementally, without disruptive architectural changes in the next few years. Today's most advanced NAND products feature 128 layers of storage capability. The 3D scaling will continue with additional layers potentially enabled by wafer-to-wafer bonding. Imec contributes to this roadmap by developing low resistance word-line metals like ruthenium, researching alternate memory dielectric stacks, improving channel current and identifying ways to control the stress that evolves due to the growing number of stacked layers. We also focus on replacing the planar logic transistors in the NAND periphery with more advanced FinFET devices. We are exploring 3D ferroelectric FETs (FeFETs) with novel wurtzite materials as 3D NAND replacement in high-end storage applications. As a replacement for traditional 3D NAND, we are evaluating the feasibility of novel types of memories.

For DRAM, cell scaling is slowing down, and EUV lithography may be needed to improve patterning. Samsung recently announced EUV DRAM production for their 10nm (1a) class. Besides exploring EUV lithography for patterning critical DRAM structures, imec provides the building blocks for true 3D DRAM solutions. And this starts with putting the memory array on top of the periphery. Such an architecture requires a low thermal budget deposited semiconductor for the array transistors. And this is where low temperature IGZO (or indium-gallium-zinc-oxide) family of transistors enter the scene. We have demonstrated 40nm gate length IGZO devices with Ion/Ioff ratio >1E12. And we continue to explore alternate low temperature semiconductors using ab-initio simulations and experiments to meet the stability, mobility and reliability requirements. Ultimate 3D DRAM implementation will also require these materials to be deposited on topography. This drives the need for atomic layer deposition (ALD) for the layer formation. Finally, just as with NAND, we look at enabling a FinFET-based periphery with high-k/metal gate structures as a replacement for planar transistors with poly-Si gates.

In the embedded memory landscape, there are significant efforts to understand – and eventually tear down – the so-called memory wall:

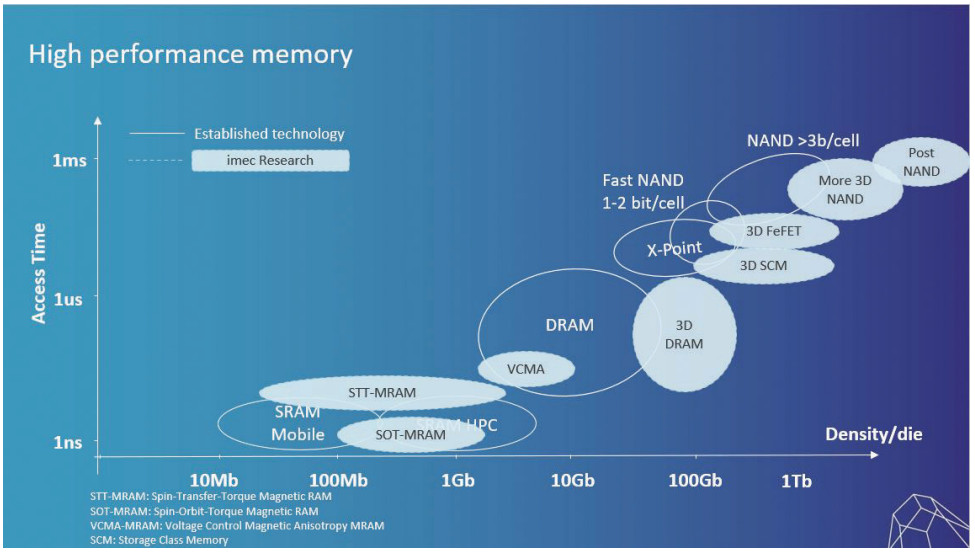


Figure 4. A view on the memory landscape and how imec contributes.

how quickly can the CPU access data from DRAM or from SRAM-based caches? How do you ensure cache coherency with multiple CPU cores accessing a shared cache? What are the bottlenecks that limit speed, and how can we improve the bandwidth and data protocols that are used to fetch the data? Imec deploys its system-level simulator platform S-EAT to gain insights into these bottlenecks. This framework also allows for evaluation of novel memories as SRAM replacements to understand the system performance for various workloads. We are studying various kinds of magnetic random access memories (MRAM) including spin-transfer torque (STT)-MRAM, spin-orbit torque (SOT)-MRAM and voltage controlled magnetic anisotropy (VCMA)-MRAM) to potentially replace some of the traditional L1, L2 and L3 SRAM-based caches (**FIGURE 4**). Each of these MRAM memories comes with its own benefits and challenges and may help us overcome the memory bottleneck by improving speed, power consumption and/or memory density. To further improve density we are also actively researching selector devices that can be integrated with the magnetic tunnel junctions – which are at the heart of these MRAM devices.

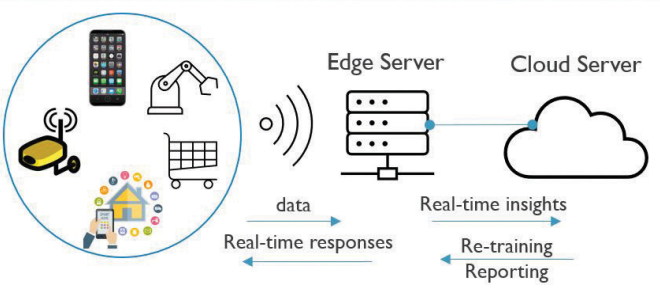


Figure 5. The edge AI framework

Trend 5: Spectacular rise of the edge AI chip industry

With an expected growth of above 100% in the next five years, edge AI is one of the biggest trends in the chip industry. As opposed to cloud-based AI, inference functions are embedded locally on the Internet of Things (IoT) endpoints that reside at the edge of the network, such as cell phones and smart speakers. The IoT devices communicate wirelessly with an edge server that is located relatively close. This server decides what data will be sent to the cloud server (typically, data needed for less time-sensitive tasks, such as re-training) and what data gets processed on the edge server.

Compared to cloud-based AI, in which data needs to move back and forth from the endpoints to the cloud server, edge AI addresses privacy concerns more easily (**FIGURE 5**). It also offers advantages of response speeds and reduced cloud server workloads. Just imagine an autonomous car that needs to make decisions based on AI. As decisions need to be made very quickly, the system cannot wait for data to travel to the server and back. Due to the power constraints typically imposed by battery-powered IoT devices, the inference engines in these IoT devices also need to be very energy efficient.

Today, commercially available edge AI chips – the chips inside the edge servers – offer efficiencies in the order of 1-100 tera operations per second per Watt (Tops/W), using fast GPUs or ASICs for computation. For IoT implementations, much higher efficiencies will be needed. Imec's goal is to demonstrate efficiencies for inference in the order of 10.000 Tops/W.

We are pursuing a different approach by looking at analog compute-in-memory architectures. This approach breaks with the traditional Von Neumann computing paradigm, that is based on sending data from memory to a CPU (or GPU) for computations. With analog compute-in-memory, computation is done inside a memory framework, saving a lot of power in moving data back and forth. In 2019, we demonstrated an SRAM-based analog compute-in-memory cell (built in 22nm FD-SOI technology) achieving 1000Tops/W efficiency. To further improve this number towards 10.000Tops/W, we are researching non-volatile memories such as SOT-MRAM, FeFET and IGZO-based memories. sD

About the author

Sri Samavedam is senior vice president of CMOS technologies at imec since August 2019. His responsibilities include programs in logic, memory, photonics and 3D integration. Prior to that, he was senior director of technology development at GlobalFoundries in Malta, NY, where he led qualification of 14nm FinFET technology and derivatives into volume production and early development of 7nm CMOS. He began his research career at Motorola in Austin, TX, working on strained silicon, metal gates, high k dielectrics and fully-depleted SOI devices. He holds a Ph.D. in Materials Science and Engineering from MIT and a masters from Purdue University.

REFERENCES

1. High-end CPU and GPU: HPC and cloud gaming are setting the bar for leadership http://www.yole.fr/xPU_HighEnd_CPU_GPU.aspx
2. DE-CIX sets a new world record: More than 9 Terabits per second data throughput at Frankfurt Internet Exchange https://www.de-cix.net/
3. Memory Market Not Forecast to Exceed 2018 High of $163.3B until 2022. IC Insights, April 28, 2020. https://www.icinsights.com/
4. After the 3D XPoint take-off, emerging NVM keeps growing. http://www.yole.fr/ENVM_IndustryUpdate.aspx
5. A view on the memory landscape and how imec contributes.

When Device Failure Is Not an Option: Reducing Latent Defects Through Proper ESD Management

MARK CAULFIELD and **BRETT REICHOW**, Entegris, Billerica, MA

A new solution, designed to remove electrostatic charge, implements carbon stripes on the inside of the tubing, as well as in fittings, valves, and other components.

EMERGING TECHNOLOGY BREAKTHROUGHS like artificial intelligence (AI), machine learning (ML), and 5G are driving the semiconductor industry to manufacture ever-finer advanced technology nodes down to 5 and 3 nm ranges, as well as implement vertical stacking and heterogeneous integration to achieve higher interconnect densities. At the same time, these applications are driving a need for higher-reliability devices as the nature of end-use markets shifts from PCs and smartphones to autonomous vehicles, medical devices, and the industrial internet of things (IIoT), where device failure is not an option.

One way to improve device reliability is to prevent defects from forming on the silicon wafer. Wafer defects come in two forms: killer defects and latent defects. Killer defects impact yield, because they are caught before the die is put into a system. Latent defects impact reliability, because they are not easily detected and fail in the field. Fifty percent of device failures are related to latent defects that do not measure out at the end of the line [1]. While managing final yields by reducing killer defects has always been critical for a semiconductor manufacturer's bottom line, eliminating latent defects could now be a matter of life and death.

Tighter process requirements

The combination of shrinking feature sizes and increased reliability is impacting semiconductor manufacturing process requirements. Reducing the possibility of wafer defects increases exponentially with each successive node. To maximize yield and maintain the reliability of electronic devices, fabs must remove smaller and smaller particles along with the metals that can contaminate wafer surfaces and interconnects. Figure 1 shows the range of defects that can result from contaminants in the chemical materials used in the chipmaking process [2].

To help prevent these defects, manufacturing in a pure environment is more critical now than ever. Purity requirements extend beyond just the cleanroom, and into the chemical and gas delivery systems, as well as the chemistries themselves.

For example, photolithography and wet etch and clean processes have become more metal sensitive at advanced nodes. As a result, the industry shifted away from stainless steel fluid handling systems that cause metals from stainless steel components to be extracted into the process chemicals and cause wafer defects. Instead, they transitioned to perfluoroalkoxy (PFA), a common fluoropolymer. While this successfully reduced the presence of extracted metals, the increased use of fluoropolymer systems creates new concerns with electrostatic discharge (ESD) in chemical transport through PFA tubing.

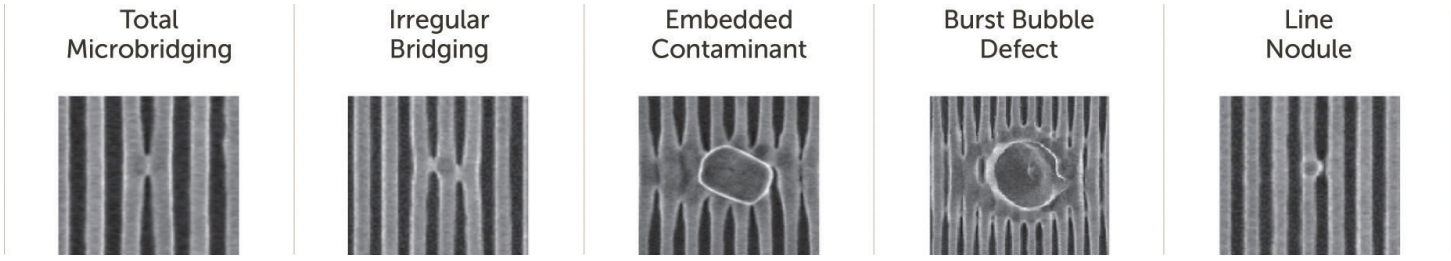


Figure 1. Photolithography wafer defect type examples.

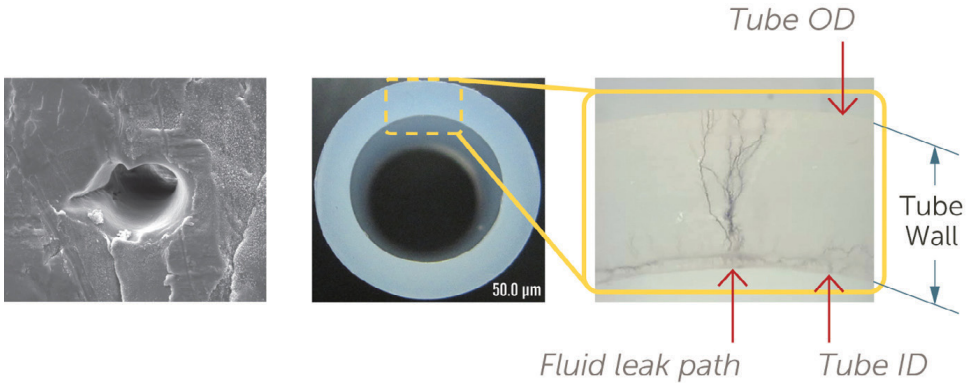


Figure 2. Left: Magnification of pinhole in a valve diaphragm. Center/right: Example of electrical discharge through a standard PFA tubing wall (0.062" diameter wall thickness).

Solvents used during clean processes have low conductivity, which enables them to generate and hold an electrical charge. When these solvents are transported in PFA systems, there is a greater risk of static charge generation and discharge due to the nonconductive nature of the PFA materials and the low conductivity properties of the solvents [3].

ESD causes latent defects

Depending on the electrical potential of the process chemistries passing through the tubing, ESD concerns range from causing industrial hazards, to shortening the life of critical process tools and components in a semiconductor fab, to damaging the semiconductor devices themselves.

Propagating ESD can be powerful enough to cause "pinhole" damage to the components, causing leaks that contaminate the process. Furthermore, ESD generated in fluoropolymer systems that are transferring flammable solvents can create leak paths through the tubing that could possibly ignite the surrounding, potentially flammable solvent-rich environment (FIGURE 2).

And lastly, when electrically charged fluid is dispensed onto a wafer, the charge randomly dissipates on the wafer, causing both killer and latent defects. Additionally, in a cleanroom environment, static buildup can attract contaminants or cause products to adhere to one another, which is also a potential source of killer and latent defects [4].

Finding the right solution

Here, we describe a number of approaches to mitigating ESD hazards that have been tried with varied results. Some approaches successfully addressed the more catastrophic issues, such as avoiding industrial disasters and preventing total tool failure. Unfortunately, preventing ESD-induced wafer defects continues to be a challenge for these methods.

Replacing PFA with Conductive Tubing and Grounding Strap: One solution involves replacing PFA tubing with tubing that uses a conductive carbon stripe on the outside of the tube only. When static charges build up on the outside of the electrostatic dissipative tube, the carbon-filled PFA conductive stripe allows dissipation of the charge to ground through a grounding strap that attaches to the tube.

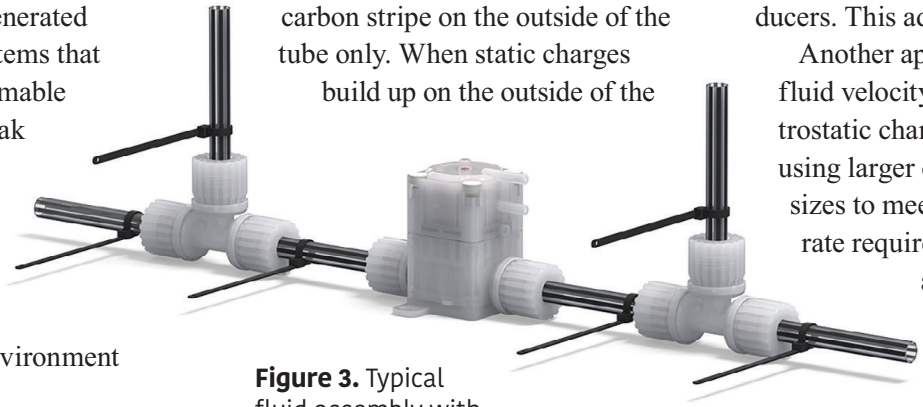


Figure 3. Typical fluid assembly with discontinuous charge dissipation path across components.

This approach addresses some of the electrostatic hazard challenges, but it is limited in usefulness. It does not prevent on-wafer damage caused by electrostatic charge accumulation on the flowing liquid itself. It is the electrostatic charge accumulation on the liquid that may cause damage to the wafer structures and impact device yield.

Initially, manufacturers adopted this solution despite its limitations because it was an improvement. However, when they realized ESD associated with charge accumulation on the nonconductive media were damaging wafers, they wanted a solution that went further, taking the charge out of the media itself (FIGURE 3).

Alternative Approaches that Failed: Several attempts have been made to mitigate electrostatic hazards in liquids by altering the chemistry itself, but have proven to be ineffective; for example, adding CO_2 into deionized (DI) water to minimize electrostatic charge. While effective for DI water, this is not an effective solution for other chemistries, because it changes the chemical composition and increases contaminant potential. Dissolving gases into the fluid, like CO_2 which adds bubbles, can also affect the reliability and accuracy of downstream subcomponents such as flow controllers and pressure transducers. This adds cost and downtime.

Another approach, reducing the fluid velocity to minimize electrostatic charge generation, means using larger orifices and tubing sizes to meet the volumetric flow rate requirements. This approach adds unacceptable cost and footprint, which contradicts the general desire of all end-users.

Conductive ESD Probes: Inserting

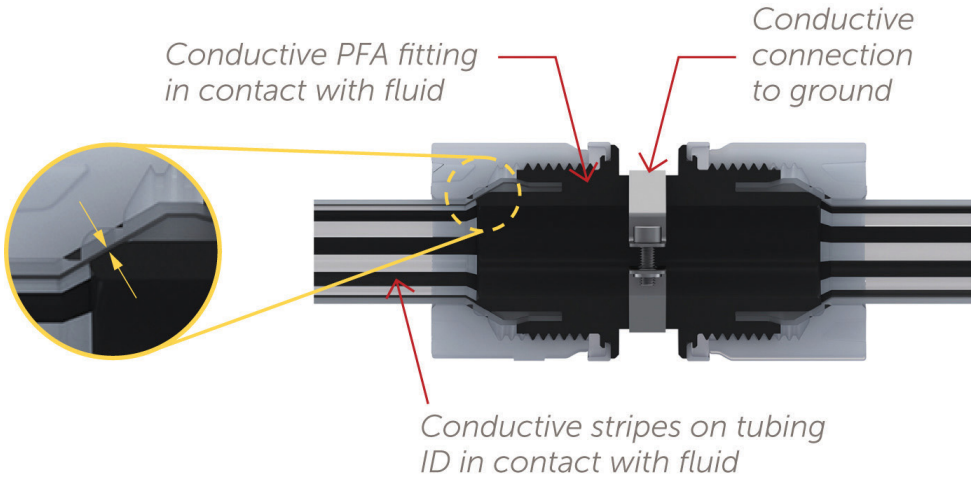


Figure 4. Conductive tubing with stripes on ID and carbon-loaded fitting to facilitate ground.

ESD probes to help mitigate the generation of electrical potential was also met with limited success in an application involving a back end of the line (BEOL) wet etch and clean (WEC) tool for wafer cleaning.

DI/CO_2 (carbonic acid) is generally used for final wafer rinse to leave a charge-free surface for the next process step. However, it does not address ESD that can occur in-situ during lower conductive chemical dispenses such as dilute hydrogen fluoride (HF). This is because the subfab-based recirculation equipment of WEC tools require several meters of PFA tubing. ESD probes can only provide a single point contact on either side of this secondary contained tubing. Unfortunately, as with the conductive tubing and grounding strap approach, it still did not fully eliminate periodic ESD events that caused arcing to the wafer during the chemical dispense, resulting in significant gate defects on the wafer.

