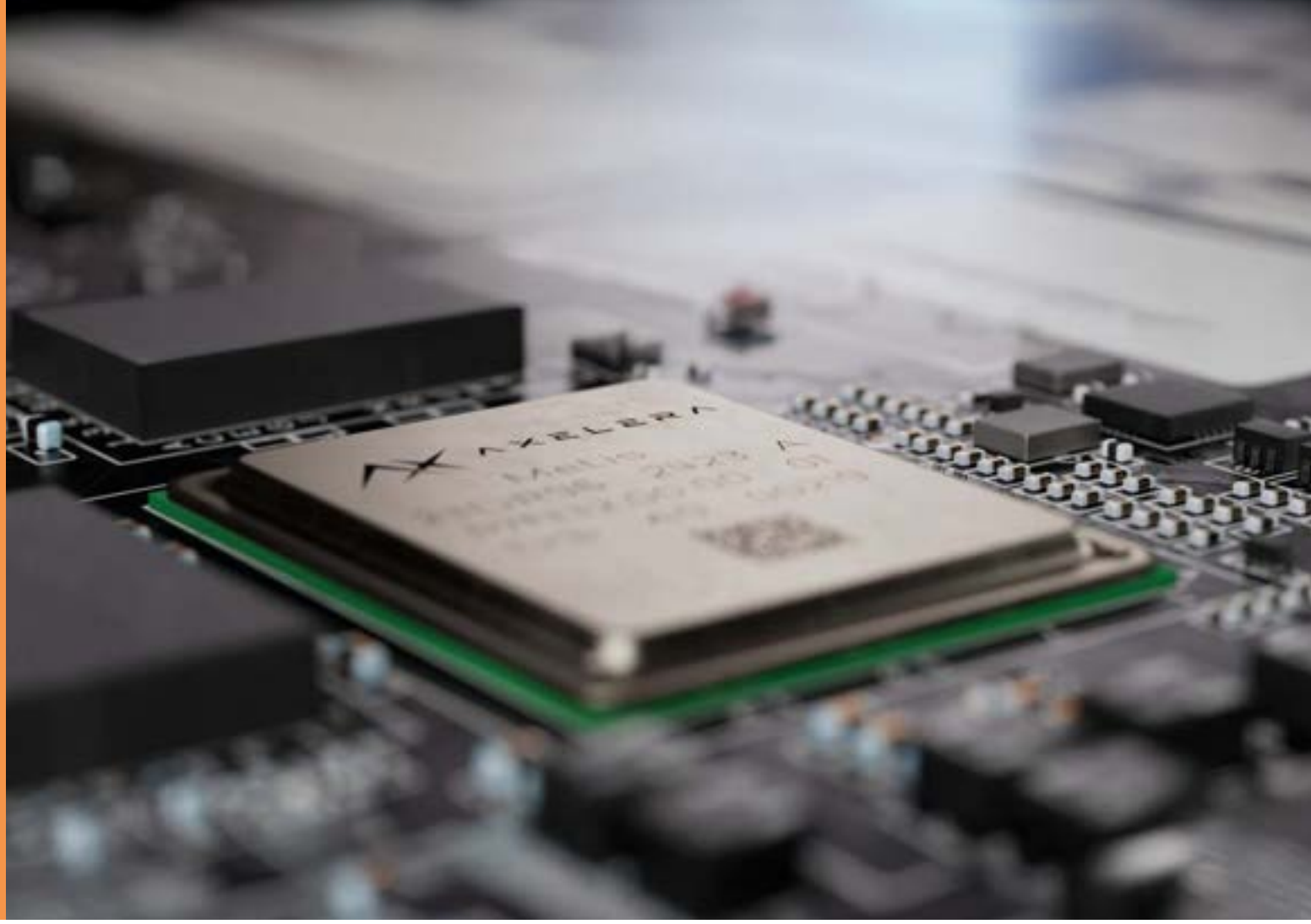
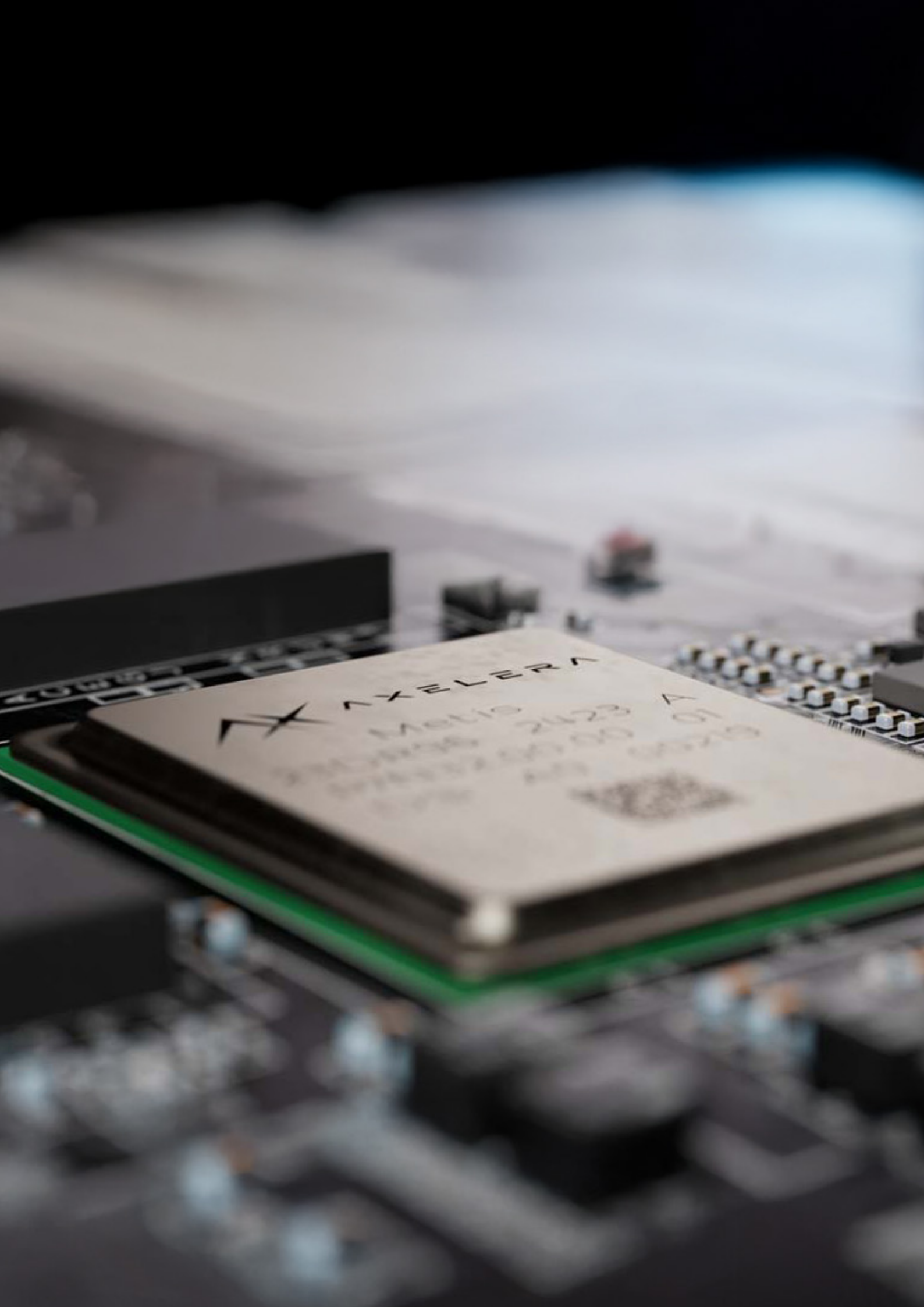


Semicon Deep Dive: Setting New Ambitions

Invest-NL Deep Tech Fund
April 2025

Deep Tech

INVESTNL


AXELERA

Contents

Introduction

The semiconductor industry is crucial for Europe's strategic independence and global competitiveness. It forms the backbone of various high-tech sectors, including automotive, telecommunications, defense, and artificial intelligence. Additionally, semiconductor technologies are vital for driving essential societal changes in energy, healthcare, and digitization.

At the same time, the European Semiconductor industry face some major challenges. Draghi points out in his report that while the EU has strengths in certain areas of the chips market, it also heavily relies on non-EU players and lacks a strong presence in high-value, innovative segments. All this against the background of growing energy demand of chip devices and an industry showing it increasingly difficult to stay on the path as envisioned by Gordon Moore in 1965.

Given the strategic Dutch position in the field of semiconductors, the Netherlands wants to play a leading role in the European Semicon Coalition at the initiative of the Minister of Economic Affairs. The ChipNL Innovation Program supports this initiative with a proposal for an industry-wide supported innovation program to strengthen our competitiveness in a future-proof manner.

With this report we want to contribute to the discussion to strengthen the international competitiveness of the Dutch and thus also the European Semiconductor sector. Invest-NL believes in a new era where the Dutch and European Semiconductor industry can design and manufacture advanced chips by mix-and-match advanced and trailing-edge chiplets made possible by novel technologies such as heterogeneous integration for which metrology & inspection is becoming increasingly important.

We also see a lot of bright new ideas at our research institutes & universities and startups coming up with new IC architectures, measuring techniques and other innovative technologies. Examples are for instance Axelera AI, Innatera or Qualinx all having designed competitive new chips reducing energy consumption while at the same time maintaining performance. Or Nearfield Instruments pushing the boundaries of metrology & inspection.

The Deep Tech Fund is strongly committed on the financing side to help Dutch semiconductor startups scale up, but we do notice that especially funding large investment rounds – as is common practice in the capital-intensive semiconductor industry – is difficult to achieve in Europe without foreign capital.

Best regards,

Gert-Jan Vaessen
Fund manager Invest-NL Deep Tech Fund

Executive summary

The Netherlands is a leading player worldwide in the field of semiconductors. The Netherlands plays an important role in Europe, particularly in the field of advanced production machines, but also in research, design and the production of chips. Additionally, the Dutch sector is also home to 49 startups in the traditional semiconductor sector, and this number amounts to 81 when incorporating emerging areas like Photonics and Quantum technologies. This strong presence places the Netherlands among the top European countries for Semiconductor startups.

The European Semiconductor industry also faces challenges. Mario Draghi in his recent report[1] pointed out that while the EU has strengths in certain areas of the chips market, it heavily relies on non-EU players and lacks a strong presence in high-value, innovative segments. The EU Chips Act aims to strengthen the semiconductor ecosystem and increase the EU's global manufacturing share to 30% by 2030. But Draghi warned that in the next decade the most value in the global chips sector will be captured by those with strong design and architectural capabilities or with research and innovation scale in advanced manufacturing. All this against the background of growing energy demand of chip devices and an industry showing it increasingly difficult to stay on the path as envisioned by Gordon Moore in 1965.

In 2021 TSMC's chairman Liu suggested a new version of Moore's Law is needed, focusing on Energy Efficient Performance (EEP): "IC's energy-efficient performance will continue to double every two years in the future". We show in this Semicon Deep Dive that the industry must address five challenges to stay on this path the next decade. In our view success will require co-innovation, collaboration and clear choices among all stakeholders. Invest-NL Deep Tech Fund proposes to guide these efforts by focusing on three promising growth areas: metrology, heterogeneous integration and new IC architectures.

Moore's Law will continue with EEP as key performance indicator. In our view improvement in EEP will be driven by four scaling engines: geometrical scaling, device scaling, circuit scaling and architecture scaling. It is our opinion that geometrical and device scaling[2] will continue as long as scaling will pay off for the most advanced digital chips. However, we also believe that novel developments such as circuit scaling and architecture scaling[3] will increasingly become important as innovative driving force the coming decade.

Circuit scaling is 3D scaling where geometrical and device scaling is 2D scaling. 3D scaling is implemented by combining chiplets in a 3D IC. Chiplet-based designs break various chip functions into distinct dies connected together, thus allowing the chip to function as one cohesive unit — albeit with trade-offs. Heterogeneous integration is the technology connecting the various chiplets in a functioning system. Besides connecting chiplets in conventional CMOS, an advantage of heterogeneous integration is the creation of crossovers: e.g. integrating mixed signal chiplets, optical components or even MEMS.

–

[1] Draghi report, 'The future of European competitiveness', September 2024

[2] Smaller geometries in the IC circuit layout respectively new transistor designs with less pixels per transistor

[3] Improvements how transistors and chips are connected respectively how transistors are used to perform calculations

The benefits of chiplets are numerous; (i) enhanced transistor density; (ii) enhanced performance and scalability; (iii) enhanced energy efficiency; (v) faster time-to-market & design flexibility; (iv) and lower manufacturing costs. In an example, we have calculated the cost price of a high performance 5nm monolithic chip versus the very same chip segmented into chiplets and connected through hybrid bonding. Our finding is that chiplet-based architecture is almost 20% less expensive which is similar to shrinking by 1-2 nodes in the IDRS[4] roadmap.

The yield in hybrid bonding of chiplets is key as rework is not possible when one die fails. Therefore we believe Metrology & Inspection will play an even more important role in packaging than in front-end processing to maintain cost savings. Shrinking geometries, new packaging techniques such as chip-on-wafer and wafer-on-wafer bonding and more complex IC designs all require high resolution, high accuracy metrology tools preferably at high throughput.

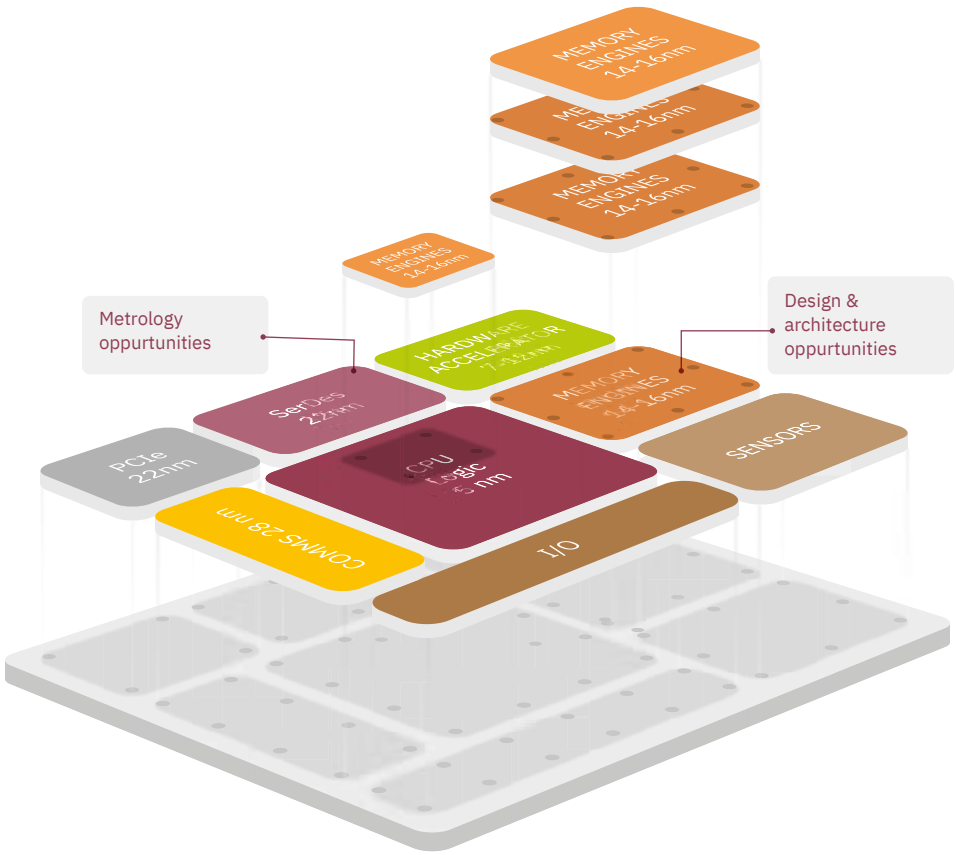
We believe in the new era in which the Dutch and European semiconductor industry can design and manufacture advanced chips by mix-and-match chiplets manufactured by established global semicon leaders such as TSMC, Samsung and Intel combined with locally produced trailing-edge chiplets which fulfil a specific function. And if we are optimistic: chiplets might even enable new IC architectures providing a path to the creation of more startups as financing becomes less challenging. For instance, chiplets lower the design costs of IC's significantly (overall design costs of a 28nm IC amount to over USD 50m versus USD 540m for the 5nm node) and trailing edge fabs are significantly less expensive than leading-edge fabs.

On balance we envision that the Dutch and European semiconductor industry can thrive by designing and manufacturing advanced chips. By combining chiplets from global leaders like TSMC, Samsung, and Intel with locally produced chiplets tailored for specific functions, we can achieve new high performing and efficient IC architectures. By pushing the boundaries of enabling technologies such as metrology & inspection and heterogeneous integration we can achieve remarkable innovation. These technologies will pave the way for fostering the emergence of numerous Dutch and European startups and driving technological progress forward along the path of EEP.

–

[4] International Roadmap for Devices and Systems (https://irds.ieee.org/editions)

[5] https://techovedas.com/cowos-tsmcs-new-secret-weapon-for-advanced-packaging/



Heterogeneous integration of chiplets will drive Moore's Law forward
and offer additional valorisation opportunities for the Dutch semicon eco-system.

1

Goal and scoping

1. Goal and scoping

Semicon industry market leaders, TSMC, Intel and Samsung, expect Moore's Law to continue in the coming decades when expressed as “Energy Efficient Performance” (EEP). EEP progress is powered by geometrical, device, circuit and architecture scaling engines.

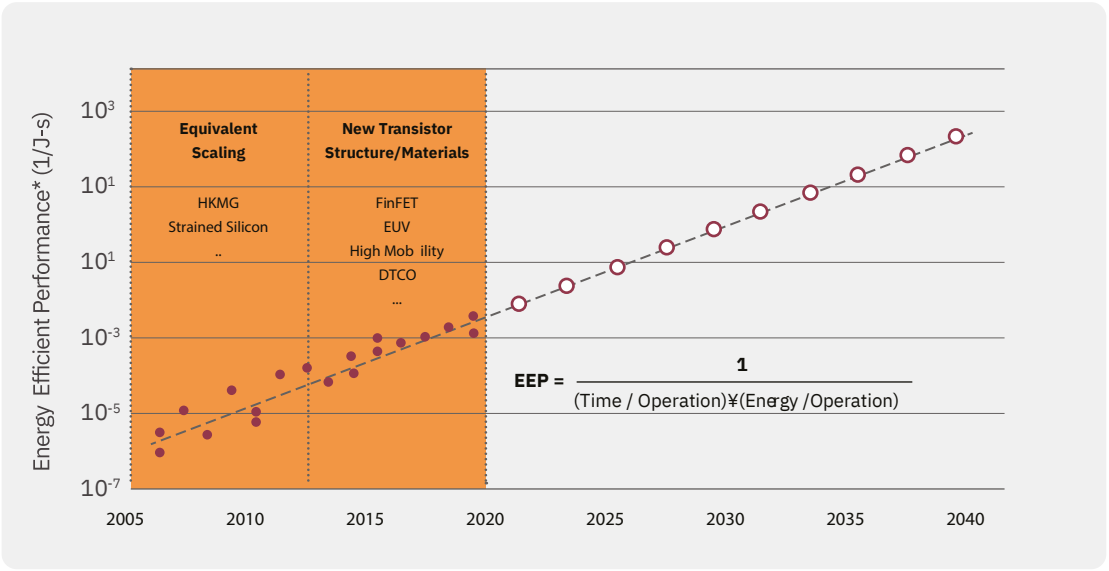
Invest-NL Deep Tech Fund believes that the relative importance of these engines will evolve over time driven by the semicon industry according to “bang for your buck”. This development offers opportunities for Dutch industry and start-ups. Our goal is to identify these promising growth areas for the Dutch semicon industry, as well as to formulate preconditions for successfully utilizing them through valorisation through entrepreneurship.

The semicon industry making Integrated Circuits (ICs) is a vast and enormous business. In 2024, the semicon industry is expected to reach over $650 billion. The semicon industry is also very innovative. The Economist, writing in 201611 about Moore’s Law, used an, often-repeated, analogy: “If cars and skyscrapers had improved at such rates since 1971, the fastest car would now be capable of a tenth of the speed of light; the tallest building would reach half way to the Moon.”

This report takes a top-down approach to identify these opportunities with focus on mainstream semicon markets and roadmaps such as memory and logic. It is expected that these opportunities address potential roadblocks for EEP development This leaves bottom-up opportunities in niche applications unaddressed. Furthermore, the traditional semicon market is expanding with novel segments using photonics and quantum technology.

The chosen approach leaves limited market uncertainty, as identified opportunities reflect the expressed needs of leaders in semicon such as TSMC, Intel and Samsung. This allows for a focussed approach while the market volume is large. Dutch semicon companies, having strong relations with these leaders, will benefit from addressing the EEP-roadblocks that will enhance their business. The goal is an even stronger and sustainable Dutch semicon eco-system.

In the next sections the semiconductor roadmap based on Moore’s Law will be discussed followed by an assessment of identified opportunities.



The future of Moore's Law according to TSMC.[6]

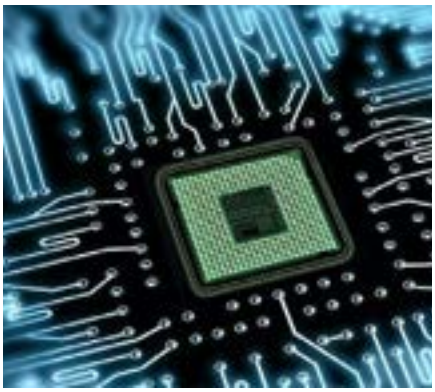
[6] https://www.youtube.com/watch?v=J5QVKCjyNG0

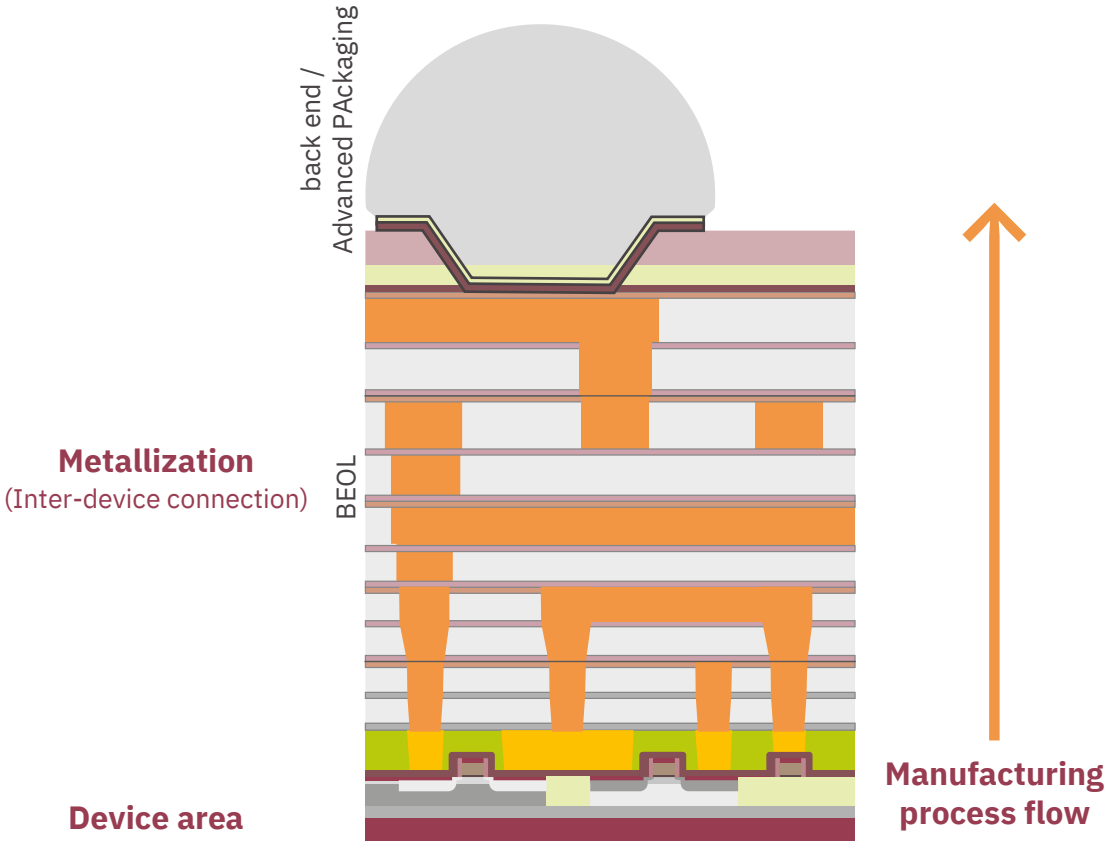
2

Introduction Integrated Circuits (ICs)

2. Introduction Integrated Circuits (ICs)

We're seeing the digital transformation of everything. Semiconductor Integrated Circuits are in almost everything: computers, cars, home appliances, medical equipment, etc. Semiconductor companies will sell over $ 650 billion worth of chips this year.





Integrated Circuit.[7]

2.1 The Semicon Ecosystem[8]

Companies in the semiconductor ecosystem design and make chips. They sell them to companies that then use these chips in systems and devices (e.g. iPhones, PCs, airplanes, cloud computing, etc.) that are sold to consumers, businesses, and governments. The revenue of products that contain chips is worth tens of trillions of dollars. The semicon industry is a business that manipulates materials an atom at a time on wafers that have many trillions transistors on it.

Looking at the semicon ecosystem one sees seven different types of companies. Each of these distinct industry segments feeds its resources up the value chain to the next until finally a chip factory (a "Fab") has all the designs, equipment, and materials necessary to manufacture a chip. Taken from the bottom up these semiconductor industry segments are:

- **Chip Intellectual Property (IP) Cores.** The design of a chip may be owned by a single company, or some companies license their chip designs – as software building blocks, called IP Cores – for wide use. There are over 150 companies that sell chip IP Cores. For example, Apple licenses IP Cores from ARM as a building block of their microprocessors in their iPhones and Computers.
- **Electronic Design Automation (EDA) Tools**. Engineers design chips (adding their own designs on top of any IP cores they have bought) using specialized Electronic Design Automation (EDA) software. The industry is dominated by three U.S. vendors– Cadence, Mentor (now part of Siemens) and Synopsys.
- **Specialized Materials**. So far, our chip is still in software. But to turn it into something tangible we are going to have to physically produce it in a chip factory called a "fab." The factories that make chips need to buy specialized materials and chemicals such as Silicon wafers, photomasks and over 100 different gasses.
- **Wafer Fab Equipment (WFE)**. These machines physically manufacture the chips. Five companies dominate the industry – Applied Materials, KLA, LAM, Tokyo Electron and ASML. These are some of the most complicated (and expensive) machines on earth.
- **"Fabless" Chip Companies**. Systems companies (Apple, Qualcomm, Nvidia, Amazon, Facebook, etc.) that previously used off-the-shelf chips now design their own chips. They create chip designs (using IP Cores and their own designs) and send the designs to "foundries" that have "fabs" that manufacture them.
- **Integrated Device Manufacturers (IDMs)**. They design, manufacture (in their own fabs), and sell their own chips. They do not make chips for other companies. There are three categories of IDMs– Memory (e.g. Micron, SK Hynix), Logic (e.g. Intel), Analog (TI, Analog Devices).
- **Chip Foundries**. Foundries make chips for others in their "fabs". They design unique processes using this equipment to make the chips, but they don't design chips. TSMC in Taiwan is the leader in logic, Samsung is second.
- **Outsourced Semiconductor Assembly and Test (OSAT)**. Companies that package and test chips made by foundries and IDMs. OSAT companies take the wafer made by foundries, dice (cut) them up into individual chips, test them and then package them and ship them to the customer.

