INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS™

2023 IRDS

SEMICONDUCTOR CRISES AND ROADMAPS RESCUES

PAOLO GARGINI, IRDS CHAIRMAN

Table of Contents

List of Figures

List of Tables

About the IRDS

The IRDS™ is a set of predictions that serves as the successor to the ITRS. The intent is to provide a clear outline to simplify academic, manufacturing, supply, and research coordination regarding the development of electronic devices and systems.

The goals of the roadmap are as follows:

- To identify key trends related to devices, systems, and all related technologies by generating a roadmap with a 15-year horizon
- To determine generic devices' and systems' needs, challenges, potential solutions, and opportunities for innovation
- To encourage related activities worldwide through collaborative events, such as related IEEE conferences and roadmap workshops

The shift and evolution of the roadmap from the ITRS to the IRDS™ has translated to an expanded focus on systems. Emphasis has been placed on architectures and applications that deviate from the traditional paradigm of device→circuit→logic gate→functional block→system.

The IRDS™ topics are a forecast of technology and systems characteristics, needed technologies, and capabilities of the microelectronics and systems industries. The intent is to focus on academic, manufacturing, supply, and research coordination regarding the development of microelectronic devices and systems. Visit us for more information about the IRDS and its roadmaps.

Supporting Organizations

The work of the IRDS is supported by the following regional organizations and IEEE societies and technical communities:

IRDS IRC Regional Comments

From Europe:

"International cooperation is key for speeding up technological innovation. The IRDS plays a decisive role in this area and helps identify challenges and possible next generation and emerging technologies in advanced computation and functionalities for future systems and applications. This joint effort will be instrumental for the new electronics' industry and in the digitalization of many domains to reduce footprint, as well as the electronic monitoring targeting societal challenges using sustainable electronic technologies. On behalf of Europe, the European Sinano Institute fully supports the IRDS roadmap activities".

From Japan:

"The new ecosystem of the electronics' industry is driven by system integrators that design products enabled by multiple semiconductor technologies. The role of the IRDS in identifying the fundamental building blocks in the electronics industry spanning from devices to systems and from systems to devices is critical for international collaboration for advanced design, function, compute and manufacturing of future devices and systems."

From US:

"The IRDS is uniquely positioned to guide the evolution of the microelectronics industry because of the expertise and caliber of its technical contributors; the depth and breadth of its content, and the continued popular global interest in its various roadmap topics. This is ever the more important when the US is in the process of implementing the CHIPS Act—they will look to the IRDS. The organization and composition of the IRDS roadmap effort is key to its success, high relevance, and industry regard. The long-standing support of its international partners in Europe and Japan adds strength to the IRDS roadmaps to create a comprehensive network of teams. Such teams, as part of their complimentary regional roadmap efforts, help to carry forward the guidance of the IRDS to focus implementation activities within their spheres of influence in academic research, R&D activities through industrial partnerships, and to progress the industry forward. Critical to its nature as an evolutionary roadmap is the commitment and attention to promising new technologies, systems and architectures on the horizon that will fuel next-generation applications and products."

Acknowledgements

International Roadmap Committee

Europe—Francis Balestra, Enrico Sangiorgi

Japan—Yoshihiro Hayashi, Norikatsu Takaura, Hidemi Ishiuchi

U.S.— Paolo Gargini-chair, Tom Conte-vice-chair

IRDS International Focus Teams

The outstanding work and the team members of the International Focus Teams are acknowledged in each of their roadmap chapters. The IRDS International Focus Teams chairs and co-chairs of these teams are as follows:

Application Benchmarking—Tom Conte

Systems and Architectures—Kirk Bresniker and Stephen Dukes

Outside Systems Connectivity—Michael Garner

More Moore—Mustafa Badaroglu

Lithography and Patterning—Mark Neisser and Harry Levinson

Factory Integration—Supika Mashiro and James Moyne

Yield Enhancement—Slava Libman and Don Wilcox

Beyond CMOS—An Chen, Shamik Das, Matt Marinella, Michael Franks

Cryogenic Electronics and Quantum Information Processing—Scott Holmes

Packaging Integration—Dev Gupta

Metrology—George Orji, Ben Bunday, Yaw Obeng

Environment, Safety, Health, and Sustainability—Leo Kenny and Steve Moffat

More than Moore—Enrico Sangiorgi

Autonomous Machine Computing—Shaoshan Liu and Jean-Luc Gaudiot

Mass Digital Storage—Tom Coughlin and Roger Hoyt

IRDS Project Manager

Linda S. Wilson

Overall Roadmap Characteristics

Alan K. Allan

SEMICONDUCTOR CRISES AND ROADMAPS RESCUES

Abstract

Every couple of decades a major semiconductor crisis arises like clockwork. Typically, some clues on the origin of the crisis could have been predicted in advance by those who were looking for them since usually the weakness and vulnerability of the technical or economical semiconductor supply chain are always quite evident. However, it is human nature to avoid addressing a problem before it dramatically arises. With past crises government intervention became a critical part of the resolution or at least it was very relevant in providing funds and triggering actions that eventually led to a solution. This was the case even though the resulting solutions often were quite different from what had been originally intended. The International Roadmap for Semiconductors (ITRS) and its evolution to the International Roadmap for Devices and Systems (IRDS) provided leadership and continues to play a key role in guiding implementation of such solutions. Four semiconductor crises that occurred over the past 60 years are notable. This narrative provides a historical perspective; reviews the good and the bad lessons learned from these crises and makes recommendations. The development and evolution of the semiconductor roadmap and its vital role in guiding the industry is presented.

The key question is this: Can we analyze what we learned from the earlier crises and study which solutions were effective in resolving each crisis to better deal with the present in-progress crisis that has already, once again, triggered unprecedented government investments in semiconductors?

Note: Those that are mainly interested in the Chips Acts should please go directly to Section 4.

1. 1957: THE RUSSIAN INFLUENCE—CRISIS #1

In 1955, following the successful demonstration of the transistor effect, Bill Shockley moved from Bell Labs to the San Francisco bay area in California to be closer to his aging mother. Collecting funds for his start up was not a problem, based on his notoriety especially after he received the Nobel prize in physics in 1956. He hired young and unknown scientists for his company and was ready to revolutionize once again the world of semiconductors. However, his management style was less than orthodox and on September 18th, 1957, a group of research engineers—later to be known as the "Treacherous Eight"—resigned from his company and formed Fairchild Semiconductors. This new company was launched on October 1st, 1957. It is well known that the destiny of newly formed companies demands that any new enterprise will be financially in the red for few years before it eventually becomes profitable, but this was not going to be the case in this instance.

Sputnik Ignited Silicon Valley!

On October 4th, 1957, (3 days later!) the Sputnik satellite was launched by the USSR. The implications of such a nation owning control of the skies led to dramatic reactions around the world, in part because science had been turned into an instrument of national power. As a result, the Unites States formed the Defense Advanced Research Project Agency, known as DARPA, on February 7th, 1958 (Figure 1).

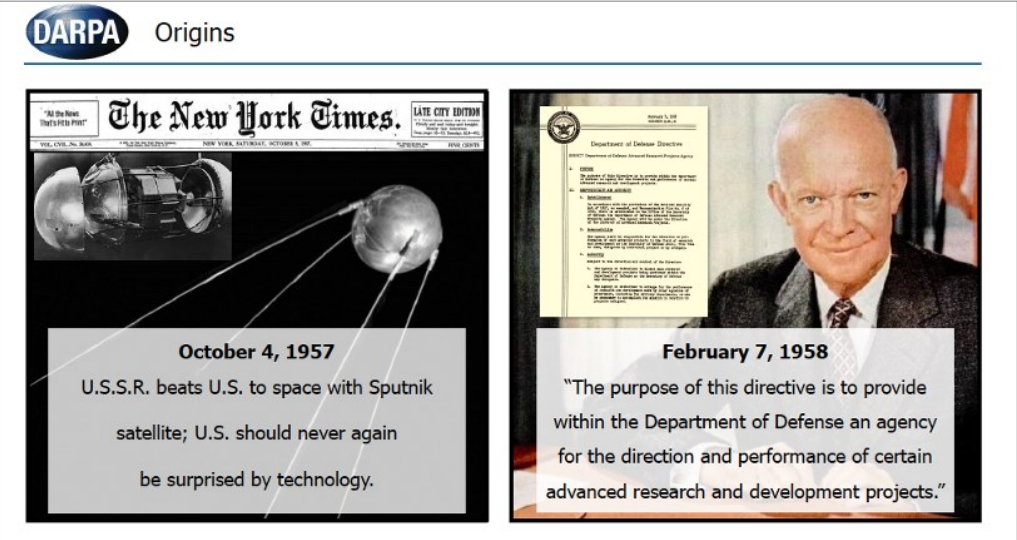
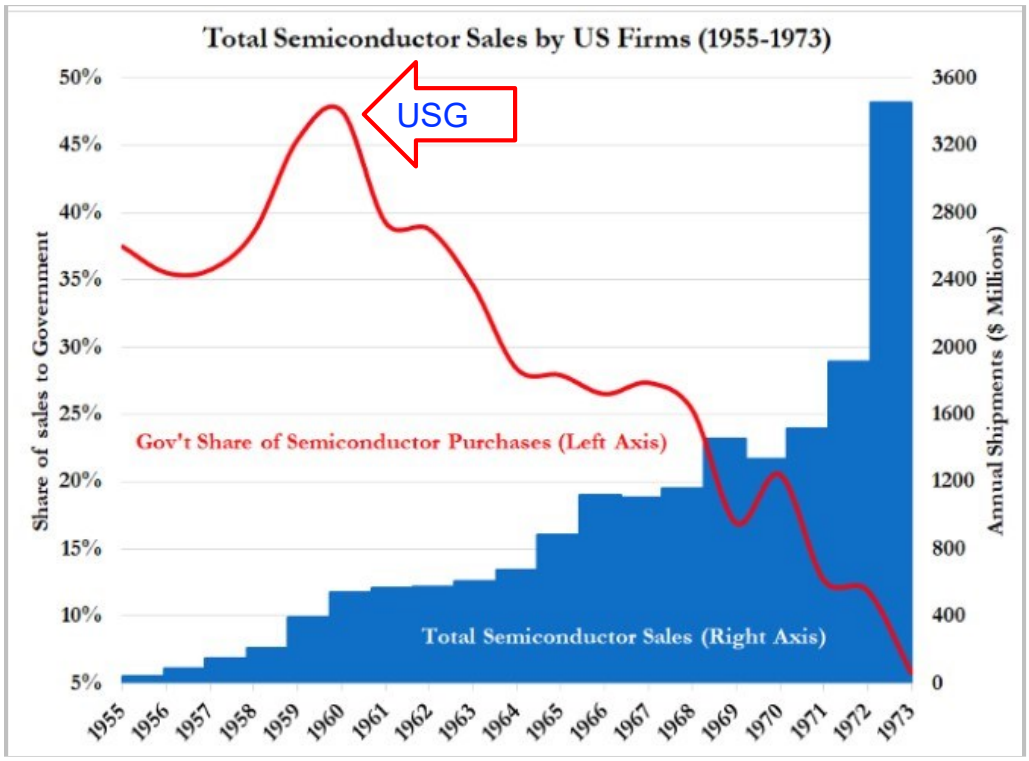


Figure 1 The launch of Sputnik shocked the world and caused the formation of DARPA.

Most of all, the US Government (USG) realizing that the weight of the satellite was a limiting factor and a major challenge for any launch, began to acquire large quantities of transistors for this task. (The weight of silicon-based transistors was much less than that of vacuum tubes that represented the standard building blocks of electronic circuits at the time.) Transistors were bought by the USG at the price of $150 each! Increasing the supply of these valuable electronics components represented the practical goal of the first ever crisis in semiconductors. It goes without saying that in 1958 Fairchild Semiconductors became a profitable division and by 1961 revenue exceeded $20 million out of a total Fairchild Camera and Instrument revenue of $93M—not bad for a start up!

USG Became the Number 1 Venture Capitalist in Semiconductors and Created Silicon Valley

By 1960 the USG market share of semiconductors in the US was close to 50%. In the following years, the number of newly formed startups exploded in the Bay Area south of San Francisco—that, in the past, had been known only for its orchards—and the myth of Silicon Valley was created. By 1973 more than 40 companies had been formed. Correspondingly, as the industry blossomed, the market share of the USG dwindled down to close to 5% (Figure 2).



Source: ICE Semiconductor Data

Figure 2 USG bankrolled the formation of Silicon Valley.

It is fair to say that Sputnik triggered the birth of Silicon Valley and USG became the venture capitalist (VC) that provided the initial and essential funds to accelerate the birth of the semiconductor industry. The planar process, the integrated circuit, the self-aligned silicon gate MOS process, and the introduction of multiple types of memories and the microprocessor, to mention a few, were invented and introduced into manufacturing following the influx of USG investments.

Thank you, USG!

2. 1985: THE JAPANESE INFLUENCE—CRISIS #2

By 1970 the silicon gate had finally made it into high-volume manufacturing and MOS transistors began to systematically replace bipolar transistors in most applications, especially in memory products. However, by 1968, Fairchild management's methods were once again a cause of frustration for some of the ambitious members of the Treacherous Team; furthermore, Fairchild Camera and Instrument division revenue was declining and losing money. On July 18th, 1968, Intel was formed by Robert Noyce and Gordon Moore with Andy Grove joining as a member.

The Many Secrets of Silicon Valley

The magic of the silicon gate process contained several trade secrets that were never fully disclosed in published papers and, in fact, Gordon Moore used to say***: "Write it on Silicon not on paper"*** to make the point. Publishing the latest development results was practically impossible at the time. For example, in 1983 when, after already shipping products with a combined tungsten/polysilicon maze of interconnections, I was finally able to file two patents and publish a paper on this subject, the result was that many people in the company at the time resented giving away "valuable information." In reality, the publication did not reveal secrets, since the process that subsequently actually made it to high-volume production used tungsten silicide instead, which was better and more functional than what was published and the actual composition and other "tricks" used in the process were never revealed.

The Myth of Silicon Valley!

In the 70s and early 80s revenue money and healthy profits continued to come into semiconductor companies because of new and exciting products and the Myth of Silicon Valley kept growing. Many people at the time believed that nobody in the world could even threaten this technological superiority because of critical unpublished trade secrets. Silicon Valley had become like a secret "members only" society (and I was glad to be part of it).

But all of this was soon going to change. Most semiconductor companies in the US were obsessed with achieving the technological leadership position and were not paying enough attention to mundane activities like enhancing wafer yields or improving overall manufacturing productivity methods as money kept flowing in. At the same time, the DRAM technology, the product with the highest wafer volume, was no longer a secret and its roadmap had become completely predictable—demanding a quadrupling of bits per die every 3 years. Also, since the timing of all the new product introductions had become completely predictable, there were no mysteries for these product classes. The result was that within a few months all the companies in the world were introducing very similar products. *So what was going to be the differentiator?*

Equipment Productivity and Manufacturing Methods enabled Lower Product Cost!

In the late 70s Japanese companies had initiated several projects aimed, among other things, at producing better equipment and better manufacturing methods with the intent of becoming self-sufficient in DRAM production. In 1981, Japanese companies accomplished this goal and openly published and declared their intentions of conquering the dominant position in DRAM internationally; however, no one paid attention to it (Figure 3).

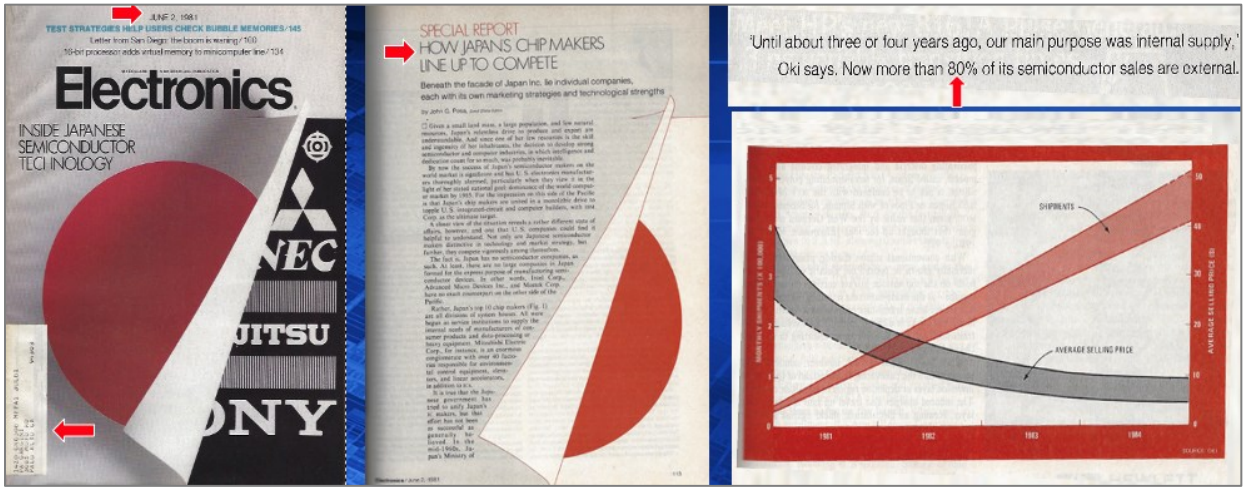


Figure 3 Japan reached self-sufficient status in DRAM and began exporting at competitive prices.

Clues of an upcoming crisis were indeed clear. By 1983, Intel aggressively began acquiring the lithographic process key enablers—steppers from Nikon, a Japanese company! These steppers and the ones made by Canon, another Japanese company, were indeed more efficient and reliable than the established leader, the US company called GCA.

It did not take too long to verify that this threat was real and yet most US IDMs producers were still not paying any attention. In 1985 it became clear that Japanese companies manufactured 92% of all the DRAMs sold in the US. Intel had dropped out of DRAM manufacturing. The second semiconductor crisis had begun!

Silicon Valley Defeated!

The myth of invincibility of Silicon Valley had been broken but nobody was willing to accept the reality that this had been accomplished in a fair way (Figure 4). Nevertheless, the USG and 14 US semiconductor companies in 1987 formed Sematech with the goal of regaining worldwide technology leadership in DRAM by 1993, with each side investing $100M/year in the effort. Processes and products to manufacture DRAMs and SRAM were donated to

Sematech to accelerate the startup but nothing ever came of it. No company really wanted to donate their secret IPs to this effort and Sematech stalled.

It was not easy convincing the US IDMs and equipment suppliers that the Japanese superiority stemmed from equipment performance and manufacturing methods. It took about three years to redirect Sematech funds towards supporting the re-engineering of reliable and efficient equipment in the US and achieving successful results.

By 1990 realty finally sunk in and the mission of Sematech was redirected towards re-energizing the equipment suppliers in the US that were forecasted as facing extinction by the mid-90s. In addition, US IDMs began to take defect reduction and manufacturing productivity seriously and this new approach also was successful.

Table 1 Yields of Japanese companies exceeded the yields of US companies allowing then to sell at lower prices

Yields in percent

	Yield						
Country	1981	1986	1987	1988	1989	1990	1991
United States	55	60	60	67	74	80	84
Japan	45	75	79	81	85	89	93

Source: VLSI Research

By mid-90s the Japan/US market share of manufacturing equipment stabilized around the 50/50 equilibrium point, while in conjunction US companies were then actively and successfully pursuing defect reductions and productivity enhancements (Figure 4).

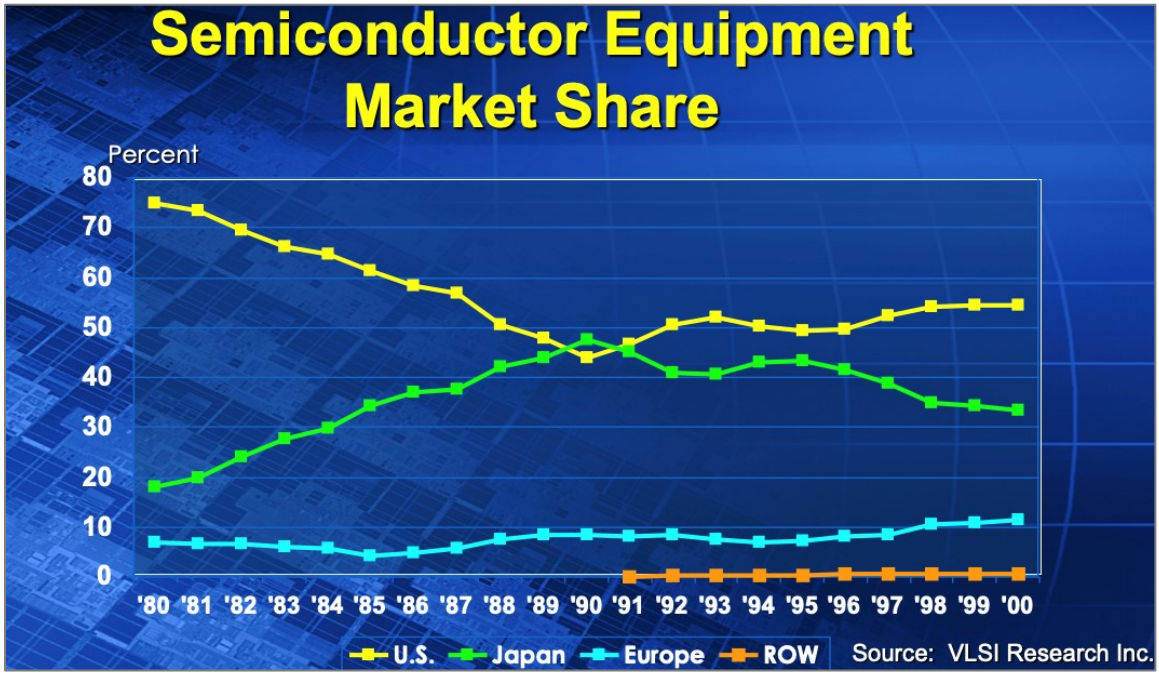


Figure 4 Sematech was instrumental at refurbishing US equipment suppliers.