To eliminate these periodic events, it is necessary to minimize the creation of electrical potential more consistently throughout a tool's fluid distribution by eliminating electrostatic charge accumulation on the flowing liquid itself. It is the electrostatic charge accumulation on the liquid that may cause damage to the wafer structures. Managing the generation of electrical potential throughout the entire system is therefore required.

A conductive system in the wetted path

A promising solution to mitigate electrostatic charge hazards in PFA fluid handling systems uses tubing with a continuous conductive element from the interior of the tubing to the exterior.

The challenge with a conductive stripe on the ID (inside diameter) of the tubing is connecting the stripe to ground. Because the stripe is not exposed like in the OD (outside diameter) striped application, the issue of connecting to ground required a different approach. This solution allows for some charge dissipation of the fluid, but the continuity is broken at fittings, valves, or other components in the flow stream. This results in significant, complex wiring to bring all the separate tube sections to ground.

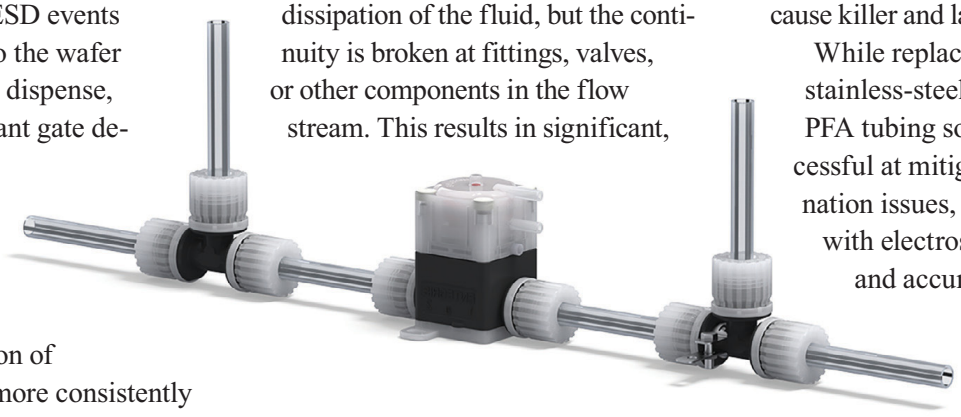


Figure 5. Fluid assembly continuously protected from charge dissipation across all components, inside and out.

A continuous conductive system

Back at the drawing board, a fully conductive system, which includes conductive tubing, fittings, and valves, was designed from a new material. The tubing has carbon stripes along the inside, and the fittings are carbon-loaded to create a path to ground when the sealing area contacts the conductive stripe (FIGURE 4).

This system has proven to successfully reduce the accumulated electrostatic charge from both the media flowing through the tubing as well as any charge that might build up inside the tube, mitigating safety risks, 5. Significant metal extraction testing has been performed as well, which shows little to no change in extraction compared to neat PFA tubing.

Summary and conclusion

As the semiconductor industry continues down the advanced-technology-node, high-density interconnect path, device reliability is also becoming a critical concern. End-use applications, particularly those involving the health and safety of citizens, are moving to zero defect initiatives. This is impacting the fab infrastructure itself, as manufacturers seek to find chemical handling and delivery solutions that eliminate both metal contamination and ESD issues that can cause killer and latent wafer defects.

While replacing industry-standard stainless-steel delivery systems with PFA tubing solutions has been successful at mitigating metal contamination issues, it has created problems with electrostatic charge generation and accumulation in the tubing and chemical flow stream. Initial solutions that use carbon stripes only on the tubing outside diameter along with Continued on page 65

Powering Ahead

Integrated RF power and impedance matching, solid-state tuning networks, non-sinusoidal bias for customized ion energy distribution, and predictive power delivery will enable process applications at advanced technology nodes.

PETER GILLESPIE, VP & GM, Semiconductor Products, Advanced Energy Industries, Inc., (AE) Milpitas, CA; **DAN CARTER**, Member of Technical Staff II, AE, Fort Collins, CO; **DENIS SHAW**, PhD, VP Semiconductor Integrated Solutions, AE, Fort Collins CO; **ISABEL YANG**, PhD, Chief Technology Officer, AE, Denver, CO

THIS IS THE THIRD ARTICLE IN A SERIES highlighting the importance and transformation of process power in semiconductor manufacturing applications. The first article, "Process Power Steps Out from the Shadows," uncovered the high-growth $1.2B process power market and how it has become a critical enabler for etch, deposition, and other processes such as ion implantation and e-beam inspection.

The second article, "Process Power: The New Lithography," dove deep into how RF power delivery systems have innovated and been evolved as they were tapped to enable etch and deposition processes to "draw-in" device patterns between photolithography processing steps, especially in 3D memory devices and complex logic features. This underscored how process power has changed from formerly being viewed as a simple "dumb black box" to now being seen as the "new lithography." The second article also revealed how RF generators and matching networks have become complex, highly synchronized systems with sophisticated controls and system-level power combinations including multi-frequency feeds, pulsing and micro-second response times.

This final article looks at the future of semiconductor manufacturing processing requirements and the new approaches that enable process applications at advanced technology nodes. It is clear that processes will require even faster speeds of response, additional control knobs, more direct ion energy distribution discrimination, the ability to transform available power data into actionable information, and, as always, remain under relentless pressure for lowest cost and smallest size. Solutions to these challenges will be discussed including integrated RF and tuning networks, solid-state tuning networks, non-sinusoidal bias for tailored ion energy distribution, and predictive power delivery.

A bigger process power toolbox

The semiconductor manufacturing industry has consistently moved forward via both incremental progress (improving and evolving what has worked) and innovation (new capabilities and disruptive approaches). Manufacturing devices near and below 7 nm technology nodes bring new challenges that require novel process power innovations and, at the very least, a bigger toolbox of process power capabilities.

The need for greater process power agility and response speed is driven by many shorter process steps, within longer process recipes. The power level is critical during every microsecond in deposition to control film stress and in etch to manage the plasma through frequent, abrupt, and profound changes in flow, pressure, power and chemistry that wildly swing plasma impedance. This requires repeatable and reproducible power delivery.

As 3D device stack layer counts have gone from tens to hundreds and features get ever deeper and narrower, the end of the horizon for simply increasing RF power is coming into focus. Very high RF bias powers (increasing to 50 kW and beyond to potentially as high as 100 kW) are likely not sustainable for clean power delivery and the tight control required; and additionally, the potential for damage and premature chamber wear is significant. Aside from these risks, the power required for a six or ten chamber system with 100 kW of RF power on each chamber raises the stakes not just in complexity, but also for bulk power distribution, platform costs, and even environmental concerns for both system and fab. This is an example where innovation is required vs. incremental scaling.

Significant and emerging enablers in the process power toolbox are

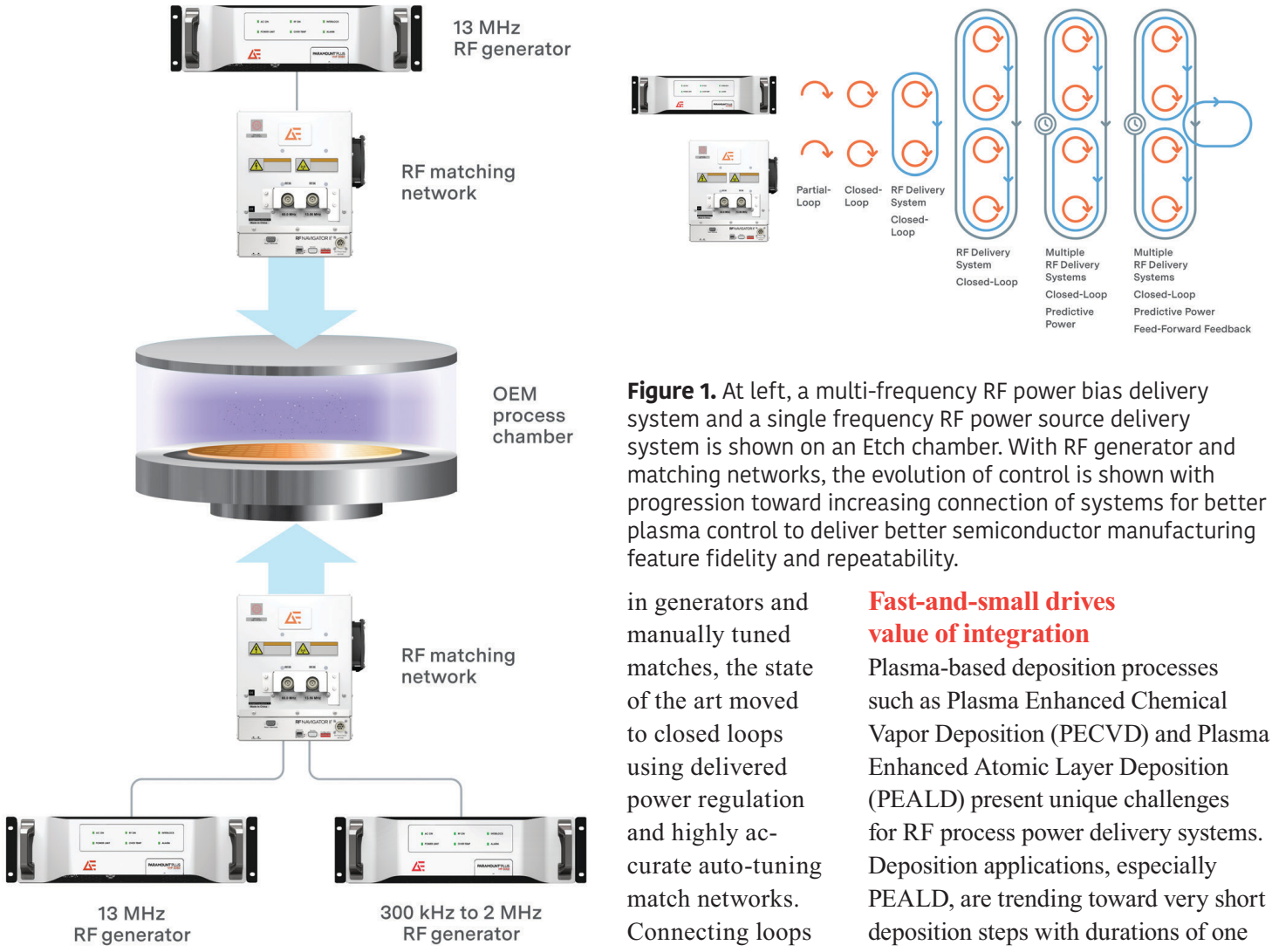


Figure 1. At left, a multi-frequency RF power bias delivery system and a single frequency RF power source delivery system is shown on an Etch chamber. With RF generator and matching networks, the evolution of control is shown with progression toward increasing connection of systems for better plasma control to deliver better semiconductor manufacturing feature fidelity and repeatability.

precision measurement and control of power, and with these, the acquisition and use of power control parameter information. As process power has become more complex, a rich array of available information has been amassed from onboard power metrology, control loops, and tuning network states. This information, parsed through the domain of power expertise, provides insight that can be leveraged to learn, optimize, and predict power parameters relevant to the process performance itself. This adds another whole dimension to the process power toolbox.

FIGURE 1 illustrates the evolution of process power control. Starting with simple command-and-response operation and open loop controls, such as forward power regulation in generators and manually tuned matches, the state of the art moved to closed loops using delivered power regulation and highly accurate auto-tuning match networks. Connecting loops to synchronize process power delivery at both source and bias enables pulsing and complex plasma control through sometimes only marginally stable plasma transitions. Predictive power machine learning can be added at the product, system, or multi-system level. Ultimately, there will be further linkage between process power information and broader system parameters for feed-forward and feedback.

Beyond process power capability and the use of control loops and predictive information, the increased cost and size of incremental solutions will also become showstoppers without implementing different approaches. Put together, these are reasons to look beyond the incremental and reach into a bigger process power toolbox that is described in this article.

Fast-and-small drives value of integration

Plasma-based deposition processes such as Plasma Enhanced Chemical Vapor Deposition (PECVD) and Plasma Enhanced Atomic Layer Deposition (PEALD) present unique challenges for RF process power delivery systems. Deposition applications, especially PEALD, are trending toward very short deposition steps with durations of one second or less. In addition, semiconductor manufacturing equipment (tools) used for these applications are often designed to process multiple wafers simultaneously while sharing a common vacuum (sub-atmospheric) management system.

These complex tools are tightly packed with pumps, valves, and other Critical Subsystems (CSubs) and hardware, limiting the available space for RF power systems. Deposition tools are also under arguably the tightest cost pressures in the plasma Wafer Fabrication Equipment (WFE) segment, and cost has been an ongoing motivation for novel approaches wherever possible to reduce overall system expense.

By their nature, deposition plasmas tend to utilize a more stable and predictable plasma impedance range compared to etch applications. The highly

cyclical character of these PECVD and PEALD processes, combined with an understanding of their impedance behavior, sets up both the need for, and opportunities to, optimize power delivery for performance, longevity, and cost effectiveness. For these applications, the unique physical integration of power generation and highly responsive impedance matching has increasingly proven to be very effective, achieving multiple benefits in these demanding and cost sensitive process tools.

FIGURE 2 shows an example of an integrated RF power delivery solution for PECVD. In this case, an electrically switchable solid-state capacitor array is physically integrated with the RF power amplifier to produce a small form-factor power delivery system in a single enclosure. The overall package size is nearly equal to a traditional RF power generator, but this design includes the matching network within the common enclosure. (Recall from the second article in this series, RF generators and matching networks are most commonly separate "boxes" connected by a power cable.) The integrated package permits mounting directly to the wafer pedestal connection and further saves space that would otherwise be necessary with traditional non-integrated, rack-mounted power supplies.

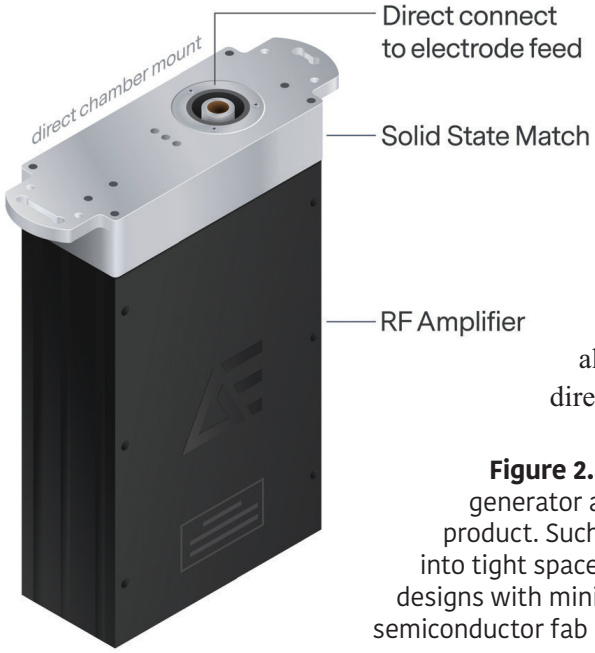


Figure 2. Example of an integrated generator and solid-state matching product. Such compact designs easily fit into tight spaces, enabling complex tool designs with minimized footprint in expensive semiconductor fab cleanrooms.

In addition to space savings, mechanically driven vacuum capacitors are replaced in the integrated matching network with solid state components that are more reliable and repeatable, both especially valuable attributes in these high-cycle, repetitive processes. More importantly, eliminating the traditional stand-alone vacuum capacitor matching network allows for very fast tuning and direct power regulation, which reduces latency in today's short-cycle, high step-count deposition processes (**FIGURE 3**).

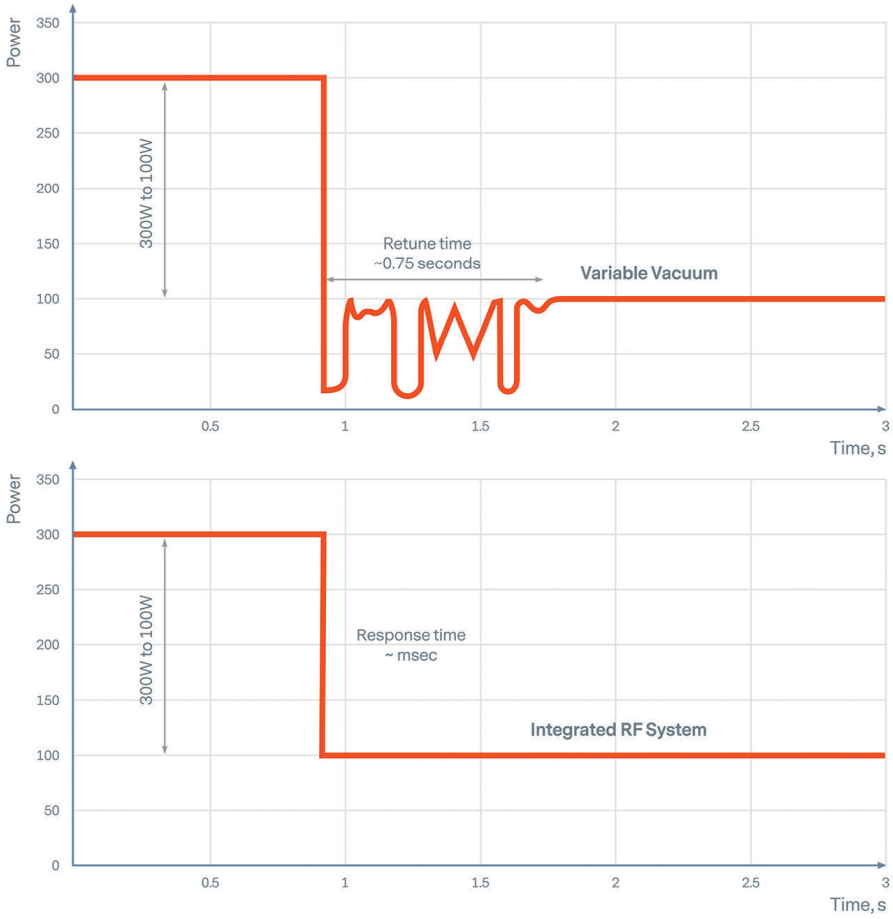


Figure 3. Shown is an integrated RF system delivering power to a setpoint change from 300W to 100W within a millisecond timeframe compared with typical vacuum capacitor matching network tune times of a second or more. This ensures constant, repeatable power delivery where the near-instant tuning increases productivity for these very short, high-cycle count plasma Deposition processes.

It is worth remembering that earlier generation integrated RF power delivery was instrumental in making 300 mm PECVD tools and Deposition processes practical. In 2000, Advanced Energy (AE) introduced an innovative 13 MHz generator with an integrated solid-state matching network (multiple PIN diodes) and the highest power density in the industry at the time (>0.27 watts per cm^3), which solved critical space issues in tool design. To date, more than 15,000 PECVD manufacturing chambers have relied on this integrated power delivery solution for critical deposition processes.

Today, there is a wide resurgence in, and expansion of, this physically integrated approach, as advancements

in controls, RF metrology, improved generator architectures, and new solid-state match switches are enabling a wider tuning range expanding the capability of this proven technology. This form of integrated process power delivery is more relevant than ever for providing high reliability, ultra-fast RF power response for the multi-chamber tools that must precisely deposit thin layers of alternating materials in vertical memory device stacks. The many short process steps produce extremely high cycle counts where solid-state tuning elements' robustness outlive vacuum capacitors in conventional matching networks.

While physical integration does not represent an entirely "new tool" in the process power toolbox, modern Deposition requirements where demand for increasingly stable plasma is critically coupled with the new architectures of extreme-productivity platforms, are driving a new class of highly engineered, integrated RF generator and match solutions into wider adoption.

Agile power with solid-state matching

Whereas many plasma deposition processes operate within a relatively narrow power and impedance space, other applications such as conductor etch—especially for etch applications defining 3D and high aspect ratio features—require a wide range of chemistries and a broader diversity of process conditions. These etch processes require power delivery and impedance matching spanning wider conditions than those targeted by the more straightforward integrated solutions discussed previously. However, like high-cycle Deposition applications, many etch recipes are migrating to shorter step intervals, with some steps down to less than a one-second range, necessitating agile, precise power delivery resembling the requirements for the short-cycle steps seen in PEALD processes.