–

[7] https://news.skhynix.com/semiconductor-front-end-process-episode-2/
https://www.automotivelogistics.media/tier-suppliers/clepa-calls-for-eu-strategy-to-tackle-semiconductor-shortage/41992.article

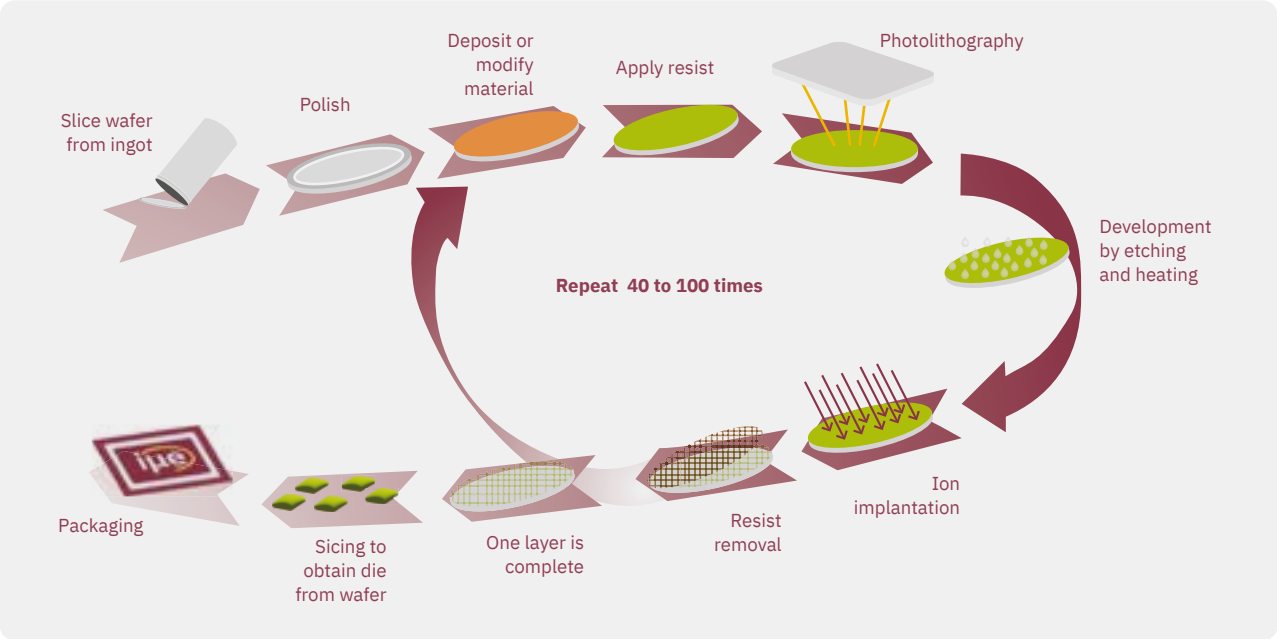
[8] https://steveblank.com/2022/01/25/the-semiconductor-ecosystem/


TNO

2. Introduction Integrated Circuits (ICs)

2.2 Manufacturing of Integrated Circuits

Integrated Circuits or chips are probably the most complicated products ever manufactured. The diagram below is a simplified version of the 1000+ steps it takes to make a chip. As chips have become denser (with trillions of transistors on a single wafer) following Moore's Law. The cost of building fabs, that enable denser chips is now >$10 billion for one chip factory.



Semiconductor production process.[9]

[9] https://www.iue.tuwien.ac.at/activities/process-simulation

The process of creating semiconductor integrated circuits can be broken down into several key steps[10]. The first step is wafer preparation. A silicon wafer is chosen as the starting material for the semiconductor process. The wafer is cleaned, polished, and prepared to be used as a substrate for the creation of the electronic components.

The second step is patterning. In this step, a pattern is created on the silicon wafer using a process called lithography. A thin layer of photoresist is applied to the surface of the wafer. Then a mask, with a single layer of the chip, is imaged on top of the wafer. The areas of the photoresist that were exposed to the light are then removed, leaving behind a patterned surface on the wafer.

The third step is doping. In this step, impurities are added to the silicon wafer to change its electrical properties. The most common type of impurity used is boron or phosphorus, which are added in small amounts to create either p-type or n-type semiconductors, respectively. These impurities are added using a process called ion implantation, in which ions are accelerated to high speeds and then implanted into the surface of the wafer.
The fourth step is deposition. In this step, thin films of material are deposited on the wafer to create the electronic components. This can be done using a variety of techniques, including chemical vapor deposition (CVD), physical vapor deposition (PVD), and atomic layer deposition (ALD). These processes can be used to deposit materials such as metals, oxides, and nitrides.

The fifth step is etching. In this step, material is removed from the surface of the wafer to create the desired shape and structure for the electronic components. Etching can be done using a variety of techniques, including wet etching, dry etching, and plasma etching. These processes use chemicals or plasma to selectively remove material from the wafer.

The final step is packaging. In this step, the electronic components are packaged into a final product that can be used in electronic devices. This typically involves testing each chip to ensure that it meets the required specifications, then separating it from the wafer and mounting it onto a package or substrate.

–

[10] https://www.renesas.com/en/blogs/semiconductor-device-manufacturing-process-challenges-and-opportunities

3

Semicon market

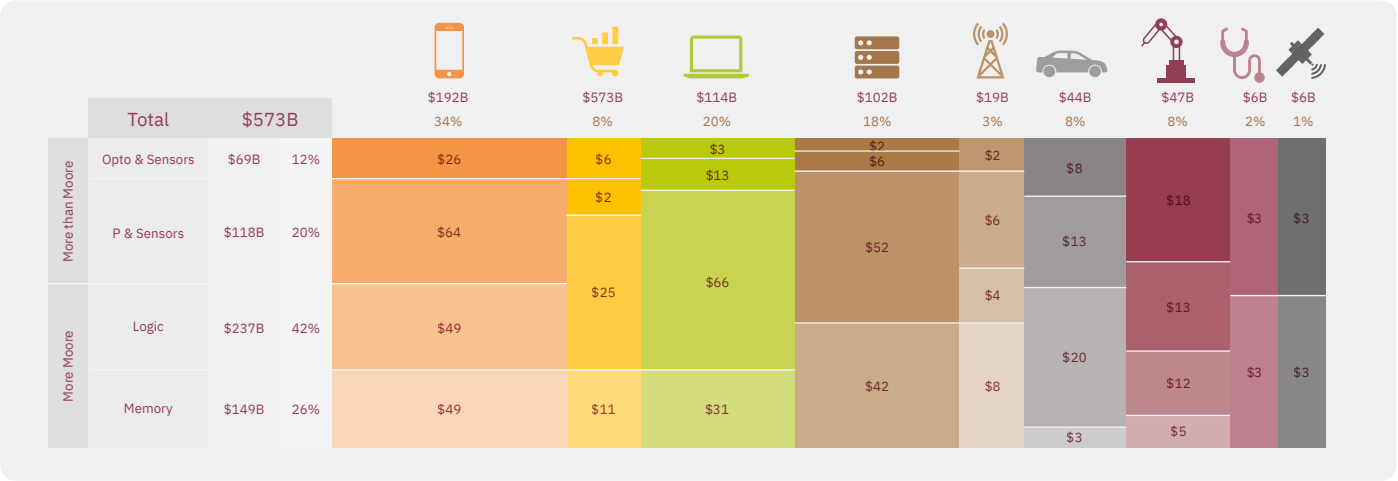
3. Semicon market

The semiconductor ecosystem has considerable means to drive innovation. According to ASML[11] the sector as a total generated $ 865 billion EBIT in 2023 and it reinvests around half of its EBIT to drive long term growth. Overall, the economic profit of the semicon industry has increased significantly over the last 20 years including that of equipment suppliers and fabless companies. Over this period the dominant position of TSMC is clear, not just from a technical perspective but also from an economical perspective.

Market capitalization gives us a glimpse of the future: the world believes that Artificial Intelligence is the next big thing in semicon with US companies accounting for 71.5% of the total industry's market capitalization end 2024[12] with Nvidia and Broadcom at the helm.

3.1 The main types of chips and their applications[13]

When looked according to functionality and type of chips, the main categories of semiconductors are (digital) memory chips, (digital) microprocessors, analog, mixed signal, MEMS, photonic chips and quantum chips. The digital chips represent the largest categories.



Semiconductor device industry – 2022 key figures by application and device family.[14]
Source: Overview of the Semiconductor Device Industry 2023, Yole Intelligence, July 2023

- Mobile
- Consumer
- Computing
- Servers
- Telecom
- Automotive
- Industrial
- Medical
- Defence & Aerospace

–

[11] https://edge.sitecorecloud.io/asmlnetherlaaea-asmlcom-prd-5369/media/project/asmlcom/asmlcom/asml/files/investors/investor-days/2024/02global-market-trends-industry-technology-roadmap-esg--christophe-fouquet.pdf

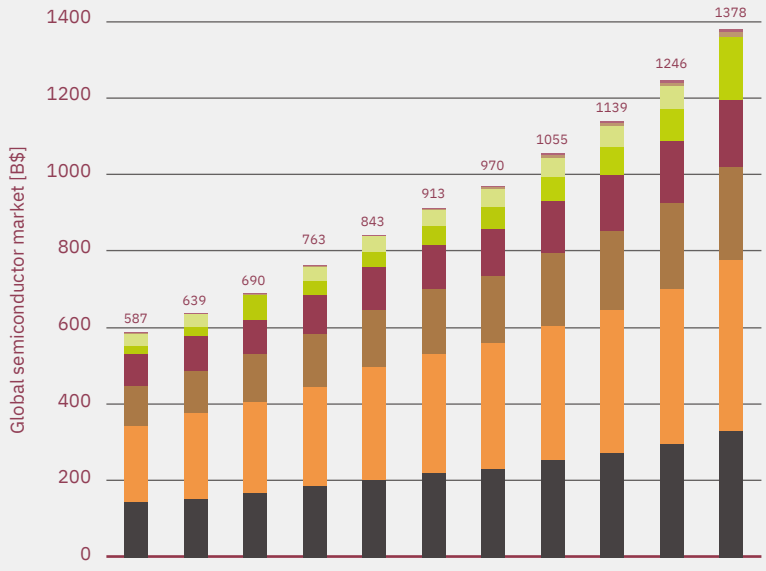
[12] Graphic shows Nvidia, US domination of chip industry ...

[13] https://www.investopedia.com/ask/answers/042115/what-are-main-types-chips-produced-semiconductor-companies.asp

[14] https://www.yolegroup.com/product/report/overview-of-the-semiconductor-devices-industry-2023/

According to a report from Market.us[15], the global semiconductor market is expected to grow significantly, reaching $690 billion in 2024, and the compound annual growth rate between 2023 and 2032 will reach 8.8%. By 2032, the semiconductor market is expected to grow to over $1.3 trillion.

The largest segments remain digital: memory, microprocessors and logic. The fastest growing segments are photonics (CAGR 23%) and quantum technology (CAGR 25%), but their volume is very limited.



	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
Quantum	1	1	1	1	2	2	3	4	5	5	8
Photonics	1	2	2	3	3	4	5	6	8	9	11
MEMS	29	31	34	36	40	43	46	50	55	59	64
RF	24	27	31	36	41	48	55	64	74	86	101
Analog	80	86	91	101	108	116	124	136	144	158	172
Logic	106	116	123	134	150	165	176	190	206	219	242
MPU & MCU	200	220	236	260	292	312	325	350	374	407	450
Memory	146	156	172	191	207	223	236	256	275	301	331

Global semiconductor market by component type
Sources: Market.us, The Business Research Company, Ameco, Roots Analysis, KBV Research, Global Market Insights, Fortune Business Insights

–

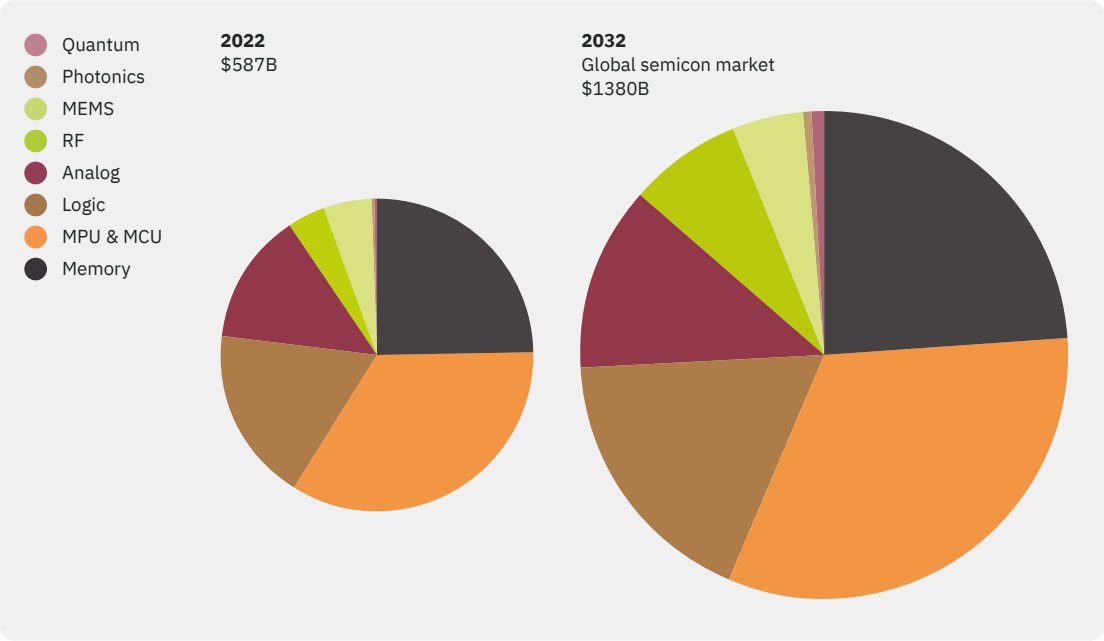
[15] Market: Six factors driving global semiconductor market - SemiMedia https://www.semimedia.cc/16039.html

3. Semicon market

3.1.1 Memory Chips

From the perspective of functionality, semiconductor memory chips store data and programs on computers and data storage devices.

Random-access memory (RAM) chips provide temporary workspaces, whereas flash memory chips hold information permanently unless erased. Read-only memory (ROM) and programmable read-only memory (PROM) chips cannot be modified. In contrast, erasable programmable read-only memory (EPROM) and electrically erasable read-only memory (EEPROM) chips can be changed.



3.1.2 Logic and Microprocessors

Microprocessors contain one or more Central Processing Units (CPUs). Computer servers, personal computers (PCs), tablets, and smartphones may each have multiple CPUs based on Von Neumann architecture.

The 32- and 64-bit microprocessors in PCs and servers today are based on x86, POWER, and SPARC chip architectures, first developed decades ago. On the other hand, mobile devices like smartphones typically use an ARM chip architecture. Less powerful 8-, 16- and 24-bit microprocessors (called microcontrollers) turn up in products such as toys and vehicles.

Technically a type of microprocessor, Graphics Processing Unit (GPU) is capable of rendering graphics for display on an electronic device. When used in conjunction with a CPU, a GPU can increase computer performance by taking on some computationally intensive functions, such as rendering, from the CPU. This accelerates how quickly applications can process since the GPU can perform many calculations simultaneously. This shift also allowed for the development of more advanced and resource-intensive software and activities such as Artificial Intelligence (AI).

3.1.3 Commodity ICs and ASICs

Commodity integrated circuits (CICs), are simple chips used for simple chips used for performing repetitive processing routines. Produced in large batches, these chips are generally used in single-purpose appliances such as barcode scanners. Characterized by razor-thin margins, the commodity IC market is dominated by large Asian semiconductor makers. If an IC is made for a specific purpose, it is called an ASIC, or application-specific integrated chip. For example, bitcoin mining today is accomplished with ASICs that only do that one function: mining. Field-programmable gate arrays (FPGA chips) are another type of commoditized IC that can be customized to a manufacturer's specifications.

3.1.4 Analog and mixed signal chips

In analog chips, voltage and current vary continuously at specified points in the circuit in contrast to digital chips. An analog chip typically includes transistors along with passive elements such as inductors, capacitors, and resistors. Analog chips are more prone to noise, or small variations in voltage, which can cause errors. Analog chips are required for wideband signals, and they are used as sensors and in power supplies.

Mixed circuit semiconductors are typically digital chips with added technology for working with both analog and digital circuits. A microcontroller might include an analog-to-digital converter (ADC) for connecting to an analog chip, such as a temperature sensor, for example. A digital-to-analog converter (DAC), conversely, can allow a microcontroller to produce analog voltages for making sounds through analog devices.
Analog chips[16] are integral to various industries such as manufacturing, automotive, IoT, and consumer devices. In recent years, the analog chip market has experienced notable consolidation, with the top companies gaining a larger share. Reasons why analog chips are important in our digital world are (a) Signal Processing, (b)Real-world Interface, (c) Power Efficiency, (d) Smooth Transitions, (e) Amplification and Filtering, (f) Communication Systems, (g) Sensor Integration.

Over the past couple of decades, the analog market has witnessed a steady consolidation trend. Since 2013, the top five analog companies (NXP, STMicroelectronics, Infineon, Analog Devices and Texas Instruments) have grown their market share from 40.4% to 62.9%. Several factors contribute to this consolidation: Commoditization, Economies of Scale, Industry Partnerships.

Texas Instruments stands out as the largest analog chip provider globally, holding a prominent position in the market. TI's revenue is predominantly generated from signal conversion chips (77%), covering sound, temperature, and pressure. The remaining 23% comes from embedded processing chips used in a range of electronics from simple devices to complex automotive applications. TI's strategic focus on capital allocation, including stock buybacks and dividends, has further solidified its standing in the industry.

3.1.5 Radio Frequency (RF) chips

Devices or modules known as RF power semiconductors are used to increase the signal strength of a radio frequency and used in telecommunication, automotive, consumer electronics, aerospace & defence, among other industries. Vehicles frequently use radio frequency semiconductors with millimetre-wave radar frequencies for both short- and long-range applications.

–

[16] https://techovedas.com/what-are-major-players-in-analog-chip-market/

3. Semicon market

The Global Radio Frequency Semiconductor Market Size valued for $20 billion in 2021 and is anticipated to reach $39 billion by 2030[17]with a CAGR of 8% from 2022 to 2030. The Business Research Company[18] expects even a steeper growth with CAGR as large as 19%.

3.1.6 Micro-electromechanical systems (MEMS)

MEMS[19] (micro-electromechanical systems) is the technology of microscopic devices incorporating both electronic and moving parts. MEMS are made up of components between 1 and 100 micrometres in size (i.e., 0.001 to 0.1 mm), They usually consist of a central unit that processes data (an integrated circuit chip such as microprocessor) and several components that interact with the surroundings (such as microsensors).

The global MEMS market is segmented[20] into various types of devices, such as accelerom-eters, gyroscopes, inkjet heads, microfluidics and optical MEMS.

Globally the MEMS market is expected to reach $64 billion by 2032, at a CAGR of 8.4% during the forecast period 2022 to 2032[21]. The rising requirement for MEMS in different markets that includes the electronics industry, automotive, manufacturing, and healthcare is driving the MEMS market. Furthermore, the expanding use of radio-frequency MEMS systems, the rising need for Internet of Things (IoT) devices, and the increasing demands for consumer electronics are importa nt factors driving the MEMS market.

The major players in the global MEMS market include TDK Corporation, Analog Devices, Panasonic Corp., NXP Semiconductors, STMicroelectronics, TE Connectivity, Robert Bosch, Texas Instruments, Broadcom, and others.

3.1.7 Silicon photonics market

The use of photonics to meet bandwidth requirements for data centres is encouraged by optics' ability to guarantee high-bandwidth data transfer. Because of this, the demand for next-generation performance in data centres can be satisfied by integrating silicon photonics technology with conventional, complementary metal-oxide semiconductor (CMOS) devices. Because of this, the market is expanding due to the demand for low-cost, high-speed interconnects that can sustain ever-increasing data rates over 100 Gbps. Other applications of integrated photonics are expected a.o. in healthcare.

The Global Silicon Photonics Market size is expected to reach $7.5 billion by 2030, rising at a market growth of 28.8% CAGR during the forecast period[22].

3.1.8 Quantum computing chips

Quantum computing is a subfield of computer

–

[17] https://www.amecoresearch.com/amp/reports-detailsytr-land.php?page=276757&slug=radio-frequency-semiconductor-market

[18] https://www.thebusinessresearchcompany.com/report/5g-radio-frequency-chip-rf-chip-global-market-report

[19] https://en.wikipedia.org/wiki/MEMS

[20] https://www.rootsanalysis.com/mems-market#:~:text=The%20global%20MEMS%20market%20is,%2C%20RF%20MEMS%2C%20and%20thermopiles.

[21] https://www.sphericalinsights.com/reports/mems-market

[22] https://www.kbvresearch.com/silicon-photonics-market/

science based on quantum theory. It is a cutting-edge technology that uses quantum mechanics to solve more difficult problems for classical computers and is rapidly developing. Quantum computers have made it possible to use concepts from quantum physics in computing. It differs from standard computing in terms of speed, data, and bits. Most of the time, the system is used to compare and select the best answer to a complex problem. Quantum computers can be utilized to create more exact and productive machine learning calculations utilized in applications such as picture and discourse acknowledgment. Quantum computer engineering is still in the nascent stage. Numerous technical obstacles are needed to overcome in order to implement quantum computing technology in real-time applications. Due to their extraordinary sensitivity to external interactions, quantum computers have the potential to cause the state function to collapse. Therefore, the uncertainty is the size of this segment is large:

- Global Market Insights[23]: Quantum Computing Market size was valued at $974 million in 2022 and is projected to record more than 10% CAGR from 2023 to 2032. This results in $2.5 billion in 2032.
- Fortune Business Insights[24]: The global quantum computing market size was valued at $885 million in 2023 and is projected to grow from $1.16 billion in 2024 to $12.6 billion by 2032, exhibiting a CAGR of 34.8% during the forecast period.

3.2 Semicon market trends

Market.us stated that the growth of data-intensive industries such as artificial intelligence, machine learning, and big data analysis has stimulated the demand for high-computing power and high-speed semiconductors. The development of the Internet of Things and 5G also requires more advanced and energy-saving semiconductors. The demand for sensors used in consumer equipment, biomedical applications, drones, robots and other fields also continues to rise. In addition, demand for semiconductors in the automotive field is also expected to grow rapidly.

3.2.1 Drivers for semiconductor market growth

Market.us lists multiple factors that are driving semiconductor market growth:

- **Technological advancements and continued innovation.** Chipmakers are focused on producing smaller, faster and more efficient semiconductors. Advances in artificial intelligence, the Internet of Things, autonomous vehicles and 5G networks bring opportunities and stimulate the pursuit of semiconductor solutions.
- **Growing demand for electronic equipment across industries.** Industries such as automobiles, healthcare, automation, and aerospace rely heavily on semiconductor components. In addition, the continued growth of the consumer electronics market and the widespread integration of electronic products in various applications have greatly promoted the expansion of the semiconductor industry.
- **Cloud computing and data centre expansion.** The growing popularity of cloud services has boosted demand for high-performance processors, memory chips and other semiconductor components.
- **Development of the automobile industry.** The development of electric vehicles, autonomous vehicles, in-car entertainment systems, and ADAS advanced driving assistance systems have increased the demand for semiconductors.

–

[23] https://www.gminsights.com/industry-analysis/quantam-computing-market

[24] https://www.fortunebusinessinsights.com/quantum-computing-market-104855

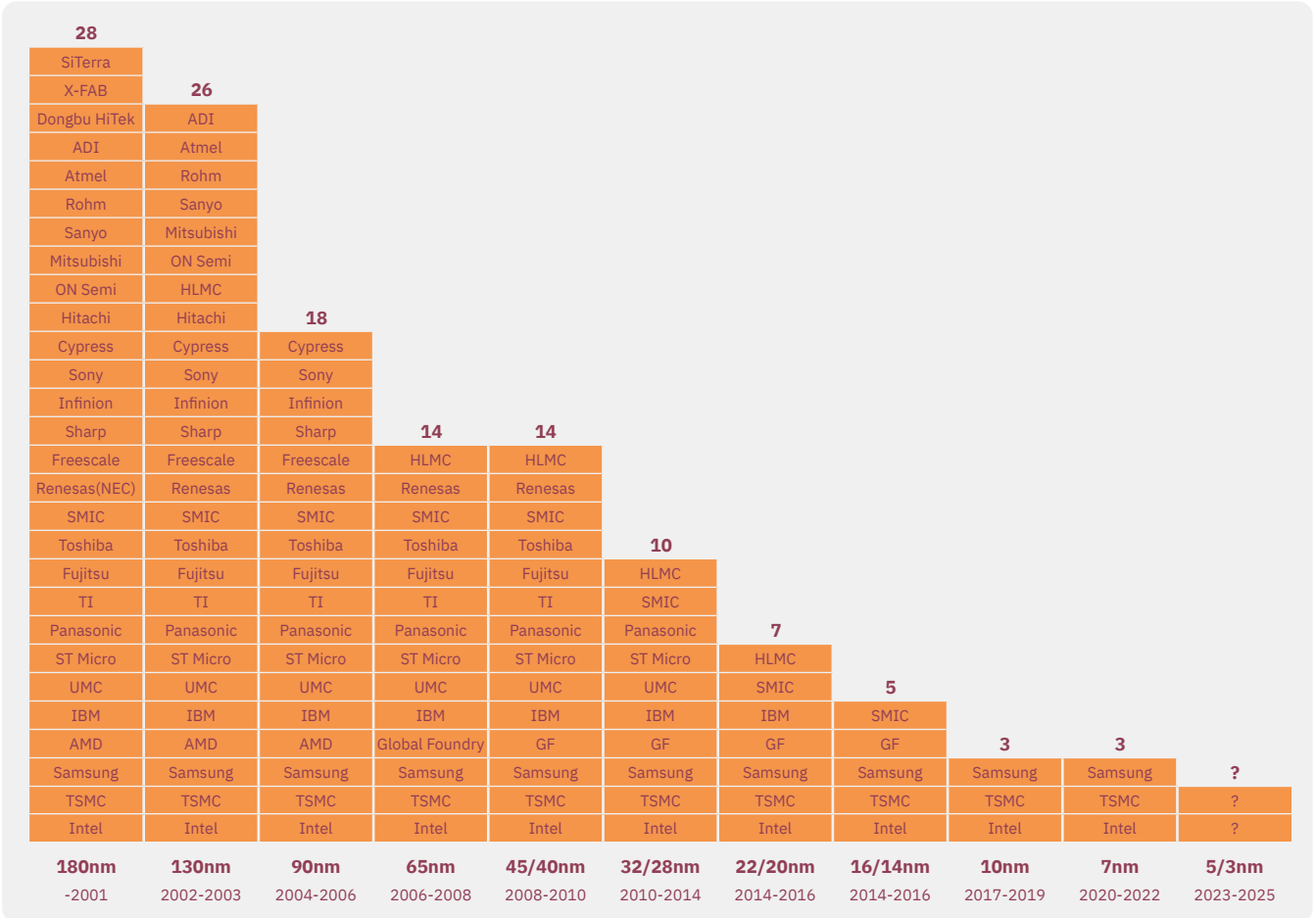
[24] https://www.fortunebusinessinsights.com/quantum-computing-market-104855

3. Semicon market

3.2.2 Consolidation

From a regional perspective, Market.us stated that the Asia-Pacific region dominated the global semiconductor market in 2023, with a share of more than 51.5%. North America also plays a key role in semiconductor design and innovation as home to practitioners and cutting-edge research institutions. The European market share is small, but it focuses more on semiconductors in the automotive and industrial fields.