Successful products supported by adequate technology and manufacturing methods make companies successful in the electronics business.

Investments in new products and new technology generations can only be sustained if the present generation can produce enough net profits to fund the subsequent one. In few words, in a very competitive market requiring massive investments this can only be accomplished if a company has a large enough market share so that sales can generate the profits necessary to sustain the investments required to fund the next round of products and technologies. Andy Grove used to say that *"The semiconductor business is very similar to the behavior of a compulsive and successful gambler in Las Vegas. Each time after a sizeable win instead of celebrating we take all the money on the table and re-invest all of it in the next bet".*

Do not fight a battle that is already lost, "if they go left, you should go right"

Since Japan Inc. owned 92% of the DRAM market in 1985 it was not likely for Intel to ever generate enough net income to become competitive in this market again, and so it was decided to forever stay out of this market and search for a new one. At the time, Intel was producing the lion share of microprocessors for the PC business, even though revenue was not very high. Intel took a gamble by bootstrapping itself in this market hoping to reach higher earnings levels. It became clear that the MPU business was the only risky but viable way on the road to redemption. In 1987 Intel was ranked at the 10th place in semiconductor revenue. The only avenue to success for a company that does not have a sufficiently large reserve of money to establish a new business consists in being fast and nibble and accumulate enough money to fund future technologies and products before anybody else could realize what was happening. Paraphrasing a statement suggested by the Art of War: ***"if they go left, you go right",*** Intel had made the right decision indeed. Technology and products acceleration was decided as the road to success in 1989 at the time when the company was introducing the 80486 on a 1-micron technology that was by then 4 years behind Japan Inc. To accomplish this goal, it was necessary for the manufacturing organization to absorb the technology transferred from the development group without making any changes, which would save one year out of the cycle. Under this strategy, technology introduction milestones were accelerated to a 2-year cycle, but most of all the introduction of new microprocessor products shifted from 4-year to a 2- and then 1-year pace or even less (Figure 5).

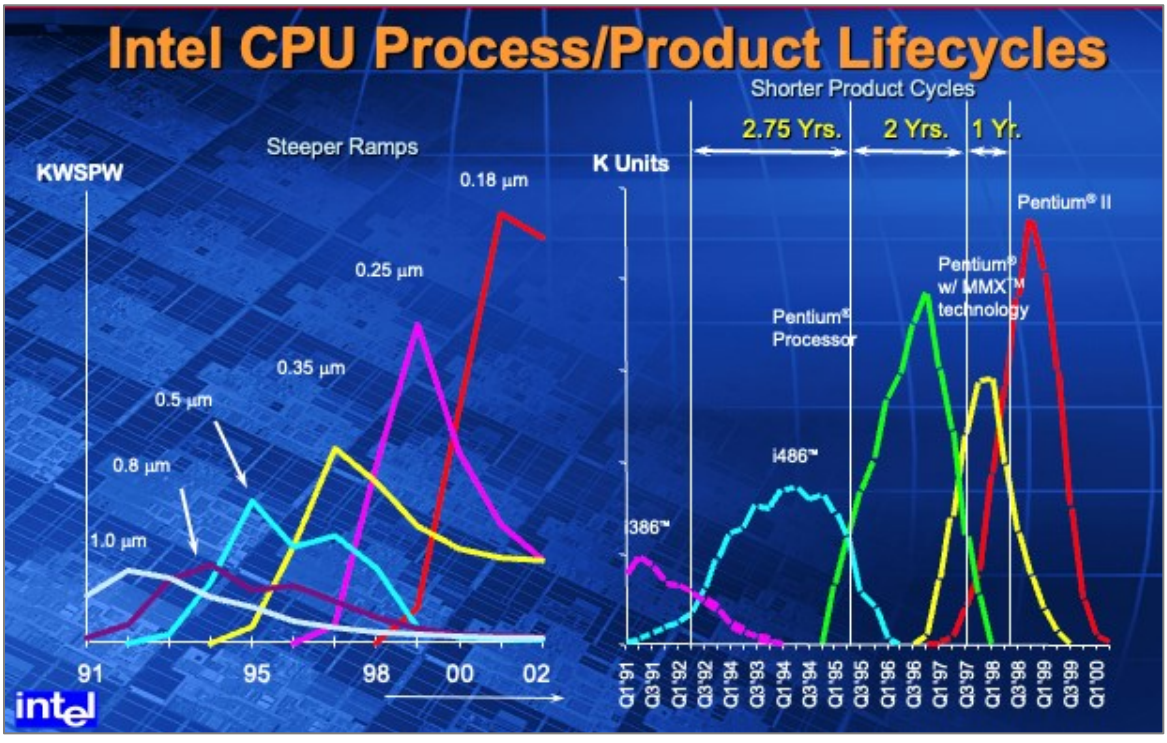


Figure 5 Intel accelerated introduction cycle of new MPUs from 4 years in the 80s to 1 year.

This strategy became known under the phrase of "copying exactly;" This strategy came at the price of boxing the Intel wafer cost in a substantially higher cost bracket than any other company since cost reduction once the technology transferred to manufacturing could no longer be made in the interest of time. However, the revenue created by the microprocessor business was more than sufficient to bear this unusually high wafer cost.

Intel Rose from the Ashes of DRAMs

By 1993, much to everyone's surprise, Intel surpassed NEC in revenue as the number 1 semiconductor company in the world (Table 2). By 1996, as Intel continued to rank number 1, it became clear that Intel had done something that had not been very well communicated or understood on how it managed to emerge as the revenue leader rising from the ashes of the DRAM demise. Finally, in April 1996, as the newly promoted Director of Technology Strategy, I got the honor of explaining for the first time to the analysts what had happened since 1989 leading to Intel success. It should have been evident to everybody that the accelerated product cycle was the secret of success and not the 2-year technology cycle, since Intel had been and remained a laggard in technology introduction throughout the 90s.

But was the story fully understood then (or now)? A revenue leader should not imply "technology leader."

Table 2 1993 Top 10 companies ranking in revenue: Intel as #1

1993			
		Sales	
Rank	Company	($B)	Share
1	Intel	$7.6	9.2%
2	NEC	$7.1	8.6%
3	Toshiba	$6.3	7.6%
4	Motorola	$5.8	7.0%
5	Hitachi	$5.2	6.3%
6	TI	$4.0	4.8%
7	Samsung	$3.1	3.8%
8	Mitsubishi	$3.0	3.6%
9	Fujitsu	$2.9	3.5%
10	Matsushita	$2.3	2.8%
Top 10 Total ($B)		$47.2	57.2%
Semi Market ($B)		$82.6	100%

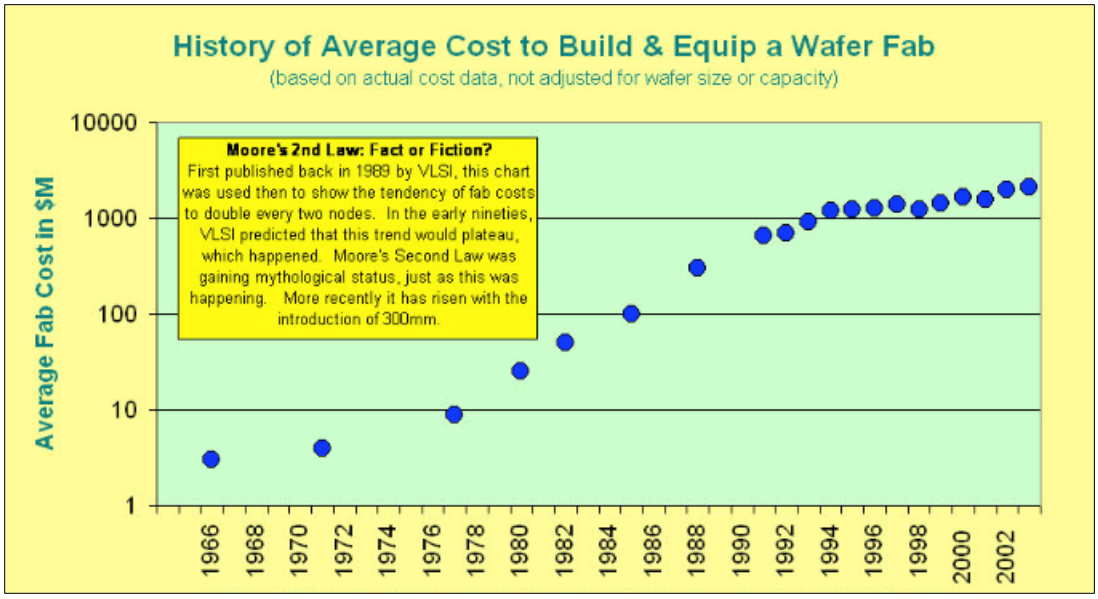
Source: IC Insights

3. 1997: THE END OF THE MOS SILICON GATE—CRISIS #3

To increase the number of transistors per die and the simultaneously increasing transistor performance, the semiconductor industry had merged these two fundamental drivers under a single motto: Scaling! By decreasing transistor dimensions by 30% per side the area of the transistor was reduced by 50% generation after generation. This nominally accomplished a doubling of the number of transistors every 2 years. By reducing the length of the transistor channel by 30% or even 40% the transistor speed was proportionally increased. As simple as that! "***Smaller, Faster, Cheaper!"*** However, many challenges were still looming ahead and were bound to upset this magic balance.

The Demise of the $1B Club

By the year 2000, multiple companies hovering around the $1B-plus revenue level were quite profitable. At the time the industry operated with 200 mm wafers. In the 90s the cost of a fully equipped manufacturing plant was around the $600-800M level and the members of the $1B club could afford to build and progressively outfit a factory at this cost over a 5 to 6-year span. All the sudden 300 mm wafer size conversions and the need to supply cost-effective products to the customers required new technological capabilities and higher capacity levels. In the 2002-2003 timeframe; the complete cost of refurbishing or building and outfitting a new factory with brand new 300 mm equipment moved the total cost of this transformation well over the $2-3B range (Figure 6). It quickly became clear that this level of investment was no longer affordable for the members of the $1B club and something needed to be done. Successful products were already in manufacturing at these companies on 200 mm wafers and the next generation of products had already been designed but they were waiting for a viable manufacturing plant.



Source: VLSI, 2003

Figure 6 Cost of a new factory escalated beyond the $1B mark in the 300 mm era.

Fabless and Foundries: a Match made in Heaven!

The $1B club soon realized that they needed a manufacturing solution outside the financial limits of their companies. It did not take long before the match between the soon to be fabless (at least at the 300 mm level) and foundries was celebrated. However, the foundries had been supported from the very beginning by the leading individual device manufacturers (IDMs) to unload their older technologies, which allowed them to reconfigure their factories for more advanced technologies. In essence and at that time, the foundries did not have the technology that was needed to support the designs of the $1B club members. In the end, business motivations always point out what is needed and reality wins: ***"The $1B club members transferred their teams and leading technology knowledge to support and accelerate the foundries manufacturing capabilities to produce their newly designed products".***

System Integrators finally Broke Loose from the IDMs!

Companies like Apple and Qualcomm realized early in the second half of the first decade of the 21st century that they did not need to associate themselves with the fortunes of the IDMs in order realize their products any longer; they could fare quite well with what the foundries could readily offer by then (perhaps not the very latest technology but good enough for their purpose). And so, the smart phone generation was born in this way.

"Verily I say to you, that no prophet is acceptable in his own country"

In 1971, Professor Carven Mead at Caltech had calculated that by the time the thickness of the gate dielectric was reaching the 5 nm value the amount of charge tunneling through it from the transistor channel to the upper gate electrode would have made the amount of leakage current too high and would have impaired many practical applications. Once again, this was a well-known fact and, once again, people decided not to pay attention to this. Furthermore, reductions in applied voltage and other technical workarounds were allowing to stretch the forecasted limit below the 5 nm level but then another problem was aggravating the situation—by 2005 the thickness of the gate dielectric would vanish (Figure 7).

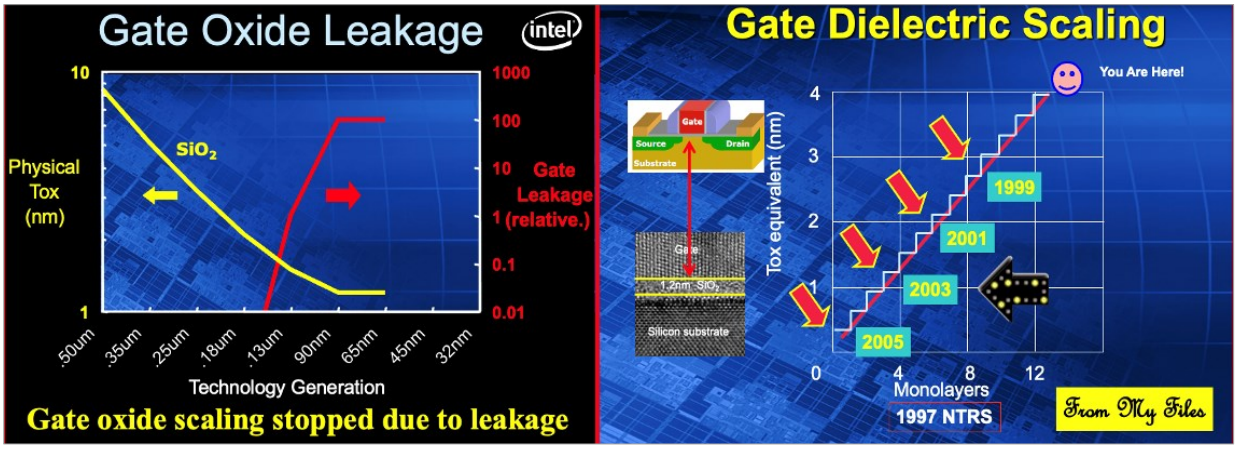


Figure 7 The end is near: gate leakage escalating and gate thickness vanishing beyond the year 2000.

Gordon Moore, who had received his PhD from Caltech, strongly believed this prediction and by 1991 began asking the question: ***"What is the industry going to do about this?"*** But nobody was willing to take any action. Paraphrasing what Machiavelli had said a long time ago in 1513, "***There is no problem more difficult to solve than challenging an existing system since those that benefit from it will be against any changes and those that may benefit from the changes will be lukewarm supporters at best."***

Once Again, Governments Stepped In

In 1998, the International Technology Roadmap for Semiconductors (ITRS), with participation from Europe, Japan, Korea, Taiwan, and the US was formed, and it was then possible to finally mobilize all the resources necessary to deal with the magnitude of the problem (Figure 8).

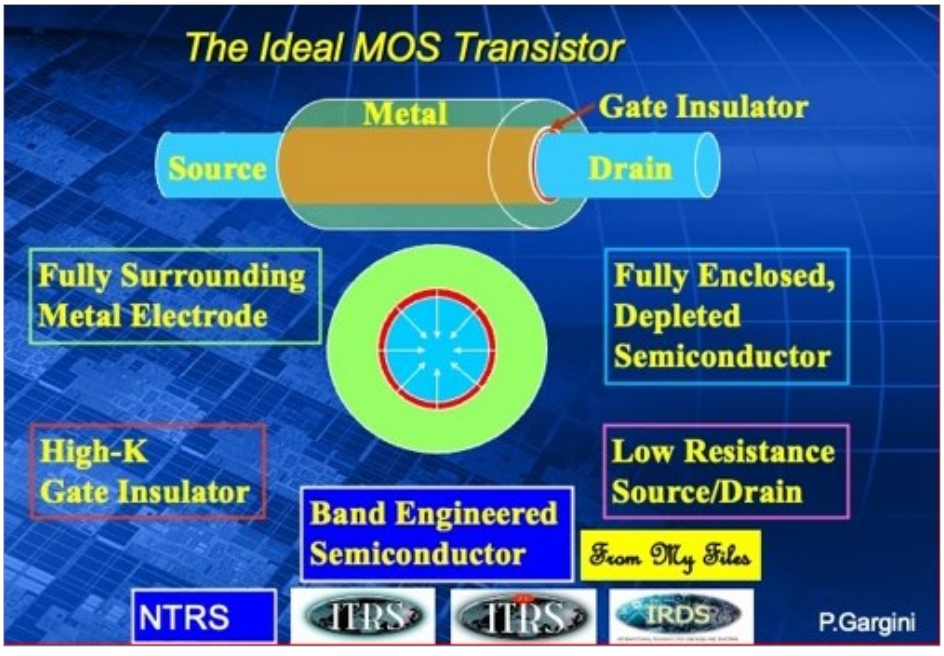
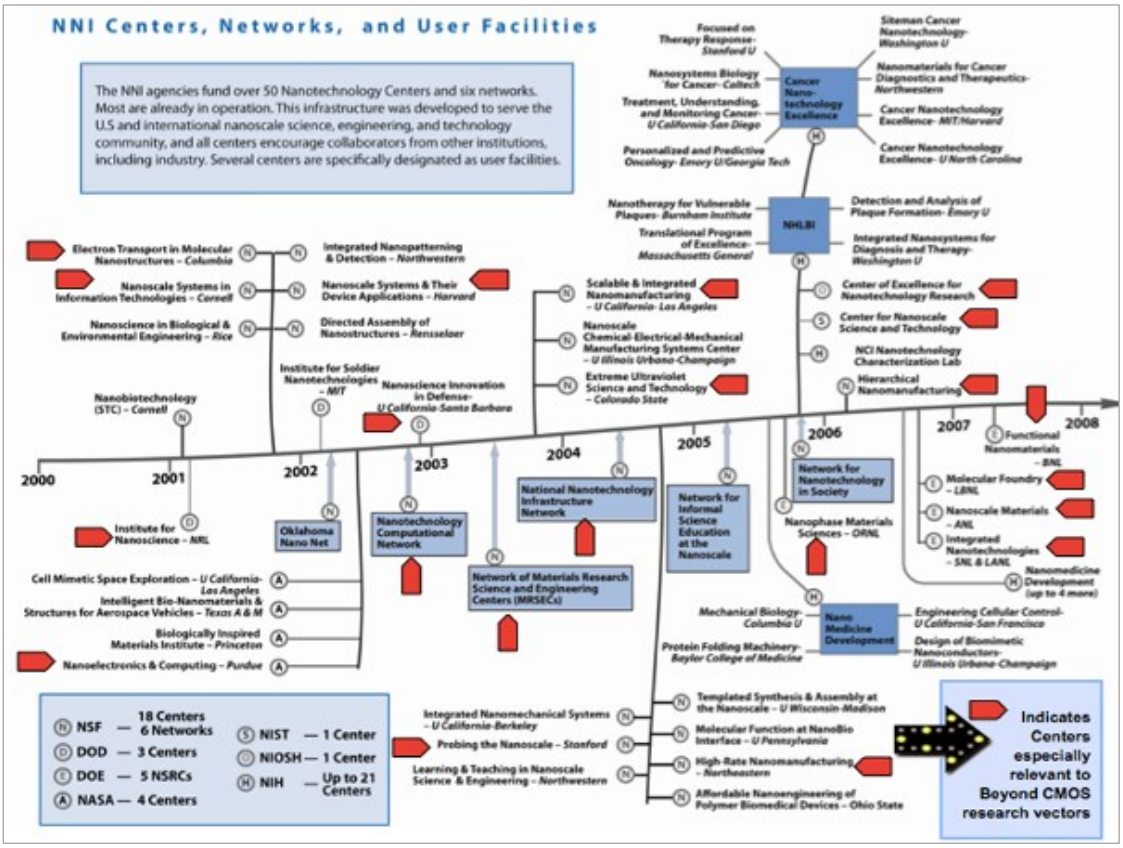


Figure 8 ITRS outlined the goals to refurbish the classical CMOS process.

In the year 2000, the US government announced major funding in technology focused on the theme of **"*Nanotechnology.*"** Research funds of over $400M that were allocated in 2001 by the USG were soon followed by similar levels of investments in research by Japan, Europe, as well as all other countries producing semiconductors. The total worldwide funding soon accumulated within the $1B range! The initial goal was often over aggressively

stating that "***The transistor will build itself from the bottom-up.***" By 2005 a compromise between industry and governments had been reached calling for a more balanced investment split between a bottoms-up and tops-down approach (Figure 9).