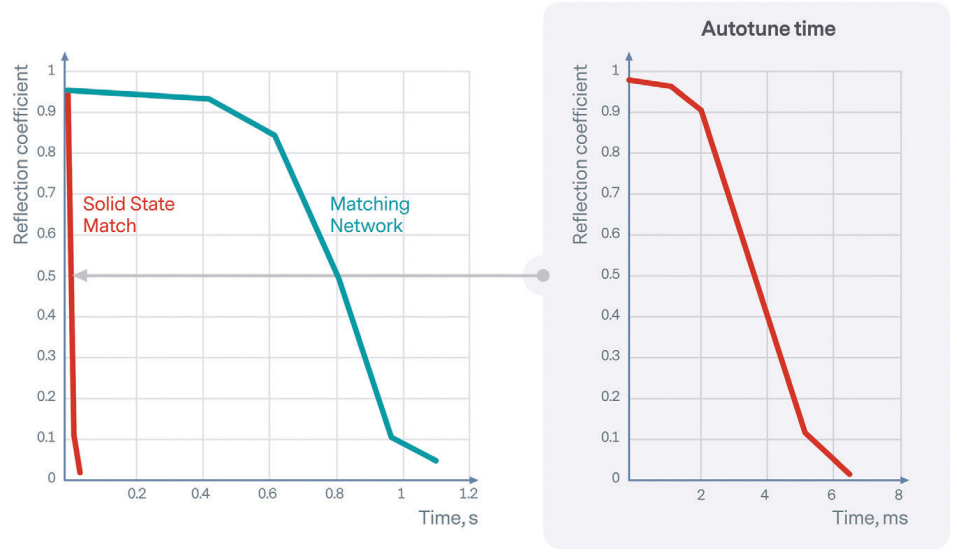


Figure 4. The purpose of a match network is to rapidly reduce reflected power to near 0W, as shown on plots of reflection coefficient (normalized reflected power) versus time. Typical tuning time of a traditional matching network (blue line) and a solid-state match (red line) on a plasma chamber are compared. The expanded view of the SSM tuning trace on the right shows that full tuning can be achieved in the millisecond timescale, compared with hundreds of milliseconds for a traditional match.

Driven by the need to achieve fast, accurate tuning across wider impedances, the integrated solid-state match solutions in cyclical deposition processes have been adopted into full-featured, stand-alone matching networks. Having similar form, footprint and tuning ranges as their traditional counterparts, these matches offer tuning times in the millisecond range compared with the one-second range typical of traditional matching networks. Illustrating this speed enhancement, tuning time of a solid-state match is compared with a variable vacuum capacitor match in **FIGURE 4**.

Solid-state matching networks (SSM), when configured for full-range tuning, can cover an impedance space essentially identical to that of a vacuum capacitor match while the compactness of solid-state circuitry allows the SSM to fit within a similar footprint of the conventional version (**FIGURE 5**). Further, SSM units have no moving mechanical parts. This eliminates wear and drift mechanisms inherent to traditional matches, enabling higher reliability and repeatability on a broad range of challenging and continually varying plasma processes.

With SSM tuning networks, accurate power delivery and tuning into critical, power sensitive, short duration plasma Etch steps are possible while still supporting the conventional longer steps and wide tuning range traditionally handled by vacuum capacitor based matching networks. Large and abrupt impedance swings, increasingly

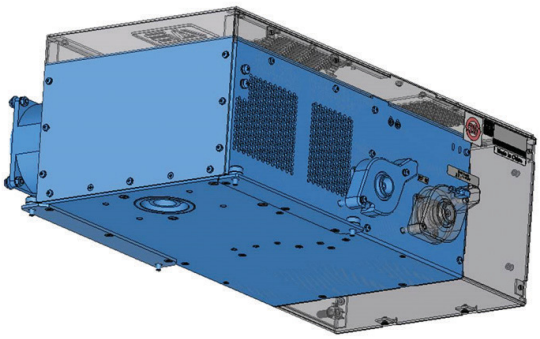


Figure 5. Solid state match networks can fit into chassis designs similar to standard match networks. The blue portion of the graphic represents the relative size of a standard match network, compared with a comparable solid -state match represented in grey (less than a third the size of conventional matching networks).

common in today's challenging 3D device Etch processes, can also be managed by merging fast switching with novel control algorithms. With these together, tuning speed and span combine to ensure consistent power delivery through process step excursions avoiding power delivery interruption and even "winking-out" of the plasma, sometimes seen in slower, vacuum capacitor-based designs. This new approach to impedance matching addresses a myriad of new challenges presented to modern process power systems by todays demanding 3D and HARC etches.

In 2010, AE introduced the first true, wide range solid-state match (Navigator II FastCap) [1] for semiconductor plasma processing applications. With millisecond tuning times, the Navigator II FastCap found initial interest in demanding high-speed matching applications like Conductor Etch, PECVD, and Metal Deposition processes because of its ability to ensure precise and repeatable power delivery through very short duration and very fast-changing plasmas. As process step times continue to contract, demand for solid-state-matching is growing and driving wider adoption of AE's even faster solid-state matches.

Narrowing ion energy distribution with non-sinusoidal (Non-RF) power

As discussed in the prior article, the complexity in 3D device manufacturing and High Aspect Ratio (HAR) features has driven the need for process power systems to provide better discrimination of targeted ion activation and control of the overall plasma environment. Specifically, the mix of ions to radicals, the energy distribution of ions incident to the wafer and respective arrival rates of various species (flux) are all key factors determining both quality and efficiency of these challenging 3D Etch process applications.

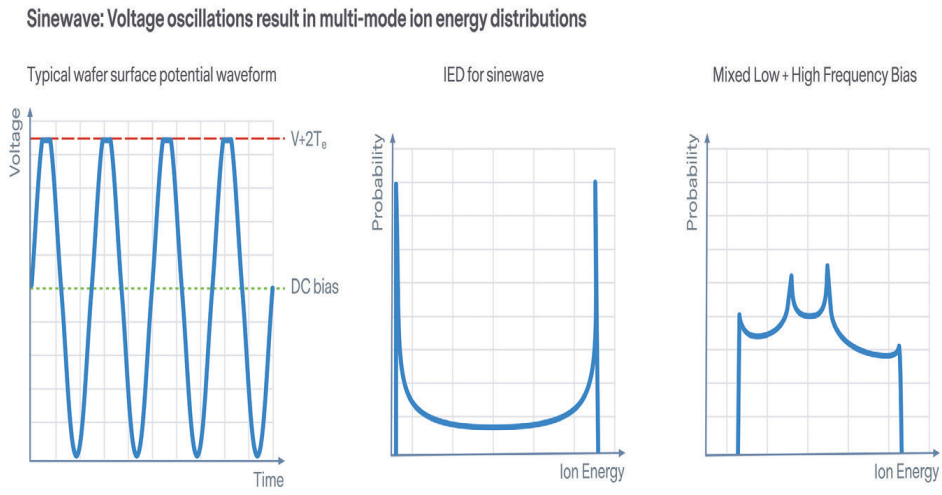


Figure 6. Oscillating voltage (left) from RF biasing translates, by virtue of its probability function, to (middle) bimodal ion energy distributions; mixing multiple frequencies leads to additional modes and more complex energy distribution functions (right).

Multi-frequency pulsed RF power systems are now mainstream largely for their ability to separate, or decouple, plasma density production (determining flux) from ion energy generation. The ion energy distribution is an especially critical factor in HAR Etch applications. Balancing High Frequency (HF, usually 13 MHz or higher) and Mid Frequency (MF, usually 300 kHz to 2 MHz) power provides process engineers more control over ion energy distributions compared to single-frequency pulsed RF power systems. However, the time-varying character of RF biasing is emerging as a significant limitation to the scaling of these features even with dual or multi-frequency pulsed RF power.

The sinusoidal nature of RF biasing produces broad, multi-peaked Ion Energy Distributions (IEDs) as seen in **FIGURE 6**. As described in the previous article, mixing HF with MF power balances plasma creation with high ion acceleration. To achieve the energies required in some of today's demanding Etch processes, increasing amounts of MF power is required. However, MF power is rising to difficult to manage levels, making controllable and/or narrow IEDs even harder to achieve. Higher MF also drives higher thermal loading, along with lower efficiencies since larger fractions of applied bias is consumed by (High Aspect Ratio Contact) accelerating ions to energies not adequate for and even detrimental to the process.

As 3D IC devices increase the number of layers and features get deeper, there will be a point where it becomes impractical to use conventional multi-frequency power delivery systems. The lack of adequate IED control and the trend towards higher and higher powers are problematic, leading to the need to explore alternate process power solutions for HAR biasing. A non-sinusoidal bias source has emerged as an advantageous alternative. This non-RF approach provides the means for direct control of bias voltages offering a leap forward in both efficiency and in the control of IEDs for complex and demanding applications such as 3D Etch, HAR Etch, and deposition applications.

More fundamentally, to control the distribution of ion energies to a substrate, the voltage across the plasma sheath above the substrate must be carefully controlled. Since dielectrics are present on both the wafer and electrostatic wafer chuck, a form of AC voltage is required. Asymmetric waveforms, when properly regulated, have

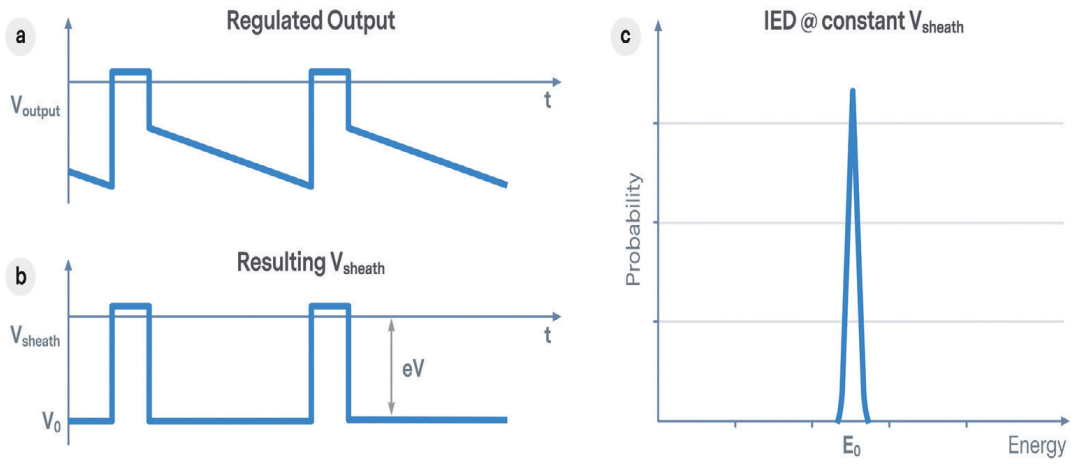


Figure 7. In asymmetric biasing, the Ion Energy Distribution can be directly controlled by regulation of the sheath voltage. Controlling the magnitudes and slopes of the output waveform (a) provides user control of sheath voltage (b). Narrow Ion Energy Distribution (c) is achieved when the sheath voltage is regulated to constant value.

been shown [2] to provide a controllable voltage drop across the sheath and therefore controllable ion energies to the surface. To achieve direct control of the IED, an asymmetric bias approach must accomplish three key tasks: 1) establish the accelerating (sheath) voltage at the surface; 2) maintain the sheath drop at a desired level; and 3) periodically reverse to remove charge and reset for the next cycle. **FIGURE 7** shows how a preferred asymmetric waveform (a) at the output (delivered to the wafer chuck) translates to producing a controllable sheath voltage (b) and, when V_{sheath} is held constant, a narrow IED (c).

Non-sinusoidal, asymmetric bias systems are new to high-volume semiconductor manufacturing process power and offer unique advantages over conventional sinusoidal bias power approaches. Beyond the ability to produce narrow IEDs (**FIGURE 8 left**), they also offer the possibility to tailor energy distributions (wide or narrow) according to specific process requirements (**FIGURE 8 right**). Controlling the effective IED based on process objectives brings the advantage of efficiency. When bias power is more directly applied to producing only the desired range of ion energies, less power is lost accelerating ions to undesirable levels. Improving bias power efficiency not only reduces waste but also avoids the growing potential for wafer damage and premature chuck and chamber wear-out.

AE introduced eVoS™ a non-RF, asymmetric-wave bias product in 2019. Initially used in specialty applications, eVoS applications are broadening to high-performance etch and atomic-scale processes where precise control of bias and resulting ion energy distribution is required. As the novel technology scales for 3D and deep, high aspect ratio processing, the eVoS solution provides the means to increase targeted ion energy without excessively increasing RF thereby avoiding untenable levels of power and associated costs for these increasingly deep and

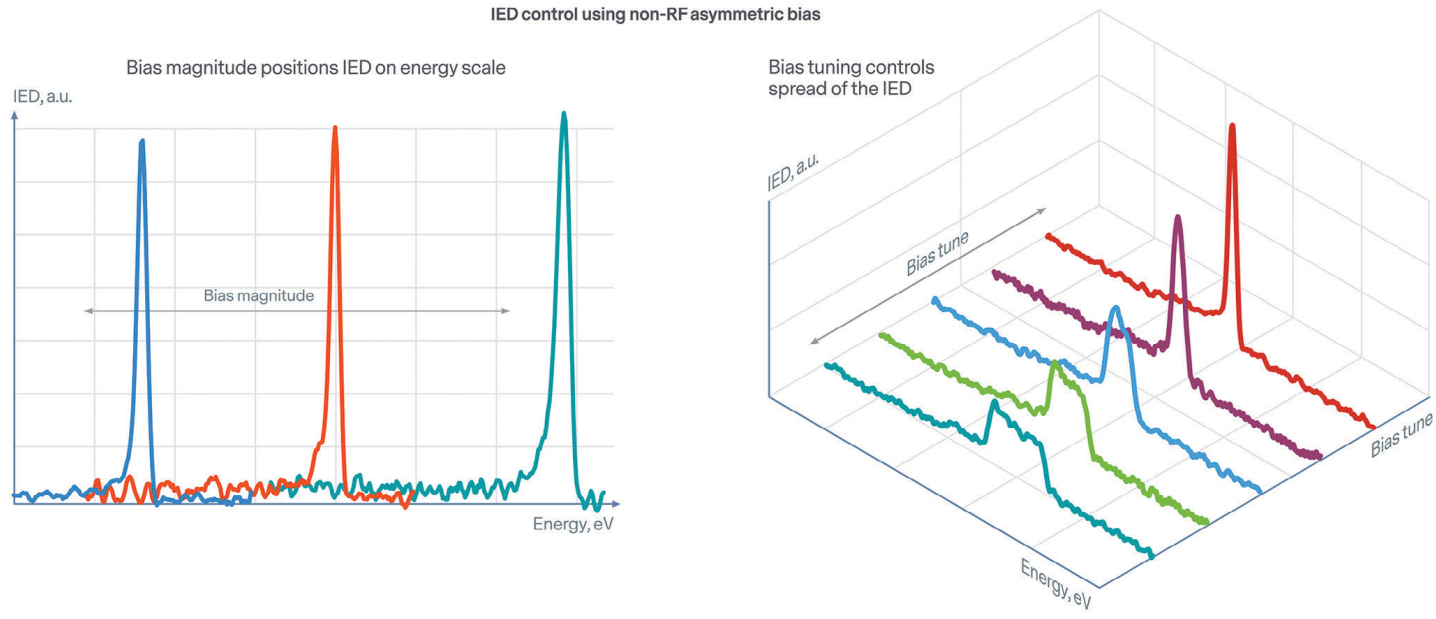


Figure 8. The AE eVoS non-RF, asymmetric bias system [3] allows independent setting of bias magnitude and bias tuning, separating controls of IED average from IED width [4].

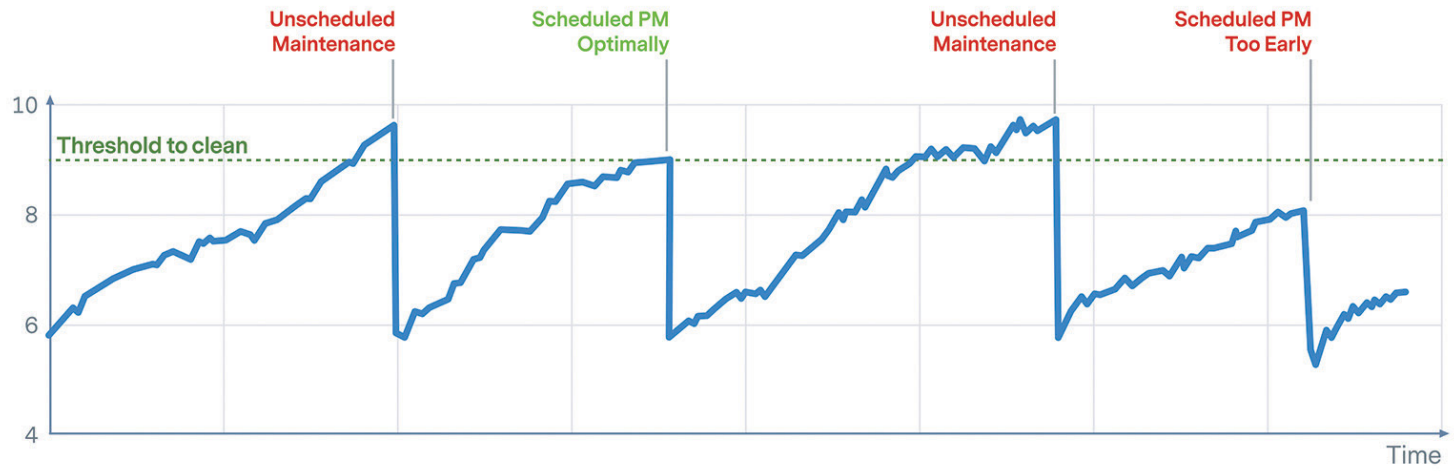


Figure 9. Time-based preventive maintenance cycle indicates non-optimized maintenance cycles. Green dashed line indicates optimal condition for PM.

challenging features. Benefiting from the asymmetric-wave application, eVoS provides control capabilities not available to RF (single or multiple frequency) bias systems and potentially opens new frontiers in this ever-evolving, new form of Etching.

The power of prediction

Increasingly, technology commentators claim, "data is the new oil" because of its growing value in many aspects of our daily lives. In the past few years, however, it has become clear that collecting and saving data for data's sake alone does not add much value. However, when domain knowledge is combined with these data, it becomes possible to add huge value by identifying the "needle in the haystack" and thereby enables useful insights to be extracted from an overwhelming sea of data. In semiconductor manufacturing, the intersection of process power engineering expertise with advanced data science methodologies is where value creation happens, empowering delivery of highly differentiated value by improving process performance and results.

The objectives to deliver data-driven, actionable insights are primarily to 1) contribute to reducing the total cost of ownership; and 2) more rapidly, and with more certainty, diagnose yield and defect containment challenges. Unique to process power, and AE specifically, AE's power generators and matching networks can be used directly to collect information that enables AE analytics without the need for additional external sensors.

An illustrative process performance solution is predictive chamber re-conditioning for point-of-use plasma abatement (of effluent process gases in semiconductor manufacturing). For this application, the chamber in a remote plasma source, used to break down the process gas byproducts downstream of the process chamber, wears over time and ultimately degrades the performance of the plasma abatement itself and requires servicing to return to targeted performance.