TSMC Is now the world's largest semiconductor company in the world with total sales of $88 billion in 2024 which is much more than Intel's $53 billion in the same year. Samsung semiconductor revenues dropped to $67 billion, while NVIDIA's sales increased by 125%



Semiconductor industry evolution.[25]

[25] High-end performance packaging: 3D/2.5D integration report. Yole Développement 2020

to almost $61 billion in 2024.
Because process development costs and capital investment costs kept skyrocketing as circuit patterns were shrunk, semicon companies began to leave the competition one after another. Around 2002/2003, there were 26 semicon companies worldwide capable of manufacturing 130-nm devices. However, the number of companies continuing with scaling competition decreased[25]. For example, there were only 18 companies at 90 nm and only 14 companies at 45 nm. Beyond 10 nm, only Intel, Samsung, and TSMC remained. They are the custodians of Moore's Law.

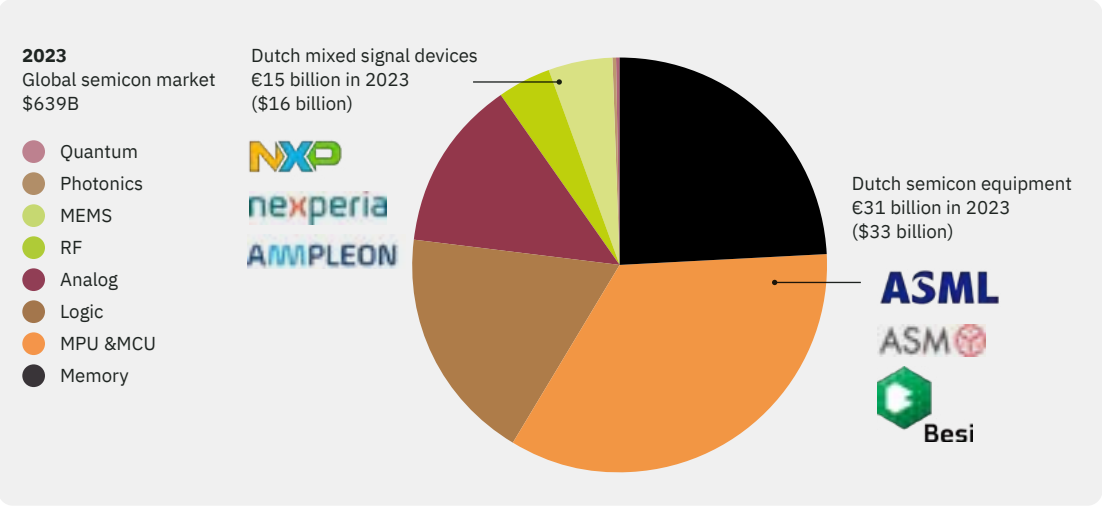
3.3 Overview Dutch Semiconductor market

3.3.1 The 'Big-6' are dominating the Dutch Semicon Landscape

The Netherlands is strongly represented in the traditional semiconductor industry in Europe, both in terms of 'going concerns' as well as start- and scaleups. The Dutch semiconductor ecosystem consists of more than 300 companies and organizations[27] with a total turnover of around EUR 57 billion in 2023 and approximately 135,000 FTEs. The Big-6 - ASML, NXP, ASM, Nexperia, Besi and Ampleon - represent approximately 75% of the total turnover[28] in the sector and is responsible for almost 70% of employment in the sector.

PWC[29] has shown that the semiconductor sector is very important for Dutch SMEs: 62% of companies have a turnover of <EUR 10m and 26% between EUR 10m - EUR 100m. In addition, the sector is a major driver of Dutch innovation capacity: the Big-6 account for EUR 6.8bn or over 45% of total R&D by companies in the Netherlands and account for 47% of all Dutch patent applications at the European Patent Office (EPO)[30].

Two Dutch Semicon companies are included in the list of top 20 Semiconductor companies worldwide: ASML and NXP which are ranked at number 8 respectively 18. ASM is ranked at nr. 42 and Besi at nr. 82. Furthermore, in the



–

[26] https://www.tel.com/museum/magazine/report/202106/
[27] Brainport development, RVO
[28] 2023 Annual reports

–

[29] PWC, Semicon in NL, May 2024
[30] 2023 annual reports, CBS Statline

3. Semicon market

Semiconductor sector a company's relative market position is important to capture economic profit. On this point the Dutch Big-6 score well:

- ASML is the global nr.1 in semi-equipment and nr. 1 in lithography with a ~83% market share[31] and being the sole supplier of EUV systems.
- NXP is the nr. 3 IC design and manufacturing company in Europe and it is nr. 2 in Automotive IC's worldwide with an 11% market share[32] and nr. 6 in Analog worldwide with a 4.7% market share.[33]
- ASM is market leader in ALD with a 55% market share.[34]
- Nexperia's global market in discrete components is 9.4%.[35]
- Besi is market leader in advanced packaging with over 35% market share.
- Ampleon is market leader in Radio Frequency power.

More detailed profiles for the Big-6 and its relative strengths are listed below.

- **ASML Holding NV (ASML)** is a microelectronics solutions provider that offers semiconductor manufacturing equipment. The company's product portfolio includes lithography systems, metrology and inspection systems and refurbished systems. Its computational lithography and patterning control software solutions enable customers to attain high yield and better operational performance. ASML revenue 2023: €27.6 billion[36]
- **NXP Semiconductors NV (NXP)** is a semiconductor manufacturing company that provides mixed signal and analog products. It offers radio frequency (RF), power management, security and digital processing products, microcontrollers, communication processors, analog and interface devices, connectivity chipsets, RF power amplifiers, sensors, security controllers and application processors. NXP revenue 2023: €12.2 billion[37]
- **ASM International NV (ASM)** is a supplier of semiconductor process equipment for wafer processing. The company offers atomic layer deposition products (ALD), epitaxy products, low-pressure chemical vapor deposition, and diffusion products, plasma-enhanced chemical vapor deposition, and plasma-enhanced ALDs. ASM revenue 2023: €2.6 billion[38]
- **Nexperia** is a semiconductor manufacturer headquartered in Nijmegen, the Netherlands. It is a subsidiary of Wingtech Technology, a Shanghai-listed company partially owned by the State-owned Assets Supervision and Administration Commission of the State Council. It has front-end factories in Hamburg, Germany, and Greater Manchester, England. It is the former Standard Products business unit of NXP Semiconductors (previously Philips Semiconductors). The company's product range includes bipolar transistors, diodes, ESD protection, TVS diodes, MOSFETs, and logic devices. Nexperia revenue 2023: €2.0 billion[39]
- **BE Semiconductor Industries NV (Besi)** manufactures, develops, markets and sells assembly equipment for global semiconductor and electronics industries. The company's product portfolio comprises die attach equipment, packaging equipment, and plating equipment. BESI revenue 2023:

–

[31] Seekingalpha.com, Khaveen investments

[32] Market Research IDC, 2024

[33] Statista, 2021

[34] Company info, investor relations

[35] Nexperia, Annual report 2021

[36] https://www.asml.com/en/investors/annual-report/2023

[37] https://investors.nxp.com/static-files/41300297-6aca-44d5-8753-8374b0ef8753

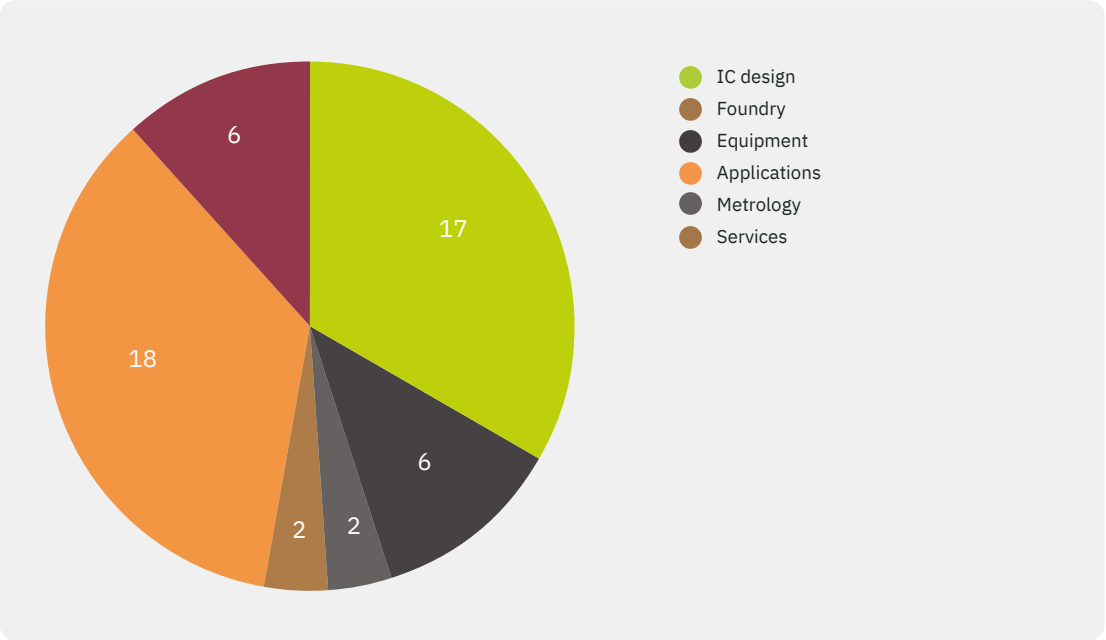
–

[38] https://www.asm.com/media/1fcpm1px/asm-annual-report-2023.pdf

[39] https://www.nexperia.com/about/news-events/press-releases/Nexperia-announces-2023-financial-results

[40] https://www.besi.com/investor-relations/press-releases/2023/details-1/be-semiconductor-industries-nv-announces-q4-23-and-full-year-2023-results/

Breakdown of Dutch Semicon startups



€579 million[40]

- **Ampleon** is a semiconductor manufacturer headquartered in Nijmegen, Netherlands and the former RF Power business division of NXP Semiconductors. Ampleon has offices in France, Finland, South Korea, the US, the Netherlands, Sweden, Taiwan and China.[13] In 2016, they opened their own package-assembly-test factory in the Philippines. They offer various LDMOS Power Transistors meant to be used primarily in aviation, and they develop 5G mmWave chips for 5G base stations. Ampleon revenue 2023: €425 million[41]

3.3.2 The Netherlands at the forefront of Semiconductor startups

Today, in the Netherlands there are 49 startups active in the 'traditional' Semiconductor market. If we also include novel market segments such as Photonics (17 startups) and Quantum Technology (15 startups), we even arrive at a total of 81[42] startups. In total, 24 startups in the Netherlands are engaged in the development of applications or services.
Without applications & services, there are 27 startups active in the following market segments: 17 in IC Design, 6 in Semiconductor Equipment, 2 in Metrology & Inspection and 2

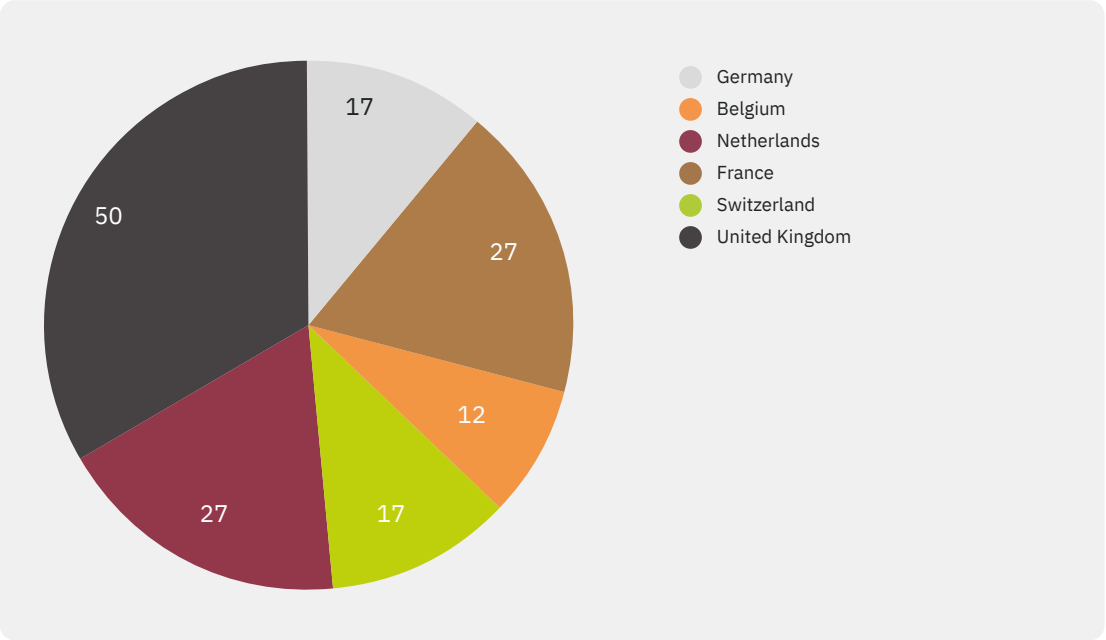
–

[41] https://www.zoominfo.com/c/ampleon-semiconductors-co/371949502

[42] Pitchbook and Brainport Development data analysis. Startup definition used is VC held companies with one or more deals completed

3. Semicon market

Semicon startups in West-Europe



startups are foundries.
IC design houses are important for the semiconductor ecosystem as companies capture 50% of the total added value in the global semiconductor value chain. Front-end and back-end factories generate 25% respectively 6% of the total added value[43]. Furthermore, if Europe wants to increase its production capacity to a global market share of 20%[44] it also needs to create future demand.

With 27 startups in traditional chip design, semi-equipment, metrology & inspection and foundry's the Netherlands is scoring relatively high in Western Europe especially on a per capita basis. In fact, although the USA has 257 startups in the same domain, on a per capita basis the Netherlands is generating two times more startups. Compared to Asia-Pacific this is even four times more. Nevertheless, given Asia-Pacific's significantly larger size this still amounts to 816 semicon startups in absolute terms. In Western Europe 142 startups are located in the same domain.
Our data analysis also reveals two challenges for the Dutch Semicon sector. First of all, the average funding round size for scale-ups (defined as funding rounds > €100 million) is significantly lower in Western Europe compared to North America and the Asia-Pacific area. In fact, in the Netherlands over the last decade the largest funding round was Nearfield Instruments with €135 million of which approximately half was funded by foreign CVC's. Furthermore, over the last decade only 4 deals larger than € 100m were closed in Europe.

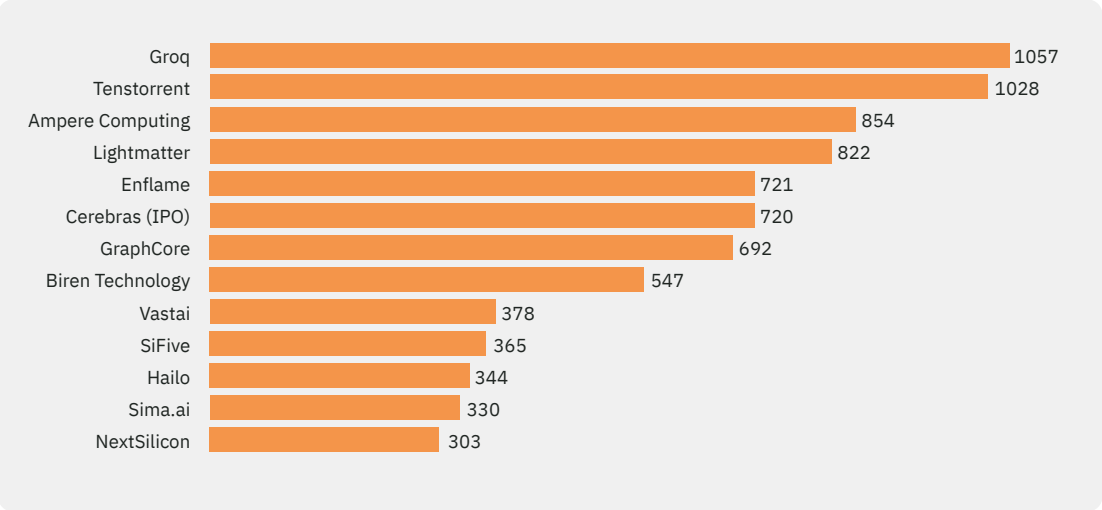
–

[43] Draghi, The Future of European Competitiveness, September 2024

[44] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en

Secondly, when we zoom in further, we see that there are two AI IC design startups in the Netherlands compared to over 150 AI hardware startups worldwide, most of which in the USA or China. 10% or 15 of these AI startups have raised over USD 300 million in funding in the last few years of which 10 has raised even more than USD 500m in funding[45]. Especially telling in our view is that of the 35 AI hardware startups with funding of more than USD 100m only 3 are located in Europe.

AI startups with funding >USD 300 million



AI startups with funding >USD 300 million

–

[45] Dr. Ian Cutress, More than Moore, CASPA AI Trends seminar, January 2025

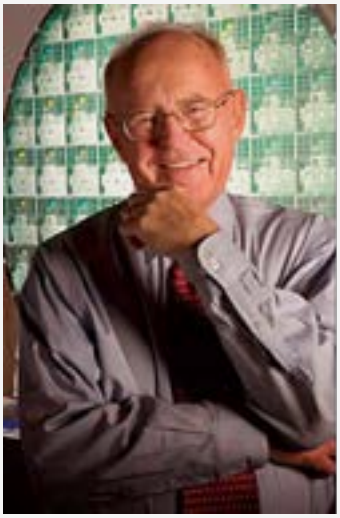
4

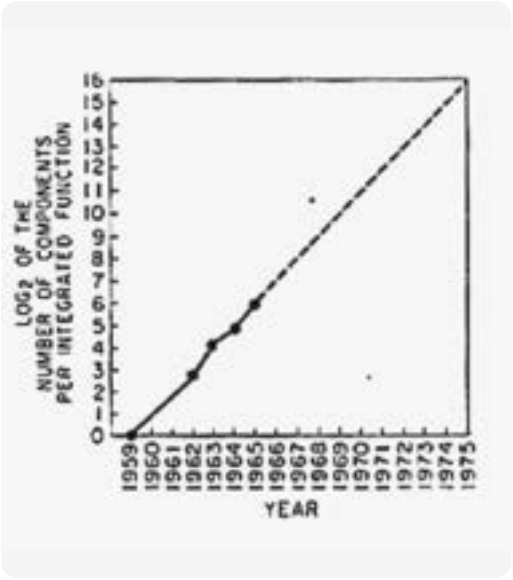
Past, Present, and Future of Moore's Law

4. Past, Present, and Future of Moore's Law

The empirical rule driving the development of the semicon industry over at least the past half-century is Moore's Law. Semiconductors have achieved astonishing development through increasing integration levels along Moore's Law and falling costs per transistor. In addition, Moore's Law has been functioning as an absolute anchor for the semicon industry to implicitly plan for predetermined harmony.

Despite the fact that it has been repeatedly declared that Moore's Law has reached its limit due to economic, technical, physical, and other factors, the law has tenaciously continued to survive. Or in the words of Martin van den Brink, retired ASML CTO: “Moore’s Law will continue, as long as we have good ideas”.





Gordon Moore and his famous graph from 1965.[48]

4.1 Past: the original text of Moore's Law

Gordon Moore, founder of Intel Corporation (U.S.), wrote an article in 1965 titled, "Cramming more components onto integrated circuits”. Using two graphs in his article, Moore tried to explain that the recently launched integrated circuits were promising electronic devices that, in the future, would dramatically increase the number of components they contained.

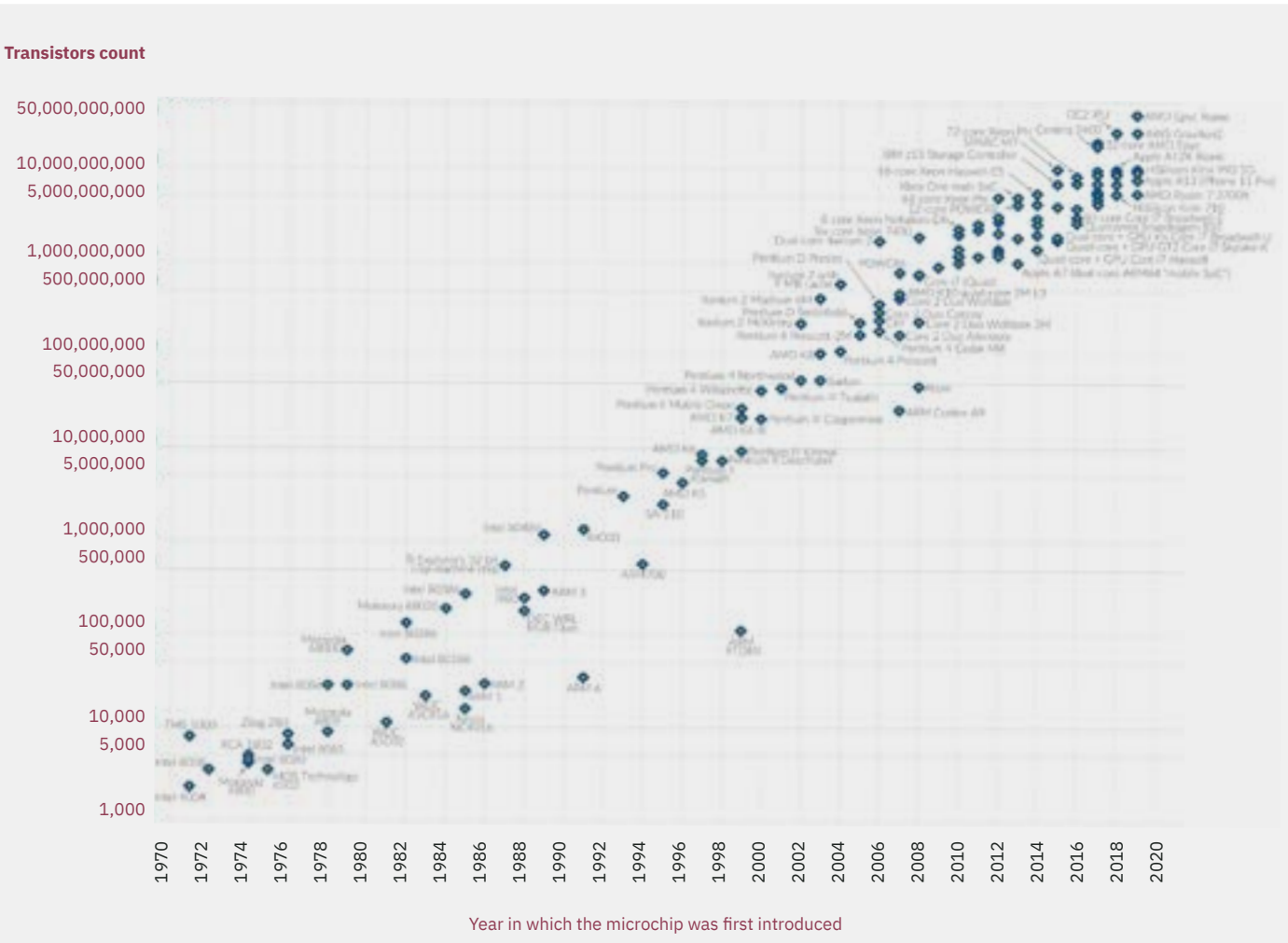
He speculated that the number of components per integrated circuit for minimum cost would reach 65.000 by 1975 and concluded with the hopeful observation, “I believe that such a large circuit can be built on a single wafer.”

At the end of 1975, Moore re-examined the trends in the integration level of ICs over the preceding 10 years and made a correction to “In the future, the integration level of semiconductors will double every two years.” From then on, this prediction came to be called “Moore’s Law.” Semiconductor devices have been following Moore’s Law in increasing the number of transistors over the past 50-plus years.

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.

This advancement is important for other aspects of technological progress in computing - such as processing speed or the price of computers.



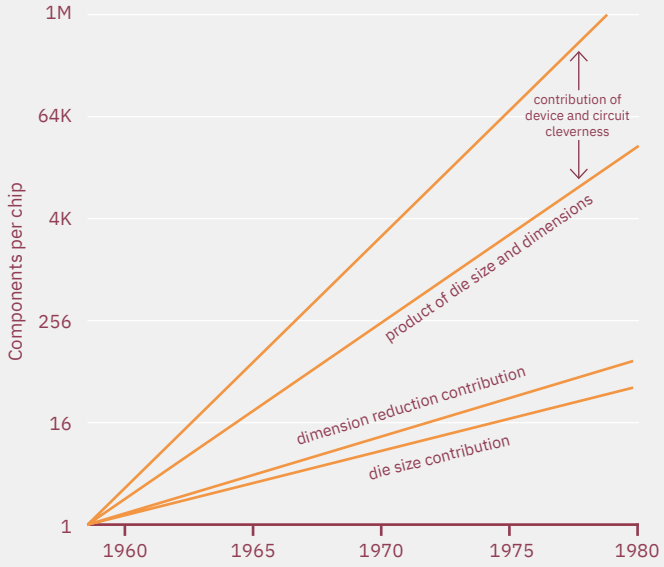
Trends in the number of transistors contained in a semiconductor chip by semiconductor device type

Source: Wikipedia

4. Past, Present, and Future of Moore's Law

The transistor count in Intel microprocessors (MPU) for PCs had been increasing at an average annual rate of approximately 40% until 2010. Since then, the rate has fallen by half. Since 2013, transistor counts for Apple's A-series application processors (APU) used in iPhones and iPads have continued to increase at the annual rate of 43%, up to the A13 processor containing 8.5 billion transistors. The high-end GPU from NVIDIA (U.S.) contains more transistors than other companies' processors, already exceeding 50 billion, and has been increasing the integration level following Moore's Law.

The transistor count in DRAM used to increase at an approximate average annual rate of 45% until the early 2000s but subsequently declined to approximately 20% until the 16-gigabit generation, which appeared in 2016. Annual growth in NAND flash memory capacity remained at 55-60% through about 2012 but has since been around 30-35%. The shrinking of two-dimensional structures was stopped at 20 nm or slightly smaller, and in the case of NAND memory, the momentum for increasing the capacity is being restored by transitioning to three dimensions ahead of other devices, as described later. Already in cutting-edge products, the number of layers in three-dimensional structures has increased from 128 to 176. Interestingly, also already in 1975 Gordon Moore envisioned that improvements will not only stem from reductions in the size of components (i.e. geometrical and device scaling, see Section 5), but also from two other factors (see the graph below for Moore's historical analysis and his short term extrapolation from 1975): namely (i) increases in semiconductor die sizes, i.e. circuit scaling; and (ii) contributions from what he called 'Device and Circuit Cleverness'. This form of device and circuit scaling is a.o. squeezing waste space out of the chip and getting rid of isolation structures.



–

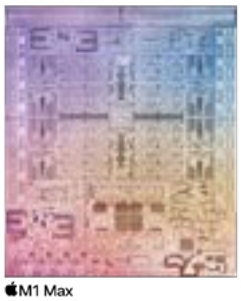
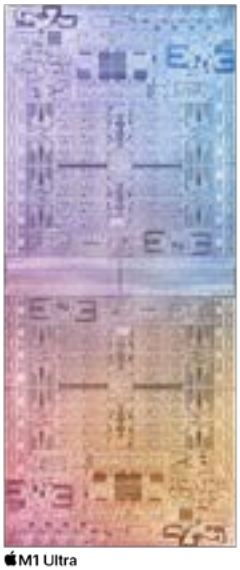
[46] https://en.wikipedia.org/wiki/Transistor_count

[47] G.E. Moore, in Plenary Address, International Electron Devices Conference (IEDM), 1 –3 (1975)

He predicted that the contribution from 'Increases in die sizes' had been close to the contribution from smaller components. In 1975 an Intel 8080 microprocessor die was 20mm^2. Today, the Apple M1 Ultra die is 864 mm^2.