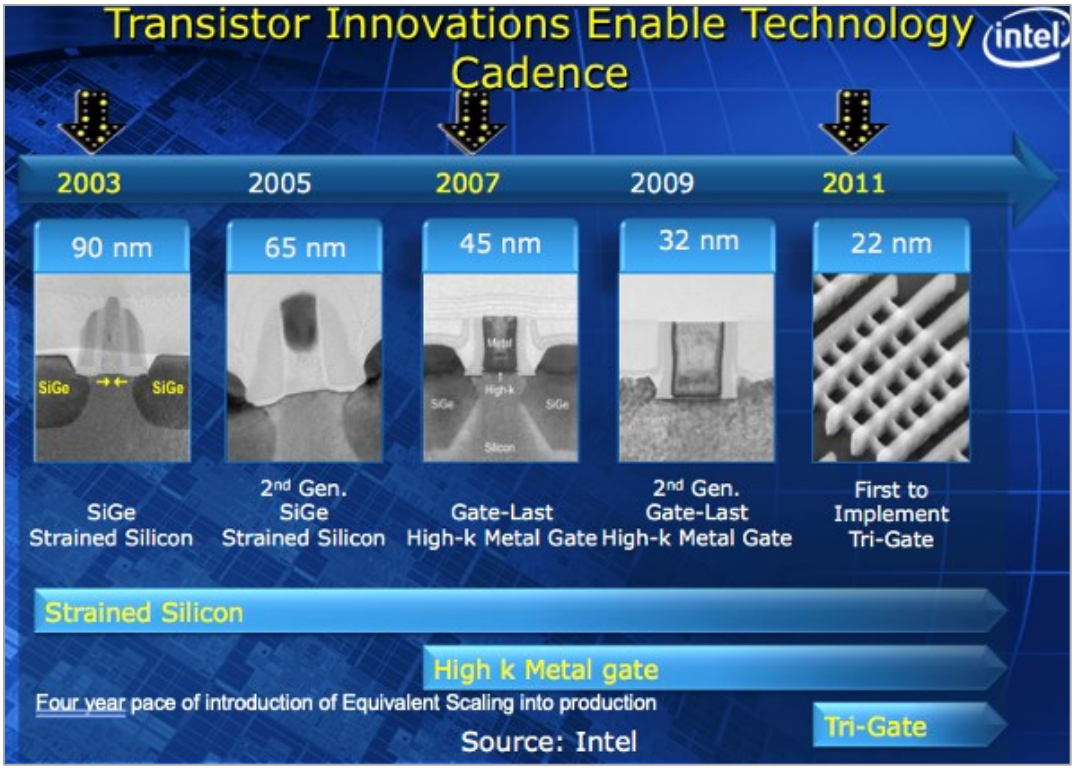
Source: National Science Foundation (NSF)

Figure 9 Nanotechnology programs supported refurbishment of CMOS process and transistor structure.

University research typically precedes industry needs by at least 10 years. However, most of the time the information goes unnoticed since the industry's attention is concentrated on the 3-5 years span driven by economically motivated goals. It is therefore essential to systematically review publications of the past to find useful clues on how to solve problems arising on the not-too-distant future. This was indeed the case for the problems outlined by the ITRS. In fact strained silicon[1] and FinFET[2 3 4] solutions had been indeed proposed long before they were needed.

Subsequent research pointed the way to viable solutions but when it came time to test them out, unexpectedly Sematech declined to make prototypes.

Demonstrating the viability of the solutions came via IMEC in Leuven, Belgium—a practically unknown Flemish organization that, in the end, proved the feasibility of viable solutions and sequentially strained silicon, high-k/metal-gate and FinFET went into manufacturing in 2003, 2007 and 2011, respectively. By the end of 2011, Intel had finally become for the first time not only a revenue leader but also a technology and manufacturing leader (Figure 10).

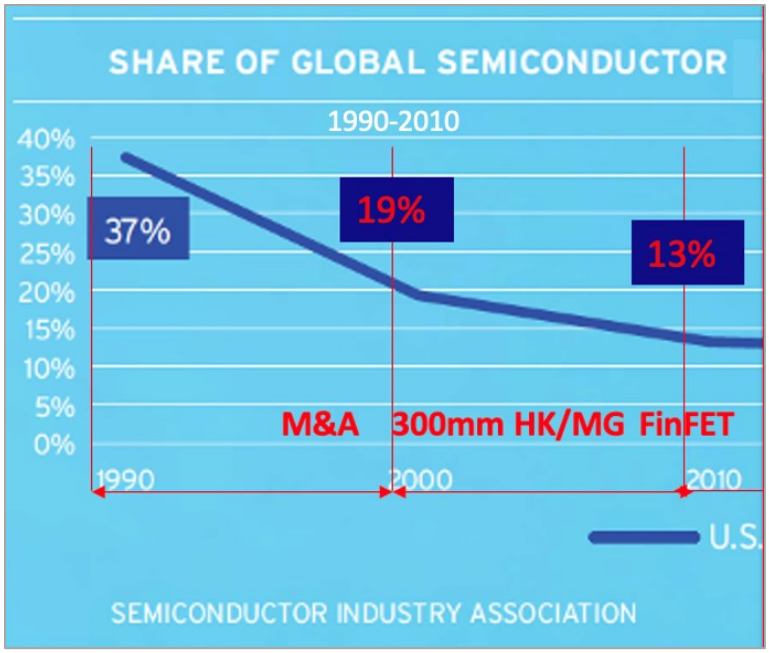


Source: Intel

Figure 10 First phase of refurbishment of CMOS device and process completed by 2011.

Finally, Financial Leadership and Technical Leadership proceed Hand in Hand

These technical solutions introduced between 2003 to 2011 into manufacturing saved the semiconductor industry and concurrently the conversion from 200 mm wafers to 300 mm wafers assured the cost effectiveness of the semiconductor business. Several companies continued designing their products, but progressively more and more companies were relying on wafer foundries to manufacture their wafers and eventually became fabless. This "Foundry/Fabless" combination was a great technological and economical solution. However, as a result the wafer volume kept on shifting and increasing outside Europe, Japan, as well as within the US, who once had been the leaders in semiconductor manufacturing. It was evident in ***2010 that only 13% of all wafers manufactured in the world were produced in the US (versus 37% in 1990). Once again, nobody paid any attention to this fact (Figure 11).***



Source: Semiconductor Industry Association, 2010.

Figure 11 The demise of the $1B club fostered migration of wafer manufacturing outside the US.

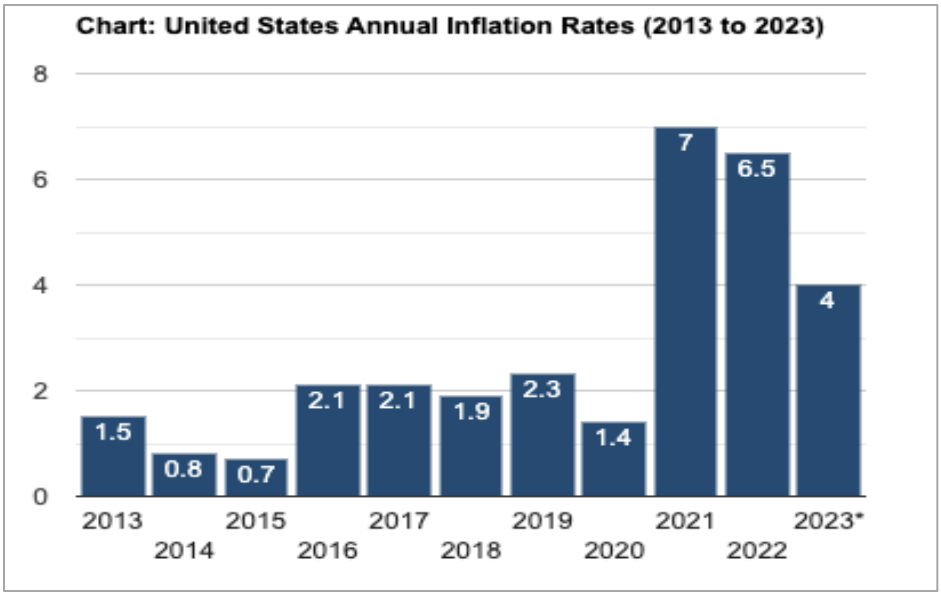
4. 2020: The Pandemic Influence—Crisis #4

By 2020, the Foundry/Fabless industry model—in conjunction with a handful of semiconductor giants dominating the memory, microprocessors and analog markets—had been solidly established for over 10 years as the foundational model of the new semiconductor industry. Everybody in the electronics industry was benefiting from this new Fabless/Foundry operational mode. Companies like Apple, Google, Facebook, Qualcomm, NVIDIA and Broadcom—just to mention a few—that had never been in semiconductor manufacturing on their own had become extremely successful at designing their own chips and having them produced outside the US. Still, nobody paid attention to this ….And then the pandemic came!

Societies around the world realized that semiconductors had penetrated every aspect of life, and nothing could really work without them.

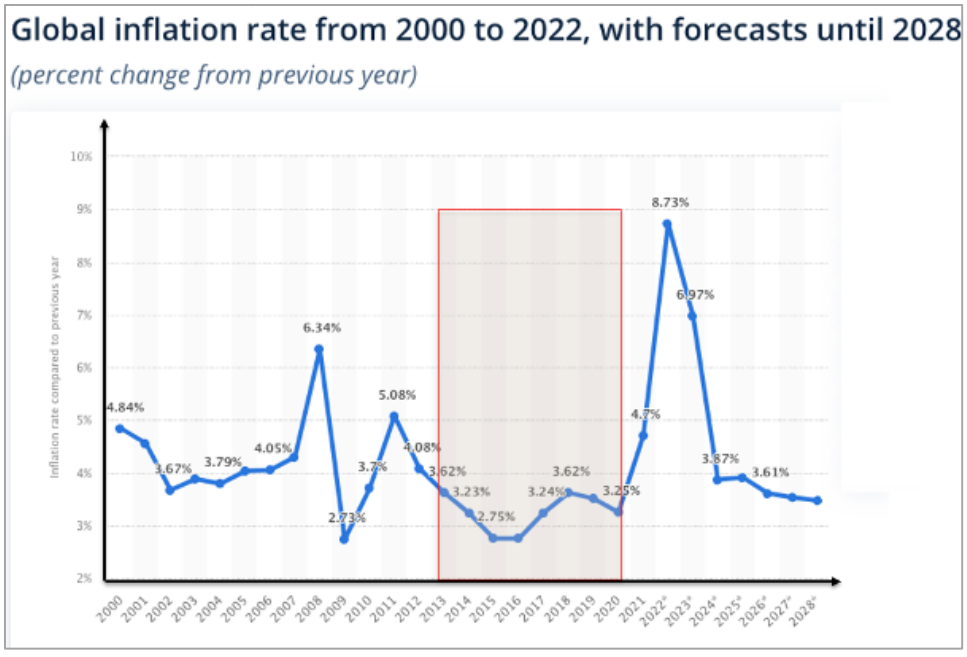
As a preventive approach (and as the only measure known at the time against the spreading of the pandemic) many nations decided to slow down or shut down most of their public activities. Individuals isolated themselves to protect from the contagion while scientists searched for and developed vaccines.

This resulted in a catastrophic disruption in the supply chains. Governments finally realized in horror what had happened for over 10 years—most, if not all, supply chains had been fully ***globalized.*** This meant that goods and services were provided from sources around the world in locations where the associated costs were lower than in Europe, Japan, and the US as well as in any of many other industrialized nations. Additionally, the profit margins made by the oversees producers were perfectly adequate for the ecosystem within which they operated, and cost of goods and services were very affordable for the beneficiary nations so everybody was benefitting from globalization. ***No wonder inflation was not a big problem for the past 8 years in the US and around the world! (See Figures 12 and 13.)***



Source: Statista

Figure 12 US inflation was under control for eight years before pandemic.



Source: Statista

Figure 13 Inflation was under control for eight years around the world before pandemic.

The initial, almost hysterical, reaction to the disruption of most supply chains in most nations led to proposals aimed at repatriating a whole spectrum of products and services. Even though this reaction may have appeared as a perfectly natural solution at the time, the repatriation proposal was only the result of a cursory evaluation and as time went by it is become clear that the amount on money and time required to execute any repatriation were completely out of reach. Furthermore, the cost of goods and services emerging out of this repatriation process would have been completely unbearable for consumers.

Nevertheless, like it happened in the previous crises, by the time this realization sank in many governments had already made their announcements and had initiated the budgeting processes aimed at funding some forms of repatriation. ***As in the previous crises the task at hand consisted in appropriately and politely redirecting these investments*** to address manageable solutions, thus making the best realistic use of these funds since these massive infusions of well-motivated massive government money occur only every 15-20 years. ***In essence, the present task consists in spending this money wisely to support the electronics and semiconductor industries and other related industries for the next decades.***

Summary of All Past Crises

Before starting to analyze what kind of actions should be taken to solve the 4th crisis it is beneficial to summarize the main events of the previous ones

1. We learned from the ***first crisis*** how the US government funded development of semiconductors beginning with 1958 and created the semiconductor industry. Between 1965 and 1975 Silicon Valley had essentially become self-supported. This implies that companies could make enough money to fund the next technology generation on their own and government funding was no longer essential.
2. The ***second crisis*** showed that products and not technologies lead to financial leadership. Intel was a technology laggard though the 90s but dominated the decade in revenue.
3. In the ***third crisis*** a well-organized ITRS, supported by the WSC, clearly pointed the way to extending the life of CMOS devices almost indefinitely. Government funded research in new technologies and most of all industry funded prototyping performed at IMEC, which extended the life of the semiconductor industry for decades to come. All the technical goals proposed by ITRS in 1998 were achieved. The Fabless/Foundry model proved that, even more so than in the past, ***products supported by adequate technologies are the key to making the economic success of electronic companies.***
4. By 2011 the semiconductor industry, after the innovative restructuring occurred in the previous decade, had once again fallen in a well understandable evolutionary mode and some people felt that the ITRS was no longer required to guide the electronics industry. The industry drivers had evolved from equipment capabilities to semiconductor technologies and was now in a period in which products designed by system integrators were produced outside Europe, Japan, and the US. A new age of the electronics industry was born where new drivers demanded new approaches and required improved and broader guidance. Thus, the International Roadmap for Devices and System (IRDS) was born in 2016 under IEEE sponsorship with support from SDRJ from Japan and from SiNANO from Europe. The IRDS has already mapped the system and technology directions for the next 15 years and now ***execution is the key to success*** like it was 25 years ago with the ITRS. And, like then, large amounts of governments' fundings are becoming available around the world. ***Would history really repeat itself?***

The pandemic has created an unprecedented challenge that triggered availability of unprecedented amounts of funds under the "Chips Acts" label. Governments, industry, and academia are all proposing solutions.

How can we find some clues on how to successfully solve the fourth crisis?

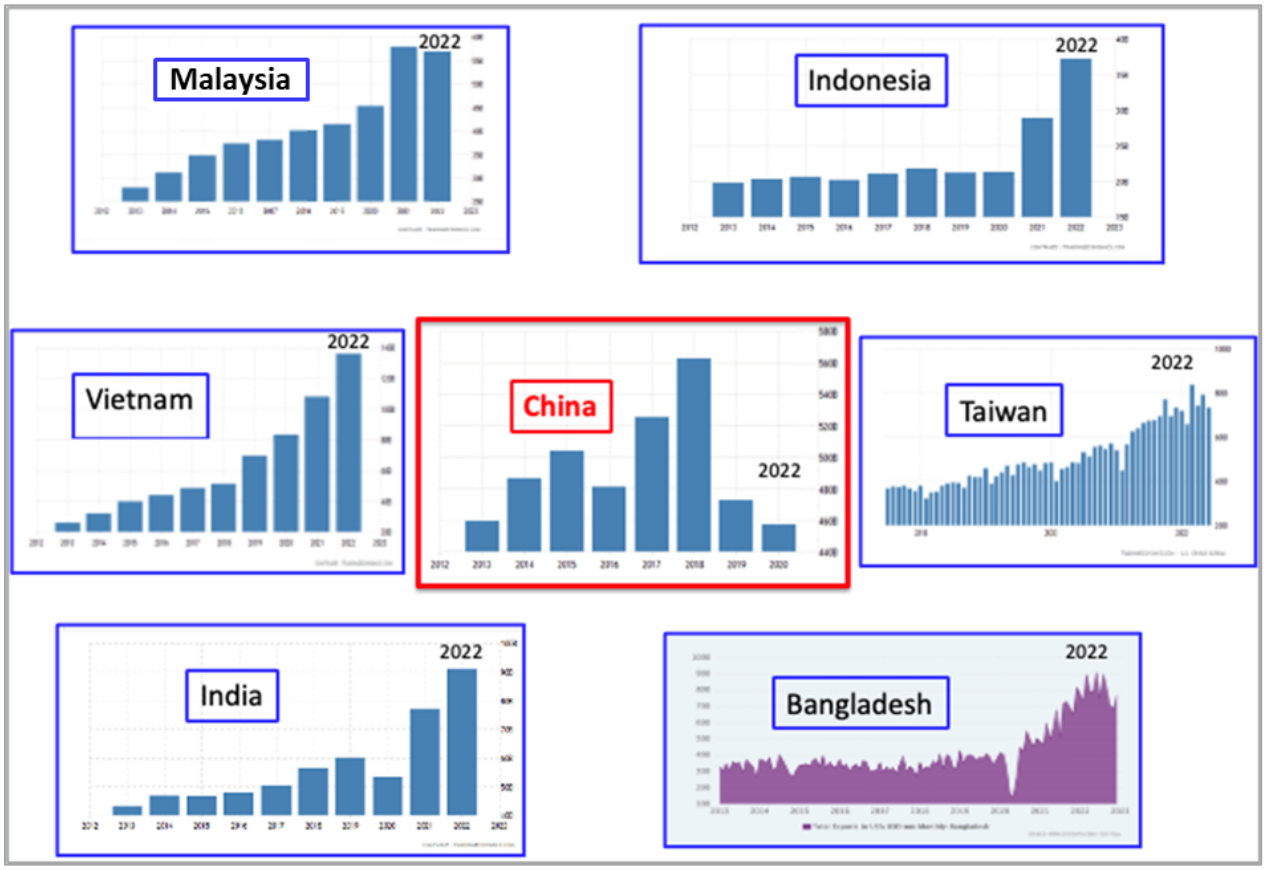
5. ADJUSTING TO THE REALITY OF THE 4TH CRISIS

Long live globalization!

The pandemic crisis induced most nations to adopt a negative attitude towards globalization and led them to assuming that the disruption in the supply chains could have been mitigated or even avoided if each nation owned a completely self-sufficient multitude of domestic manufacturing lines producing almost any items. ***This is a practical impossibility in the present economic era*** since many ingredients that are fundamental for several supply chains are almost inexistent in some nations and must be imported. Furthermore, global production of many items had been optimized in the most cost-effective way for over two decades and so most nations have been lacking most of the fundamental building blocks of many supply chains for many years. It is true that China, as a supporter of several critical supply chains, has been substantially affected by the pandemic, but system integrators had already set up a web of potential emergency contingency plans that were activated and imports levels from alternative suppliers were expeditiously raised up to unprecedented levels in a year. This behavior corresponds to a well-known business practice called: ***"multi***

source supplier strategy" *(Figure 14)*. Nowadays, after 3 years the economic effects of the pandemic are almost invisible. As a result of the unexpected competition created by the many nations willing to step into the supply chain, China is now coming back and imports to the US are now exceeding historical trends.

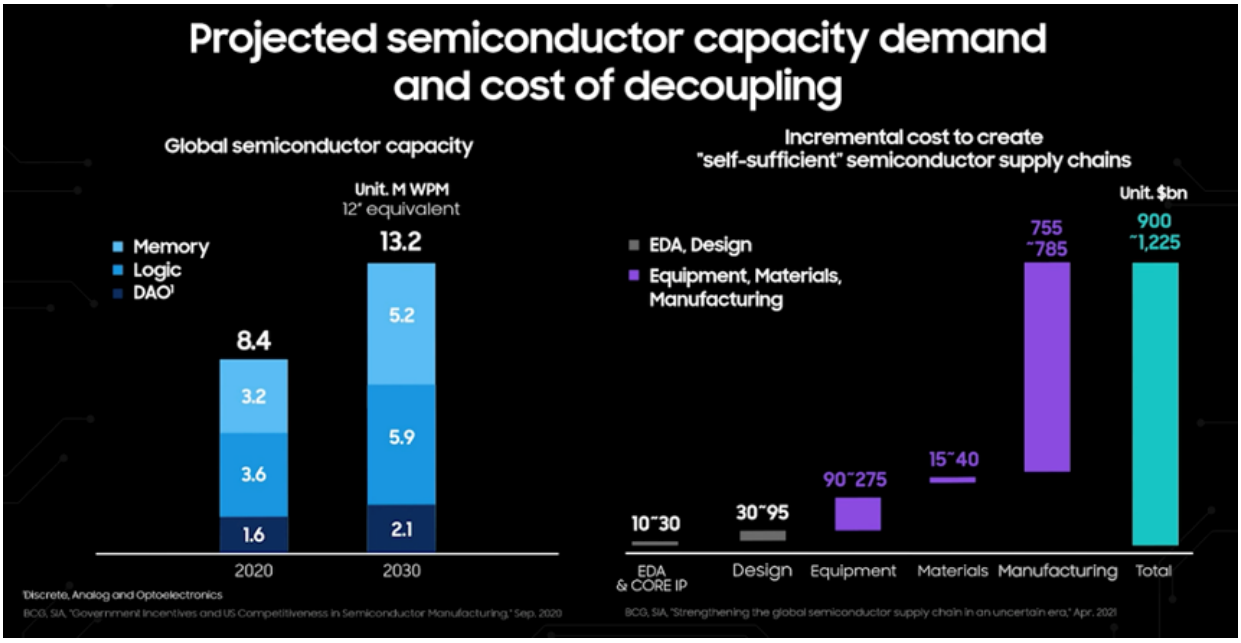
Will the benefits of this refurbished and extended supply chain keep by itself inflation close to historical levels?