The typical preventative maintenance approach is a fixed-time intervention, such as monthly cleaning/reconditioning depending on the estimated usage. As illustrated in **FIGURE 9**, fixed-time maintenance is often not optimal. Sometimes, the cleaning of the chamber is done prematurely (wasting money), and sometimes it is too late, impacting process performance or worst-case causing process drift, resulting in unplanned downtime to service the plasma abatement chamber. In high-performance fabs,

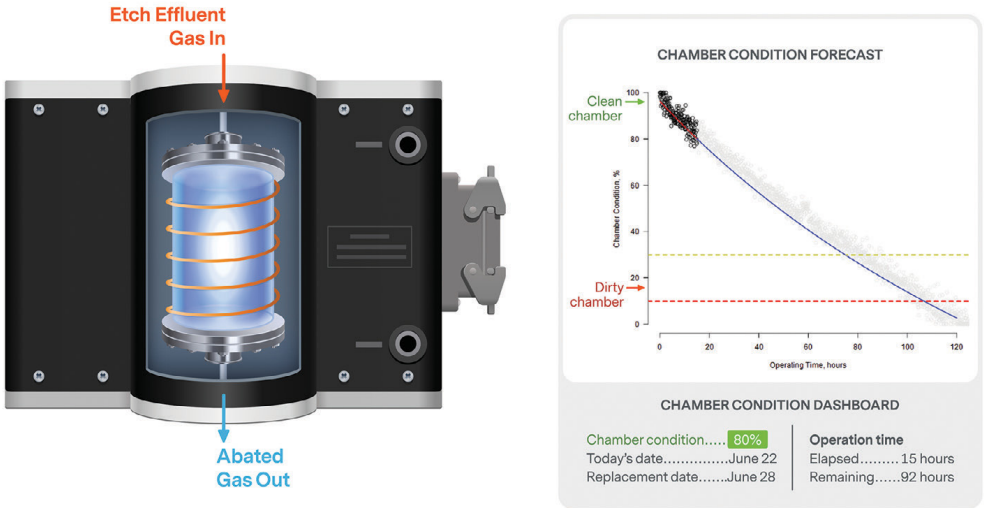


Figure 10. Abatement chamber (left) with chamber condition dashboard (right) showing figure-of-merit (chamber condition 0-100%) on vertical axis, and operation time on horizontal axis.

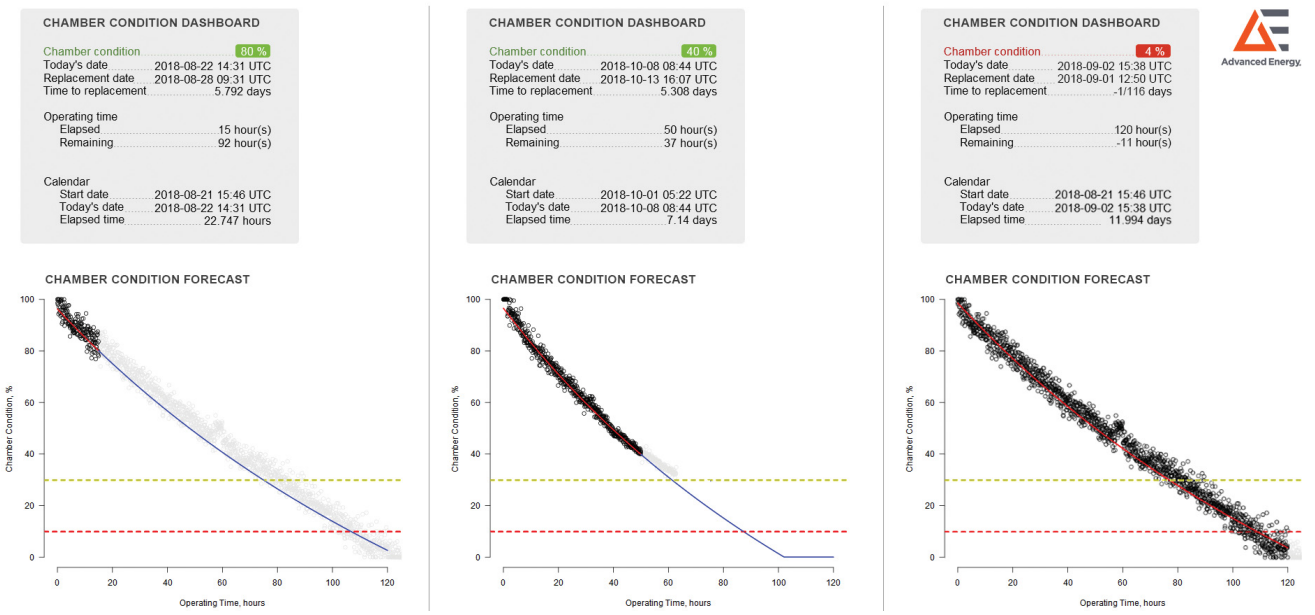


Figure 11. 11a: chamber in good condition. The blue line indicates the predictive chamber condition based on ingested data. 11b: chamber still in good condition, the model blue line adjusts the prediction based on usage and more ingested data. Notice slope of line has changed. 11c: chamber must be serviced as the threshold for maintenance has been crossed (red line). The yellow line indicates an earlier warning which can be set by the customer.

there is little tolerance for unexpected maintenance events that reduce productivity and raise costs.

In contrast to time-based maintenance cycles, AE developed a predictive chamber clean solution (PowerInsight by Advanced Energy™) [5] that alerts the user when the chamber should be cleaned, based on real-time usage conditions, regardless of gas species breakdown, the previous cleaning cycle, or the starting condition of the chamber. Put simply, the predictive power methodology, deployed as a Machine Learning solution and integrated into customer platforms, provides a figure-of-merit annunciation for chamber condition status and time-to-servicing (**FIGURE 10**).

In **FIGURES 11A**, **11b** and **11c**, the figure-of-merit trend as the number of hours left to chamber cleaning is shown. The yellow line is the threshold for warning, and the red line indicates when cleaning/chamber reconditioning must be done. AE's process power knowledge in combination with proprietary data analytics provides unique value differentiation to customers to optimize process performance and maintenance cycles.

With the explosion of Big Data, analytics, and the exponential rise of computational power, fabs are increasingly dedicating resources to use Artificial Intelligence (Machine Learning or Deep Learning) to extract actionable insights for improving productivity and yield. AE's solutions are enabling the "last mile" of actionable insights, marrying the physics of process power applications to big data analytics. Feedback from end users (fabs) is increasingly validating this approach as a wider range of applications are adopted.

Out from the shadows to power ahead

Sub 7 nm technology nodes drive new process requirements Continued on page 65

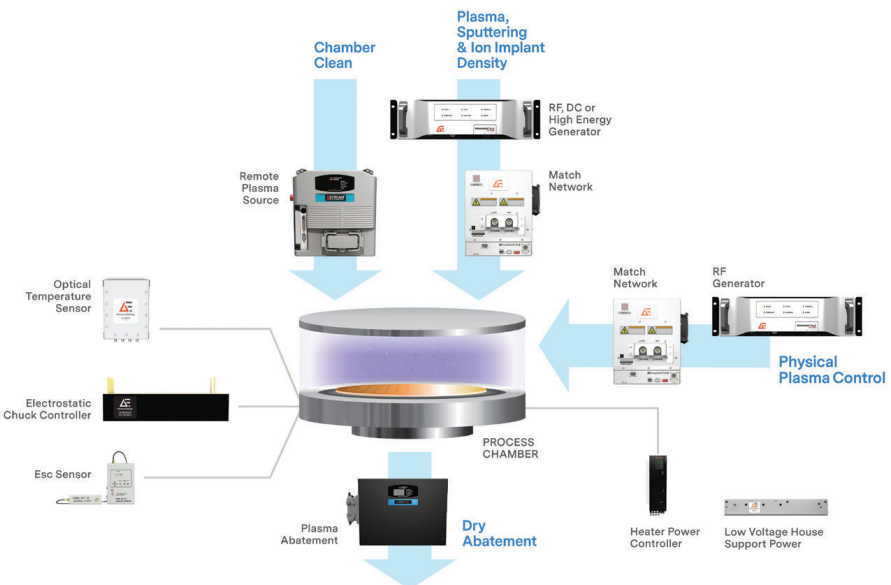


Figure 12. Example of the range of power and temperature measurement and control solutions for Wafer Fab Equipment. Going forward the interrelationship of these products will increasingly be sharing information for system optimization.

A Novel Approach for Nanopatterning Using AFM Lithography

JAKE KIM, **CHARLES KIM** and **CATHY LEE**, Park Systems, Research Application Technology Center

Park SmartLitho software is ideally suited for novel device surface structure development based on nanolithography and for investigations into ever decreasing feature sizes and line spacings for advanced nanoelectronics.

SINCE THE INVENTION OF ATOMIC FORCE microscopy (AFM) [1], it has found widespread application in non-destructive sample surface imaging and significant interest in its electrical, magnetic and mechanical properties. However, in addition to those, AFM offers vast potential for local surface modification and patterning, using either excessive cantilever loading force resulting in mechanical scratching, ferroelectric switching or by oxidizing the surface via the application of AFM tip bias. Local oxidation, also known as 'bias mode AFM nanolithography,' has been widely used in the customization of nanoscale conducting or semiconducting surface patterning [2, 3]. Bias mode AFM nanolithography offers many advantages for customized patterning: it circumvents the diffraction limit present in optical lithography methods, it does not require optical masks and its procedure is straight forward. By applying a voltage bias between a conductive AFM tip and a substrate, the tip-sample contact region forms an oxidized layer (**FIGURE 1**). Nanopattern control by oxidation occurs by manipulating experimental parameters, including the applied AFM tip bias, the tip material/geometry, scan speed and humidity.

In this application note, we present nanopatterning via oxide growth on a bare silicon wafer [4] using bias mode AFM nanolithography with Park SmartLitho, the new nanolithography software developed by Park Systems [5]. Furthermore, we demonstrate the capabilities of Park SmartLitho for lithography on ferroelectric samples by patterning local domains of a PZT (lead zirconium titanate) film on silicon wafers [6]. Using Park SmartLitho, we fabricated complex structures with detailed and elaborate features, via both oxidation and ferroelectric domain switching. Finally, we propose optimized process parameters for successful nanopatterning. The software is ideally suited for novel device surface structure development based on nanolithography and for investigations into ever decreasing feature sizes and line spacings for advanced nanoelectronics.

Materials and methods

A bare silicon wafer and a PZT film on a silicon wafer were used to demonstrate nanopatterning via oxidation and ferroelectric domain switching, respectively. The root mean square roughness (Rq) for the bare silicon wafer was > 1 nm so that

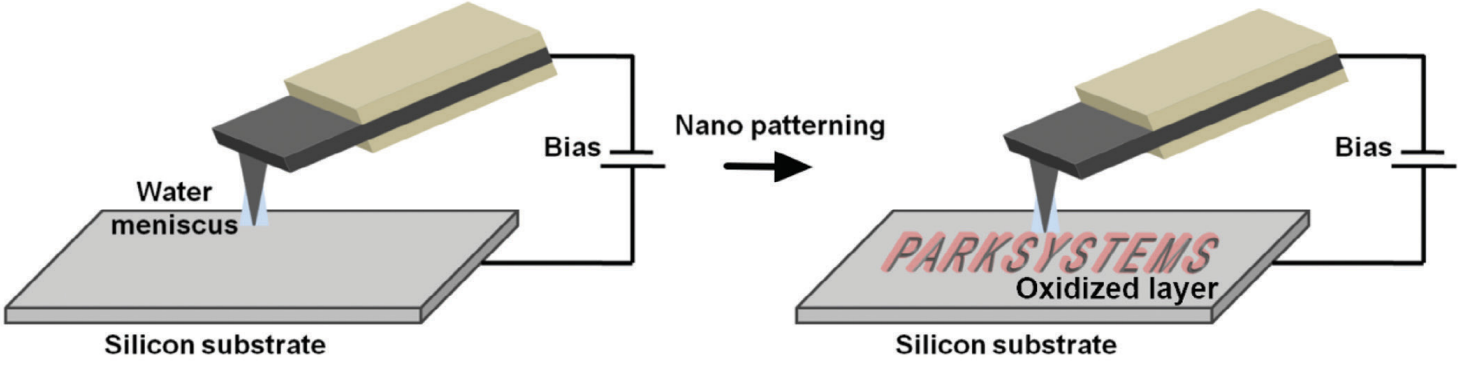


Figure 1. Schematic representation of bias-assisted AFM nanolithography via local surface oxidation. Patterns are created on the surface by oxidation upon bias application between tip and sample.

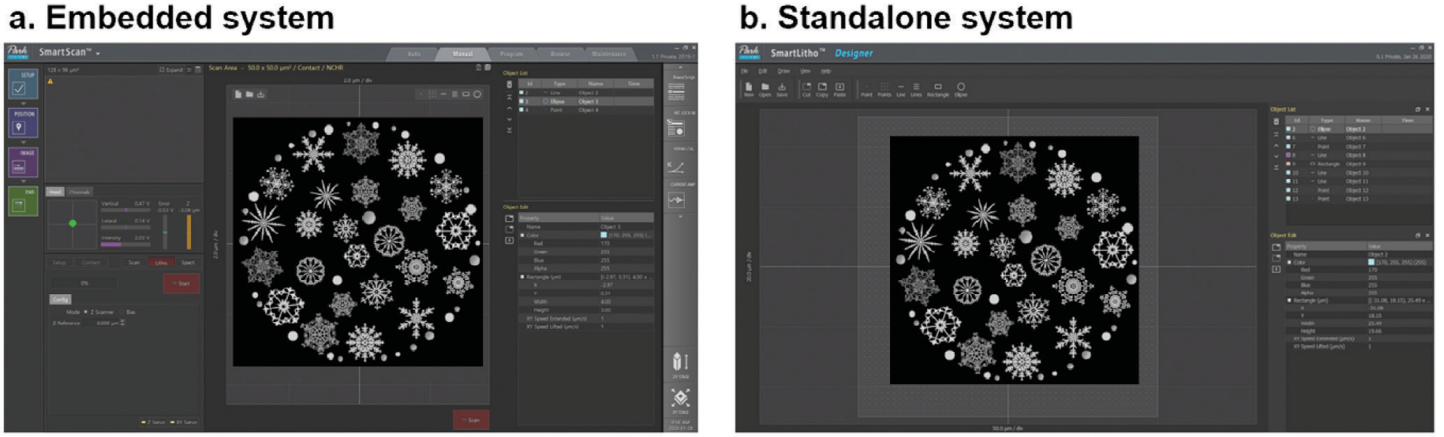


Figure 2. Captured images of Park SmartLitho software, (a) embedded in Park SmartScan imaging software and (b) the stand-alone version.

the height contrast between the oxidized layer and the substrate could be identified. We used a CDT-NCHR cantilever with a nominal spring constant of 80 N/m and a resonance frequency of 400 kHz for oxidation nanopatterning. The CDT-NCHR tip contains a conductive diamond coating (< 10 kOhm at the platinum surface) and the tip curvature radius was 100 nm ~ 200 nm. The tip radius of the CDT-NCHR was slightly larger than tips with other conductive metal coatings; however, diamond coated tips show improved operational stability and higher sensitivity. For nanopatterning via ferroelectric domain switching on a PZT, we used a PPP-EFM cantilever with a nominal spring constant of k = 2.8 N/m and a resonance frequency of 75 kHz. The PPP-EFM tip has a platinum-iridium coating and a radius of approximately 25 nm. Since ferroelectric domain switching on PZT does not require a high loading force between the AFM tip and the sample surface, a soft conductive cantilever was used. The detailed scan parameters are found in Table 1.

All measurements were conducted on a Park NX10 AFM system utilizing Park SmartLitho, the new AFM nanolithography software from Park Systems. Park SmartLitho combines a variety of nanolithography modes, including constant Z scanner mode, constant force mode, bias (and other) modes with a user-friendly graphic editor. The user can access many template types for customized nanopatterning such as bitmap images, polygons, polylines and polydots. Park SmartLitho is available either as a stand-alone version or embedded in the Park SmartScan™ operating software (**FIGURE 2**). The stand-alone and embedded versions offer the same functions, like drawing file creation as well as bitmap or pattern editing, and can operate the user's chosen/preferred nanolithography mode.

Table 1. AFM nanolithography parameters

	NANOPATTERNING BY OXIDATION		NANOPATTERNING BY FERROELECTRIC DOMAIN SWITCHING	
SAMPLE	Bare Silicon wafer		PZT on Silicon wafer	
CANTILEVER	CDT-NCHR		PPP-EFM	
ENVIRONMENT	High humidity (> 60%)		Ambient	
	NANOLITHOGRAPHY	**AFM IMAGE**	**NANOLITHOGRAPHY**	**AFM IMAGE**
SCAN SIZE	35 μm×35 μm	40 μm×40 μm	30 μm×15 μm	35 μm×25μm
PIXEL RESOLUTION	400×400	1024×1024	300×300	1024×1024
SCAN RATE	20 μm/s	0.5 Hz	10 μm/s	0.5 Hz
BIAS RANGE	0 ~ –10V	-	0 ~ –10 V	AC bias: 2V DC bias: 0V

Results and discussion

FIGURE 3 shows the original bitmap image used as the nanopatterning template as well as the height

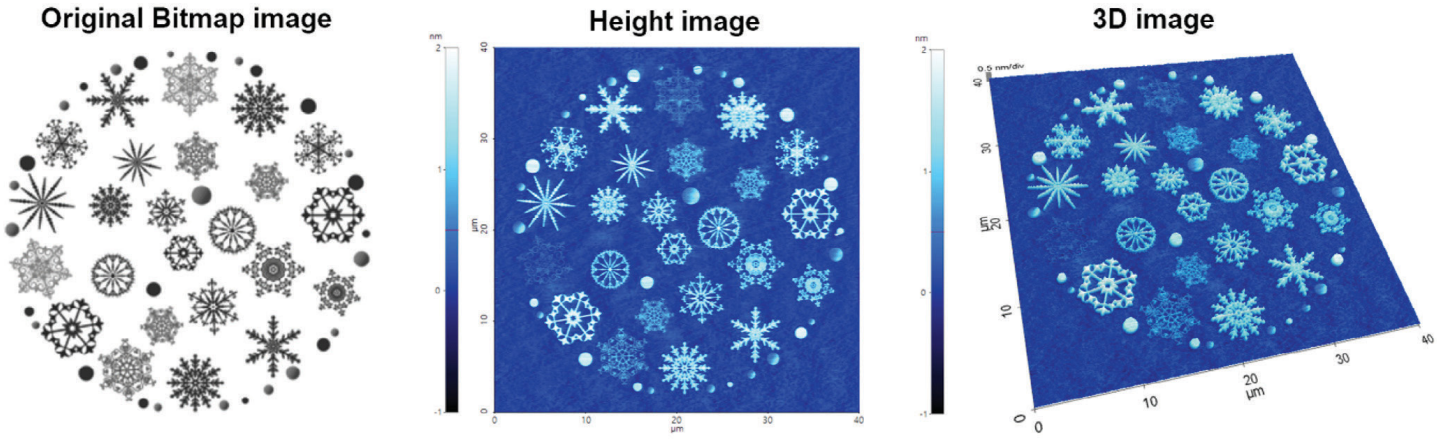


Figure 3. Original bitmap image used as the template (left), AFM height image after bias assisted AFM nanolithography (center) and corresponding 3D image of the sample surface (right). The nanopatterning was generated by surface oxidation.

and corresponding 3D images of the sample surface after bias mode AFM nanolithography. The nanopattern on the surface of the bare silicon wafer resembled the template of the Christmas ball closely. The pattern was generated using bias assisted silicon oxidation by applying a sufficiently high voltage between the tip and sample. The oxide layer thickness was controlled by the applied bias magnitude and environmental humidity. The oxide layer thickness was directly proportional to the applied tip bias. Furthermore, oxidation required a water meniscus at the AFM tip end that was influenced by environmental humidity [7]. Therefore, the measurement was conducted from a 0 V ~ 10 V tip bias under high humidity conditions (> 60%). The oxide nanopattern featured heights that ranged from 0.8 nm ~ 1.5 nm. FIGURE 4 shows two of the constituent structures (5μm×5 μm) with their corresponding 3D images and line

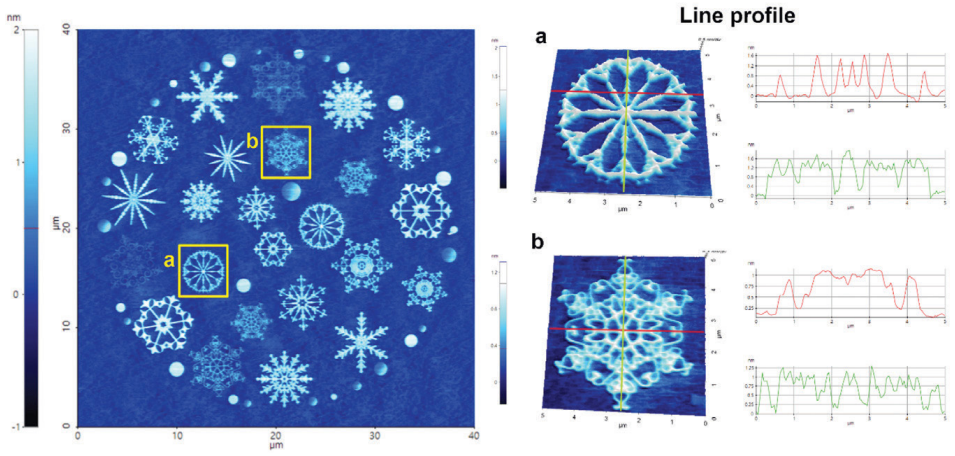


Figure 4. AFM height images after nanolithography and line profile analysis on two of the constituent structures. AFM line profiles in a and b show the ~1 nm height of the oxide layer from the substrate.

profile information. From these line profiles, we confirmed structure heights of approximately 1 nm. The distinct height image contrast after nanolithography illustrated the successful and high-resolution nanopatterning available by bias assisted oxidation using the new Park SmartLitho software from Park Systems.

