EE Times 50 1972 2022

THE NEXT SILICON FRONTIER

EE TIMES 50TH ANNIVERSARY SPECIAL EDITION

CONTENTS

FOREWORD

EE Times is 50, and to celebrate the milestone, we invite you to rewind, play, and fast-forward through electronic engineering times.

The integrated circuit was only six years old when Gordon Moore observed — nearly 60 years ago — that the transistor count per silicon die area doubled roughly every year (which he later revised to every two years). Moore's Law transistor scaling pushed the semiconductor industry forward, bringing ever smaller, denser, and more powerful generations of chips to market that touched every facet of life and thereby changed society itself.

In 1971, Intel's first "micro-programmable computer on a chip," the 4004 microprocessor, held 2,300 transistors. By 2010, an Intel Core processor with a 32-nm processing die and second-generation high-k metal gate silicon technology held 560 million transistors. In 2021, IBM's 2-nm chip could fit up to 50 billion transistors while achieving 45% higher performance or 75% lower power consumption than state-of-the art 7-nm chips in production.

Node shrinking is getting even more extreme with TSMC's 1-nm process node, developed in collaboration with the Massachusetts Institute of Technology and the National University of Taiwan. But can we go beyond 1 nm? Convinced that "Moore's Law will not stop," imec is working on a sub-1-nm process, and its latest transistor roadmap extends to 2036. With industry partners ASML, TSMC, Intel, Samsung, and many others, imec is preparing to move beyond nanometer geometries to the angstrom era of semiconductors.

Despite the continued progress in traditional transistor scaling, the semiconductor industry has reached an inflection point. The demand for faster, smaller, smarter, and more energy-efficient chips calls for new design and manufacturing paradigms.

While it is still difficult to imagine a world without silicon, its physical limitations have shifted the focus to other semiconductor materials and compounds. Wide-bandgap semiconductors like silicon carbide and gallium nitride have received considerable attention, as they enable higher frequencies, higher voltages, and more complex electronic products than silicon.

To collectively prepare for the future, the "EE Times 50th Anniversary Special Edition: The Next Silicon Frontier" explores emerging themes like:

- Are carbon nanotubes, graphene, and other 2D materials change agents for the post-silicon age?
- Can photonics light the path to high-speed, energy-efficient data processing?
- Is heterogeneous integration opening new possibilities in IC packaging?
- Will neuromorphic technologies pave the way for brain-like computers?
- How does quantum computing work, and why is it important?
- Is sustainable electronics attainable?

By asking and answering these questions, this eBook reflects 50 years of journalistic tradition at EE Times.

The name "Electronic Engineering Times" was first used on Sept. 11, 1972, according to the Official Gazette of the United States Patent and Trademark Office. In pursuit of the mission set forth by founder Gerard Leeds — connecting the global electronics industry — EE Times' reporters and editors quickly established the publication's voice, which echoed far beyond California's Silicon Valley to other meccas of electronics innovation, from Tel Aviv and Munich to Bangalore and Shanghai.

Over the decades, EE Times has endured by reinventing itself, just as the industry we cover has done. We have been agile, responsive, and innovative so that we would continue to meet our audience where they were. In the 1990s, EE Times was the first trade publication to go digital. In the early 2000s, we expanded globally, with a constellation of local-language websites in Asia and Europe. We developed eBooks in the 2010s and embraced books, podcasts, special projects, and virtual events in the 2020s.

What hasn't changed are our journalistic values and ethics. We still work to inform by going beyond the obvious, evaluating the different angles of a story, and contacting the right sources so that "the news" becomes a meaningful analysis of and for the industry. We tell the stories of the visionaries, pioneers, and innovators who have architected the digitization of society and who are preparing for its future. We forge intergenerational bonds and create intercultural connections to strengthen the global electronics community.

No engineer is an island. Innovations can only succeed if team members work toward common goals, if they meet a need or solve a problem, and if end users embrace them.

At EE Times, no editor is an island. Over the course of half a century, our organization would not have sailed through the still and the storms without business leadership and the hundreds of talented editors, copy editors, graphics designers, audience developers, doers, and go-getters who have populated our masthead.

EE Times' 50th-anniversary celebration is also yours, dear EE.

Together, we connect the dots between the past and the future. We connect science and society through technology.

Together, we write the history of the electronics industry.

— *Anne-Françoise Pelé*
EDITOR-IN-CHIEF, **EE TIMES EUROPE**

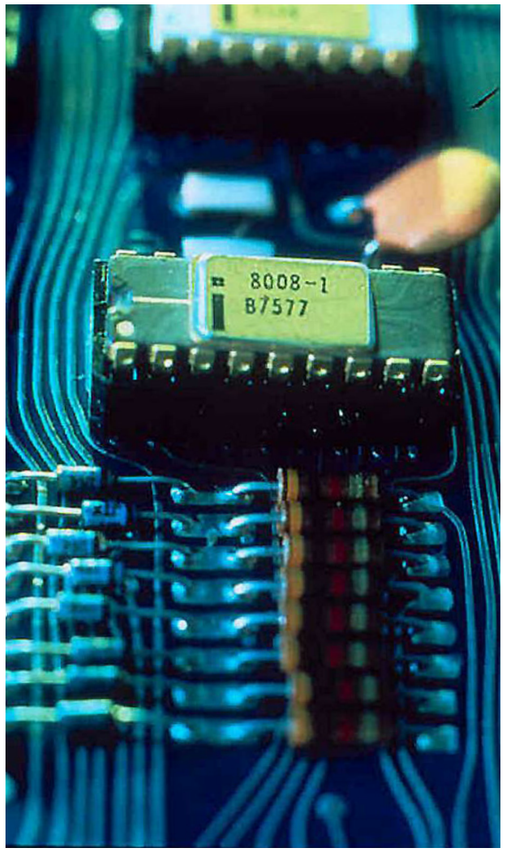
CHAPTER ONE:

THE FOUNDATION

1972: THE YEAR EVERYTHING CHANGED

BY BARBARA JORGENSEN

“The Godfather” was playing in theaters everywhere, and “American Pie” was riding the FM airwaves. China was completely isolated from the West, and no one could imagine carrying a phone in their pocket. It was 1972, the year everything changed.



Intel’s 8008 microprocessor
(Source: Intel)

Hindsight being what it is, a series of seemingly unrelated events that year set the stage for the electronics industry we know today. U.S. President Richard Nixon visited China that year. Atari released the first-ever video game, “Pong.” Intel introduced the 8008 processor, and the first epic rivalry for dominance in the computing market was just beginning.

“I think the early ’70s were pivotal for the electronics industry,” said Michael Knight, president of TTI Inc.’s Exponential Technology Group, SVP for corporate business development, and self-acknowledged high-tech history buff. “Gaming was the next killer app. TV and radio were the first killer apps driving electronic components, but back then, computing was a brute-force timesharing system. The early ’70s set the stage for modern electronics. You had the first truly programmable microprocessor out of Intel.”

The high-tech battleground at that time was handheld calculators, and the combatants were HP and Texas Instruments. As the story goes, HP co-founder Bill Hewlett issued a challenge to his engineers in 1971: Fit all of the features of their desktop scientific calculator into a package small enough for his shirt pocket. They did. The HP-35

(Source: TTI Inc.)

could not only add, subtract, multiply, and divide but compute trigonometric functions, logarithms, and exponents. It sold for $395.

The following year, Texas Instruments countered with the SR-10. TI’s calculator did not give values for trigonometric functions, but it cost only $150.

Semiconductor technology was enabling smaller and faster devices. Intel’s 8008 was the world’s first 8-bit programmable microprocessor and only the second microprocessor from the chipmaker. It featured 50% more transistors and 8× the clock speed of its predecessor, the 4004, and it was capable of data/character manipulation. Semiconductor historians credit the 8008 with cementing the future of microprocessor development and production as a business avenue, which paved the way for the modern computer age.

That release was illustrative of another technology battle that was in full swing: Japan’s rise as an electronics powerhouse. Handheld calculators were introduced to the U.S. by Japanese companies Busicom (Nippon Calculating Machine Corp.) and Sharp (Hayakawa Electric). Chips in early Busicom calculators were made in the U.S. by Mostek and Intel, while Texas Instruments supplied ICs to American calculator competitor Bowmar.

By 1985, the solar-powered Sharp EL-345 calculator sold for $5.95.

Japanese companies and Japan’s government, via its Ministry of International Trade and Industry (MITI), had been steadily investing in the nation’s technology and manufacturing capabilities since the 1960s. Japan was already the price leader in consumer electronics, but until the early 1970s, quality had

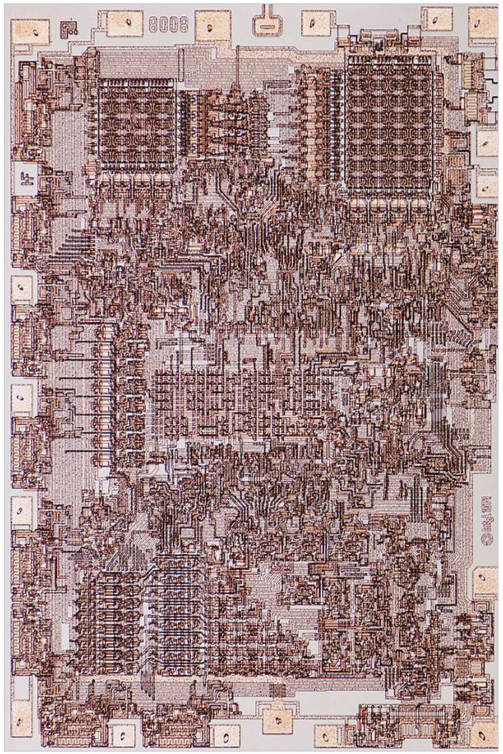
been pretty shabby. Japan's automotive industry — another focus of the nation — had adopted leading-edge practices such as lean manufacturing, which minimized inventory investment, and total quality management. Automakers had also begun to engage external suppliers, while the U.S. industry was still largely vertically integrated. These disciplines were easily transferrable to electronics.

Japan had built its edge in electronics by importing technology from other nations and then out-innovating the competition. The Sony Walkman, the VCR, and digital watches were all Japanese inventions. Two decades later, China would emerge as the low-cost option for electronics products and manufacturing services.

THE CHIPS

By 1972, Texas Instruments was accustomed to being top dog in the chip market. TI's Jack Kilby is credited with the invention of the integrated circuit — as are Jean Hoerni and Robert Noyce of Fairchild Semiconductor — in the late 1950s. Kilby is usually credited with having developed the concept of integrating device and circuit elements onto a single silicon chip, while Noyce is given credit for having conceived the method for integrating the separate elements.

Intel's 8008 had been developed on a separate track from the 4004, and the newer MPU's higher transistor count, higher performance, and data-/character-manipulation capabilities gave it broader market appeal. Developed for Busicom, the 4004 had 2,250 transistors and could



Intel's 8008 microprocessor, April 1972 (Source: Intel)

perform up to 90,000 operations per second in 4-bit chunks, but it could only handle arithmetic. The 8008 was a precursor to the x86 architecture, said TTI's Knight. "All these years later, the x86 is not a dead technology. Like MOSFETs, [the 8008 was] foundational technology."

Large-scale integration was also introduced in 1972. LSI enabled tremendous reductions in the cost, size, weight, and power consumption of components, along with increased speed and reliability. These features stemmed from the physical structure of the chip; more gates could be added without enlarging it — **the personalization of computing had begun.**

GAMING

The first PC was still years away, but gaming was driving electronics innovation. Magnavox introduced the first gaming console, the Odyssey, in May 1972. The Odyssey had a number of removable circuit cards that switched between the built-in games. A few years later, with the Odyssey 2, each game could be customized, with its own background and foreground graphics, gameplay, scoring, and music. Game players could purchase a library of video games tailored to their interests.

The Odyssey 2 included a full alphanumeric membrane keyboard, intended for educational games, selecting options, or programming.

Image by Evan Amos, own work, CC BY-SA 3.0

"Gaming drove a bunch of innovations," said Knight. "It got people focused on on-board memory — the 8008 was programmable and could address 16 KB of memory — which was important for gaming. Somewhere along the line, the gaming industry made a mistake: It became more focused on software. Then there was this rebirth, thanks to Nintendo, and gaming went back to being the driver of chip technology. It gave birth to Nvidia and the GPU, which is now critical to autonomous driving."

By the mid-1970s, the first ROM cartridge-based consoles arrived, including the Atari Video Computer System (VCS). Coupled with rapid growth in arcade video games, including "Space Invaders" and "Pac-Man," the home-console market flourished.

THE RISE OF CHINA

Nixon's 1972 arrival in Beijing ended 25 years of no communication or diplomatic ties between China and the U.S. and was the key step in normalizing relations between the U.S. and the PRC.

"There is a lot of correlation between what happened in Japan and what's happening in China," said Knight. "Japan decided to move upstream in electronics, and the government got behind it and was very deliberate about enabling technology. That mantle got passed to China, which is on the same journey. From a government standpoint, China moved from being a second source to truly innovative with its own brands.

"Nixon going to China, which laid the foundation for its entry into the World Trade Organization [WTO], followed by electronics outsourcing, allowed China to start leveraging what [U.S. tech companies] were teaching them," he added. "It was a brilliant move on Nixon's part; he saw bringing China into the first world was the best way to defeat communism. Where we are today, regarding competitiveness with China, began with that decision."

Barbara Jorgensen is managing editor and co-founder of supply chain publication EPSNews.

THE ROOTS OF SILICON VALLEY, PART I: **FOUNDERS, LEGEND, LEGACY**

BY MALCOLM PENN

As part of our EE Times 50th Anniversary Special, this three-part series looks at the 74-year history of Silicon Valley, including how it came to be; why the captains of the nascent industry set up shop in California; and the legacy of William Shockley Jr., Fairchild, and the "Fairchildren" who laid much of the foundation for the semiconductor industry we know today. Part 1 considers the birth of the transistor, how Shockley ended up in Silicon Valley, the origins of Fairchild Semiconductor, how the pioneering startup was funded, and what eventually happened to Shockley.

WILLIAM SHOCKLEY JR. AND THE BIRTH OF THE TRANSISTOR

The transistor was successfully demonstrated on Dec. 23, 1947, at Bell Laboratories (Murray Hill, New Jersey), the research arm of American Telephone and Telegraph Co. The three Bell researchers credited with its invention were William "Bill" Shockley Jr.; John Bardeen, the department head and group leader; and Walter Brattain. Shockley continued to work on development at Bell Labs until 1955 when, having foreseen the transistor's potential and looking to work for more than a salary, he quit to set up the world's first semiconductor company, becoming a de facto industry father.

Shockley was born in London on Feb. 13, 1910, the son of William Hillman Shockley, a mining engineer born in Massachusetts, and his wife, Mary (née Bradford), who had also been engaged in mining as a deputy mineral surveyor in Nevada.

The family returned to the United States in 1913, setting up home in Palo Alto, California, when Mary joined the Mining Engineering Department faculty at Stanford University. But for this twist of fate — given that both Shockley's parents were mining engineers — the family could have instead settled in Colorado, Nevada, or West Virginia.

William Jr. earned his B.S. degree at the California Institute of Technology (CalTech) in 1932 before moving to the East Coast to study at the Massachusetts Institute of Technology (MIT) under J.C. Slater. He obtained his Ph.D. there in 1936, submitting a thesis on the energy band structure of sodium chloride, and joined Bell Telephone Laboratories, where he remained until his resignation in 1955.

Shockley Semi-Conductor Laboratory (Source: Arnold and Mabel Beckman Foundation)

Upon leaving Bell Labs, Shockley moved back to Palo Alto (where his ailing mother still resided), initially as a visiting professor at Stanford but with the vision to establish his own semiconductor company making transistors and four-layer (Shockley) diodes. Had he decided instead to remain on the East Coast — close to Bell Labs, MIT, or IBM in Vermont — Silicon Valley might well have developed on the East Coast rather than the West Coast of the United States. The geographical difference almost certainly would have shaped an industry with a markedly different personality.

In Palo Alto, Shockley found a sponsor in Raytheon, a pioneer in what came to be known as electronic warfare. But Raytheon's support was short-lived. Undeterred, Shockley, who had been one of Arnold Beckman's students at CalTech, turned to him for advice on how to raise $1 million in seed money. Beckman was an American chemist, inventor, entrepreneur, founder, and CEO of the hugely successful Beckman Instruments — and now also a budding financier who believed that Shockley's new inventions would be beneficial to his own company. So rather than pass the opportunity to his competitors, he agreed to create and fund a laboratory on the condition that the lab would work to bring its discoveries to mass production within two years.

Beckman and Shockley signed a letter of intent to create the Shockley Semi-Conductor Laboratory (the hyphenation was then common practice) as a Beckman Instruments subsidiary under Shockley's direction. The new group would specialize in semiconductors, beginning with the automated production of diffused-base transistors. Shockley's original plan was to establish the laboratory in Palo Alto, close to his mother's home, but that changed when Fred Terman, provost at Stanford

University and central figure in the rise of Silicon Valley, offered him space in Stanford's new industrial park at 381 San Antonio Rd. in Mountain View. Beckman bought licenses on all necessary patents for $25,000, and the company was launched in February 1956.

STANFORD SOWS THE SEEDS

The seeds for Stanford's high-tech relationship with industry were sewn much earlier. In 1936, Sigurd and Russell Varian — together with William Hansen, Russell's ex-college roommate and by then a professor at Stanford — approached David Webster, head of Stanford's Physics Department, for help in developing the Varian brothers' idea of using radio-based microwaves for aircraft detection in poor weather conditions and at night. Webster agreed to hire them to work at the university in exchange for lab space, supplies, and half the royalties from any patents they obtained. The group's work eventually led to the August 1937 development of the klystron, subsequently adopted by Sperry, and the formation of Varian Associates in 1948.

In 1938, shortly after the klystron's development, Bill Hewlett and David Packard, who had graduated three years earlier with degrees in electrical engineering from Stanford University, formed Hewlett-Packard in a garage at 367 Addison Ave. in Palo Alto under the mentorship of Fred Terman. The garage is often referred to as the "Birthplace of Silicon Valley," understating the contributions of Terman and Stanford in creating the catalytic environment for Californian high-tech ventures, as well as the explosive role that Shockley Semiconductor would subsequently play. From a semiconductor perspective, 381 San Antonio Rd. in Mountain View, Shockley's address, is more appropriately the real birthplace of Silicon Valley, as recognized by IEEE.

SHOCKLEY SEMICONDUCTOR

Given his own prodigious IQ, Shockley embarked on an ambitious hiring campaign, seeking to employ the brightest scientists available — not just Ph.D.s, but Ph.D.s from the finest universities who were at the very top of their class — bringing together a veritable brain trust of brilliant engineers. The hiring process was not that straightforward, however; because most electronics-related companies and professionals at that time were based on the East Coast, Shockley's startup had to post ads in The New York Times and the New York Herald Tribune. Shockley initially tried to recruit from his Bell Lab peers, but knowing his reputation as a difficult manager, none would join him.

Early respondents included Sheldon Roberts of Dow Chemical, Robert Noyce of Philco, and Jay Last, a former intern of Beckman Instruments. Each was required to pass a psychological test, followed by an interview. Julius Blank, Gordon Moore, Last, Noyce, and

Roberts started working in the April–May timeframe, and Eugene Kleiner, Victor Grinich, and Jean Hoerni during the summer. By September 1956, the lab had 32 employees, including Shockley.

Though never medically diagnosed by psychiatrists, Shockley's state of mind has been characterized as paranoid or autistic. All phone calls were recorded, and staff were not allowed to share their results with each other — not exactly feasible, given that they all worked in a small building. At some point, Shockley sent the entire lab for a lie-detector test, although all refused. Shockley also lacked business experience and industrial management, unilaterally deciding that the lab would pursue an invention of his own — the four-layer diode — rather than develop the diffused silicon transistor that he and Beckman had agreed upon.

"...SHOCKLEY'S STATE OF MIND HAS BEEN CHARACTERIZED AS PARANOID OR AUTISTIC."

Barely six months passed when discontent boiled over, prompting seven employees to voice their concerns to Arnold Beckman — not to get rid of Shockley but to put a more rational boss between him and them. Their request might well have been granted had Shockley's Nobel Prize not been announced in November 1956, extending Shockley's fame and inflated ego. Rather than rock the boat, Beckman chose not to interfere, instead telling the seven to keep their heads down. Future Intel founders Noyce and Moore stood on different sides of the argument, with Moore leading the dissidents and Noyce standing behind Shockley, struggling to resolve conflicts. Shockley considered Noyce his sole supporter, but the team started to disintegrate, starting with Jones, a technologist, who left in January 1957 because of a conflict between Grinich and Hoerni.

ARTHUR ROCK INVESTS

In March 1957, Kleiner, who was also beyond Shockley's suspicions, asked permission to attend an exhibition in Los Angeles. Instead, he flew to New York to seek investors for a new company that he and the six others were by now contemplating. Kleiner's father, an investment banker, introduced Eugene to his broker, who in turn introduced Kleiner to Arthur Rock at Hayden Stone & Co. The team's original idea was to join an existing company. Rock, already investing in new companies — what today would be called startups — together with Alfred Coyle, backed Kleiner's proposition of a seven-strong, pre-packaged team, believing that trainees of a Nobel laureate were destined to succeed. Finding prospective investors, however, proved to be difficult, given the U.S. electronics industry was at that time concentrated on the East Coast. The California Group — as the seven became known — wanted to stay near Palo Alto. Rock presented the group to 35 prospective employers; all declined.

Dollar bills signed by each founding member were part of the contracts among the California Group, which formed the basis of Fairchild Semiconductor. (Source: Computer History Museum)

Lacking financial backing, the group led by Moore, as a last resort, presented Arnold Beckman with an ultimatum in May 1957: Solve the "Shockley problem" or they would leave. Moore proposed finding an academic position for Shockley, replacing him in the lab with a professional manager. Beckman again refused, believing that Shockley could still succeed — and shortly regretted that decision.

A month later, Beckman finally inserted a manager between Shockley and the team, but by then, it was too late, as the seven were now committed to leaving and embarking on Plan B — namely, creating their own startup. Recognizing they were followers, not leaders, the group persuaded the charismatic Noyce to join them. The now-expanded California Group met up with Rock and Coyle at the Hill Hotel in California. These 10 engineers became the core of a new company. Coyle, fond of ceremony, produced 10 $1 bills and laid them carefully on the table. "Each of us should sign every bill," he said. "These dollar bills covered with signatures would be our contracts with each other."

In August of that year, in a final throw of the funding dice, Rock and Coyle met with inventor and businessman Sherman Fairchild, founder of Fairchild Aircraft and Fairchild Camera and Instrument Co. The son of a wealthy entrepreneurial father who had made his fortune as an early investor in IBM, Fairchild was a bright and equally entrepreneurial engineer who had amassed a small fortune during World War II selling cameras for reconnaissance planes. Given that he had already developed an interest in semiconductors, Fairchild sent Rock to meet his deputy, Richard Hodgson. Risking his reputation, Hodgson accepted Rock's offer. Within weeks, paperwork and funding for the new company, Fairchild Semiconductor, had been completed.

The capital was divided into 1,325 shares, with each member of the California Group of eight receiving 100 shares, 225 shares going to Hayden Stone & Co., and

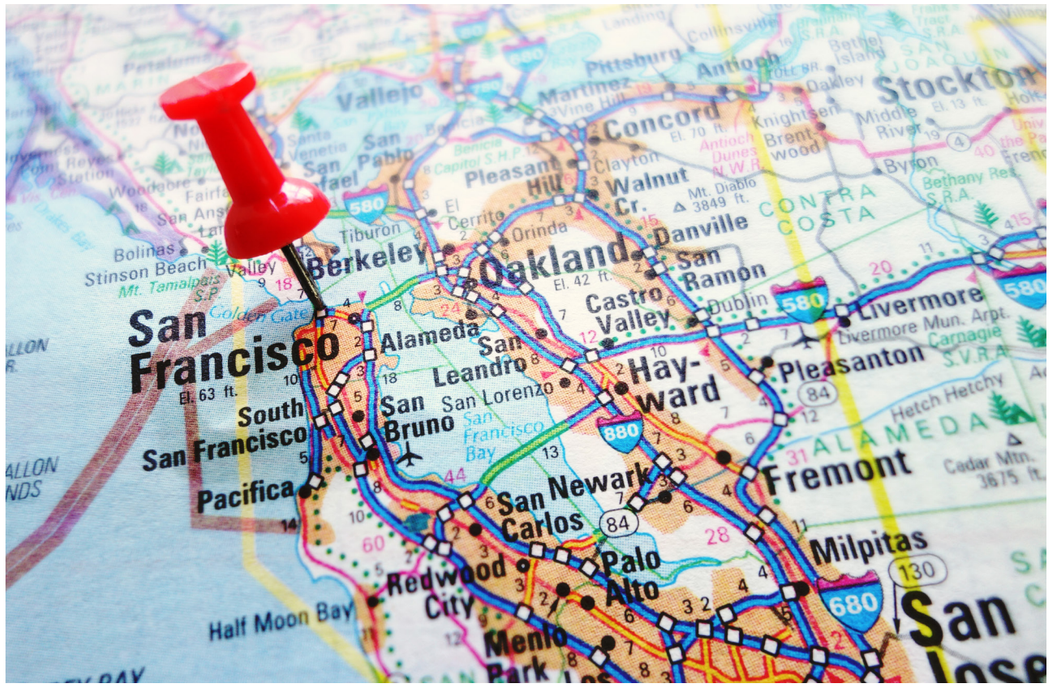
the remaining 300 shares held in reserve. Fairchild provided a loan of $1.38 million and, to secure the loan, the eight gave Fairchild the voting rights on their shares with the option of buying them back at a fixed total price of $3 million.

'THE TRAITOROUS EIGHT'

The eight left Shockley on Sept. 18, 1957, and Fairchild Semiconductor was born. While there is no documentary evidence, the group quickly became known as "The Traitorous Eight." Shockley never understood the reasons for their defection, considering it a betrayal, and allegedly never again spoke to Noyce or the others. With the help of a new team, Shockley brought his own diode to mass production the following year, but by then, time had been lost and competitors were already close to developing integrated circuits. In April 1960, Beckman sold the unprofitable Shockley Labs to the Clevite Company, based in Waltham, Massachusetts, bringing his association with the semiconductor industry to an end.

On July 23, 1961, Shockley was seriously injured in a car crash and, after recovery, left the company and returned to teaching at Stanford. Four years later, Clevite was acquired by ITT. In 1969, ITT decided to move the lab to West Palm Beach, Florida, where it had an already-established semiconductor plant. When the staff refused to move, the lab ceased to exist.

Malcolm Penn is chairman, CEO, and founder of semiconductor industry analysis firm Future Horizons.



THE ROOTS OF SILICON VALLEY, PART 2: PLANAR TECHNOLOGY, THE FAIRCHILDREN

BY MALCOLM PENN

Part 2 of this three-part series looks at the evolution of planar technology; the "family tree" of semiconductor startups that evolved from Fairchild (the "Fairchildren"), including Intel; and the competition with Texas Instruments.

FAIRCHILD SEMICONDUCTOR

Founded in intrigue, Fairchild set up shop on 844 E. Charleston Rd., on the border between Mountain View and Palo Alto, and went on to record a long history of innovation, producing some of the most significant technologies of the second half of the 20th century. It quickly grew to be among the top semiconductor industry leaders, spurred on by the successful development of the silicon planar transistor.

Transistors, however, were already presenting a new challenge, dubbed the "tyranny of numbers." If you wanted to make a simple flip-flop, it needed four transistors. About 10 wires were needed to connect them. Interconnecting two flip-flops required not only twice the number of transistors and wires but also four or five additional wires to connect the two devices. So four transistors needed 10 wires, eight needed 25, and 16 needed 60 to 70 wires. In other words, as the transistor count increased linearly, the number of connections grew exponentially, whereby the exponential was greater than one but less than two.

While transistors were relatively easy to mass-produce, connections were much more difficult, as wires had to be soldered by hand and took up a lot of space. The industry's desire to build bigger and more complex systems was stymied by the difficulty in wiring everything together. To this point, few had paid much attention to wiring, but connections would soon become a potential showstopper, driving the need for the integrated circuit.

PLANAR TECHNOLOGY

In 1958, Jack Kilby of rival semiconductor company Texas Instruments demonstrated the ability to integrate a pair of transistors on a semiconductor substrate. Kilby's transistors were wire-bonded, however, leaving the connection problem

unresolved. That problem was solved by Bob Noyce, with the help of Jean Hoerni (who provided the technique) and Jay Last (who eventually made it work).

Hoerni had been working on a fix for reducing transistor defects. Defects were traced to unprotected transistor surfaces inside a package, allowing particles to contaminate and degrade the device over time. Hoerni's solution was to protect the transistor surface with a passivation, or protection, layer of silicon dioxide (SiO_2), grown or deposited on top of the structure. Rather than deposit the emitter and base regions on top of the substrate, as with the current mesa process, Hoerni saw another way: If the surface was completely covered with SiO_2, the emitter and base areas could then be selectively diffused. The net result was a much flatter surface, allowing for greater automation during production.

Planar technology, announced in January 1959, would become the second-most–important invention in the history of microelectronics — after the invention of the transistor — laying the foundation for future integrated circuits. At the time, the advance went virtually unnoticed, with the key exception of Noyce, who recognized that a glass layer was an insulator, providing a means for connecting wires laid on top and patterned like a printed-circuit board.

Noyce filed his patent in April 1959, triggering a legal battle between Texas Instruments and Fairchild (Kilby and Noyce remained friends, with high regard and respect for each other). Texas Instruments claimed that Kilby's patent claim — "electrically conducting material such as gold laid down on the insulating material to make the necessary connections" — was a preexisting description of Noyce's patent claims and that Kilby had only used wire bonds as the quickest way to a prototype. Had this assertion been upheld, Noyce's later-dated patent would have been declared invalid.

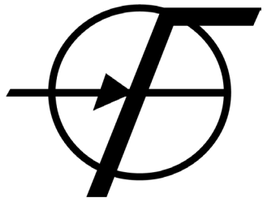
Texas Instruments lost the argument, both patents were declared valid, and the two companies reached a cross-licensing agreement.

Kilby was humble by nature, and even though his patent pre-dated Noyce's, he generously announced that both he and Noyce had invented the integrated circuit, contrary to the position of Texas Instruments' management.

In 1959, Sherman Fairchild exercised his right to purchase the founding members' shares, an event that turned former entrepreneurs and partners into ordinary employees, thereby undermining the company's team spirit and sowing the seeds of future friction.

Isolation was another big problem yet to be solved before integrated circuits could become a commercial reality. The problem was how to stop adjacent transistor interference. Noyce delegated this thorny problem to Last, who was running the R&D group. It was no easy task, taking some 18 months before the first working device was produced on Sept. 27, 1960.

TROUBLE AT FAIRCHILD



Development also met with strong internal resistance. Tom Bay, Fairchild's vice president of marketing, accused Last of squandering resources. In November 1960, Bay demanded termination of the project, with the resultant savings allocated to transistor development. Moore refused to help, and Noyce declined to discuss the matter, leaving Last to fight the battle on his own. The conflict flared up barely a month after Fairchild announced the transition of its transistor production from mesa to planar technologies. Moore refused to credit this achievement to Hoerni, fanning the flames of the already-developing tensions among the eight founding partners.