Sources: TradingEconomics.com, US Bureau of Economic Analysis (BEA)

Figure 14 Imports to the US from multiple countries escalated as China imports dwindled down.

Looking back at the past 3 years, we may ask if the economic impact of the pandemic was nothing more than a glitch (even though a big one) in the globalized supply chain system, much like when the electricity goes out for an extended amount of time or when an unpredicted flood occurs, or even when an earthquake hits a region. People normally suffer the consequences for a while, but they always rebuild after any disastrous events and inevitably, having learned the lesson, take steps for resiliency and as insurance against the next event. This realization puts us in the position of formulating practical recovery and insurance plans for the electronics industry. Let's keep in mind that, as an example, the cost of just duplicating semiconductor wafer capacity by repatriating would easily exceed $1T! ***Who has that kind of money available on the sideline?*** (See Figure 15.)



Source: Kinam Kim, Chairman Samsung Electronics, Monday December 14, 2021, IEDM Plenary

Figure 15 Over $1T investments would be required to repatriate the global semiconductor capacity.

Buying Manufacturing Insurance

The electronics industry landscape is constituted by the Fabless/Foundry that design and produce the full supply chain spanning from design to wafer manufacturing to assembling to manufacturing final products in parallel with few IDM giant companies dominating the memory, logic and analog business; these are the backbones of the electronics industry. Any strategy addressing semiconductor manufacturing must take this fact into account and devise strategies that include this reality. Upsetting this model would have severe implications for the economics of many nations. ***This implies that it is not realistic to repatriate a substantial portion of domestic semiconductor manufacturing in either Europe, Japan, or the US in any reasonable amount of time with the expectation of utilizing exclusively domestic suppliers.*** However, similar to buying insurance (e.g., earthquake, flooding, severe weather events, etc.) for individuals and local businesses to alleviate but not eliminate the effect of any future shortages, the only viable and timely insurance for the semiconductor industry consists in inviting the semiconductor manufacturing giants to locally establish or increase the number of manufacturing plants in Europe, Japan, US, and China.

The Insurance Tickets

- In ***Europe***—Intel and GlobalFoundries have large local manufacturing facilities and both companies are now planning of building additional wafer and assembly plants in Europe. TSMC is building its first chip plant in Europe in a joint venture with European companies.
- In ***Japan***—TSMC is building a plant in Kumamoto prefecture and has established a 3DIC research facility in Tsukuba.
- In the ***US*** —TSMC is building 2 wafer plants in Arizona and Samsung is building one wafer plant in Texas. Intel and Micron Technologies are also building additional wafer plants in the US.
- In ***China*** —TSMC is building a second wafer factory in Nanjing
- In ***Taiwan*** —TSMC is building additional 3 wafer and 2 packaging manufacturing plants

These are the insurance tickets! Like any other form of insurance, it is good to have them while hoping that they are never to be used.

The Innovation Opportunity

There is also a second and very relevant point that needs to be made. Typically, when a large influx of government funds become available for any reason, it represents a unique opportunity to make the best of it by addressing real fundamental problems that may have become apparent, but nobody has wanted to focus on them because they do not represent an imminent business threat. Does this sound familiar? ***Is the situation of the alarm sounded by 1998 ITRS repeating again with the alarm sounded by the IRDS?***

Again, we can learning from the past. The first and third crises can be characterized as the two largest transformational inflection points for the semiconductor industry supported by large government fundings.

In the ***first crisis*** the industry started with manufacturing individual transistors avidly acquired by the USG and emerged from the crisis with an extensive menu of products realized with fully integrated circuits (IC) made of millions of interconnected transistors on a single die.

In the ***third crisis*** the semiconductor industry was facing multiple "Red Brick Walls" looming around the beginning of the new century and identified by the ITRS. These issues were addressed with the funds of the Nanotechnology initiatives provided by multiple governments. The electronics industry successfully emerged with a fully refurbished CMOS process and a multitude of related technologies with enabling new capabilities (e.g., MEMs, sensors, LEDs, imagers, etc.) lead to novel products that will continue to extend the menu of the electronics industry via the semiconductor industry for decades to come.

Governments come to the Rescue during Inflection Points.

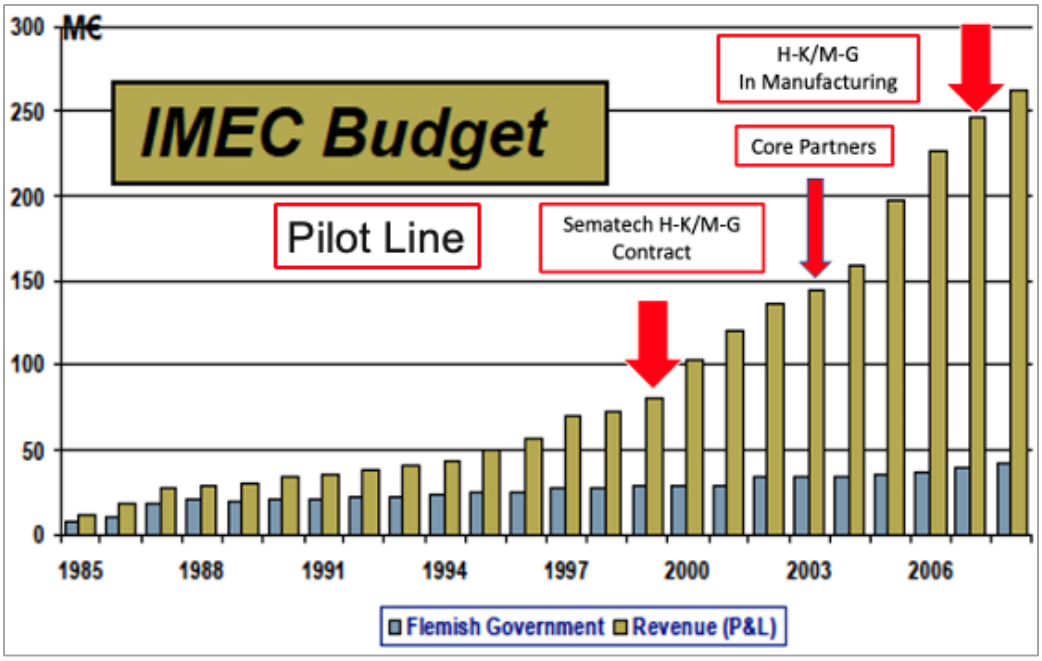
In both crises, massive government investments augmented industry investments that were looking for a real problem to be solved. Inflection points can represent therefore an opportunity for ***new, untested technologies and products*** to be tried, developed, and then introduced into manufacturing. This can occur without being subjected to all the financial constraints and limitations faced by new start-ups trying to mature and compete with new high-risk innovative approaches against the well-established ecosystem of existing electronics and semiconductor giants that dominate most sectors of the electronics and semiconductor business today. Also, in this environment, unproven and risky technologies and products can be evaluated without any restraints or restrictions. Thus, governments become the best friendly ***venture capitalists.*** This is what we are going to analyze next as well as provide a practical roadmap to success.

The Model already Exists!

During the ***third crisis,*** the blueprint of a well-organized model was laid out followed by a well-executed implementation plan on how to test, develop and introduce into manufacturing many new high-risk technologies. The first step of this plan consisted with a well-thought-out roadmap (ITRS) laying out the challenges and the possible solutions to crash any of the Red Brick Walls looming ahead. The ITRS—with the full support of the World Semiconductor Council (WSC)—offered the opportunity for the first time to synchronize research around the world aimed at solving the fundamental problems limiting the future of the semiconductor industry in a timely fashion. Sequentially, copper interconnects, strained silicon, high-k/metal-gate and FinFET technological and structural innovations were introduced into manufacturing as outlined and scheduled by the ITRS. As indicated before (i.e., Machiavelli) introduction of new technologies remains primarily an organizational challenge since none of the well-stablished companies want to take the risks of evaluating them. That is the time when governments can play a big role.

Imec to the Rescue

In the year 2000, when Sematech refused to prototype the above technologies, it was possible to convince Gilbert Declerck, CEO of an almost unknown university facility named IMEC to rise to the task. Ironically, it was also possible to convince Sematech to pay $12M for a 3-year prototyping contract to IMEC for the demonstration of high-k/metal-gate on manufacturing equipment. By 2003, IMEC stood on their own forming the Core Partners Program (CPP), supported with a brand new 300 mm clean room. This program was highly successful and while IMEC budget stood at $75M in 1999, it reached over $800M in 2022 (Figure 16).



Source: IMEC

Figure 16 IMEC budget escalated as refurbishment of CMOS process and structures continued

As mentioned, systematic reviews of past publications can result with useful clues on how to solve problems arising in the not-too-distant future. This is because it is typical for university research to precede industry needs by 10 or more years. Most of the time such information goes unnoticed since the industry's attention is concentrated on the 3-5 years span driven by economically motivated goals. Such was the case with problems and solutions outlined in the ITRS; e.g., strained silicon and FinFET were proposed solutions in the roadmap long before they were needed.

Let's be clear, research was done everywhere in the world, prototyping was done at IMEC and then the proven technologies were transferred to the industry for optimization and introduced into high-volume manufacturing. At that time, ***everybody had a role to play. This is the case today. So, let's then assign roles and responsibilities now!***

6. OUTLINING THE FUTURE PATH WITH THE IRDS

The ITRS

The ITRS was instrumental in guiding the industry in refurbishing the silicon gate process and promoting new transistor structures for over 15 years, but by the middle of the second decade of the 21st century, the drivers of the electronics industry were no longer solely limited to supporting the evolution of traditional technologies and the need for new innovative technologies became necessary.

The IRDS

In 2016, the International Roadmap for Devices and Systems (IRDS) was formed as an evolution of the ITRS broadening the industry coverage by addressing new drivers of the electronics industry beyond technology by including ***devices and systems*** (https://irds.ieee.org/editions). This roadmap, sponsored by IEEE, the SDRJ and SiNANO is not a complete departure from the ITRS, but it is a ***necessary*** evolution and an ***extension*** of the ITRS. In fact, a lot of the original ITRS working groups members (now operating as the International Focus Teams, or IFTs) continue today as well as working with new IFTs. In fulfilling its role, the IRDS has been anticipating upcoming challenges for the electronics industry and pointing to possible solutions. Therefore, let's put into prospective which next steps need to be taken by using full advantage of experiences accumulated from past crises and especially by studying the related solutions. The IRDS is recognized as a global guidance for the industry and the industry is listening—the IRDS passed the 1,000,000 visits in 2022 (Figure 17).

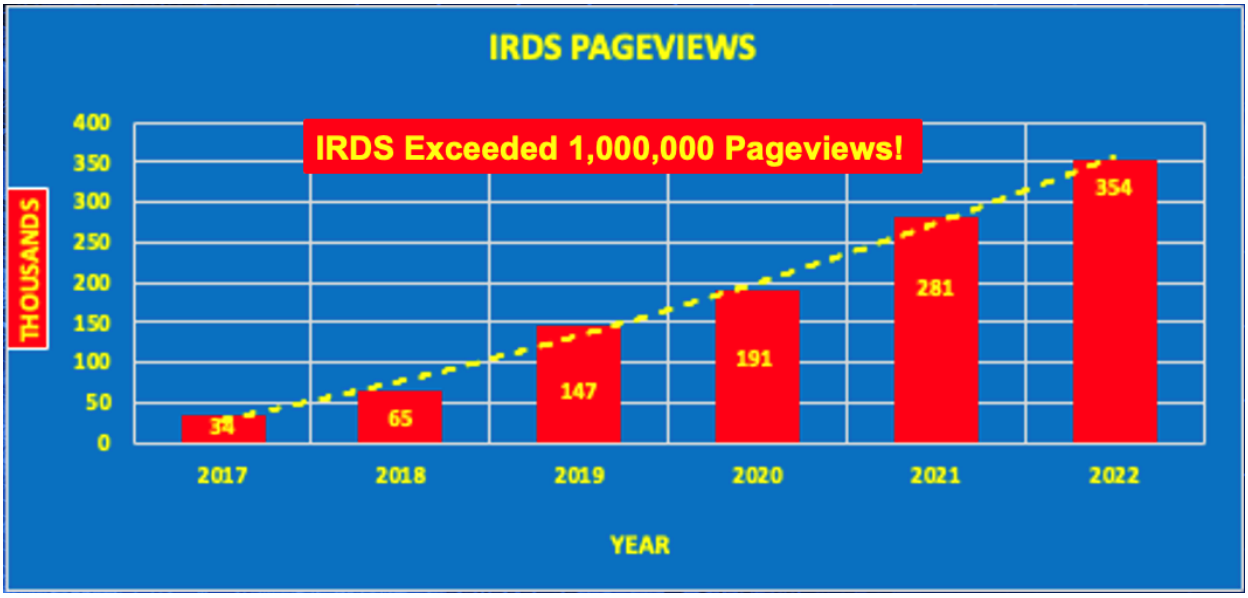


Figure 17 Page visits of IRDS escalated to pass the 1,000,000 marks in 2022.

The Myth of Leadership in Transistor Scaling

The definition of "technology node" was clearly articulated in 1994 (Figure 18). This is defined as the half-pitch of the tightest layer that is typically represented by an interconnect layer. This definition has, however, been largely abused by companies that opportunistically chose an alternative definition of their choice for their benefit. To remain consistent with the initial one and only definition of technology node (while the industry is talking of 5 nm, 3 nm and now 2 nm technologies) the IRDS highlights at the top of its technology assessment tables the technology nodes based on the correct 1994 definition (Table 3). Under these conditions the present, most aggressive technologies are in the ***12 nm technology*** node range with the 10 nm technology node expected to be introduced in the year 2025 (Figure 19).

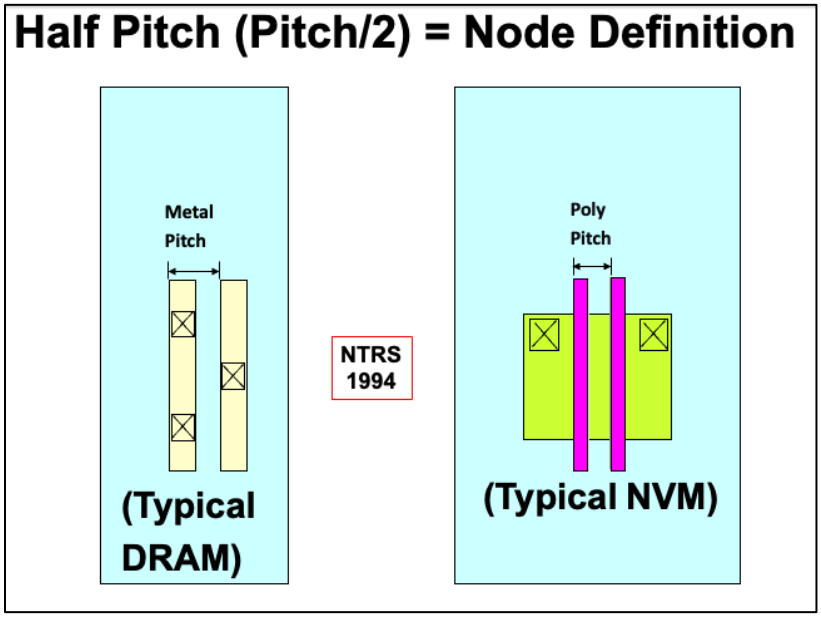
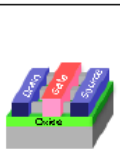
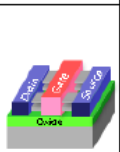
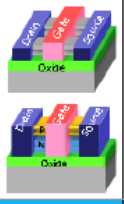
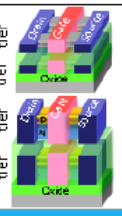
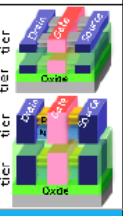
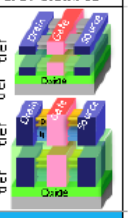


Figure 18 Original definition of technology node

Table 3 ITRS and IRDS continued to use the correct technology node definition.

YEAR OF PRODUCTION	2022	2025	2028	2031	2034	2037
	G48M24	G45M20	G42M16	G40M16/T2	G38M16/T4	G38M16/T6
Logic industry "Node Range" Labeling	"3nm"	"2nm"	"1.5nm"	"1.0nm eq"	"0.7nm eq"	"0.5nm eq"
Fine-pitch 3D integration scheme	Stacking	Stacking	Stacking	3DVLSI	3DVLSI	3DVLSI
Logic device structure options	finFET LGAA	LGAA	LGAA CFET-SRAM	LGAA-3D CFET-SRAM	LGAA-3D CFET-SRAM	LGAA-3D CFET-SRAM
Platform device for logic	finFET	LGAA	LGAA CFET-SRAM	LGAA-3D CFET-SRAM-3D	LGAA-3D CFET-SRAM-3D	LGAA-3D CFET-SRAM-3D
						
LOGIC DEVICE GROUND RULES						
Mx pitch (nm)	32	24	20	16	16	16
M1 pitch (nm)	32	23	21	20	19	19
M0 pitch (nm)	24	20	16	16	16	16
Gate pitch (nm)	48	45	42	40	38	38
Lg: Gate Length – HP (nm)	16	14	12	12	12	12
Lg: Gate Length – HD (nm)	18	14	12	12	12	12
Channel overlap ratio – two-sided	0.20	0.20	0.20	0.20	0.20	0.20
Spacer width (nm)	6	6	5	5	4	4
Spacer k value	3.5	3.3	3.0	3.0	2.7	2.7
Contact CD (nm) – finFET, LGAA	20	19	20	18	18	18
Device architecture key ground rules						
Device lateral pitch (nm)	24	26	24	24	23	23
Device height (nm)	48	52	48	64	60	55
FinFET Fin width (nm)	5.0					
Footprint drive efficiency – finFET	4.21					
Lateral GAA vertical pitch (nm)		18.0	16.0	16.0	15.0	14.0
Lateral GAA (nanosheet) thickness (nm)		6.0	6.0	6.0	5.0	4.0
Number of vertically stacked nanosheets on one device		3	3	4	4	4
LGAA width (nm) – HP		30	30	20	15	15
LGAA width (nm) – HD		15	10	10	6	6
LGAA Width (nm) – SRAM		7	6	6	6	6
Footprint drive efficiency – lateral GAA – HP		4.41	4.50	5.47	5.00	4.75
Device effective width (nm) – HP	101.0	216.0	216.0	208.0	160.0	152.0
Device effective width (nm) – HD	101.0	126.0	96.0	128.0	88.0	80.0
PN seperation width (nm)	45	40	20	15	15	10

Source: IEEE, More Moore, 2022 IRDS

This aside, feature scaling has been predominantly enabled by the capability of printing lines and spaces of progressively smaller dimensions by means of lithographic techniques. The introduction of EUV into manufacture in 2017 assures the viability of reaching 8 nm technology node around the end of this decade. Furthermore, by applying the double exposure techniques developed for 193 nm exposure technology and/or by utilizing exposure tools with NA of 0.5 or even higher, the next decade of scaling is completely under control.

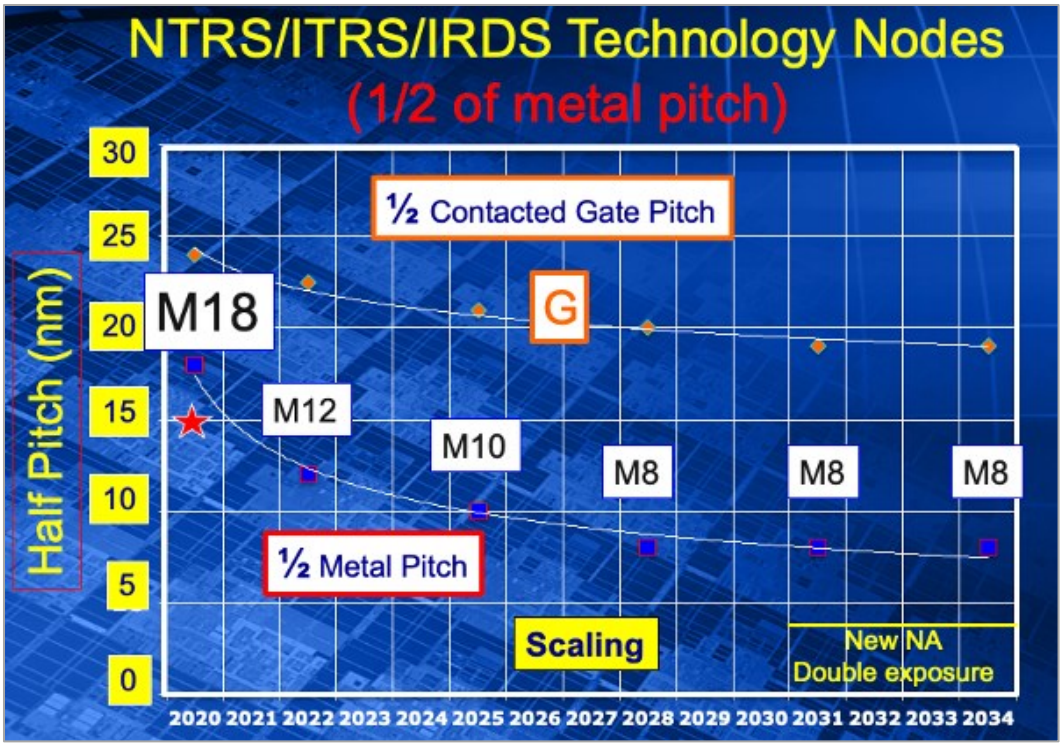


Figure 19 Double exposure and introduction of EUV tools with higher NA will support scaling beyond 2030.