FIGURE 5 shows height and piezoelectric force microscopy (PFM) quad images, their line profile information and 3D images after lithography. Ferroelectric domains in PZT can be switched by applying different biases between the tip and sample. We performed both AFM non-contact imaging for height information and PFM imaging for PFM quad information after bias mode AFM nanolithography. As shown in both, the height images were flat surfaces without a pattern. However, the structure resolved in the PFM quad strongly resembled the original traditional Korean pattern template, illustrating successful ferroelectric domain switching. While nanopatterning by oxidation requires humidity control for high-quality nanolithography, nanopatterning via ferroelectric switching requires slow scan speeds and lower loading force scanning to avoid sample damage due to contact mode.

Conclusion

In this application note, we demonstrate the nanopatterning via oxide growth on a bare silicon wafer using bias mode AFM nanolithography with Park SmartLitho, a novel nanolithography software from Park Systems on a Park NX10 AFM system. We successfully showed that using Park SmartLitho, a bias assisted nanopatterning of complicated structures can be readily done by locally oxidizing the surface of a silicon wafer and switching ferroelectric domains of a PZT film. Furthermore, we proposed optimized parameters including humidity values, a tip bias range, scan rate and suitable cantilevers to generate accurate and distinct nanoscale oxide patterns on a silicon substrate as well as ferroelectric domain patterns on PZT. This study demonstrates that nanopatterning by AFM nanolithography, in particular oxidation patterning on silicon substrates and Park SmartLitho, offers significant potential for customized nanosensor surface modification as well as nanodevices for semiconductor research and industry.

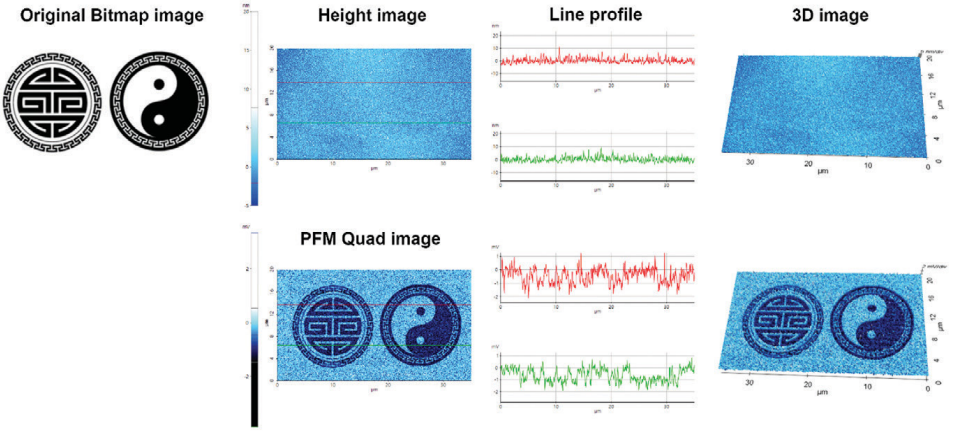


Figure 5. Original bitmap images used as templates (left), AFM height and PFM quad images after ferroelectric switching in AFM nanolithography bias mode, line profile analysis (center) and the corresponding 3D images (right). The nanopattern on PZT was generated by ferroelectric domain switching.

REFERENCES

1. Binnig, G., Quate, C. F., & Gerber, C., Atomic force microscope. Physical review letters, 56 (9) (1986) 930.
2. Minne, S. C., Flueckiger, P., Soh, H. T., & Quate, C. F., Atomic force microscope lithography using amorphous silicon as a resist and advances in parallel operation. Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 13 (3) (1995) 1380-1385.
3. Garcia, R., Martinez, R. V., & Martinez, J., Nano-chemistry and scanning probe nanolithographies. Chemical Society Reviews, 35 (1) (2006) 29-38.
4. Garcıa, R., Calleja, M., & Pérez-Murano, F., Local oxidation of silicon surfaces by dynamic force microscopy: Nanofabrication and water bridge formation. Applied Physics Letters, 72 (18) (1998) 2295-2297.
5. Advanced Vector Nanolithography Using Closed Loop Scan System (2010, Online Article). Retrieved from:https://parksystems.com/park-spm-modes/97-mechanical-properties/251-nanolithograpy
6. Li, D., & Bonnell, D. A., Ferroelectric lithography. Ceramics International, 34 (1) (2008) 157-164.
7. Abdullah, A. M., Hutagalung, S. D., & Lockman, Z., Influence of room humidity on the formation of nanoscale silicon oxide patterned by AFM lithography. International Journal of Nanoscience, 9 (04) (2010) 251-255.

Challenges to Replacing Hard-Disk Drives

BEN WHITEHEAD, Storage Product Specialist, Mentor, a Siemens Business, and **LAURO RIZZATTI**, Verification Expert

Computational storage devices are the new must-have peripherals for intensive storage applications.

In the world of computing, moving data is an expensive proposition, slowing performance and increasing power consumption. The challenge is magnified as storage developers attempt to replace hard-disk drives (HDDs) with solid-state drives (SSDs).

The design community has looked for ways to decrease data movement between storage and compute engines since the early days of SSDs. In the quest, a new approach called computational storage device (CSD) emerged that will reduce overall energy consumption and increase performance.

When comparing movement of data in storage versus compute engines, storage data access consumes orders of magnitude more energy than data processing in the computing engine.

Power consumption of various computations versus memory accesses is recapped in Table 1. An 8-bit fix-integer addition consumes 0.03 picoJoules (pJ), while the same addition based on 8-bit floating-point burns 10 times more energy. Reading from an 8 kilobyte SRAM consumes 5 pJ or 10X more power, and two orders of magnitude (640 pJ) more than when reading from a large DRAM.

As shown in FIGURE 1, CSD eliminates backward and forward data transfers with the host computer by performing computations locally within the SSD.

The top blue arrow in an SSD example represents the request data from the host to storage. The two orange arrows represent the read data from the SSD moving across the data bus to the host, and the manipulated data by the host moving back to the SSD. Orange arrows highlight massive power consumption and performance degradation.

The concept encapsulated in the CSD adds a small processing element next to the storage. The request data from the host can be computed locally by that processing element, eliminating the data transfer with the host, saving power and accelerating execution.

Two implementations of the CSD architecture are available today, either using a dedicated field programmable gate array (FPGA) or an application processor (AP).

Data can be processed locally by attaching an FPGA to an SSD on the same board and configuring the programmable device with compute elements, such as an Arm core or DSPs with advantages and drawbacks. The benefit is improved performance, while the disadvantage is the need to make the FPGA file-aware of the activity inside the SSD to be able to retrieve and process files. A system for file management with reading, processing and writing files back to the storage fabric is needed.

The alternative is to deploy an

Operation:	Energy (pJ)	Area (µm²)
8b Add	0.03	36
16b Add	0.05	67
32b Add	0.1	137
16b FP Add	0.4	1360
32b FP Add	0.9	4184
8b Mult	0.2	282
32b Mult	3.1	3495
16b FP Mult	1.1	1640
32b FP Mult	3.7	7700
32b SRAM Read (8KB)	5	N/A
32b DRAM Read	640	N/A

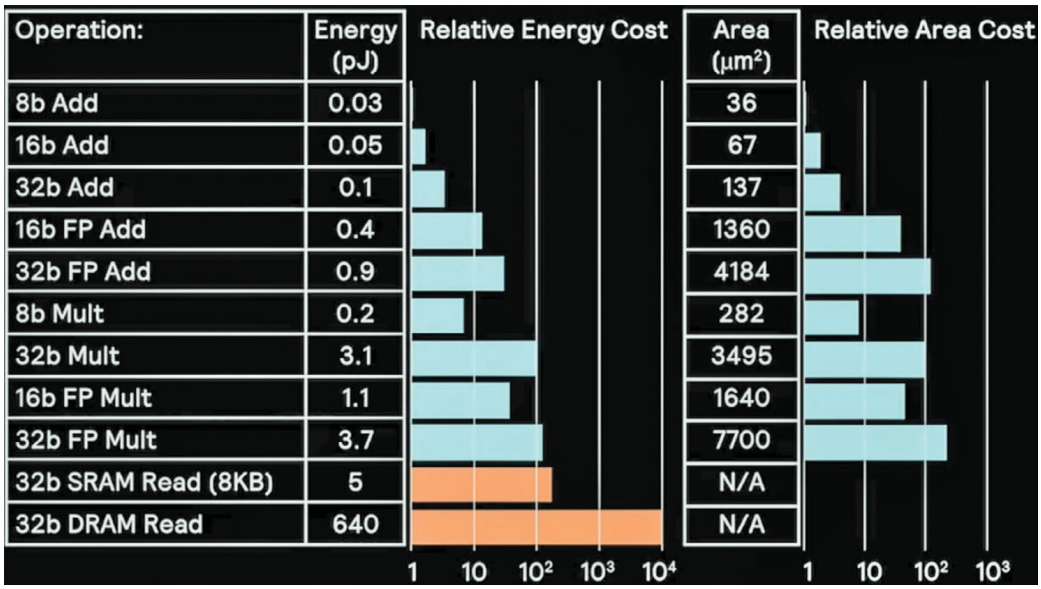


Table 1. Data movement is ruled by energy consumption.
Source: Mark Horowitz, ISSCC presentation 2014

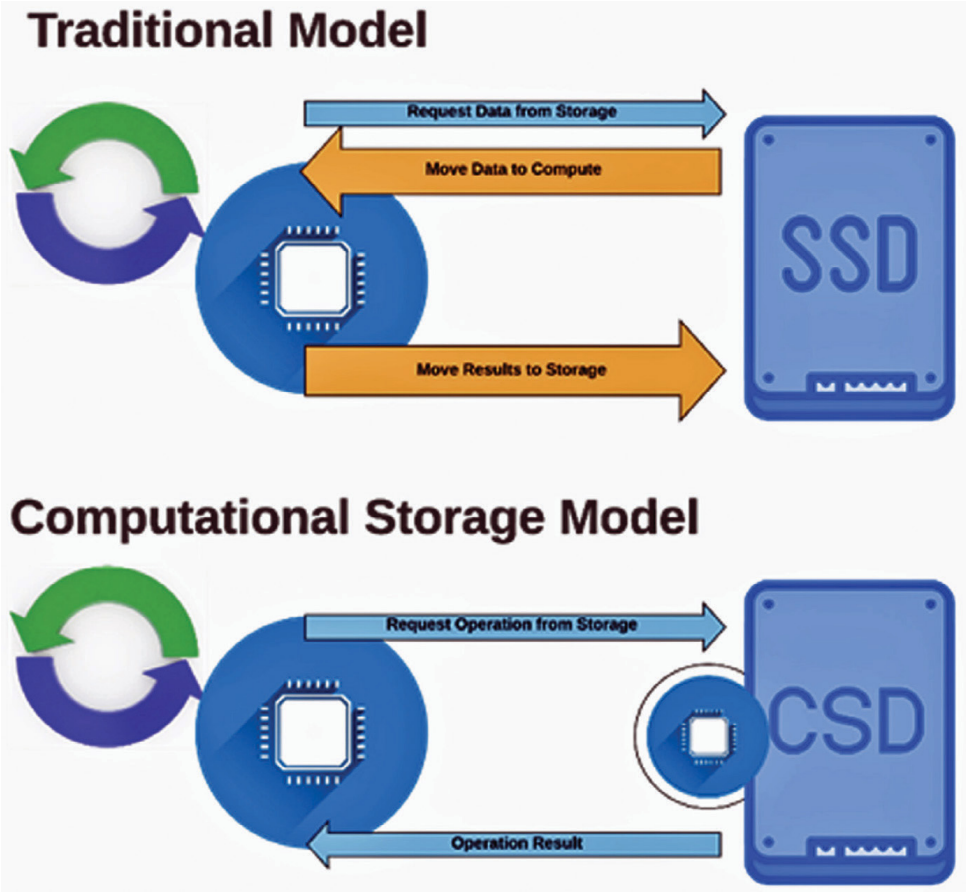


Figure 1. CSD eliminates data transfers with the host computer by performing computations locally within the SSD.
Source: Mentor, a Siemens Business

AP that runs Linux. Linux is natively file aware, readily available, ensures the consistency of the file system and is served by an open-source developer community, existing infrastructures, a range of development tools and many applications and protocols. The setup is known as "on-drive Linux."

The host could view a CSD running as a standard non-volatile memory express (NVMe) SSD when installed in a PCI port. The user can create a secure-shell (SSH) connection into the drive, view it like any other host on the network because Linux runs locally as the CSD becomes a network-attached — in effect, a headless-server running inside an SSD. Applications can directly manipulate files stored on the NAND.

Disaggregated storage will be even more important as storage systems move to a non-volatile memory express over fabrics (NVMe-oF) protocol to connect hosts to storage across a network fabric using the NVMe protocol. Designers can disaggregate and move compute "in situ" to improve performance, lower power usage and clear PCIe bandwidth for the rest of the system using CSD. NVMe-oF reduces some storage network bottleneck.

While adding SSDs to servers in data centers scales the amount of storage, it does not scale processing power. Instead, adding CSDs scales linearly total capacity as well as performance, the result of a local processor. By increasing the number of CSDs in data centers, the number of processors increases linearly leading to performance scaling.

The TERASORT benchmark in FIGURE 2 compares deployment of an SSD with a CSD by increasing their number of cores from 1 to 8. The number of SSDs does not change the performance (orange line), but performance is accelerated by adding CSDs with the cross point at four units (blue line). If one CSD takes approximately 850 milliseconds (ms) to run a TERASORT benchmark, eight CSDs run the same benchmark in about 500ms.

CSD verification

Adding processors plus an entire Linux stack and applications to the already complex hardware/firmware of the traditional SSD make CSD verification a challenging task.

Traditional verification approaches are inadequate due to a recent discovery that found the non-deterministic nature of the SSD storage interfered with hyperscale data center requirements. The solution to both is hardware emulation-based virtual verification that allows for pre-silicon performance and latency testing within 5% of actual silicon. Veloce VirtuaLAB from Mentor, a Siemens Business, is an example of the virtualization methodology. A new set of tools for CSD design verification from block level to system level were needed to build on top of tools and expertise developed to support a networking verification methodology.

System-level verification includes six parts starting with a PCIe/NVMe standard host interface setup:

1. Virtual NVMe/PCIe host running real-world applications on Quick EMUlator (QEMU) to implement host traffic
2. Veloce Protocol Analyzer for visibility on all interfaces such as NMVe, PCIe and the NAND fabric
3. Hybrid configurable platform with software stack to boot Linux and run applications virtually with the ability to save the state of the system at any point and re-start from as needed
4. Veloce emulation platform to emulate the CSD design under test (DUT) in pre-silicon with real-world traffic
5. Soft models for

Continued on page 64

Developing a UHF RFID Reader RF Front End

VAN YANG, **EAGLE ZHANG**, and **AARON HE**, Analog Devices, Inc., Norwood, MA

Two implementations showing how engineers can trade off receiver sensitivity for reduced design complexity, component count, and board space in UHF RFID applications.

ULTRAHIGH FREQUENCY RADIO FREquency identification (UHF RFID) systems have been widely adopted for applications such as asset management and apparel retail. Recently, they have gained attention for use in unmanned supermarket applications and for the electronic identification of motor vehicles. This article focuses on a target application for the electronic identification of motor vehicles in China, which must be compliant with the Chinese standards GB/T 29768-2013 "Information Technology—Radio Frequency Identification—Air Interface Protocol at 800/900 MHz" [1] and GB/T 35786-2017 "General Specification for Read-Write Equipment of the Electronic Identification of Motor Vehicles." [2] This article describes two implementations, showing how engineers can trade off receiver sensitivity for reduced design complexity, component count, and board space. Although the RF front end described in this article is application specific, both the analysis method and the front end itself are applicable for general UHF RFID reader solutions.

Standards summary

According to the GB/T 29768-2013 and GB/T 35786-2017 standards relating to the electronic identification of motor vehicles, the key air interface parameters and the performance requirements of high performance Type 2 readers for these applications are summarized in Table 1 through Table 3.

System link budget analysis

Passive RFID systems have two fundamental link limits: The forward link normally limited by the minimum RF to dc power to supply the tag electronics and the reverse link limited by the reader receiver sensitivity. The forward and reverse link budget formulas [3,4] are described in Equation 1 through Equation 3:

$$Prip = \frac{Ptx \times Gtx}{FSPL} \quad (1)$$

$$Prx = \frac{Ptx \times Gtx \times Grx \times Gtag^2}{FSPL^2} \eta_{\mathrm{mod}} \quad (2)$$

$$FSPL = \left(\frac{4\pi d}{\lambda}\right)^2 \quad (3)$$

Prip: tag receive isotropic power
Ptx: reader transmit power
Gtx: reader transmit antenna gain
Gtag: tag antenna gain
FSPL: free space pass loss
Prx: reader receive signal power
Grx: reader receive antenna gain
η_{mod}: tag modulation efficiency
d: distance between reader and tag
λ: signal wavelength in free space

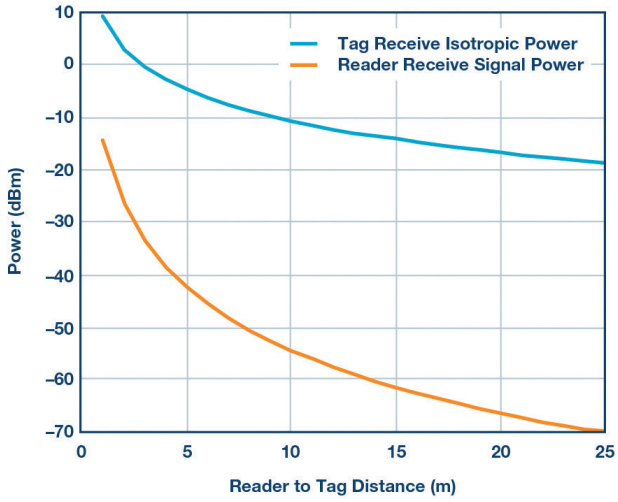


Figure 1. Forward and reverse link budget calculation.