It is clear: the last decade or two prove it has become increasingly difficult for the semiconductor industry to stay on the path as envisioned by Gordon Moore. Or in the words of market research agency IC Insights[49]: "The powerful drive of the IC industry to innovate its way over and around technology barriers can never be underestimated, but there are some very dramatic shifts underway regarding how ICs are designed and manufactured."

Apple processors[48]

Chip name	M1	M1 Pro	M1 Max	M1 Ultra
Chip Photo Drawn at almost the same scale				
Announce date	November 10, 2020	October 18, 2021	October 18, 2021	March 8, 2022
Semiconductor technology node	TSMC N5	TSMC N5	TSMC N5	TSMC N5
Chip area (W x H)	118 mm^2 (10.9x10.9)	245 mm^2 (18.9x12.9)	432 mm^2 (19.3x22.4)	864 mm^2 (19.3x44.8)
Transistor count	16 billion	33.7 billion	47 billion	114 billion

–

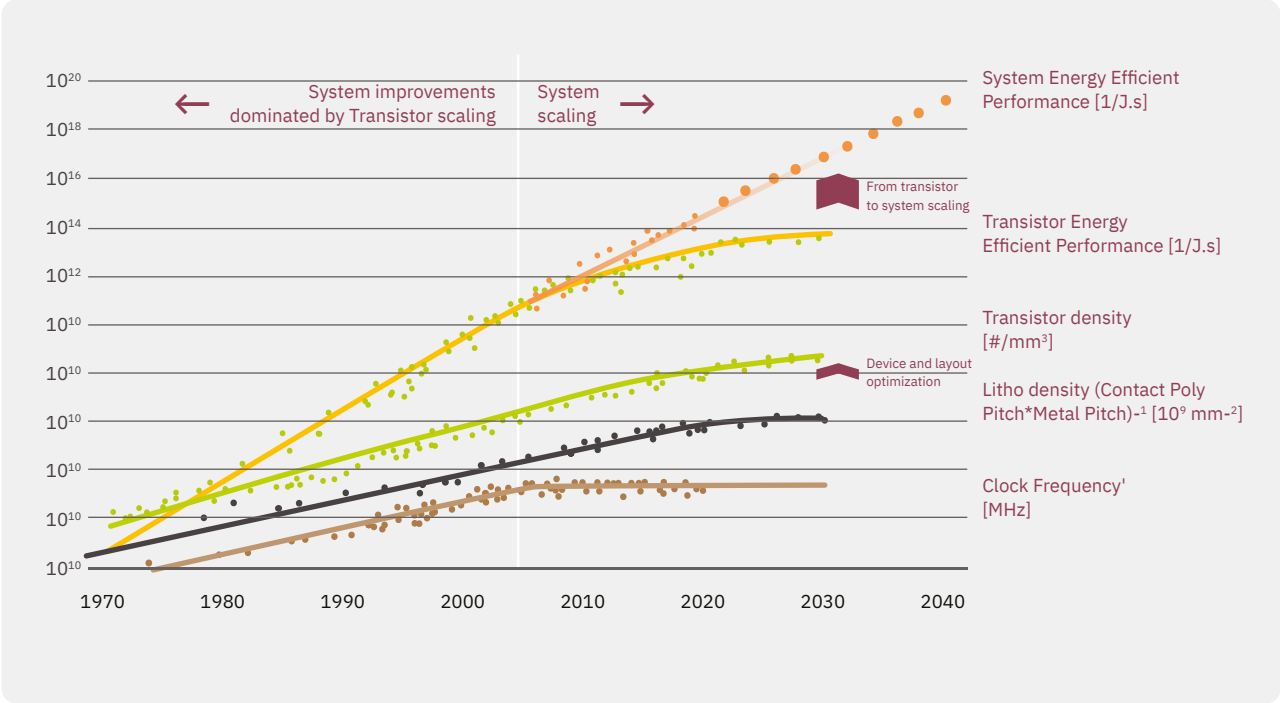
[48] https://apple.fandom.com/wiki/Apple_M1_Max

[49] https://xtech.nikkei.com/dm/article/NEWS/20110519/191938/

4. Past, Present, and Future of Moore's Law

4.2 Present: rising semiconductor manufacturing and energy cost[50]

There's a quiet upheaval happening in the semiconductor industry. The rules that have always governed the industry are fraying, undoing assumptions that we took for granted. The irreproachable Moore's Law, that exponential progress will make things cheaper, better, and faster over time, is changing. People are starting to appreciate that making a chip is not easy. While we continue to improve transistor density through new techniques, each one layers additional costs.



ASML's view on the future of Moore's Law.

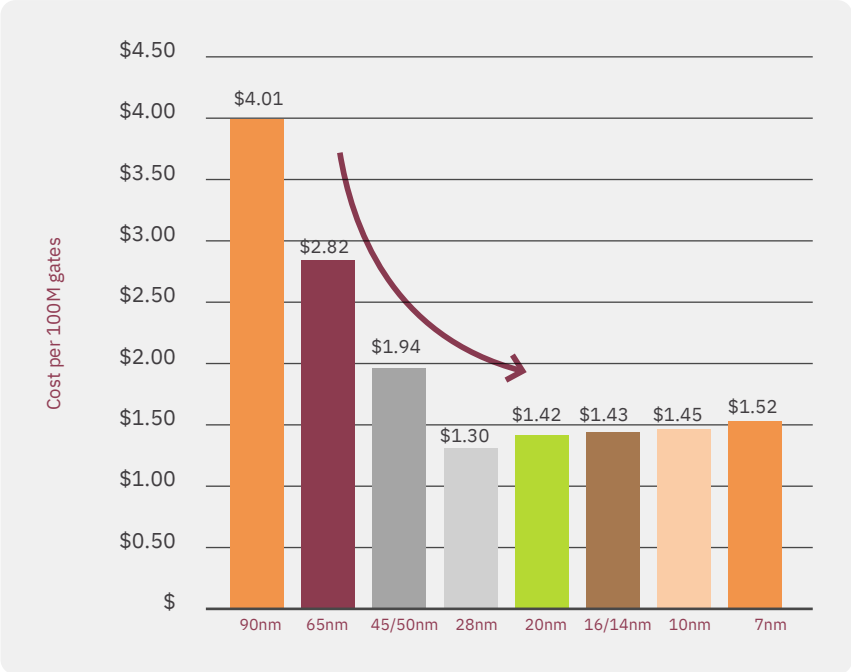
[51] https://tweakers.net/reviews/9468/3/asmls-toekomstplannen-hoe-ziet-asml-de-chipindustrie-tot-2030-moores-law-nog-steeds-levend.html

4.2.1 Rising manufacturing costs

One of the major historical assumptions of Moore's Law is that not only would the number of transistors double every two years, but the cost of the transistors would decline. That is no longer true. The chart below is from Marvell's 2020 investor day. The bar for 28nm was approximately 2011-2012.

What's interesting is that a qualitative change happened around 28nm, as it was one of the last planar nodes. As we switch FinFET and GAA or the next iteration of gate technology, it is to be expected that the cost per transistor will continue to increase. This is driven by the increased complexity of making these chips — namely, the added number of steps in manufacturing.

Gate cost trend



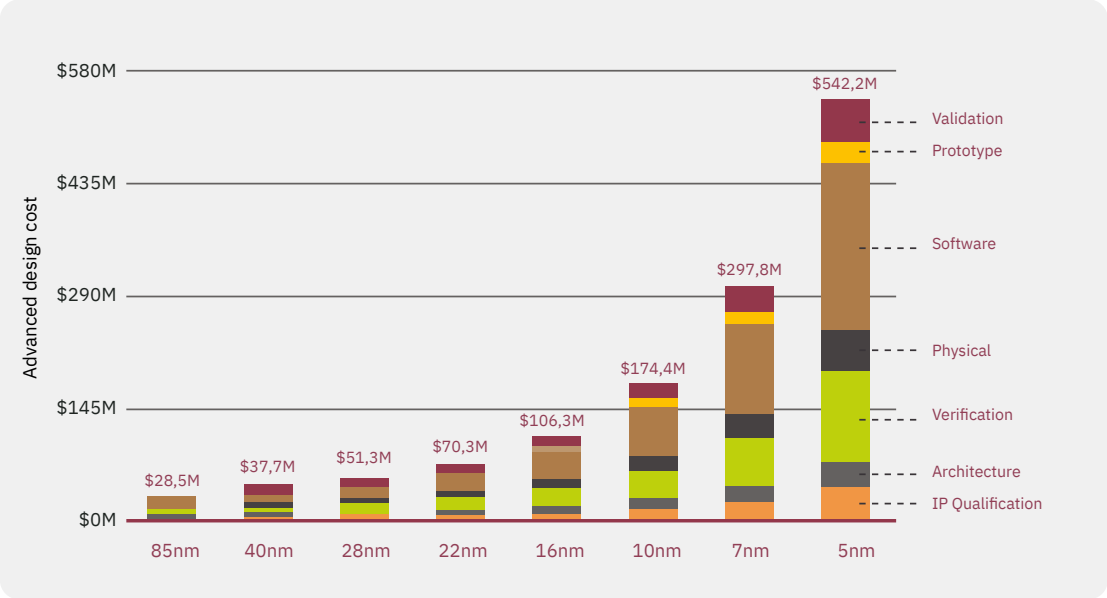
Marvell 2020 Investor Day - Slide 43[52]

[52] https://www.fabricatedknowledge.com/p/the-rising-tide-of-semiconductor

4. Past, Present, and Future of Moore's Law

The drastic price increase is broad-based and is across DRAM, NAND, and Logic. The primary driver is not only the rising costs of tools such as EUV, but the rising number of steps to make a chip. For instance, the number of logic process step increased from approximately 400 at 28nm to over 1200 process steps at 5nm[53].

The rising complexity of chips with smaller nodes is also reflected in the design cost of the chip. At 28 nm node this was just $51 million while at 5 nm node it is 10x higher. These starting costs limit the number of new designs in advanced nodes drastically.



Advanced design cost is more than $ 500 million in 5 nm node[54]

4.2.2 Energy demand by semiconductor devices[55]

For quite some years the total energy consumption by semiconductor devices worldwide was around 2000 TWh. This is about 16x the total Dutch energy consumption of 120 TWh. In this decennium semicon energy consumption is expected to increase by 450% mainly due to data centres and networks.
It isn't that our demand for data has been meagre in the recent past, according to Goldman Sachs Research. In fact, data centre workloads nearly tripled between 2015 and 2019. Through that period, though, data centres' demand for power remained flattish, at about 200 terawatt-hours per year. In part, this was because data centres kept growing more efficient in how they used the power they drew, according to Goldman Sachs Research. But since 2020, the efficiency gains appear to have dwindled, and the power consumed by

–

[53] https://www.priz.guru/strategies-for-effective-cost-reduction-in-production-processes/

[54] https://www.researchgate.net/figure/Chip-Design-and-Manufacturing-Cost-under-Different-Process-Nodes-Data-Source-from-IBS_fig1_340843129

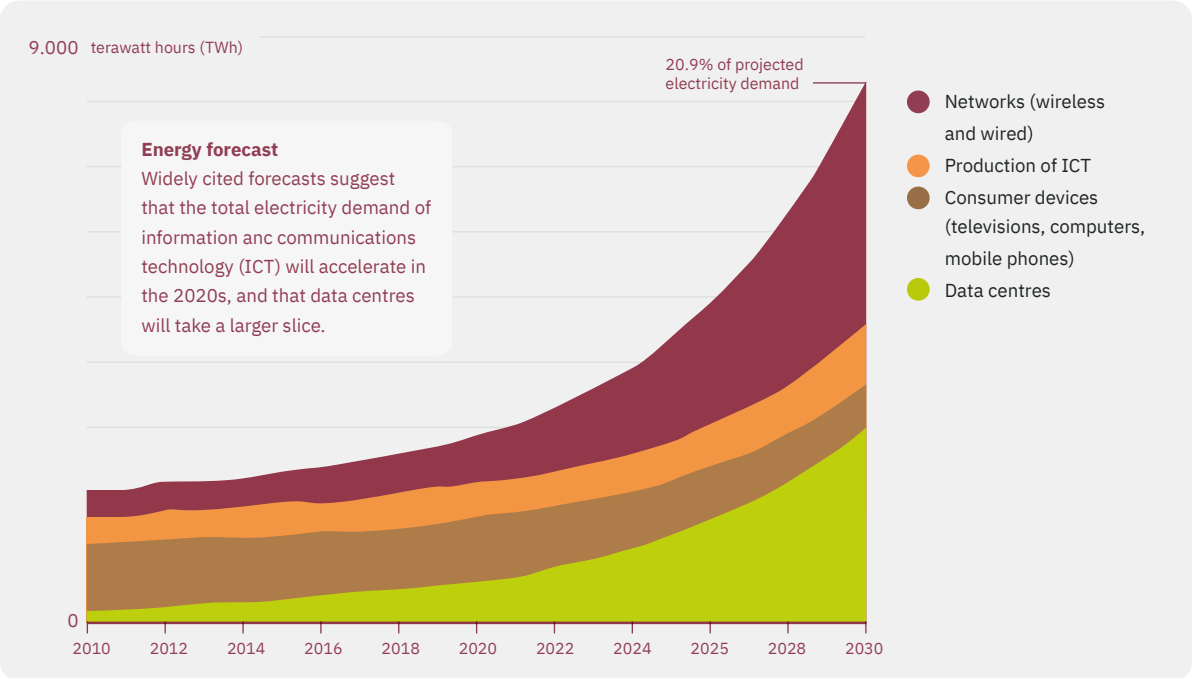
[55] https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand

data centres has risen exponentially. Some AI innovations will boost computing speed faster than they ramp up their electricity use, but the widening use of AI will still imply an increase in the technology's consumption of power. A single ChatGPT query requires 2.9 watt-hours of electricity, compared with 0.3 watt-hours for a Google search, according to the International Energy Agency. Goldman Sachs Research estimates the overall increase in data centre power consumption from AI to be in the order of 200 terawatt-hours per year between 2023 and 2030[56]. By 2028, AI is expected to represent about 19% of data centre power demand.

The energy hunger of data centers can also be seen in the Netherlands. Here, data centers account for approximately 4% of national energy consumption, compared to 1.5% in 2017. "Giga data centers" currently under construction could double power consumption again in the coming years.[57]

4.3 Future: Energy Efficient Performance (EEP) as metric of Moore's Law?

Nations around the world have set out toward decarbonization and the demand for improving the power efficiency of semiconductor devices is heightening more than ever, especially with respect to reducing the power consumption of data centres. In his keynote address at ISSCC 2021, TSMC's Chairman Liu said, "IC's energy-efficient performance (throughput × throughput/Watt) will continue doubling every two years in the future."



–

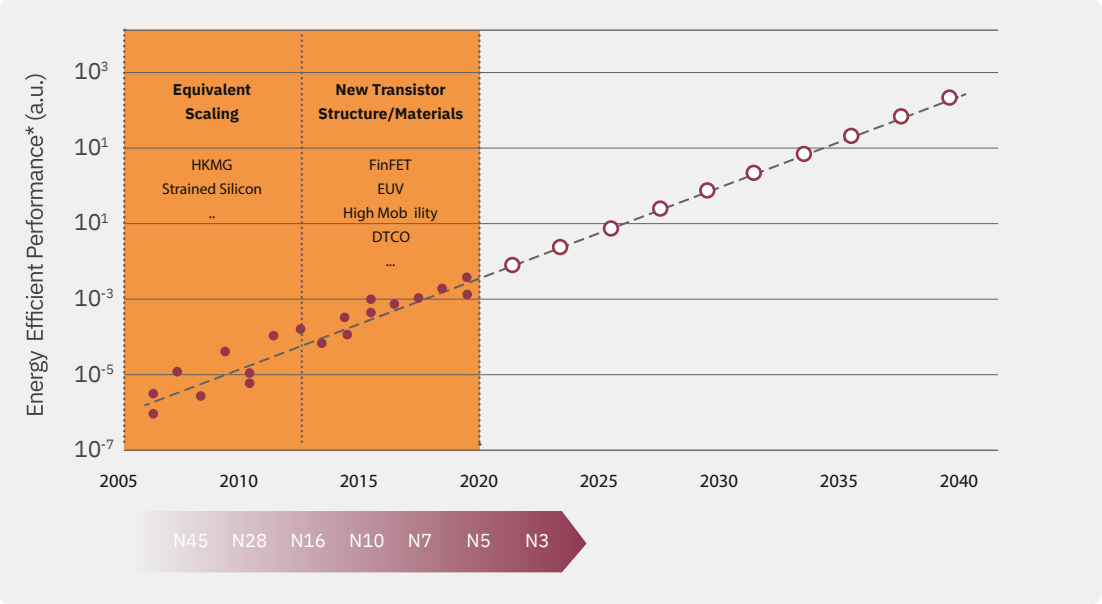
[56] https://seekingalpha.com/article/4706513-iberdrola-earnings-beneficiary-of-surging-power-demand

[57] https://seekingalpha.com/article/4706513-iberdrola-earnings-beneficiary-of-surging-power-demand

4. Past, Present, and Future of Moore's Law

When the energy efficiency of the GPUs announced so far is plotted, it can be seen that the introduction of new transistor structures and new materials one after another has successfully doubled the energy-efficient performance relative to performance every two years. Chairman Liu is calling on semiconductor materials makers, semiconductor manufacturing equipment makers, process engineers, circuit designers, system architects, and global academia, all of whom comprise the ecosystem, to collaborate in the future to double the energy-efficient performance every two years. In this new version of Moore's Law the metric is on energy efficient performance (EEP) or performance, power area and cost (PPAC).

It is that metric of Moore's Law that semicon market leaders expect to continue for the next decades. In other words: the original story of Moore was just about the number of transistors on a chip. This resulted in higher performance (GFlops), cost reduction (cost per transistor) and higher energy efficiency (J/transistor). Now we are in an era where we have to work hard on cost and energy efficiency while increasing the performance.



History and projection of energy-efficient performance (throughput × throughput/Watt) of GPUs
Note: Names such as N45 indicate scaling technology nodes (nm) at TSMC.
Source: TSMC[58]

–

[58] https://www.nu.nl/klimaat/6330479/datacenters-slurpen-energie-door-ai-komen-klimaatdoelen-in-gevaar.html

[59] https://www.imec-int.com/en/articles/smaller-better-faster-imec-presents-chip-scaling-roadmap

[60] https://www.appliedmaterials.com/jp/ja/stories/ppact-enablement-company.html

4.4 For EEP improvements we need smaller, better, faster[59]

Computing power needs are exploding due to the rapid rise of digital applications and data processing. With the growing use of artificial intelligence to tackle the major challenges of our time, like climate change or food shortage, the computing need is expected to double every six months from now on. To handle the exponentially growing amounts of data in a sustainable way, we need dramatic improvements in chip energy efficient performance (EEP) also stated as performance, power, area, cost (PPAC)[60].

As traditional Moore's Law scaling (i.e. transistor density) slows, the industry needs new ways to increase chip performance and increasing chip performance while lowering power consumption is a growing challenge. Innovations help the semiconductor industry to create faster, more efficient chips. In order to achieve that, we need to address five challenges or walls simultaneously. *"While no company in the world can accomplish this alone, co-innovation and collaboration across the semicon ecosystem will enable the continuation of Moore's law"*, according to imec.

The five challenges or walls for EEP improvement are:

- The **geometrical scaling wall**: pure lithography-enabled scaling is slowing down. It is becoming increasingly difficult since individual structures of microchips and transistors are approaching the size of atoms, where quantum effects begin interfering with the operation of microchips.
- The **memory wall**: system performance is confronted with data path limitations between the cores and the memory. In fact: memory bandwidth cannot keep up with processor performance. We have more flops per second than gigabytes per second. So, memory and processor have to be closer together.
- The **power wall**: it is becoming more challenging to bring power into the chip and extract heat from the chip package, so we will have to develop improved power delivery and cooling concepts.
- The **sustainability wall**: the manufacturing of semiconductor devices contributes to an increasing environmental footprint, including greenhouse gas and water, natural resources, and electricity consumption.
- The **cost wall**: obviously, chip manufacturing costs may explode with the complexity increase, together with the costs for design and process development.

The following sections reveal how Moore's Law, with Energy Efficient Performance (EEP) as metric, can live on in the next decades.

–

[56] https://seekingalpha.com/article/4706513-iberdrola-earnings-beneficiary-of-surging-power-demand

[57] https://seekingalpha.com/article/4706513-iberdrola-earnings-beneficiary-of-surging-power-demand

[58] https://www.nu.nl/klimaat/6330479/datacenters-slurpen-energie-door-ai-komen-klimaatdoelen-in-gevaar.html

5

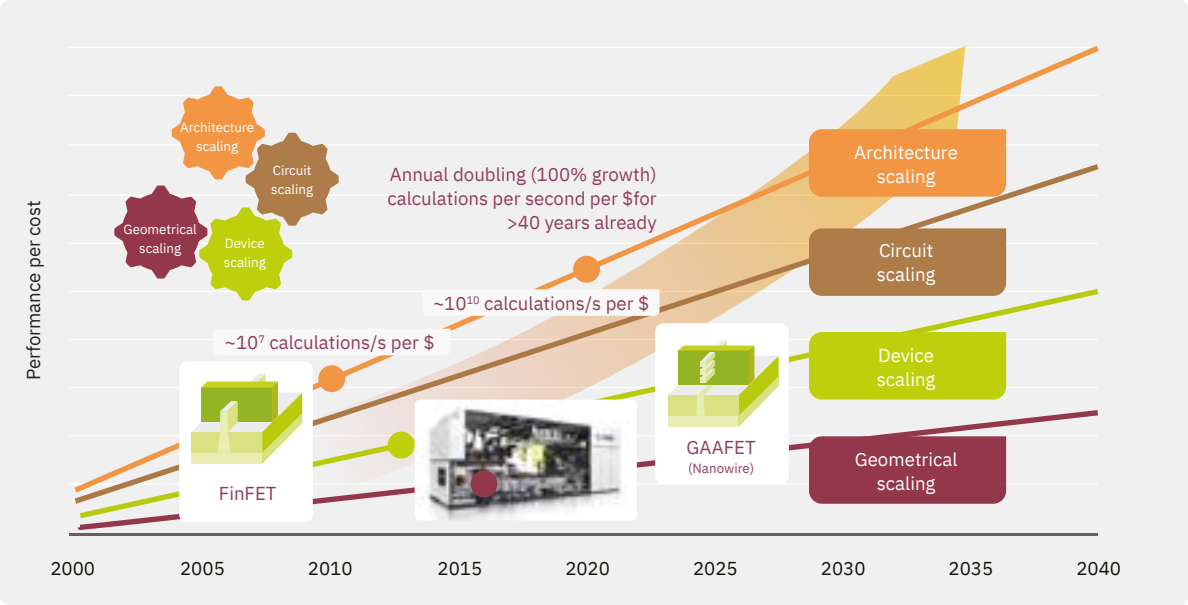
The scaling engines that drive Moore's Law into the future

5. The scaling engines that drive Moore's Law into the future

Moore's Law is facing five walls: as described in the previous section. So at first sight, things don't look great for Moore's Law in the metric of transistor density. This prognosis is especially true if we stubbornly stick to Dennard scaling and traditional Von Neumann compute architectures. In its scaling roadmap, imec proposes an alternative path for the future of chip technology, with fundamental changes in architectures, materials, new basic structures for transistors, and ... a paradigm shift.

Moore's Law in the metric Energy Efficient Performance (EEP) scaling has and will be driven forward by four scaling engines. The paradigm shift is that the emphasis in future scaling will shift towards system-level scaling: the combination of circuit scaling and architecture scaling.

- **Geometrical scaling**. Smaller pixels in the layout due to e.g. improvements in lithography.
- **Device scaling**. Improved layout of a t ransistors that requires less pixels per transistor.
- **Circuit scaling**. Improvements how transistors and chips are connected.
- **Architecture scaling**. The architecture reflects how transistors are used to perform calculations.



Moore's Law is driven by four scaling engines.

In this section, we provide a brief overview to the role each scaling engine plays in enabling EEP improvements in the coming decades. We follow the roadmaps of imec and International Roadmap for Devices and Systems (IRDS) that will guide us towards 2036. In the next sections developments in the four scaling engines will be discussed in more detail.

Geometrical scaling (See Section 6 for more details) will continue unabated for the next 10 years and beyond as a means to increase transistor density in ICs. The continuous advances in lithography will be key to further geometrical scaling: traditional lithography uses light, and, today, the wavelength of light is greater than the required accuracy of the patterns. That's why Extreme UV (EUV) lithography has been introduced. It is now appearing on more and more functional production belts for volume manufacturing. To go smaller we need an updated version of EUV, high-NA EUV or even hyper-NA, with bigger lenses.

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.

At the same time, we will also need innovation in the transistor: **device scaling** (See Section 6 for more details). Today almost all chip manufacturers build microchips with FinFET transistors. However, when entering the 3nm-generation, FinFETs suffer from quantum interference, causing disruptions in the operation of microchips.

Next in line is the Gate-All-Around (GAA) or nanosheet transistor, built up as a stack of nanosheets, it will offer improved performance and improved short channel effects. This architecture will be essential from 2nm node onwards. Major chip manufacturers like Samsung, Intel, and TSMC will introduce GAA transistors in their 3nm and/or 2nm nodes.

Further scaling can be realized by putting the negative and positive channels on top of each other, referred to as the Complementary FET (CFET) transistor, a complex vertical successor to the GAA. It significantly improves density but comes at the expense of increased process complexity, especially to contact the source and drains of the transistors. This enables more transistors per mm2 without making the transistor or its pixels smaller.

In time, CFET transistors will incorporate new ultra-thin 2D monolayer materials with an atomic thickness, like Tungsten disulfide (WS2) or molybdenum. This device roadmap, combined with the lithography roadmap, will bring us to the ångström age.

According to IRDS, the limit of what can be done with lithography will be reached. After that, "the way forward is to stack.... That's the only way to increase density that we have." 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. This is **Circuit scaling** (See Section 7 for more details).

Circuit scaling is addressing two other challenges that are playing at the system level sub 2nm-node. The memory bandwidth cannot keep up with CPU performance. The processor can't run faster than the pace at which data and instructions become available from the memory. To knock down this 'memory wall', memory must come closer to the chip. Circuit scaling delivers the solution here. An interesting approach for tearing down the memory wall is 3D system-on-chip (3D SOC) integration, which extends today's popular chiplet approaches. Following this heterogeneous integration approach, the system is partitioned into separate chips (called 'chiplets') that are concurrently designed and interconnected in the third dimension. It will allow for example to stack a SRAM memory layer for level-1-Cash

5. The scaling engines that drive Moore's Law into the future

right on the core logic devices, enabling fast memory to logic interaction. To achieve extreme high bandwidth off-module connectivity, optical interconnects, integrated on photonics interposers are being developed.

Regarding system-related challenges, getting enough power into the chip and getting the heat out becomes more difficult. However, a solution is in sight: the power distribution now runs from the top of the wafer through more than ten metal layers to the transistor. The semicon industry is currently working on a solution from the backside of the wafer. Power rails are sunk into the wafer and connected to the backside using nano-through-silicon vias in wider, less resistive materials. This approach will decouple the power delivery network from the signal network, improving the overall power delivery performance, reducing routing congestion, and, ultimately, allowing further standard cell height scaling.