This implies that government-sponsored research on this subject should not represent a major focus of attention. Suppliers and leading semiconductor manufacturers have this item well under control. In addition, the center of gravity of many widely successful consumer products is around the so called 28 nm technology node according to the industry definition. (This is closer to 45 nm of even higher if the correct technology node definition is used.) In conclusion, feature scaling should not be the center of attention for all the Chips Acts. However, access to leading-edge technology should be regarded only as means of enabling a convenient prototyping pilot line to carry on tests on new innovative and badly required technologies on manufacturing-worthy equipment.

If it is not scaling, then what are the performance and technology enablers?

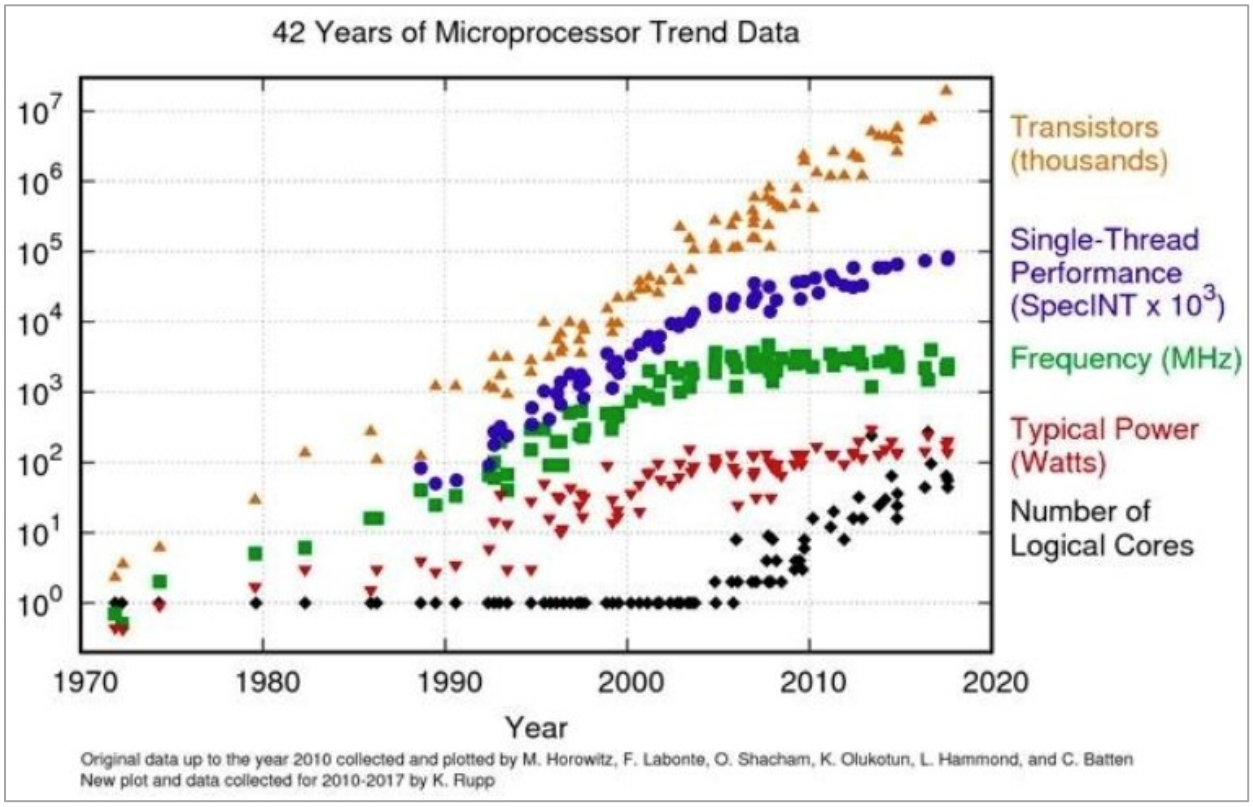
The Architectural Component

The MPU architecture provided the most uncommitted and flexible architecture capable of adjusting to almost any request. The operating system and the program instructions guided the MPU to provide the right answer to almost any product specifications. As time went by, the list of requirements became longer and more complex, as represented by the continuously increasing length of the necessary instructions. None the less, faster, and faster MPUs could make up for this increased complexity by higher speeds of execution. These execution speeds made this underlying process invisible for the users because lengthier instructions and faster MPU compensated each other and made execution time acceptable for the users. Increasing the number of interconnect layers (e.g., more than 17 layers are already in production) contributed also to improving performance by providing interconnects of limited length and thus minimizing signal propagation delays.

Nobody can fool Mother Nature

However, the laws of physics cannot be fooled forever, and dynamic power dissipation kept on increasing as time went by, as well as the die temperature. As predicted as far back as 1989, eventually an upper power limit was reached, and frequency could no longer be increased. Even though with each technology generation faster transistors were being manufactured; it was not possible to increase the speed of the MPU beyond 5-6 GHz. The culminating event occurred in 2003 when Intel introduced the Prescott MPU, that reached the physical limit of 115-130 W dynamic

power dissipation. When this occurred, scaling transistor channel to enhance transistor speed no longer represented a viable way of increasing MPU performance. Overall feature scaling was still highly beneficial to reduce transistor footprint and interconnects' length to maximize components density in a single die. From then on "***faster***" was gone as a salient feature of scaling and the industry motto was forever reduced from three to only two components: "***Smaller and Cheaper***."



Source: M. Horowitz, et al. 2010, K. Rupp, 2017.

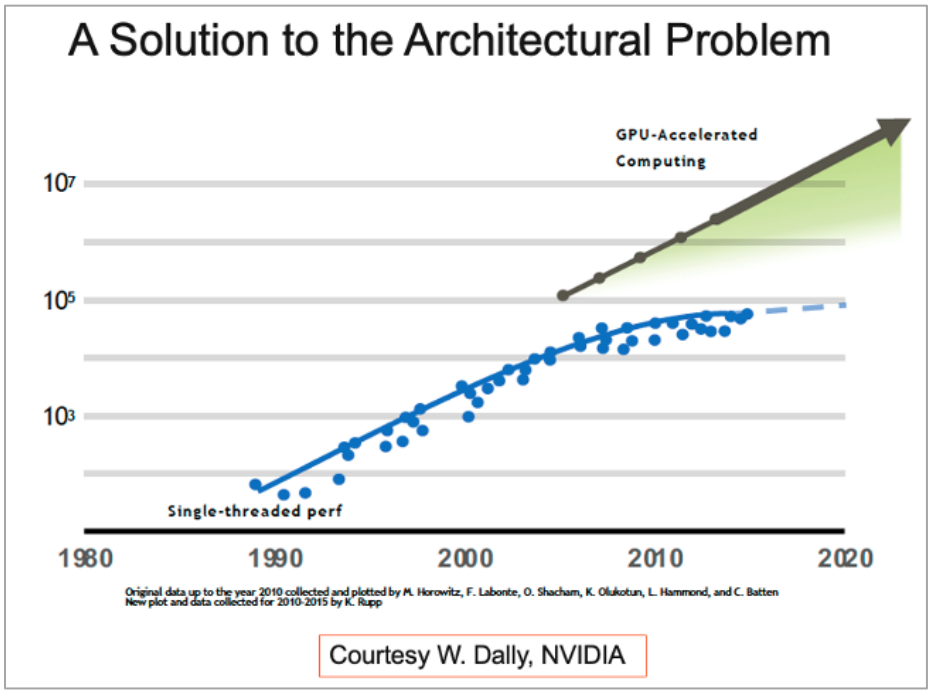
Figure 20 Tradeoffs in MPU performance after the power limit was reached.

Immediately after reaching the operating frequency limit, as predicted, the palliative solution consisted in subdividing the core of the MPU in several smaller cores with each operating at a reduced frequency. By combining the output of all of them it was possible to produce an output operating at the individual core frequency multiplied by the number of cores—this architecture was introduced in 2005. This may appear simple, but it implied that any problem at hand could be completely partitioned into essentially independent portions with each assignable to a specific core practically operating independently and in parallel with all the other cores. However, this is not possible in many cases. So, the number of cores kept on increasing and performance improved but not at historical rates. This architectural approach could not completely compensate for the inability of increasing operating frequency (Figure 20).

Customized MPUs

To continue increasing performance despite the frequency limitation imposed by power dissipation constraints, it became necessary to abandon the completely flexible MPU architecture and start defining it in accordance with the task at hand. In few words, the hardware and the software started to be optimized as specific task and performance resumed climbing once again closer to historical rates. This specialized MPU were referred to as **"*accelerators*"** (Figure. 21).

An accelerator is defined as a piece of specialized hardware circuitry that implements various functions to perform a set of operations with higher performance or greater energy efficiency than a general purpose MPU. Accelerators are nowadays commonly used to significantly improve the performance of certain workloads. They can either be a separate component attached to a system or can also be integrated directly into the processor.

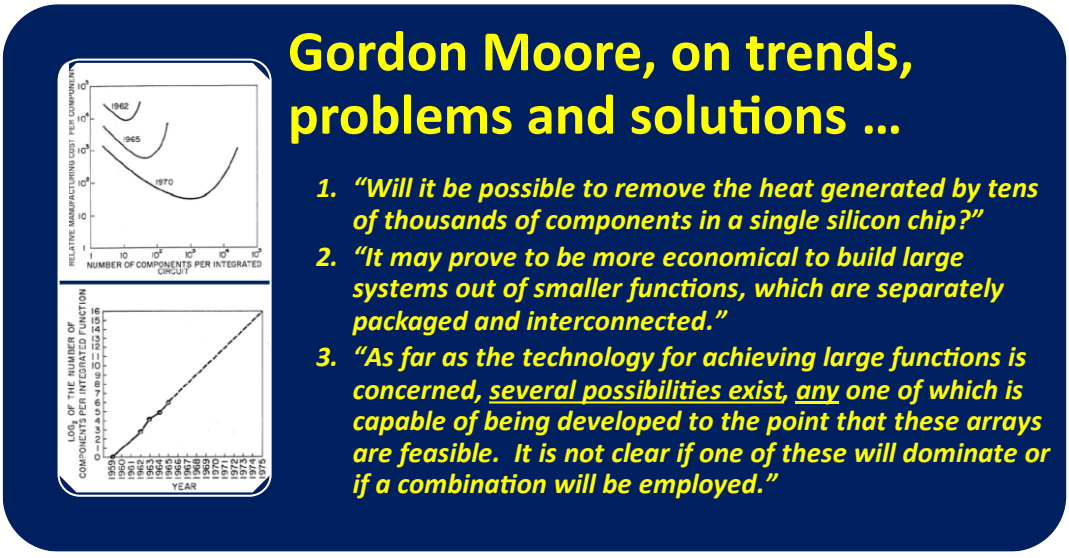


Source: NVIDIA

Figure 21 Introduction of GPU accelerators resumed performance progress at historical trends.

The Technology Components: Feature Scaling, Die Partitioning and Topology Innovation

In 1965, in his classical paper Gordon Moore had anticipated some future trends, problems and outlined possible solutions. (Figure 22).



Source: Gordon Moore papers, M1965. Dept. of Special Collections, Stanford University Libraries, Stanford, Calif. Box 3, folder 13.

Figure 22 Gordon Moore, on trends, problems and solutions-3 quotes

The first two phrases are included in the published paper, but the editor decided to omit the third sentence, which proved to be a ***critical part*** of what Gordon Moore had submitted for publication (Figure 23).

THE FUTURE OF INTEGRATED ELECTRONICS

Gordon E. Moore
Fairchild Semiconductor
A Division of Fairchild Camera & Instrument Corporation
Palo Alto, California

If allowed the luxury of a broad definition of integrated electronics to include all the various technologies which are referred to as microelectronics today, as well as any additional ones which result in electronics functions supplied to the user as irreducible units, then the future of integrated electronics is the future of electronics itself. Only a few areas in the realm of electronics do not obviously benefit from the advantages that can be expected from integration.

This paper examines the driving force which will result in integrated functions pervading the electronics art and broaden its scope beyond my imagination. Its purpose, however, is not to anticipate these extended applications, but rather to predict the development of integrated electronics technology for perhaps the next ten years. If one subscribes to the theory of the increasing rate of technical revolution, even such a relatively modest objective is almost certain to result in gross errors.

FAIRCHILD SEMICONDUCTOR®

The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wristwatch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics may be distributed throughout the machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

Present and future

By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

Both approaches have worked well and are being used in equipment today.

The author

Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild Semiconductor and has been director of the research and development laboratories since 1959.

Electronics, Volume 38, Number 8, April 19, 1965

Sources: Gordon Moore papers, M1965. Dept. of Special Collections, Stanford University Libraries, Stanford, Calif. Box 3, folder 13; Electronics, Volume 38, Number 8, April 19, 1965.[5] [6]

Figure 23 Gordon Moore, on trends, problems and solutions. original Fairchild paper and edited for Electronics magazine

As shown in previous paragraphs, scaling of device features is still progressing unabated and increasing components' density in a die in accordance with Moore's Law. However, as anticipated by Gordon Moore in 1965: "…*for achieving large functions is concerned* ***several possibilities exist***" ***and now the time has come to review how some of them are finally utilized.***

And so, by 1991 as the possible evolution of the planar silicon-gate—the workhorse of the semiconductor industry—seemed to be heading for its demise, Gordon Moore wanted to verify the viability of the above point. ***Could we still increase functionality when it will no longer be possible to scaled down device features or even when the final feasible die size would eventually be reached?*** The ensuing practical test demonstrated the long-term viability of a multi dice approach as a possible contributor to continuously increasing system performance; with this result successfully demonstrated (Figure 24) the attention was then placed back on refurbishing the silicon gate process.

Source: Intel

Figure 24 Viability of Multi-Chip Module via flip-chip technology using Sn-Pb collapsible bumps on the bond pads demonstrated in 1992-94 for future use.

Multicore and Multi-Dice Architecture

Typically, the multicores architecture described early on implied that all the cores resided within a single MPU, but this is not the only possible case.

First, it is necessary to point out that the die size of any product is limited by the maximum reticle size of lithographic exposure tools to 33x26 mm corresponding to an area of 858 mm^2. This limitation imposes that any larger die size (as dictated by product requirements) must be partitioned in several smaller dice. Since this partitioning becomes a mandatory requirement for designs exceeding the 858 mm^2 area there are also advantages derived by this imposed partitioning.

a. By placing individual core dice next to memory dice reduces signal delays

b. Power management becomes easier since the total power per individual die is reduced accordingly to the number of cores

c. Furthermore, the amount of useable silicon yielded on a wafer for a fixed number of defects is higher as the die size is reduced. While one defect per cm^2 would prevent a die of this size for functioning and the wafer yield would be zero, on the contrary once the one cm^2 required die area is subdivided in 4 dice of equal area it follows that only one die out of four would be negatively affected by the one defect but the other three dice would be fully functional; so by partitioning a single large core into multiple smaller ones increases the useful number of dice per wafer due to better die yields. This is extremely important especially in the early phases of manufacturing because when a new technology is firstly introduced the yields are still low.

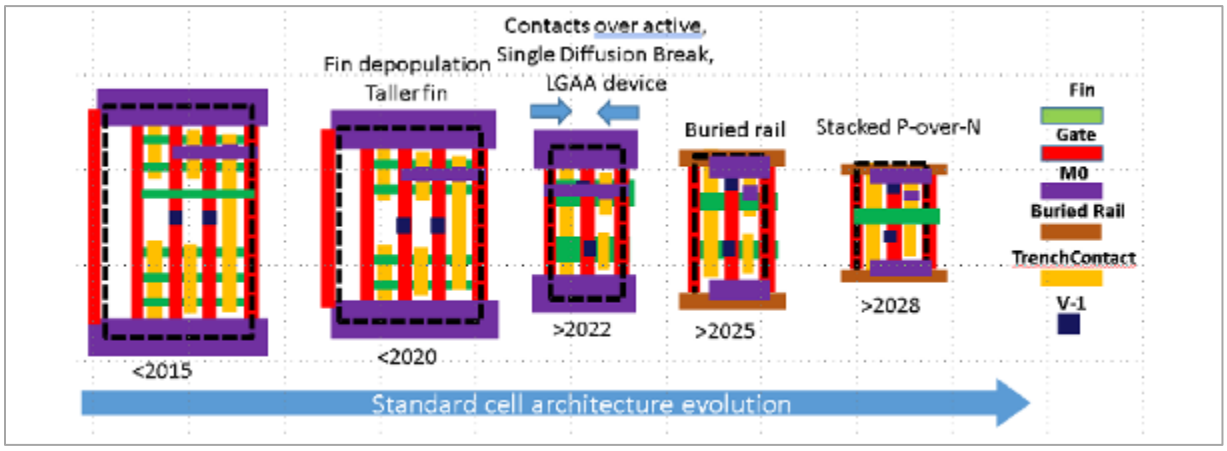
d. However, in practical cases the silicon area required by four cores, to use the example above, instead of a single large one, is slightly higher by 5-10%. This occurs because while the single large die has I/O devices on the four sides of the die periphery the smaller cores will double the amount of silicon required for the I/Os, since each one of them has I/Os on all four sides of each die and therefore doubling the area dedicated to this function.

So indeed, hand in hand coexistence of both approaches as forecasted by Gordon Moore is already progressing. But this is not all there is!

The IRDS clearly articulates how the topological scaling factor is contributing to increase transistor density. The following are comments extracted from the More Moore chapter of the 2022 IRDS. See the full chapter for additional insight.

Ground rule scaling alone is not adequate to scale the cell height

It is necessary to bring the design scaling factor into practice. For example, standard cell height will be further reduced by scaling the number/width of active devices in the standard cell as well as scaling the secondary rules such as tip-to-tip, extension, P-N separation, and minimum area rules. Similarly, the standard cell width can be reduced by focusing on critical design rules such as fin termination at the edge fin, etc., and enabling structures such as ***contact-over-active.*** Also, the contact structure needs to be carefully selected to reduce the risk of increased current density at the junctions. It is expected that beyond 2028 P and N devices could be stacked on top of each other allowing a further reduction. This trend in standard cell scaling is shown in Figure 25.



Source: 2022IRDS, More Moore

Figure 25 Scaling of standard cell height and width through fin depopulation and device stacking.

After 2031 there is no room for 2D geometry scaling, therefore 3D very large-scale integration (VLSI) of circuits and systems using sequential/stacked integration approaches will be necessary. This is since there is no room for contact placement as well as worsening performance because of gate pitch scaling and metal pitch scaling. It is projected that physical channel length would saturate around 12 nm due to worsening electrostatics while gate pitch reduction will continue until a width (~14 nm) to still provide sufficient space for the device contact while providing acceptable parasitics. This drawback of pitch scaling represents a compromise derived by dual-gate-pitch processing where relaxed pitch devices are used for high-performance cells while concurrently tight pitch devices are used for high-density cells.

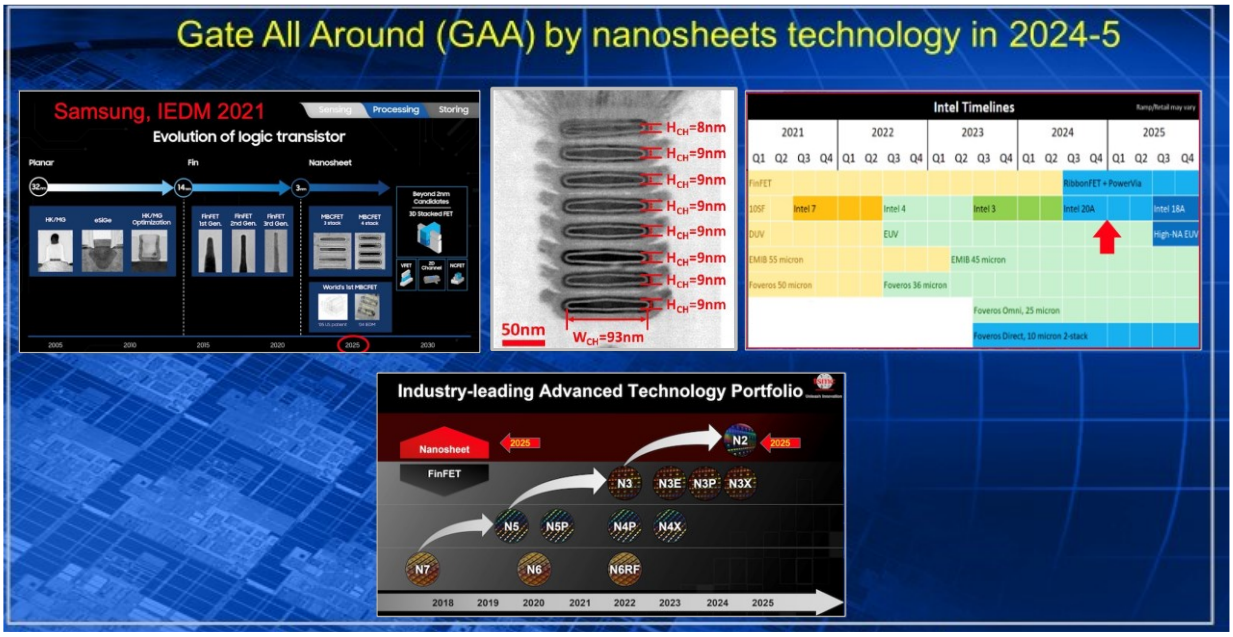
3D Everything

1. Die Level

Nanosheets Transistors

As stated in a previous paragraph making transistor footprints smaller still remains a goal of the technology effort to increase the density of components in a single die, but why limit the individual transistor layout to a 2D scheme? In 2013, the ITRS pointed to the third dimension as an avenue to reduce transistor footprint. This was actively pursued by NAND manufactures that began stacking layers of memory cells one on top of each other and the first commercially available products were introduced in 2015. NAND leaders are nowadays manufacturing products with as many as 232 layers of memory cells and this number will increase for the foreseeable future! The introduction of FinFET intro manufacturing in 2011 opened the way to exploring the benefits of the 3rd dimension for manufacturers. It was also forecasted by the ITRS that logic device makers would eventually follow a similar approach of the NAND

manufacturers by 2025. This date has been indeed confirmed to be correct as many logic leaders have announced early introduction of stacked nanosheets transistors in manufacturing in the next couple of years (Figure 26).