Last continued to develop six more parts, but ongoing conflicts were the last straw. Flush with their planar and isolation process success, Last and Hoerni left Fairchild on Jan. 31, 1961, to launch Amelco in Mountain View, with financing from Teledyne Corp. arranged by Arthur Rock. Their plan was to develop ICs to support Teledyne's military business. Eugene Kleiner and Sheldon Roberts joined the pair a few weeks later. With this high-level defection, the eight founding members had been split into two groups.

Fairchild announced the world's first standard logic family of ICs, direct-coupled transistor logic, in March 1961. The device was based on Hoerni and Last's resistor-transistor logic (RTL) planar process under the µLogic trademark. The µL903 three-input NOR gate became the basic building block of the Project Apollo guidance computer. Designed by MIT and built by Raytheon, the lunar navigation computer required 5,000 devices and was the first major IC application. Miniaturization for space applications was driving early scaling.

Fairchild's lead, however, was short-lived. David Allison, Lionel Kattner, and others also left at around the same time as Hoerni and Last to launch Signetics (Signal Network Electronics). One year later, in 1962, the startup announced a much-improved, second-generation logic family, the SE100 Series diode-transistor logic (DTL). Fairchild quickly responded with its own DTL family, the 930 series, undercutting Signetics and rendering that company unable to compete against Fairchild's marketing juggernaut.

NE555 TIMER: MOST POPULAR IC EVER?

Signetics' most famous legacy part was the NE555 timer. Designed in 1971, the 555, along with the ubiquitous TTL 7400 quad two-input NAND gate, was probably the most popular IC ever sold. Signetics was acquired by Philips in 1975.

Early ICs were housed mainly in either TO-5 or TO-18 adapted metal can transistor packages. These worked fine for three-lead devices, but scaling them to provide more connections proved to be limiting, given that they could be made only so large and the radial leads packed only so tightly. Ten leads were about the practical limit and

Fairchild Semiconductor founders, 1957 (Source: Computer History Museum)

would not support the more complicated ICs in the pipeline. It fell to Fairchild's Don Forbes, Rex Rice, and Bryant "Buck" Rogers to provide a fix in 1964, via the invention of the now-familiar dual in-line package, the tiny oblong "millipedes" that would crawl across circuit boards for the next 40 years.

The packaging innovation stemmed from a ceramic flatpack design devised in 1962 by Yung Tao, a Texas Instruments engineer, as an industry standard for surface-mount ICs for the U.S. military. The concept was adapted for through-hole rather than surface mounting, with an eye toward ease of handling for electronics manufacturers and easier PCB layout design for delivering power to the ever-increasing number of ICs, routing their signals around the board. Another consideration was cost, given the growing consumer IC market. The 0.1-inch (2.54-mm) package pin spacing left plenty of room for PCB tracks to be routed between pins, and the 0.3-inch (7.62-mm) spacing between rows of pins left room for other tracks.

Fairchild launched its dual in-line package in 1965, originally in ceramic, but the package took off with a vengeance when Texas Instruments introduced a plastic resin version, driving the unit cost down dramatically. As a result of great design, low cost, and support for increasingly complex ICs, the plastic dual in-line package became the industry standard, with its basic 14-pin design extended to support more leads, up to 64 pins in a 0.6-inch–wide form factor, and more complex ICs. It was eventually surpassed by second-generation surface-mount devices in the late 2000s as chip complexity and pin count requirements surpassed the capability of dual-in-line packages.

With as many as 15,000 dies on a wafer, assembly and test now outweighed wafer fab costs — hence, the need to reduce labor costs as a matter of survival. After some early failed ventures — for example, in Shiprock, New Mexico, at a Navajo reservation — along with early attempts at automation, offshoring test and assembly to Asia proved successful, at least in the short term. Noyce, an investor in a small radio company in Hong Kong, suggested to Charlie Sporck that he and Jerry Levine scout the region.

They were attracted by the low labor cost, non-unionized facilities, Western-educated technicians, good engineering schools, and tax incentives and other government subsidies. In 1963, Fairchild set up the industry's first East Asian assembly and test operation in a former shoe factory on the Kowloon side of Hong Kong. Other semiconductor manufacturers subsequently followed Fairchild to the region, primarily Malaysia.

FROM HOTEL CHAIN TO INTEL

Julius Blank, Victor Grinich, Moore, and Noyce stayed with Fairchild until 1968. In March of that year, Moore and Noyce decided to leave, turning to Rock for funding and launching NM Electronics in the summer of 1968. One year later, NM Electronics bought the naming rights from the hotel chain Intelco, thereby launching Intel Corp.

Grinich also left in 1968, first to teach at UC-Berkeley and Stanford, where he published the first comprehensive textbook on integrated circuits. But Grinich never lost the startup itch and quit academia in 1985 to co-found and run several startups, including Escort Memory Systems, developer of industrial RFID tags.

Blank, the last of The Eight, left Fairchild in 1969 to become a consultant to tech startups. Seeking a more hands-on role, he co-founded Xicor in 1978 to make E^2PROMs.

As for the original four defectors, Hoerni headed Amelco until the summer of 1963, when, after a conflict with the Teledyne owners, he left for Union Carbide Electronics. In July 1967, supported by watch company Société Suisse pour l'Industrie Horlogère (SSIH), the predecessor of Swatch Group, Hoerni founded Intersil. The startup pioneered the market for low-power custom CMOS circuits, some of which were developed for Seiko. The combination kickstarted the Japanese digital watch industry.

Hoerni next launched the European version of Intersil, called Eurosil, financed in part by SSIH's desire to build a fab in Munich, not far from Swiss watch manufacturing.

Eurosil was eventually sold to Diehl in late 1975. Hoerni left in 1980, returning to the West Coast to form Telmos. The startup produced semicustom products covering the linear interface between sensors to microprocessors and digital logic cores along with high-voltage, high-current drivers.

Last continued at Amelco, completing a 12-year tenure as VP of technology at Teledyne, Amelco's parent. In 1982, he founded Hillcrest Press, specializing in art books. Roberts also left to set up his own business and later served as a trustee at Rensselaer Polytechnic Institute.

That left just Kleiner, who departed to pursue a career financing the many early-stage companies springing up on the West Coast, teaming with Thomas Perkins, head of R&D at Hewlett-Packard, to form Kleiner Perkins. They opened an office on Sand Hill Road in Palo Alto, the locale that would become the home of U.S. venture capitalists. While Rock and Hayden Stone could arguably be credited with establishing the first venture capitalist firm, Kleiner Perkins was the first investor with a physical office in Silicon Valley. Kleiner Perkins would go on to fund Amazon, Compaq, Genentech, Intuit, Lotus, Macromedia, Netscape, Sun Microsystems, Symantec, and dozens of other companies.

As for today, Amelco — after numerous mergers, acquisitions, and rebrandings — no longer exists, but its IP portfolio survives, now owned by Microchip.

Malcolm Penn is chairman, CEO, and founder of semiconductor industry analysis firm Future Horizons.

THE ROOTS OF SILICON VALLEY, PART 3: STARTUP FEVER AND VENTURE CAPITAL

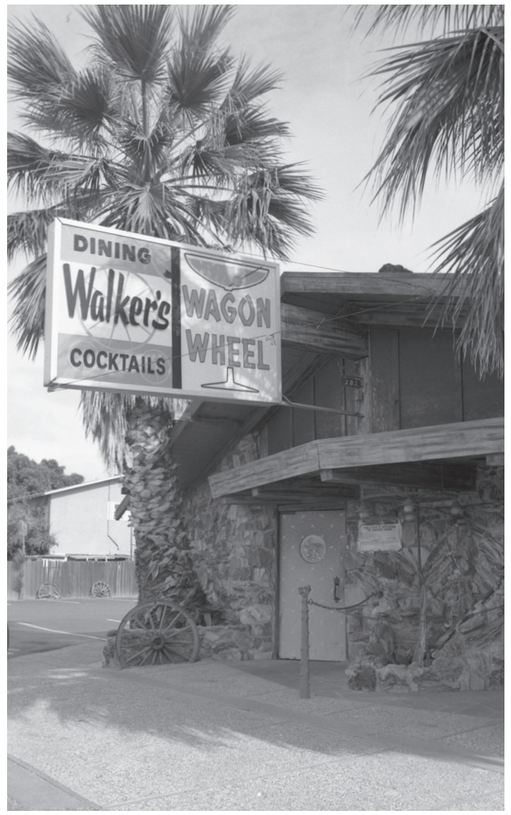
BY MALCOLM PENN

The third and final part of our series explores the legacy of Silicon Valley as a result of the three key inventions that changed the world in the 1960s: the integrated circuit, startup fever, and venture capital.

MELTING POT FOR THE FAIRCHILDREN

Sheldon Roberts, Eugene Kleiner, and Jean Hoerni's collective decision to leave and compete against Fairchild, just over three years after the company was founded, was the first of what would be many subsequent defections and spinouts. The "Fairchildren" would directly or indirectly create dozens of corporations, including Intel and AMD. Fairchild thus sowed the seeds of innovation across multiple companies in the region that would eventually become known as Silicon Valley.

Local watering holes, restaurants, and other hot spots provided venues for Silicon Valley's "work hard, play hard" ethos, where industry folk gathered after work to drink, gossip, brag, trade war stories, talk shop, exchange ideas, change jobs, and develop new contacts. Key venues included the Wagon Wheel, Lion & Compass, and Ricky's, along with the Peppermill and the Sunnyvale Hilton.



Industry folk would meet, gossip, trade ideas, and change jobs at watering holes like the Wagon Wheel (pictured), Lion & Compass, and Ricky's. (Source: Computer History Museum)

Stanford University, and particularly Fred Terman, also played a catalytic role, propelled by the engineering department chair's vision for academia to develop a new relationship with the science- and technology-based industries dependent on brainpower as their greatest asset. Terman further recognized the need to develop local industry, not just by building a community of interest between faculty and industry but also by encouraging new enterprises — what we would call startups today — to cluster around the university. To that end, Stanford provided intellectual property and office space, often rent-free other than the local property taxes.

While it is unclear who came up with the moniker "Silicon Valley," Don Hoefler, a technology reporter for the industry publication Electronic News, is often credited with popularizing the name in a 1971 column about the region's chip industry. Hoefler also promoted the area's innovative qualities and was one of the first writers to chronicle the Northern Californian technology industry as a community.

THE FAIRCHILD LEGACY

Throughout the first half of the 1960s, Fairchild was the undisputed semiconductor leader, forging ahead across all industry segments, be it design, technology, production, or sales. Early sales and marketing efforts were modest and military-oriented; that changed in 1961 when Robert Noyce and Tom Bay recruited a group of aggressive salesmen and marketing specialists, including W.J. "Jerry" Sanders III and Floyd Kvamme.

The newcomers transformed Fairchild's sales and marketing departments into one of the industry's legends.

Among the pivotal moments was Fairchild's entry into the consumer TV market. Attracted by potential high volumes, Sanders wanted to replace the tube (valve) CRT driver with a transistor, but the target price was $1.50. Transistors at that time were selling to the military for $150. In what can only be regarded as a massive leap of faith, Noyce's instructions to Sanders were, "Go take the order, Jerry. We'll figure out how to do it later. Maybe we'll have to build it in Hong Kong and put it in plastic, but right now, let's just do it."

In 1963, Fairchild hired Robert Widlar to design analog operational amplifiers using Fairchild's digital IC process. Despite its unsuitability, Widlar, in partnership with process engineer Dave Talbert, succeeded in adapting the process to produce two revolutionary parts: the world's first operational amplifiers, the μA702 in 1964 and μA709 in 1965. With these two parts, Fairchild now dominated both the analog and digital IC markets, first with its μLogic RTL family and then with its 930 series DTL. In April 1965, Gordon Moore famously published his article "Cramming More Components onto Integrated Circuits" in Electronics. Later to be known as Moore's Law, it was basically an extrapolation of four plots on a graph showing IC transistor density over time.

Fairchild's digital technology lead was, however, being overtaken by Texas

Instruments. Having fallen behind in RTL and DTL, Fairchild's chief rival decided to copy Sylvania's ultra-high–performance (SHUL) transistor–transistor logic (TTL) circuit design, adapting it to its own process to counter the announcement of Fairchild's third-generation 9000 series TTL logic.

"GO TAKE THE ORDER, JERRY. WE'LL FIGURE OUT HOW TO DO IT LATER."

Headed up by Stewart Carroll, Texas Instruments set up a "design factory" that could churn out several new designs a week, mostly by guessing the W/L ratios, laying out the circuits, correcting them if the prototypes did not work, and zeroing in on a specification that manufacturing could support. The design factory was supported by an optical photomask generator, as opposed to a manual rubylith layout, that could quickly create a photographic chip layout, as well as a "quick turn" fab line to churn out parts.

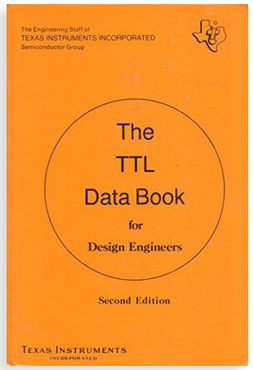
TTL DATA BOOK

To strengthen its attack, Texas Instruments masterminded a marketing coup by persuading other semiconductor companies to second-source its TTL rather than Fairchild's competing product. In this single masterly move, Texas Instruments established its 74 Series version of TTL as the de facto third-generation industry standard, leaving Sylvania's SHUL, Fairchild's 9000 Series, and other proprietary alternatives behind. It then proceeded to neutralize the entire second-source movement by providing every engineer with a copy of its ubiquitous orange book ("The TTL Data Book for Design Engineers"). Its twice-yearly "must attend" TTL seminars, not just in the U.S. but globally, were supported by an aggressive new product introduction program.

By always ensuring any bill of materials included at least one TTL part that was only available from Texas Instruments, the company was able to stay one step ahead of the competition and own the TTL market for the best part of 30 years, until standard logic eventually fell victim to the 1980s application-specific IC revolution.

In the meantime, starved of capex, Noyce's position on Fairchild's executive staff was consistently being undermined by Sherman Fairchild's corporate interference and his lack of support. The Fairchild management team was increasingly upset by Sherman's corporate focus on unprofitable ventures at the expense of the semiconductor division. The company suffered its ultimate humiliation in July 1967, when the semiconductor industry fell victim to the first of its cyclical recessions, during which the company lost money and was forced to concede its technology leadership to Texas Instruments.

Charles Sporck, Noyce's operations manager — often credited with running the industry's tightest ship — left in early 1968 along with Pierre Lamond to join Widlar and Talbert at National Semiconductor. That triggered Noyce's and Moore's departures that same year — a pivotal moment in the eventual demise of the firm. The collective exodus of Sporck, Noyce, and Moore, along with so many other executives, signaled the end of an era, prompting Sherman Fairchild to bring in a new management team led by C. Lester Hogan, then vice president of Motorola Semiconductor.



Of the eight original founders, only Julius Blank remained, although he, too, would be gone within a year.

HOGAN'S HEROES

Hogan's arrival, and the subsequent displacement of Fairchild managers, demoralized the company even further, prompting a further exodus of employees who would launch a host of new companies. Leading a group dubbed "Hogan's Heroes," the ultra-conservative Motorola executives immediately clashed with Sanders, Fairchild's flamboyant sales chief.

While initially slow to respond to the changing market under Sanders's direction, Fairchild had embarked on a strategy of leapfrogging Texas Instruments by focusing on more complex, large-scale parts with 30 gates or more, instead of simpler, small- and medium-scale devices under 30 gates — a strategy that was proving popular and successful with engineers. The move forced Texas Instruments to recognize the threat and copy all of Fairchild's 9300 Series parts under 74 series numbers (for example, the 9300 became the 74195 and the 9341 the 74181.)

Sanders's entire strategy collapsed, however, when Hogan capitulated to Ken Olsen, founder and CEO of Digital Equipment Corp. and a key Fairchild customer. Olsen wanted Fairchild to give up on its proprietary TTL technology and instead second-source Texas Instruments' 74 Series TTL. Against Sanders's wishes, Hogan agreed, signing the death warrant for Fairchild's TTL strategy. Sanders was, understandably, livid. "You've just killed the company, Ken," Sanders fumed.

Hogan's betrayal was the last straw for Sanders. He, together with a group of Fairchild engineers, quit to start Advanced Micro Devices. When Sanders was installed as president, one of his first moves was to establish the mantra: "People first, revenues and profits will follow." Sanders also gave every employee stock options in the new company, an innovation at the time.

Wilf Corrigan, who had moved with Hogan to Fairchild as director of Discrete Product Groups, succeeded Hogan as president and CEO in 1974. Fairchild continued to decline, however, dropping to sixth place in the semiconductor industry by the end of the decade.

In the summer of 1979, with the semiconductor market again riding high on its fourth year of successive double-digit growth, Fairchild fell victim to a hostile takeover bid from Gould, a major U.S. producer of electrical and electronic equipment, hell-bent on a diversification strategy.

Unable to fend off the buyout, Corrigan sought the best price for shareholders. Fairchild was eventually sold to Schlumberger, a French oil services industry company, for $350 million, or $66 per share (Gould went as high as $57 per share).

Schlumberger was unable to revive the deteriorating company, and it continued to lose money. Corrigan departed in February 1980, and once his non-compete clause expired, he and Rob Walker co-founded ASIC pioneer LSI Logic Corp. in 1981.

Schlumberger initially replaced Corrigan at Fairchild with one of its own managers, Tom Roberts, who unsuccessfully ran the firm like a heavy-equipment company. Two years later, in 1983, it recruited Donald W. Brooks, a Texas Instruments veteran, to reverse its decline. By then, Fairchild was a legend in trouble, lagging in leading-edge technologies and losing money, even as the rest of the semiconductor industry was booming.

The company was eventually sold to National Semiconductor in 1987 for one-third of the price paid by Schlumberger eight years earlier. With the Fairchild brand now dead, Brooks left, and the company was back in the hands of former Fairchild general manager Charlie Sporck. Kirk Pond became COO at National Semiconductor in 1994, where he led the successful management buyout in 1997. With the Fairchild name revived, Pond continued as president and CEO until 2005, when he became chairman, before retiring a year later in 2006.

Pond was succeeded by Mark Thompson until the company was acquired by ON Semiconductor (now onsemi) in September 2016. ON Semiconductor was the discrete, standard analog and logic device division spun out from Motorola's Semiconductor Components Group in 1999.

SILICON VALLEY LEGACY

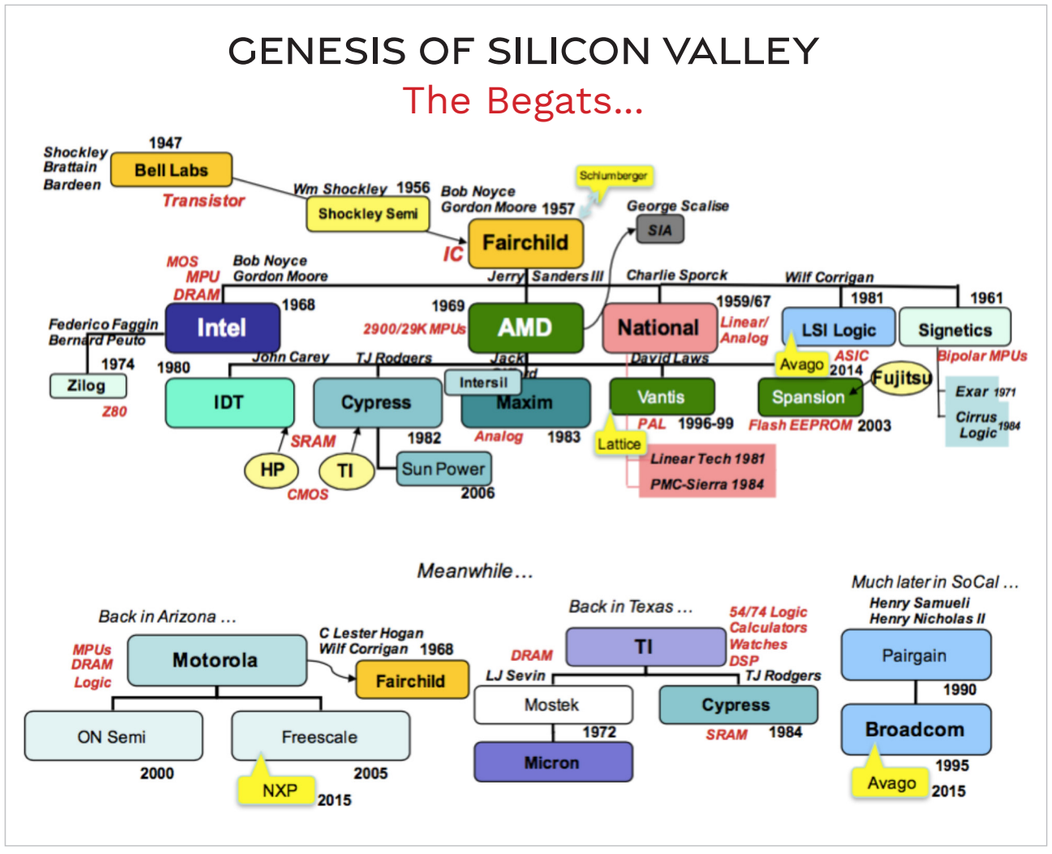
On Feb. 14, 1956, Arnold O. Beckman and William B. Shockley announced to a luncheon audience of scientists, educators, and the press at San Francisco's St. Francis Hotel that they were founding Shockley Semiconductor Laboratory in Palo Alto. Not long thereafter, three critical advances in the 1960s — integrated circuits, startup fever, and venture capital — changed the world.

Perhaps these inventions would have happened somewhere else, at some other time, by somebody else. The fact that they all occurred within a short space of time — in and around Palo Alto — driven by the entrepreneurial spirit of the "Traitorous Eight" and many other key players, along with the startup ethos of Stanford University, made Silicon Valley unique in the history of technology.

But what if Shockley's parents had moved to Colorado, Nevada, or West Virginia to pursue their mining careers on their return to the U.S. rather than Palo Alto? Would Silicon Valley have developed instead in those places?

What if Shockley had chosen to set up Shockley Semiconductor on the East Coast, where there was an already-well-developed infrastructure, rather than Palo Alto, which had none? From the perspective of infrastructure, the East Coast was far better positioned to forge a Silicon Valley.

What if Terman hadn't had the foresight to develop a community of interest between Stanford's faculty and a fledgling industry, encouraging new enterprises to cluster around the university?



Key players in the creation of Silicon Valley's semiconductor ecosystem (Source: Jeff Software)

What if the Russians, Europeans, or Japanese had invented the integrated circuit first? Could Silicon Valley have sprung up in the USSR, Europe, or Tokyo instead?

What would the world look like today had any of these scenarios not unfolded?

Clearly, fate played a role in bringing Shockley and semiconductors to Palo Alto, but the West Coast proved a far more fertile environment for the risk-taking spirit of the "Traitorous Eight" and their peers than the more risk-averse East Coast business and financial communities.

All eight of the original founders eventually left Fairchild and went on to become serial entrepreneurs, co-founding a wide variety of new startups, both in semiconductors and venture capital. They surrounded themselves with brilliant engineers who also wanted to start new companies, prove themselves, and change the world — stoking the startup boom driven by Shockley Semiconductor as the embryo, Fairchild Semiconductor as the incubator, and the Palo Alto ecosystem as the catalyst.

The rest, as they say, is history.

Perhaps Walker, the co-founder of LSI Logic, best summed up the entrepreneurial spirit of the Valley, and the rise and fall of Fairchild, with this comment: “It’s amazing what a few dedicated people can accomplish when they have clear goals and a minimum of corporate bullshit.”

Malcolm Penn is chairman, CEO, and founder of semiconductor industry analysis firm Future Horizons.

EE Times 50

PARTNER CONTENT

THREE LESSONS FROM VALLEY’S ANALOG HISTORY, AND A LOOK FORWARD

BY DAVID ANDEEN, *SENIOR DIRECTOR, BUSINESS DEVELOPMENT AND MARKETING, ANALOG DEVICES*

LESSON I: STICK WITH YOUR PASSIONS

Energetic and freshly graduated from Stanford with an electrical engineering degree, Bob Reay had all the opportunities in the world. The burgeoning economic engine of Silicon Valley meant jobs. “When I graduated, I had a job offer from Intel to work on the 286 and a job offer from Intersil to work on A/D converters,” recalled Reay, who today is vice president and Fellow, Power Products, at Analog Devices. “I decided to go to Intel because they were the big thing. Then the guys at Intersil called and said, ‘Bob, do you like to windsurf?’ I said yes. They said, ‘Come and work for us; we’ll buy you a windsurfer.’ I said, ‘Oh, really? Well, how about you throw in a roof rack and a wetsuit?’ The next thing you know, I had been bribed by windsurf equipment and became an analog designer.”

Thus launched Reay’s amazing analog design career, one of many that connect the rich history of analog semiconductor innovation. At Intersil, Reay joined future Maxim leaders Tunç Doluca (CEO) and Pirooz Parvarandeh (CTO), both early in their careers, developing A/D converter products and other core analog products that continue to provide revenue and innovation streams to this day.

Maxim’s founding team (from left): Dave Fullagar, Jack Gifford, Fred Beck, Sam Ochi, Dave Bingham, Beverly Fuller, Roger Fuller, Rich Hood, Lee Evans, Steve Combs

Intersil fits into Silicon Valley history in a unique way. Founded in 1967 by Jean Hoerni, Intersil’s claim to fame is Hoerni’s planar process, the exact foundation upon which semiconductor fabrication takes place today. “Back then, there was a tight connection between the process and the designer — the process was like another design tool,” recalled Reay. The

nature of planar devices, such as the values of passive components, meant designers needed to redevelop classic analog circuits, which originally consisted of large discrete resistors and capacitors. At the time, and continuing to this day, demand for analog designers fueled the fate of companies across the industry.

The seeds of today's Analog Devices, the global leader in analog semiconductors, grew in many ways from Intersil. General Electric owned Intersil, which organizationally resided in the lighting division. The pairing seemed odd, especially to casual West Coast engineers, who periodically received visits from white-shirted East Coast executives. "At one point, Jack Welch came out to visit," said Reay. "They actually rolled out a red carpet for him!"

The boiling point developed with one of the analog industry's central figures: Jack Gifford, founder and first (and longtime) CEO of Maxim. At the time, the topic of stock options was a consistent issue with GE management, who did not grant them to other divisions. Employees at Intersil, conversely, viewed stock options as a fundamental component of compensation, the fuel and excitement of a startup, and the potential for riches. Gifford agreed to join Intersil only if he received stock options. The issue exploded when the two Jacks, Gifford and Welch, argued at a party and Gifford told Welch to "f--- off." Those who knew Gifford were not surprised. Welch didn't wait long to fire Gifford, who promptly responded by launching Maxim in 1983.

In founding Maxim, Gifford brought together some of the finest minds in analog, with the goal of creating an employee-owned company that put analog first.

Two years earlier, the genesis of Linear Technology, another legendary analog company, formed from National Semiconductor. National had long been the standard for an aggressive, innovative Silicon Valley company, with characters like Charlie Sporck, the hard-charging, cigar-toting CEO with the wry smile. Meetings regularly were punctuated by shouting and the pounding of fists on tables in what now seems like an odd way to keep the lines of communication open. But the culture worked for Silicon Valley. Bob Swanson, the founder and primary CEO of Linear Technology, left National with four brilliant colleagues: Bob Widlar, Bob Dobkin, Brian Hollins, and Brent Welling. The story goes that National became increasingly political, requiring Swanson, a VP at the time, to spend half his days justifying analog to the centralized teams, who believed that if a process worked for digital, it should work for analog, too. At one point, the analog product line yields dropped to zero because the production team had removed a critical die coat step.

After gathering the team to start Linear Technology, a few months went by before the company got moving. Swanson recalls the five spending their evenings shooting pool and drinking beer in his garage. Finally, one day, Welling asked, "Are we going to start a company or just drink beer and play pool every night?" With that, Swanson and the team took action. Linear sometimes gets referred to as "The House of Bob" — but not because of all the Bobs. Although Widlar and Dobkin are legends, Linear was the House of Bob Swanson.

On the other side of the country, Analog Devices was steadily building its own analog franchise. Reay remarked, "ADI always seemed more academic, and also more open" with technology and information. Around the time that Maxim and Linear Technology were just starting up, ADI was celebrating more than 15 years of existence and over a decade of operation as a public company.

Analog Devices founders Ray Strata (left) and Matthew Lorber

LESSON 2: DON'T BE AFRAID OF FAILURE

Reay's technical education began early. "My father was a math professor, so I grew up doing math problems at the dinner table," he recalled. "I had no idea that this wasn't normal until I actually got out of high school and met real people! I got a Radio Shack electronics kit while I was in high school. I started playing with that and building all sorts of circuits. It was fun, but [the kit] didn't really explain how [the circuits] worked.

"Dad bought us an Apple 2 computer, which came out after my freshman year in college," he continued. "That summer, my brother and I took apart the Apple 2 and decided to build a speech synthesizer for it. We bought this 40-pin DIP chip and built the card. We wrote all of this code, and we're finally getting it ready. As we plug in the card, smoke comes pouring out! So Dad walks by and says, 'OK, boys, what did you do?'

"We didn't have a lot of money at the time, so my dad made us do a failure analysis to find out what had happened. We finally figured out that we plugged the card in without turning off the power. The board we plugged in had these big bypass capacitors, so when you plugged it into the backplane, you had this big inrush of current. I'm telling you this because you can't let the fear of blowing stuff up stop you.

"You cannot be afraid of failure," Reay asserted. "Failure is part of the process."

Indeed, Silicon Valley and failure intertwine and connect uniquely and consistently. While the world celebrates technology breakthroughs, successful IPOs, and the winners of history, progress completely stalls without the many, many failures that precede any level of success. This truth holds even more for the analog industry. Dobkin once remarked that "an analog design engineer with 10 years of experience was more valuable than an engineer with five years

of experience, because he had made more mistakes."

The process of failure, and amazing success, for Silicon Valley begins with William Shockley, the inventor of the transistor, Nobel laureate, and founder of Shockley Semiconductor.

Reay recalled, "At Stanford, we had these dorm seminars where we'd invite Nobel laureates to come and talk to us. When I was an early undergrad, we had Bill Shockley come in and talk to us. Then we took him to lunch at the food service. It was really interesting to hear him talk about the early years of Silicon Valley, and even before, with the early work in semiconductor physics.

"The thing I can say is, it seemed to be this really exciting time when stuff was developed. Stuff that today you take for granted: simple things, like a planarized process or just integrated circuits or low-dropout [LDO] regulators. There are all these things that you think have been around forever, but they weren't around forever. There were a lot of fundamental breakthroughs going on."

While an amazing technologist who could spot talent and reportedly "see electrons," Shockley managed poorly and exhibited extreme paranoia. In 1957, frustrated with Shockley, the legendary "Traitorous Eight" — Gordon Moore, C. Sheldon Roberts, Eugene Kleiner, Robert Noyce, Victor Grinich, Julius Blank, Jean Hoerni, and Jay Last — all left Shockley. The group found funding and started Fairchild Semiconductor in Palo Alto, California. That seminal moment spawned the era of semiconductors, arguably the birth of Silicon Valley. A multitude of innovation, processes, products, and companies were formed from the seed of Fairchild, from the household-name semiconductor companies like Intel and AMD to companies pursuing other products, such as color TVs. A familiar Silicon Valley pattern emerged: Innovate, fail, succeed, and try again.