Already in 2019 TSMC talked about circuit scaling as the way forward. They called it 'System Integration': creating larger systems out of smaller functions that are connected sideways or stacked. So once again, almost six decades after his original paper, Moore has proved to be prescient.

Node shrinks will continue for a while yet at a slower pace but are getting more expensive. So, what happens then when gains from node shrinks, chiplets, stacking and other sources of improvement really do come to an end? The answer is **architecture scaling** (See Section 8 for more details).

For many decades almost all computing has been based the Von Neumann architecture. Professor Von Neumann saw the digital computer as a system with input, a central processing unit, and an output. But we will need to evolve toward domain-specific and application-dependent architectures, with massive parallelization comparable to the way our human brain works. This implies that the CPU will have a smaller role in favour of custom-made circuits for specific workloads.

Nvidia is an example of how architecture scaling can change the semiconductor industry while increasing EEP. For instance, Nvidia's Chief Scientist Bill Dally stated in 2023 that they booked a gain of over 1000x in 'single chip inference performance' over the last decade. However, only 2.5x of that has come directly from process improvements. In other words, maintaining the progress of Moore's Law is the result of billions and billions of dollars of investments, some extremely complex engineering, but is in the end responsible only for a fraction of Nvidia's GPU gains. The performance gain is primarily related to, what has been called Huang's Law named after Nvidia's CEO Jensen Huang:
Huang's law is an observation in computer science and engineering that advancements in graphics processing units (GPUs) are growing at a rate much faster than with traditional central processing units (CPUs), so much faster than traditional Moore's Law.

Basically, although GPU's and CPU's both use billions of transistors, their architecture and application are different. Where CPUs are general purpose processors, GPUs allow for 'parallel processing' and that makes it ideal for AI applications. The novel parallel processing architecture is responsible for most of the stellar inference performance of Nvidia's chip over the last decade.

Therefore, the end of Dennard scaling and the deceleration of performance gains for standard CPU microprocessors are not problems that must be solved but facts that, recognized, offer breathtaking opportunities not only for the Dutch Semiconductor industry, but also for Europe.

The next decade will see a Cambrian explosion of novel computer architectures, meaning exciting times for computer architects in academia and in industry. With tens of billions of transistors on a chip there are still many, many opportunities to explore and new ways to use them. Let along when we can combine these billions of transistors with photonic devices and even quantum devices that all bring their own unique capabilities.

Developments in the various scaling engines will be discussed in more detail in the upcoming sections.

6

Geometrical and device scaling

6. Geometrical and device scaling

Imec[61] recently shared its sub-1nm silicon and transistor roadmap. The roadmap gives us an idea of the timelines through 2036 for the next major process nodes and transistor architectures. Note that the node-names, such as “5 nm node”, are just names with no meaning of the actual feature size in the transistor.

Imec’s logic scaling roadmap.[62]

Lithography, which is the key technology for dimensional or **geometrical scaling**, has been increasing the resolution level by shortening the wavelength of the light source used, from g-line (436 nm) to i-line (365 nm), to KrF (DUV at 248 nm), and then to ArF (DUV at 193 nm). Furthermore, the introduction of ArF liquid immersion lithography, which uses ArF excimer laser as the light source and water as the emersion liquid between the lens and the wafer, has improved the resolution level. Subsequently, extreme ultraviolet (EUV at 13.5 nm) lithography, which many people had considered impossible to implement, was put into practical use, opening a path toward shrinking logic devices to the smallest metal pitch in the roadmap.

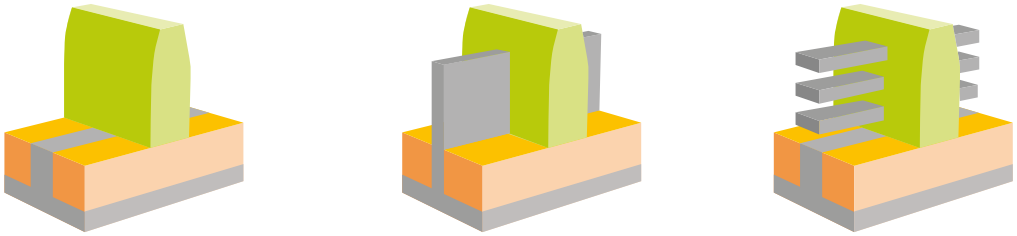
–

[61] Luc van den Hove, IMEC, ITF World, May 21, 2024

[62] https://spectrum.ieee.org/stco-system-technology-cooptimization

In fact, the minimal metal pitch is not limited by the capability of lithography, but rather by the conductivity of the metal structure. Smaller interconnect lines will have higher resistivity and that can limit the RC-time of the structure and therefore the clock-frequency of the processer. Once a material with higher conductivity than Cu can be used, the minimal metal pitch can be further reduced. Graphene has this conductivity. The challenge is on deposition and contacting this 2D material.

For decades shrinking planar transistors drove the **device scaling** roadmap until FinFET transistors were introduced about 10 years ago that give better control over the leak current in the channel. The roadmap outlines that standard FinFET transistors will last until 3nm node but then transition to the new Gate All Around (GAA) nanosheet designs that will enter high-volume production in 2024. So, we are now on the verge of GAA, which is an even more 3D-intensive structure than FinFET.



Three generations of transistors.[63]

Imec charts the course to forksheet designs at 2nm and A7 (0.7nm), respectively, followed by breakthrough designs like CFETs and atomic channels at A5 and A2. Complementary FET (CFET) transistors will shrink the footprint even further when they first arrive with the 1nm node (A10) in 2028, allowing more densely packed standard cell libraries. Eventually, we'll see versions of CFET with atomic channels, further improving performance and scalability. CFET transistors stack NMOS and PMOS devices on top of each other to enable higher transistor density without improved lithography resolution.

2D materials are being researched as transistor channel materials in 1nm node processes. These are two-dimensional (2D) atomic layer, nanomaterials, such as graphene and transition metal dichalcogenides. Imec maintains that these new technologies and materials can now be expected to extend Moore's Law beyond 1nm or A10 node.
Meanwhile the insulating film/gate materials changed from the conventional SiO2/SiN (silicon nitride insulation film)/poly Si (polysilicon) gate to a high-k (high relative dielectric constant) insulating film/metal gate, further suppressing the gate leak current. The wiring material was also changed from conventional Al to Cu, possessing high electrical conductivity, and Co and Ru will also begin to be used in the future. So, more and more elements of the Periodic Table are used in chip manufacturing.

–

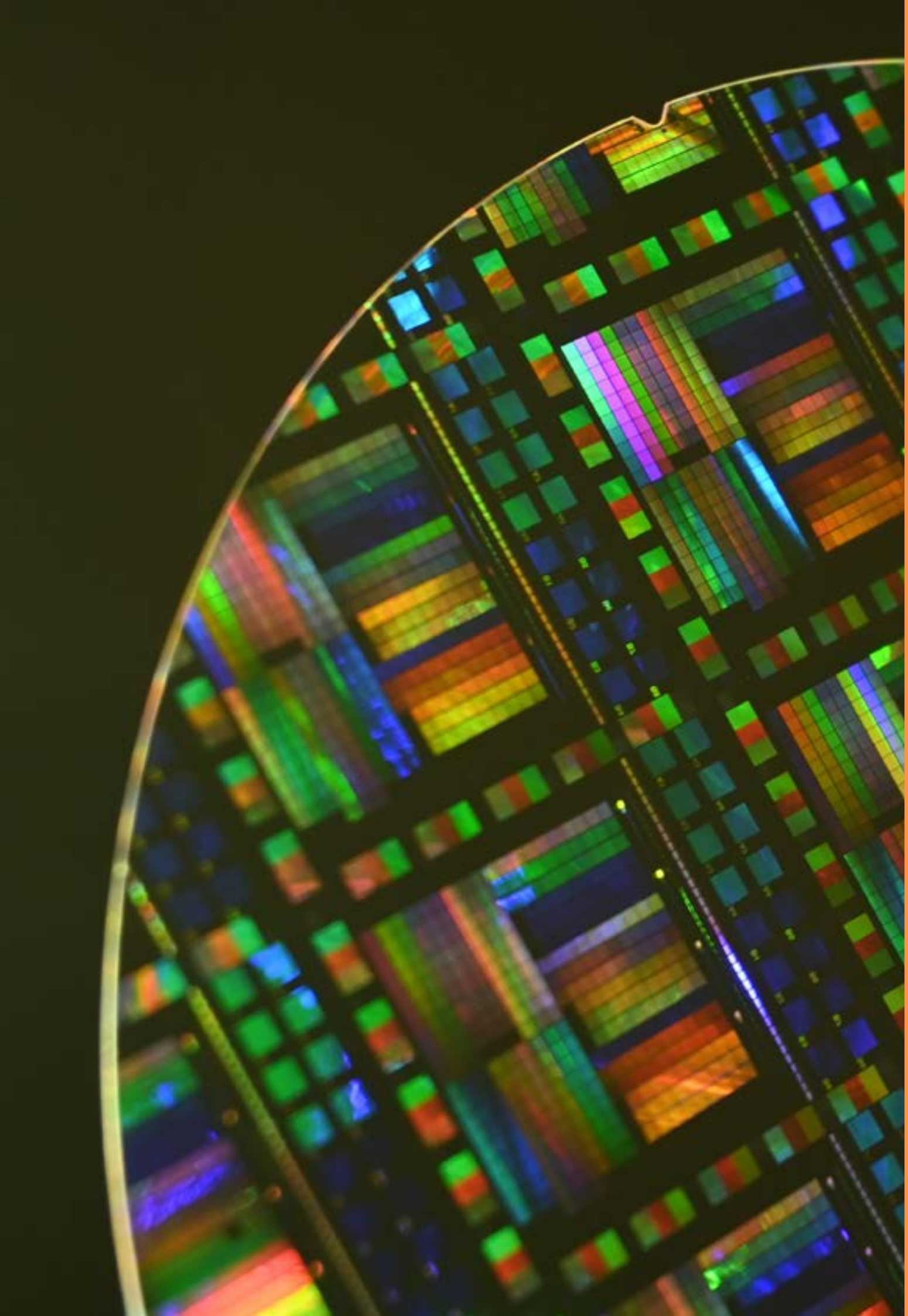
[63] https://www.researchgate.net/figure/Comparative-structures-of-MOSFET-FinFET-and-GAAFET_fig3_379259370

6. Geometrical and device scaling

This requires novel deposition and etching techniques.

Another step in the device layout that enhance the transistor density is the introduction of backside power delivery network (BPDN). Unlike today's designs that deliver power from the top of the chip down to the transistors, BPDN route all power directly to the backside of the transistor with TSVs, thus separating power delivery from the data transmission interconnects that remain in their normal location on the other side. Separating the power circuitry and the data-carrying interconnects improves voltage droop characteristics, allowing for faster transistor switching while enabling denser signal routing on the top of the chip. Signal integrity also benefits because the simplified routing enables faster wires with reduced resistance and capacitance.

Moving the power delivery network to the bottom of the chip enables easier wafer-to-wafer bonding at the top of the die, thus unlocking the potential to stack logic on memory. Imec even envisions possibly moving other functions to the backside of the wafer, like the global interconnect or clock signals.

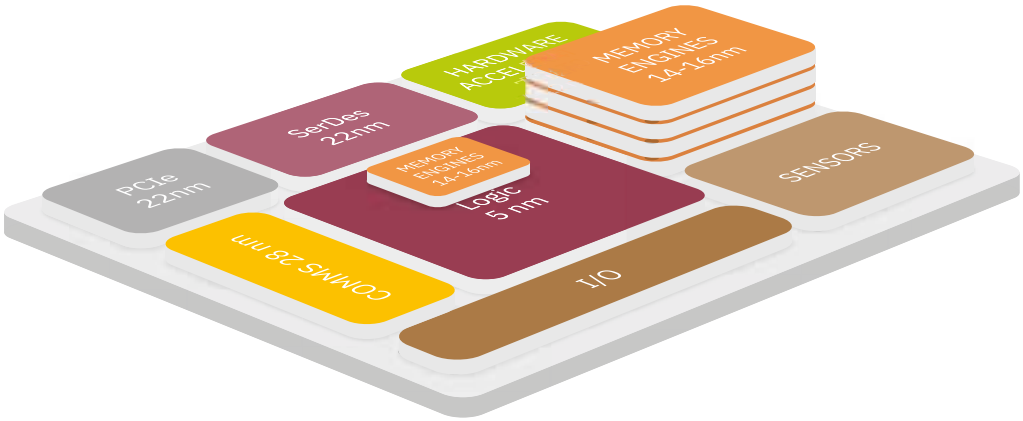


7

Circuit scaling

7. Circuit scaling

Moving to smaller nodes is becoming more expensive over time, and the standard approach of building monolithic chips with ever more transistors per area has given way to 'circuit scaling'. Circuit scaling is 3D scaling where geometrical and device scaling is 2D scaling. 3D scaling is implemented by combining chiplets in a 3D IC. Chiplet-based designs break various chip functions into distinct dies connected together, thus allowing the chip to function as one cohesive unit — albeit with trade-offs.



Chiplets offer a modular system that combines separate chips from different vendors and technology nodes instead of designing all functions into one monolithic system on chip.[64]

Chiplets are small, modular chips serving a specific function, such as CPUs or GPUs that can be mixed and matched into a complete system. The Lego-like approach hands manufacturers the flexibility to compose a system cost-effectively with lower entry costs for new chip designs and increased efficiency and performance.

7.1 The benefits of circuit scaling technology

Chiplet technology represents a paradigm shift in semiconductor design, addressing several critical challenges faced by traditional monolithic ICs. Here are the primary benefits[65]:

7.1.1 Enhanced transistor density

Today most ICs have all transistors in a single layer. When circuit scaling is implemented multiple of these layers can be stacked on top of each other, enhancing the transistor density. In the previous section the CFET was introduced where a n-type and a p-type transistor are on top of each other, while the power to these transistors is delivered in a third layer (BPDN). In NAND memory the 3rd dimension is already used for years with several 100 layers to store information. DRAM memory is expected to into the 3rd dimension as well.

[64] https://www.researchgate.net/figure/Chiplet-partitioning-concept_fig1_347759902

[65] https://www.micro-ip.com/en/chiplet-technology/

7.1.2 Enhanced Performance and Scalability

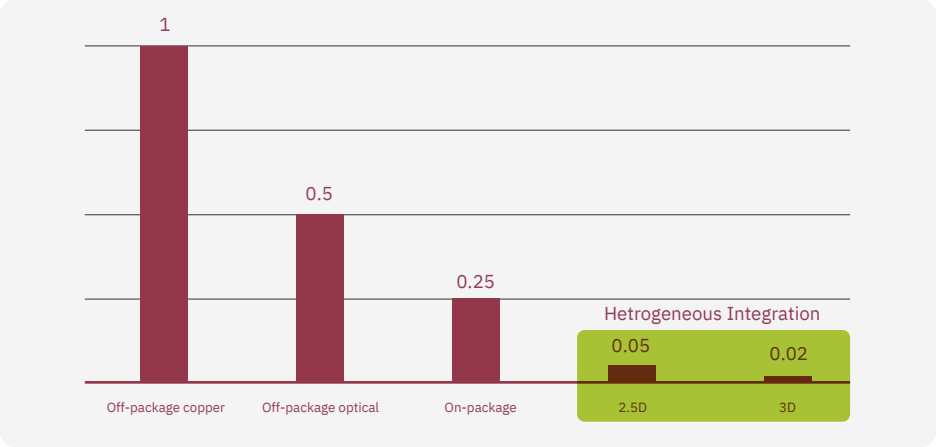
Chiplets allow for the integration of different functional blocks, each optimized for specific tasks, onto a single package. This modular approach enhances performance by enabling higher bandwidth and lower latency interconnects between chiplets, crucial for AI and HPC applications that require rapid data processing and transfer.

As chiplets can also be stacked, the distance data has to travel is reduced. In this way the performance of the system can be enhanced compared to a monolithic chip . In AI applications, where high bandwidth communication between memory and processor is required, stacked memory and processor chiplets will enhance the performance significantly.

There are indications that performance can be further enhanced by combining electronics with photonics. UDN reports that NVIDIA and TSMC developed the first silicon photonic chip prototype in 2024 and that NVIDIA and TSMC are also working on optoelectronic integration technologies and advanced packaging.

7.1.3 Enhanced energy efficiency

The shorter distance does not only reduce latency, but it also reduces power dissipation as the resistance is lower due to shorter interconnect length when using chiplets compared to a monolithic chip. Heterogeneous stacking of chiplets can reduce the interconnect power dissipation by a factor 20-50x compared to interconnect with a traditional printed circuit board (PCB).



Power dissipation in interconnects is reduced.[68]

–

[66] https://semianalysis.com/2023/02/02/iedm2022p1/#samsung-advanced-packaging-hybrid-bond-logic-4um-hybrid-bond-hbm-paper-3-6

[67] https://wccftech-com.cdn.ampproject.org/c/s/wccftech.com/nvidia-tsmc-develop-advanced-silicon-photonic-chip-prototype-says-report/amp/

[68] Dr. Sundar Ramamurthy, Applied Materials, 'Heterogenous integration, ICAPS and epitaxy', Semicon West, 2023

7. Circuit scaling

7.1.4 Improved Yield and Cost Efficiency

By fabricating smaller, individual chiplets instead of a large monolithic die, the yield rates improve significantly. Defects are confined to smaller areas, reducing the need to discard entire wafers. This approach also enables cost savings, as chiplets can be produced using different process nodes, optimizing cost and performance for each function.

The picture on page 39 shows that processors have grown significantly with negative impact on yield when manufactured monolithic. This can be improved when the processors are segmented in chiplets

The table on page 65 shows an example of potential cost saving using chiplets versus monolithic dies. In the Annex on page 118 a more detailed explanation including the assumptions used in our calculation is provided. In the example above a cost saving of 19.5% is realized which is similar to 1-2 nodes in IDRS. It is the result of reduced die cost and higher bonding cost. Here yield and direct cost per placement are the main parameters . Yield of hybrid bonding is key as rework is not possible. This means that an error in bonding will result in loss of at least two expensive dies or even more when the package contains more chiplets. Metrology in packaging will play an even more important role than in front-end processing to maintain cost saving.

7.1.5 Faster Time-to-Market, Design Flexibility and Innovation

Chiplets offer unparalleled flexibility in design, allowing designers to mix and match IP blocks from different vendors or process technologies. This flexibility accelerates innovation, enabling the integration of cutting-edge functionalities without the constraints of monolithic designs.

Chiplets allow for affordable application specific ICs (ASICs): mix and match chiplets at advanced and slightly older or more mature nodes from different vendors with unique characteristics, for e.g. sensors, made at legacy nodes. This enables the manufacturing of high-performance systems without the need to invest in leading-edge node factories and/or for startups to face extremely high upfront IC design costs. For instance, the figure on page 25 reveals design costs for advanced 5nm chips amount to well over USD 500 million versus USD 50 million for 28nm respectively USD 70 million for 22nm chips.

7.2 Heterogeneous integration

So, segmenting a monolithic chip into chiplets has several advantages. For example, it can reduce the investment needed to scale up European production, while at the same time reducing chip design costs for startups. However, it also brings a new challenge: the chiplets have to be integrated in a functioning system. This is what heterogeneous integration does: semiconductor chiplets, mixed signal chiplets, optical components and/or even MEMS can be combined in a high performance system.

Two major industry directions are emerging in 3D system integration: 2.5D chiplet integration connecting chips side-by-side through a common substrate (also known as an interposer) and 3D-SoC, where the chiplets are stacked on top of each other.

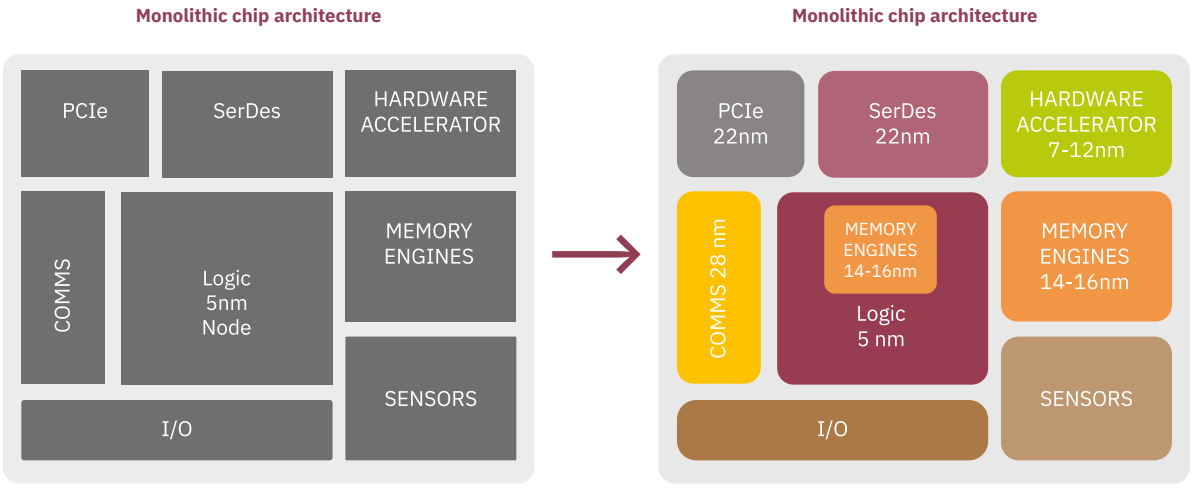
7.2.1 2.5D interposer technology

In 2.5D integration, the chiplets are connected via a common substrate in silicon, organic polymers, glass, or laminated.

[69] https://stateofthefuture.substack.com/p/e13- -how-chip-lego-is-driving

[70] https://arxiv.org/html/2203.12268v4

Segmenting a monolithic chip into chiplets also manufacturing these chiplets with different node technology.[69]



Monolithic chip	Parameter	Chiplet-based
400 mm^2	Chip area	440 mm^2 10% extra for interconnects
N5: 400 mm^2	Chip area per node	N5: 105 mm^2 (24%) N7: 50 mm^2 (11%) N14: 60 mm^2 (14%) N22: 130 mm^2 (30%) N28: 95 mm^2 (21%) (% reflects fraction of total chip area)
$ 68.5	Bare die price total [$]	$ 38.5 43,7% lower price than monolithically
$ 100.9	Known Good Dies price total [$]	$42.4 58% lower price than monolithically
$ 106.2	Known Good Bonded Dies price total [$]	$ 63.9 40% lower price than monolithically
$ 109.2	Packaged chip price total [$]	$ 87.9 19.5% lower price than monolithically

7. Circuit scaling

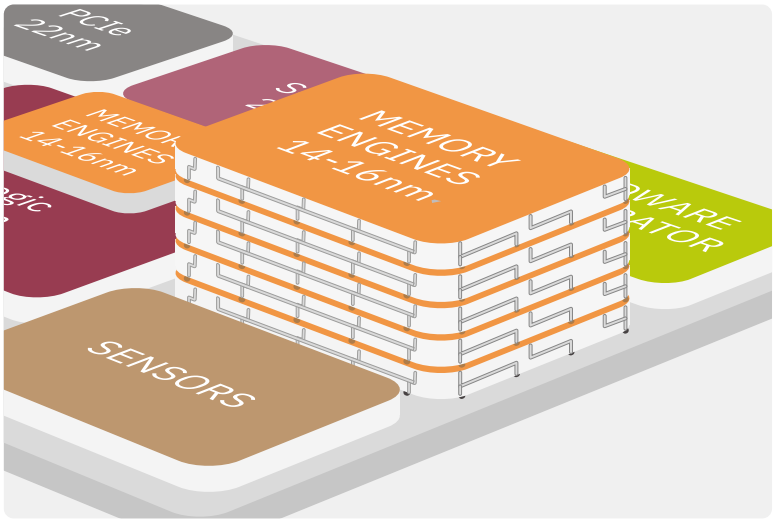
Early chiplet integration is focused on using silicon interposer substrates for interconnect between dies. It involves placing two separate chiplets very closely together (<50µm apart) on the common interposer, a substrate with micrometer-scale wiring that establishes the connections. Silicon interposers utilize traditional BEOL Cu/oxide damascene to realize µm and sub-µm level interconnect pitches with very high yield.

Imec is currently focusing on silicon and organic substrates. While silicon interposers are an established technology for high-performance applications featuring the finest pitch and good thermal and electrical properties, they also come with higher cost and complexity. Organic substrates are therefore researched and optimized as alternatives.

7.2.2 3D system-on-chip: hybrid bonding realizes submicron pitches

Some applications, such as high-performance computing, may demand high performance, smaller form factors, or a higher level of system integration and, therefore, prefer a complete 3D approach. Instead of establishing sideway connections, chiplets can be stacked on top of each other, forming a 3D-System on Chip (SoC). This approach does not add additional blocks but rather co-designs the chiplets together and lets them operate as if they would be the same chip.

Die-to-wafer and wafer-to-wafer hybrid bonding is a key technology for integrating 3D-SoC at the µm interconnect density level. It involves connecting two silicon chip(let)s with a low-temperature expansion coefficient together. The critical component in this process is the dielectric that planarizes and activates the surface of stacked layers for



Chiplets can be integrated using a silicon interposer.
Imec is also researching alternatives like silicon bridges or organic RDL.[71]

–

[71] https://www.cadence.com/en_US/home/explore/electromagnetic-simulation.html

[72] https://www.tomshardware.com/news/imec-reveals-sub-1nm-transistor-roadmap-3d-stacked-cmos-20-plans

effective bonding and electrically insulates the different chiplets in the stack. The roadmap even forecasts pitches of 400nm and 200nm.

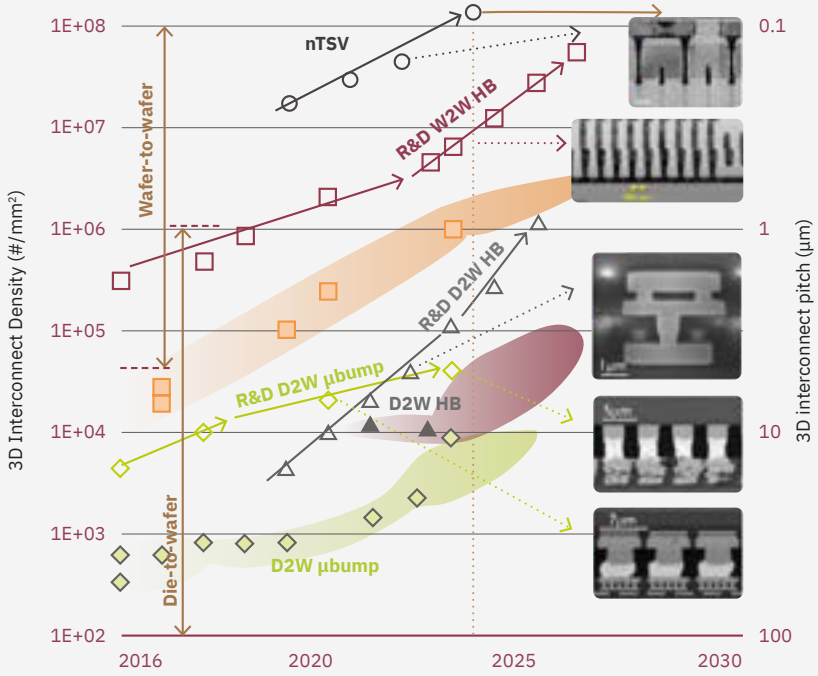
7.3 Future of circuit scaling

Separating functions and technology nodes into different chiplets proves more cost-effective and has space, performance and yield benefits over huge chips in cutting-edge process technologies.