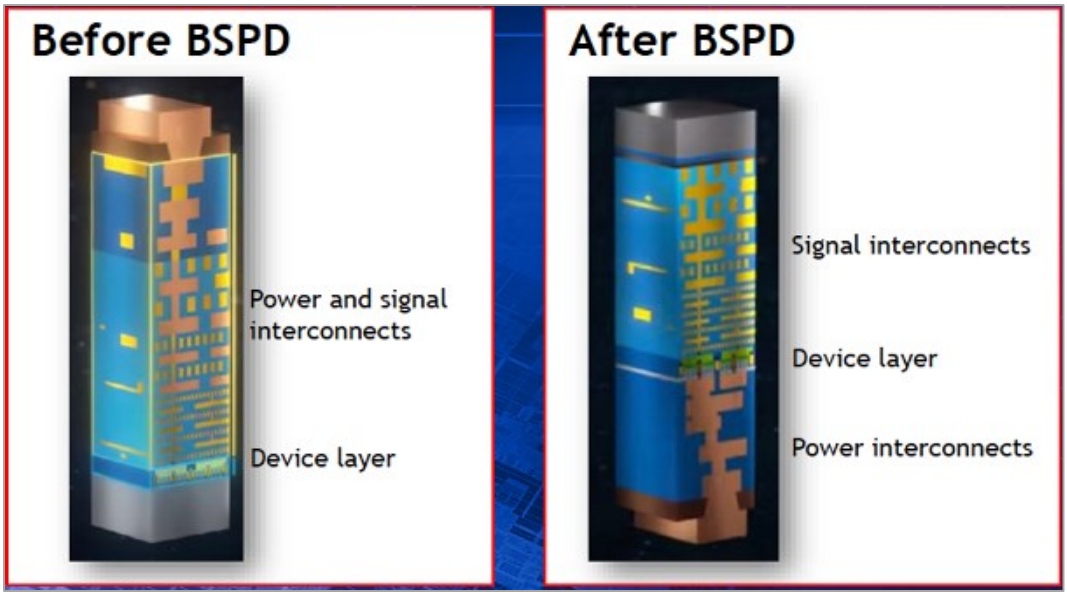
Sources: Samsung, Intel, TSMC

Figure 26 Introduction of nanosheets in manufacturing anticipated by ITRS in 2013 is scheduled for 2024-5

The nanosheet approach consisting of building multiple nanosheet layers on top of another to make transistors of reduced footprint will become the standard building block in IC designs. Eventually CMOS structures will be built by stacking the PMOS and NMOS transistors on top of each other and so on.

Back Side Power Distribution

An additional technological approach to increase density and performance is represented by the newly introduced concept of back side power distribution (BSPD). Complexity of the MPU has kept on increasing over the years and so has the number of interconnects. These are used to carry signals and deliver power. Feature scaling rapidly decreases the pitch of local interconnects whose resistance increase negatively affecting signal propagation delays. On the other hand, power delivery requires the largest possible interconnect lines to minimize resistance and consume excessive space as they make their way down to the lower layers of interconnects (top-down). Using power vias to deliver power from the back of the die relax the pitch of the lowest interconnect layer and enable an independent optimization of front-side interconnect layers for improved signal delivery and backside delivery of power on reduced resistance lines. Overall, this approach leads to better area utilization and shorter wires (Figure 27).



Source: 2023 VLSI Symposium, Intel

Figure 27 Area reduction, reduced signal propagation delay and lower resistance power delivery with BSPD

In summary, cells with back-side device contacts (BSCON) and cells with transistor stacking can provide large area scaling and performance improvements.

But there is more to this!

2. Package Level

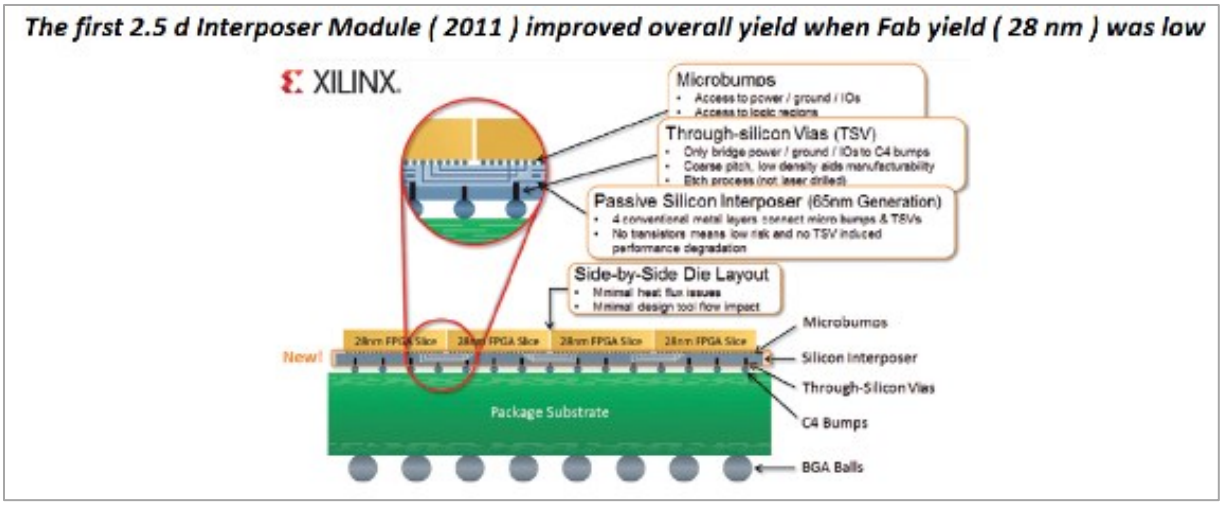
a. In addition to 3D integration at the die level a multitude of packaging technologies have reemerged in the past 10 years. Flip chip bonding utilizing Sn-Pb collapsible bumps was already used by system companies as far back as the early 80s (Figure 24). Once the above alloy was deposited in the bond pads the chip was flipped over, the bond pads were aligned to corresponding landing pads and then a short temperature cycle "reflowed" the Sn-Pb mix that would then form a stable bond. Nowadays, Bismuth (Bi) has replaced Pb due to health concerns but with very similar properties.

b. Package-on-Package (PoP) was also used as far back as the 80s to double the amount of available memory at a reduced footprint. In this case a fully packaged die was placed on top of another packaged die and the two were then connected to each other by wire bonding or even lead to lead soldering.

c. Nowadays it has become quite common to place two dice with one next to the other (2.5D) typically on an organic interposer and then connect them to each other by means of multiple layers of interconnections built in the organic or any other material types of substrates. Communications between the dice occur via the circuitry built in the interposer (Figures 28 and 29).

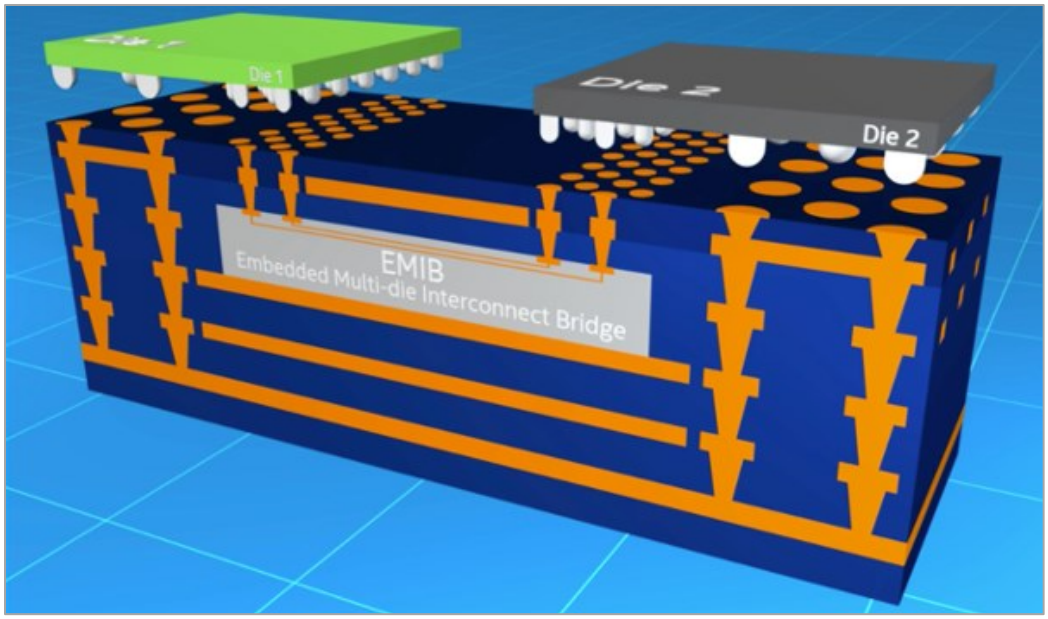
d. A multitude of approaches whereby memory dice are mounted on top and connected to some type of MPU or silicon on chip (SOC) are also widely used. Here the nomenclature characterizing multiple variations around this theme begins to be quite extensive (Figure 30).

e. Finally, wafer-to-wafer bonding has become an enabling semiconductor technology in the 3D arsenal. In a typical wafer bonding process, two face-to-face wafers are permanently joined (i.e., bonded) to one another by applying physical pressure, temperature, and/or an electric field. This approach enables substantially different technologies to be combined in a single die. As an example, in NAND products it has been demonstrated that by separately producing CMOS high-speed wafers (incompatible with high-temperature array processing) and separately producing NAND cell arrays, it is possible to obtain highest speed products (Figure 31).



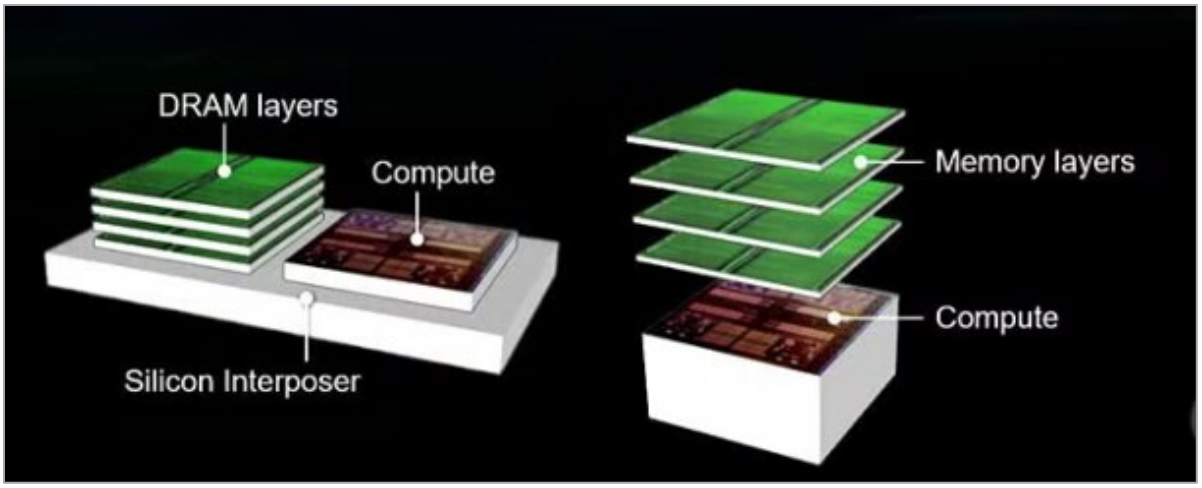
Source: XILINX

Figure 28 2.5D introduction heralded the use of multidice for leading products



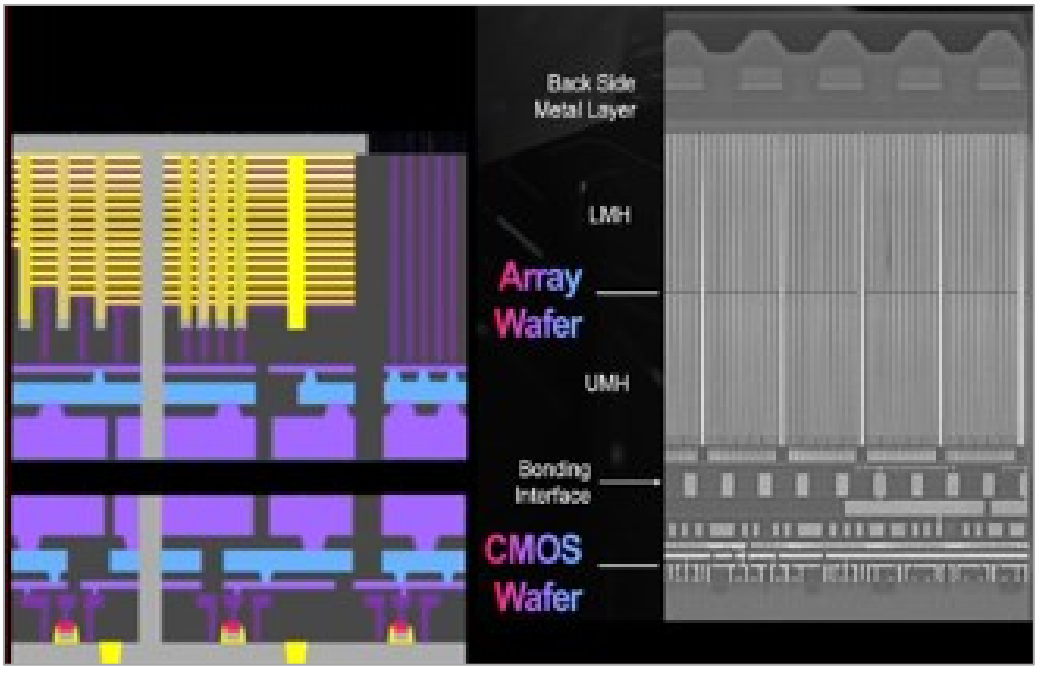
Source: Intel

Figure 29 Evolution of 2.5D supporting innovative interconnection processes



Source: SEMI

Figure 30 Vertical stacking memory dice on top of MPU reduces length of interconnects.



Source: Siva Sivaram, President Western Digital, VLSI Symposium, June 12, 2023, Plenary

Figure 31 Wafer-to-wafer bonding is used to maximize array density and CMOS periphery speed.

3D VLSI expects to bring performance power area cost (PPAC) gains for the target node as well as to pave ways for heterogeneous and/or hybrid integration. The challenge of such integration in 3D is ***how to partition the system to result in better utilization of devices, interconnects, and sub-systems such as memory, analog, and I/O.*** That is why functional scaling and/or significant architectural changes are required after 2031. This would potentially be the time where "beyond CMOS" and specialty technology devices/components would bring up the system scaling towards high-system performance at unit power density and at unit cube volume.

Optical interconnects will provide solutions as new architectures are developed. A critical issue is signal propagation delay. Establishing the most expeditious means of communications among billions of transistors in a chip; among

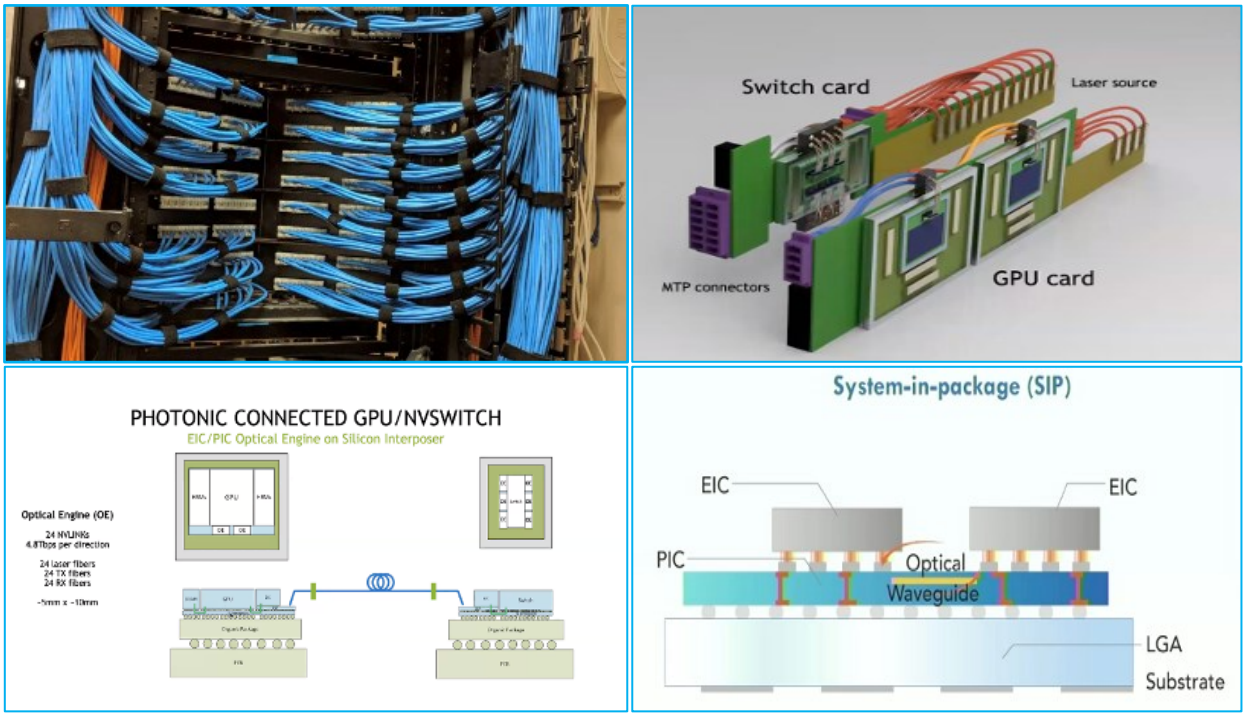
multiple dice located in a single package, or when communicating from package-to-package or board-to-board is challenged with the delay with the signal travelling on interconnects from one location to another. Switching transistors or circuits at higher speeds may become irrelevant when the speed of signal propagation is limited by the ability of channeling the signal via interconnects.

Additionally, signal propagation delay of interconnects also depends on their resistivity and capacitance. Replacement of aluminum with copper (due to the lower resistivity of the latter) has achieved resistivity reduction to date. There are no other materials that can further reduce this resistivity.

Today's solutions focus on minimizing the length of interconnections as the only available avenue to reduce further signal propagation delay of interconnections. At the chip level this approach has led to the multiplication of layers of interconnections, with as many as 17 layers of interconnections successfully introduced into manufacturing in logic products. At the package level, placing layers of memory dice on top of computing chips reduces the length of interconnections to the shortest possible physical limits (Figure 30).

The use of fiber-optics technology is well known as the best interconnect technology for communications. Fiber optics propagating optical signals and not electrical signals have been used for long-range communications due to their minimal signal loss. Typically, fiber-supported communications reach distances as far as 100 km (62 miles) without requiring amplification and 200 km (124 miles) can be accomplished also with special fibers.

An electrical signal traveling back and forth on the internet from any source typically spends 70% of the total round-trip time within the data center. For this reason, data centers were the first to adopt fiber optics communications from rack to rack. Now this technology can be applied to semiconductors interconnection at the chip and on the package (Figure 32).