As defined in GB/T 35786-2017 section 6.2 and section 6.5.2.2, Ptx is 30 dBm and the feeder cable insertion loss is less than 1 dB, so the real Ptx is around 29 dBm. An antenna of 10 dBi to 12 dBi gain is used in the field test, so Gtx is assumed to be 12 dBi. As for Grx, in an electronic identification of motor vehicles application, the reader normally uses the monostatic configuration, whereas a single antenna is used at the reader for both transmission and reception, so Grx = Gtx = 12 dBi. A tag antenna is usually similar to a dipole and it is reasonable to assume Gtag = 2 dBi. ηmod represents the modulation efficiency of the tag, which depends on the tag antenna matching and the tag impedance shift that occurs during modulation and it is reasonable to assume that $\eta_{mod} = -8$ dB. The center frequency is 922.5 MHz, so λ = 0.33

Table 1. Reader to Tag Physical MAC Layer Key Parameter Summary

PARAMETER	DESCRIPTION
Frequency Range	920 MHz to ~925 MHz
Occupied Bandwidth (OBW)	250 kHz
Channel Center Frequency	920.125 + 0.25 n (0 ≤ n ≤ 19) MHz
Adjacent Channel Leakage Ratio (ACLR)	Adjacent channel: <–40 dB Alternate channel: <–60 dB
Reader Maximum ERP	20 dBm at Channel 0 and Channel 19 33 dBm from Channel 1 to Channel 18
Reader Out of Band Emission	See Table 2
Modulation Type	DSB-ASK, SSB-ASK
Modulation Depth	30% to ~100%
Data Coding	Truncated pulse position (TPP)
Tari	6.25 μs or 12.5 μs

m. The system link budget shown in **FIGURE 1** is based on the previously described formulas and parameters.

To support the link range at 25 meters as defined in the standard, the tag sensitivity should be better than –18.7 dBm and the reader sensitivity should be better than –70.4 dBm. In the standard, the tag sensitivity requirement is defined as –18 dBm, which matches the analysis result quite well. However, the reader sensitivity requirement is defined as –65 dBm, which has considerable deviation when compared to the analysis result. This deviation may come from the tag antenna gain value. In electronic identification of motor vehicles applications, it is not necessary to design the tag antenna to be omnidirectional. Adding a reflector will result in an additional 3 dB antenna gain. And since the tag antenna gain (Gtag) is squared in Equation 2, the reader sensitivity analysis result will increase by 6 dB to be at –64.4 dBm. In this case, the analysis result will match with the standard requirement.

Self jammers in UHF RFID readers

In a UHF RFID system, the reader transmits a continuous wave (CW) signal to power the passive tags while simultaneously receiving the backscattered signal from the tag at the same frequency. Due to the poor transmitter-to-receiver isolation, the strong CW signal, together with the related transmitter noise, will leak into the receiver. Usually this leakage signal is called a self jammer (SJ) signal and this self jammer signal will

Table 2. Reader Out of Band Emission Requirement

	FREQUENCY RANGE	LIMIT (DBM)	MEASUREMENT BANDWIDTH	DETECTOR MODE
Maximum Output Power Mode	30 MHz to ~1 GHz	–36	100 kHz	rms
	1 GHz to ~12.75 GHz	–30	1 MHz	
	806 MHz to ~821 MHz 825 MHz to ~835 MHz 851 MHz to ~866 MHz 870 MHz to ~880 MHz 885 MHz to ~915 MHz 930 MHz to ~960 MHz	–52	100 kHz	
	1.7 GHz to ~2.2 GHz	–47	100 kHz	
Standby Mode	30 MHz to ~1 GHz	–57	100 kHz	
	1 GHz to ~12.75 GHz	–47	100 kHz	

Table 3. Type 2 Reader Key Performance Requirement

ITEM	LIMIT
Receiver Sensitivity	≤–65 dBm
Read Distance in Static Mode	≥25 m
Write Distance in Static Mode	≥12 m
Dynamic Identification Performance	Car speed ≤ 150 km/h: Successfully read information in a chip identifier data bank and vehicle registration data bank 150 km/h < car speed ≤ 200 km/h: Successfully read information in a chip identifier data bank

degrade the reader's sensitivity.

In an RFID reader for the electronic identification of motor vehicles, a directional coupler is normally used as the duplexer for the transmitter and receiver. The SJ signal occurs mainly because of the reflection of the antenna, the limited isolation of a directional coupler, and the reflection of the circuit connected to the coupler port.

To overcome this SJ signal issue, two methods could be used. The first one is to design a self jammer cancellation (SJC) circuit before the receiver LNA. The second method is to use direct conversion receiver architecture while using the same local oscillator (LO) that's used by the transmitter and the receiver. In this case, the self jammer signal will convert to dc at baseband, and then dc block capacitors will be used for the ac coupling of the signal. After this dc blocking point, the SJ signal is removed and the dynamic range requirement on the subsequent components is loosened. This means that enough gain could put on baseband to decrease the receiver noise figure (NF). These two methods could be used separately or combined. A typical SJC circuit is shown in FIGURE 2. [5]

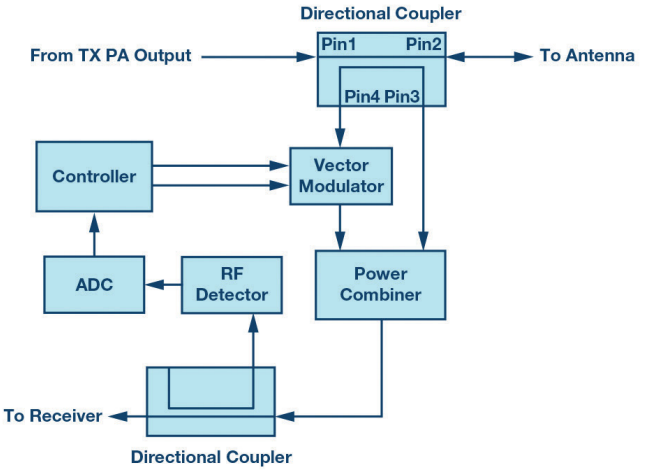


Figure 2. Typical self jammer cancellation circuit.

Reader key RF performance analysis

An SJC circuit including a UHF RFID reader RF front-end block diagram is shown in 3. Analog Devices' AD9963 integrates a dual-channel DAC and a dual-channel ADC. ADF9010 integrates a transmitter modulator, PLL/VCO, and receiver baseband filter and PGA. The demodulator ADL5382 is included on the ADF9010 evaluation board. The ADL5523 is used as an LNA as it provides low noise figures, high gain, and high linearity. The 75 dB high dynamic range RF detector LT5538 is suitable for the SJC RF power detector.

For the transmitter, in the digital domain, the signal should be low-pass filtered to meet both the frequency domain ACLR requirement and the time domain RF envelope requirement. And in the analog domain, both PA linearity and LO phase noise will impact the ACLR performance. The low-pass filtered, TPP coded ASK signal has around a

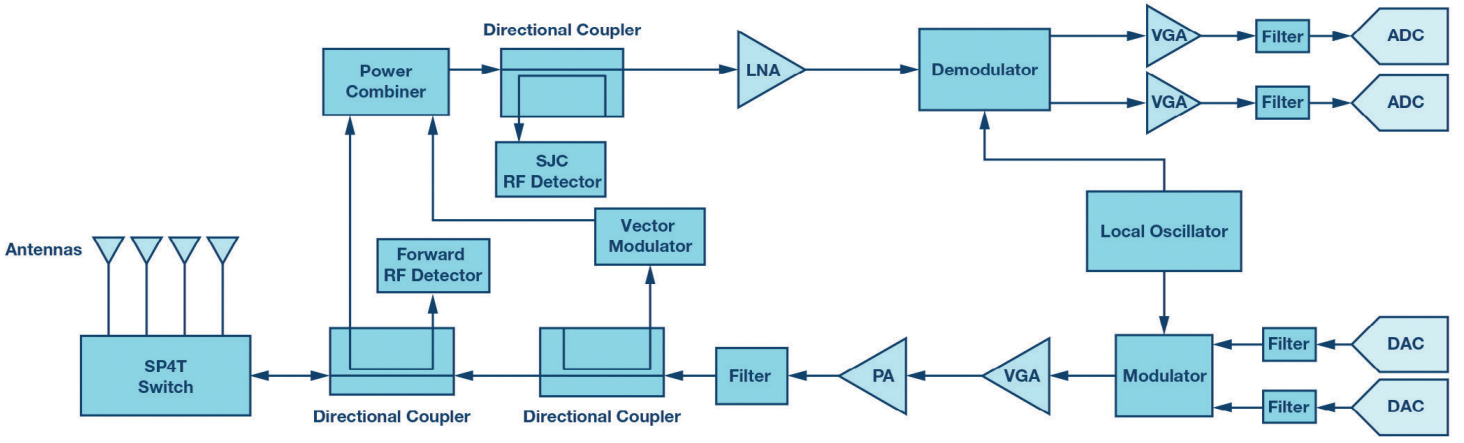


Figure 3. UHF RFID reader RF front-end block diagram.

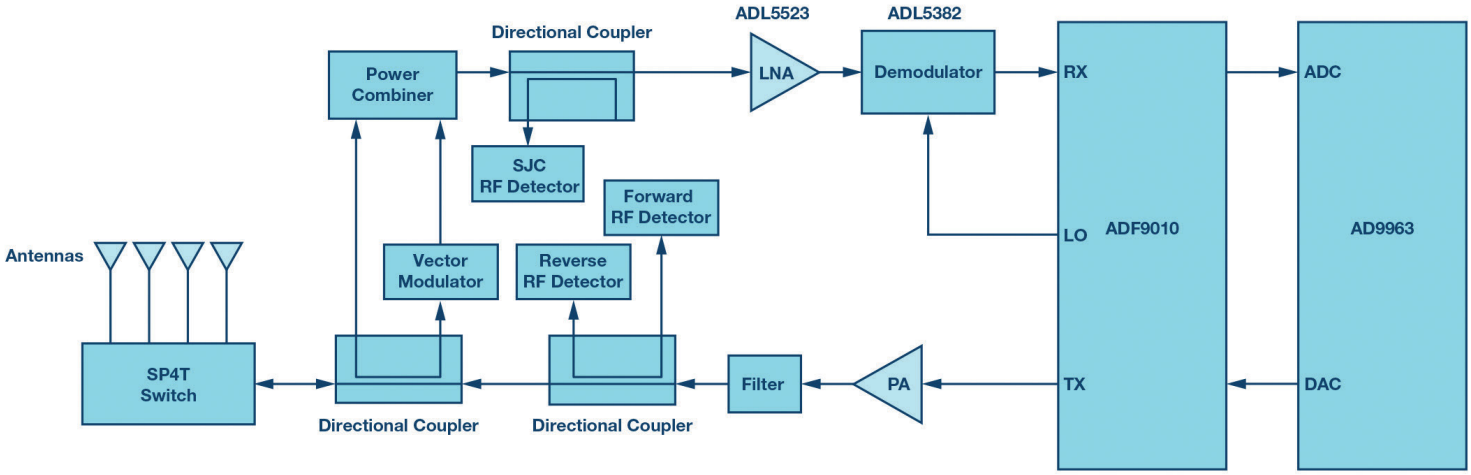


Figure 4. UHF RFID reader RF front end using the ADF9010 and AD9963.

2 dB peak-to-average ratio (PAR). The average PA output power is around 32 dBm, with a 1 dB margin, so a PA of more than 35 dBm P–1 dB should be chosen. As for the LO phase noise, the phase noise integration from 125 kHz to 375 kHz should be less than –40 dBc and the phase noise integration from 375 kHz to 625 kHz should be less than –60 dBc. As for the out of band emission requirement, an RF filter is needed to meet the requirement at the transmitter harmonics frequency. For the requirement close to working frequency such as the requirement of –52 dBm of 100 kHz measurement bandwidth at 915 MHz and 930 MHz, the RF filter normally has no attenuation yet, so the noise floor requirement for the modulator at 0 dBm output power is around $–52 – 10 \times \log10 (10^5) – 30 = –132$ dBm/Hz. And the phase noise requirement at a 5 MHz offset should be less than –132 dBc as well.

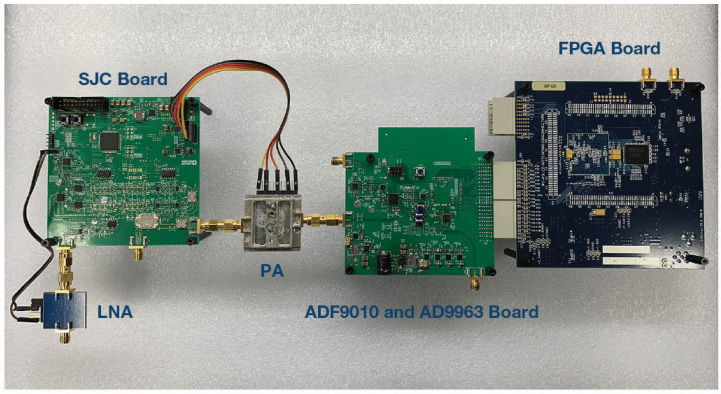


Figure 5. Test setup photo.

For the receiver, the receiver sensitivity is specified as –65 dBm in the GB/T 35786-2017 standard. It is assumed that the reader should meet this –65 dBm sensitivity at all possible data rates and that the back link frequency (BLF) of 640 kHz is the worst case. For an SJC that includes an RFID reader, the insertion loss from an antenna port to the SJC output is around 15 dB, so that the sensitivity requirement at the SJC output point is –80 dBm and assume the tag backscattered signal power not including dc is $–80 – 3 = –83$ dBm. The ASK modulation signal demodulation threshold is around 11 dB and the signal bandwidth of BLF 640 kHz uplink signal is 2.56 MHz. So the total NF requirement is $NF \leq –83 – (–174 + 10 \times \log10 (2.56 \times 10^6) + 11) = 15.9$ dB. This total NF requirement includes the impact of the noise of the receiver circuit after SJC, the SJC circuit induced noise, and the transmitter leakage noise. Assuming the delay is matched between the vector modulator signal branch and self jammer branch, which means both the CW self jammer signal and the transmitter leakage noise are cancelled. The transmitter leakage noise includes three parts: the phase noise, the amplitude noise, and the white noise. Normally the amplitude noise and the white noise will be cancelled to the noise floor of –174 dBm/Hz. For the residual phase noise, since the transmitter and receiver use the same LO, it will convert to dc during down conversion because of the range correlation effect. [6] In this case, the vector modulator branch noise will be the only extra induced noise. Assume the vector modulator branch noise floor is –162 dBm/Hz, so at the SJC circuit output, the effective NF is $–174 – (–162) = 12$ dB, then the NF requirement for the receiver circuit after SJC is $10 \times \log10 (10^{1.59} – 10^{1.2}) = 13.6$ dB.

Receiver sensitivity

This first implementation takes a two-chip approach that focuses on receiver sensitivity combining an RF transmitter with converter. Specifically, the ADF9010 is a fully integrated RF transmitter modulator, local oscillator (LO), and receiver analog baseband front end that operates in the frequency range of 840 MHz to 960 MHz. The AD9963 is a 12-bit, low power MxFE® converter that provides two ADC channels with sample rates of 100 MSPS and two DAC channels with

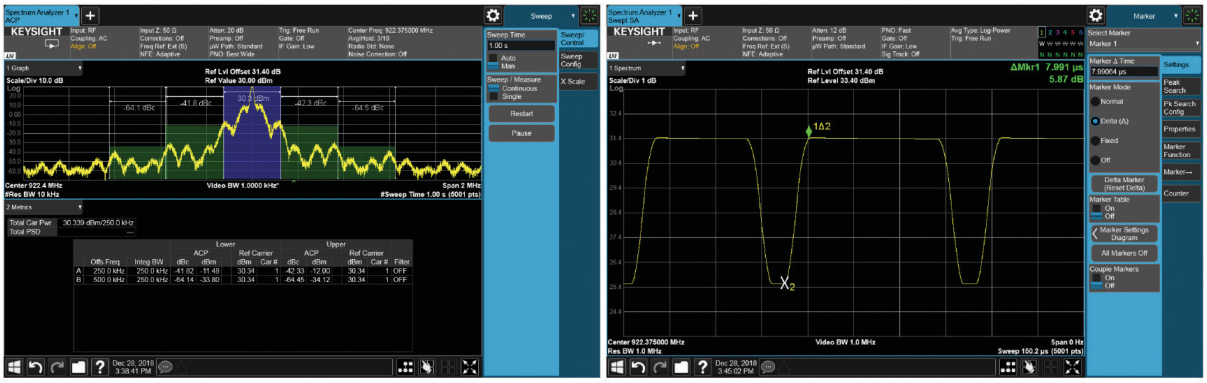


Figure 6. Transmitter test result.

sample rates to 170 MSPS. FIGURE 4 shows the block diagram for the UHF RFID reader RF front end\. The NF of the ADL5523 cascaded with the ADL5382 and ADF9010 receiver gain setting of 24 dB is less than 3 dB.

To implement the UHF RFID reader RF front end, both an SJC board including the adaptive SJC algorithm and the ADF9010 and AD9963 board are built. The ADF9010 and AD9963 board integrated the demodulator ADL5382 as well. The two boards are cascaded to test the transmit and receive system-level RF performance.

For the transmit test, a TPP coded, 50% modulation depth, DSB-ASK with Tari set to 12.5 µs RFID downlink waveform is built in Python® and downloaded to the FPGA board. The spectrum domain ACLR and time domain RF envelope are tested at the antenna port with a PA output power of 32 dBm. The test result is shown in FIGURE 5. For the ACLR test result, the adjacent channel is around –42 dBc which has 2 dB margin and the alternate channel is –64 dBc which has a 4 dB margin. For the RF envelope, the ripple is less than 1%, which has enough margin when compared to the 5% limit and the rise time and fall time are in the range limit of 1 µs and 8.25 µs.

For the receive test, a Tag simulator is built using the Analog Devices SPDT RF switch HMC545A and it is controlled by a microcontroller unit. The control pattern is an RFID uplink FM0 coded data list. An ASK decoding program is built by MATLAB.® By using this program

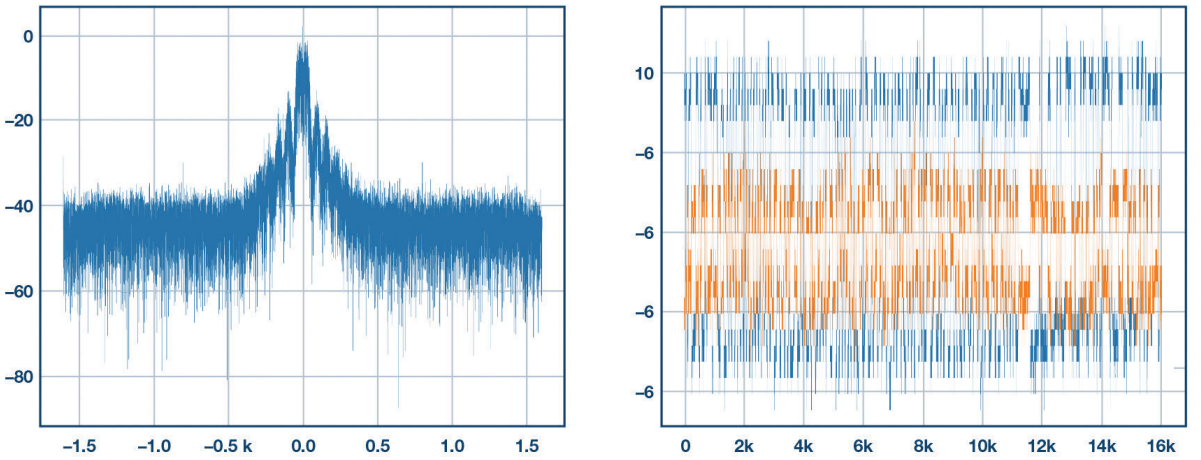


Figure 7. Received receiver data FFT plot and decoded data.