While a modular approach provides answers to the complexity and cost of multichip packaging, this paradigm shift brings about specific technological challenges. Size is just one challenge. A substantial part of chiplet research is dedicated to making the interconnections smaller and/or exploring different concepts of bringing the pieces together. Other challenges are thermal issues caused by stacking chiplets together and power delivery becomes more critical (tackled by novel architectures such as backside power delivery networks). By the way, further standardization efforts are needed to ensure compatibility and communication between different chiplets.

Imec's vision goes even further and includes breaking the chips into smaller pieces, with caches and memories split into their own units with different transistors, then stacked in a 3D arrangement atop the other chip functions. This methodology will also lean heavily on backside power delivery networks (BPDN) that route all power through the backside of the transistor.

Finally, 3D system-on-chip or hybrid bonding also allows for shorter interconnects and therefore higher bandwidth which has limited the computational capabilities of modern CPUs and GPUs. Circuit scaling therefore also offers a (partial) solution for reducing the effectiveness of dimensional scaling by adding a few extra transistors.



Focus imec R&D roadmap: 5-10 year ahead of industry adoption

- W2W HB - imec R&D
- W2W HB - industry adoption
- D2W HB - imec R&D
- A D2W HB - industry adoption
- D2W uBump - imec R&D
- D2W Bump - industry adoption

DW2 = Die-to-wafer bonding
W2W = Wafer-to-Wafer bonding
HB = Hybrid bonding
TSV = Through-Si-via
nTSV = Backside contact at standard cell level (eg.Backside Power Delivery)

The 3D interconnect roadmap from imec summarizes the different approaches to interconnect chiplets and projected interconnect densities and pitches.[72]

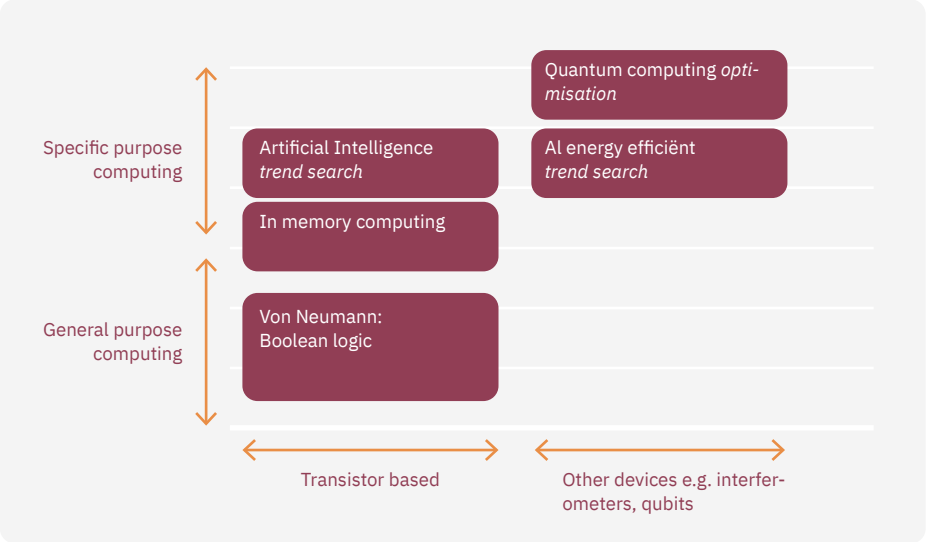
8

Architecture scaling

8. Architecture scaling

Architecture scaling is another path to extend Moore's Law when 2D and 3D scaling reaches its physical or economical limit. The architecture reflects how calculations are performed using an integrated network of devices, such as transistors. Today the calculations are performed in a serial process while in future more processes are performed in parallel similar to the human brain.

In the future (>2032) other devices than transistors might play a role for specific architectures and applications. Examples are photonic Mach-Zehnder interferometers and quantum qubits. How important their role will be in microprocessors will depend strongly on the ability to scale EEP with these novel devices.



8.1 Von Neuman architecture

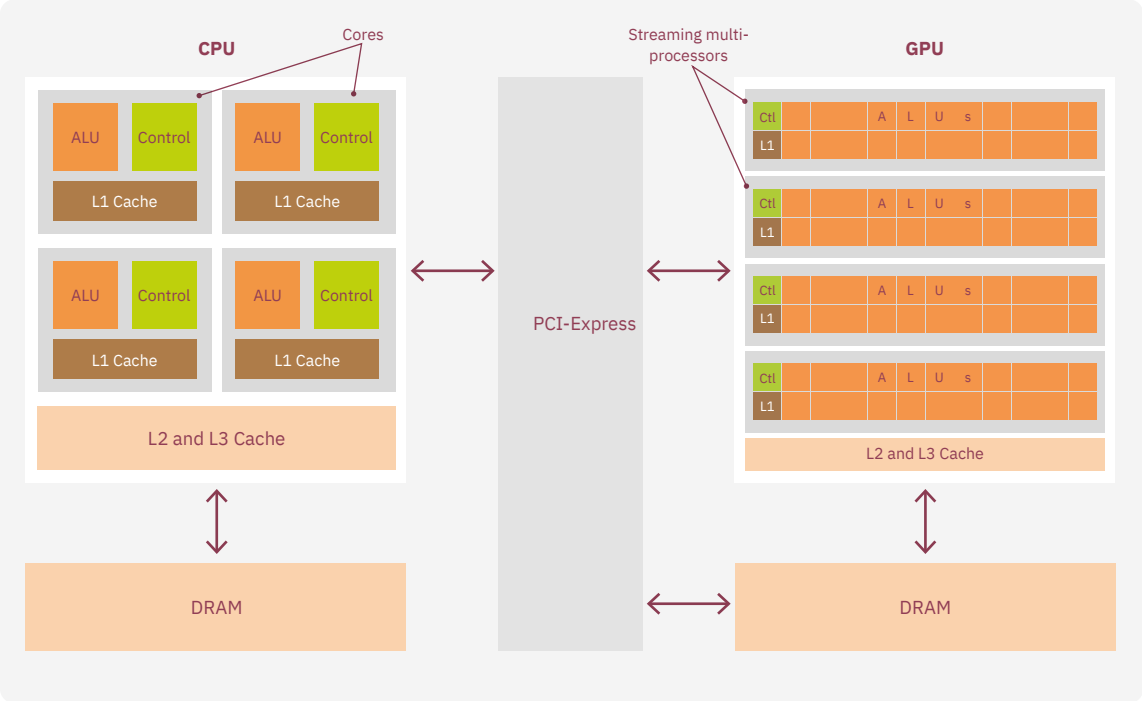
Since the beginning of computing the Von Neuman architecture has been used, implemented in the CPU. This general-purpose computer architecture design consists of a Control Unit, Arithmetic and Logic Unit (ALU), Memory Unit, Registers and Inputs/Outputs. Buses are the means by which data is transmitted from one part of a computer to another, connecting all major internal components to the CPU and memory.

Von Neumann architecture is based on the stored-program computer concept, where instruction data and program data are stored in the same memory. Every problem is broken down in an endless stream of Boolean (yes/no) questions, such as NAND and OR operations.

These operations are implemented by transistors. This design is still used in most computers produced today.

8.2 Artificial Intelligence

More recently, the artificial intelligence (AI) architecture has been implemented in the GPU. The GPU has many small "cores" that can work in parallel. The Arithmetic and Logic Unit (ALU) handles a limited set of instructions: vector-matrix multiplications. Also, the application is limited to primarily image recognition where the image is in the matrix while the vector represents the characteristics for recognition.



CPU versus GPU.[73]

NVIDIA is very successful designing AI chips with more than 200 billion transistors in their latest Blackwell architecture.
GPUs can also be made with other devices than transistors such as Mach-Zehnder interferometers. Here the matrix is the optical signal that is externally influenced in each interferometer as an implementation of the vector. For instance, the US startups Light-Matter has developed such an architecture in cooperation with Global Foundries. They claim that their array of just 65.000 interferometers outperformance the EEP of NVIDIA's transistor implementation.
Other examples of novel AI architectures are Axelera AI and Innatera. Axelera AI is Dutch start-up that develops Edge AI processors based on a Digital In-Memory Computing architecture also allowing for parallel processing (see page 110 for more information).

–

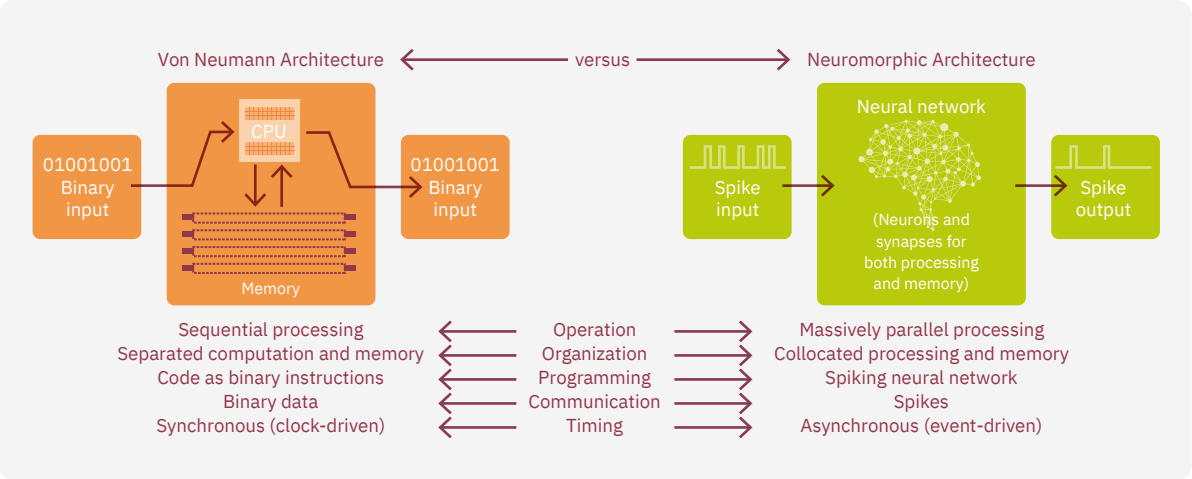
[73] https://www.linkedin.com/posts/gopal-chakraborty-b6a7b91b_gpu-architechture-memory-activity-7246032579190013952-1AEw/

8. Architecture scaling

8.3 Neuromorphic architecture

Neuromorphic computing is an approach to computing that is inspired by the structure and function of the human brain. A neuromorphic computer/chip is any device that uses physical artificial neurons to do computations. The implementation of neuromorphic computing on the hardware level can be realized by oxide-based memristors, spintronic memories, threshold switches, transistors, among others. Training software-based neuromorphic systems of spiking neural networks can be achieved using error backpropagation.

The Dutch startup Innatera is developing neuromorphic IC's (see also page 109).



CPU versus neuromorphic architectur[74]

8.4 Quantum architecture

A quantum computer is a computer that exploits quantum mechanical phenomena. On small scales, physical matter exhibits properties of both particles and waves, and quantum computing leverages this behaviour using specialized hardware. Classical physics cannot explain the operation of these quantum devices, and a scalable quantum computer could perform some calculations exponentially faster than any modern "classical" computer. Theoretically a large-scale quantum computer could break widely used encryption schemes and aid physicists in performing physical simulations; however, the current state of the art is largely experimental and impractical, with several obstacles to useful applications.

The basic unit of information in quantum computing, the qubit (or "quantum bit"), serves the same function as the bit in classical computing. However, unlike a classical bit, which can be in one of two states (a binary), a qubit can exist in a superposition of its two "basis" states, which loosely means that it is in both states simultaneously. When measuring a qubit, the result is a probabilistic output of a classical bit. If a quantum computer manipulates the qubit in a particular way, wave interference effects can

[74] https://medium.com/chain-reaction/brain-computers-976ae7ec19a8

amplify the desired measurement results. The design of quantum algorithms involves creating procedures that allow a quantum computer to perform calculations efficiently and quickly.

Currently several types of qubits are in the race to form the heart of a quantum computer. For more information on Quantum technology see Invest-NL Deep Tech Fund report 'Venturing into Quantum Technology in the Netherlands', October.

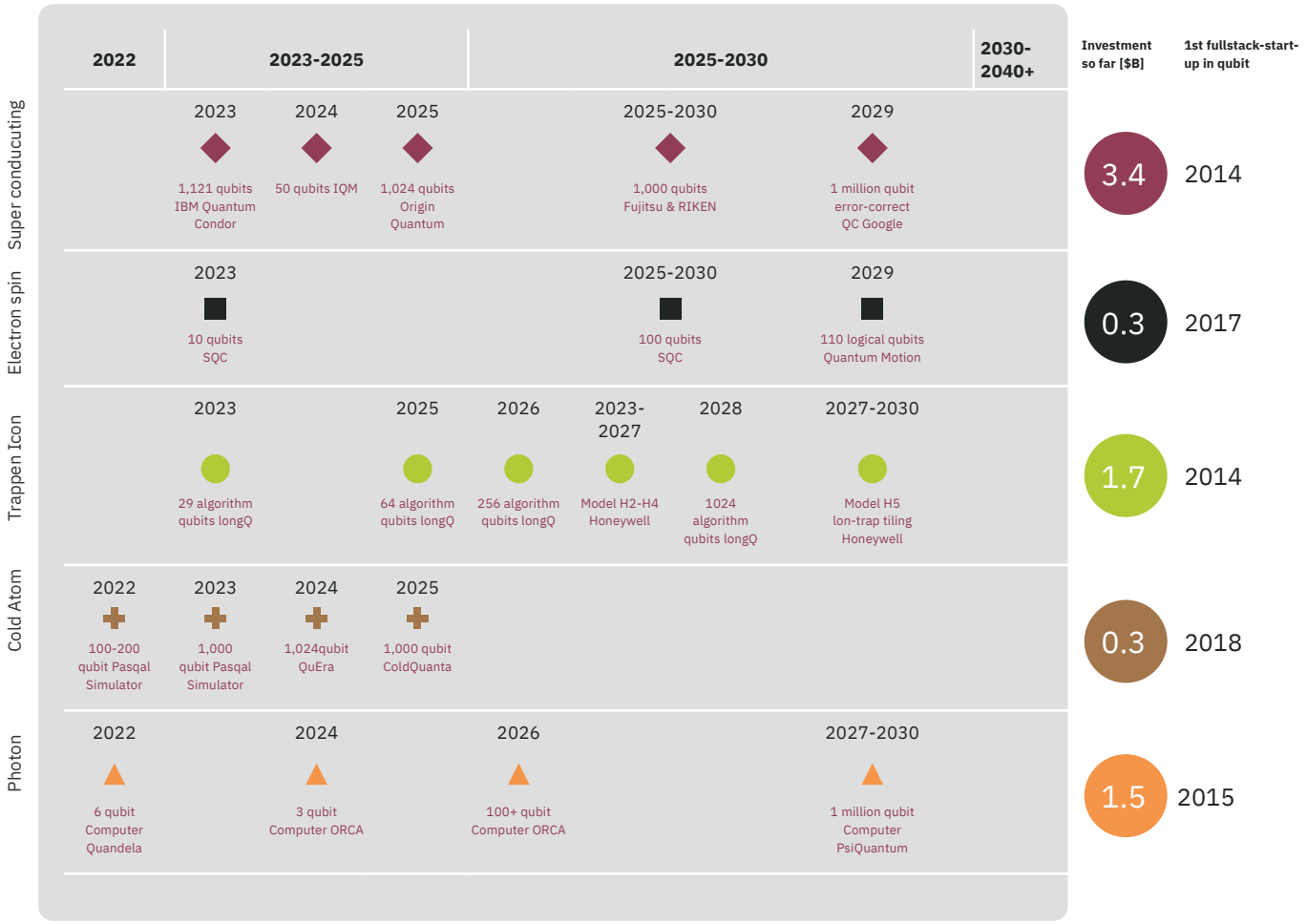


Chart from Arthur D Little Quantum Computing report May 2022, showing different full-stack players racing to develop the first LSQC based on 5 different types of qubits, complemented with data on total investment in qubit type and starting year of the first start-up in the respective qubit.[1]

[1] Venturing into Quantum Technology in the Netherlands, Deep Tech Fund, Invest-NL, October 2023

9

Intel, Samsung and TSMC roadmaps

9. Intel, Samsung and TSMC roadmaps

In this section it will be described how the general roadmap of IDRS and IMEC are implemented by the leading-edge foundries such as TSMC, Samsung and Intel.

Unlike in the past, when a single industry roadmap dictated how to get to the next process node, the three largest foundries increasingly are forging their own paths. They all are heading in the same general direction with 3D transistors and packages, a slew of enabling and expansive technologies, and much larger and more diverse ecosystems.

Roadmaps for all three show that transistor scaling will continue at least into the 18/16/14 angstrom range, with a possible move from nanosheets and forksheet FETs, followed by complementary FETs (CFETs) at some point in the future. The key drivers are AI/ML and the explosion of data that needs to be processed, and in most cases these will involve arrays of processing elements, usually with high levels of redundancy and homogeneity, in order to achieve higher yields.

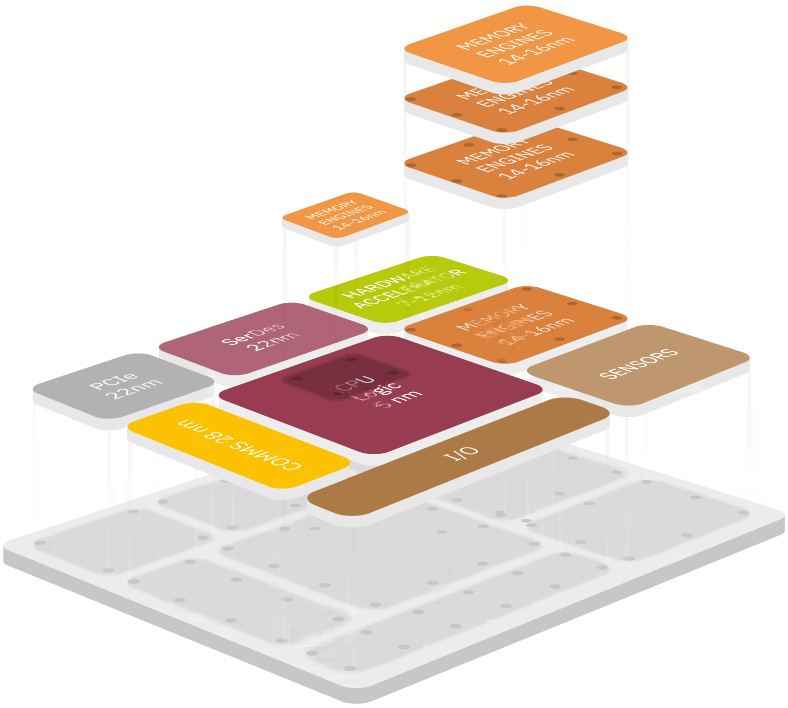
In other cases, these designs may contain dozens or hundreds of chiplets, some engineered for specific data types and others for more general processing. Those chiplets can be mounted on a substrate in a 2.5D configuration, an approach that has gained traction in data centres because it simplifies the integration of high-bandwidth memory (HBM), as well as in mobile devices, which also include other features such as image sensors, power supplies, and additional digital logic used for non-critical functions. All three foundries are working on full 3D-ICs, as well. And there will be hybrid options available, where logic is stacked on logic and mounted on a substrate, but separated from other features in order to minimize physical effects such as heat — a heterogeneous configuration that has been called both 3.5D and 5.5D.

9.1 Rapid and mass customization

One of the biggest changes involves bringing domain-specific designs to market much more quickly than in the past. Mundane as this may sound, it's a competitive necessity for many leading-edge chips, and it requires fundamental changes in the way chips are designed, manufactured, and packaged. Making this scheme work demands a combination of standards, innovative connectivity schemes, and a mix of engineering disciplines that in the past had limited interactions, if any.

Sometimes referred to as "mass customization," it includes the usual power, performance, and area/cost (PPAC) trade-offs, as well as rapid assembly options. That is the promise of heterogeneous chiplet assemblies, and from a scaling perspective it marks the next phase of Moore's Law. The entire semicon ecosystem has been laying the groundwork for this shift incrementally for more than a decade.

But getting heterogeneous chiplets — essentially hardened IP from multiple vendors and foundries — to work together is both a necessary and daunting engineering challenge. The first step is connecting the chiplets together in a consistent way to achieve predictable results, and this is where the foundries have spent much of their effort.
Intel Foundry's current solution, prior to fully integrated 3D-ICs, is to develop what industry sources describe as "sockets" for chiplets. Instead of characterizing each chiplet for a commercial marketplace, the company defines the specification and the interface so that chiplet vendors can develop these limited-function mini-chips to meet those specs. That addresses one of the big stumbling blocks for a commercial chiplet marketplace. All the pieces need to work together, from data speed to thermal and noise management.



Intel's Ponte Vecchio processor with chiplets, also called tiles.[75]

Intel's Ponte Vecchio chiplet design relies heavily on its Embedded Multi-Die Interconnect Bridge (EMIB), first introduced in 2014. "The really cool thing about an EMIB base is you can add any amount of chiplets," said Lalitha Immaneni, vice president of technology development at Intel. "We don't have a limitation on the number of IPs that we can use in design, and it won't increase the interposer size, so it's cost-effective and its agnostic of the process. We have given out a package assembly design kit, which is like your traditional PDK for the assembly. We give them the design rules, the reference flows, and we tell them the allowable constructions. It will also give them any collaterals that we need to take it into our assembly."

Depending upon the design, there can be multiple EMIBs in a package, complemented by thermal interface materials (TIMs), in order to dissipate heat that can become trapped inside a package. TIMs typically are pads that are engineered to conduct heat away from the source, and they are becoming more common as the amount of compute inside a package increases and as the substrates are thinned to shorten the distance signals need to travel. But the thinner the substrate, the less effective it is at heat dissipation, which can result in thermal gradients that are workload-dependent and therefore difficult to anticipate. Eliminating that heat may require TIMs, additional heat sinks, and potentially even more exotic cooling approaches such as microfluidics. Both TSMC and Samsung offer bridges, as well.

–

[75] https://www.extremetech.com/computing/intel-sunsets-its-massive-ponte-vecchio-data-center-gpu

9. Intel, Samsung and TSMC roadmaps

Samsung has embedded bridges inside the re-distribution layer (RDL) — an approach it calls 2.3D or I-Cube ETM — and it's using them to connect sub-systems to those bridges in order to speed time to working silicon. Instead of relying on a socket approach, some of the integration work will be pre-done in known-good modules.

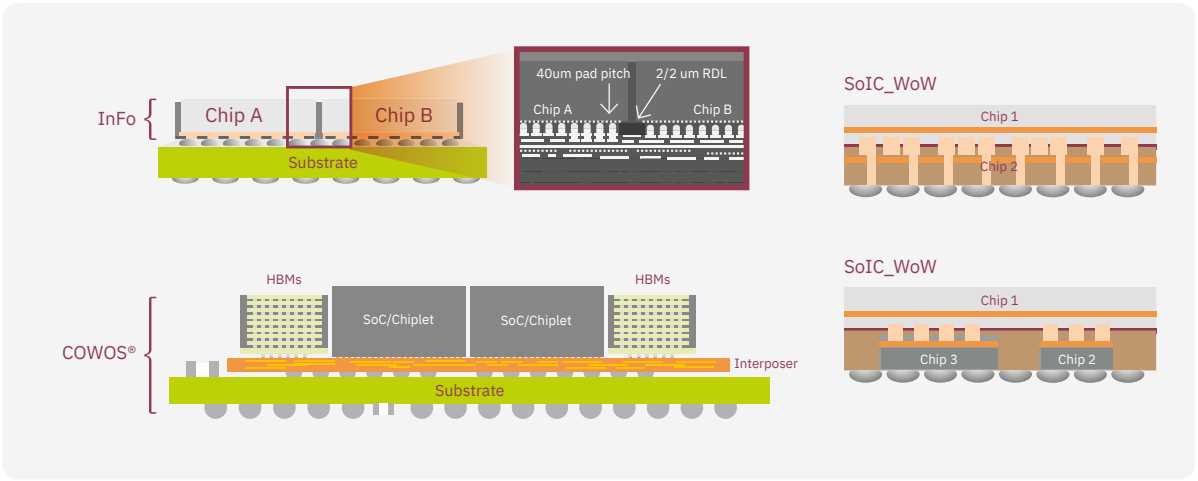
"Putting together two, four, or eight CPUs into a system is something that very sophisticated customers know how to go out and do," said Arm CEO Rene Haas, in a keynote speech at a recent Samsung Foundry event. "But if you want to build an SoC that has 128 CPUs attached to a neural network, memory structures, interrupt controllers that interface to an NPU, an off-chip bus to go to another chiplet, that is a lot of work. In the last year and a half, we've seen a rush of people building these complex SoCs wanting more from us."

Samsung also has been building mini-consortia[76] of chiplet providers, targeted at specific markets. The initial concept is that one company builds an I/O die, another builds the interconnect, and a third builds the logic, and when that is proven to work, then others are added into the mix to provide more choices for customers.

TSMC has experimented with a number of different options, including both RDL and non-RDL bridges, fan-outs, 2.5D chip-on-wafer-on-substrate (CoWoS), and System On Integrated Chips (SoIC), a 3D-IC concept in which chiplets are packed and stacked inside a substrate using very short interconnects. In fact, TSMC has a process design kit for just about every application, and it has been active in creating assembly design kits for advanced packaging, including reference designs to go with them. The challenge is that foundry customers willing to invest in these complex packages increasingly want very customized solutions. To facilitate that, TSMC rolled out a new language called 3Dblox, a top-down design scheme that fuses physical and connectivity constructs, allowing assertions to be applied across both. This sandbox approach allows customers to leverage any of its packaging approaches — InFO, CoWoS, and SoIC. It's also essential to TSMC's business model, because the company is the only pure-play foundry of the three[77] — although both Intel and Samsung have distanced their foundry operations in recent months.

"We started from a concept of modularization," said Jim Chang, vice president of advanced technology and mask engineering at TSMC, in a presentation when 3Dblox was first introduced in 2023. "We can build a full 3D-IC stacking with this kind of language syntax plus assertions."

[77] TSMC also is the largest shareholder (35%) in Global Unichip Corp., a design services company



Device input/output (I/O) densities and the interconnect pitch, achieved using various packaging technologies
Source: TSMC[78]

The figure on above shows the history and future projection of the device input/output (I/O) densities, and the interconnect pitch achieved using various packaging technologies at TSMC. Since 2011, TSMC has been increasing system integration density by applying CoWoS, i.e. silicon interposer, to many devices. In the future, the company says it is planning to dramatically increase the I/O density by stacking multiple chips vertically using 3D IC. The ideal monolithic 3D structure (an IC structure in which process technology is utilized to stack transistors on a single silicon substrate instead of stacking multiple silicon substrates), which was suggested a long time ago, is ultimately expected to be realized, but its materialization will likely take a long time.

Samsung followed with its own system description language, 3DCODE, in December 2023. Both Samsung and TSMC claim their languages are standards, but they're more like new foundry rule decks because it's unlikely these languages will be used outside of their own ecosystems. Intel's 2.5D approach doesn't require a new language because the rules are dictated by the socket specification, trading off some customization with a shortened time to market and a simpler approach for chiplet developers.