The characters in this story descend from Fairchild. Gifford worked at Fairchild, having been handpicked by Widlar to run the nascent linear circuit business. Gifford then co-founded AMD with Jerry Sanders in 1968. After conflict with Sanders, Gifford briefly went into farming before returning to semiconductors at Intersil. Swanson worked at Fairchild from 1963 to 1967, primarily in manufacturing and operations, before ultimately joining National.

The paths of Gifford and Swanson differ and coincide in many ways. Gifford grew up in Torrance, California, a baseball-playing West Coast native who attended UCLA and spent his life in California. Swanson, by contrast, hails from Wilmington, Massachusetts — the modern-day headquarters of Analog Devices, by coincidence — and spent time on both coasts, as well as stints in Scotland and Germany while at National. Despite these differences, the two shared a strong passion for technology, their companies, and, most importantly, getting the most out of their employees.

LESSON 3: DIG DEEP

"Swanson was a taskmaster," said Reay. "His genius was understanding where everyone's peak performance was. He would push each person to be their best — it's a different level for each person. We used to go into meetings where we knew there would be bullshit. We used to take bets to see how many questions it took Swanson to ferret out the bullshit. He would always figure it out. If you made stuff up, you were verbally killed. It was tough!"

Reay offered an example from his own experience: "In one particular instance, I had just come out with the LTC485. One of the issues was latch-up. When it gets zapped with ESD [electrostatic discharge] while transmitting data, the part injects all the energy back into the substrate and latches up. The easy way to solve this was to change the starting material to a richly doped P-type wafer, then deposit the lightly doped P-type epi substrate. It's a very well-known process.

"I go to this meeting where I'm proposing this epi process. Swanson looks at me and says, 'Reay, I can hire anybody off the street to come and put this on epi. In fact, if that's all it takes, I don't need a high-priced Stanford guy like you. I hear Maxim's hiring; maybe you oughta go over there!'

"After that meeting, I couldn't sleep for about three days. Eventually, I figured out how to solve it. What I didn't realize until later is that while most products were on the bipolar process, most new designs were on CMOS. So in production, the CMOS-to-bipolar ratio was about to change. Epi costs an extra $30 per wafer. Our CMOS wafer costs, as a company, would have gone up 30%. By having a public beating of me, in front of the whole design group, Swanson made sure that we didn't start using epi for another two to three years, and the company made a lot of money, which we handed back as profit sharing. That was Bob's genius."

Gifford shared the same ability as Swanson. From Doluca's book, "Maximum Impact":

> *As always, whether negotiating for an event, a product price, or a company purchase, Gifford always walked into the deal with a strategy in mind. "He's trying to get the best out of everybody," says [Steve] Combs. "He's going to take whatever they said was their best and he's going to tell them it's crap … knowing that they were going to do better when the real deal came through." No matter the situation, "he had his game face and he used it. He knew how to push people."*

Said Reay, "I never worked for Gifford, but he was the same way. They're both good at reading people. They're really tough. Stories of throwing chairs against the wall, Swanson putting a pen through his fist — they're all true."

Doluca's account mirrors Reay's recollections. "Yes, Jack was a tough manager who perfected the art of tough love," Doluca writes. "Many are the unique and legendary water cooler stories shared

about him at tech companies in recent decades, and most are completely true."

LOOKING FORWARD

The analog thread of Silicon Valley history added another stitch in August of 2021, when Analog Devices completed the acquisition of Maxim, bringing together the three legendary high-performance analog companies (ADI acquired Linear in March 2017). ADI's legendary leader, Ray Stata, retired as chair of the board at ADI early in 2022, a role he served since 1973. He still frequently visits the company. The imprints of Swanson and Gifford, who passed away in 2008, remain in the people who carry forward the torch of analog semiconductors and other industries to which they have migrated.

Silicon Valley, always pushing forward, remains the center of the technology universe. Filling the vacuum of the Fairchild of the '60s are now Facebook, Google, Apple (the old guy of the group), and multitudes of software and internet companies, as well as those pushing the edge of technology in health care, artificial intelligence, and green energy, among other industries.

With analog semiconductors firmly embedded in the history of Silicon Valley, challenges and opportunities continue to emerge. The climate crisis, brought on by human-induced greenhouse gas emissions, now threatens human habitats across the globe. Silicon Valley suffers from drought and wildfires that plague California and many parts of the world. This crisis, however, means a challenge and an opportunity for passionate people to try and fail, to dig deep, to be entrepreneurial, and to solve problems. How can analog semiconductors help?

Reay turned again to history to answer that question: "Bob Dobkin did the first three-pin LDO linear regulator. He did it as a skunkworks project — really innovative. Now, LDOs are everywhere, just a standard thing. When I started, we did some of the first switching regulators, [and] TI was also doing them — it was an arms race. A lot of the innovation was driven by the digital side. The first digital processors were powered by linear regulators, but as the power levels kept going up, it just became obvious that was not a good approach, and you had to get more clever with the power architectures.

"There's been a nice steady progression of new types of architectures. The basic premise of converting one voltage to another as efficiently as possible, without heat and energy dissipation, and trying to shrink the size down as much as possible — that hasn't changed.

"There are lots of new and interesting architectures. The Silent Switcher is an interesting breakthrough because buck regulators have been around for a long time. Then all of a sudden, through some very clever observations, you can use magnetic fields to cancel each other out, and get the electromagnetic interference way down. Or you have the coupled inductor technology that came from Volterra and through Maxim, another example of really clever architecture.

"What I've seen as the biggest change in the last 10 years is that the innovation in power design used to be silicon-based, and the packaging didn't change much," Reay said. "There were new architectures, processes, and materials — you bring in gallium arsenide and gallium nitride, for example. But about 10, maybe 15 years ago, the silicon wasn't changing as rapidly as the packaging was. All of a sudden, we started hiring a lot more packaging engineers. A lot of the innovation in trying to get the density in power was not coming from the silicon innovations; it was just as important to have innovations in how you did the package, such as going to copper pillars to get rid of the bond wires or using laminate substrates where you could put the bypass caps right on the die or using bus bars."

Reay noted, "You look at the density of what's going on in the µModules, where you're putting in discrete components, bare dies, bus bars, and it's all incorporated into a tiny space. The mechanical part of the design is just as important as the electrical part. I see the same thing with increasing voltages and the requirements of electric vehicles. There's been a lot of innovation in the high-voltage, high-power device types."

Where is all of this going? "More and more, you need to expand beyond the basic building blocks," Reay said. "But we can't forget that we have all of this analog expertise. We still have to invest in the basic analog functionality. The thing we [ADI] have going for us is this 50-year history of innovative analog design."

How about solving the climate crisis? "I think there are a lot of young people coming out of school — including my son, who works at a green energy startup — who view the energy space as something we need to solve at a global level," said Reay. "I see a lot of young people who really think about this; they're thinking, 'Is the product I'm working on going to have an advantage in the long run?' I didn't think about that at all.

"There are a lot of interesting developments that we don't even hear about — a lot of people, startups, and excitement," he said. "The solution to these problems involves looking at human behavior: Where is all of the energy going? Some of this is not rocket science.

"I talk to young people; I look at them and I'm so excited for them," he added. "They're on the ground floor of a bunch of new stuff. I hope they can all feel the same excitement that I felt at that age. It makes me feel confident about the future."

The untold future of Silicon Valley will undoubtedly rely on characters — those who stick to their passions, are willing to fail, and dig deep to solve difficult problems. As the analog industry and technology at large follow an unpredictable, unique path, the future of climate solutions and energy efficiency will also be unique and unpredictable. The hope, excitement, opportunities, and challenges will always be there. So here's to another 50 years of passion, failure, digging deep, and solving problems.

Reay concluded with some sound advice: "Whatever it is that got you into the business in the first place, that you really love, and you're really good at, keep doing that." ■

A 50-YEAR HISTORY OF THE MICROCONTROLLER

BY TIM BURGESS AND BERND WESTHOFF

Microcontrollers (MCUs) have been with us since the early 1970s and are now celebrating 50 years of existence. During that half-century, MCUs have become essential components in the electronics toolbox. Starting from humble beginnings as programmable calculator chips, MCUs are now integral to the design of electronic products including automobiles, medical products, consumer devices, toys, televisions, radios, appliances, and many other embedded applications occupying every electronics niche. MCUs punch far above their weight class. That's why they've become ubiquitous.

Intel introduced the first commercial, single-chip microprocessor, the 4004, on Nov. 15, 1971. Although it was designed for embedded applications (what we now call "edge" applications), the 4004 microprocessor required many additional chips to create a complete system. The 4004 lacked on-chip RAM, ROM, and I/O functions required by an MCU.

The first chip that could be called an MCU was the Texas Instruments (TI) TMS1802NC. It was developed because many TI customers planned to build electronic calculators. They all wanted calculator chips, and they all wanted their chips to be just a little bit different. Each set of calculator chip requirements demanded a new chip design, so TI defined a mask-programmable calculator chip with a 4-bit processor, 182 bits of serial RAM to hold the calculator data, a 3,520-bit ROM that held a custom calculator program for each customer, and the I/O circuitry needed to drive a seven-segment display and read a matrix keyboard. TI named the chip the TMS1802NC and announced it on Sept. 17, 1971 — two months before Intel's 4004 introduction. Figure 1, showing the TI calculator chip patented under U.S. patent 4,074,351, clearly shows all the essential MCU elements.

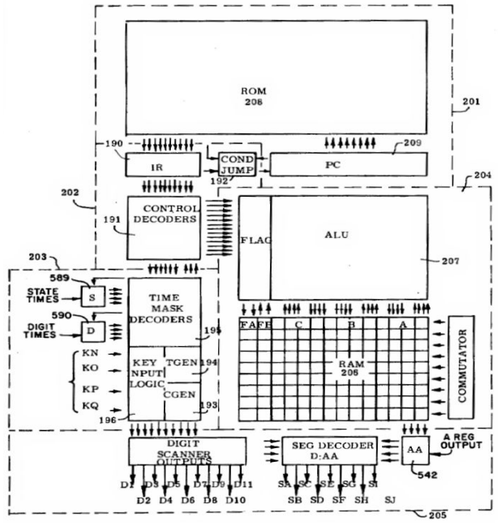


Figure 1: TI calculator chip (Source: U.S. Patent and Trademark Office)

TI rebranded the TMS1802NC as the TMS0102 in 1972, and it became the first member of the company's TMS0100 ROM-programmable calculator chip family. By 1974, the TMS0100 family included more than 15 devices. These devices incorporated all the essential elements of a single-chip MCU: CPU, RAM, ROM, and I/O. However, the TMS0100 chips were not general-purpose MCUs; rather, they had highly specialized I/O, suited only for building calculators. TI addressed that problem by retrofitting general-purpose I/O circuits and announcing the 4-bit TMS1000 MCU family in 1974.

Intel announced its first MCU, the 8048, in 1976 — five years after the company announced its first microprocessor. The 8048 had an 8-bit CPU, 64 bytes of RAM, and a 1-KB mask ROM. Its CPU architecture and assembly code bore no resemblance to any of the company's earlier microprocessors. Like all early MCUs, the 8048 had a mask ROM. One major innovation introduced with the 8048 was the field-programmable EPROM version, the 8748, which had a clear window carved into the package so that the EPROM could be erased with ultraviolet light. Intel's 8048 met with some success as a peripheral or keyboard controller as well as in the automotive market, but the company's second-generation 8051, introduced in 1980, became hugely successful and remained so for decades.

Other semiconductor vendors also jumped onto the MCU bandwagon during the 1970s. In 1977, Motorola Semiconductor introduced the 6801, an MCU version of its 6800 microprocessor. The 6801 had an 8-bit CPU, 128 bytes of RAM, and a 2-KB mask ROM. Like the 8048, the 6801 was offered in an EPROM version with a windowed package called the 68701. The 6801 found early success with General Motors.

Fairchild Semiconductor announced the multichip, MCU-like F8 family in 1974. Mostek, an official F8 second source, introduced a single-chip MCU based on the F8 called the MK3870 in 1977. Although little known today, that device became successful as an embedded controller. Mostek offered a bond-out version called the MK38P70, which accommodated a piggyback EPROM through an IC socket built into the top of its ceramic package, which some equipment makers used for limited production.

Zilog introduced its Z8 MCU in 1979. Federico Faggin founded Zilog in 1974 after having led the design of Intel's first four microprocessors. He initially planned to offer an MCU but decided to challenge his former employer's microprocessors instead. Zilog's first product, introduced in 1976, was the Z80 8-bit microprocessor. The Z8 appeared three years later and bore no resemblance to the Z80.

The MCU market exploded in the 1980s and 1990s as semiconductor makers in the U.S., Europe, and Japan raced to introduce new 8-, 16-, and 32-bit MCUs. Japanese semiconductor suppliers, including Hitachi with its 8-/16-/32-bit H8 MCU platform, Mitsubishi Electric with its 16-bit M16C MCU family, and NEC with the 32-bit V850 MCU, became market leaders in many embedded product categories because of their advanced flash technology and wide product portfolios. Those three major MCU players have since merged to become Renesas Electronics Corp.

These early MCUs all used proprietary CPU cores, but that situation is now evolving. In 2004, ARM introduced small, 32-bit Cortex-M processor IP cores aimed at the MCU market.

ARM's compressed, 16-bit Thumb-2 instruction set makes these CPU cores attractive for MCUs because they make more efficient use of limited, on-chip program memory. More recently, history may be repeating itself with the 32-bit RISC-V architecture, which also offers a compressed instruction set that makes this CPU attractive for use in MCUs. Many vendors offer MCUs based on ARM and RISC-V CPU architectures in addition to MCUs based on proprietary, high-performance CPU architectures.

Vendors continue to add new capabilities to their MCUs, including three of the most requested features: cloud and wireless connectivity, security, and artificial intelligence and machine learning (AI/ML). MCUs with Ethernet ports, support for Wi-Fi and Bluetooth wireless connections, and the required software networking stacks have already appeared. Equally important are newly announced partnerships with major IoT cloud service providers such as AWS and Microsoft Azure. The need for integrated security features increases with the addition of cloud connectivity, so some MCU vendors have made security integral to the design of their latest devices. AI/ML capabilities open the door to a slew of new MCU applications, such as smart sensors, predictive maintenance, video processing, sound/voice processing, and vibration analysis. Figure 2 illustrates the type of MCU available today and the results of 50 years of MCU development.

The current-generation MCU shown in Figure 2 incorporates a 200-MHz, 32-bit ARM Cortex-M33 CPU with floating-point and DSP extensions and ARM's TrustZone security, 512 KB of SRAM, 1 KB of battery-backed SRAM, 2 MB of code flash, 8 KB of data flash (used like E^2PROM), multiple serial I/O ports including Ethernet and USB, 12-bit ADCs and DACs for analog I/O, a security/cryptographic engine, timer/counters, and a real-time clock. By every measure, this MCU can handle tasks that are orders of magnitude greater than those original MCUs from the 1970s.

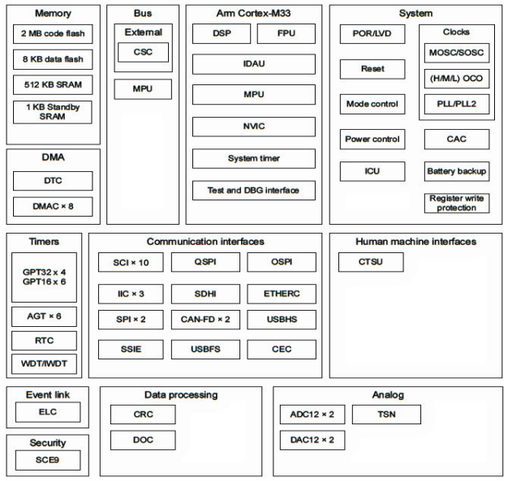


Figure 2: Block diagram of a Renesas RA6M5 MCU (Source: Renesas Electronics)

The new capabilities illustrated in the MCU diagram, and the new applications that such devices enable, point the way toward the next 50 years of MCU development.

Fifty years ago, MCUs were nowhere to be found. Now, we see them everywhere, and there's no sign that this will change.

Tim Burgess is senior director for the high-end MCU business and **Bernd Westhoff** is director of global MCU product marketing, both at Renesas Electronics.

EXCLUSIVE: AN INTERVIEW WITH CARVER MEAD

BY NITIN DAHAD

While in Silicon Valley for the 59th Design Automation Conference, I had an exclusive interview with legendary electronics engineer and applied physicist Carver Mead, the Gordon and Betty Moore Professor Emeritus of Engineering and Applied Science at California Institute of Technology. Mead had just been named the 2022 Kyoto Prize laureate in Advanced Technology. In 1997, the prize went to Federico Faggin, Mead's fellow Synaptics founder.

According to the nonprofit Inamori Foundation, which awards the Kyoto Prize to honor those who have contributed significantly to humankind's scientific, cultural, and spiritual betterment, Mead's pioneering contributions to electronics include proposing and promoting a methodology to divide the design process of very large-scale integration (VLSI) systems into logic, circuit, and layout designs and separating them from the manufacturing process. He also contributed greatly to the advancement of computer-aided design technology, paving the way to VLSI design automation, which in turn led to the rapid development of the semiconductor industry.

I sat down with Mead to find out more about the achievements that had earned him the award, the standout moments in a career that has included being Badge No. 5 at Intel and launching Synaptics, and what he is up to now. He also shared his views on today's neuromorphic computing and whether it can ever reach anywhere near the efficiency levels of the human brain.

Q: First of all, congratulations on the Kyoto award. How do you feel about that?

Carver Mead: It's very satisfying, because it's the first time that it's been noticed that there was a lot of work early on to get the content that went into the VLSI courses. That was hard work, and there was nobody around watching. The way people were doing it was nuts. I mean, they would figure out some system definition and then they'd hand that off to somebody who'd make some logic equations for it, and then they'd hand it off to somebody who'd go and make logic diagrams for it. And then they'd hand that off to somebody who'd turn those into circuit diagrams. And they'd hand that off to somebody who'd go and make a layout for that circuit diagram. And it was all done by hand, on Mylar.

And then when they went to make a mask, they had this drawing that had all the process layers on it, and they had to make masks for each separate one. So they would take a rubylith. You can see through the red layer to the Mylar

Carver Mead (Source: Caltech AMT)

below, but what they do for each layer is cut along the edges of the shapes very precisely with, basically, a razor blade — for the whole chip, just that layer! And then they'd give that to someone who had to go around with tweezers and pull out the little strips that had been cut.

I took a look at that, and I said, "There's no way that I could do that myself, and there's no way that scales." I had just done the scaling stuff for how far you could go because Gordon had asked me how small we could make the transistors.

Q: That's Gordon Moore?

Mead: Yes. And I had figured out we could at least get down to the 10-nm range. Well, what I actually did was figure out we could get to maybe 3-nm–thick gate oxides. They were at a 100 at the time. So we could go a factor of 30 in scale, a factor of 1,000 in density. And so that meant we were going to make integrated circuits with millions of transistors on them. Well, there's no way you're going be able to do that with the process they were using. So I had to think through not only how to make masks but how to do the whole design process. You're not going to draw a logic diagram for something with a million transistors. You need a more structured approach to the whole design process. So I had to figure that out for myself. And I chose to do my own chip. That was all in the late '60s.

So finally, by 1971, I had figured out enough to make my own chip. And then I got it fab'ed. Fortunately, I had some former students at Intel who would run it through the fab for me. And when that chip worked, it was just astounding because there are so many levels of abstraction. [At Intel,] I was a consultant; I wasn't an employee, but I was Badge No. 5. It was Bob Noyce and Gordon Moore, and Jean Jones was the admin, and Arthur Rock. So that was the original founding group. It wasn't called Intel then. They didn't get the name Intel until Andy [Grove] joined.

It was thrilling to be part of that, but when I saw the way they were doing the design, it didn't make any sense. It was not going to scale.

Q: It's actually EE Times' 50th anniversary this year. Was there a landmark you achieved in 1972, when EE Times was born?

Mead: In '71, I got my first chip working. There are so many levels of abstraction to get from a system idea to a working piece of silicon. Until you do it, you're not sure if you've missed something someplace. So when that chip worked, it gave me confidence. And then, of course, the students had been watching what I was doing, and they said, "We want to learn how to do that." Dick Pashley came and said, "Will you teach a course in that?" And I said, "Well, if you can get a dozen students, sure, I'll do it." Well, he got eight students. Of course, if you're going to teach people how to do it, you have to enable them to do it. So we did this multi-project chip in 1971, which came back in January 1972. And all the students had this big "aha" moment when their chip actually worked. That was the first VLSI class, '71–'72 — and that class had the seeds for what became the structured design methodology and the use of pattern generators instead of hand-drawn things.

Q: What inspired you to get into either electronics or chip design?

Mead: That's a really good question. For me, it was in 1968. I got invited to give a talk at the Device Research Conference, a little workshop that was done every year by the IEEE. They invited people doing leading-edge device work in the U.S. There were only maybe 30 of us then, and we could all sit in one room and hear about the newest stuff that people were doing. They forbid you to take pictures of anything, so it was just people talking about the latest stuff. That year, they invited me to give a talk, so I talked about the scaling, and in the process, I discovered this thing about how scaling was going to go: The devices got smaller and didn't draw any more power per unit area. And they got faster. I mean, it was the biggest violation of Murphy's Law that I think there's ever been! And on the flight on the way home, I was thinking, I've been working on the physics of the transistors, but that's not the problem. The problem is, how do you make a thing with a million moving parts? It's never been done. It just changed my life. I had to do it and figure it out.

Q: That is quite visionary. I mean, who would imagine we could get a million transistors on a chip at that time, when the geometries were so large?

Mead: Well, I had lots of arguments, of course, because people didn't believe it. So I actually spent quite a bit of my time going around giving talks, just to try to get people to believe that it was within the laws of physics that you could make transistors that size.

Q: Tell us about the birth of Synaptics.

Mead: That story starts in '81, when Dick Feynman and John Hopfield and I started the Physics of Computation course at Caltech, because we thought that there were deeper ways of understanding computation than just Turing machines. We were having lunch one day and arguing about this, and they said, "The sure way to learn about it is to teach a joint course on it." So for three years ['81–'83], we taught a joint course where we rotated who would give the lecture. And of course, none of us had finished ideas. This was all trying to get our heads around an impossibly enormous question. But it was thrilling.

Q: You'd been teaching. So what was the reason for starting a company like Synaptics then?

Mead: From my former lifetime, I had known all the people at Fairchild and the people who had moved over to Intel. I had become friends with Federico when he was working with M. Shima on the 4004. When they formed Zilog, I would go by and see what they were doing and try to talk them into doing structured design.

Q: And have more arguments?

Mead: Yes. It's the way it is. Federico and I had been friends for years. One night, we went up to — I think it was the Mountain House — for dinner, and while driving back, we were talking. Federico had already kind of gotten it in his head that there's a company here. And I think he had a little start on one. So he said, "Well, let's do this together."

Q: One of the things we talk about a lot is how everyone is doing all these neural networks. How closely should we be copying the neuron in silicon? Neurons evolved within the constraints of their biology. Is it wise to copy that, given the constraints of silicon? How do we know we aren't just copying neuron housekeeping functions that keep the neuron alive?

Mead: Nobody knows the answer. I mean, the simplest idea of a neural network could be something that learns with examples, and backpropagation, [proposed in the late '80s], was a brilliant insight. ... The idea there was to learn from examples, which we do in neurons, and that one idea, with a bunch of insights having to do with implementation, has turned into big business in mainstream computing today because things got to a scale and the techniques got good enough that it could do useful work.

People are just getting to the point where they're using vision chips that look for the relevant information in the image instead of just scanning out every image and then trying to figure stuff out from that, which is insane — it doesn't scale well. It took 30 years for there to be an urgent need for vision systems that didn't have big latency. But once people decided they wanted to make self-driving cars, then you needed vision systems that didn't have big latency. It takes that long before there's some connection between the technique and a commercially viable product direction.

Those things can happen fast in software because everything's digitized already. But even in software, it took that long before the deep-learning stuff took off. There were basically no new ideas there, [just] the path of evolution of how you

do it and how you use silicon to do it, which wasn't at all obvious to people 30 years ago. If it's just the application of stuff we know already, that can happen very fast, because the platform is there. But if you have to build the platform — the intellectual platform as well as the physical platform — that takes longer, because as part of that evolution process, there has to be a commercial product at each step, or else it can't keep going.

Q: Where do you think we are with neuromorphic chips today?

Mead: Vision systems have pioneered the important idea that it's the changes in the information that are meaningful; it isn't the mass. In your visual image, the picture is nice, but what you act on are the changes. That became the beginning of event-driven computing. And that's a deep idea. It sounds trivial, but we're just at the beginning of building the event in as part of the way it works. … It sounds obvious, but it isn't at all obvious how you actualize [it]. Dynamic vision is the first place where it's kind of hooked in. But it takes that long.

Q: With neuromorphic computing and trying to emulate neurons, can you get to the efficiency level of the brain?

Mead: It is astounding how much effective computation gets done in the 20 W in our brain. And that is really what we [wanted] to figure out when we started the whole neuromorphic thing. We wanted to understand that phenomenon: How can it possibly be? Once you've tried to make applications that do anything even remotely like what animals do — even insects can do better than any of our self-navigating robotic [systems], and they're little things and run on a milliwatt.

It's astounding. We still don't understand it. We've got some insights, and it's helped the interface between neurobiology and synthetic computing — making chips that do stuff is a very rich area. It has just begun to generate things that are commercially viable.

Q: So does analog computing play an important part in that?

Mead: It's difficult to see what should be done in analog and what should be done in digital. In the neural system in brains of animals, the signals that go over any appreciable distance are all digital — the nerve spikes. The computation in the dendritic tree of neurons is all analog, or it's a combination; you have signals that come from the nerve spikes of other neurons and you're aggregating those in an analog way, but they're quasi-digital in nature.

No one has yet been successful in building a thing that works like the dendritic tree of neurons. As a technical achievement, to realize a thing that works like a real dendritic tree requires a level of gain control and stability that's beyond anything that has been done. When I finally gave up, I was trying to do that.

And of course, the technology has evolved to be more digital. We still have analog stuff at the sensory end, so maybe that's where the next thing is going to happen.

Q: What are you up to today, and what gives you the most joy?

Mead: The thing I did after the neuromorphic stuff was a simpler and more unified way of looking at electrodynamics and quantum physics. They are really one discipline [but] have been taught as separate disciplines … so then the students never quite get them to fit. I've done a first pass-through of what you would do to make that one discipline, and it turns out you can do it at the first level. It works much better for both disciplines, and they fit together. So that was very satisfying, but that was in the year 2000.

I just did a new version of that a couple of years ago with John Cramer. We have a paper on it that came out two years ago in Symmetry and has some insights beyond what's in the little book "Collective Electrodynamics" that I wrote in 2000.

Q: You've been awarded this lifetime achievement prize. What is the one thing that you feel really proud of as your legacy?

Mead: For the period that was addressed by the Kyoto prize, it was the development of a new way of looking at digital design that recognized that it would scale to a very large scale, so it had to be a more system-level design. It had to incorporate the physical properties of microelectronic technology, and that had to be done as a unified thing, not as separate disciplines. Each step of the way had to fit with the one before it, or you didn't end up with a thing that worked, and getting that all fit together was really what was honored in the Kyoto prize. That was very satisfying because it was a period when nobody cared about it: It just had to be done, and once it was done, it looked obvious.

But the thing I'm the most pleased about is what I call collective electrodynamics — the development of electrodynamics from a quantum basis rather than from some funny mechanical ideas.

Q: You were quite a visionary when you understood the potential of scaling transistors and materials. What's your vision for any period of time in the future now for silicon, and are we doing the right things or is there something that we should be doing differently?

Mead: I'm not close enough to everything that's going on in microelectronics to make a cogent statement about that. It's become a huge field. It's wonderful what's happening. But as always, there needs to be a next important idea. And if I knew what that was, I'd be doing it.

Nitin Dahad is a correspondent for EE Times and EE Times Europe and is editor-in-chief of embedded.com.

PARTNER CONTENT

AN ELECTRONICS INNOVATION JOURNEY WITH EE TIMES

BY K.T. MOORE, *VICE PRESIDENT OF CORPORATE MARKETING, CADENCE*

EE Times is 50 years old this year, having first published in 1972. As it happens, it is almost Cadence's 40th anniversary, at least if you squint when you look at it. So in this piece, let's take a look at the early history of Cadence and how it grew to be the company it is today through wise acquisitions that changed the industry. And all along the way, EE Times reported on these developments.

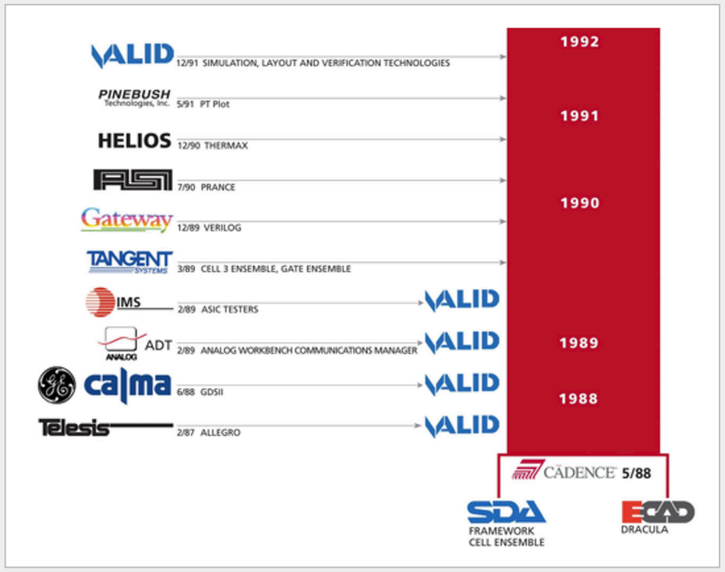
Cadence was created by the merger of SDA and ECAD. Even Cadence executives think that Cadence went public in 1987, and it did indeed file the documents to do so, but the initial public offering (IPO) never took place. Nonetheless, Cadence is a public company, so obviously there is more of a story there.

The company got its start when Jim Solomon left National Semiconductor, where he had been director of IC design for analog and mixed-signal, to found SDA (which, depending on whom you talk to, either stands for Solomon Design Automation or Silicon Design Automation). SDA was based in Silicon Valley. He funded the company by securing half of the capital from VCs and half from four companies (including his former employer). In exchange for their investments, the companies were able to send two engineers apiece to work there and had access to the software that resulted. At the time, there wasn't really an EDA industry, and top-tier semiconductor companies developed their own tools, leaving companies that had fewer resources without many good options. In that era, funding a company for access to design tools was an attractive proposition.

In 1987, SDA filed to go public, but the day that its IPO was to take place turned out to be Black Monday, the biggest one-day fall in the stock market ever. The IPO was pulled.

Meanwhile, Glen Antle and Paul Huang had launched a company called ECAD to commercialize physical verification — in particular, a design rule checker (DRC) called Dracula. ECAD went public in 1987, before the IPO window closed on Black Monday.

Cadence came into existence in 1988 when SDA merged with ECAD. Since ECAD was already a public company, this was a sort of way for SDA to go public. The merged company rebranded as Cadence, and Joe Costello, who had been SDA's COO, became its first CEO.