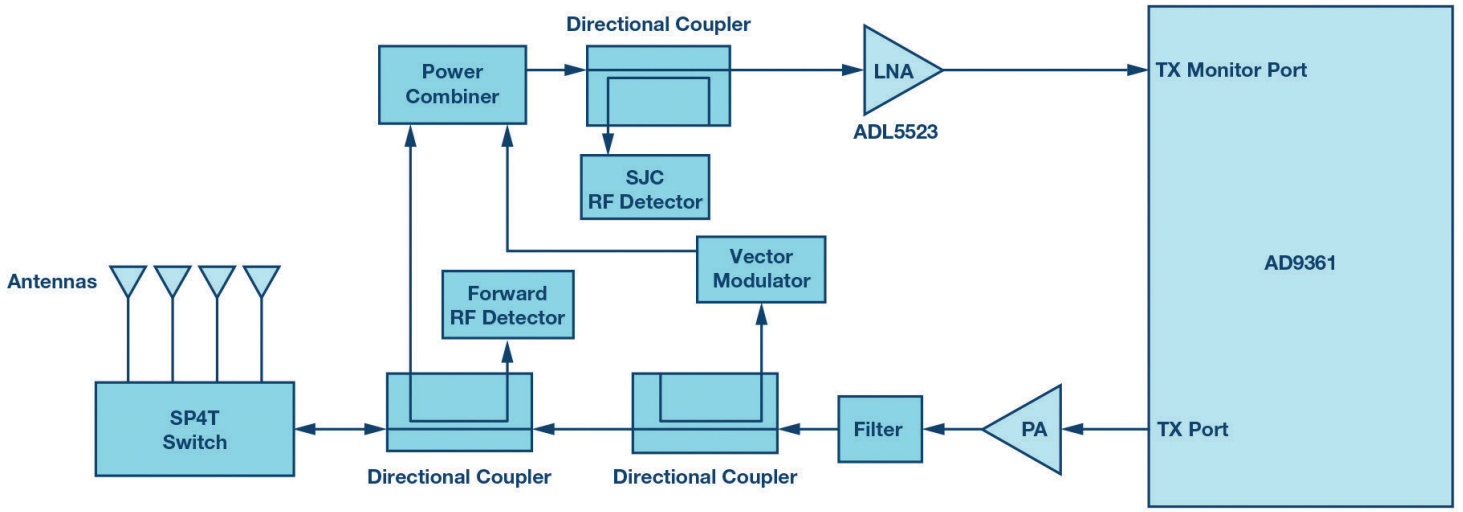


Figure 8. A UHF RFID reader RF front end using an AD9361 block diagram.

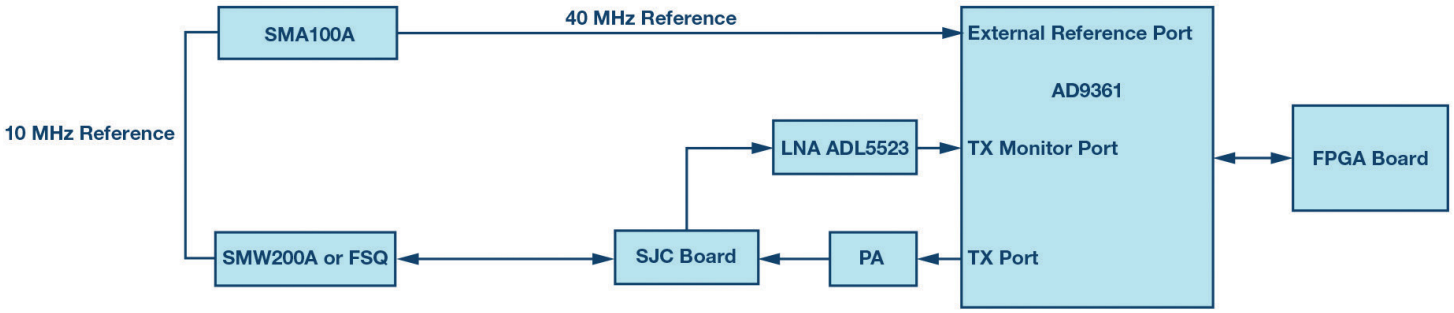


Figure 9. Test setup block diagram.

to decode the IQ and comparing it to the original data in the data list, the BER and the receiver sensitivity can be calculated. FIGURE 6 shows the received IQ data and the FFT plot. The figure shows that a –74 dBm RFID uplink signal with 320 kHz BLF was successfully decoded by the program.

Reducing design complexity, component count, and board space

For some applications, engineers may have the option to choose to trade off receiver sensitivity for design complexity, component count, and board space. In these cases, a more integrated RF transceiver like the AD9361 can be used that integrates all RF, mixed-signal, and digital blocks necessary to provide all transceiver functions in a single device. To implement a UHF RFID reader, the transmitter and receiver should use the same LO to take advantage of the range correlation effect, so the AD9361 transmitter monitor path, rather than the normal receiver path, will be used. The transmitter monitor path bypasses the internal LNA, so an external LNA, such as the Analog Devices ADL5523, is added. The ADL5523 is a high performance GaAs pHEMT LNA with 0.8 dB NF and 21.5 dB gain. The block diagram in FIGURE 8 shows the implementation of the UHF RFID reader RF front end. Compared to the discrete component implementation (FIGURE 4), this approach is significantly simplified.

The AD9361 baseband is dc coupling, rather than ac coupling. In this case, it is required that the SJC circuit can decrease the self jammer signal to a low enough level—for example, less than –35 dBm—that will not saturate the analog circuit. This allows the self jammer converted dc signal to be removed in the digital domain.

The AD9361 transmitter monitor path gain distribution is comprised of two gains: front-end gain (transmitter monitor gain) and receive low-pass filter gain (GBBF). The transmitter monitor gain could be set to 0 dB, 6 dB, or 9.5 dB. GBBF could be set from 0 dB to 24 dB with 1 dB step. With this flexible gain configuration, receiver AGC functionality could be implemented easily. For this UHF RFID reader application, a transmitter monitor gain setting of 3 dB and a GBBF setting of 6 dB is chosen. When the AD9361 gain setting is 3 dB, then the cascaded NF of the ADL5523 and AD9361's transmitter monitor port is around 12.6 dB. This setting has a 1 dB margin compared to the analysis requirement of 13.6 dB, while the digital domain power is –7 dBfs if the residual self jammer is –35 dBm.

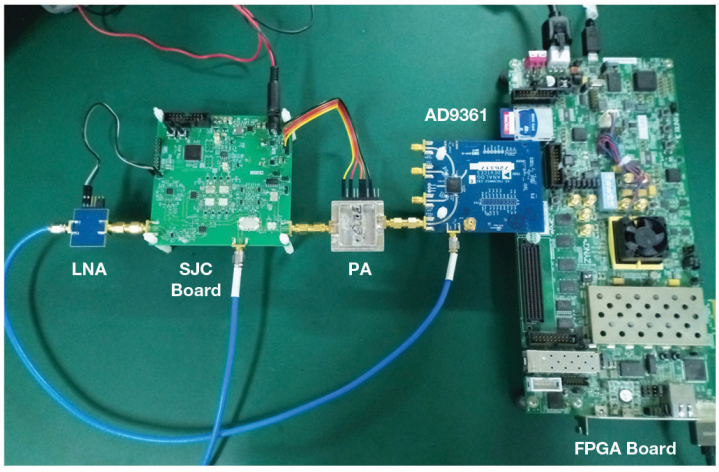


Figure 10. Test setup photo.

To implement the UHF RFID reader RF front end, an SJC board including an adaptive SJC algorithm is built. This is cascaded with AD9361 to test transmitter and receiver system-level RF performance. The test setup block diagram and photo are shown in FIGURE 9 and FIGURE 10.

The test result is shown in FIGURE 11. For the ACLR test result, the adjacent channel is around –42 dBc, which has a 2 dB margin, and the alternate channel is –61 dBc, which has a 1 dB margin. For the RF envelope, the ripple is less than 1%, which meets the margin requirements for a 5% limit. The rise time and fall time are in the range limits of 1 µs and 8.25 µs.

For the receiver test, an RFID uplink FM0 coded data list is built and downloaded to the signal generator SMW200A and then configured for SMW200A to transmit a DSB ASK signal with this data list. The AD9361 received IQ data are stored in the FPGA board and fetched to a PC using an FTP tool. An ASK decoding program is built by MATLAB. Using this program, decoded data is compared to the original data in the data list, which then allows for the BER and receiver sensitivity to be calculated. FIGURE 12 shows the FFT plot

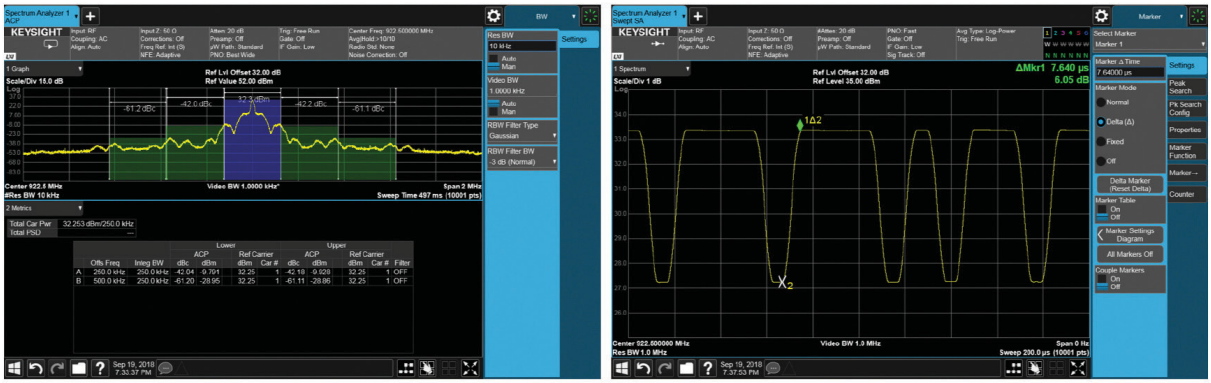


Figure 11. Transmitter test result.

and decoded data of the MATLAB program. It is tested that –65 dBm RFID uplink signal with a 640 kHz BLF is successfully decoded by the program.

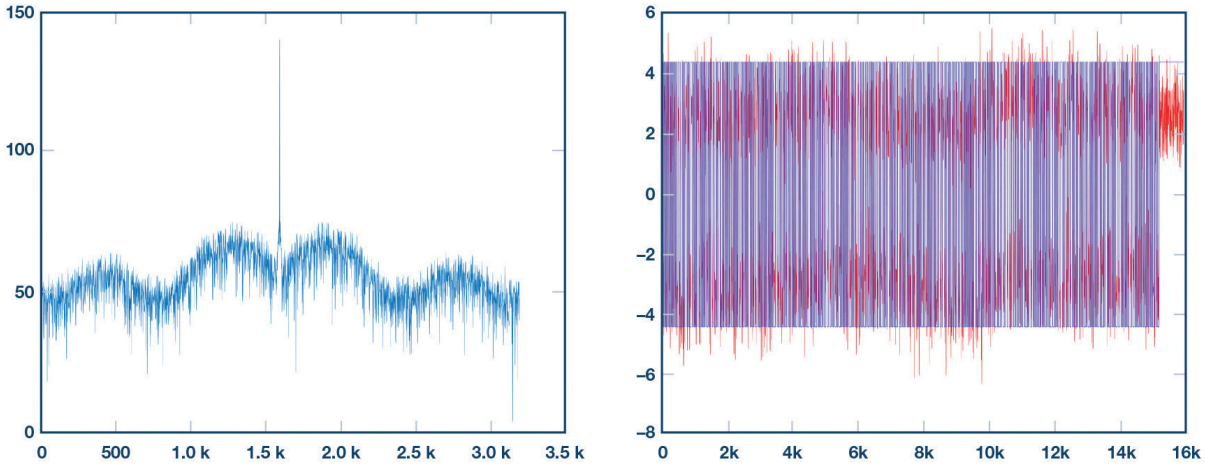


Figure 12. Received receiver data FFT plot and decoded data.

Engineers have many options when designing a UHF RFID reader RF front end. For applications that need high receiver sensitivity, a discrete implementation provides better results. For other applications, an integrated approach balances the trade-off of receiver sensitivity degradation with a greatly simplified design, lower component count, and reduced board space. Although the RF front end described in this article is application specific, both the analysis method and this front end itself are applicable for general UHF RFID reader solution.

REFERENCES

1. GB/T 29768-2013 "Information Technology—Radio Frequency Identification—Air Interface Protocol at 800/900 MHz."
2. GB/T 35786-2017 "General Specification for Read-Write Equipment of the Electronic Identification of Motor Vehicles."
3. Monza X Antenna Application Note. Impinj.
4. Gregor Lasser and Christoph F. Mecklenbrauker. "Self-Interference Noise Limitations of RFID Reader." *2015 IEEE International Conference on RFID (RFID),* June 2015.
5. Alírio Boaventura,João Santos, Arnaldo Oliveira, and Nuno Borges Carvalho. "Perfect Isolation: Dealing with Self-Jamming in Passive RFID Systems." *IEEE Microwave Magazine,* Vol. 17, No. 11, November 2016.
6. Byung-Jun Jang and Hyun-Goo Yoon. "Range Correlation Effect on the Phase Noise of an UHF RFID Reader." *IEEE Microwave and Wireless Components Letters,* January 2009.

About the Authors

Van Yang is a system applications engineer manager at ADI Shanghai. Van joined ADI in 2015 as an FAE for industrial and healthcare customers support. Prior to joining ADI, Van worked as an FAE at TI for over four years. He graduated from Huazhong University of Science and Technology with a master's degree in communication and information systems. He can be reached at *van.yang@analog.com*.

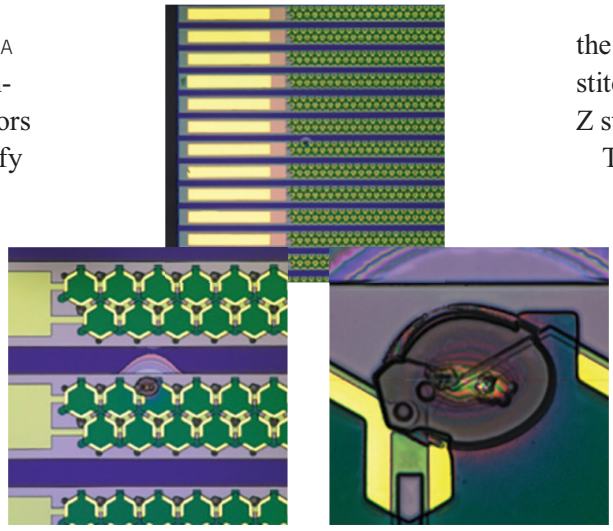
Eagle Zhang is a field applications manager at ADI Shenzhen. Eagle joined ADI in 2001. He started his role as field applications engineer at ADI, and then worked as an ADI China core market technical support manager and ADI southern China field applications manager. During his time as a field applications manager, Eagle built the ADI southern China field technical support team. Eagle earned his bachelor's degree and master's degree in engineering thermal physics from Tsinghua University. He can be reached at *eagle.zhang@analog.com*.

Aaron He is a system applications engineer at ADI Shanghai. Aaron joined ADI in 2017. Prior to joining ADI, Aaron worked as a senior RF engineer at Ericsson. He has more than 10 years of experience working on wireless communication base station design, integration, and production test system development. Aaron received his B.S. in telecommunication engineering from Xi'an Jiaotong University in 2001, and his M.S. in microwave engineering from Huazhong University of Science and Technology in 2006. He can be reached at *aaron.he@analog.com*.

Going from Macro to Micro in Semiconductor Inspections

HAMISH ROSSELL, Olympus Product Applications Manager, Scientific Solutions Group, Waltham, MA

Macro-to-micro viewing capabilities of digital microscopes can help overcome common challenges in the inspection process.



MICROSCOPY HAS ALWAYS PLAYED A pivotal role in the semiconductor industry as inspectors need to quickly and efficiently identify and image defects on wafers in a repeatable way. Until recently, the wide range of contrast techniques and magnifications needed to complete these inspections made the process difficult.

With optical microscopes, semiconductor inspections required multiple systems, skilled operators, and complex work instructions to achieve the necessary results—taking up valuable space and resources. However, with modern systems inspections are faster and more repeatable than ever before.

In this article, we will explore how the powerful macro-to-micro viewing capabilities of digital microscopes overcome common challenges in the inspection process.

The challenges of using conventional microscopes for semiconductor inspections

Conventional compound microscopy inspection relies on switching from one fixed magnification to another, and the lowest magnification has a field of view around 8 mm. To see more of your sample, you would need to manually stitch the image, significantly modify the microscope, or move the sample to a stereo microscope.

In contrast, the largest field of view available on the most recent DSX series digital microscope is 19.2 mm. This is usually enough to capture an entire overview in one image. Keep reading to learn more about the macro observation benefits of digital microscopes.

Macro

The capabilities of digital microscopes extend beyond a wide magnification range. Adding motorized components to the microscope extends the system's functionality even further.

For larger samples that extend beyond your system's field of view, you can quickly and easily capture images of the entire sample using the automated stitching function and a motorized X, Y, Z stage on a low-power objective.

The result is a fully focused, high-resolution image you can use to show larger features of interest or provide context for further analysis. With a maximum stage travel range of 300 mm × 300 mm, entire wafers can be imaged quickly to provide an overview map.

This map will both show macro scale defects and serve as a navigation aid as the operator increases magnification. It may be useful to initially image at this low magnification level using different observation techniques to highlight contamination.

With the ability to work and image with different techniques at the press of a button, you can save significant time in your inspection workflow.

Micro

Typically, the next step following macro observation is to change systems or disassemble and reassemble some other digital microscopes. This causes a loss of position and focal point.

Using the Olympus DSX1000 digital microscope, you can exchange the objective lens without the risk of exposing the camera to contamination or the sample to unnecessary movement, all

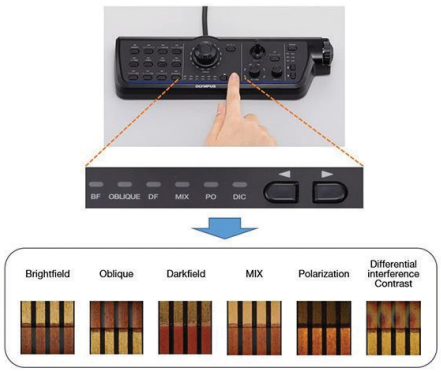


Figure 1. Easily switch contrast techniques by pressing a button on the optional controller for the DSX1000 digital microscope.

while maintaining position data and focus. The captured macro map can now be used to quickly navigate the sample and provide a view of multiple image acquisition areas.

With the addition of zoom optics, you can use the full optical range of each objective. To quickly magnify a feature of interest, there is no need to change the objective or rely on digital zoom. The DSX zoom optics will bridge multiple conventional objective magnifications, reducing the need for an objective change, parts to store, and movement over the sample.

It is nearly impossible to present a perfectly planar sample perpendicular to the viewing plane—something that has caused issues for microscopists for decades. The conventional methods to do this were either significantly compromised, such as using an inverted microscope system, or used complex and space-consuming vacuum and levelling systems.

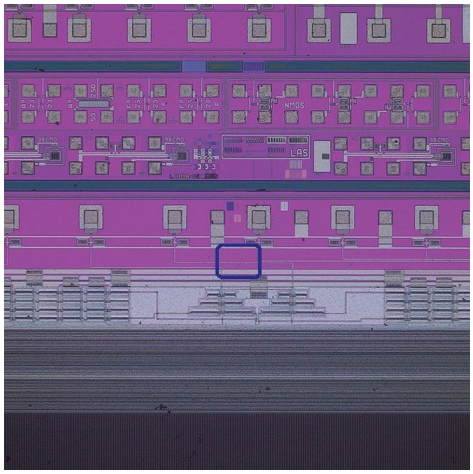


Figure 2. Brightfield observation: Defect detection at low magnification (70X).

This is not the case in modern digital microscopy, where developments in 3D acquisition and autofocus have made it possible to acquire and analyze even the roughest or most uneven samples using a wide range of techniques.