–

[78] https://www.tel.com/museum/magazine/report/202106/
https://xtech.nikkei.com/atcl/nxt/column/18/00065/00575/

10

Consolidated Moore’s Law roadmap

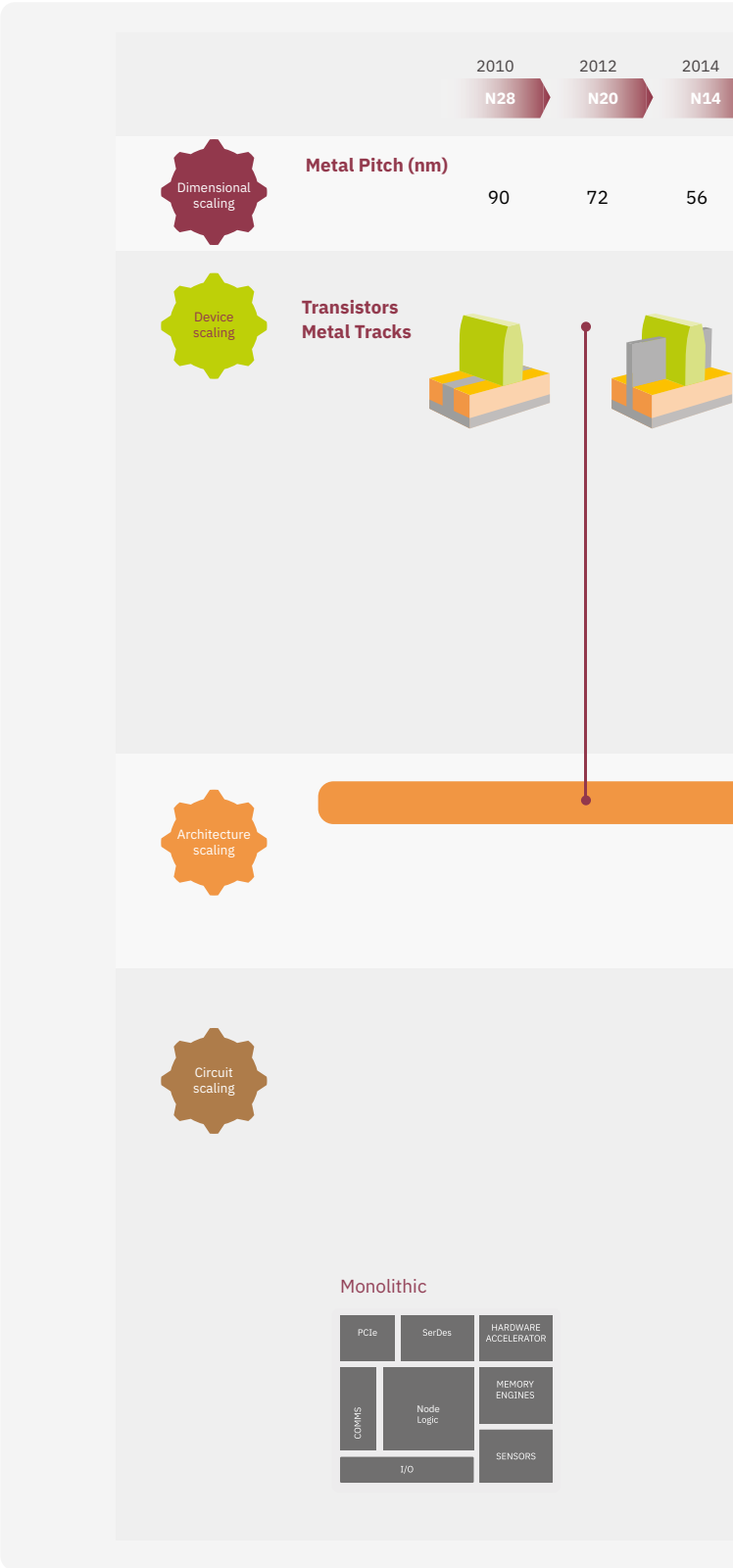
10. Consolidated Moore's Law roadmap

In previous sections it is described that Moore's Law will continue with EEP (Energy Efficient Performance) as Key Performance Indicator. It will be driven by the four scaling engines (1) geometrical scaling, (2) device scaling, (3) circuit scaling and (4) architecture scaling.

Together these scaling engines has to address the scaling walls (See Section 4) that semiconductors are facing: (1) geometrical scaling wall, (2) memory wall, (3) power wall, (4) sustainability wall and (5) cost wall.

Dimensional and geometrical scaling will continue as long as these developments pay off in one or more of the main segments in digital chips.

The main novel development for the next decade is system scaling: the combination of circuit scaling and architecture scaling. In circuit scaling, large chip designs are segmented in chiplets. Chiplets are small, modular chips serving a specific function, such as CPUs or GPUs that can be mixed and matched into a complete system. The Lego-like approach hands manufacturers the flexibility to compose a system cost-effectively with lower entry costs for new chip designs and increased energy efficiency and performance.
Chiplets break down complex chips into smaller, specialized components, providing enhanced flexibility and customization, leading to faster time to market and shorter upgrade cycles. This modular approach allows manufacturers to integrate components tailored to specific tasks, resulting in versatile and efficient chips. Chiplet technology overcomes constraints like reticle size and the memory wall, which traditionally hinder performance and scalability. It also reduces manufacturing costs, improves yield rates, and enhances supply chain resilience by enabling sourcing from multiple suppliers.




2016 2018 2020 2023 2025 2027 2029 2031 2033 2035 2037 2039
N10 N7 N5 N3 N2 A14 A10 A7 A5 A3 A2 SUB A2
Continued dimensional scaling
48 40 28 23 22 21 18 18-16 18-14 16-12 16-12 14-10
Device and innovations
FinFET
NanoSheet
CFET
2DFET, ?, ...
finFET finFET finFET GAANS FET GAANS FET FSFET Introduction Scaled Extended Introduction Scaled
Metal Tracks →
6 6 6 5-6 5 5 4-5 4-5 4-3 4-3 3
Other devices
Photonic: Mach-Zehnder
Quantum: qubits
Von Neumann
Artificial Intelligence (AI)
Neuromorphic
Quantum
Monolithic
Back-side Power Delivery Network
Context aware interconnected
Back-side Power
Back-side Power + signal
160-65 115-45 90-42 40-34 40-28 tbd tbd
...
Chiplets on interposer
PCIe 22nm
SerDes 22nm
HARDWARE ACCELERATOR 7-12nm
COMMS 28 nm
MEMORY ENGINES 14-16nm
Logic 5 nm
MEMORY ENGINES 14-16nm
SENSORS
I/O
Chiplets 3D integrated

10. Consolidated Moore's Law roadmap

Due to these advantages, leading companies like AMD and Intel have adopted chiplet designs in products such as AMD's EPYC processors and Intel's Ponte Vecchio GPUs. This widespread adoption underscores the growing importance of chiplets in high-performance computing applications.

Chiplets are also key in some new architectures. In Artificial Intelligence processors chiplets for memory and logic can be stacked to minimize the distance the data has to travel. Also, for other new architectures combinations of e.g. electronic and photonic devices are combined.

New architectures will also lead to new devices. GPUs for Artificial Intelligence can be made with transistors as devices but also with photonic Mach-Zehnder interferometers as device. US company LightMatters[79] claims that their GPU with 65.000 photonic devices is 5-10x faster than Nividia's GPU with 54 billion transistors.

Also, quantum computing comes with their own type of devices, called Qubits of which many different types do exist.

10.2 Chiplet Market Forecast

In 2025, the chiplet market is forecasted to reach a value of $5.7 billion, according to a report published by Markets and Markets[80]. This represents a compound annual growth rate (CAGR) of 18.9% from 2020 to 2025. In 2032, the chiplets market is expected to reach a value of $58.5 billion, according to a report published by Transparency Market Research. This represents a CAGR of 23.9% from 2021 to 2031. This forecast takes into account the increasing demand for high-performance computing and data analytics, as well as the growing trend towards modularity and customization in electronics design.

The IDTechEx report[81] on the topic, "Chiplet Technology 2025-2035: Technology, Opportunities, Applications", highlights the transformative potential of chiplets. It predicts the market will reach $411 billion by 2035, driven by high-performance computing demands across sectors such as data centres and AI. This means that about 1/3 of all chips will be manufacturing from chiplets with heterogeneous integration in 10 years.
The IDTechEx report analyses key sectors where chiplet technology plays a crucial role:

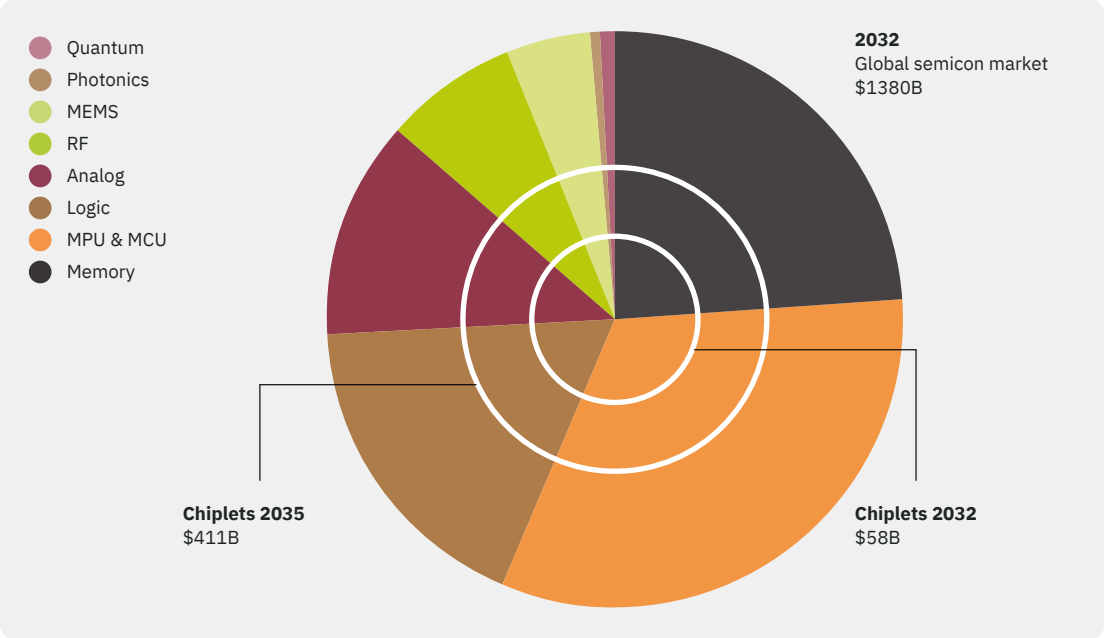
- **Server (57%)**: High-performance computing demands drive significant adoption.
- **PCs (26%**): Enhanced performance and customization are key drivers.
- **Automotive Industry (8%)**: The integration of diverse functionalities supports automotive innovations.
- **Mobile Phones (7%)**: Chiplets contribute to advanced functionality and efficiency.
- **Telecommunications, 5G, IoT (2%)**: Chiplets enable efficient network solutions.
- **Others**: Various applications benefit from the modularity of chiplets.

–

[79] https://spie.org/news/photonics-focus/marapr-2022/harnessing-light-for-photonic-computing

[80] https://anysilicon.com/the-ultimate-guide-to-chiplets/

[81] https://www.idtechex.com/en/research-article/chiplet-market-growth-forecast-to-us-411-billion-by-2035/31905


Quantum
Photonics
MEMS
RF
Analog
Logic
MPU & MCU
Memory
2032
Global semicon market
$1380B
Chiplets 2035
$411B
Chiplets 2032
$58B

11

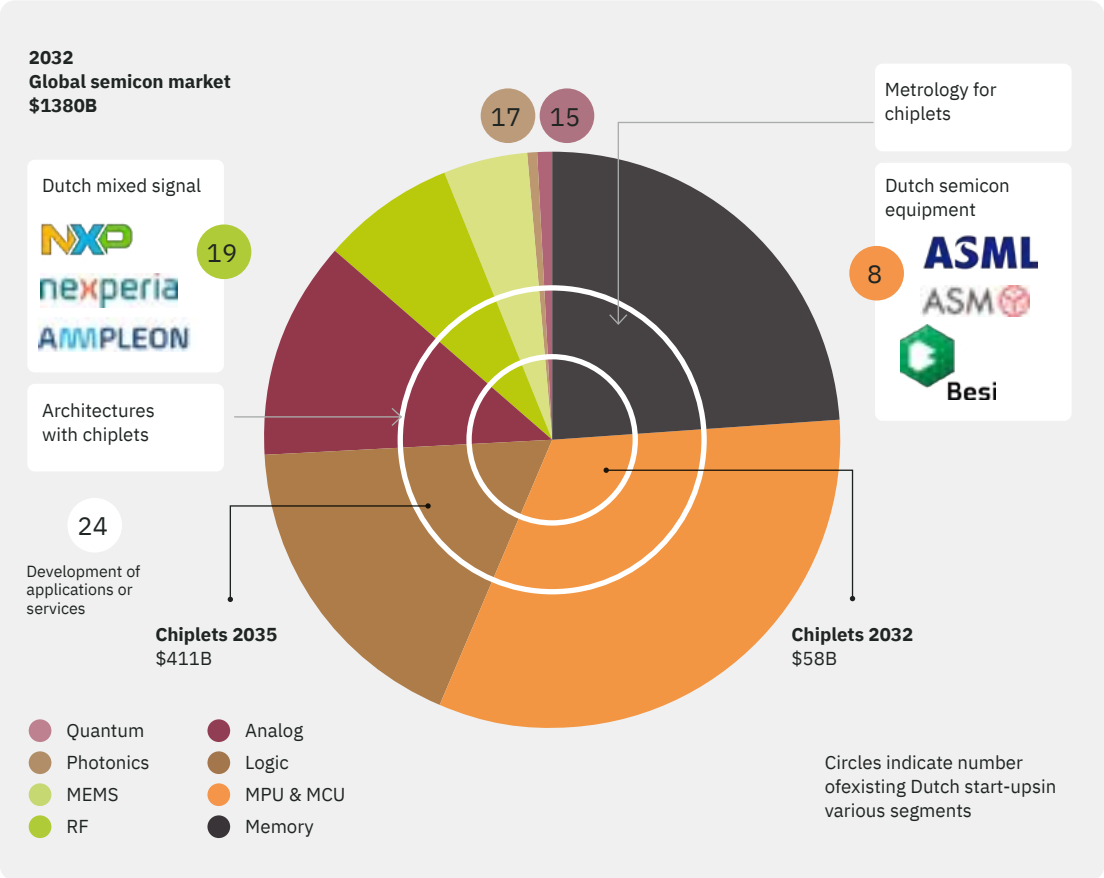
Top-down Dutch opportunities

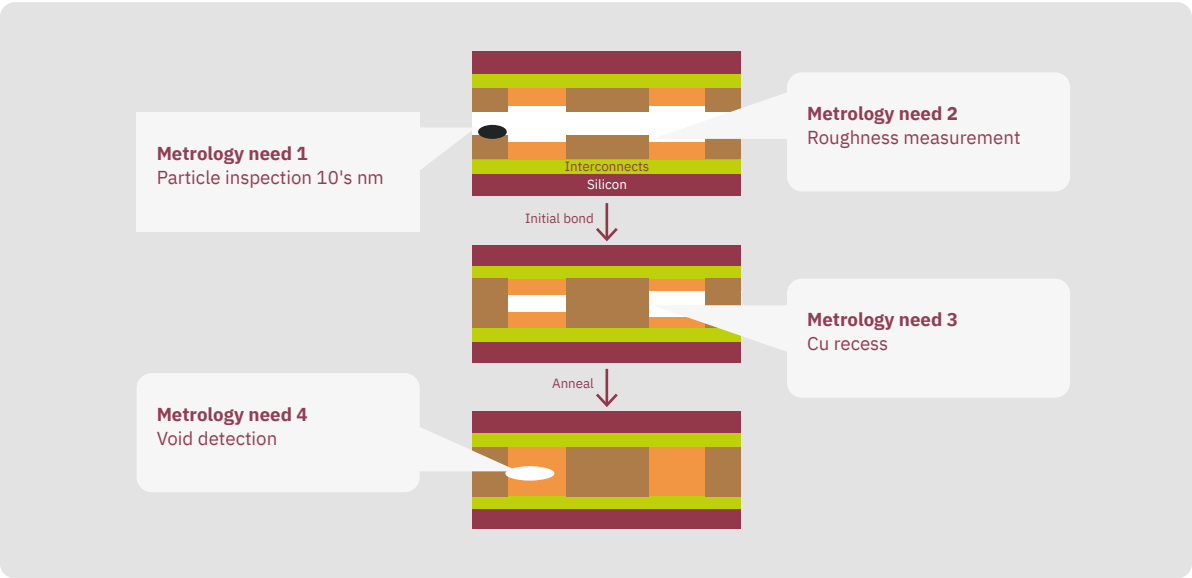
11. Top-down Dutch opportunities

In previous sections the future of Moore's Law has been discussed that identifies new opportunities. In this section they will be linked to the Dutch semicon ecosystem. Many of new developments in geometrical and device scaling are well addressed by leading Dutch companies such as ASML and ASM, while in the upcoming circuit scaling BESI is leading in hybrid bonding.

Therefore, opportunities developed by start-ups should be complementary to these parties, preferably even help to strengthen their position. In this way the start-up can enter the market with support from Dutch world leaders. Metrology and inspection equipment have this opportunity.

Another opportunity area is opened by chip-lets where new chip designs are realized by combining chips into a single package like LEGO-bricks. This opens opportunities for IC designs and architectures even for chips that are needed in relatively small series.





In heterogeneous integration of chiplets novel metrology methods are required to achieve high yield.[82]

11.1 Metrology opportunities

An opportunity area is metrology and inspection to ensure yield control of the primary processes such as lithography (ASML), deposition (ASM) and packaging (BESI). Inspection equipment looks for unwanted defects in chips. Metrology equipment measures the tiny structures and materials in chips, ensuring the devices meet a given specification. In the next section examples will be described of metrology and inspection that are needed in the various nodes of Moore's Law with main focus on chiplets.

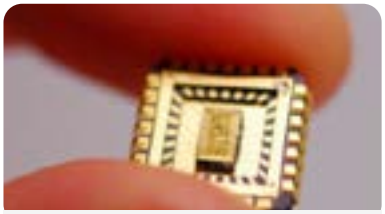
- **Impact on semicon**: yield enhancement in novel devices and chiplets that are needed in high volume.
- **Probability**: metrology is needed in all phases of a device lifecycle. In research (yield) learning can be based on low frequency sampling. As learning progresses in development towards high-volume manufacturing, the throughput has to increase in these phases. This makes it possible to start with a minimal viable product in research.
- **Fit to Dutch ecosystem**: the Dutch semicon ecosystem is very strong when it comes to equipment manufacturing with many suppliers in opto-mechatronics such as VDL-ETG, Prodrive, Demcon and Sioux.
- **Benefit for existing Dutch semicon parties**: the Dutch equipment companies are well position in the primary manufacturing process of new devices and chiplets. They can benefit from new parties that make equipment to enhance the yield without having to develop it themselves. The start-ups can benefit from the domain knowledge of the big parties that can also coach the start-up in the market.

–

[82] https://spectrum.ieee.org/hybrid-bonding

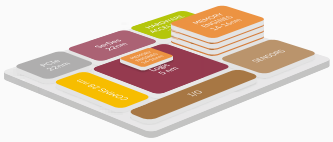
10. Consolidated Moore's Law roadmap

In Section 12 examples will be described of metrology opportunities that arise with upcoming scaling trends.



Monolithic Application Specific IC

+ High performance, due to full integration
- Long lead time, including design verification
- High start cost, so no small series



Chiplets

+ Very high performance, as chiplets can be stacked
+ Short lead time with standard chiplets
+ Low start cost, so good for small series

Chips on Printed Circuit Board

- Low performance, due to long distance over PCB
+ Short lead time, as standard chips are combined
+ Low start cost, so good for small series

11.2 Chiplet design and architecture opportunities

Chiplets create the opportunity for "democratic" chip manufacturing. In monolithic chips the starting costs for e.g. lithography reticles, can be hundreds of million Euros and that will be a barrier to realize new designs. When a new design is made from chiplets these Lego-bricks can be purchased from various foundries and the entrance barrier cost is limited to the integration. This enables high performance chips (that would require application specific ICs (ASICs) in monolithic design) in small series at relative low entrance barrier cost. Examples of this approach will be described in Section 13.

Chiplet-based products have recently emerged across multiple industry-leading semiconductor vendors and cloud providers, including AMD, Intel, Broadcom, Amazon and Ampere, as a potential technology to extend Moore's law . Their remarkable growth has been driven by a mixture of motivations: (1) to avoid very large die sizes for improved yields; (2) to enable product diversification through mix-and-match strategies for reuse and cost savings, and (3) to promote heterogeneous integration - integrating chiplets from multiple process nodes for function, cost and/or performance optimization. One expects these motivations are equally relevant to all companies developing semiconductor products. As an example, only 60% or less of the logic in domain-specific accelerators is actually domain specific. Companies could,

[83] https://www.opencompute.org/blog/building-an-open-chiplet-economy

in theory, buy IP in the form of chiplets, to reduce both their costs and time-to-market.

In practice, technical and business challenges have largely restricted the development of chiplet-based designs to organizations with the financial resources to control the entire development stack for their product.

Analysts from Accenture segment silicon IC consumption across five primary markets in order of size: Data Processing (40%), Communication (30%), Automotive, Consumer devices and Industrial (each about 10%). As with other applications for software and hardware, open standards and activities can accelerate the development and growth of robust multi-vendor chiplet supply chains with lots of choice for the end user to create the customizations they desire.

Dutch companies NXP and Nexperia have experience in the approach in e.g. automotive chips. Start-ups can benefit from this experience while their products can be manufactured in cooperation with the big companies. Chiplets can be a 'blessing in disguise' for Dutch and EU semicon industry: something that seems bad or unlucky at first but results in something good happening later. Let's face it: continuation of traditional Moore's Law, where the figure of merit is the number of transistors per mm2, is not easy as huge technical barriers have to be become such as energy demand. Chiplets are crucial to overcome these barriers, and we are entering an era where Energy Efficient Performance (EEP) is the new figure of merit of Moore's Law. EEP is enhanced not just by increasing the number of transistors per mm2, but by the ability to integrate advanced chiplets. This enables new designs and architectures to outperform even if they are not based on highest resolution technology.

In this new era Dutch and EU semicon industry can design and manufacture advanced chips by mix and match chiplets from existing semicon leaders such as TSMC with locally produced chiplets that fulfil a specific function.

If we are optimistic: chiplets might even enable new Dutch or EU IC architectures. When an IC design has revolutionary impact, it is a called a new architecture. For example, Nvidia revolutionized computing with their AI chips that use processor cores and memory in a different way. Most likely the next architecture will be using chiplets.

It all starts with an IC design that solves a relevant problem even if it has a specific application.

- **Impact on semicon**: the number of designs realized per node is decreasing due to ever higher design cost. Chips made from chiplets, that are produced in high volume at low cost, have a much lower entrance barrier. This is especially of interest for start-ups that want to enter the market with new design.
- **Probability**: when the design process of a new chip is started based on the chiplet philosophy, the probability of a successful launch is much higher as entrance barrier cost are lower and time to market is shorter.
- **Fit to Dutch ecosystem**: there are many Dutch companies that make deep-tech products that can potentially benefit from dedicated chips as they also use dedicated opto-mechatronics. This is not affordable with monolithic chip layout as the series are small.
- **Benefit for existing Dutch semicon parties**: novel designs from start-ups have to be manufactured. As chiplets are made worldwide, integration can be realized locally by Dutch parties such as NXP and Nexperia.

In Section 13 examples will be described of chiplet design and architecture opportunities that arise with upcoming scaling trends.

12

Metrology opportunities

12. Metrology opportunities[84]

During this entire manufacturing process in both the fab and packaging house, a chip line undergoes a number of inspection and metrology steps using different equipment. Inspection equipment looks for unwanted defects in chips. Metrology equipment measures the tiny structures and materials in chips, ensuring the devices meet a given specification. Both inspection and metrology equipment help ensure that a given chip line can be manufactured with good yields and few, if any, defects. This means cost of metrology and inspection has to be in line with value of added yield by the very same metrology and inspection.

The current metrology equipment types are capable to support current semicon technology. New semicon technologies are emerging as indicated in the previous sections, which require novel metrology and inspection equipment. For example, migration from today's FinFET's, to a next-generation gate-all-around (GAA) transistors. GAA is a complex 3D-like transistor structure. This requires metrology and inspection that can look inside the device not just from the top. In the metrology world, there is also a growing number of challenges with the next wave of 3D NAND devices, DRAMs and other chips.

In fact, seven grand challenges that must be addressed in the metrology field have been addressed by the CHIPS for America office. They are:

- **Metrology for Materials Purity, Properties, and Provenance.** Meet increasingly stringent requirements for semiconductor materials purity, physical properties, and provenance across a diverse supply chain through development of new measurements and standards.
- **Advanced Metrology for Future Microelectronics Manufacturing.** Ensure critical metrology advances keep pace with cutting-edge and future microelectronics and semiconductor manufacturing.
- **Enabling Metrology for Integrating Components in Advanced Packaging.** Provide enabling metrology spanning multiple length scales and physical properties for acceleration of advanced packaging for future-generation microelectronics.
- **Modelling and Simulating Semiconductor Materials, Designs, and Components.** Improve tools needed to effectively model and simulate future semiconductor materials, processes, devices, circuits, and microelectronic system designs.
- **Modelling and Simulating Semiconductor Manufacturing Processes.** Seamlessly model and simulate the entire semiconductor manufacturing process, from materials inputs to chip fabrication, system assembly, and end products.
- **Standardizing New Materials, Processes, and Equipment for Microelectronics.** Standardize the methods that will support and accelerate the development and manufacturing of microelectronics and advanced information and communications.
- **Metrology to Enhance Security and Provenance of Microelectronic based Components and Products.** Create the metrology advances needed to enhance the security and provenance of microelectronic components and products across supply chains and increase trust and assurance.

[84] https://marklapedus.substack.com/p/us-chips-program-launches-metrology?r=3qbuhb&utm_campaign=post&utm_medium=web&triedRedirect=true

As stated, today's metrology tools are capable. But to address these grand challenges, the industry needs new breakthroughs. In response, research and development is world-wide on-going for several advanced metrology tool types[85].

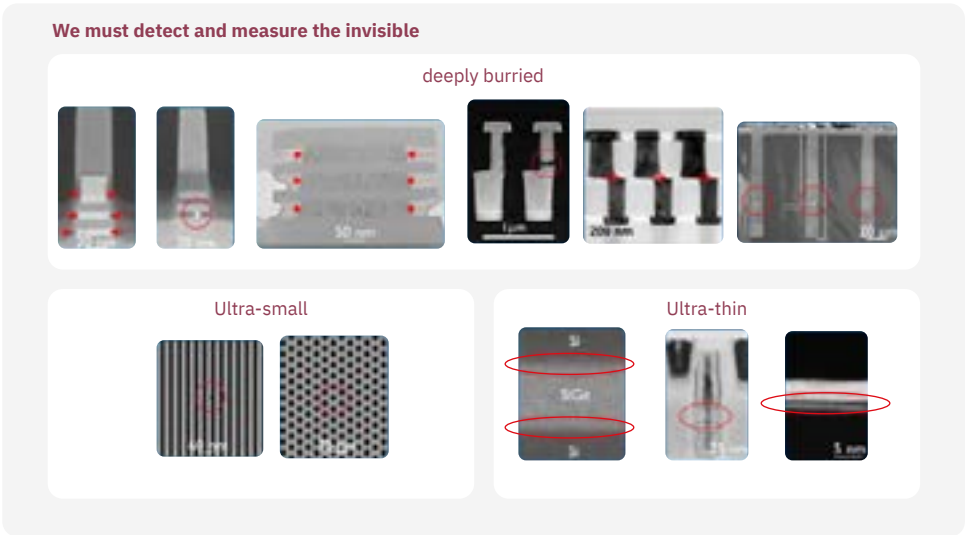
"Metrology and inspection have entered a new era," said Anne-Laure Charley, R&D manager at imec's ITF, in a recent presentation. "We are indeed transitioning from a world where metrology was the first step to be reduced or even removed, to a world where it has become a real technology enabler. And we have new challenges in front of us that drive new, innovative approaches."