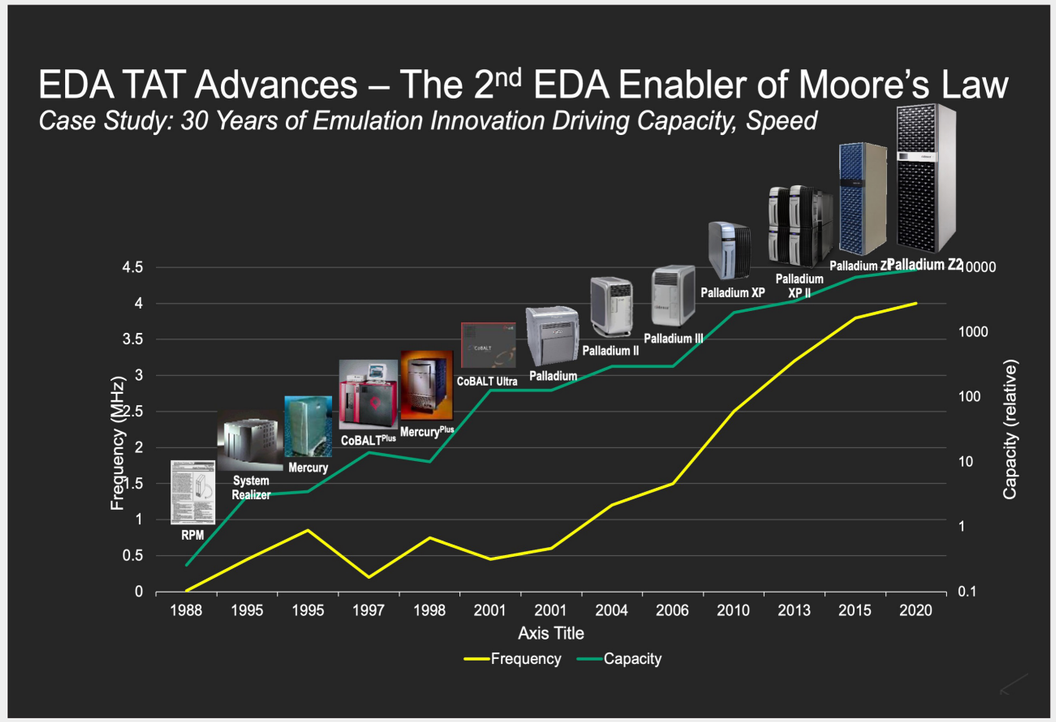
Four key products from that era were Analog Artist, the Virtuoso layout editor, the SKILL scripting language, and the Spectre circuit simulator. All four products are still widely used today. Another key product was the Dracula DRC, which became the industry standard for decades, although it has been superseded by Cadence DRCs with other names.

The first key acquisition that the young Cadence made, in 1989, was Tangent Systems. The Tangent products were Tancell (for standard cell with two layers of metal) and Tangate for gate array. When a third level of metal was added, Tancell became Cell3. Cadence renamed them to Cell Ensemble and Gate Ensemble.

Next, in 1990, Cadence acquired Gateway Design Automation. Gateway was the creator of the Verilog language and register-transfer level (RTL) simulator and also had an extremely fast gate-level simulator, Verilog-XL. This would turn out to be a prescient acquisition as design abstraction moved from the gate level up to the RTL level, and the simulation business grew explosively. Verilog became a de facto standard and eventually an IEEE standard. Since then, Verilog has been joined by VHDL and SystemVerilog as significant languages, and that initial simulator has grown into today's Xcelium simulation platform.

Cadence had a schematic capture product for analog design called Composer, but it was not widely used for digital design. In 1991, Cadence acquired Valid, which had a front-end design system widely used for both gate-array and PCB design. With that move, Cadence was no longer just an IC design tool provider; it became a PCB design tool company as well. The acquisition also made Cadence the largest EDA company at the time.

During this period, Cadence was dominant in the merchant market for tools for semiconductor design, and the company grew fast. Tools whose names are still famous today, such as



(Source: Cadence)

Dracula, Virtuoso, Silicon Ensemble, and Verilog-XL, had explosive growth as semiconductor companies switched from internally developed tools to commercial tools and as system companies started to design their own semiconductors. This typically was accomplished using an ASIC methodology, where the design close to the system was done by the systems companies and the design close to the silicon was done at the semiconductor companies. This was also a period of explosive growth for the PC industry and its surrounding ecosystem, which had a lot of semiconductor content.

OTHER KEY DEALS

There were many smaller acquisitions during the following years under Costello and the next two executives to hold the CEO post: Jack Harding and Ray Bingham. Here are a few of the key ones.

By the late 1990s, synthesis had become the dominant design methodology, and Cadence had internally developed a product in the space called Synergy. In 1998, Cadence acquired Ambit Design Systems, with its BuildGates synthesis product. Cadence would also acquire Get2Chip in 2003.

Cadence entered the emulation business in 1999 with the acquisition of Quickturn but, more significantly, in 2002 acquired the emulation business of IBM. This technology has been the heart of Cadence's emulation product line ever since, including the current Palladium Z series.

A FEW MORE KEY ACQUISITIONS

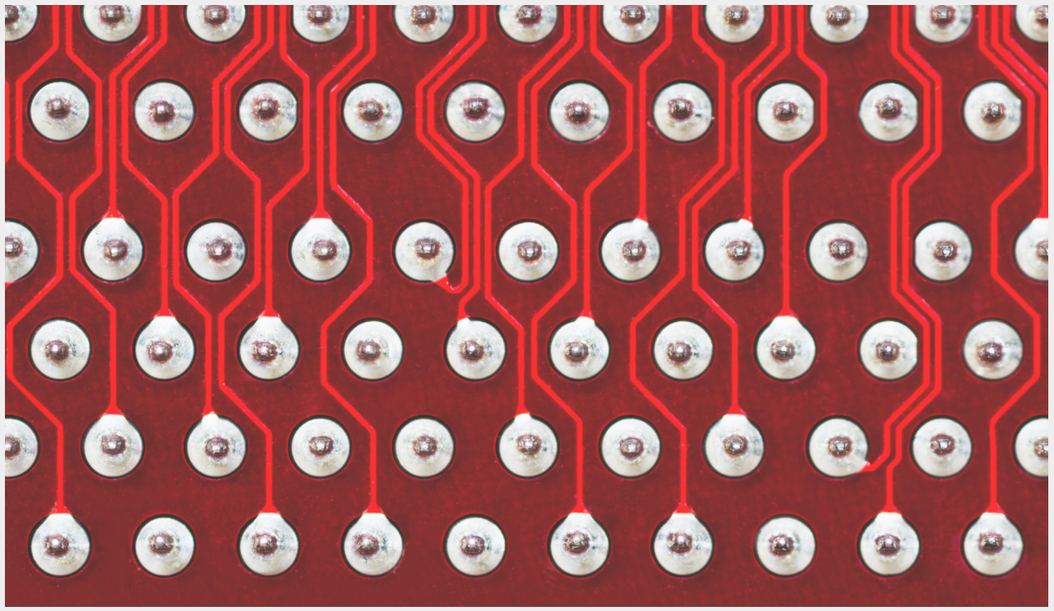
- **Silicon perspective** (floorplanning)
- **CadMos** (signal integrity)
- **Simplex** (extraction and analysis)
- **Verisity** (constrained random verification and the "e" language)
- **Denali** (memory compiler VIP and DIP, Cadence's first foray into IP)
- **Tensilica** (processor IP for vision, audio, AI, and more)

During this period, the key journalist covering EDA was EE Times' Richard Goering. Every marketing executive in the EDA industry would hope that Richard's piece on their latest product release would be on the front page of EE Times when it showed up in their mailbox (a physical one in that era) on Monday.

Perhaps surprisingly, Richard's background was journalism school; he was not an engineer who had decided to become a journalist. Given how technical EDA was and is, this was an achievement. If you google most of the acquisitions mentioned earlier, the first result will usually be an EE Times article written by Richard. He would later go on to work at Cadence as an in-house journalist and writer before he eventually retired in 2015 to a farm in Watsonville, California.

Of course, Cadence has a more recent history under CEOs Lip-Bu Tan and now Anirudh Devgan. But the 50th anniversary of EE Times seems like a good opportunity to look at the earlier, perhaps less familiar period of its history.

It is also the perfect moment for the 40-year–old "upstart" company to congratulate the mature, 50-year–old publication for its longevity in such a fast-changing industry as electronics. ■



CHAPTER TWO:

THE SILICON GOLD RUSH

SILICON AND THE SEMICONDUCTOR INDUSTRY: WHAT LIES AHEAD

BY JEAN-CHRISTOPHE ELOY

While silicon remains a constant in our industry, there are already signs of significant changes to the industry models we know today. Let me begin by saying that silicon content in everyday products will continue to increase. But an evolution is under way, driven by four factors, all of which integrate functionality and are reshaping the industry.

The first factor is front-end integration to integrate more transistors — and thus more functionality — per square millimeter of silicon.

Two other trends are more recent and are closely related. One is advanced packaging. Classic packaging was merely a way to interconnect the device to the model board, but now, packaging is used to support further integration. The other is advanced substrate integration, which is replacing the conventional PCB with advanced substrates. We saw the first real system-in-package emerge about five years ago, as the industry moved on from flip-chip packaging and die integration to integrate more functions at the substrate level.

Finally, the supply chain is shifting as smartphone and vehicle manufacturers introduce their own chip design teams rather than rely on Tier 1 suppliers. System makers are also going deep into the supply chain, interacting and integrating directly with Tier 2 or even Tier 3 suppliers.

STRUCTURAL CHANGES: SHIFTING SUPPLY CHAIN

OEMs are diving deeper and deeper into their supply chains, foundries are providing advanced packaging services as part of their global support portfolios, and EMS companies are moving into the lower-cost end of design and manufacturing. The result is that multiple players are either moving down the supply chain and integrating suppliers, or moving up and providing services and products to their customers' customers. At the top of the chain, OEMs are bypassing Tier 1 suppliers and directly accessing Tier 2 suppliers.

This shift started in the smartphone industry, with Apple integrating a lot of design and other smartphone makers following suit. It is also happening in the automotive industry, where Toyota, Volkswagen, Tesla, and other new car makers no longer rely solely on Tier 1 suppliers.

This aggregation of the supply chain is still new but will have a long-term impact on the industry. It will mean that there will be more IDMs providing modules and subsystems directly to customers. It will also mean that more OEMs will be fabless and will design their own chips. Direct access to foundries will allow them to bypass IDMs and module makers.

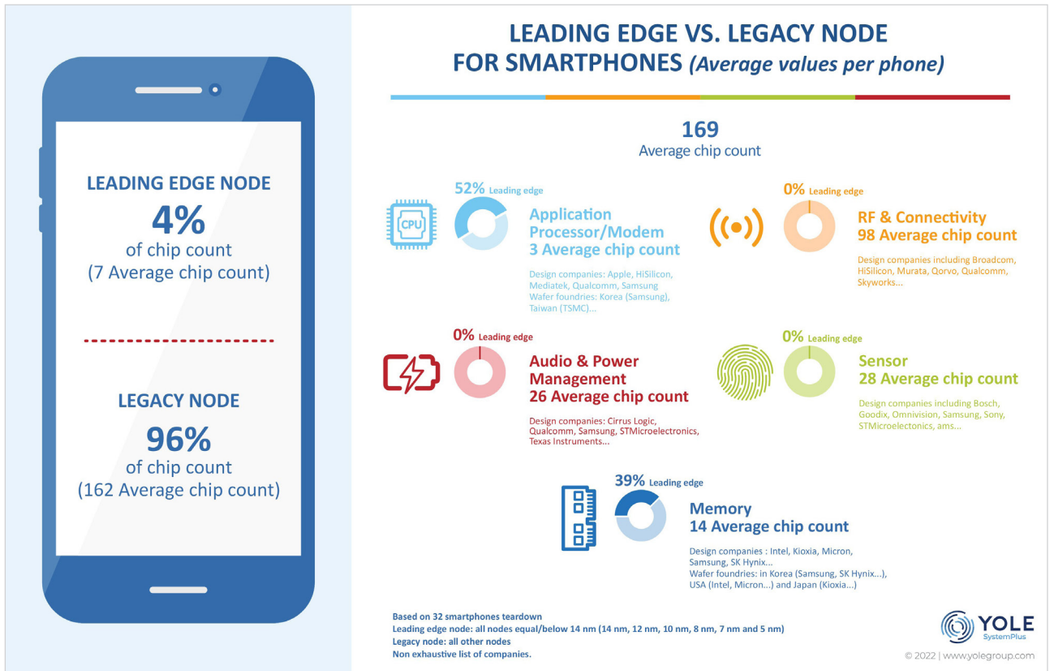
There will be some strategic devices for which OEMs will have a role in the design to obtain functionalities that will enable differentiation in the market.

There are also changes in the foundry business. We see Intel moving into the foundry space and Samsung and STMicroelectronics both increasing their foundry activities. I think this multiple-IDM foundry model will proliferate to ensure continuing business.

For the same reasons, an EMS like Foxconn, which is increasingly heavily involved in the manufacture, packaging, and assembly of devices, will gain new business in areas where it can aggregate value for the customer.

Consider Qualcomm, for example, which once provided application processors in high volumes to smartphone makers but has seen that activity gradually decrease as Apple, Samsung, Oppo, Vivo, Huawei, and others have moved design in-house. As one customer base shrinks, it is important to find new markets, such as augmented/virtual reality or, for the automotive market, broadening out into module development for ADAS.

I think being able to diversify and being able to provide other value to other industries will be the means to finding growth, or at least stabilizing sales, as the supply



2022 European semiconductor ecosystem — localization of the main 150 semiconductor and photonics fabs and 50 main R&D centers (in units) (Source: Yole Group internal data, 2022)

chain evolves. This is a smart approach because all fabless companies rely on a supply chain that is open, so an OEM can create a design team and use the existing supply chain to increase its margins. The OEM is able to design more functionality into its products and work directly with a simplified supply chain.

GEOPOLITICAL CHANGES: CHINA'S CHIP INDEPENDENCE

Another critical shift is the isolation of China. China will inevitably have a semiconductor industry that is quite separate from what we see in Europe, the U.S., Korea, Japan, or Taiwan. New laws introduced in the U.S. to isolate China are, in fact, pushing China to be independent from the rest of the world.

China does not operate on the basis of the next financial quarter's profits; instead, it looks ahead to the next 50 years. This gives it the luxury of time and the ability to find paths to overcome the technology ban decided by the U.S. government.

This is a big and important change because it will mean that we will have two semiconductor industries: a Chinese and a non-Chinese one.

TECHNOLOGY EVOLUTION: COMPOUND SEMICONDUCTORS

Silicon remains dominant, but devices based on compound semiconductors like gallium arsenide, silicon carbide, gallium nitride, and indium phosphide are evolving quickly as performance requirements diversify. SiC is reemerging because its characteristics are required to optimize electric-vehicle batteries and thereby increase EV range; silicon is just not good enough here.

Silicon has done a good job for multiple applications in which data rate is important, but now, there is a movement toward optics and silicon photonics. Photonics, in particular, is growing rapidly. Whereas silicon once accounted for 99% of industry substrates, its share today is closer to 75% or 80%. I am not saying silicon will disappear — indeed, it is still growing — but other substrates for power devices, optoelectonics, and RF applications are growing even faster.

Compound semiconductors are becoming an important business, but their supply chain differs widely from silicon's. RF, photonics, and power devices are forming their own structure, with distinct supply chains, material manufacturers, intellectual property, test equipment, and manufacturing equipment.

LOCAL PRODUCTION: REINDUSTRIALIZATION IN EUROPE, U.S., JAPAN

Yes, semiconductor reindustrialization can happen in the regions that have moved most chip manufacturing offshore. I see Europe, the U.S., and Japan moving toward local semiconductor production rather than being dependent on China, Taiwan, or any other part of the world for supply.

This is a political push, and it's extremely complex to implement. Not every country's domestic industry has a business need for 2-nm fabs, though. Automotive, industrial, and medical applications — even smartphones — are relying primarily on 28 nm and above. Remember, there are more than 150, mostly small and midsized semiconductor fabs in Europe, and they're focusing on semiconductor process reuse to manufacture photonic devices, RF, power modules, sensors, advanced packaging, and the like.

Instead of moving full speed to 5-nm fabs, a business model is evolving under which a fab is owned partly by an IDM and partly by a foundry. An example is the arrangement between ST and GlobalFoundries. Benefits for the IDM are that it can integrate multiple technologies, including more advanced ones. For the foundry, the benefit is that the fab is always working at capacity. This trend started several years ago with Tower and Panasonic, and partnership activity has recently increased.

I think this is a super-innovative move. Sharing between players that are at different parts of the supply chain is key to profitability. It's a matter of sharing the risk and cost of developing a new manufacturing site facility. A facility can be used for an IDM and by a foundry that may serve the same customer.

An added factor is that cross-border mergers and acquisitions are becoming so complex that closing them is increasingly difficult. Recently, there was a failed attempt by GlobalWafers to buy Siltronic, Nvidia failed to buy Arm, and there have been multiple failed acquisitions of French, German, and Italian semiconductor companies by Chinese entities (the aborted deals for France's Unity SC and Italy's LPE are two examples).

I do not believe this will ease over the next five years, because the semiconductor industry is of strategic importance for all countries. There will be more M&As, but they will be lower in terms of pace and frequency, and they will tend to occur between entities within the same country.

Jean-Christophe Eloy is president and CEO of Yole Group.

WHAT IT WILL TAKE TO KEEP CHIP SCALING ON TRACK

BY LUC VAN DEN HOVE

In its industry roadmap, imec makes the case for a paradigm shift away from the von Neumann architecture and charts a path to get there.

Computing power needs are exploding as the volume of digital applications and data processing continues to rise. With the growing use of artificial intelligence to tackle the major challenges of our time, such as climate change and food insecurity, computing requirements are expected to double every six months for the foreseeable future. To handle the exponentially growing volume of data in a sustainable way, we need improved high-performance semiconductor technology. In order to achieve that, we need to address five challenges simultaneously: scaling, memory, power, sustainability, and cost.

No company can achieve the feat alone, but co-innovation and collaboration across the semiconductor ecosystem will enable the continuation of Moore's Law. That is the key message in imec's roadmap for the upcoming 15 to 20 years.

FIVE WALLS AT ONE TIME

- **The scaling wall:** Purely lithography-enabled scaling is becoming increasingly difficult as individual structures of microchips and transistors approach the size of atoms, where quantum effects begin interfering with the operation of microchips.
- **The memory wall:** System performance is confronted with data path limitations between the cores and the memory. In fact, memory bandwidth cannot keep up with processor performance. We have more floating-point operations per second than gigabytes per second.
- **The power wall:** It is becoming more challenging to bring power into the chip and extract heat from the chip package, so we will have to develop improved power delivery and cooling concepts.
- **The sustainability wall:** Semiconductor manufacture has a growing environmental footprint, including greenhouse gas emissions and consumption of water, natural resources, and electricity.
- **The cost wall:** Obviously, chip manufacturing costs, together with the costs for design and process development, may skyrocket as complexity increases.

TEARING DOWN THE WALLS

At first sight, things don't look great for the perpetuation of Moore's Law — Gordon Moore's observation that the number of transistors in a dense integrated circuit doubles about every two years. This prognosis is especially true if we stubbornly stick to Dennard scaling and traditional von Neumann compute architectures.

In its scaling roadmap, imec proposes an alternative path for the future of chip technology, with fundamental changes in architectures and materials, new basic structures for transistors, and no less than a paradigm shift. The imec roadmap will take us from 7 nm to 0.2 nm, or 2 angstroms, by 2036, keeping an introductory pace of two to two-and-a-half years.

First, continual advances in lithography will be key to further dimensional scaling. Traditional lithography uses light, and the wavelength of light is greater than the pattern accuracy required today. That's why extreme ultraviolet (EUV) lithography was introduced and is now appearing in a rising number of functional production belts for volume manufacturing. EUV will take us from the 5-nm generation to 2 nm. To go smaller, we need an updated version of EUV: high-NA EUV, with bigger lenses. These will have a diameter of 1 meter, with an accuracy of 20 picometers. The first high-NA EUV prototype, which is being developed by ASML, will be available in 2023. Insertion in high-volume manufacturing is expected in 2025 or 2026. To de-risk the manufacturing introduction, imec, together with ASML, has set up an intensive program to develop all the key enabling building blocks, such as the mask technology and materials using wet or dry UV resist, metrology, and optics characterization.

At the same time, we will need innovation in the transistor architecture.

Today, almost all chip manufacturers build microchips with FinFET transistors. However, starting with the 3-nm generation, FinFETs suffer from quantum interference, causing disruptions in the operation of microchips.

Next in line is the gate-all-around (GAA) or nanosheet transistor. Built up as a stack of nanosheets, it will offer improved performance and improved short-channel effects. This architecture will be essential from 2 nm onward. Major chip manufacturers including Samsung, Intel, and TSMC have already announced that they will introduce GAA transistors in their 3-nm or 2-nm nodes.

The forksheet transistor, which originates from imec research, is even denser than the nanosheet transistor, extending the GAA concept to the 1-nm generation. The forksheet architecture introduces a barrier between the negative and positive channels, enabling the channels to be closer together. This architecture is expected to enable a cell-size shrink of 20%.

Further scaling can be realized by putting the negative and positive channels on top of each other, referred to as the complementary FET (CFET) transistor, a complex vertical successor to the GAA. The CFET significantly improves density but comes at the expense of increased process complexity, especially to contact the source and drains of the transistors.

In time, CFET transistors will incorporate new ultra-thin 2D monolayer materials with an atomic thickness, such as tungsten disulfide (WS_2) or molybdenum. This device roadmap, combined with the lithography roadmap, will bring us to the angstrom age.

Two other challenges are in play at the system level of these sub-2-nm transistors. The memory bandwidth cannot keep up with CPU performance. The processor can't run faster than the pace at which data and instructions become available from the memory. To knock down this memory wall, memory must come closer to the chip.

An interesting approach for tearing down the memory wall is 3D system-on-chip integration, which goes beyond today's popular chiplet approaches. Following this heterogeneous integration approach, the system is partitioned into separate chips that are concurrently designed and interconnected in the third dimension. This will allow, for example, the stacking of an SRAM layer for Level 1 cache right on the core logic devices, enabling fast memory-to-logic interaction. To achieve extreme high-bandwidth off-module connectivity, optical interconnects integrated on photonics interposers are being developed.

Regarding system-related challenges, getting enough power into the chip and getting the heat out become more difficult. However, a solution is in sight: The power distribution now runs from the top of the wafer through more than 10 metal layers to the transistor.

Imec is currently working on a solution from the backside of the wafer. We will sink power rails into the wafer and connect them to the backside using nano-through-silicon vias in wider, less resistive materials. This approach will decouple the power delivery network from the signal network, improving the overall power delivery performance, reducing routing congestion, and, ultimately, allowing further standard-cell height scaling.

Finally, semiconductor manufacturing comes at a price. It requires large amounts of energy and water and creates hazardous waste. But the entire supply chain needs to commit to tackling this problem, and an ecosystem approach will be essential. Last year, imec launched its Sustainable Semiconductor Technologies and Systems (SSTS) research program, which brings together stakeholders of the semiconductor value chain — from large system companies like Amazon, Apple, and Microsoft to suppliers including ASM, ASML, Kurita, Screen, and Tokyo Electron. The goal is to reduce the carbon footprint of the entire industry. The program assesses the environmental impact of new technologies, identifies high-impact problems, and defines greener semiconductor manufacturing solutions early on during technological development. It allows us to make informed choices when we move to future technology generations and develop new processes.

PARADIGM SHIFT

In the long term, the von Neumann architecture needs an overhaul. John von Neumann saw the digital computer as a system with input, a central processing unit, and an output. But we will need to evolve toward domain-specific and application-dependent architectures, with massive parallelization comparable to the way our human brain works. This implies that the CPU will have a smaller role in favor of custom-made circuits for specific workloads.

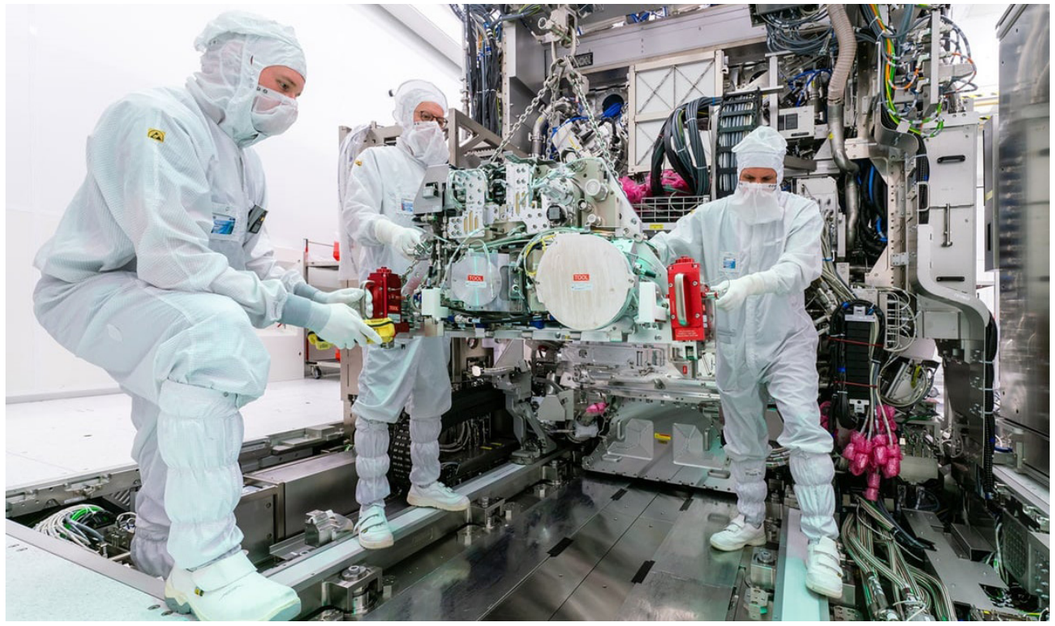
This paradigm shift, in addition to the walls ahead, marks the beginning of interesting times in the semiconductor industry. We will need co-innovation and collaboration across the entire semiconductor ecosystem: foundries, IDMs, fabless and fab-lite providers, and equipment and material suppliers. We must make this shift not just to meet Moore's Law for the sake of it, but because semiconductors are at the core of the performant deep-technology applications that can make impactful progress in tackling the challenges of our time — climate change, sustainable mobility, environmental pollution, and food insecurity.

The stakes are high.

Luc Van den hove is president and CEO of imec.

CHINA'S EUV CONUNDRUM IN THE **CHIP COLD WAR**

BY MAJEED AHMAD



(Source: ASML)

The advent of extreme ultraviolet (EUV) lithography has been an industry fairytale come true, enabling chip manufacture at smaller-nanometer geometries like 5 nm — and thrusting the company that cracked the EUV code into the ranks of tech royalty. Yes, we're talking about Advanced Semiconductor Materials Lithography (ASML), which some industry insiders call the most important tech company you've never heard of.

The lithography technology puts chip patterns on a wafer surface by exposing it to light only once. However, EUV, with its promise of an instant, 14× leap forward in reducing chip size, was much harder to achieve than anyone expected and arrived nearly 20 years late. It required the reinvention of almost everything, from tools to techniques.

EUV was finally ready for deployment just as the trade war between China and the United States began bubbling up, and Veldhoven, Netherlands–based ASML thus found itself at the center of the chip cold war. One of the most prominent trade restrictions leaves ASML unable to export EUV machines to China, which represents nearly 15% of the Dutch company's business.

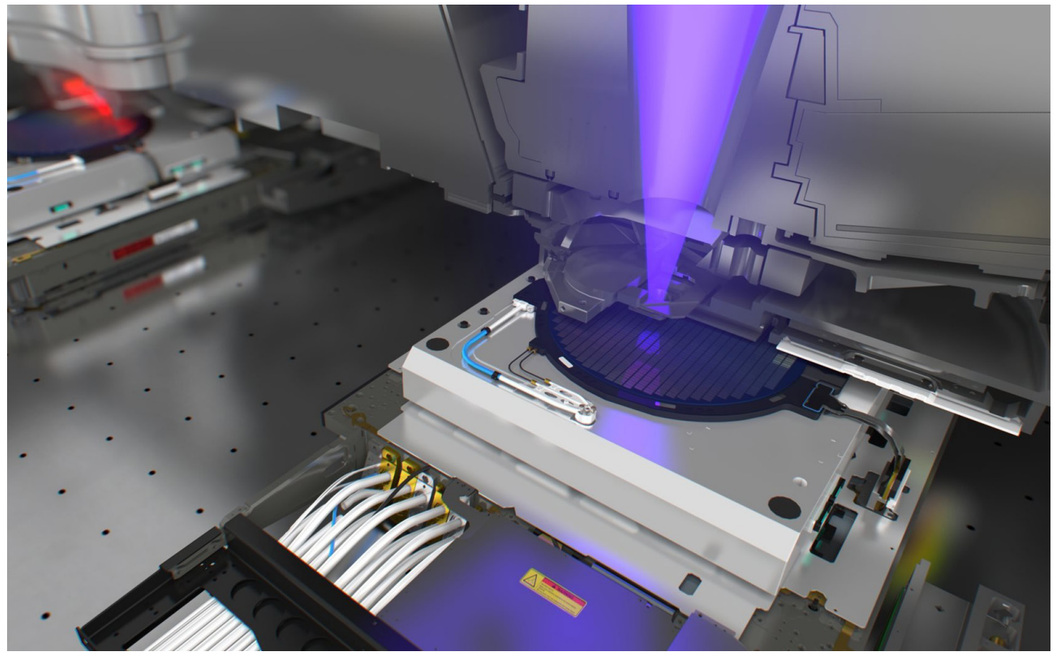
What came next shows the gravity of the situation around the adoption of cutting-edge chip-manufacturing technology and its availability to China's semiconductor fabs. According to industry reports, the U.S. government is also restricting ASML and Japan's Nikon from delivering deep ultraviolet (DUV) lithography machines to fabs in China.

The DUV lithography machines, which have been widely available in the market for years, might be able to deliver the same capacity, albeit at a higher cost. Indeed, DUV technology, also known as immersion lithography, has been serving the 7-nm nodes at TSMC fabs. TSMC's N7 node doesn't use the 13.5-nm–wavelength EUV lithography process to print the wafer patterns that yield chips.

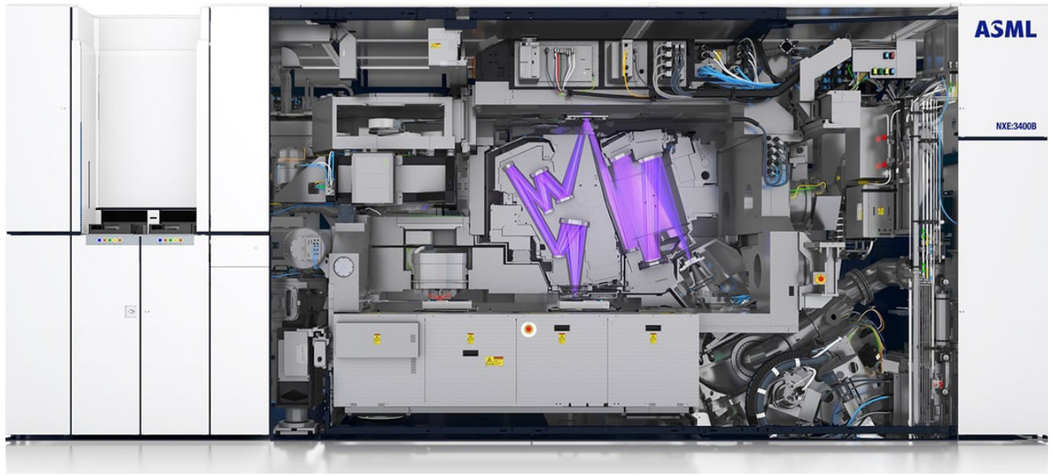
Industry observers believe the DUV-related restrictions won't have much impact, as China's semiconductor fabs presumably have already bought enough DUV machines to serve their needs. The question nonetheless remains why export restrictions were levied on DUV lithography, an old technology that dates back to 1980.