On the DSX1000 digital microscope, a field of view of as little as 55 µm is achievable with submicron resolution. This can be done using an Olympus UIS2 objective with planar and apochromatic correction and a 0.95 numerical aperture (NA). In other words, an objective built to provide a perfectly flat, undistorted image with color correction beyond the visible spectrum and the highest level of detail and resolution from an air objective.

Advanced analysis

As mentioned earlier, a big plus of modern digital systems is that they can take the guesswork out of imaging by combining repeatability with ease of use. In the DSX1000 digital microscope, multiple observation techniques (brightfield, oblique, darkfield, MIX, polarized light, and true differential interference contrast) are available at the touch of a button. As a result, finding the correct technique is a straightforward process, especially when using the desktop controller to increase ergonomics and efficiency.

To choose your observation method, you can either sequentially select it or find the optimal one using the Best Image function (**FIGURE 1**). With this function, the system automatically images the sample under preset conditions and displays thumbnail images of your sample under the different methods. This makes it quick and easy to choose the right one, since you can simultaneously view them and choose the method that best shows the features or defects.

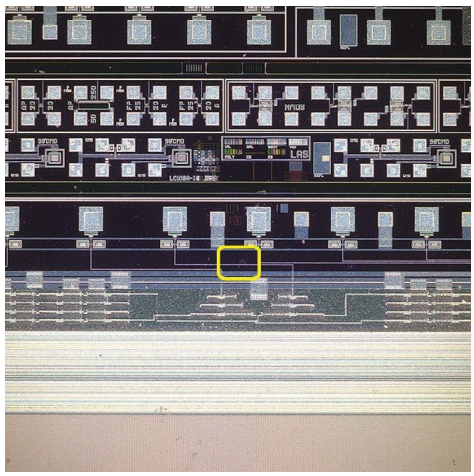


Figure 3. Darkfield observation: Defect detection at low magnification (70X).

Now let's put these observation methods into practice by exploring a typical use case.

Example of a typical use case

FIGURE 2 is a brightfield image of a semiconductor wafer with a defect. This defect is hard to see under some conditions, so finding the correct observation method can be time-consuming on a conventional system.

By selecting darkfield, you can see that the defect is slightly more visible, as shown in **FIGURE 3**.

Certainly, the defect is more visible—but what about using DIC? With the push of a button, the microscope will insert the polarizer, analyzer, and motorized and encoded DIC prism needed for this observation

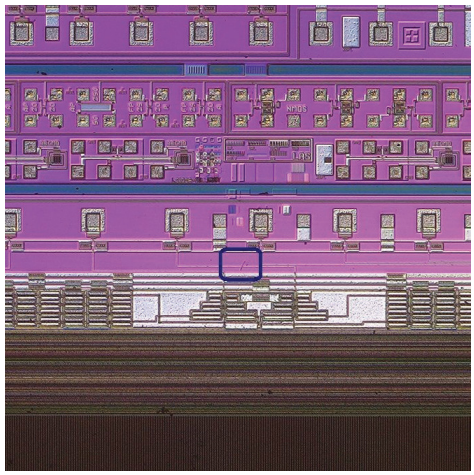


Figure 4. DIC observation: Defect detection at low magnification (70X).

method (**FIGURE 4**).

In DIC observation, the defect is now highly visible, and the conditions required to achieve this are recorded in the image. These conditions can be saved with the image and as an option for the operator to use in the future (**FIGURE 5**).

Guaranteed accuracy and precision

In semiconductor inspections, the integrity of data and measurements is critical. Without a guaranteed measurement, the microscope is simply a subjective observation tool. The DSX1000 digital microscope has integral calibration tools that, when performed by an Olympus certified technician, will calibrate every objective at every magnification to a traceable standard.

While the focus of this piece has been on semiconductor inspection, the versatility of the digital microscope is limitless. From macro inspection of test pads, to micro inspection of contamination, to optional fluorescence observation, digital microscopes are assets to any inspection environment.

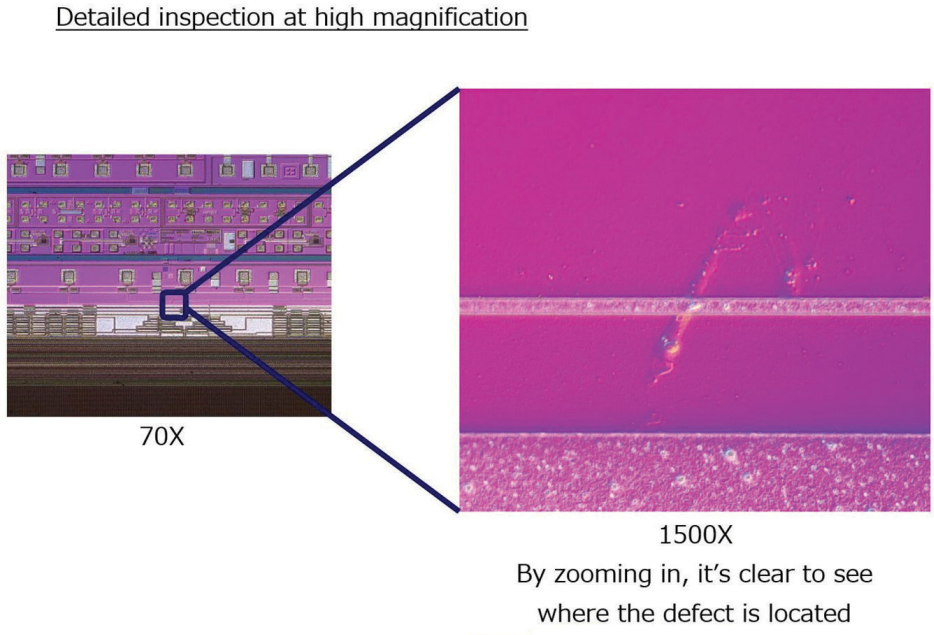


Figure 5. Defect viewed at 1500x under DIC observation

Continued from page 54

memory interfaces like NAND, DDR4, and SPI NOR

6. Veloce Codelink to run and debug the full system firmware at offline speeds of about 50 megahertz

The result is a seamless hardware and software debug environment with 100% visibility of all signals and software.

Conclusions

Several applications, including image recognition, edge computing, AI/ML, real time analytics, data base query, 5G and self-driving vehicles benefit from CSD making the new device a must-have peripheral for intensive storage requirements.

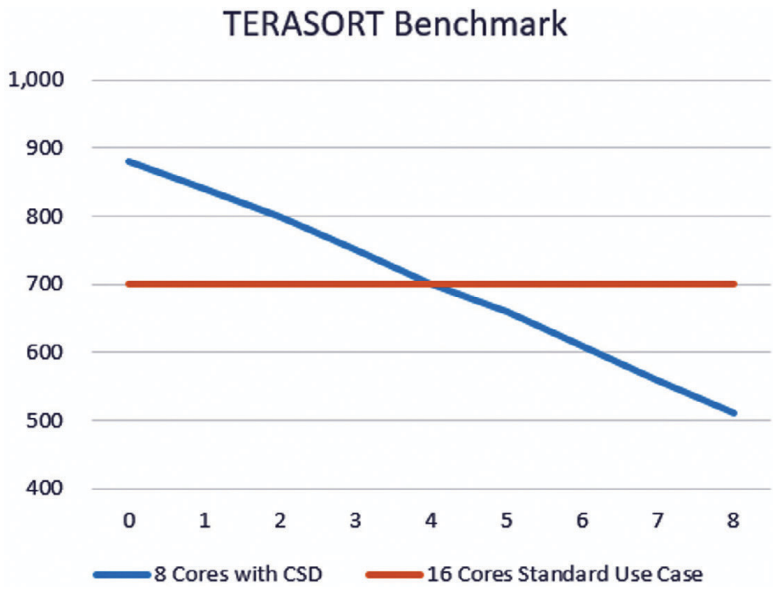


Figure 2. Deployment of SSDs versus CSDs using the TERASORT benchmark highlights where performance is accelerated with additional CSDs. *Source: Mentor, a Siemens Business*

CSD's complexity challenges the traditional pre-silicon verification approach increasing the risks of missing the critical time-to-market window. A new system verification environment based on virtualization of the testbed to drive a CSD design meets the challenge with full firmware validation at high speed to accelerate time to market and perform architectural explorations.

About the authors

Ben Whitehead is the product and solutions expert for the storage market in the Emulation division at Mentor, a Siemens Business.

Dr. Lauro Rizzatti is a verification consultant and industry expert on hardware emulation. Previously, Dr. Rizzatti held positions in management, product marketing, technical marketing and engineering.

Continued from page 40

an array of grounding straps can help reduce electrostatic charges but are limited in effectiveness.

Answering a call from manufacturers to more effectively remove electrostatic charge from the entire flow path, a new solution that implements carbon stripes on the inside of the tubing, as well as in fittings, valves, and other components, was developed that helps in reducing static charge accumulation on the liquid, all without affecting the cleanliness of the fluid path. This approach provides an end-to-end fluid handling system to overcome the many challenges presented by electrostatic hazards, reducing safety and financial risks.

Early uses of this continuous conductive system in fabs show promising results. The system has proven to successfully reduce the accumulated electrostatic charge from both the media flowing through the tubing as well as any charge that might build up inside the tube, mitigating not only safety risks, but damage to the wafers themselves.

REFERENCES

1. F. von Trapp, *Ensuring Device Reliability in Automotive Electronics,* February 13, 2020, 3D InCites.
2. *The Fight for Purity in Chemical Manufacturing,* Entegris White Paper.
3. M. Caulfield, J. Leys, J. Lindner, B. Reichow, *Reducing Electrostatic Discharge in Semiconductor Fluoropolymer Fluid Handling Systems while Maintaining Chemical Purity,* Entegris Technical Note.
4. EOS/ESD Association, *Electrostatic Discharge: How Important Is It?* InCompliance, January 31, 2019.

About the authors

Brett Reichow is the VP of the Fluid Management business at Entegris, Inc. within the Advanced Materials Handing (AMH) division. The Fluid Management business provides chemical delivery component solutions to the semiconductor and life sciences markets including valves, fittings, tubing, instrumentation, dispense, and container products. Brett has 20+ years of leadership experience in engineering and business management providing solutions to semiconductor equipment manufacturers.

Mark Caulfield is a senior technology manager at Entegris, Inc. supporting the Advanced Materials Handling (AMH) division. In this position, he leads a team of applications engineers who provide technical support for AMH products. Mark has 30+ years of field experience in semiconductor processing, specifically supporting semiconductor equipment and process applications. Mark has held various leadership positions at an OEM cleaning equipment manufacturer and at Entegris, www.entegris.com.

Continued from page 48

which, in turn, demand new power solutions. Going forward, a bigger process power toolbox is being used to overcome challenges such as speed of response, ion energy distribution, and enabling measurement and control with predictive insights. These new solutions are being adopted in ways that deliver these advanced capabilities at low costs and compact sizes, as power continues to be a leading process enabler.

Process power control and system-wide synchronization is becoming critically enabling. In **FIGURE 12**, semiconductor Wafer Fab Equipment (WFE) process power elements are shown for plasma-based processes. The core importance of power also applies to Ion Implantation and E-Beam inspection and SEM processes, where high-voltage power is fundamental to controlling electron-beam performance. Temperature control and measurement is also shown due to its tight coupling with the process power delivery.

It is interesting to think that this series of articles on process power has most likely been read from remote locations as many in our industry work from home. Ironically, this shift to working from home has increased the demand for data services, which has driven the need for more compute, storage, connectivity, telecommunication, and data capacities. This, in turn, has driven the need for more high-performance computing and memory chips. This, finally, brings things full circle back to the advances in process power that are enabling chip manufacturing steps.

Many of us working from home are also newly remotely linked to our work development facilities and labs. This has additionally sped adoption of Machine Learning and Artificial Intelligence as we have used IIoT devices to interact with our equipment, provide predictive insights and even optimize our systems from a distance. This, too, is mirrored in process power as more demanding requirements and higher complexity is increasing reliance on using power parameter information to predict and optimize processes – and making advanced semiconductor manufacturing possible.

REFERENCES

1. Associated Patents include US Patent 8416008B2 , 9,287,092, 9,287,086, 9,208,992, 9,767,988, 9,362,089, 9,105,447
2. Wendt, et. al., J .App. Phys, 88, 2, 2000
3. Patent pending. Chamber clean is one of many applications that leverage PowerInsight by Advanced Energy.
4. Carter, et. al., Proceedings ALD/ALE Technical Symposium, 2018
5. Patent pending. Chamber clean is one of many applications that leverage PowerInsight by Advanced Energy.

Continued from page 68

shared right now among the SEMI members and across the industry.

At a micro level, companies are giving every used cleanroom gown its own bag to limit transmission. Workers are propping open interior doors to avoid touching and prioritizing the use of automatic doors wherever possible. Unexpected shortages of facemasks have spurred a rush to qualify second-source providers. To limit cross-shift exposure, most member companies have begun staggering shifts to rotate schedules and insert time gaps between shift transitions. Employees admitted on sites are not allowed to move among buildings, and most members change the building entry location every day.

At a macro level, SEMI asked 16 governors and the chairs of the National Governors Association, U.S. Conference of Mayors, National League of Cities, and National Association of Counties to consider the semiconductor industry "essential business." This was coordinated with the Semiconductor Industry Associations worldwide to reinforce the importance of the microelectronics business in many countries.

And there are more lessons on the list.

How else is the industry specifically building on what's been uncovered? Among takeaways already being surfaced during the pandemic is the need for accelerated development of standards with ever-more comprehensive cybersecurity specifications. In an increasingly competitive world, such standards will help to sustain uninterrupted progress during a next crisis, including another quarantine. Underscored by the pandemic, two draft standards are underway.

Intel and Cimetrix are leading SEMI Draft Document 6566, *Specification for Malware Free Equipment Integration*, refining it to define protocols for pre-shipment scans of equipment as well as various types of ongoing support, including file transfers, maintenance patches, and component replacement. Second, TSMC and ITRI are leading SEMI Draft Document 6506, *Specification for Cybersecurity of Fab Equipment,* which defines a common, minimum set of security requirements for fab equipment.

Turning back to the big picture, while market demand may pause, most companies in the supply chain should continue to be in solid shape – partly due to the determined analysis, refinement and certification of standards designed to avoid interruptions during both good times and bad.

As my former boss, Bob Noyce, said, "It is impossible to do business in the world today without having standards that both the customers and the suppliers agree to."

One more of the key lessons repeated during this worldwide lockdown is that the microelectronics industry represents the very core of the most productive half century in history. And, with the collaboration that is our hallmark, that we collectively will be the ones to carry the torch into the future.

For this reason, the upcoming Virtual SEMICON West, July 20-23, will feature the broadest array of futurists – from both within, and beyond, the industry – to shed much light on what will be the strategies and tools for managing the next disruptions.

In addition to the specific lessons and actions I've noted here, we hope you'll join us next month to hear – and tell – what the next solutions may look like.

Continued from page 18

will ultimately have to build fault-tolerant quantum computers that can correct any errors that result from this instability. Lux Research doesn't expect a fault-tolerant quantum computer to become available for at least 10 years.

"Quantum computing is not currently providing business value that could not be achieved with today's existing computers, and it's not clear when it will. For this reason, we advise companies not to make it a priority right now, unless your work is already bottlenecked by today's supercomputing," says Roberts. For companies that must pursue quantum computing now, research projects that estimate when quantum advantage can be achieved will be key. Lux Research advises forming partnerships for these projects based on the level of internal expertise, as this greatly affects which players will be most helpful for your unique projects.

2020 IEEE International Electron Devices Meeting to Be Held Virtually

The organizers of the upcoming 66th annual IEEE International Electron Devices Meeting (IEDM), the world's premier forum for technological breakthroughs in semiconductor and electron device technology, have decided to hold the conference virtually this year given the ongoing uncertainties posed by the prevailing COVID-19 pandemic.

The virtual 2020 IEDM is scheduled for December 12–16, 2020, and both live and recorded sessions will be used to showcase and discuss the world's best original work in all areas of microelectronics research and development.

"The IEDM Executive Committee has decided that in the interest of prioritizing the health and safety of the scientific community, a virtual approach is the best option for this year," said Dina Triyoso, IEDM 2020 Publicity Chair and Technologist at TEL Technology Center, America, LLC.

More details about the 2020 IEDM will be announced in late summer.

Unsung Heroes Shine from the Chip Industry

DAVE ANDERSON, President of SEMI Americas

"WHAT DO YOU SEE CHANGING TO BE better prepared before the next crisis?" That's an increasingly popular – and pressing – topic of conversation as the world lurches through the first half of this eventful year.

Recently, the Wall Street Journal published a report on a few of the unsung technologies playing a critical role during the terrible pandemic, including SEMI standards. Yes, standards.

As part of the ongoing preparation for the next waves of challenges and opportunities – including stop-work global crises – standards are on a roll. That's because the stakes are so high.

During this global shut-down, fab production nevertheless has been officially deemed "essential business," but social-distancing sharply limits human presence in the factory. As we all know, high-tech manufacturing, and the markets served, can be highly vulnerable to a hiccup in the factory tool set, where time is of the essence when repairs are needed.

It's in this context that standards have risen to the occasion when production has been most susceptible during these past months of skeleton crews. Factory automation, including the port where information securely flows in and out, the data packets that are exchanged, the interface that allows all the data to be collected and analyzed – and then implemented in AI-like fashion – all rely on standards.

As a quick recap, the industry's first collaborative deliverables a half century ago were standards for early data analysis, which led to increasingly rapid improvements in fab operations. Today that's called machine learning and AI.

As factories became more sophisticated in their automation and development of equipment communications interface protocols during the past few decades, chipmakers didn't want a hodgepodge of interfaces and languages – various remote connectivity solutions across different suppliers – nor the accompanying accidental or intentional cybersecurity threats. Different recipes for an array of chip types were to be accommodated, but cost efficiency and yield were a top priority across the business.

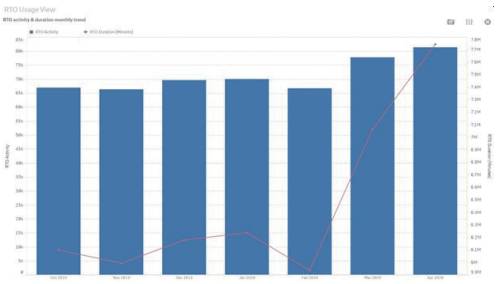


Figure 1. The red line in the chart indicates that duration of time spent in work sessions of Remote Tool Operation (RTO) software climbed 3-4x during March and April compared to previous non-quarantined months. *(Image credit: Telit IoT May 2020)*

SEMI and its members organized to address the issue. SEMI Standards committees – today numbering 5,000 experts from member companies, who have implemented more than 1,000 standards – teamed from across the global supply chain. Led today by James Amano, senior director of International Standards, SEMI guides the industry's disparate interests to develop broadly agreed international data, communications, traceability and cybersecurity standards.

With only a bare minimum of people allowed into the quarantined cleanroom of 2020, fast corrections have been able to remain the norm in environments like the leading-edge lines where recipes or deliverables can change every 3-4 months. If we didn't have precise remote support, operations would suffer. As a result, due in great part to standards previously implemented, we're so strategically automated that the fabs are performing well despite the pandemic.

For example, suppliers of software that is used to diagnose and manage semiconductor manufacturing tools remotely, saw its usage more than double between February and April (FIGURE 1). It has remained at record-high usage levels during May and into June. A lot of folks have continued to handle vital diagnostics and control, but from home.

Equipment makers never missed a beat as their technicians peered from offsite locations into customers' active fabrication lines. Fab owners have reaped the benefits of remaining on schedule and in the black.

But the pandemic has revealed there's still room for improvement. For instance, specific lessons from those with cleanrooms are being Continued on page 66

1000 SEMI STANDARDS & COUNTING

From wafers to packaging to everything in between...

The microelectronics industry counts on SEMI standards to drive innovation in every aspect of product design and manufacturing.

Explore the wide variety of SEMI standards and get involved in setting standards for your segment with our open, consensus-based process.

Learn more and get involved at:

semi.org/standards

semi™