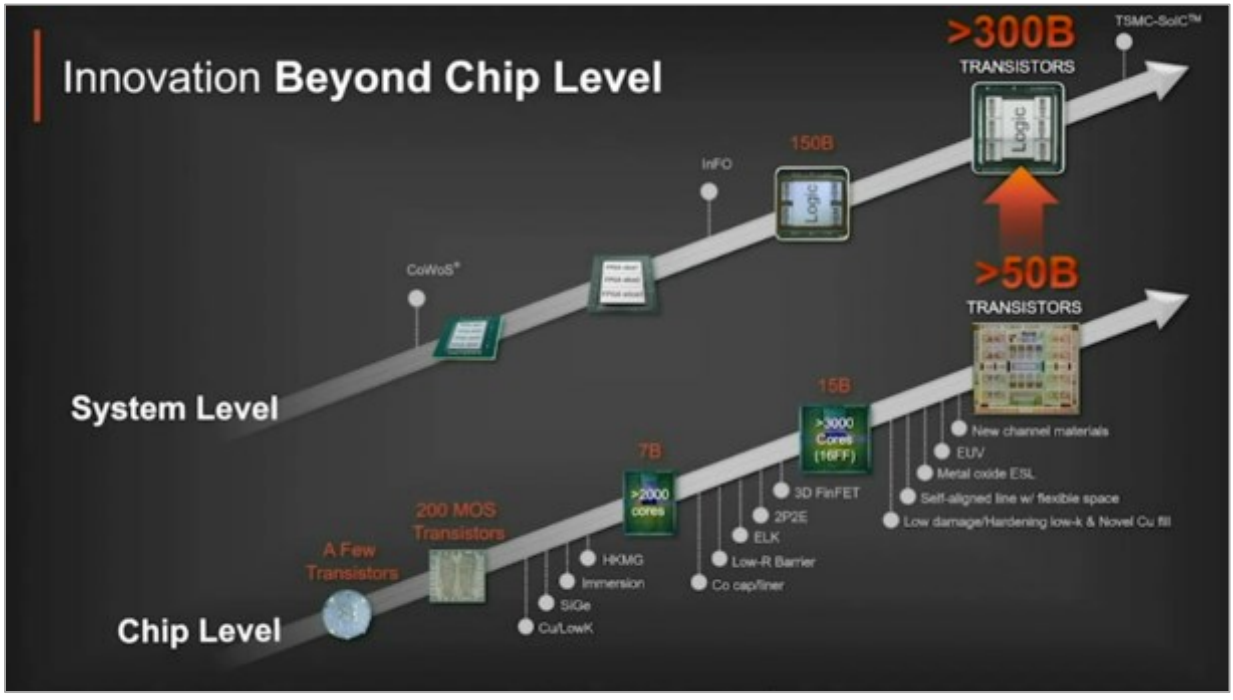
Sources: Fiber Broadband Association, NVIDIA, Lightelligence

Figure 32 Evolution of fiber optics from data centers (meters) to board (centimeters) to package-to-package (millimeters) to chip-to-chip (tens of microns)

Summation of Die and Package Level Contribution to Increasing System Functionality

From these considerations, it appears obvious that integration of multiple functions will continue unabated for the foreseeable future. For about 50 years, increasing transistor count on a single die was the main vehicle to increasing functionality.

However, recommendations on several enhancements to this approach to further increase functionality were anticipated and outlined in the original paper submitted in Gordon Moore's classical 1965 publication that was not included in the publication. The viability of his recommendation on multichip technology for high-volume manufacturing was verified as far back as 1992-94. However, one approach does not preclude from using the other approach also, so they conjunctively contribute in very different ways to enhance system integration via monolithic implementation at the die and the package level (Figure 33).



Source: Mark Liu, Chairman Board of TSMC, February 15, 2021, IEDM Plenary

Figure 33 Increase in transistor density at the chip level and at the system level (via package integration) proceeding in lock steps.

The advent of FinFET demonstrated that the third dimension was a viable approach to increase transistor performance and layout optimization. Stacked memory cells in NAND and nanosheets being implemented for logic products represent the evolution of utilization of the third dimension at the die level.

Multi-dice packaging is already optimizing system performance. Wafer-to-wafer bonding offers the possibility of integrating multifunctional but completely different technologies merged in a single die. We can therefore confidently formulate a long-term forecast on total transistor integration for the foreseeable future (Figure 34).

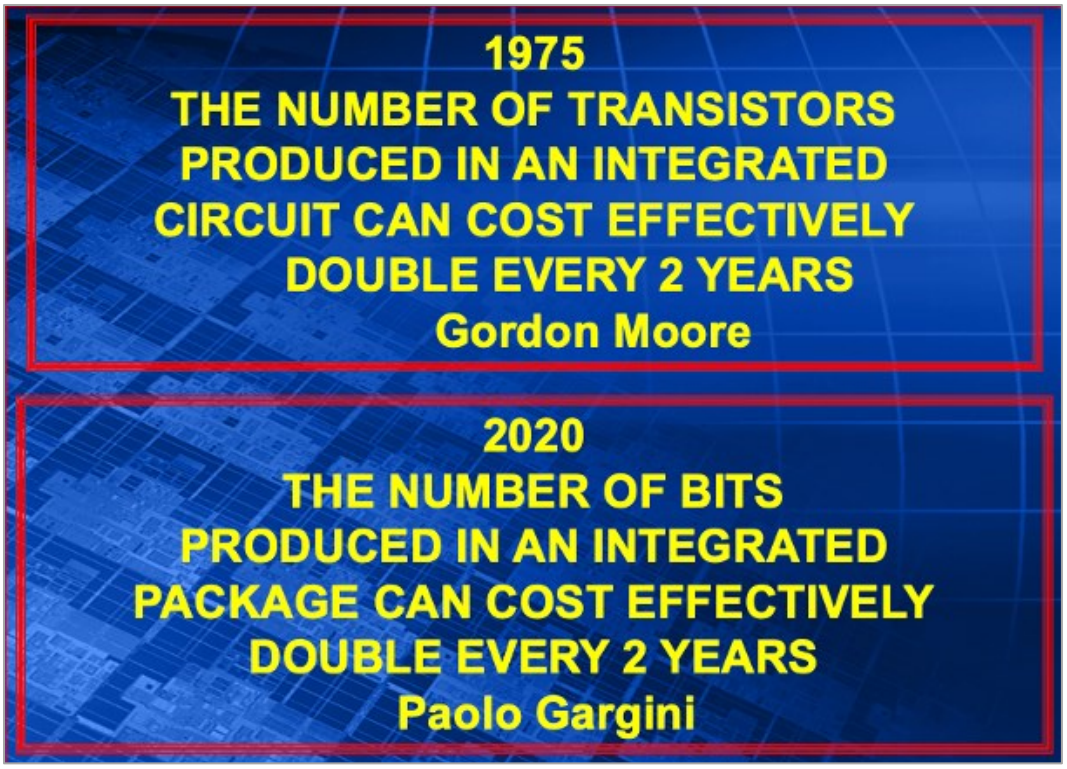


Figure 34 Transistor integration and package integration proceeding in parallel trends.

Summing it All Up!

- ***Equipment and materials suppliers emerged as the winners at the end of the second crisis.***
  - To this day and for the foreseeable future, ***equipment and material*** capabilities ***will not*** be a limiter to the progress of the semiconductor industry.
- ***Semiconductor manufacturers emerged as the winners at the end of the third crisis.***
  - Semiconductor technology will not be a limiter of the electronics industry in the foreseeable future.
- ***System integrators will emerge as the winners of the fourth crisis.***
  - System integration requirements will guide the adoption of new technologies for the benefit of the electronics industry.

Feature scaling will continue unabated for the next 10 years and beyond as a means of increasing transistor density in ICs. However, transistor channel length scaling is no longer a "must do" to meet performance requirement as maximum operating frequency is limited to 5-6 GHz due to limitations imposed by dynamic power dissipation. Multilayer NAND memory cell are solidly in production and nanosheet transistors will follow FinFET transistors and then stacked NMOS/PMOS transistor will follow. A variety of 2.5D and 3D structural approaches will increase components density and integration of many homogeneous and heterogenous technologies in new revolutionary systems.

7. PRACTICAL GUIDE ON HOW TO BENEFIT FROM PREVIOUS CRISES' EXPERIENCES TO ORGANIZE FOR THE 4TH CRISIS

Prototyping New Technology Capability is Key to Success.

The IRDS has already identified the upcoming challenges for the electronics industry, and it is pointing the way to the research community. ***Prototyping*** will be more important than ever due to the increased interdependency of process, transistor, topology, packaging, circuit complexity and overall system requirements. Before starting any project, it is essential to clearly understand what the goal of the overall system is and to identify who the beneficiaries of any successful outcome would be. The first crisis benefitted the semiconductor industry at large. The second crisis benefitted the equipment and material suppliers. The third crisis benefitted the semiconductor manufacturers.

System integrators operating in the fabless/foundry model will be the ***beneficiaries of the programs funded in the fourth crisis.***

Unless this last point is fully realized it is not possible to map a path to success.

The Fabless/Foundry Leaders of the Electronics Industry Meet Their First Real Challenge

Fabless system integrators and foundries are the economic engines and the true giants of the electronics industry. They have emerged in the past 15 years and have reached unprecedented market capitalizations. The products generated with this methodology have been highly successful, but the electronics industry is now facing a classical situation where ***successful products have reached their limits of growth.***

The PC industry shipments peaked at around 300 million units in 2012 and had been declining since then. The demand for video conferencing during the pandemic created an unexpected but short-lived resurgence of PCs that is already over. Cell phone shipments peaked at around 1.5 billion units in 2014. Services are now the most profitable revenue contribution to successful mobile phone companies.

It is becoming clear that new products are badly needed to continuously grow the electronics industry at or better historical trends.

But what are the products that will dominate the electronics industry in the next 10 years? The capabilities of these products have been identified, but the products do not yet exist. This is not an unusual situation.

How can the IRDS Help the Customers Before They Even Know What They Need?

System integrators have nowadays outlined what kind of products they would like to have manufactured on their behalf by the globally integrated supply chain in the next five years. However, there is ***not a method or even an organization responsible for the development of the necessary but not yet existing technology building blocks.*** In essence, most of these technologies either do not exist at this time or have not been developed on 300 mm tools for high-volume manufacturing.

The ***epidemy*** of the present situation was made remarkedly clear during the keynote address delivered at IEDM in 2021 by Michael Abrash, Reality Labs Chief Scientist of META, when, after outlining a series of exciting new products he concluded by saying, ***"The technology does not exist…but it will."*** It may appear as simple as represented by the level of confidence with which this was stated in bright daylight, ***but this is the core of the present problem—who is going to take the initiative towards making these technologies real? (See Figure 35.)***

Figure 35 New missions of the electronics industry partners

Today's Challenge: Reverse Engineering Products' Requirements to Identify the Missing Technology Modules

System integrators have repeatedly announced the type of products they are intending to introduce in the foreseeable future, so it is imperative to identify from these products specifications which technologies are required since several of them are not readily available yet. All these elements need to be coherently prototyped to support the introduction of new products.

Here is a simple list:

1. ***6G and Wi-Fi 7 and beyond***—These will require devices operating well beyond the 500 GHz frequency and reaching the Terahertz regime. None of the existing devices can efficiently operate in this frequency range.
2. ***Mixed reality and virtual reality***—Several types of video products resembling divers' masks may be used as introductory products but normally looking glasses with lens populated by LED or similar displays capable of blending reality and virtual imaging need to be created and optimized.
3. ***Computing at the edge and self-driving automobiles***—Multiple sensors using very low power while utilizing visible and invisible electromagnetic frequencies to detect almost everything that need to be developed.
4. ***Holography and Avatars***—During the pandemic people that had never used a PC learned how to use one to make video meetings. In the future 3D projections (avatars) will become viable but a long list of devices and optical technologies need to be generated to make this real.
5. ***Fiber optics to everything***—All the activities listed above have an insatiable requirement for bandwidth. Copper coaxial lines are not very efficient above tens of Gigahertz and so fiber optics need to be deployed everywhere. A multitude of optical devices needs to be developed for this purpose.
6. ***Power management***—This task requires devices like SiC and GaN that are only available on small size wafers and the engineering to take them up to 300 mm wafers is still lacking.
7. ***Biotechnology***—Multiple demonstration of implanting devices in humans and contactless sensing require optimization of many technologies to become commonly useable.

This is nothing more than an initial list of reverse engineered tasks that require development of new devices and technologies to support the next generations of products (Figure 36).

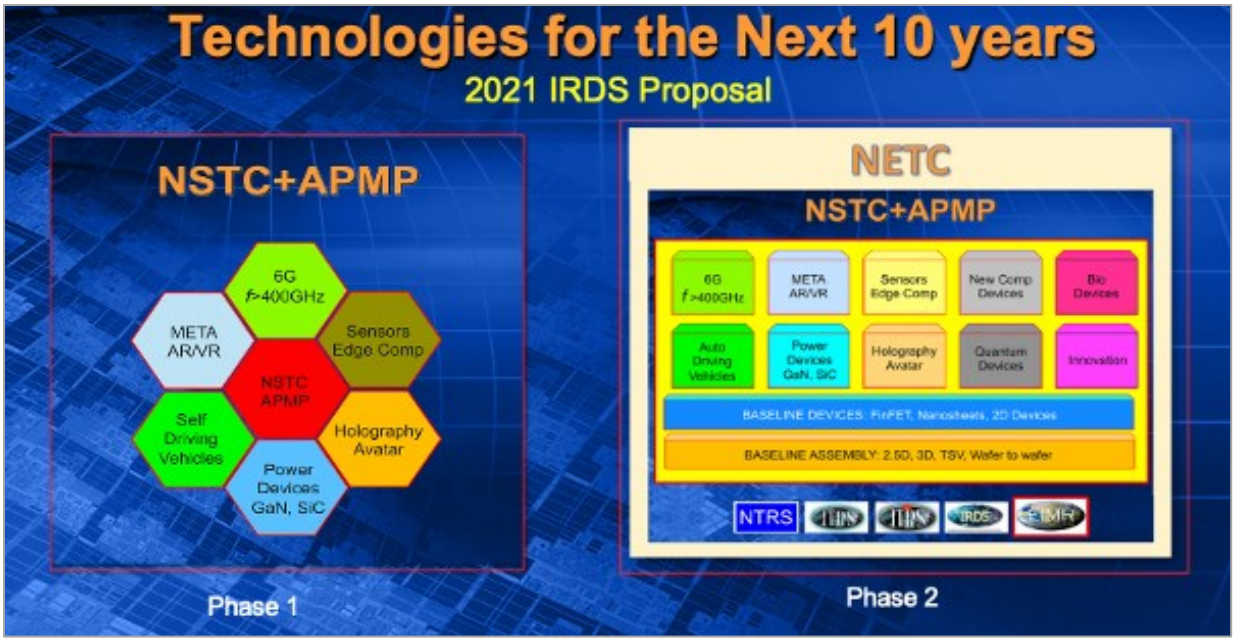


Figure 36 IRDS estimates of technologies needed to enable innovative systems for the next 10 years.

Once these technologies and devices are fully developed then it is the role of semiconductor producers to absorb all of these into high-volume manufacturing but ***who will develop and prototype the building blocks necessary to assemble the next generations of electronics products?***

8. NUTS AND BOLTS

Let's now address the practical implementation! All the above chapters would mean nothing unless practical solutions are recommended and adopted.

Prototyping is not Research nor Manufacturing

The following recommendations should be considered as a blueprint of what is needed to perform a successful prototyping operating supporting the needs of system integrators for the next 10 years. Each region already has some of these elements in place so the list of items reported below should be considered for planning purpose only and it is expected that each region will customize it based on existing or already planned buildings, capabilities, and organization (Figure 37).

1. Design center

The close interaction and integration of technologies and systems requires planning the whole system from the very beginning to generate top-down requirements for technologies and devices. The bottom-up approach of the past is no longer viable. Prototyping is no longer limited to testing novel individual transistors or circuits, but it requires testing the building blocks of future complex systems. A design center may require up to $100M if fully equipped with adequate design tools and computers capabilities.

2. New universities' clean rooms are needed

To provide the systems integrators with the necessary technology outlined above it is necessary to channel university research along these novel directions. However, the feasibility of this research must be eventually demonstrated on 200 mm wafers to provide a necessary level of confidence on the viability of the technologies to the system integrators. At present no single university has this capability! It is therefore recognized that at least three university clean rooms need to be upgraded with these capabilities in each region. This possibly implies new buildings and a brand-new equipment set. Minimal individual investment should be at least in the $120M range for each university.

3. The prototyping center should consist of 3 basic elements.

a. A clean room constituted of at least an initial 50K sq. ft. size and extendable to 100K sq. ft. and equipped with the latest 300 mm equipment. This should cost around $3B.

b. An annexed assembly plant working as a seamless extension of the wafer prototyping plant. Initial cost should be in the range of $500M.

c. A material research center. Many new materials will be required to demonstrate the products outlined above. This will require about $300M.

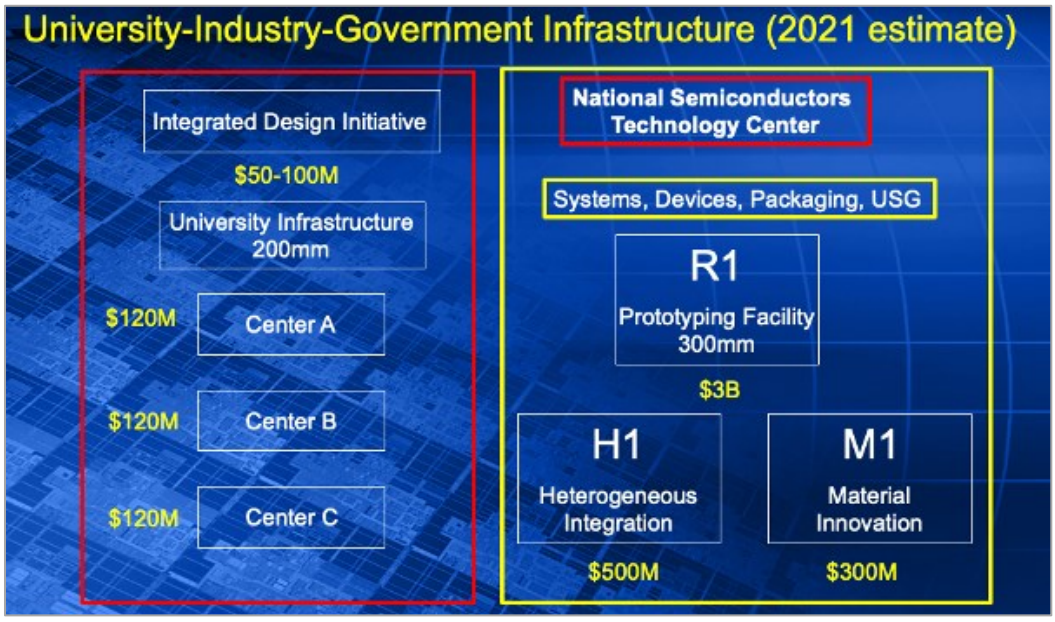


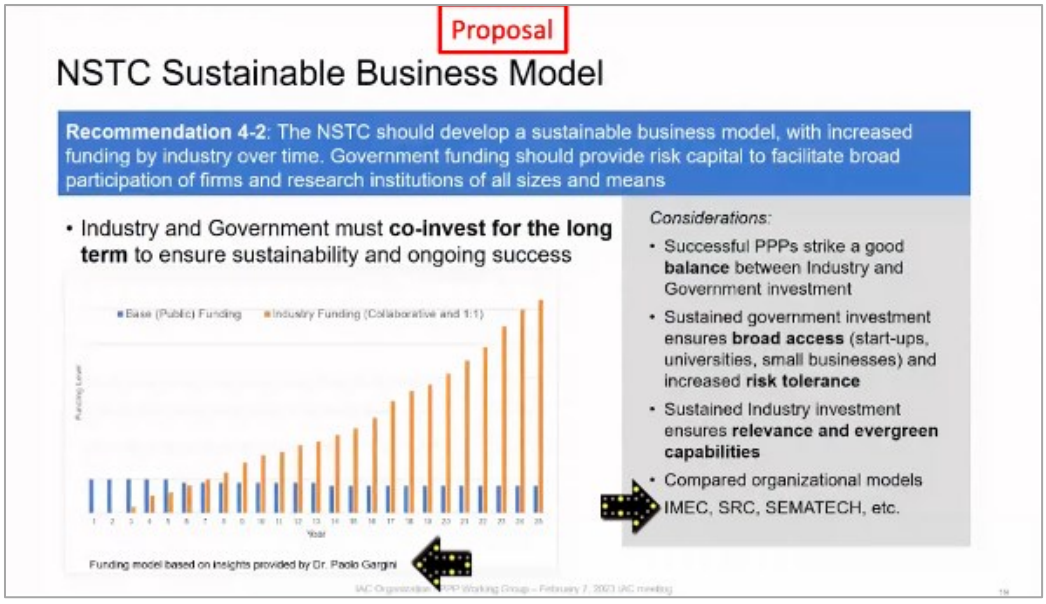
Figure 37 IRDS recommendation on how prototyping of systems' building blocks should be organized for success

4. Management

a. Most consortia are dominated by boards composed of major stakeholders that, for obvious reasons, seek benefits for their own organization. For this reason, the board members are reluctant to provide valuable information on problems that are essential for their business plans and so in most cases the least common denominators become the programs for the consortium (Figure 38).