A recent jaw-dropper from Semiconductor Manufacturing International Corp. (SMIC) might partly answer this question. SMIC has reportedly used DUV lithography to manufacture a 7-nm chip for a crypto miner, with extra exposure making up for the lack of EUV.

It's important to note that, in theory, EUV technology isn't essential for



An EUV machine, pushing the boundaries of physics to manufacture chips at smaller geometries, comprises nearly 100,000 parts sourced from about 5,000 vendors, many of them in China. (Source: ASML)



The EUV lithography ban is a major test for China's semiconductor ambitions. (Source: ASML)

manufacturing chips at advanced nanometer nodes. As already noted, TSMC has manufactured 7-nm chips using the 193-nm–wavelength DUV process. DUV machines may require more mask layers, however, translating to more exposure time and added complexity. That, in turn, lowers yield and increases the cost of manufacturing chips.

That puts to the test China's strategic importance for semiconductors. Will the capacity to manufacture advanced chips matter more to China's industry than achieving economies of scale for manufacturing those chips? SMIC's 7-nm chip-manufacturing revelation shows that China is willing to pay the cost of technological breakthroughs. First and foremost, SMIC and Hua Hong Semiconductor were established to manufacture the most advanced chips at smaller nodes. Second, the way the U.S. sanctions nearly bankrupted ZTE and crippled Huawei underscores China's appetite for advanced chips.

SMIC and Hua Hong Semiconductor will likely bear the brunt of the EUV sanctions, and how those companies react to the EUV and DUV technology restrictions will hint at China's broader response to the semiconductor technology export bans. For now, ASML, the sole provider of EUV lithography machines, is adhering to the U.S. export restrictions. The Dutch company has also asked its U.S. employees to refrain from servicing, shipping, or providing any support to any customer in China until further notice.

The EUV export strictures are at the center of a geopolitical storm, and it will be interesting to watch how fabs in China get around this technology conundrum.

Majeed Ahmad is editor-in-chief of EDN and Planet Analog.

EE Times 50

BEYOND SILICON: EXPLORING COMPOUND SEMICONDUCTORS

BY CHRIS McGRADY

As emerging applications raise the bar for power-device performance and as materials manufacturers bring down costs, engineers are adopting recent compound-semiconductor market entries and gaining a renewed appreciation for more mature options.

In recent history, silicon has been the semiconductor material of choice in electronics manufacturing. Silicon's relative abundance keeps its costs relatively low and makes it easy to obtain. Its stability, particularly in its refined state, makes it an excellent choice for use in electronics of all varieties. In recent years, however, compound semiconductors have become increasingly common. This rise in popularity is due in part to process improvements that have driven down the materials' manufacturing costs, but the biggest market driver has been the growth of applications that benefit from compound semiconductors' unique properties. This article looks at the most popular compound semiconductors on the market today, summarized in Table 1.

Gallium arsenide

GaAs is a III-V compound semiconductor that provides very high electron mobility as well as a high saturation of electron velocity, making the material stable even at high frequencies. It is considered a wide-bandgap (WBG) semiconductor compared with silicon and, because of its WBG properties, outperforms silicon at high temperatures as well as high frequencies. GaAs is also a direct-bandgap semiconductor, which means it emits light more efficiently than traditional silicon. As a direct-bandgap material, GaAs is resistant to radiation.

Because of these physical features, GaAs semiconductors are found in many products that involve light and radiation, such as X-ray machines, microwaves, and LEDs.

Gallium nitride

GaN is another direct-bandgap III-V semiconductor that is considered a WBG material. GaN is a very hard, very stable semiconductor with excellent thermal and conductivity properties. In fact, GaN can conduct electrons with as much as 1,000× more efficiency than silicon.

The ability to grow GaN crystals with relative ease on various substrates has driven down the cost of this high-performance material. With increased demand for semiconductor materials that perform well in high-voltage applications like electric vehicles, GaN is seeing a surge in popularity.

MATERIAL NAME	SYMBOL	PRIMARY APPLICATIONS	MELTING POINT
Gallium arsenide	GaAs	X-ray machines, microwaves, LEDs	1,240˚C
Gallium nitride	GaN	Industrial uses, electric vehicles	2,500˚C
Indium phosphide	InP	Optoelectronics, photovoltaics	1,062˚C
Zinc sulfide	ZnS	Photocatalytic applications	1,185˚C
Zinc selenide	ZnSe	Flat-panel displays, lasers, LEDs	1,525˚C
Silicon carbide	SiC	Electric vehicles, industrial uses, UPS	2,730˚C
Silicon germanium	SiGe	Consumer electronics, automotive, telecommunications	~1,300˚C

Table 1: Common compound semiconductors

Today, applications for GaN are typically found in designs that require excellent switching speed, high breakdown strength, strong thermal conductivity performance, and lower on-resistance. For this reason, GaN is becoming increasingly common in power devices — particularly in industrial and EV applications — as well as in RF components and LEDs.

Indium phosphide

Like GaAs and GaN, InP is a III-V compound semiconductor. It shares several physical characteristics with GaAs, most notably its classification as a direct-bandgap material. Also like GaAs, it has a face-centered cubic crystal structure and is highly stable, even at frequencies measured in the hundreds of gigahertz. However, compared with traditional silicon and even GaAs, InP offers superior electron velocity, making it a star in applications that are both high-power and high-frequency. In fact, it was once used to make a bipolar transistor that could operate at 604 GHz.

While InP is not yet a commonplace material, as its benefits are fairly specialized, it is currently used in optoelectronics and photovoltaics. More interestingly, InP's properties suit it for applications in a very specific range of the electromagnetic spectrum. That is, InP unlocks a zone in the electromagnetic spectrum that resides between microwave and infrared. This terahertz region holds great promise for use in space applications.

Zinc sulfide

While ZnS is the primary form of zinc found in nature, the ZnS formulation used as a semiconductor material for electronics is its dense synthetic form. ZnS is a II-VI material, but it exhibits some similarities with the III-V materials mentioned above and is also considered a WBG semiconductor. Because it is resistant to oxidation and hydrolyzation, even when reduced to nanometer scale, ZnS is a good choice for photocatalytic applications.

It is also a low-cost compound, with high strength and a stable direct bandgap. Easily made by igniting a mixture of zinc and sulfur, ZnS offers relatively high performance for its price.

Zinc selenide

ZnSe is a WBG II-VI semiconductor and was one of the first semiconductor materials discovered. It can be made in both a hexagonal and a cubic crystal structure, is very stable, and performs well in high-temperature applications.

Applications for ZnSe include optoelectronics like flat-panel displays, lasers, and LEDs. The material emits blue light and is particularly effective in optical applications because of its resistance to thermal shock and its stability.

Silicon carbide

Among the materials profiled here, SiC has gotten the most buzz in recent years. This IV-IV semiconductor occurs naturally in an exceedingly rare mineral known as moissanite (which is found in meteorites) and first appeared in electronics applications in the early 1900s, when it was used in radios. Today, SiC's unique properties are making it a material of choice in many high-power applications.

SiC's stability and WBG composition give the material incredible electrical, thermal, and conductivity properties. New applications for power electronics have fueled demand for materials with these properties, and SiC has risen to the challenge. With its efficiency, robustness, and power properties, SiC has found use in EV drivetrains, industrial applications (such as trains, data centers, and uninterruptible power supplies), and RF applications (which benefit from its high-switching, high-frequency performance).

Silicon germanium

Another IV-IV compound is SiGe, which is valued for its favorable power consumption performance and ability to handle high frequencies. With faster electron mobility than silicon, SiGe finds common use in consumer electronics, automotive applications, and telecommunications — any design where small size and high performance are imperative.

While SiGe remains more expensive than silicon, it is lower in cost than other WBG materials, such as GaN, and it offers more flexible bandgap tuning than traditional silicon. Recent advancements in production processes have made this compound semiconductor an increasingly popular choice for manufacturers.

A GROWING NEED FOR COMPOUND SEMICONDUCTORS

Years ago, applications that required compound semiconductors were rare. Today, however, requirements for modern high-power and high-frequency devices make these WBG materials an increasingly popular choice for a variety of applications. As production processes continue to improve and the materials become cheaper and more readily available, expect to see compound semiconductors deployed in more of the most commonly used devices.

Chris McGrady is a contributing writer to EE Times with more than 10 years of experience in the electronics industry.

SiC AND GaN: AN INDUSTRY DRIVEN BY DIFFERENT ENGINES

BY EZGI DOGMUS, POSHUN CHIU, AND TAHA AYARI

Over the last several decades, developments in silicon carbide and gallium nitride technologies have led to growing industry acceptance and the promise of revenues in the billions of dollars for devices based on the two wide-bandgap (WBG) materials.

The first commercial SiC device arrived in 2001 in the form of a Schottky diode from Germany's Infineon Technologies. Rapid development followed, and the industry sector is now poised to exceed $6 billion by 2027. GaN first wowed the industry in 2010, when U.S.-based Efficient Power Conversion (EPC) delivered super-fast switching transistors. Market adoption has not matched that of SiC, but come 2027, power GaN revenues could hit more than $2 billion.

The secret to future market success for each technology rests with electric and hybrid vehicles.

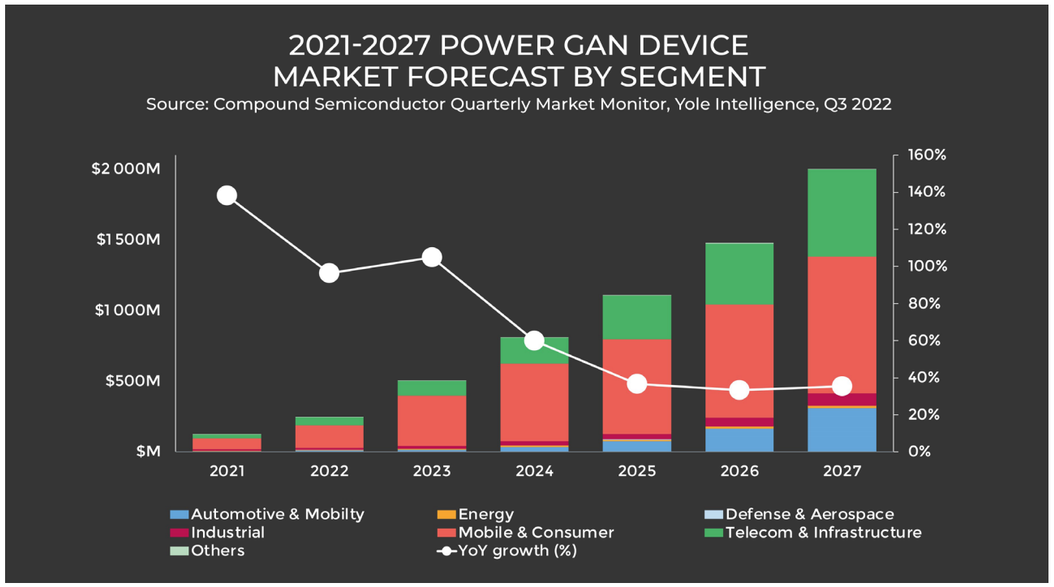
SWEET SPOT FOR SiC

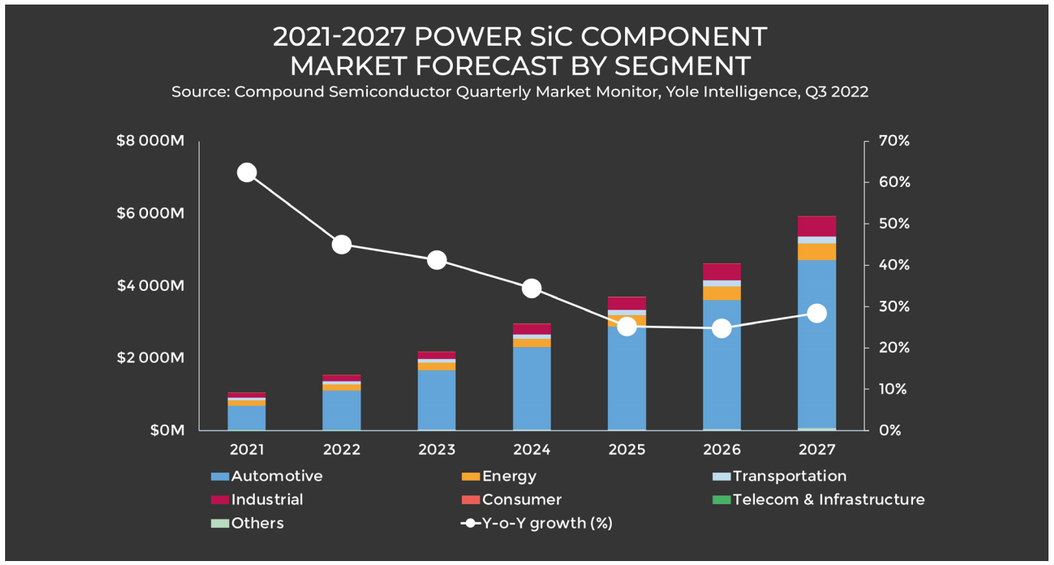
The EV/hybrid-vehicle market is truly the sweet spot for SiC power components: More than 70% of revenues, equating to $4.7 billion, are expected to come from this sector. Tesla kickstarted the SiC power device market in 2017, when it became the first automaker to use SiC MOSFETs in a car (the Model 3). Sourced from STMicroelectronics, the device was integrated with an in-house–designed main inverter. Other automakers have been quick to follow, including Hyundai, BYD, Lucid, Nio, General Motors, and Geely, among others.

Today, the transition of battery voltage from 400 V to 800 V is under way, sparking a new wave of competition among OEMs to achieve the battery-electric vehicle with best-in-class driving range and fast charging. Accordingly, a voltage rating of 1,200 V is required for SiC components. BYD's Han EV and Hyundai's Ioniq 5 are shipping in high volume today. And as every quarter delivers record sales, more high-end BEVs are expected to come to the market and drive even stronger demand for SiC in 2023.

From a supply perspective, major SiC device manufacturers — such as Wolfspeed, Infineon Technologies, STMicroelectronics, Rohm, and onsemi — provide automotive-grade discrete products or modules. The surge in demand also motivates IDMs to adjust their strategies to meet the requirements of automotive OEMs.

One key trend has been to build in-house substrate supplies for device manufacturers, vertically integrating from SiC substrate to device manufacturing in order to manage the supply of critical SiC technology. Rohm acquired SiCrystal a decade ago to integrate substrate growth capability internally, and onsemi made a similar and significant move in 2021 by acquiring GT Advanced Technologies (GTAT), an American SiC boule supplier. The leading company in SiC components, STMicroelectronics, is also expanding substrate activities after having acquired Norstel in Sweden in 2019. Infineon Technologies is sourcing SiC substrates from multiple suppliers; the German company is also developing boule and wafer-splitting technology.





For substrate players, entering device manufacturing means catching higher values and, thus, higher revenues in the future. The leading SiC substrate player, Wolfspeed, is expanding its activity. With 200-mm SiC substrates seen as a strategic resource for the next-generation platform, offering a better cost proposition than the current 150-mm substrate platform, Wolfspeed opened the first 200-mm wafer fab in New York in 2021. Very recently, the company announced the addition of a materials plant in North Carolina with an investment of $5 billion. It is believed that 200-mm SiC adoption is accelerating, and more players are announcing their plans for 200 mm in the coming years. Coherent Corp. (formerly II-VI), another leading substrate supplier, revealed its ambition to enter device manufacturing in the coming decade. Another emerging SiC supplier, SK siltron css, is also expanding its capacity significantly after the recent ribbon cuttings at its plants in South Korea and Michigan.

SiC faces competition from the Si-based IGBT, which is more mature, cheaper, and more accessible for automotive OEMs to deploy in their upcoming EVs. Nonetheless, innovations and incremental improvements at each level of SiC are being proposed to tackle issues such as cost and yield. In general, SiC is the solution for 800-V BEVs, and the scaling up of SiC also makes this technology more available for industrial applications, such as fast-charging stations and photovoltaics, to drive what is expected to be a multibillion-dollar business in the coming years.

GaN'S NEXT GROWTH WAVE

So where does this leave GaN? This WBG semiconductor has yet to witness the success of SiC in the EV sector. Thanks to its high-frequency operation and efficiency, it is already well-established in less demanding consumer applications. For high-end and high-power applications, OEMs are eyeing the technology with intense interest or have development programs underway, but there are few design wins in that market.

For the past three years, the GaN power device market has been driven by consumer fast-charger applications. In 2021, several companies doubled or even tripled their GaN device revenue, resulting in a 138% increase in total device market revenues compared with 2020. GaN power devices are increasingly being used in fast chargers for a range of mobile devices, including smartphones. Indeed, U.S.-Ireland-based Navitas Semiconductor, the United States' Power Integrations, and China's Innoscience all manufacture GaN power products for the fast-charger market.

Given this activity, GaN power device revenues are estimated to reach about $225 million in 2022. But as GaN device suppliers look to enter other markets to raise volumes, this figure is expected to swell to $2 billion by 2027. The second wave of GaN growth is expected to be driven by the datacom/telecom and EV/hybrid-vehicle markets.

New power consumption efficiency regulations in the data center business are expected to be imposed in the coming two to three years to meet CO_2 emission objectives, and that activity would favor the adoption of GaN-based 80 Plus Platinum-grade power supplies. Several players are focusing on this opportunity: Infineon Technologies with its CoolGaN technology, GaN Systems with its GaNPX products, and Transphorm. Power supply providers such as Delta Electronics, Compuware, and Solum are already involved.

It is still early days for GaN in EVs. Many power GaN players, such as EPC, GaN Systems, and Transphorm, have developed automotive-qualified GaN devices for on-board chargers, DC/DC conversion, and even traction inverters in EVs/hybrid vehicles, and myriad partnerships have been formed with automotive businesses.

For example, VisIC supplies its devices to automotive supplier Marelli for its 800-V multilevel traction inverter and has also partnered with Hofer for the same application. Transphorm continues to work with Marelli to provide devices for on-board charging and DC/DC conversion. EPC, currently delivering automotive-qualified low-voltage GaN devices, is working with French-based BrightLoop to develop affordable power supply converters for off-highway and commercial vehicles.

On the ecosystem side, a significant development comes from Navitas, which became a publicly traded company with a market value of $1.04 billion by combining with special-purpose acquisition company Live Oak Acquisition. The company recently announced two acquisitions: VDDTech, specialized in advanced digital isolators for next-generation power conversion, and GeneSiC, an important provider of SiC devices. Navitas no longer defines itself as a pure GaN IC player but as a new-generation power electronics company.

More generally, Yole Intelligence has witnessed the entry of new players to the supply chain. Notably, Rohm is offering a 150-V GaN product for telecom/datacom applications. BelGaN, a new GaN foundry based in Belgium, has recently acquired onsemi's fab. Focusing on the Chinese ecosystem, multibillion-dollar investments have been allocated in recent years. A domestic Chinese supply chain for GaN power is well-developed, especially for the consumer market. It is noteworthy that Innoscience, a privately owned company, has been investing more than $1 billion to strengthen its position as a worldwide GaN IDM and to expand its current capacity of 10k 8-inch wafers per month to 70k wafers per month by 2025.

So what's next for both SiC and GaN? As manufacturers of power SiC devices anticipate a multibillion-dollar market, will GaN experience the same success? Widespread OEM adoption of GaN in telecom/datacom and EV applications would radically impact market forecasts.



Ezgi Dogmus is team lead analyst and **Poshun Chiu** and **Taha Ayari** are technology and market analysts, all with the Compound Semiconductor & Emerging Substrates team at Yole Intelligence, part of Yole Group.

GaN'S EVOLUTION FROM SCIENCE PROJECT INTO **MAINSTREAM POWER SEMICONDUCTOR**

BY ALEX LIDOW

Forty-five years ago, power MOSFETs were first launched commercially as EE Times celebrated its fifth anniversary. Now, on the 50th anniversary of this industry journal, power-conversion technologies are experiencing the first tectonic shift since the move from bipolar to MOS. That shift, of course, is due to the viral adoption of wide-bandgap power devices. This article will focus on gallium nitride, as there are others taking up the torch for the complementary technology of silicon carbide.

WHERE IS GaN TODAY?

GaN hit commercial shelves in March 2010, and the first major application, LiDAR, adopted enhancement-mode GaN transistors from Efficient Power Conversion (EPC) 16 months later. LiDAR was the perfect "killer app" in that it used the high speed and tiny size of chip-scale GaN transistors at their maximum performance to achieve 3D digital mapping with better resolution and higher speed than radar. Thus was born the essential sensor for autonomous motion.

LiDAR was followed by 48-V DC/DC converters for artificial intelligence and bitcoin-mining systems. The trend toward 48-V distribution buses in advanced computers and servers has accelerated ever since. Around 2013, the designers of satellite systems became aware of the excellent properties of GaN transistors in environments that involve multiple forms of radiation. After a few years of extensive reliability and radiation testing, GaN devices left Earth for the first time. Since then, hundreds of thousands of GaN transistors and hybrid modules have found homes anywhere from low Earth orbit to the more stringent geosynchronous Earth orbit. Aging rad-hard MOSFETs are no match for the faster, smaller, less expensive, and very rad-hard GaN devices in applications such as DC/DC converters, reaction wheel motor drives, LiDAR systems, and ion thrusters.

Higher-voltage (650-V) GaN transistors and ICs from companies such as Navitas Semiconductor, Power Integrations, and GaN Systems pried open the first high-volume consumer product, the cellphone fast charger, beginning in 2018. This expanded the awareness of GaN's potential from power-system design engineers to the wide world of cellphone consumers. E-bikes, drones, and robots soon adopted GaN transistors to reduce weight, size, cost, and EMI. Automotive applications such as 48-V DC/DC converters, headlamps, cabin fans, seat heaters, and on-board chargers are coming soon. At this point, GaN is more than a specialty technology (see Figure 1); it is a broad-scale replacement for silicon MOSFETs in applications ranging from 30 V up to 650 V — a multibillion-dollar market.

WHY GaN IS A MARKET THAT FAVORS STARTUPS

The key players in GaN all started their existence in this millennium as pure-play GaN companies. The early entrants were EPC and Transphorm. GaN Systems and Navitas followed a few years later, then Innoscience and VisIC. Most recently, NXP Semiconductors and Cambridge Electronics (CEI) joined the club.

Along the way, established semiconductor companies such as Infineon Technologies, STMicroelectronics, Texas Instruments, and Panasonic apathetically tested the market with sample volumes and mixed messaging.

GaN startups are largely in the "wafer-fab–recycling business." The equipment needed to make GaN transistors and integrated circuits that far outperform their silicon counterparts is trailing-edge technology: 150-mm (now going to 200-mm) wafer foundries with 0.5-µ technology found life-extending business in the fabrication of GaN devices. This low capital cost is rare in the semiconductor industry and strips large companies from their advantages of scale. In addition, GaN technology is fast evolving, thus favoring fast-acting companies that have short cycles of learning. The fastest companies mature the technology the fastest while developing fundamental intellectual

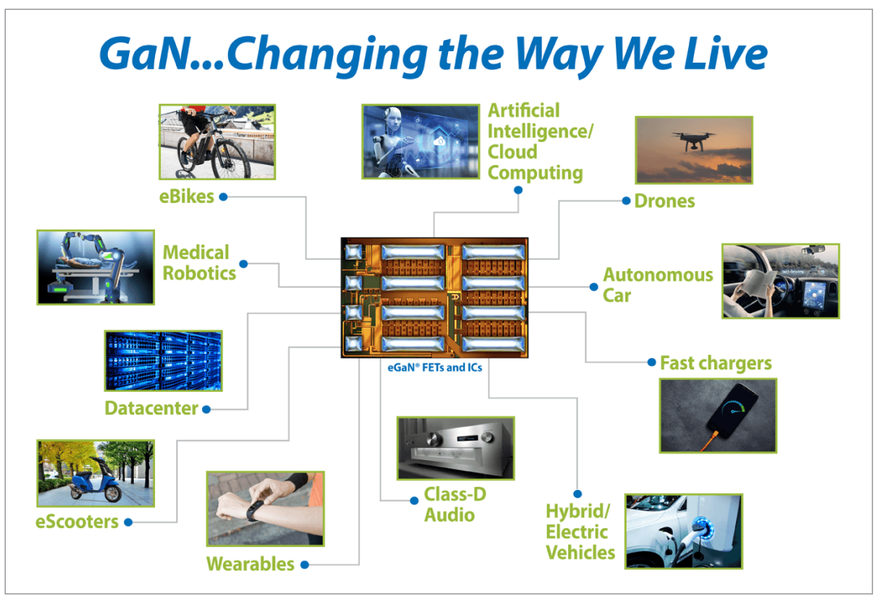


Figure 1: Applications for GaN FETs and ICs today

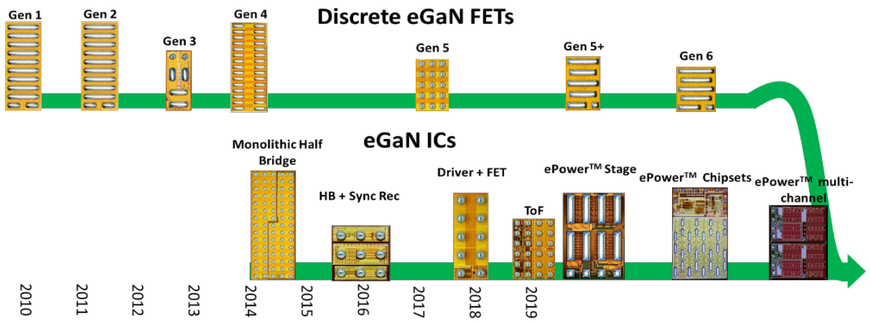


Figure 2: The evolution of GaN from discrete to IC to full system-on-chip

property and establishing de facto standards in specification, packaging, and performance. Once the technology matures, big players will once again have the advantage. If history is any indication of the future, however, those big players will not be the established silicon goliaths of today's markets.

WHERE IS GaN GOING?

GaN is a better power semiconductor than silicon in the range of 30 V to 650 V. Today's best-performing GaN is still 300× larger in size than physics says it can eventually achieve with comparable electrical properties. Silicon, on the other hand, is at the physical limit, so future improvements require disproportionate increases in capital investment. GaN, however, will remain in the fab-recycling business for many years to come, as improvements in the technology are more dependent upon device structure and material improvements and less dependent upon wafer diameter or feature size.

There is another advantage to GaN technology that cannot be matched by silicon: the ability to integrate power devices with signal and digital devices. GaN integrated circuits first started showing up in September 2014, when EPC launched the first family of monolithic half-bridge ICs. Since then, Navitas has entered the market with ICs designed to reduce cost and decrease the size of fast chargers for phones and tablets, while EPC has expanded the family to include monolithic power stages for DC/DC converters and motor drives, as well as hyper-fast ICs for LiDAR applications (see Figure 2).

CONCLUSION

When we celebrate EE Times' 75th anniversary, there is little doubt that GaN power devices will have displaced power MOSFETs (and much of the ecosystem that supports power MOSFETs), much like power MOSFETs displaced bipolar transistors starting almost 50 years ago. There is a vitality in the power-conversion industry that has not been felt for many years. That vitality is coming from the new generations of innovative products made with wide-bandgap semiconductors that create performance, size, and cost possibilities that have never been possible with silicon.

Alex Lidow is CEO of Efficient Power Conversion (EPC).

EE Times 50

EFFICIENT SiC POWER DEVICES **BUILD ON SILICON'S LEGACY**

BY VICTOR VELIADIS

2022 marks EE Times' 50th anniversary and the 75th anniversary of the discovery of the transistor. This monumental invention by Shockley, Bardeen, and Brattain at Bell Labs in 1947 set a course of massive semiconductor electronics development that has touched every aspect of our lives. I remember, as a young engineer, being inspired seeing that first transistor on a wall display as I entered the Bell Labs building in Murray Hill, New Jersey, for work.

Early transistors were made with germanium, but silicon became the preferred semiconductor material around 1960 because its larger bandgap enabled lower leakage current. For high-power applications that condition high voltages and currents, silicon power MOSFETs and IGBTs were developed, optimized, and inserted into a plethora of systems. Indeed, silicon power devices today are entrenched in power electronics, thanks to their low-cost volume production, excellent starting material quality, ease of fabrication, and proven reliability and ruggedness.

Nonetheless, although Si power devices enjoy a device/circuit design legacy and a large-volume, streamlined fabrication infrastructure, they are approaching the operational limits imposed by silicon's fundamental material properties. Silicon's relatively low bandgap and critical electric field result in high conduction and switching losses and compromise high-temperature performance. Thus, other promising materials systems have been explored for efficient high-power applications.

Silicon carbide is an IV–IV wide-bandgap compound material with strong chemical bonding between the Si and C atoms, allowing for high hardness, chemical inertness, and high thermal conductivity. Wide n- and p-type implantation doping ranges and a relatively thin thermal native SiO_2 are possible in SiC. With respect to silicon, SiC exhibits a 3× higher wide bandgap, a 7× to 9× higher critical electric field strength, and a 2.5× higher thermal conductivity. These favorable material properties, which allow for highly efficient power devices with reduced form factors and simplified cooling, represent a quantum leap in performance evolution over Si. This was recognized early on and led to significant

investments in SiC that culminated with the introduction of the first commercial SiC Schottky barrier diode in 2001, the first commercial SiC MOSFET in 2010, and the commercial insertion of SiC MOSFETs in electric vehicles starting in 2018.

Given SiC's compelling benefits for high-power applications, sustained efforts have been directed toward developing SiC material and device technology since the 1980s. Important prerequisites for the commercialization of power devices manufactured on a novel semiconductor material include the availability of large-diameter, low-defect substrates and high-quality epitaxial-layer growth; streamlined fabrication that exploits the existing, expansive Si fab infrastructure; and the demonstration of superior device performance, with reliability and ruggedness rivaling those of incumbent semiconductors.

Relatively pure SiC crystals were grown by the Lely sublimation technique in 1955. The crystals were mostly 6H-SiC but inadvertently included other polytypes. Tairov and Tsvetkov invented a reproducible method for SiC boule growth in 1978; the approach involved insertion of a 6H-SiC seed crystal into a sublimation growth furnace and controlled mass transport from the SiC source onto the seed crystal. This growth process, known as the modified Lely method or seeded sublimation method, was improved upon and further developed by several groups to obtain SiC boules with large diameters and reduced defect densities. Today, 150-mm SiC substrates are primarily used in production; 200 mm is also utilized, but at a smaller scale. Killer defect densities have been drastically reduced by aggressively addressing basal plane dislocations, the last remaining major catastrophic defect.

In 1987, Kuroda et al. demonstrated high-quality SiC CVD homoepitaxial growth on off-axis substrates at relatively low growth temperatures of 1,500°C to 1,650°C. This is the standard epitaxial technique on 4H-SiC substrates, which are preferable for power devices because of their higher electron mobility in the vertical direction, lower intrinsic carrier concentration, and shallower dopant ionization energies. Horizontal hot-wall (Kordina et al., 1993) and planetary warm-wall reactors are proven platforms for highly successful SiC epitaxial production. SiC wafers have a silicon and a carbon face. Devices are typically made on the silicon face because of the better gate oxide quality. Currently, the SiC wafer represents 45% to 65% of the overall SiC device cost, a consequence of its unique complex fabrication specifics. As a result, several disruptive SiC wafer technologies are being explored, including engineered substrates and more efficient boule utilization.

Leveraging the expansive Si fab infrastructure is paramount for achieving cost-effective SiC manufacturing.

Numerous mature processes from silicon technology have been successfully transferred to SiC. However, SiC material properties necessitate optimization of specific processes, including wafer thinning, dry etching, heated implantation and anneal, low-resistivity ohmic contact formation, high-quality gate oxide interfaces, metrology and inspection of transparent wafers, and handling of wafers with relative lack of flatness. SiC device manufacturers have developed IP for several high-yielding fabrication processes and, unlike in silicon, compete on both design and processing. Today, SiC manufacturing is mature, and its fab infrastructure mirrors that of Si. Integrated SiC device manufacturers coexist with foundries and fabless companies, and design houses provide know-how and IP that can be licensed to accelerate entry to market.

From a historical perspective, 1-kV p-n and Schottky barrier diodes were reported in 1991 and 1993 by Matus et al. and Urushidani et al., respectively. With respect to transistors, the first vertical-trench MOSFET was demonstrated in 1993 by Palmour et al. and the first planar double-implanted power MOSFET in 1997 by Professor Cooper's group at Purdue University.

My SiC journey started in the early 2000s, when I was fabricating SiC JFETs for RF radar power amplifiers. Their device area was 1.2×10^{-3} cm^2, and the yield reached 92% inclusive of JFETs situated at the very edge of the R&D wafers (Figure 1). The small RF device area alleviated the detrimental impact of defects on yield. Wafers at the time were 3 inches in diameter, and my design consisted of 134 reticles. Fabricating large, 0.27-cm^2 devices for power applications would require the area of a full reticle. Of the 134 reticles shown in the wafer map of Figure 1, only the five outlined in yellow have no failed devices, rendering the 0.27-cm^2 device yield at 3.7%. It remains the case today that the presence of material and processing defects has a disproportionally catastrophic impact on yields as the area of the SiC device increases.

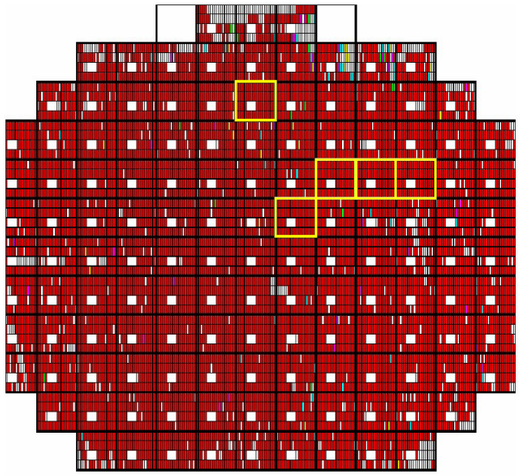


Figure 1: Wafer map of JFET gate-to-source breakdown voltage showing a yield of 92%, which includes the contribution of JFETs located at the very edges of the wafer. Good devices appear in red; failed devices appear in gray. The large, uniformly situated white squares in each reticle contain experimental devices and characterization structures that are not measured in the automated probe card testing. Only five reticles, denoted with yellow outlines, contain no failed devices.

As early as 2006, I fabricated 1,200-V 0.19-cm^2 JFETs capable of 54-A current output at a forward voltage drop of 2 V and a gate bias of 2.5 V (Figure 2). This motivated me to continue work toward SiC commercialization. In fact, my

0.19-cm² JFET received a "Device New Records" mention at the International Conference on SiC and Related Materials 2007 in Otsu, Japan.

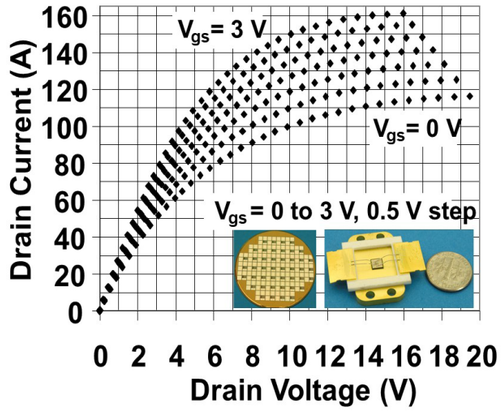


Figure 2: Representative forward-current characteristics of a 0.19-cm² 4H-SiC JFET fabricated in 2006 on a 3-inch wafer. The device outputs a current of 54 A at a forward-voltage drop of 2 V and a gate bias of 2.5 V. It is capable of blocking 1,270 V.

When a promising technology is introduced, a suitable high-volume application is needed to achieve cost reductions that stem from economies of scale. For SiC, hybrid and electric vehicles constitute the mass-commercialization application that is driving the explosive SiC growth we are now experiencing. Per Yole Group's projections, power SiC devices will be a $6.3 billion market by 2027, with automotive accounting for $5 billion of that. Overall, SiC is expected to capture more than 20% of the power device market by 2027.

In the next five years, complex and labor-/time-intensive SiC substrate growth will continue to drive vertical integration, "wafering" innovations that yield more substrates per boule, and engineered substrate advances. Wafers at 150 mm will represent over 50% of market share, as 200-mm wafers approach cost and defect-density parity per square centimeter. Non-CMOS-compatible processes like ohmic contact formation, heated implantation and anneal, substrate thinning, and metallization will be close to standardization. Wafer planarity will improve, facilitating fabrication. Gate oxide optimizations will increase mobility and reduce threshold voltage instability. Because of its higher level of complexity, the trench MOSFET configuration, more meaningful for devices rated below 1,700 V, will not be the choice of companies entering SiC manufacturing. As processing IP complicates SiC foundry fabrication, IDMs will continue to dominate production. Manufacturing SiC in mature, fully depreciated Si fabs has emerged as a cost-effective model and will persist moving forward. Defect density reductions will allow for larger SiC device current ratings closer to those of Si. The price of 650- to 1,700-V SiC MOSFETs will be 1.5× to 2× higher than that of silicon options. However, system-level cost savings in major applications like automotive and photovoltaics will outweigh the higher cost of procuring SiC devices.

Victor Veliadis is executive director and CTO at Power America and a professor of electrical engineering at North Carolina State University.

EE Times 50

SILICON PHOTONICS: BREAKING THE INTERCONNECT

BY SALLY WARD-FOXTON

Different forms of optical connection are boosting bandwidth between today's big chips.

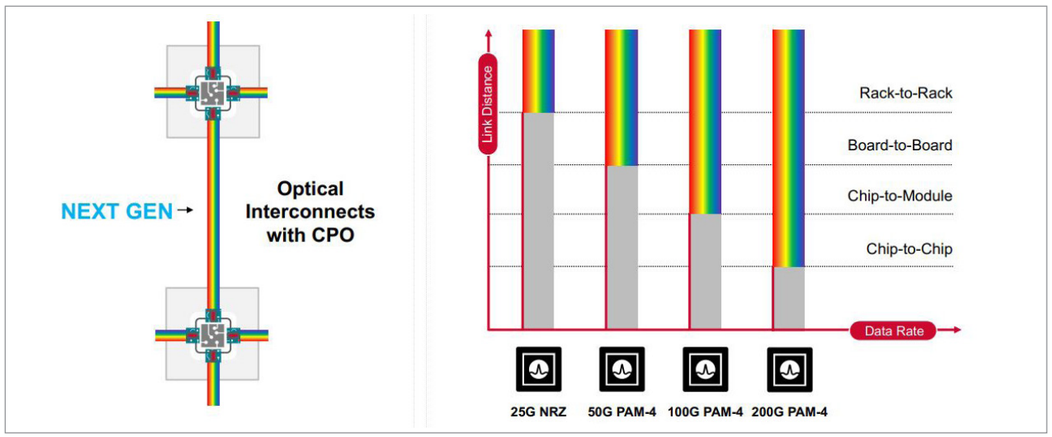
Data center workloads are rapidly growing in scale and demand unprecedented compute and latency performance. Moving data around quickly and efficiently is therefore a critical piece of the puzzle for data center infrastructure. Data transfer down copper wires is simply not able to keep up; copper is fast becoming the bottleneck for data communications.

The data center has long used optical fibers to communicate at very high speeds between systems. Now, advances in silicon photonics that allow optical components and waveguides to be built from silicon and integrated into chips (or chiplets) are helping to bring optical communication to shorter links in various forms. Silicon photonics promises to bring optics' bandwidth, consumption, and latency to more parts of the data center.

There are several slightly different architectural ideas for silicon photonics in the data center. Here's a closer look at the approaches and some of the companies pursuing them.

CO-PACKAGED OPTICS

The idea behind co-packaged optics (CPO) is to replace electrical SerDes links



As the optical portion of the interconnect gets closer to the chip, data rates rapidly improve. (Source: Broadcom)

between pluggable optics modules and switch ASICs with optical links.

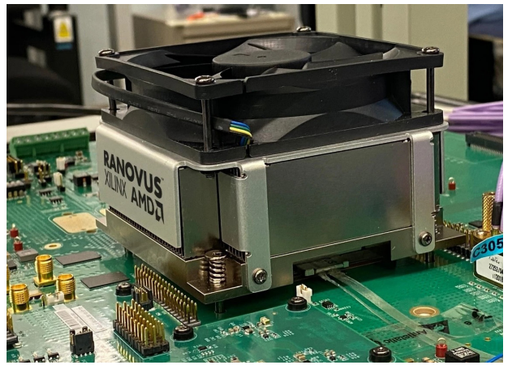
Pluggable optics are well-used, standardized components that sit at the edge of a board and effectively convert electrical signals from chips to the optical domain for transport off the board using optical fiber. Using this system, the link between the chip and the edge of the board uses electrical signals down copper wires.

Co-packaged optics would replace familiar pluggable optics modules with optical links that go directly to the switch ASIC or processor, integrating the laser source and an I/O chiplet within the same package as the switch ASIC or processor. Moving these functions close to the chip in this fashion can boost bandwidth and improve energy efficiency.

One company working on co-packaged optics is Ranovus. The company showed its CPO chiplet, **Odin, working with an AMD/Xilinx data center FPGA** earlier this year. This solution co-packages the optical I/O chiplet and lasers with the FPGA silicon. Ranovus is targeting switch ASICs as well as big data center processors (such as those used for AI acceleration), but it also plans to target disaggregated server architectures with pooled memory.

Because Ranovus can embed lasers directly into the silicon, the Odin photonics chiplet scales from 800 Gbps to 3.2 Tbps in the same footprint. The Ranovus laser is a quantum-dot multi-wavelength laser that can generate up to 96 wavelengths simultaneously. This means interface design can be simplified compared with approaches using multiple discrete lasers. The module draws roughly 4 W, replacing electronics that require roughly 14–17 W.

Ranovus also offers a chiplet version without lasers, responding to market calls for plug-and-play components that limit customers' dependence on any particular vendor.



Ranovus's optics co-packaged with an AMD/Xilinx Versal ACAP FPGA die (Source: Ranovus)

LASERS OUT

Broadcom offers Humboldt, a 25.6-Tbps CPO switch that co-packages its Tomahawk 4 switch ASIC with four 3.2-Tbps silicon photonic chiplets. Broadcom says this CPO solution saves more than half the power consumed by pluggable optics.

While co-packaged optics can work with lasers either in the package or external to it, Broadcom is firmly in the external-laser camp. The company has settled on a design that uses external lasers plugged into the front panel of the server. This means the laser can run at higher efficiency and stay cooler. Retaining pluggable lasers also

Broadcom's Humboldt CPO chiplets co-packaged with its Tomahawk 4 switch ASIC can replace traditional switches, saving space and improving power consumption and bandwidth. (Source: Broadcom)

means they are easily replaceable once deployed (so redundancy isn't required), and the customer isn't tied into any particular CPO design. While integrated solutions require fewer optical connectors (and therefore reduce loss) and enable wafer-scale processing and test, they actually require more power than the pluggable laser, according to Broadcom. They are also sensitive to temperature and need to be thermally isolated from the ASIC.

Broadcom's Humboldt switch is set to be installed in Chinese hyperscaler Tencent's data centers, per a new system architecture defined by Tencent.

MULTISOURCE

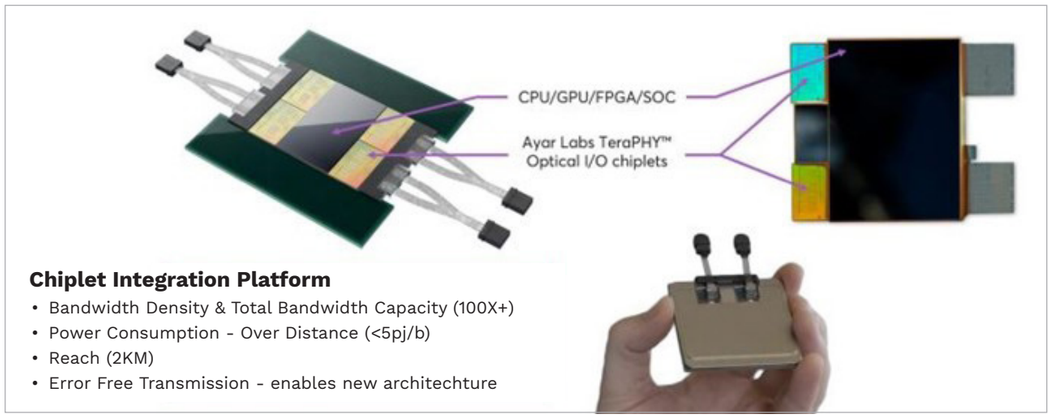
Also in the external-laser camp is **Ayar Labs**, which has developed an optical chip-to-chip communications chiplet called TeraPHY. The company supplies the chiplet in known-good die form to compute-chiplet companies for co-packaging with large ASIC dies bound for the data center.

"Generation after generation, we're still scaling compute per reticle," Ayar Labs CTO and co-founder Mark Wade told EE Times. "But at the same time, I/O bandwidths were very much struggling to keep pace, even in the 2010s … How will electrical I/O technologies fare in attempting to escape that bandwidth limit?"

Electrical links are starting to require retimers, which means increased cost and complexity. Optics can solve this problem; TeraPHY chiplets offer 1,000× the bandwidth density improvements at one-tenth of the power compared with electrical I/O. The result is ASICs that can communicate with each other over distances from a few millimeters up to 2 km.

Ayar's decision not to include a laser in the same package was made right at the start. "The physics of lasers are disconnected from the physics of CMOS microelectronics; they don't like to operate at high temperatures," said Wade. "They rapidly lose power efficiency, and their reliability gets exponentially worse."

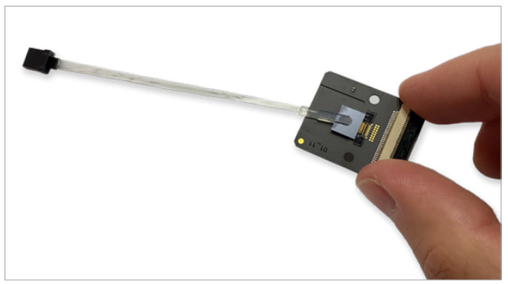
High thermal design power compute nodes use hundreds of watts; the temperature inside an SoC package might be 80°C or above. Disaggregating



Ayar Labs' optical I/O chiplets pictured alongside a host SoC and showing the final packaged module (Source: Ayar Labs)

the light source from the SoC allows it to be placed farther away, keeping its temperature below 55°C.

Ayar has designed its own laser light source, called SuperNova, which provides up to 16 wavelengths of light — enough to power 256 channels of data, or 8.192 Tbps. The disaggregated laser offers platform flexibility and field replaceability. While Ayar plans to support multiple TeraPHY chiplets with a single laser source to save costs, there are tradeoffs, including having to rely on just one source without redundancy.



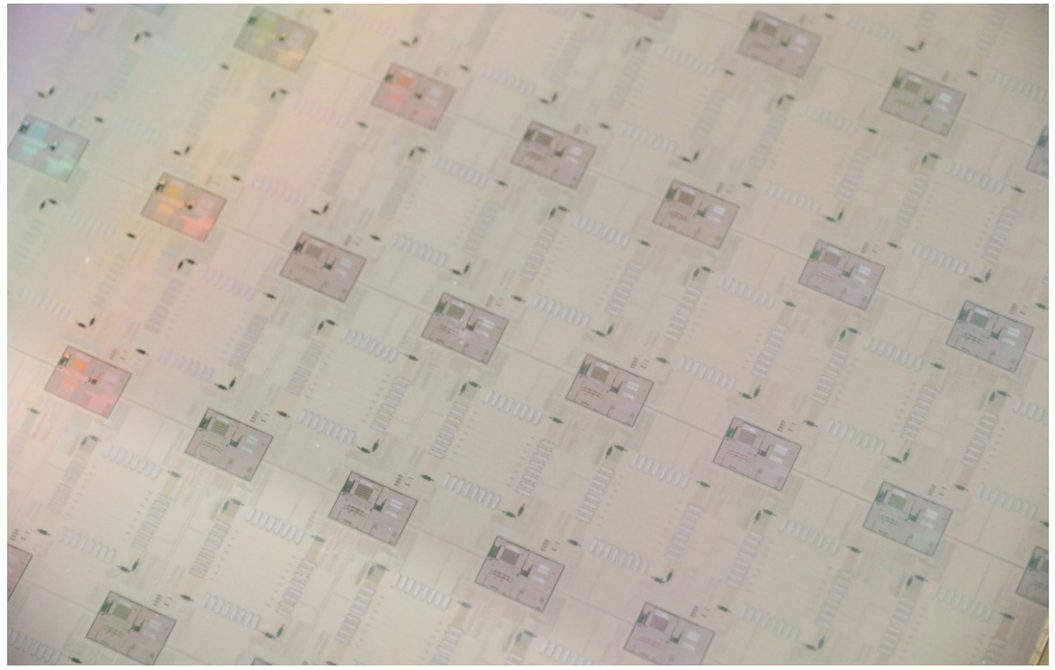
Ayar Labs' SuperNova laser source can produce 16 wavelengths, enough for 256 channels. (Source: Ayar Labs)

SuperNova is compatible with the continuous-wave wavelength-division–multiplexing multisource agreement among laser makers.

WAFER-SCALE

Lightmatter is another silicon photonics company hoping to introduce optical chip-to-chip interconnect to the data center. The company has developed a **wafer-scale device, Passage**, which is effectively a programmable photonic interconnect through which ASICs or high-bandwidth–memory chiplets packaged on top can communicate. Up to 48 processor dies can be mounted on the wafer using a standard chip-on-wafer process, and they don't all have to be the same; for example, a wafer might hold a mixture of CPUs and accelerators in a supercomputer.

"Imagine you have a data center and you want to do AI training," Lightmatter CEO Nick Harris told EE Times. "We could take arrays of, for example, Google's TPUs, or Nvidia's GPUs, or Lightmatter's processors,



Lightmatter's Passage device uses a wafer-scale die for optical interconnect between compute/memory dies mounted using a chip-on-wafer process. (Source: Lightmatter)

and put them on top of Passage, increase the interconnect bandwidth by a factor of 100, and reduce the interconnect energy costs by 10×. You'd start to see racks get collapsed into Passage platforms."

Using Passage, each die gets 96 Tbps of bandwidth at a maximum latency of 2 ns between chips.

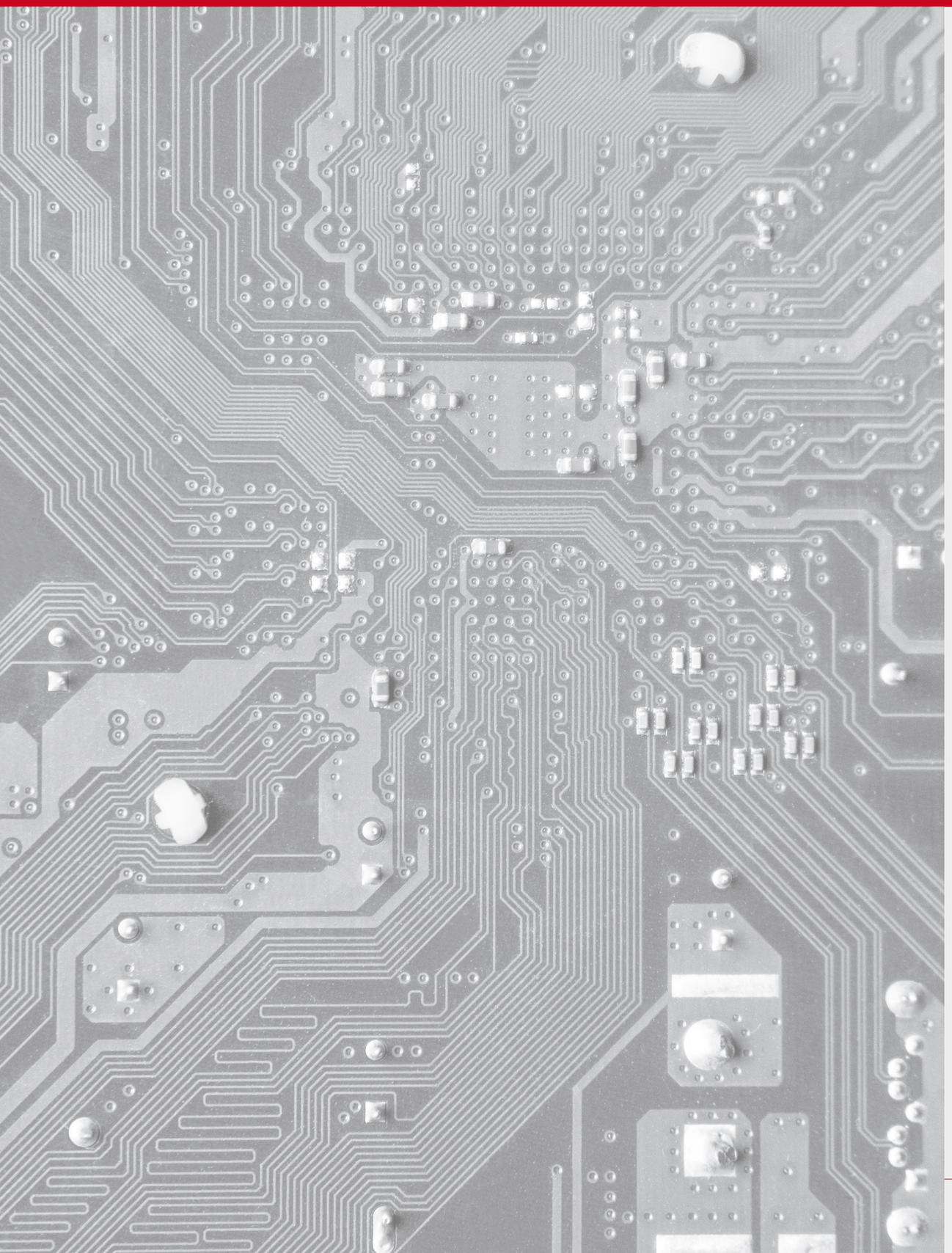
Passage is a combination of switch and interconnect that is entirely reconfigurable in 1 ms, and reconfiguration can be done remotely — no need to plug and unplug fibers. The idea is to enable a reconfigurable supercomputer while avoiding the yield issues that come with attaching optical fibers to chips. The Passage wafer includes lasers, optical modulators, photodetectors, and transistors.

STILL PLUGGING AWAY

While these exciting technologies promise a future of high-bandwidth, low-power data transfer in the data center, co-packaged optics are not expected to replace plug-gable optics completely in the next 10 years, **according to analyst Yole Intelligence**. Pluggables have the advantage of being a well-established, mature technology with multiple vendors in the market.

While CPO's technology story is clear, adoption also depends on the right combination of manufacturing yields, reliability, multivendor business models, and cost.

Sally Ward-Foxton is a correspondent for EE Times.



EE Times 50

PARTNER CONTENT

ROCHESTER KEEPS BUSINESSES MOVING, **THEN AND NOW**

BY TRACEY CORBITT, *DIRECTOR OF GLOBAL MARKETING, ROCHESTER*

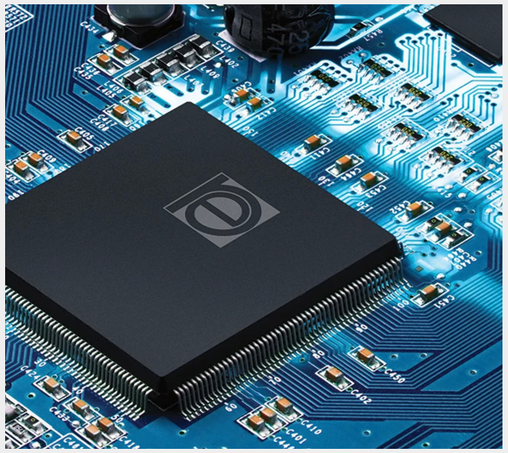
Supporting customers and suppliers for 40 years and counting.

The evolution of Rochester Electronics tells a story of a company that has advanced to support the changing needs of its customers as well as the industry's top semiconductor suppliers. Rochester's 40-year history is rich with innovations to support obsolete semiconductors.

Curt Gerrish founded Rochester Electronics in 1981, when the pace of technological advancements within the semiconductor industry had begun to result in a significant increase in end-of-life (EOL) components. Customers with long product life cycles and a need for ongoing "spares and repairs" were increasingly met with the market obsolescence of components critical to their applications. What were initially thought to be inexpensive component replacements often resulted in costly redesigns and requalifications. With more than 20 years of experience working at Motorola Semiconductor, Curt identified this need and established a vision that became Rochester Electronics. Rochester provides customers with an ongoing source of supplier-authorized components following EOL, ensuring a traceable, certified, and guaranteed supply to support customers in high-reliability and industrial sectors with longer product life cycles.

The innovative model Curt envisioned ensured that both suppliers and their valued customers would receive the same exemplary level of product-life–cycle service required to keep their businesses moving.

Through the years, Rochester has continued to increase supplier authorizations while expanding its



product offerings to include both active and EOL semiconductors. Today, Rochester has more than 70 supplier authorizations, which include acquisitions. Rochester stocks over 15 billion devices, encompassing more than 200,000 part numbers warehoused in the United States at its Newburyport, Massachusetts, and Portsmouth, New Hampshire, locations.

Supply issues in the past two-plus years have undermined the certainties of normal delivery. Covid-related manufacturing, shipping disruptions, and even natural disasters have led to supply chain uncertainty and lengthening lead times. Component discontinuation notices have risen by 15% over the same period, as third-party fab priorities have changed and as the industry refocuses its fab investments to address a lower-powered, battery-dominated landscape.

Lean supply chains have met a demand upswing and supply constraints, and all market sectors are affected.

The customer is under pressure to guarantee supply, and typically, gray-market or unauthorized sources are seen as the only solution. The counterfeit business is huge and sells through these gray-market channels to infiltrate the end customers. When times are tight, and products are not available, the risk increases significantly that an end customer will become a victim of a counterfeit product. There are testing and checks that can be done to ensure that products are genuine, but those take time and are costly to perform, and in some cases, the results are still not fully guaranteed.

> *The only way to ensure genuine products is to buy from an authorized source so that you are guaranteed the pedigree of the devices.*

Authorized suppliers like Rochester Electronics offer risk-free sourcing and are the only truly safe option for keeping customers' production lines operational during shortages, allocation, and obsolescence.

Buying from an authorized source that partners with the original component manufacturer (OCM) as a fully authorized distributor, Rochester Electronics identifies itself as compliant with the SAE Aerospace Standard, AS6496. Rochester is authorized by the OCM to provide traceable and guaranteed products with no quality or reliability testing required because the parts are sourced from the OCM.

Those providers who are not fully authorized may market themselves as AS6171/4-compliant, indicating that they follow standardized inspections and test procedures with minimum training and certification requirements to detect suspicious or counterfeit components. While better than no compliance at all, this is an indication that the parts are not sourced to the supplier from the OCM but have passed testing merely to minimize, not eliminate, risk.

EE Times 50 1972 2022

The only way to eliminate these risks is to purchase from a fully authorized source that is AS6496-compliant.

Moving beyond in-stock inventory, the Rochester vision expanded to include licensed manufacturing, which began with the transfer of an Intel military product line in 1995. This led to an enhanced customer quality and reliability program, resulting in Rochester's achieving ISO-9001 certification in 1998.

Fast-forwarding to today, Rochester has manufactured more than 20,000 device types. With more than 12 billion dies in stock, Rochester has become the world's largest die bank and has the capability to manufacture over 70,000 device types. Rochester's build-to-order devices use information and technology transferred directly from the OCM. All products are sold with the OCM's full approval under the original manufacturer's part number.

Also available at its Newburyport location is Rochester's Manufacturing Services business. The portfolio includes a hermetic assembly line fully developed with QML certification to MIL-PRF-38535, and the manufacturing services operation has achieved plastic assembly qualification. Rochester's in-house testing services, DSCC QML–approved to MIL-STD-883, received QML Space Level Certification in 2007. Providing high-quality testing solutions for both suppliers and customers alike is Rochester's reliability testing lab, offering QML certification, archive, and analytical services.

Supporting customers who require design and authorized product replication, Rochester established its first U.S. Design and Technology office, located in Rockville, Maryland, in 2008. This team can replicate an original device with full support from the original device manufacturer. The result is a form, fit, and functional device replacement

guaranteed to the original datasheet performance. No software changes required. A second team was added in 2016, located in Burnsville, Minnesota, and specializing in analog design capabilities. Examples of completed product replications include the Intel 80C196 processor, the ADI ASDP-2101 digital signal processor, and the NXP 80C592 microcontroller.

At the end of 2021, Rochester proudly announced receiving a letter of conformance confirming that its quality management system complies with the requirements of the IATF 16949:2016 standard for the design and manufacture of semiconductor components. Obtaining this letter demonstrates the commitment to providing customers with the highest standards of products and services within the automotive industry.

While developing and expanding its product-focused service solutions, Rochester is also increasing warehouse capacity, bettering automating processes' efficiency, and further expanding global sales support, both physically and digitally.

Rochester's rapidly expanding sales office footprint includes its global sales headquarters, centrally located on the Newburyport campus, along with APAC headquarters in Singapore; Japan headquarters in Tokyo; and EMEA headquarters in St. Neots, U.K. Branch sales offices are in Arizona; St. Petersburg, Florida; Guadalajara, Mexico; Beijing; Shanghai; Chengdu, China; Shenzhen, China; Osaka, Japan; Ramonville-Saint-Agne, France; Warsaw, Poland; Kassel, Germany; and Munich.

On the topic of digital sales and service, 2021 marked the five-year anniversary of Rochester's online program, which has evolved well beyond its initial unveiling. Today, rocelec.com content is available in 11 different languages, serving customers around the world.

2022 is the year the Rochester team transforms this platform into a full-service, easy-to-use online customer portal, meeting the needs of an audience beyond its traditional e-commerce buyers. Plans include diversifying payment options and enhancing transaction and fulfillment services such as order status, order history, product list management, online quoting, inventory notifications, and much more.