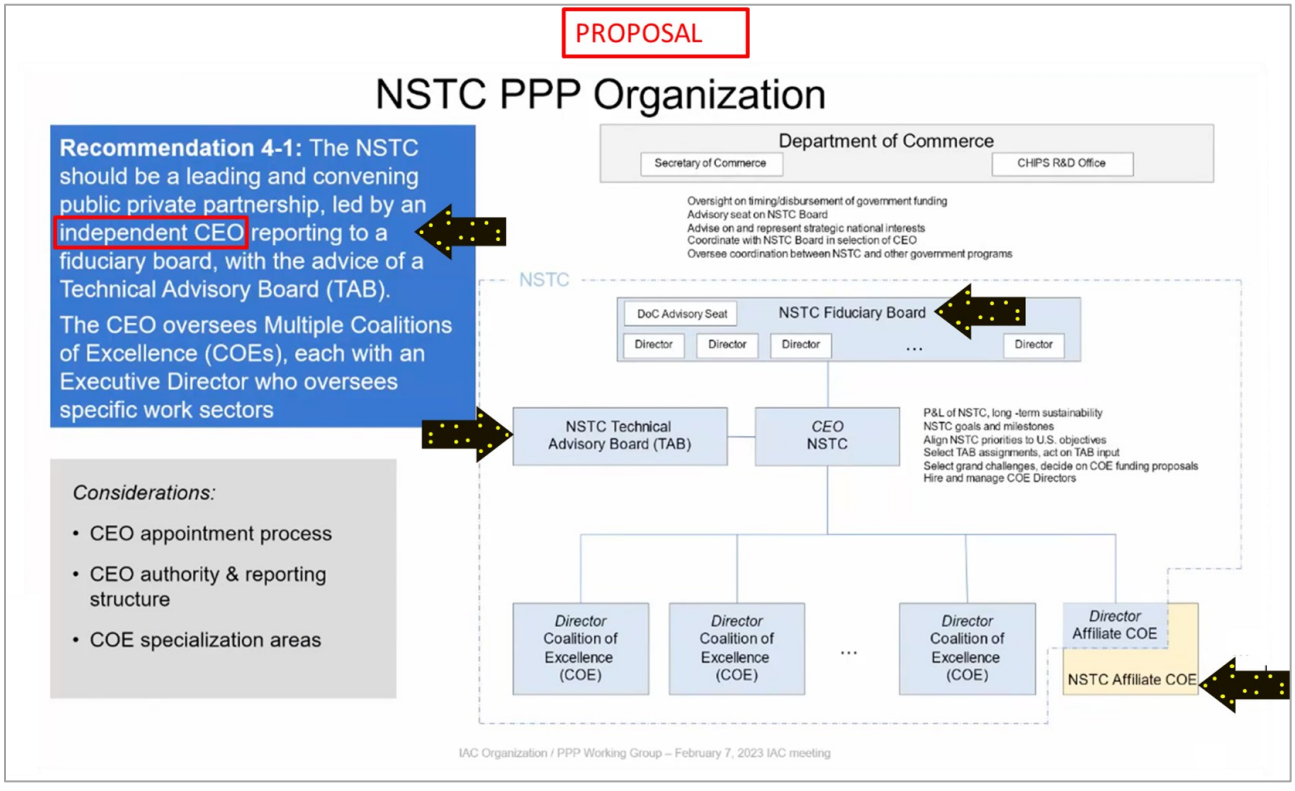
b. The IMEC model has demonstrated the viability of a technology driven organization. IMEC has a fiduciary board but does not have a board of stakeholders. The CEO and staff decide the content and goals of the programs. The Scientific Advisory Board provides technical advice based on industry trends and roadmap forecasts (Figure 39).

c. Technology transfer to member companies needs to be done by a dedicated organization augmented by the support of a receiving organization within the member company.



Source: IAC Organization PPP Working Group meeting, February 7, 2023

Figure 38 IRDS and IAC proposal on industry and government co-investment in prototyping efforts Feb 7, 2023

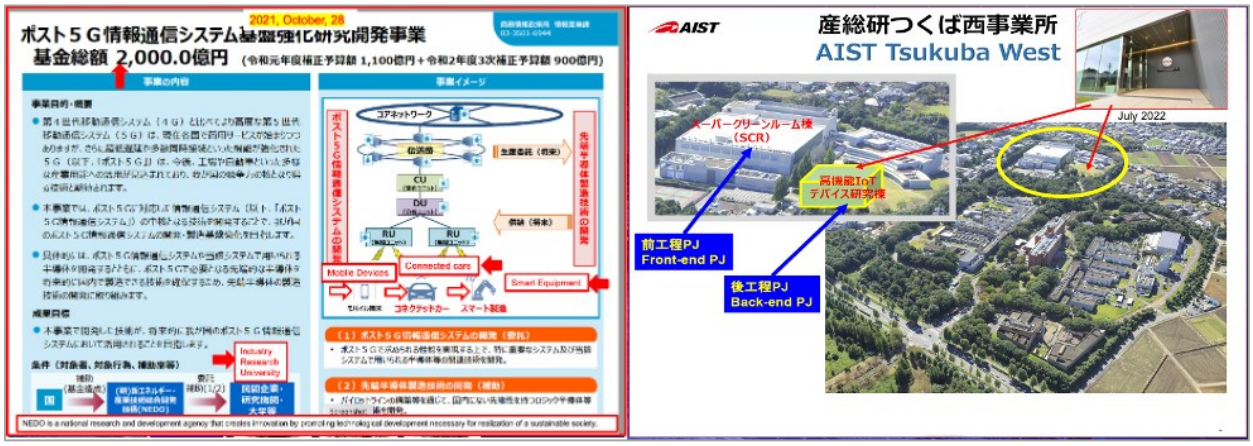


Source: IAC Organization PPP Working Group meeting, February 7, 2023

Figure 39 IRDS and IAC proposal for success of NSTC organization: Fiduciary Board, Technical Advisory Board and Independent CEO. Feb 7, 2023

Present Status

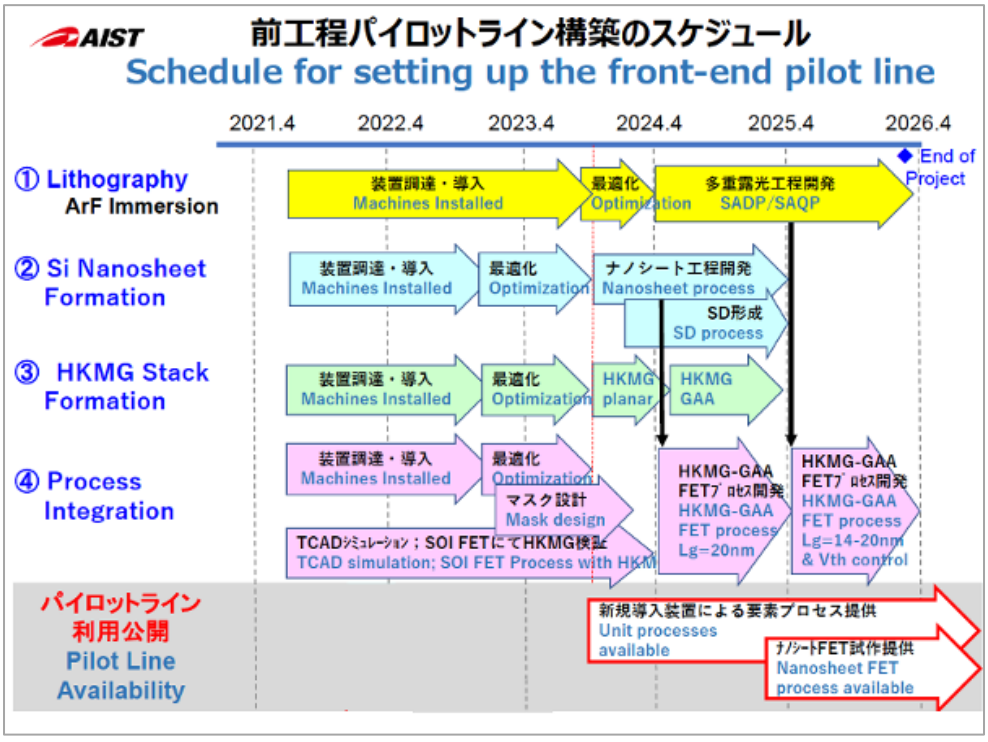
1. ***Japan.*** The super clean room (SCR) in Tsukuba was selected in October 2021 as the site offering the best opportunities to realize most of the basic capabilities outlined above (Figure 40). New equipment has been moving in since 2022 and device demonstration is expected in 2024 (Figure 41).



Source: Advanced Industrial Science and technology (AIST), Japan

Figure 40 Japan Chips Act targets development of technologies aimed at mobile devices, connected cars and smart equipment in SCR plus the new 3DIC integration center in cooperation with TSMC.

Overall Program Schedule



Source: Advanced Industrial Science and technology (AIST), Japan

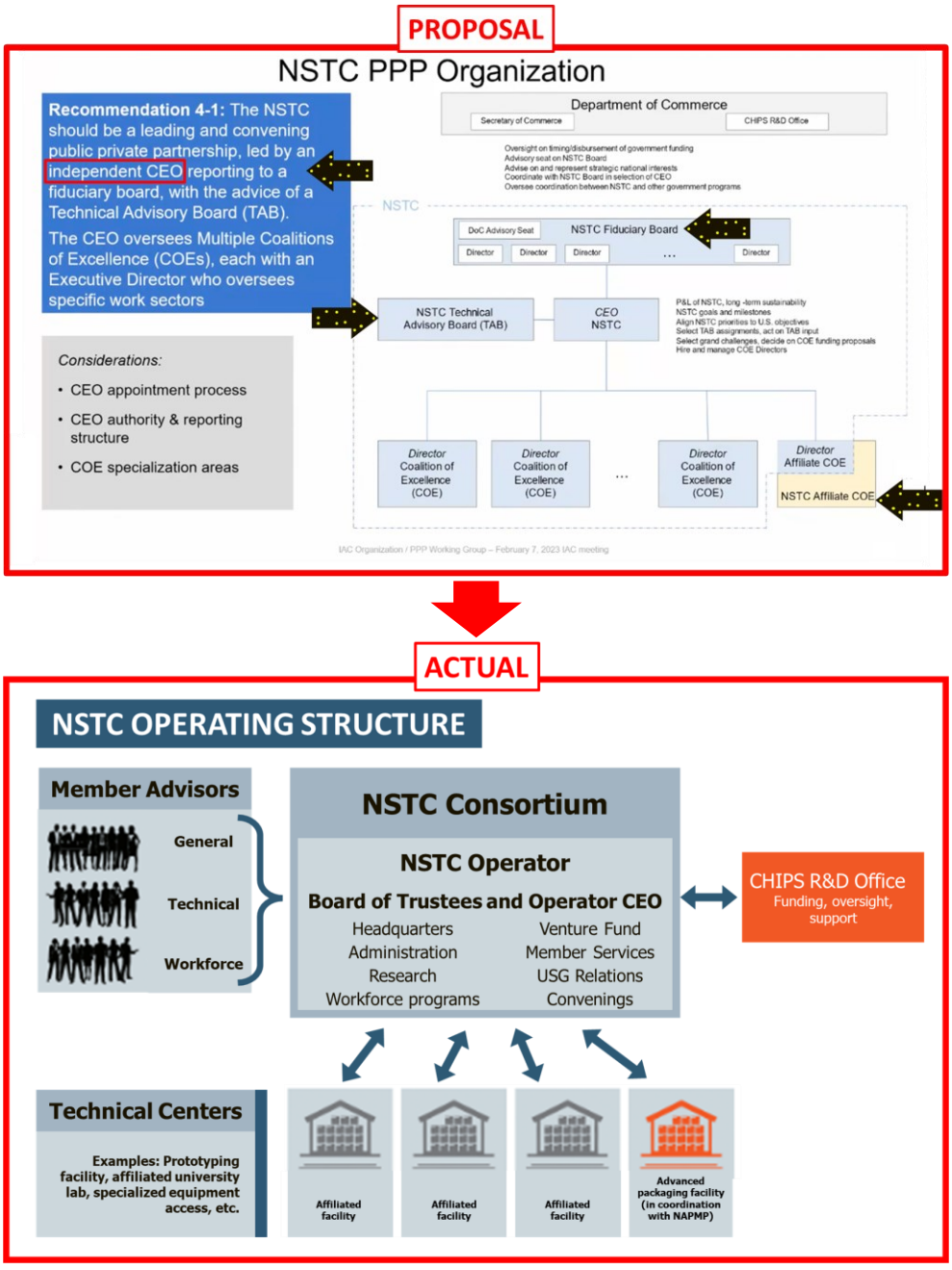
Figure 41 Integration of new devices planned for 2024.

2. ***Europe***. IMEC and LETI are already capable of processing 300 mm wafers, Fraunhofer institutes have excellent packaging capabilities and new innovative technologies consistent with the ones outlined above are under discussion. In addition, Europe has formed an organization on International Cooperation on Semiconductor (ICOS) to harmonize research, coordinate prototyping and facilitate manufacturing of semiconductors around the world (Figure 42).

Source: Sustainable Electronics and ICOS, F. Balestra

Figure 42 International Cooperation on Semiconductors (ICOS) launched by European Chip Act.

3. ***US***. The outline of the National Semiconductor Technology Center (NSTC) including the organization and a list of technologies under consideration was published on April 25, 2023, consistent with IRDS proposals (Figures 43, 44, 45 and 46).



Source: A Vision and Strategy for the NSTC.pdf. April 25, 2023

Figure 43 Organizational structure of NSTC proposed by IRDS via the IAC has been accepted.

Source: A Vision and Strategy for the NSTC.pdf, April 25, 2023

Figure 44 A Vision and Strategy for the National Semiconductor Technology Center.

TECHNICAL CENTERS: COMMUNITY INPUT

The semiconductor community has provided extensive input to the requirements for potential technical centers. The NSTC will embark on a prioritization process to ensure that the highest priority needs are met with the funds available. Identified needs for the NSTC to consider include:

- Baseline CMOS (complementary metal-oxide semiconductor): Fully functional and supported CMOS process flow at 22 nm or below with a capacity of 10,000 wafer starts per month on 300 mm wafers
- CMOS R&D process: Front-end short loops supporting < 3 nm technology R&D at a capacity of 2,000 wafers per month using extreme ultraviolet technology enabling the development of leading-edge materials, devices, and process and metrology tools
- Manufacturing test vehicles that provide low-cost patterned and functional substrates that can be used to provide data through electrical test, to enable materials, equipment, process, and device development and optimization, especially for CMOS+X enabled technologies
- Extended metrology capacity to enable R&D in a production environment including rapid failure analysis to shorten prototype development cycles, extensive in-line process monitoring capabilities, and off-line characterization facilities
- Space and flexibility to accommodate next-generation or prototype processing and metrology tools so that they can be demonstrated in a production environment
- Back-end short loop processing from specialized capabilities enabling "fab-to-lab"[24] finishing of R&D devices and high-quality processing of novel materials and devices while maintaining process and material segregation

Source: A Vision and Strategy for the NSTC.pdf, April 25, 2023

Figure 45 IRDS proposed prototyping line structure via IAC adopted for NSTC.

- Power electronics: Power management devices often require non-silicon substrates (e.g., silicon carbide, gallium nitride) and specialized designs, tools, and processes
- Radio frequency, mixed signal, and analog: Communication and sensing applications require diverse capabilities distinct from leading-edge CMOS
- Photonics: Advancements in quantum, sensing, and interconnect are all possible at the intersection of light and electronics
- Microelectromechanical systems: Sensors for mobile, automotive, health care, and internet-of-things are all growth areas that require resources distinct from traditional CMOS flow
- Bioelectronics: The convergence of microfabrication and biotechnology brings new opportunities, but also increased complexity and significant integration challenges
- Mature node: The NSTC may seek to have capacity at a mature node (e.g., 130 nm), with such a facility well suited to certain research programs and workforce education
- Design tools: New design tools and methodologies to accelerate the generation of circuit IP; virtualize devices, circuits, and processes; and enable co-design, simulation, and heterogeneous integration

Source: A Vision and Strategy for the NSTC.pdf, April 25, 2023

Figure 46 Technologies reverse engineered from products' requirements by IRDS and adopted by NSTC via IAC.

9. CONCLUSIONS

The electronics industry has undergone 4 major crises (including the present one) in the past 60 years in which governments' interventions around the world have been very important contributors in funding viable and innovative solutions.

1. The Sputnik crisis of 1957 triggered major USG investments that funded the startup of Silicon Valley.
2. The demise of the US IDMs at the hand of the Japanese DRAM producers in 1985 triggered the formation of Sematech in 1989 supported by $200M of joint USG and 14 IDMs investments. The initial mission consisted in producing more advanced DRAMs than Japanese companies by 1993, but it was later redirected to refurbish the US suppliers' industry. The goal of 50/50 Japan/US equipment market share was accomplished by the mid-90s.
3. The CMOS crisis identified by the ITRS in 1998 shortly preceded the Nanotechnology revolution in the year 2000 that was supported by governments' funding that eventually reached the $1B level in the US followed by similar funding levels in Europe and Japan. These governments' initiatives became instrumental in funding and promoting the necessary technologies to refurbish the CMOS transistors. IMEC emerged as the international prototyping center of excellence in the past decade.
4. The Foundry/Fabless model emerged as the driver of the electronics industry in the past 15 years. Up to now system integrators have been very skilled at integrating in a single product a multitude of technologies already existing. However, the yearly volume of most of these successful products has already reached a plateau or has even declined in the past few years. Many ideas about exciting new products have been proposed in many public presentations but, in most cases, the required technologies are not yet available or are seating at the bottom of the valley of death presently existing between technology providers and system integrators—***who is responsible for developing them? Can the Chips Acts provide funds and support to solve these problems?***
5. Governments around the world have already committed large number of investments in semiconductors aimed at increasing domestic capacity after the realization during the pandemic that semiconductors have penetrated every aspect of life. This goal may not be realized in the way most people expected but international leaders in semiconductor manufacturing have already committed to the construction of brand-new plants in Europe, Japan, and the US. Additional plants are under study.
6. The imminent and most important challenge consists in developing technologies in support of prototyping building blocks that are not available at present but are badly required by system integrators for new products.

10. Acronyms and Abbreviations

Term	Definition
2D	2-dimensional
3D	3-dimensional
BEA	Bureau of Economic Analysis
BGA	Ball grid array
BSCON	Back-side device contacts
BSPD	Back side power distribution
CFET	Complimentary field effect transistor
CMOS	Complementary metal-oxide-semiconductor
CPP	Core Partners Program
DARPA	Defense Advanced Research Project Agency
DOD	Department of Defense
DOE	Department of Energy
DRAM	Dynamic random access memory
EDA	Electronic design automation
EUV	Extreme ultraviolet
FinFET	Fin-shaped field-effect transistor
FPGA	Field-programmable gate array
GAA	Gate all around
GHz	Gigahertz
I/O	Input/output
IAC	Industry Advisory Council
IC	Integrated circuit
ICOS	International Cooperation on Semiconductors
IDM	Individual device manufacturers
IEEE	Institute of Electrical and Electronics Engineers
imec	Interuniversity Microelectronics Centre
IP	Intellectual property
ITRS	International Technology Roadmap for Semiconductors
LED	Light emitting diode
LETI	Laboratory of Electronics and Information Technologies
LGAA	Lateral gate all around
M1	Metal 1
MCM	Multi-chip module
MEMS	Micro electro-mechanical systems
MHz	Megahertz
mm	Millimeter
MOS	Metal-oxide-semiconductor
MPU	Microprocessor unit
Mx	Metal layer
NA	Numerical aperture

NAND	A type of gate; the abbreviation for "NOT AND;" non-volatile memory with NAND type memory cell"
NASA	National Aeronautics and Space Administration
NCI	National Cancer Institute
NIN	Nanotechnology Infrastructure Network
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
nm	Nanometer
NNI	National Nanotechnology Initiative
NSTC	National Semiconductor Technology Center
NTRS	National Technology Roadmap for Semiconductors
NVM	Non-volatile memory
PC	Personal computer
P-N	Positive-negative
PoP	Package on package
PPAC	Performance, power, area, cost
R&D	Research and development
SCR	Super Clean Room
SDRJ	Systems and Devices Roadmap of Japan
SIA	Semiconductor Industry Association
SiNANO	SiNANO Institute, European Academic and Scientific Association for Nanoelectronics
SOC	Silicon on chip
SRAM	Static random access memory
SRC	Semiconductor Research Corporation
TSMC	Taiwan Semiconductor Manufacturing Company
TSV	Through silicon via
USG	United States Government
VLSI	Very large scale integration
WPM	Wafers per month
WSC	World semiconductor council

11. REFERENCES

[1] K. Rim, J. Welser, J.L. Hoyt, and J.F. Gibbons, "Enhanced hole mobilities in surface-channel strained-Sip-MOSFETs," in IEDM Tech. Dig., 1995,p.1026.

[2] 1987_Balestra_IEEE-EDL_VI-1987

[3] D. Hisamoto; Wen-Chin Lee; J. Kedzierski; H. Takeuchi; K. Asano; C. Kuo; E. Anderson; Tsu-Jae, FinFET-a self-aligned double-gate MOSFET scalable to 20 nm, Published in: IEEE Transactions on Electron Devices, Volume: 47, December 2000

[4] Xuejue Huang, Wen-Chin Lee, Charles Kuo, Digh Hisamoto, Leland Chang, Jakub Kedzierski, Erik Anderson, Hideki Takeuchi, Yan-Kyu Choi, Kazuya Asano, Vivek Subramanian, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu, "Sub 50-nm FinFET: PMOS", IEDM 99-67.

[5] Gordon Moore papers, M1965. Dept. of Special Collections, Stanford University Libraries, Stanford, Calif. Box 3, folder 13;

[6] Electronics, Volume 38, Number 8, April 19, 1965