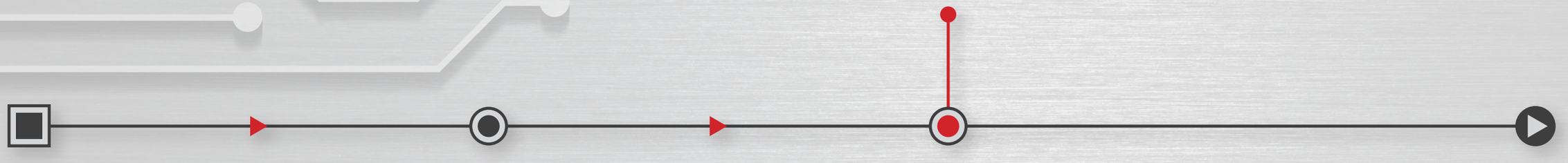
In founding Rochester Electronics as the world's first authorized distributor of EOL components, Curt Gerrish is respected as an industry pioneer. The industry landscape has changed in 40 years, but Rochester's determination to make customers glad they called has not. This principle is engrained in everything Rochester does.

Rochester is focused on providing authorized solutions to meet the ever-growing needs of the semiconductor industry anytime, around the globe, now and in the future.

Rochester is here to keep your business moving. ■

CHAPTER THREE:

THE NEXT SILICON FRONTIER



GALLIUM OXIDE: A NEXT-GEN SEMICONDUCTOR FOR POWER DEVICES

BY MAURIZIO DI PAOLO EMILIO

In the past decade, gallium oxide has seen fast technical development, propelling it to the forefront of ultra-wide–bandgap semiconductor technologies. The major targeted application space is power electronics, in which gallium oxide's intrinsic material properties — high critical field strength, widely tunable conductivity, low mobility, and melt-based bulk growth — promise to deliver the required high performance at low cost.

To maximize the potential of a novel semiconductor technology, the industry must make a concerted effort to solve technical obstacles that hinder performance. Significant technological advancement has taken place in the field of ultra-wide–bandgap semiconductors since 2016, when Flosfia, a spinoff from Kyoto University specialized in R&D and commercialization of gallium oxide thin films, concluded that gallium oxide warranted development.

The semiconductor industry is increasingly moving toward implementing devices built from wide-bandgap materials such as silicon carbide and gallium nitride, but the cost of those materials remains relatively high. In response, researchers more recently have pursued development of beta-gallium oxide (β-Ga_2O_3), a stable phase of the compound. β-Ga_2O_3 development is a result of an increased focus on materials research to improve the overall performance of power-electronic devices over the junction-based approaches of the past. β-Ga_2O_3 stands out for its intrinsic properties, including an ultra-high bandgap (an energy gap of 5 eV), good conductivity and field-holding capacity, and high critical field strength, with the highest ever demonstrated being 5.5 MV/m.

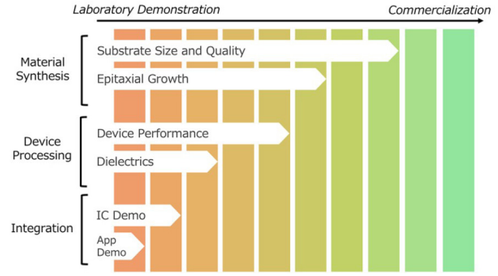


Figure 1: A top-level view of various sectors' technological progress toward commercialization of β-Ga_2O_3 applications as of October 2021 (Source: AIP Publishing)

Processing the material in different ways can result in a variety of properties, demonstrating its flexibility. For example, doping the material from a melt results in a resistivity of 10 mΩ-cm; silicon implantation can decrease it to 1 mΩ-cm. Halide vapor epitaxy on the material can be controlled to have a doping concentration in the range of 10^{15} to 10^{19} cm^{-3}.

Fabricating standard features onto β-Ga_2O_3 is also relatively easy. For example, ohmic and Schottky contacts can be made using standard metals like titanium, aluminum, and nickel at relatively low annealing temperatures. Wafering and lapping of the material can be done using standard production tools. Different dielectric materials, such as Al_2O_3 deposited using the atomic layer deposition method, can be used as gate dielectrics (Figure 1).

THE PROPERTIES OF GALLIUM OXIDE

The high critical field and relatively low mobility of β-Ga_2O_3 enable it to demonstrate better performance than SiC and GaN. The properties of the material grown from a melt make it possible to fabricate high-quality crystals at a lower-cost bulk GaN, SiC, and diamond. The prime transistor figure of merit for β-Ga_2O_3 is approximately 3× that of 4H-SiC and 20% better than the prime transistor FoM for GaN.

These advantages over existing wide-bandgap materials position β-Ga_2O_3 as a viable and low-cost alternative with increased performance. There are challenges, however, that are holding back its large-scale commercialization.

In the context of material properties, the very low thermal conductivity of β-Ga_2O_3 stands in the way of efficient heat transfer, a critical aspect of power-electronic devices. Achieving thin β-Ga_2O_3 dies will be central to the efforts to improve the material's thermal conductivity and develop better heat-removal techniques for β-Ga_2O_3 devices.

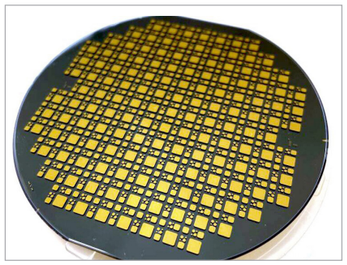


Figure 2: A typical semi-conductor wafer (Source: Tip3X)

The material also has a flat valence band, which results in negligible hole transport, meaning a lack of p-type junctions. This prevents the formation of avalanching p-n junctions and therefore is a concern for devices deployed in regions with noisy power supply or applications involving the need to rapidly take over large inductive loads, such as uninterruptible power supplies. The electric fields at the die edge can affect the device rating: Poor management can result in decreased performance and reliability, and the lack of a p-type may worsen the problem. Lack of a p-type also imposes restrictions on the design of e-mode transistors.

Various die termination methods are under investigation, such as bevel termination and termination using p-type oxides. However, current solutions to mitigate this problem involve tight process controls, casting a ray of doubt over the material's viability (Figures 2 and 3).

The small wafer size of β-Ga_2O_3 relative to other semiconductors is also a problem, as larger wafer sizes can help decrease the cost of fabrication while increasing crystal quality and lowering the defect rate. Current methods to fabricate β-Ga_2O_3

devices use a maximum wafer size of 100 mm, whereas the industry-standard semiconductor wafer diameter is 150 mm, with more companies heading toward 200 mm. β-Ga_2O_3 fabrication must move toward these larger wafer sizes to take advantage of the advanced fabrication infrastructure already in place. Moreover, there is no reliability data available for devices made from β-Ga_2O_3, as research into this aspect remains in its infancy.

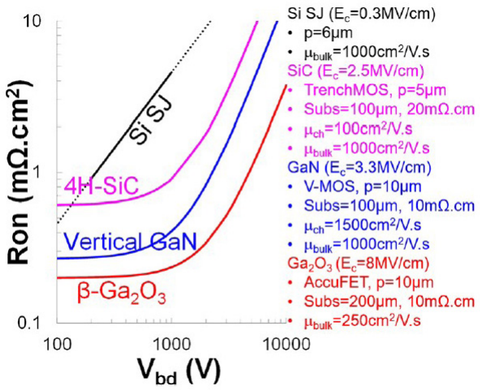


Figure 3: Calculated theoretical performance (R_{ON} vs. V_{BK}) for vertical power devices. Assumptions used in the calculation are listed on the right. The model takes into account the contact, channel, drift, and substrate resistances; p denotes the cell pitch. (Source: AIP Publishing)

There are also economic factors that need to be addressed, such as the loss of some parts of the expensive precious-metal crucibles (in the case of manufacturing methods like edge-defined, film-fed growth [EFG], and the iridium crucibles used with the Czochralski [CZ] method) during the bulk manufacturing of β-Ga_2O_3 crystals. Increasing the size of the substrate, as demanded by the state-of-the-art technologies of other semiconductor materials, tends to worsen the problem and accelerate the decline of these crucibles.

Researchers in China have reportedly developed methods that can mitigate the problem and thereby decrease the cost of the manufacturing process by about 10×, but large-scale implementation of this technology is yet to be seen. Methods suitable for epitaxial layer growth for vertical β-Ga_2O_3 devices require technology not present on state-of-the-art machines.

THE ROAD TO VIABLE DEVICES

There is a great deal of interest and research in the design, development, and commercialization of devices fabricated with β-Ga_2O_3. This interest is driving the impressive growth of substrate-fabrication technologies as companies move toward commercialization.

Although there have been many device demonstrations, considerable optimization work remains to be done to address the aforementioned challenges. Nonetheless, β-Ga_2O_3 technology has reached an exciting moment in its maturity, where the material is readily available and the challenges holding it back from being utilized in devices are well-known and well-documented.

A concerted industry effort must first take root, but successful development of technologies for large-scale, economically viable β-Ga_2O_3 manufacturing would open the door to commercialization, yielding high-reliability devices that fully utilize the material's advantages.

Maurizio Di Paolo Emilio is editor-in-chief of Power Electronics News and EEWeb.

EE Times 50

NANOTUBES AND 2D MATERIALS: CHANGE AGENTS FOR THE SILICON AGE?

BY RICHARD COLLINS

Are these materials the next generation, the long-term future, or 'never in a million years'?

The relevance of Moore's Law in 2022 and beyond has been a topic of exhaustive discussion. Emerging commercial developments, announced by major semiconductor manufacturers to support our data-hungry world, all surround packaging improvements that enable higher interconnect densities and improve integration capabilities. Despite what may appear to be waning interest, major companies and academics around the world continue to look further afield at the potential for nanotubes and 2D materials to bring another stepwise change to the silicon age.

In this article, we will discuss carbon nanotubes (CNTs), graphene, and other 2D materials. CNTs are the older sibling, having been known since the 1990s, whereas graphene was first isolated in 2004, and the researchers who achieved it went on to win the Nobel Prize in 2010. The superlative properties attributed to graphene do not need to be repeated here. It is also a broad family with limited standardization and is regularly misunderstood as a result. Among the numerous forms and grades of graphene, nanoplatelets are exhibiting the most commercial traction for composites, heat spreaders, batteries, and beyond, and although there is some opportunity for these conductive powders in electronics (e.g., printed sensors), this is not the main area to note. Instead, the main discussion is on wafers and films.

Roll-to-roll, chemical vapor deposition (CVD)-grown films presented an exciting prospect and have been extensively explored, most notably by major East Asian electronics companies, but with little uptake. If these major electronics companies have struggled to find a use case with all their experience and reach, then there is a larger problem. The issues are numerous, but the central considerations are the transfer challenge (e.g., from copper to silicon) and the quality of the incumbent products, meaning some innovation is required to see these films take off.

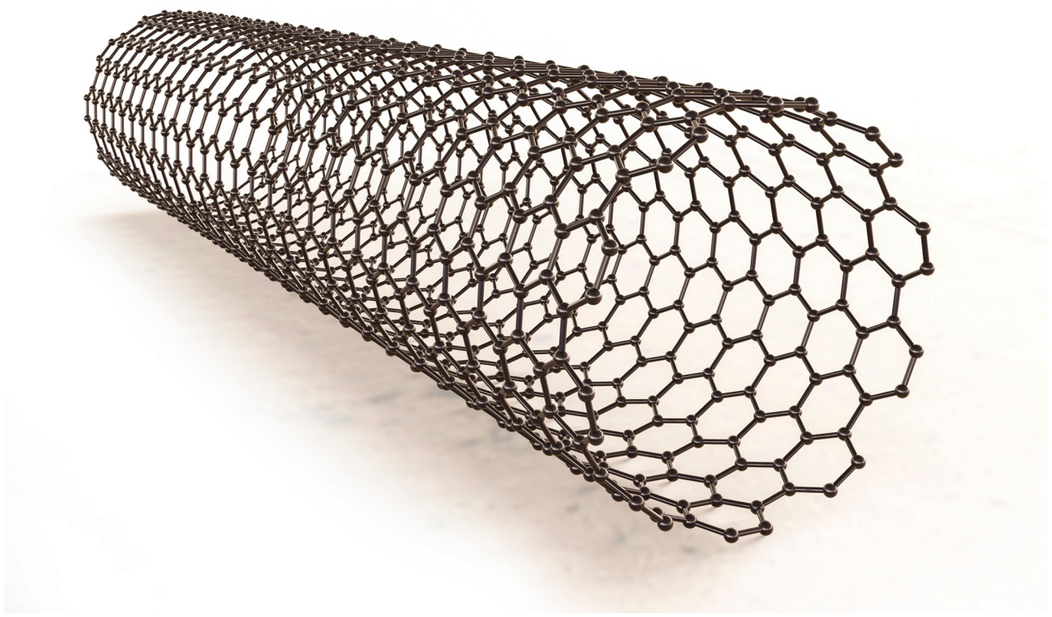
Wafers have a different story: There are major commercial advancements in graphene field-effect transistors (GFETs), with graphene grown on a metal substrate via CVD and transferred to silicon or with graphene grown directly

on the substrate of choice. However, before everyone rushes to build a fab, note that graphene does not have a bandgap; instead, GFETs are gaining interest in the worlds of sensors and optoelectronics for their mobility and surface-area properties. To achieve a truly revolutionary advance for semiconductor integration, a stepwise change is needed. This keeps developments predominantly in the realm of academia for now, with the likes of spintronics taking center stage.

CNTs have a similar story in electronics to that of their younger sibling, with one crucial difference: Single-walled carbon nanotubes (SWCNTs) can have different chirality, which crucially means they are either metallic or semiconducting. In terms of opportunities beyond silicon for nanocarbons, the most promising are CNT FETs, which can theoretically outperform MOSFETs given the superior electrical and thermal properties. IBM, TSMC, and countless other companies and research institutes worldwide have made announcements, filed patents, and generally explored this space. The design approaches vary widely, with many looking at the use of CNT FETs in the construction of monolithic 3D systems with low-temperature fabrication processes.

One challenge, among many, for CNT FETs is achieving consistency in incorporating the identical, highest-quality nanotubes.

Although this work is mostly at the research level, some developers are taking the first steps toward commercialization. This is most notably seen with Nantero, which has claimed in a 2022 press release that its CNT-based nonvolatile memory will be cost-competitive with DRAM in two years.



Finally, there is the broader world of 2D materials beyond carbon. Expanding the materials-engineering toolkit to the point where engineers can play "atomic Lego" and pick layers based on the desired properties is an enticing proposition. There are numerous materials being explored. Boron nitride, for example, offers dielectric yet high thermally conductive properties, and MXenes like Ti_3C_2 have a whole range of fascinating properties; these are just the tip of the iceberg of hundreds of proposed analogs. But perhaps the two of greatest note are the transition metal dichalcogenides (TMDs) and phosphorene.

Given their bandgap, TMDs — mainly MoS_2, WS_2, and WSe_2 — are the main area of interest for 2D materials' role in semiconductors; manufacturing processes are improving, and there is evidence of maintained interest from TSMC, Intel, Samsung, and others. Phosphorene was first isolated in 2014, but the academic publication rate has dramatically accelerated since then, with researchers drawn by the material's anisotropy, high mobility, and tunable bandgap. Whereas 2D MoS_2 is an n-type semiconductor, phosphorene is a p-type. Researchers exploring heterostructures are publishing an increasing number of papers reporting promising results. And we would be remiss not to mention the next generation of these next-generation materials: antimonene and 2D-alpha-germanium. Antimonene is notable in that it demonstrates good stability and a predictable bandgap. The wonder material is dead; long live the wonder material.

IDTechEx believes there is a role for nanotubes and 2D materials beyond carbon and SWCNTs in the semiconductor industry, but the path is long and tortuous. It will take patience, innovations, and significant capital. "Graphene and other 2D material-based technologies" were mentioned in the second pillar of the EU Chips Act, specifically with the development of pilot lines for innovative production. We have already seen European public funding in this space, most notably with the 2D Experimental Pilot Line (2D-EPL), a €20 million project launched in October 2020 by the Graphene Flagship, whose partners include imec, AMO, VTT, Aixtron, Graphenea, and others. Private funding is mostly for in-house R&D behind closed doors, but there is increasing attention on GFETs as the first sensor products come to market.

IDTechEx has been following the CNT, graphene, and 2D materials industry for more than a decade and has released unbiased market reports on these topics.

Richard Collins is research director for North America at IDTechEx.

ADVANCED SEMICONDUCTOR PACKAGING: FROM TECHNOLOGIES TO MARKETS

BY YU-HAN CHANG

The slowing of Moore's Law and the rising development and manufacturing cost of monolithic silicon integrated circuits are the two primary forces driving development of advanced semiconductor packaging technologies. This article recaps developments thus far and summarizes IDTechEx's research on where technology and market trends are headed for advanced semiconductor packaging.

How have semiconductor packaging technologies evolved?

Semiconductor packaging approaches in which components are individually packaged and integrated at the printed-circuit–board level date back to the 1970s, but with electronic device design miniaturization and rising demand for data-processing module capabilities, packaging technology requirements evolved to push beyond board-level integration. The first advancement was package-level integration, followed by the arrival of wafer-level integration. Wafer-level integration can yield at least 10× higher connection density, a substantially smaller footprint suitable for size-sensitive applications, and overall superior performance.

Which semiconductor packaging technologies should be considered "advanced"?

Wafer-level integration encompasses fan-in, core fan-out, high-density fan-out, 2.5D IC, and 3D IC packaging technologies. However, not all of them will be considered "advanced" semiconductor packaging technologies; according to IDTechEx, only those with a bumping pitch size of less than 100 µm are included. This is because only with such scale can the requirement for high-data–processing applications be met. Therefore, only high-density fan-out, 2.5D IC, and 3D IC are considered to be advanced semiconductor packaging.

What are the development trends for these technologies?

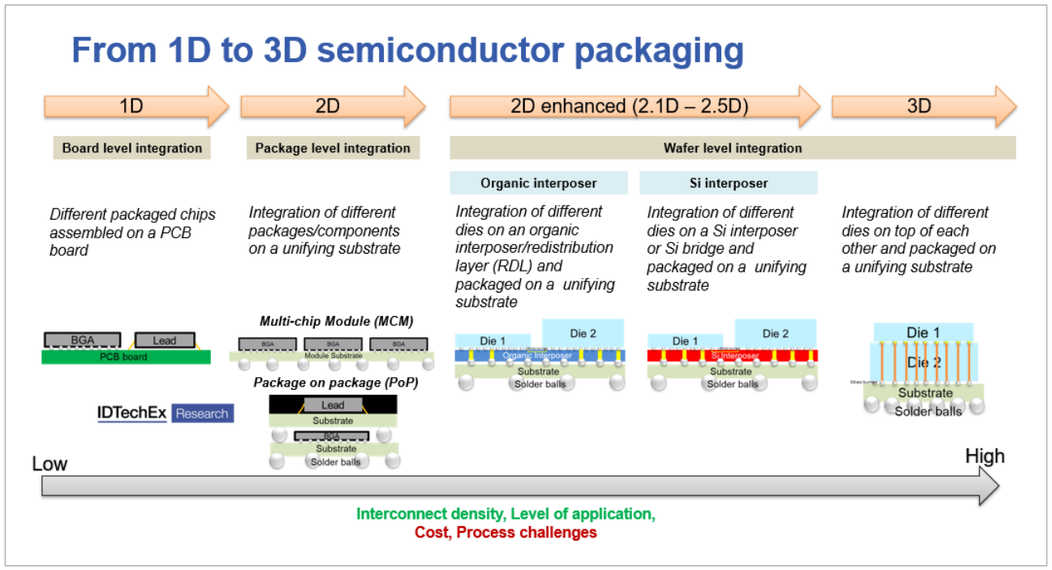
Here, we will focus on the transition from 2.5D hybrid integration (horizontal plus vertical integration) to complete 3D vertical integration. This shift is an essential piece of the puzzle for future data-centric applications. The initial task is to scale the size of the bumping pitch. The bump pitch size in a 2.5D IC package ranges between 25 µm and 40 µm, depending on whether the interposer material is silicon or polymer. For a 3D IC stacking package, the bump size must be scaled down to a single-digit–micrometer dimension. According to TSMC, the bump pitch for stacking N7/N6 chips is 9 µm, whereas the current state of the art is 6-µm bond pitch for N5 chip stacking. That figure will drop to 4.5 µm for N3 chip stacking and continue to decline for future-generation ICs.

It is challenging to stack two chips with such a small pitch size. High-precision alignment on bonding dielectric materials at ambient temperature must be accomplished, and the copper filling materials must be properly controlled to prevent overflow during the annealing process. Furthermore, thermal management becomes critical for packages at such a small bumping scale. Hence, package design that allows for better heat transmission, along with plausible liquid-cooling technologies, must be considered.

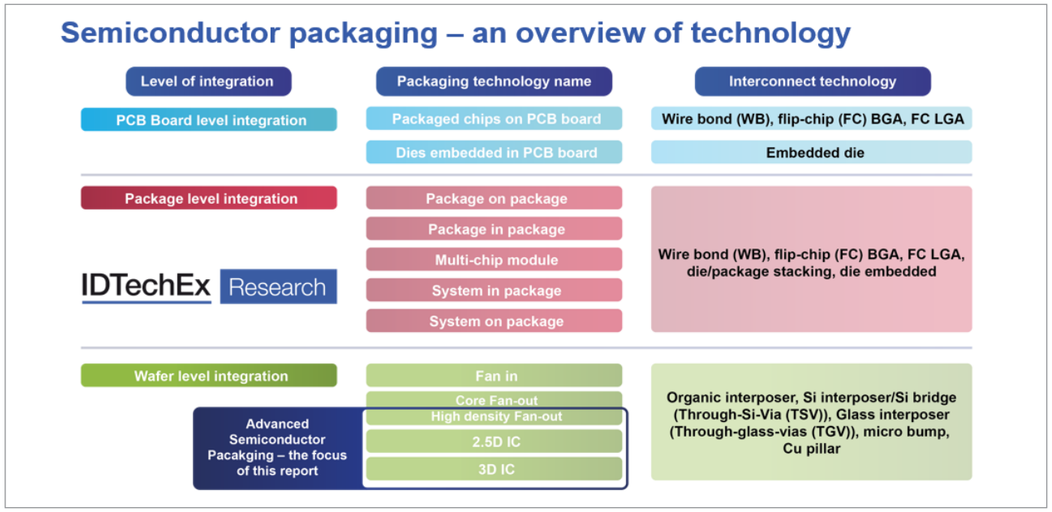
An ultimate 3D IC system will incorporate many logic and memory devices stacked on top of each other. To realize such a system, the logic and memory devices must be fabricated at low temperatures that are compatible to back-end-of-line (BEOL) processes, and each layer of the device needs to be thin enough to allow for the fabrication of ultra-dense vias for connection purposes.

What are the growth drivers in key application areas?

IDTechEx identifies high-performance computing (HPC) applications/data centers, communication networks, autonomous cars, and consumer electronics as four main application areas for advanced semiconductor packaging. The constantly expanding data-processing requirement is the unifying growth factor in all of these applications. Despite this unifying growth driver, however, each application area has its own set of needs that dictate the selection of advanced semiconductor packaging technologies.



(Source: "Advanced Semiconductor Packaging 2023–2033," IDTechEx)



(Source: "Advanced Semiconductor Packaging 2023–2033," IDTechEx)

For example, in HPC applications/data centers, the goal is to provide excellent data-processing capabilities; hence, despite their higher cost, 2.5D IC technologies that employ a silicon interposer or silicon bridge are still the way ahead. On the other hand, for consumer electronics such as smartphones or smartwatches, the focus is on miniaturization and cost; therefore, organic-based packaging technologies are still the top choice.

For 5G communications and beyond, where transmission loss is one of the key challenges, advanced packaging technologies are utilized to bring the antenna closer to the RF IC chip and thereby minimize transmission loss. Examples include antenna-in-package, which is currently the most viable option for 5G mmWave, and antenna-on-chip/-wafer, which is under intense development to drive down cost.

For future autonomous vehicles, meanwhile, heterogeneous integration of central processing units and other components, such as high bandwidth memory and reliable power delivery systems, will bring new opportunities for advanced semiconductor packaging and innovation.

IDTechEx has released a market research study titled "Advanced Semiconductor Packaging 2023–2033." This study examines the market prospects, key player analysis, and most recent developments in advanced semiconductor packaging technologies. IDTechEx leverages years of research in data centers, autonomous vehicles, 5G, and consumer electronics to provide the reader with an understanding of how advanced semiconductor packaging is influencing various industries and what the future may hold. The report also provides a comprehensive examination of the semiconductor market as a whole.

Yu-Han Chang is a technology analyst at IDTechEx, leading research in wireless tech and semiconductors.

EE Times 50

DIGITAL INSPIRATION FROM THE **ANALOG BRAIN**

BY SALLY WARD-FOXTON

Neuromorphic computing could be coming into its own, but should brain-inspired computing be digital or analog?

The human brain is the most efficient computer we know of, using 80 billion neurons to enable us to carry out complex and varied tasks, all while learning continuously. Brain-inspired, neuromorphic computing has been a tantalizing possibility for decades, starting with the work done in the 1980s in the lab of Caltech professor Carver Mead. Commercially viable neuromorphic chips are emerging, but there's a lingering divide on the best route to implementation.

"It is astounding how much effective computation gets done in the 20 W in our brain, and that is really what we set out to try to figure out when we started the whole neuromorphic thing," Mead told EE Times in an exclusive interview with Nitin Dahad earlier this year. "We wanted to understand that phenomenon: How can it possibly be?"

In the human brain, neurons communicate via short voltage pulses, called spikes, from cell to cell. Information can be encoded as sequences of spikes from particular neurons. The magnitude of these spikes does not matter, but their timing is critical. If a sufficient number of spikes arrives at a neuron at around the same time, the neuron will fire, sending its own spikes to its neighbors. A single neuron may be connected to thousands of others.

Mead said there is no clear consensus on whether digital electronics will provide the solution or whether neuromorphic chips should be analog or mixed-signal, like the brain. "In the neural system in brains of animals, the signals that go over any appreciable distance are all digital — the nerve spikes," he said. "The computation in the dendritic tree of neurons is all analog, or it's a combination. You have signals that come from the nerve spikes of other neurons and you're aggregating those in an analog way, but they're quasi-digital in nature. No one has yet been successful in building a thing that works like the dendritic tree of neurons. It's a little surprising, but it's a very difficult thing.

"The challenge, as a technical achievement, to realize a thing that works like a real dendritic tree, requires a level of gain control and stability, and that's beyond anything that has been done. When I finally gave up, I was trying to do that," Mead added.

ANALOG SPIKING

Innatera's spiking neural network processor is a mixed-signal design the

company describes as “digitally assisted analog.” Neurons and synapses are implemented in analog silicon to maintain ultra-low power consumption. Analog electronics also allow continuous time networks (digital electronics would require discretization). This is important because spiking algorithms must be able to hold particular states over a period of time.

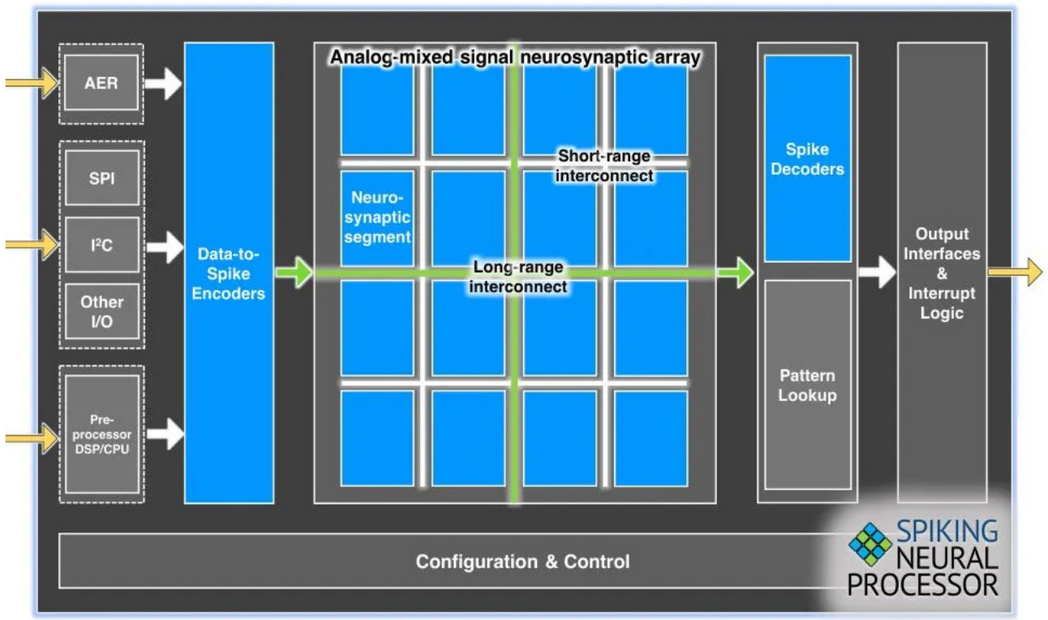
“Doing this is much easier in the analog domain — you don’t have to shift the complexity of keeping state into the network topology,” **Innatera CEO Sumeet Kumar told EE Times**. “Our compute elements naturally retain that state information. This is the reason why we do things in the analog domain.”

Innatera’s 28-nm chip consumes just 200 femtojoules for each spike sent, which approaches the actual amount of energy used by biological neurons and synapses. When you consider that a typical audio keyword-spotting algorithm uses fewer than 500 spikes per inference, it becomes clear that analog designs can work with extremely small amounts of power.

However, there are some downsides to using analog. Minor inconsistencies between devices on the chip can be a problem for accuracy, and inconsistencies between chips can cause problems when the same trained model is deployed. Innatera tackles device inconsistency by grouping neurons into segments, which are carefully designed to match features like path lengths. Inconsistencies between chips are taken care of in software.

ASYNCHRONOUS DIGITAL

Asynchronous digital electronics can be used to get the benefits of going digital while still preserving the data encoded in the timing of spiking signals. The most famous example of asynchronous digital processing for spiking networks is Intel’s Loihi processor, now in its second generation.

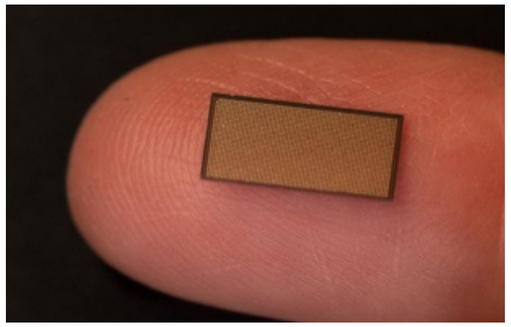


Innatera’s chip uses an analog/mixed-signal array of neurons. (Source: Innatera)

“In neuromorphic computing, the computation is emerging through the interaction between these dynamical elements,” **said Mike Davies, director of Intel’s Neuromorphic Computing Lab**. “In this case, it’s neurons that have this dynamical property of adapting online to the input received, and the programmer may not know the precise trajectory of steps that the chip will go through to arrive at an answer.

“It goes through a dynamical process of self-organizing its states and it settles into some new condition,” he added. “That final fixed point, as we call it, or equilibrium state, is what is encoding the answer to the problem that you want to solve. So it’s very fundamentally different from how we even think about computing in other architectures.”

The second-generation Loihi chip gave neurons a full instruction set, rather than just configurability, which means they can be programmed to emulate different models of the neuron. The new architecture has also departed from its strict biological inspiration and is allowing spikes to have 32-bit integer magnitude (rather than just 1 or 0). Being able to send a number along with the timing data for each spike means the same problems can be solved with fewer resources, Davies said.



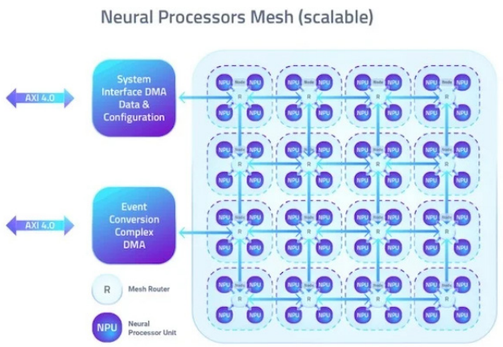
Intel’s Loihi 2 chip fits a million neurons in half the die size of Loihi’s first generation. (Source: Intel Labs)

Loihi 2 has 1 million neurons and consumes about 100 mW in typical operation.

Another company using asynchronous digital electronics to mimic biological spiking networks is BrainChip. To get around the issue of how best to train spiking networks, which remains an open research question, BrainChip’s device is optimized for convolutional neural networks (CNNs). CNNs are widely used in computer vision today. BrainChip trains these CNNs in the usual way and then converts them into spiking networks for the power and latency benefits. Like Loihi 2, BrainChip’s architecture is set up to handle spikes with a magnitude, which is required for converted CNNs.

In one example, a keyword-spotting model running on BrainChip’s development board consumed as little as 37 µJ per inference (the equivalent of 27,336 inferences per second per watt). BrainChip has also shown demos of its Akida chip performing image-processing tasks, such as person detection and face recognition, but it also works with other sensor types, including gustatory and vibration sensors.

BrainChip's Akida chip "is ready for tomorrow's neuromorphic technology, but it solves today's problem of making neural network inference possible on edge and IoT devices," **Anil Mankar, BrainChip co-founder and chief development officer, told EE Times**.



BrainChip's Akida processor uses asynchronous digital electronics to run spiking neural networks, converting from deep-learning networks for the power benefits. (Source: BrainChip)

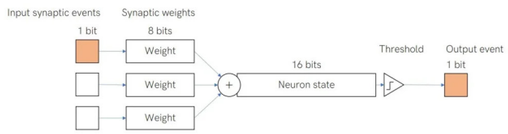
INTEGRATE AND FIRE

Also working on asynchronous digital hardware is SynSense. This startup has developed a spiking convolutional core, tailored to event-based versions of CNNs. The approach differs from BrainChip's, however; SynSense trains spiking networks directly by using approximations to perform backpropagation, the algorithm widely used to train deep-learning networks.

While SynSense uses single-bit spikes, like the brain, the company's neuron uses 8-bit synaptic weights, a 16-bit neuron state, and a 16-bit threshold. The neuron uses the simplest model of the neuron, "integrate and fire" (as opposed to more complex models, such as "leaky integrate and fire," where the internal state decays when there is no input). The SynSense neuron adds an 8-bit number to a 16-bit number, then compares it with the 16-bit threshold.

"It was somewhat surprising to us at the beginning that we could cut down the neuron design to this degree of simplicity and have it perform really well," **Dylan Muir, SynSense's senior director of global business development for algorithms and applications, told EE Times**.

In a demonstration of a system monitoring the intention to interact (whether or not the user was looking at the device), the SynSense stack processed inputs with latency below 100 ms and with less than 5 mW of dynamic power consumed by sensor and processor.

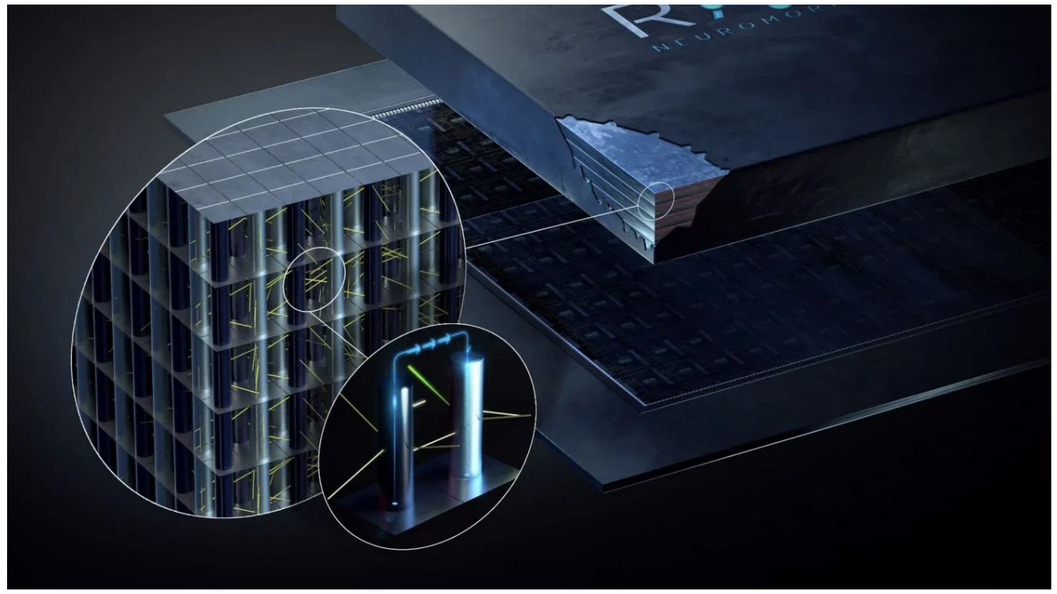


SynSense uses the integrate-and-fire model of the neuron, with biology-inspired single-bit spikes. (Source: SynSense)

SynSense is targeting applications like smartphones and smart-home devices. The company is also working with event-based camera company Prophesee to exploit synergies between Prophesee's retina-inspired camera and SynSense's brain-inspired processor.

ANALOG COMPUTE

Of course, it is also possible to take direct inspiration from the brain without using spiking networks at all. Rain Neuromorphics is building an analog compute chip that uses random connections between memristors to mimic the connections between neurons in the brain. Rain hopes to exploit the sparse connections seen in the brain to make a power-efficient chip.



Rain's analog chip uses random connections, analogous to dendrites, between neurons. The columns are coated in a memristive material that imitates the effect of the synaptic weights. (Source: Rain Neuromorphics)

"The reason randomness is important is that if you have a very large neural network, you want to maintain a certain level of sparsity," **Rain Neuromorphics CTO Jack Kendall told EE Times**. "But which neurons do you pick? If you pick in a controlled way, in a lattice or a regular pattern, you're introducing a huge amount of bias or assumption into how you think that information should be processed, but that's contrary to the entire goal of learning. The goal of learning should be to discover that pattern."

Rain's chip will go hand in hand with the company's algorithmic work. The company has demonstrated that complete analog neural networks are possible, including training certain types of networks using analog processor-in-memory chips like Rain's.

Rain's demo chip combines a tenfold reduction in power footprint with cuts in inference speed from hundreds of microseconds to hundreds of nanoseconds. Together, the result could be a thousandfold reduction in energy consumption compared with GPU solutions, the company said. Rain expects samples of its chips to become available in 2024.

Sally Ward-Foxton is a correspondent for EE Times.

GOING BEYOND MOORE'S LAW WITH **QUANTUM COMPUTING**

BY MAURIZIO DI PAOLO EMILIO

The concept of replicating quantum mechanical systems on a conventional computer was first proposed at the California Institute of Technology four decades ago by Nobel laureate Richard Feynman. One of several crucial events that led to the concept of a general quantum computing device was Feynman's concept of a computer that could serve as a quantum mechanical simulator.

At that time, the practical realization of this concept would have required huge compute power. Feynman, like many other scientists and academics, realized that the path to quantum computing had to be pursued despite Moore's Law scaling.

Just as the bit is the elementary unit of information in a classical computer, the quantum bit (qubit) is used by quantum computers to represent information in its quantum form. A classic binary bit can only represent a single binary value (0 or 1) and can only be found in one of the two possible states. A qubit, on the other hand, can represent a 0, a 1, or any proportion of 0 and 1 in the overlap of both states, with a certain probability that it is a 0 and a certain probability that it is a 1. Theoretically, quantum computers can carry out some tasks significantly faster and more effectively than digital computers.

Stability, however, is an issue for qubits: They don't retain data for long and lose their information quickly. Making quantum bits more stable is a prerequisite for maximizing performance.

In the next few years, quantum computers with more than 1,000 qubits will become available, thanks to the investments of tech giants including IBM, Amazon, Google, and Microsoft. Device characteristics around which companies will compete for differentiation and market share include qubit count, port types, communication between qubits, error rates, and operating temperature (Figure 1).

QUANTUM CONCEPTS

Quantum computers are based on three basic concepts.

- **The first concept** is quantum superposition, which is the idea behind Schrödinger's cat paradox.
- **The second concept** is entanglement, which binds quantum particles together in both time and space. The intertwining and combination of several qubits through entanglement greatly accelerates the computational process. Thanks to this property, even qubits that are physically distant from each other can influence each other. Additionally, by measuring the state of one qubit, it is possible to know the state of the other automatically, without having to measure it directly.

  In a quantum processor, multiple qubits in the superposition state are connected to each other to the point that they form a single entanglement. Entanglement is at the root of the incredible computing power offered by quantum computers and is the source of their potential to solve complex tasks beyond the capabilities of traditional supercomputers.
- **The third quantum concept** is related to the probability amplitude, which is a complex number whose modulus squared represents a probability and is used to describe an uncertain quantity. Probability amplitude is associated with each quantum state, and each possible state is given with a certain probability.

An easy way to understand quantum computing is to think of a bit as a coin. At rest, it can be indifferently in the head state or in the tail state. Now let's imagine that the coin is in rotation. During the rotation, the coin is simultaneously in both heads and tails. That is, it is in an overlap of the two states. Applying the coin analogy to qubits, if we think of putting two qubits together and intertwining them, in this way, we will have obtained four states at the same time; hence, two intertwined qubits represent a combination of four states at the same time. More generally, n qubits will represent 2^n states. The computing power of a quantum computer grows exponentially with the number of available qubits.

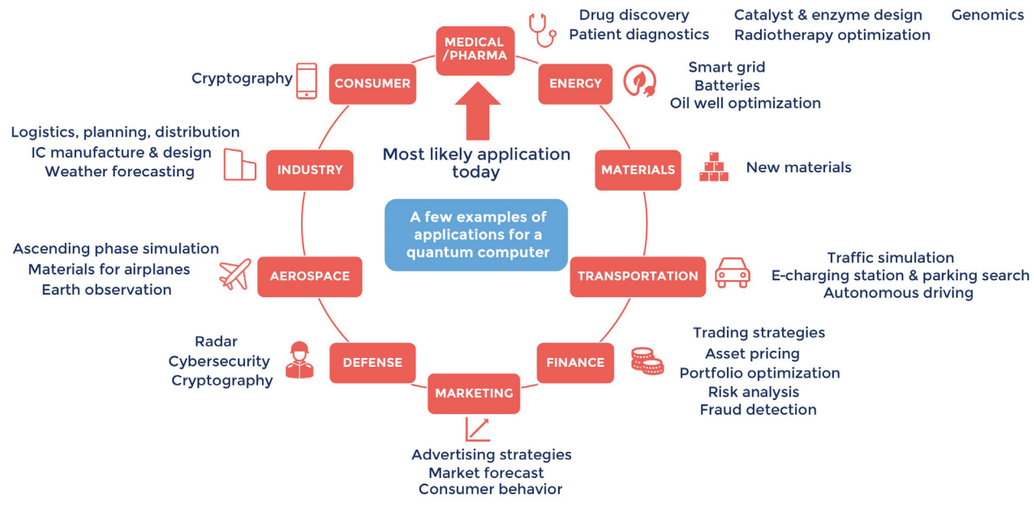


Figure 1: Quantum computing could address complex application challenges in multiple markets. (Source: Yole)

Different techniques can be used for maintaining entanglement. Qubit isolation from environmental noise is the first step. For instance, operating qubits at superconducting temperature reduces environmental noise dramatically. Fault tolerance is another strategy at the system level. Some quantum technologies have built-in tolerance to environmental noise. The trapped-ion approach appears to outperform superconducting technology in this area.

QUBIT CHALLENGES

The control of qubits is one of the difficulties presented by quantum computing. To protect the delicate qubits from thermal and electrical noise, they are built inside a cryogenic refrigerator and controlled by multiple racks connected to them by complex wiring. A huge number of these connections must be introduced as qubit counts rise, which leads to an extremely complex hardware environment. Therefore, reducing the wiring is a challenge that must be addressed. Future chips will increase the scalability of quantum computers by supporting hundreds or even millions of qubits, thereby lowering the complexity of the interconnections in the quantum system, which is one of the major obstacles to the commercial feasibility of quantum computers.

Error correction is another challenge that requires extensive work and is currently being researched. Error correction is critical in most quantum computer–related projects because it helps preserve the delicate quantum states on which computation depends. The operations needed for the correction of errors are highly complex, as they must keep the quantum information unaltered. One way to improve fault tolerance is to delegate some of the computation to a CPU.

THE PHOTONICS ROUTE TO MILLIONS OF QUBITS

In some applications, quantum computing might provide exponential acceleration over traditional computing. The attainment of fault tolerance, which means that calculations of any length or scale can be executed in presence of noise if it does not exceed a specific threshold, is a key and outstanding barrier to making quantum computing feasible.

A quantum computer is essentially a device that produces qubits in particular states, transforms them using quantum gates, and then measures them. Entanglement, which fuses qubits so that a description of each qubit's state is impossible, is used by quantum gates that act on states.

In addition to this gate paradigm, a quantum computer can be conceptualized as a cluster state or as beads connected on a string. The wire in Figure 2 represents the entanglement between the qubits (represented by beads). The measurement of the qubits is the actual computation, not merely a reading of the computation. A different gate will be applied to the qubit depending on the measurement configuration.

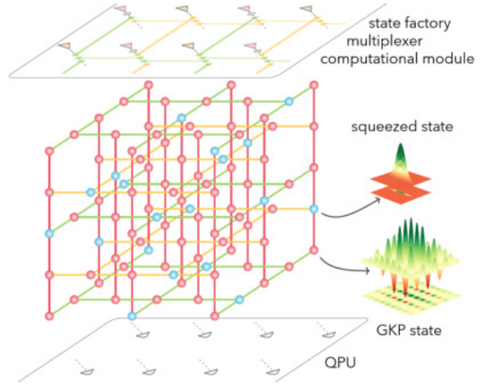


Figure 2: GKP states (Source: Xanadu Quantum Technologies)

Xanadu Quantum Technologies has developed a photonic architecture essentially consisting of four blocks: the state preparation factory, the multiplexer, the calculation module, and the photonic quantum processing unit (QPU). The multiplexer performs many generations of states in parallel to increase the probability of producing a Gottesman-Kitaev-Preskill (GKP) state. The QPU performs the necessary measurements to implement any quantum algorithm and correct errors.

The cluster state must be at least two-dimensional for the computer to be universal and three-dimensional for it to be fault-tolerant. Through repetition or redundancy, we can successfully protect each piece of information from faults in conventional computers. Redundancy in quantum computers is prohibited, however. To enable topological quantum discrete-variable error correction, 3D clusters are employed.

GKP states make good qubits because they have built-in but difficult error-correcting capabilities. The GKP code encodes logical qubits into harmonic oscillator grid states, offering a potential method for fault-tolerant quantum processing. On the other hand, for the code to be fault-tolerant, the quality of the grid states needs to be extremely high.

To synthesize qubits rather than single photons, silicon nitride enables the creation of compressed states. The compressed states can be used to produce error-tolerant qubits because they are formed deterministically (GKP states).

The method used by Xanadu has a number of benefits, including room-temperature computing and scalability up to 1 million qubits through the optical network. Cryogenic requirements are reduced, and the homodyne detectors set the device clock speed in the QPU, which operates efficiently and fast.

Photonics is an intriguing technology, and this area of R&D is bustling right now. The design, production, and integration of quantum and other optoelectronic solutions are made easier by the modular architecture of photonic chips. Laser and photonic components that have been incorporated into a chip may be utilized to minimize the size of the quantum computer.

The versatility, speed, and room-temperature functioning of the next designs will allow photonics to aid in the quick development of quantum computers.

Maurizio Di Paolo Emilio
is editor-in-chief of Power Electronics News and EEWeb.

A SUSTAINABLE PATH FOR MASSIVE DATA GENERATION

BY JEAN-RENÉ LÈQUEPEYS

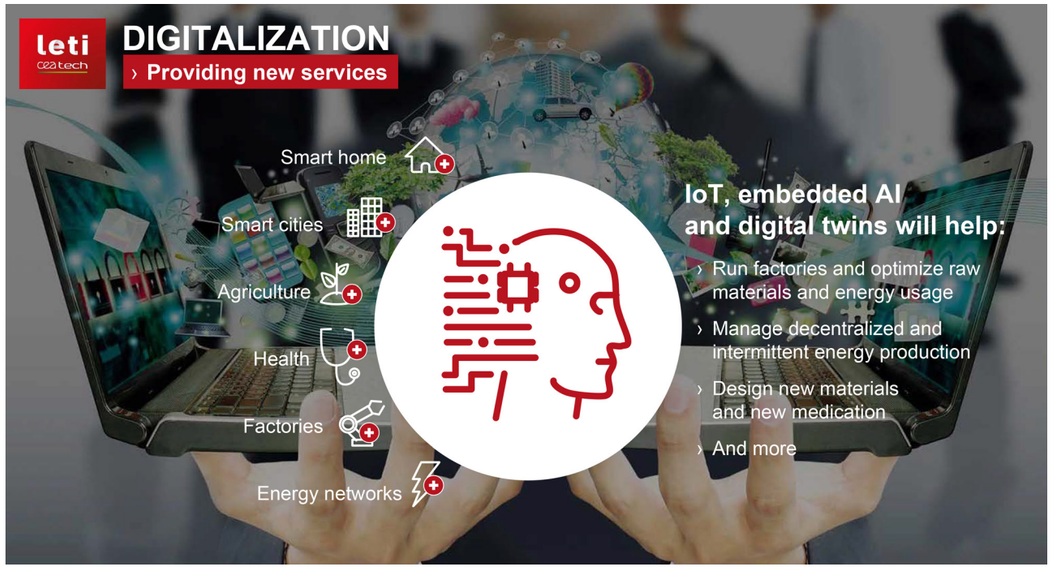
Global digitalization is fueling rapid and exponential growth in the amount of data generated — from 64 zettabytes in 2020 to an expected 2,000 zettabytes in 2035.

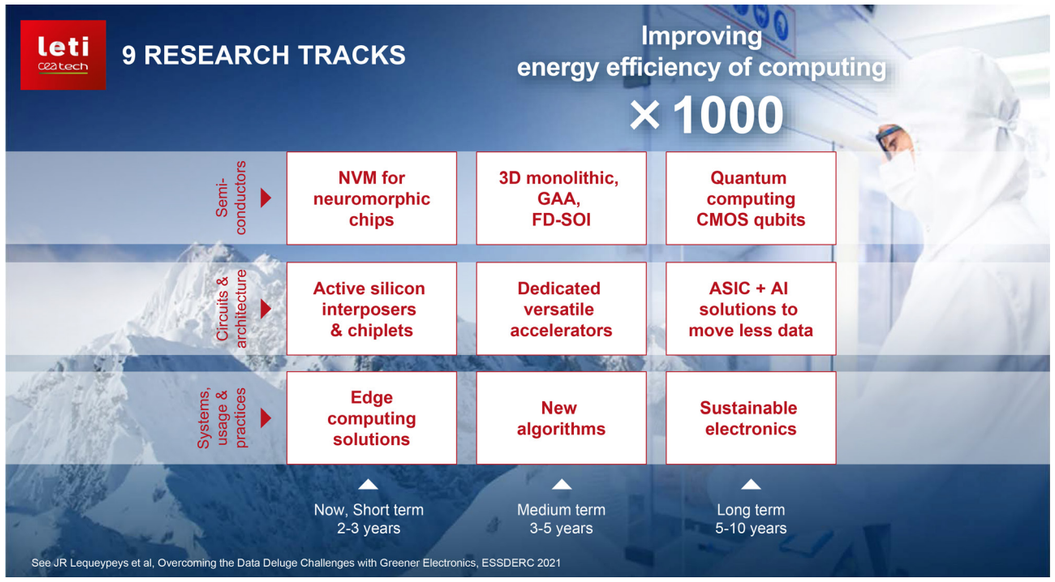
By 2030, 60 billion connected objects will be in service, and many billions of them will support a broad range of computing and processing solutions, from industrial and manufacturing processes, energy networks, and factory management to sustainable agriculture, better health care, and housing.

The internet of things, embedded artificial intelligence, and digital twins increasingly will help companies and researchers manage factories by optimizing raw materials and energy usage and by making it possible to manage and tap intermittent and decentralized energy sources. We can expect improved design for new materials and new medications with further digitalization.

MICROELECTRONICS IS THE WORKHORSE

Until now, humans and their gadgets and portable devices have generated most of the data. The so-called use stage — user devices, networks, and data centers — represented 55% of the information and communication technology (ICT) carbon footprint in 2018, according to the Shift Project, while machines and equipment used for designing and manufacturing microdevices accounted for 45%.





But that ratio is changing profoundly with the exponential increase in data generation. By the end of this year, the data produced automatically by machines is expected to account for 90% of data generation. Hardware, especially microelectronics, is the workhorse of all this digitalization. We need materials and tools to process wafers and to develop the components, and the electronics components must be packaged efficiently. The semiconductor industry plays a key role in this value chain, so the focus must be not only on building energy-efficient microchips sustainably but also on making sure its own design-and-manufacturing tools and processes leave the smallest carbon footprint possible.

Without energy-efficiency innovations, global ICT could end up with an uncontrolled energy consumption of up to 14% of all uses in 2030, compared with about 4% today. We already anticipate needs coming from Bitcoin, 4K TV, Earth observation nanosatellites, genomics for DNA sequencing and storage, and advanced driver-assistance systems for autonomous vehicles. Innovation will be key to maintaining electricity demand at an acceptable level. We need to avoid all unnecessary energy losses and to reject the concept of electronics obsolescence, instead extending the lifetimes of electronic end products by maximizing component and system life.

In countries with a good electricity mix, reducing the manufacturing-phase impact appears to be essential. Re-industrialization is relevant from this point of view because the impact of manufacturing (CO_2) comes largely from the electricity used in the fab.

ECO-INNOVATION

It is imperative that our research activities be developed with sustainability in mind for both the manufacturing and use phases. CEA-Leti created

an eco-innovation strategic program, which informs our traditional focus on innovation with economic, environmental, and social criteria to produce sustainable technologies and products. For example, we are developing semiconductor technologies with a classical PPAC approach — power, performance area, and cost — to which we have added an "e" for environment and eco-innovation in the process.

APPLICATION TECHNOLOGY CO-OPTIMIZATION

CEA-Leti encourages researchers to think beyond performance-/cost-based analysis and take life-cycle analysis into account from design through application. The driver is eco-efficiency, or the ratio between the provided functions and their environmental impacts. To the concept of design technology co-optimization, or system technology co-optimization, we have added a concept called application technology co-optimization. This improves end-of-life management of our electronic devices, including the relevant databases to ensure life-cycle analyses, and improves optimization of the environmental impact at both the system and usage levels.

REUSING SiC WAFERS

A second example concerns the technologies required for power components based on silicon carbide. The market need is high for these components, and SiC substrates are in short supply. Together with Soitec, we have devised a solution that reuses standard SiC wafers by leveraging Soitec's SmartCut process. This technology slices the wafer by ion implantation, lifting off a thin layer from a donor substrate and transferring it onto a new substrate.

We use a standard SiC wafer as the donor. After adequately preparing the surface, we implant hydrogen ions under the surface to create a fracture line. The donor wafer is then turned over and bonded to a conductive substrate — in practice, a polysilicon carbide wafer. Then via a thermal process, we separate the assembled wafer along the fracture zone to yield two separate wafers. Finishing steps make it possible to reuse the donor wafer and to process the new wafer conventionally.

This approach makes it possible to reuse the donor wafer 10×, reduces the size of the chips by 15%, and improves yields by 20%. The combination of these gains saves 20,000 tons of CO_2 for 500,000 wafers compared with standard SiC technologies. Soitec has announced a new fab in Bernin, France, to manufacture these wafers.

BOOSTING FAB ENERGY EFFICIENCY

Our teams have chosen nine research tracks to improve the energy efficiency of computing by a factor of 1,000 by 2030. We have launched some working groups to support these research tracks and to identify promising R&D topics. For example, at the beginning of the Covid-19 pandemic, we launched an initiative called Leti2030.

More than 200 people participate in this innovation brainstorming, with some external contributors sharing their insights and expertise on the evolution of big markets like health care, 6G, augmented reality/virtual reality, computing, quantum technology, and automotive.

We also share our vision with our key industrial partners to have a good understanding of their mid- and long-term needs. Our roadmaps are updated every year to consider new factors.

OPTIMISTS VS. PESSIMISTS

In microelectronics, the trend is toward a continuous increase in performance in terms of speed, power, and miniaturization. Two philosophical approaches are possible: declinism and sobriety.

Declinism is the belief that a society or institution is in decline. Applied to technology, this pessimist's approach assumes that electronic devices cannot contribute to solutions that deliver useful services.

Sobriety, the athlete's approach, aims at maximizing performance with a given budget or limited resources. Applied to the industry, this philosophy holds that electronic devices and services will contribute to society by providing technology that helps achieve carbon neutrality.

I am clearly in favor of the sobriety approach for bringing useful technologies and services to market. But we have to change our practical approach, break the molds, and think differently if we are to succeed in slashing the energy and environmental footprint of electronic devices.

Jean-René Lèquepeys is CTO of CEA-Leti. He has more than 30 years of scientific and managerial leadership at both CEA and CEA-Leti.

CYRUS KROHN
GROUP PUBLISHER OF **ASPENCORE**

BRETT BRUNE
EDITOR-IN-CHIEF, **EE TIMES**
MANAGING GLOBAL EDITOR, **ASPENCORE**

EDITORS

NORTH AMERICA

STEFANI MUÑOZ
MANAGING EDITOR, **EE TIMES**

GINA ROOS
EDITOR-IN-CHIEF, **ELECTRONIC PRODUCTS**

MAJEED AHMAD
EDITOR-IN-CHIEF, **EDN** & **PLANET ANALOG**

AALYIA SHAUKAT
ASSOCIATE EDITOR, **EDN**

BARB JORGENSEN
EDITOR-IN-CHIEF, **EPSNEWS**

GARY HILSON
DAN JONES
EGIL JULIUSSEN
DON SCANSEN

EUROPE

NITIN DAHAD
EDITOR-IN-CHIEF, **EMBEDDED.COM**

MAURIZIO DI PAOLO EMILIO
EDITOR-IN-CHIEF, **EEWEB** &
POWER ELECTRONICS NEWS

SALLY WARD-FOXTON
CORRESPONDENT, **EE TIMES**

ANNE-FRANÇOISE PELÉ
EDITOR-IN-CHIEF, **EE TIMES EUROPE**

ASIA PACIFIC

YORBE ZHANG
HEAD OF **ASPENCORE APAC**

ECHO ZHAO
CHIEF ANALYST, **ASPENCORE CHINA**

JUDITH CHENG
CHIEF EDITOR, **EE TIMES** & **EDN**,
TAIWAN/ASIA

TAIWAN/ASIA

JIMMY ZHANG
LEFENG SHAO
ZENGDE XIA
LUFFY LIU
RICARDO XIE
AMY WU
CHALLEY PENG
MOMO ZHONG
CLOVER LEE
ILLUMI HUANG
FRANCLIN ZHAO
ELAINE LIN
DEMI XIA
SUSAN HUANG
ANTHEA CHUANG
RYAN TSAI
STEPHEN LAS MARIAS
ALAN PATTERSON

LORI O'TOOLE
CHIEF COPY EDITOR

DIANA SCHEBEN
SENIOR COPY EDITOR

PETER CANNIZZARO
CREATIVE DIRECTOR

JAY DONALDSON
GRAPHIC DESIGNER

ALEXIS BICKFORD
JOE TRENTACOSTI
PRODUCTION TEAM

ACKNOWLEDGMENTS

To Gerard and Lilo Leeds, who escaped Nazi Germany in 1939 and founded CMP Publications in Manhasset, New York, in 1971. The following year, the first issue of Electronic Engineering Times was printed.

To the lineage of editors-in-chief — Girish Mhatre, Steve Weitzner, Richard Wallace, Brian Fuller, Junko Yoshida, Susan Rambo, Dylan McGrath, Brian Santo, and now Brett Brune — who have followed the news traditions and left a lasting mark on the publishing scene.

To the hundreds of seasoned newswomen and newsmen who have joined the EE Times family since the early days.

To Cyrus Krohn, group publisher of AspenCore, who has supported the creation of this eBook.

To Catherine Lanucha, director of content and audience, and Deborah Ray, content operations manager, who have led the overall EE Times 50th Anniversary project.

To Malcolm Penn, chairman, CEO, and founder of Future Horizons; Tim Burgess, senior director for the high-end MCU business at Renesas Electronics; Bernd Westhoff, director of global MCU product marketing at Renesas Electronics; Jean-Christophe Eloy, president and CEO of Yole Group; Luc Van den hove, president and CEO of imec; Ezgi Dogmus, team lead analyst at Yole Intelligence; Poshun Chiu and Taha Ayari, technology and market analysts at Yole Intelligence; Alex Lidow, CEO of Efficient Power Conversion; Victor Veliadis, executive director and CTO of PowerAmerica and professor of electrical and computer engineering at North Carolina State University; Richard Collins, research director for North America at IDTechEx; Yu-Han Chang, technology analyst at IDTechEx; and Jean-René Lèquepeys, CTO of CEA-Leti.

To Analog Devices, Cadence, and Rochester, who have placed their trust in our publication and penned a verse to our anniversary celebration.

To Barbara Jorgensen, editor-in-chief of EPSNews; Sally Ward-Foxton, correspondent for EE Times; Gina Roos, editor-in-chief of Electronic Products; Maurizio Di Paolo Emilio, editor-in-chief of Power Electronics News and EEWeb; Nitin Dahad, editor-in-chief of embedded.com; Majeed Ahmad, editor-in-chief of EDN and Planet Analog; and Chris McGrady, contributing editor to EE Times, who have shared their technical expertise and their understanding of the industry in top-notch analysis pieces.

To Lori O'Toole, chief copy editor, and Diana Scheben, senior copy editor, who have used their linguistic and stylistic skills to bring fluidity and clarity to our prose.

To graphic designer Jay Donaldson, creative director Peter Cannizzaro, and his design team, who have conceived the eBook design and meticulously executed each layout.

And, more personally, to Richard Wallace, who hired me in 2003, and to Junko Yoshida, Bolaji Ojo, and Victor Gao, who convinced me to come back in 2019. I am proud to be part of EE Times' 50-year history and perpetuate the values and legacy of this respected and authoritative publication.

— *Anne-Françoise Pelé*
EDITOR-IN-CHIEF, **EE TIMES EUROPE**