Some of the greatest challenges facing metrology and inspection involve the detection of hidden defects or features in increasingly 3-dimensional structures, both at the front and back end of line. The need for faster yield ramping depends on the early identification of systematic defects, which can be design- or process-related.

Leading-edge fabs also are integrating advanced data analytics platforms for critical measurement and inspection to enhance precision and make the most of data from various sources. "Data is said to be the gold of the 21st century, but really understanding the data is the gold," said Dieter Rathei, CEO of DR Yield. "What's happening all around the semicon industry is we have far too much data. The value is hidden in the data, and to mine the gold from it, you need tools."

Novel metrology and inspection techniques are often introduced in the research phase of new device types. Here significant yield enhancement is achieved even when a small area can be measured due to limited throughput of the equipment. In ramp-up and High Volume Manufacturing phase of the device, it is important that the throughput is high. Only when a large area can be measured the limited number of defects can be found.

We must detect and measure the invisible



–

[86] https://www.reddit.com/r/hardware/comments/1g9321b metrology_advances_step_up_to_sub2nm_device_node/

12. Metrology opportunities

In the following sections a kaleidoscopic overview will be given of several novel metrology and inspection techniques that can be real technology enablers for emerging semicon devices.

12.1 Critical Dimension Small Angle X-ray Scattering (CDSAXS)

In the fab, it's important to measure the critical dimensions (CDs) of a given chip line. You need to measure the height, length and width of a structure. Generally, the industry uses various tools to measure the CDs of a chip structure. For complex devices, one metrology tool type, called optical CD (OCD), is the primary system used to measure the CDs in chips. Other metrology tools are used, as well.
OCD uses a technique called scatterometry to measure the feature sizes and shape structures in chips. But optical scatterometry is reaching its fundamental limits, according to NIST, as there is a fundamental relationship between the detail that can be detected and the wavelength of the light used in scatterometry.
In response, NIST and others are developing a technology called CDSAXS for next-generation dimensional metrology. CDSAXS , an X-ray technique, is used for 3D NAND applications today. But the technology is limited and too slow for logic devices. Under the CHIPS Metrology Program, researchers will expand the development effort at NIST for CDSAXS to address critical industry needs for in-line dimensional metrology.

12.2 IR scatterometry

"IR scatterometry extends from nanosheet to CFET architectures," said Nick Keller, director of applications development at Onto Innovation. "And CFETs are an interesting case, because you're moving up vertically. From an optical standpoint, you're actually getting more signal because you have more material volume per unit area, so more interaction with the light. But the rub of that is that customers want to extract more parameters. So, the challenges may balance out. You're getting more sensitivity, so more information, but since more parameters are important, there's potentially more correlation between parameters."

Others agree. "Scatterometry is a powerful metrology technique, which can extract many parameters of interest," said imec's Charley. In addition, correlation of scatterometry method results with reference data from AFM, for instance, can be improved with appropriate machine learning algorithms. "When we introduce machine learning on top of our standard methods, we improve significantly machine-to-reference correlation."

12.3 EUV scatterometry

Today, chipmakers use a combination of electron scanning modalities, OCD and other techniques to measure structures. TEMs are also used. CDSAXS is used in some cases. But the industry needs a new system, which can measure beyond visible and ultraviolet wavelengths. This involves the spectral region from 10nm to 150nm. Researchers are developing novel extreme ultraviolet (EUV) optics, tabletop sources, and methodologies for inspection and process control. Today, though, no such tool exists.

[87] https://www.nist.gov/image/diagram-cdsaxs-method

12.4 Grain boundary imaging

Nanosheet materials will play an important role in emerging transistor layouts such as CFET. Graphene can be used as conductor, theoretically 40% better than copper. Other 2D-materials such as MoS and WSe can be used as channel materials. It is of prime importance that the material has large grains as grain boundaries ruin the superior performance.

Grain boundaries in 2D-materials are formed by the joining of islands during the initial growth stage, and these boundaries govern transport properties and related device performance. Although information on the atomic rearrangement at graphene grain boundaries can be obtained using transmission electron microscopy and scanning tunnelling microscopy wafer scale information regarding the distribution of graphene grain boundaries is not easily accessible. Special types of optical microscopy[88] can potentially be used to observe the grain boundaries of large-area graphene (grown on copper foil) directly, without transfer of the graphene.

12.5 Advanced analytical electron tomography

Other CD metrology tools are also used in fabs. For example, the semicon industry uses transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM). Both are used for identifying structural and morphological characteristics of a device. However, TEM and STEM imaging methods struggle with complex semiconductor device architectures, according to NIST. Under the CHIPS Metrology Program, researchers will develop atomic-resolution methods for complex semiconductor devices via electron tomography.

12.6 Voltage Contrast Scanning Electron Microscopy

While the Scanning Electron Microscopes (SEM's) creates a detailed image, the low throughput has restricted where it is used. A fab manager once estimated that to scan a 1cm² die with an e-beam could take close to a week. And therein lies the reason why e-beam technology continues to be used sparingly. Coulomb interactions between the electrons are causing this low throughput. Multibeam Scanning Electron Microscopes (SEM's) are used to enhance the throughput.

Since e-beams also enable voltage contrast inspection, they are used in-line in circumstances when only a voltage clamp can detect defects, such as the extremely deep features of a 3D NAND memory channel hole.
"We are shipping to production fabs," said Gary Zhang, vice president of HMI Product Management at ASML. "They run a number of e-beam systems for voltage contrast inspections. And that is true for both memory and logic customers."

Voltage contrast inspection relies on the build-up of surface potential difference exerting field to influence the trajectories of secondary electrons.

12.7 Down sampling optical and e-beam inspection

"The optical inspection process, often considered the workhorse in defect detection, faces limitations in terms of wavelength and resolu-

[88] https://www.nature.com/articles/nature11562

12. Metrology opportunities

tion. As critical dimensions continue to shrink in advanced nodes, optical inspection is being pushed to its limits. And despite throughput improvements, full die and full wafer e-beam inspection still have a long way to go before they are ready for high volume manufacturing," said Le Hong, director of fab Solutions, Calibre Semi Solutions at Siemens EDA. Furthermore, optimizing the sensitivity of optical inspection to capture genuine defects while minimizing false/nuisance ones has become increasingly challenging."

To address these challenges, Hong points to growing demand for software capable of intelligently downsampling from optical to scanning electron microscope (SEM) review, particularly in high nuisance regimes. "This software must also possess the performance required for inline use in HVM. Siemens EDA's Calibre SONR product offers a cutting-edge solution that leverages AI-driven algorithms for optical to SEM review downsampling. This methodology is not only design and process aware but also boasts performance that is fully inline ready for HVM applications," said Hong. "The feature-driven downsampling algorithm is well-suited to effectively handle the common occurrence of high nuisance counts during hot scans.

12.8 Element-Specific X-ray Imaging

Today, X-ray computed tomography (CT) is used for 3D characterization of subsurface feature sizes and shapes in chips. The most advanced spatial resolution is achieved using X-ray synchrotron beamlines, which is expensive.

Under the CHIPS Metrology Program[89], researchers hope to develop an advanced X-ray CT instrument for 3D, element-sensitive imaging at the micrometer to nanometer ranges.

12.9 Scanning Probe Microscopy (SPM)

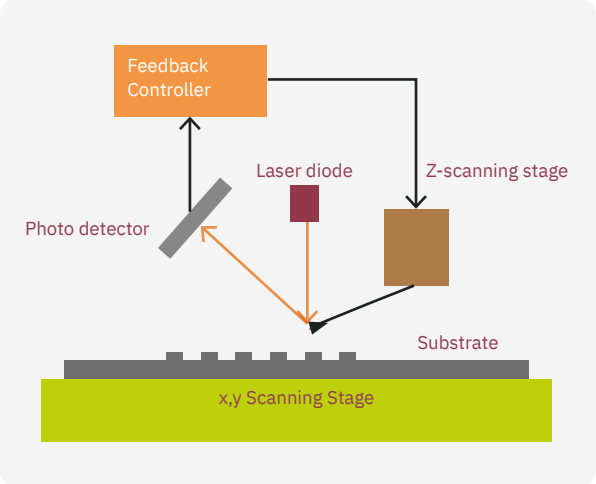
Other metrology tools are used for more specific applications in the semicon industry today. One tool type, Scanning Probe Microscopy (SPM), is a very-high-resolution type of scanning microscopy, with demonstrated resolution on the order of fractions of a nanometre, more than 1000 times better than the optical diffraction limit. The information is gathered by "feeling" or "touching" the surface with a mechanical probe. Piezoelectric elements that facilitate tiny but accurate and precise movements on (electronic) command enable precise scanning.

Advanced packages, such as system-in-package (SiP) and others, are prone to thermal challenges and mechanical incompatibilities among the various interfaces and building blocks of the package. Currently, there are no established methods to perform microscale thermomechanical measurements here. In R&D, researchers are working on indentation and scanning probe-based measurement techniques for the thermo-mechanical properties of materials and structures used in advanced packages.

Also with this scanning microscope throughput is a challenge. Nearfield Instruments , a Dutch start-up, has developed an architecture of multi-head MiniSPMs that are positioned extremely fast and accurate in parallel. They set a new record in 3D SPM metrology Move-Acquire-Measure time. This enables high sampling rate for determination of lot to lot, wafer

[89] https://www.nist.gov/programs-projects/nanoscale-element-specific-x-ray-imaging-integrated-circuit-metrologyd

to wafer, and within wafer process variation Also with this scanning microscope throughput is a challenge. Nearfield Instruments[90], a Dutch start-up, has developed an architecture of multi-head MiniSPMs that are positioned extremely fast and accurate in parallel. They set a new record in 3D SPM metrology Move-Acquire-Measure time. This enables high sampling rate for determination of lot to lot, wafer to wafer, and within wafer process variation.



Principe of a Scanning Probe Microscope.[91]

12.10 Infra-Red Atomic Force Microscopy (IR-AFM)

Infrared atomic force microscopy-based spectroscopy (IR-AFM) is an emerging technique that provides chemical analysis and compositional mapping with spatial resolution far below conventional optical diffraction limits. IR-AFM works by using the tip of an AFM probe to locally detect thermal expansion in a sample resulting from absorption of infrared radiation. AFM-IR thus can provide the spatial resolution of AFM in combination with the chemical analysis and compositional imaging capabilities of infrared spectroscopy.

TNO is developing this technology for metrology of Line Edge Roughness, Edge Placement Error and 2D-materials.

12.11 GHz half wavelength contact acoustic microscopy (HaWaCAM)

Materials in 3D semiconductor devices may be optically opaque, posing problems for traditional optical metrology methods. One solution is to use acoustical waves, which present the double advantage of not being hampered by optically opaque layers and allowing for penetration depths of 10's of µm at sub-µm wavelengths; which is considerably larger than most traditional optical methods.

–

[88] https://www.nature.com/articles/nature11562

12. Metrology opportunities

TNO works on a novel acoustic metrology method using GHz ultrasound waves to measure deeply buried subsurface features (>5 µm). The concept consisted of a GHz acoustic transducer integrated above the tip of a custom designed probe of an Atomic Force Microscope (AFM), which is then scanned across a sample. The method uses non-damaging solid-solid contact without the need for liquid coupling layers – in contra contrast to conventional acoustical microscopy. This allows for the use of much higher acoustic frequencies, hence higher on-axis resolutions.

TNO is developing this technique for alignment and overlay measurements, deep 3D defect imaging and characterization. Examples includes void detection in hybrid bonding.

12.12 Nanocalorimetry

Nanocalorimetry measures the thermal properties of small samples at fast rates. MEMS-based differential scanning calorimeters (DSC) and scanning probe calorimeters are used for this work.

Under the CHIPS Metrology Program, researchers would like to build upon today's nanocalorimeters. The goal is to develop systems that directly measure transition temperatures, reaction temperatures and specific heat capacity of nanoscale materials.

12.13 Preparing for hybrid bonding: recess

A number of fabs are exploring which metrology/inspection methods are best used for hybrid bonding prior to and after the bonding process. Hybrid bonding brings together small copper pads (<10 µm) that are slightly recessed in dielectric fields (typically SiCN). White light interferometry, a type of optical profiler, can be used to characterize the CMP edge roll-off at the wafer edge, but it may also be used to measure the copper recess depth prior to bonding.

Phase-shift interferometry (PSI) mode in WLI is used to monitor topography at the wafer level, including copper recess depth. There are strict specifications on recess depth across the wafer. Too little copper can cause opens, while too much can cause copper extension beyond the barrier oxide and potential shorts.

When it comes to measuring copper recess, overlap exists between metrology techniques especially in case of WLI profiler and the other leading method of Atomic Force Microscopy (AFM). While WLI profiler combines 4x throughput with the ability to map millions of copper pads in the same die, AFM provides the exact offset between oxide and copper compensating WLI measurements. AFM also expands in range of scan speed and scan length, covering entire die flatness post-CMP as well as pad recess.

12.14 Preparing for hybrid bonding: particles

Particles as small as tens of nm's can hamper the forming of proper electrical contacts between the chiplets in the second part of hybrid bonding. As rework is often not possible, inspection of the surface of the chiplets at production speed is needed for high yield.

FastMicro is a Dutch start-up that develops optical inspection techniques to detect the particles prior to bonding. Challenge is the combination of finding the small particles (<50 nm) at high speed (many wafers per hour). Potentially techniques like Optical Coherence

Tomography or white-light interferometry enable this.
It is an optical surface measurement method wherein the localization of interference fringes during a scan of optical path length provides a means to determine surface characteristics such as topography with height resolution of 10's of nanometers.

12.15 Void detection after hybrid bonding

Once the hybrid bond between two chiplets has been established it might be of interest to detect if voids do exist as this will hamper proper electrical contact. In more advanced products, where many chiplets are combined in a single package, detection of the voids gives an early warning of yield loss, minimizing the sunk cost of more good dies added to the package.

12.16 X-ray diffraction for advanced packaging

Both optical metrology and SEM-based tools are mainstream and in production today, while X-ray diffraction imaging is meeting specific production needs, including in advanced packaging.

"We have experience with customers using X-ray diffraction imaging with CoWoS (TSMC's chip on wafer on substrate), where they're effectively stacking chips on top of each other and then grinding silicon from the substrate, because it is effectively dead mass in the structure," said John Wall, UK site manager at Bruker. "What they found is that the XRDI technique can detect cracks, edge defects, and multiple problems that can cause the device to fail catastrophically during the back-end process and before packaging."

12.17 Quantum-enables Scanning Probe Microscopy

Quantum sensors turn the extreme vulnerability of quantum systems to external perturbations into an asset: interactions with physical quantities lead to a change in the transition energy between quantum states, allowing a sensitive and quantitative measurement of the interaction strength.

An example of such a quantum system is the spin of an electron in a magnetic field, which has two possible states: Spin up or spin down. The two states are separated by an energy difference that depends on the strength of the magnetic field. The stronger the magnetic field, the larger the energy difference. By precisely measuring the energy difference between these two spin states, it is therefore possible to draw direct conclusions about the strength of the magnetic field.

The manipulation capabilities of scanning probe microscopes are used to fabricate molecular quantum sensors on the probe tips to detect the tiny electric and magnetic fields of quantum systems at the atomic level. TNO is developing this technique for measuring extreme small currents flowing through semiconductor devices.

[93] https://www.fz-juelich.de/en/pgi/pgi-3/groups/research/atomic-scale-quantum-sensing

12. Metrology opportunities

12.18 Combining metrology and analytics

One of the greatest concerns for process and yield engineers today is controlling process variability, which affects what goes on inside a wafer, as well as wafer-to-wafer, and lot-to-lot results. In fact, across-wafer signatures from many wafer processes are not uncommon. “The location of the die on the wafer is paramount to understanding any type of variation that you are seeing, because on a typical wafer, the optimal performing die (considering both performance and power) form a donut shape,” said Nir Sever, senior director of business development at proteanTecs. “Dies at the centre and the edge of the wafer behave worse than the rest.”


QUADRA
NEARFIELD
INSTRUMENTS
VISION
LEAP
VISION
LEAP

13

Chiplet design and architecture opportunities

13. Chiplet design and architecture opportunities

Chiplets can be a 'blessing in disguise' for Dutch and EU semicon industry: something that seems bad or unlucky at first but results in something good happening later. Let's face it: continuation of traditional Moore's Law, where the figure of merit is the number of transistors per mm2, is not easy as huge technical barriers have to be become such as energy demand. Chiplets are crucial to overcome these barriers, and we are entering an era where Energy Efficient Performance (EEP) is the new figure of merit of Moore's Law. EEP is enhanced not just by increasing the number of transistors per mm2, but by the ability to integrate advanced chiplets. This enables new designs and architectures to outperform even if they are not based on the highest resolution technology.

In this new era Dutch and EU semicon industry can design and manufacture advanced chips by mix and match chiplets from existing semicon leaders such as TSMC with locally produced chiplets that fulfil a specific function. It all starts with an IC design that solves a relevant problem even if it has a specific application.

In the following sections a kaleidoscopic overview will be given of several novel chiplet design and architecture techniques that can be real technology enablers for emerging semicon devices.

13.1 Chiplets for future automotive applications[94]

Autonomous vehicles and associated ADAS systems are driving vehicle electronics content to unprecedented levels. These systems require significant processing power via a centralized system. Chiplets are one solution to distribute functionality on different technologies and can help to secure the supply chain and reduce rising development costs in advanced IC nodes. Chiplets will be a main driver for future systems to enable the maximum performance of a system with a wide variation of functionality and with a heterogeneous integration of different technologies, package types and interconnection. To create the most gain, besides the integration of functions into dedicated technologies, it is also necessary to have standardized interfaces, which allow a proper coupling of the functions, synchronization and performant low-power data submission. To have these standards and implementation as IP in dedicated technologies is a prerequisite, that these technologies are able to be used in chiplet approaches with suitable time-to-market. The availability of these IP is a key enabler for usage of technologies for chiplets.

13.2 Chiplets make case for more apps[95]

AI, high-end computing, and even consumer applications will require the scalability and flexibility of chiplets to meet future packaging needs.

Many chiplet packaging discussions have focused on applications in high-end computing where high density and low latency are needed. But the flexibility and other advantages chiplets offer are also of appeal to makers of complex consumer products, such as AR/VR headsets.

At the recent DesignCon show in Santa Clara, these issues came front and center during a

[94] https://semiengineering.com/chiplets-for-future-automotive-application/

[95] https://chiplet-marketplace.com/news/chiplets-make-case-for-more-apps

panel session titled, “Chiplets in consumer electronics and infrastructure.” What immediately became obvious was that existing silicon-on-chip technology could no longer scale in a cost-efficient manner for consumer electronics.

13.3 Chiplets for inference AI compute[96]

Chiplet packaging is catching on with companies designing high-performance processors for datacenter and AI applications. While familiar names such as Intel and AMD are in this space, so are some smaller startup companies. One of them is D-Matrix, a young company developing technology for AI-compute and inference processors.

D-Matrix has developed patented technologies to solve the physics of memory-compute integration using innovative circuit techniques, ML tools, software and algorithms. Chiplet technology was the way to go in order to develop AI and high-end server processors. “We are able to scale up and rapidly deploy our platform using smaller chips. We can get a lot more silicon and more easily test and verify these parts,” he added.

The chiplet platform, called Jahawk, provides energy-efficient die-die connectivity over organic substrates. The 8-chiplet AI system uses 6nm process technology from foundry company TSMC. The system provides 16 Gbps per wire bandwidth and achieves an energy efficiency of better than 0.5 picojoules/bit.

13.4 Chiplets for Edge AI applications[97]

YorChip is a Silicon Valley start-up focused on Chiplets for Mass Markets. They are leveraging proven partner IP and novel die-to-die technology to deliver off-the-shelf, low cost, secure chiplets at scale. They are developing a complete ecosystem of off the shelf Chiplets. YorChip announces its first Chiplet for Edge AI applications with IP licensed from Semidynamics, the leader in RISC-V IP based in Barcelona. The Semidynamics High-Performance High-Bandwidth Quad Core IP can provide 10 Int8-TOPS per chiplet. Edge AI requires high performance, high bandwidth, and low cost, the target technology is 12nm with a target die size sub 25mm2, delivers with scalable performance and low costs.

Chiplets are expected to sample in Q2 2025 with volume production in early 2026.

“It all starts with an IC design that solves a relevant problem even if it has a specific application.” Several Dutch start-ups are making IC designs. Leading edge IC designs are limited due to high initial cost of e.g. the reticle. Chiplets can boost the number of designs as of lot of initial costs are shared. Therefore, IC design companies are considered as “quartermasters” for the revival of Dutch and European semiconductors.

–

[96] https://www.designnews.com/semiconductors-chips/ai-compute-company-banks-on-chiplets-for-future-processors

[97] https://www.design-reuse.com/news/55551/yorchip-chiplet-ai-semidynamics-risc-v-ip.html

13. Chiplet design and architecture opportunities

13.5 IC design by QBayLogic[98]

QBayLogic, a start-up from Twente, excels at Field Programmable Gate Array (FPGA) and Application Specific IC (ASIC) design. Their scope is from RTL design in VHDL or Verilog, to the use of advanced testing frameworks, to manual layout. On top of being skillful with the conventional approaches, they also harness the power of their innovative method, called “Clash”, to offer results earlier.

13.6 IC design by Sencure[99]

Sencure is a startup from Twente developing chip technology for medical devices and health-tracking wearables. Their product, SNCE800, is an integrated Analog Front-end, Analog-to-Digital-Converter, and Digital Signal Processor chip aimed at high-quality electrophysiological measurements against a low power consumption. It has been specifically designed to capture Brain (EEG), Heart (ECG) and, Muscles (EMG) signals with the highest quality possible.

13.7 IC design by Chain-IC[100]

Chain-IC, a start-up from Twente, is a mixed-signal system design company specialized in analog and mixed-signal IC design. They offer services and support in the entire development process of an integrated circuit, ranging from specification and design to qualification and supply of tested chips. Chain-IC wants to make system integration accessible for any electronic system, even when volumes are relatively low.

The architecture of CPU processors, which is powering computing already for decades, primarily scales with the number of transistors. It is a match made in heaven with traditional Moore’s Law scaling, where the figure of merit is the number of transistors per mm^2. However, in the new era of Moore’s Law, where Energy Efficient Performance (EEP) is the new figure of merit, new IC architectures emerge that outperform even if they are not based on highest resolution technology.

Below are a few examples of Dutch start-up companies with an innovative IC design, digital, low power and cost-effective solution despite not being manufactured on a leading-edge technology node.

–

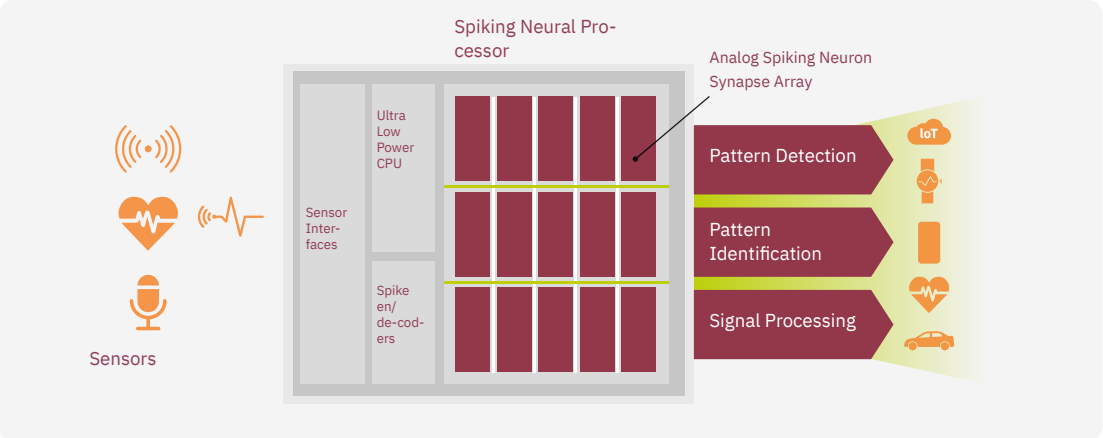
[98] https://qbaylogic.com/

[99] https://sencure.com/

[100] https://www.chain-ic.com/

13.8 Innatera or neuromorphic chips[101]

From smart thermostats in our homes to the latest wearables on our wrists, sensors have become ubiquitous, generating a staggering amount of complex data. Traditionally, this data has been sent to the cloud for processing. However, high latency and mounting costs are driving a shift in how sensor data is processed, namely, on the edge where it is collected. Innatera, a start-up from Delft, leverages spiking neural networks to efficiently and quickly process sensor data on the edge. Spiking neural networks mimic how the human brain works, using the timing of spikes to identify correlations, i.e., patterns, in data. The human brain represents and processes information using discrete events called spikes, brief surges in electrical activity when a neuron fires. If an event is important, a neuron spikes early, which is how correlations are formed, so the timing of spikes is important. The neuromorphic chips of Innatera are manufactured at 28nm node technology by TSMC.



The Spiking Neural Processor (SNP) family of ultra-low power processors enables high-performance pattern recognition at the sensor edge.[102]

–

[101] https://www.future-of-computing.com/innatera-shaping-the-future-of-neuromorphic-computing-for-the-sensor-edge/

[102] https://innatera.com/products

13. Chiplet design and architecture opportunities

13.9 Qualinx or Digital Reconfigurable RF chips

Qualinx, a start-up from Delft, develops the world's lowest power GNSS solutions, enabled by breakthrough Digital RF (DRF) and Software Defined Radio technology. Global Navigation Satellite System (GNSS) refers to a satellite constellation that provides global positioning, navigation, and timing services.

“Our technology will use as much as ten times less energy than currently available GNSS devices in the market. As a result of the improved efficiency, the battery life of, for example, fitness trackers and smartwatches can be extended from hours to several days.”, according to Tom Trill, CEO of Qualinx.

Qualinx builds on one of the Dutch key strengths – analog mixed-signal and RF – and shows how a unique and innovative IC design can lead to a competitive chipset in terms of price, performance and energy consumption even at lower technology nodes.

13.10 Axelera AI or Digital In-Memory Computing chips

Axelera AI from Eindhoven is another example of a Dutch IC design startup which develop a highly competitive chip at lower technology nodes. Axelera has designed an Edge AI chip offering best-in-class performance, efficiency, and ease of use for AI inferencing of computer vision workloads at the Edge. Its IC architecture is based on Digital In-memory Computing or D-IMC which is a radically different approach to data processing.
In D-IMC crossbar arrays of memory devices can be used to store a matrix and perform matrix-vector multiplications “in-place” without intermediate movement of data. Digital In-Memory Computing (D-IMC) technology is key to delivering high energy efficiency and outstanding performance. Based on SRAM (Static Random-Access Memory) combined with digital computations, each memory cell effectively becomes a compute element. This radically increases the number of operations per computer cycle (one multiplication and one accumulation per cycle per memory cell) without suffering from issues such as noise or lower accuracy.

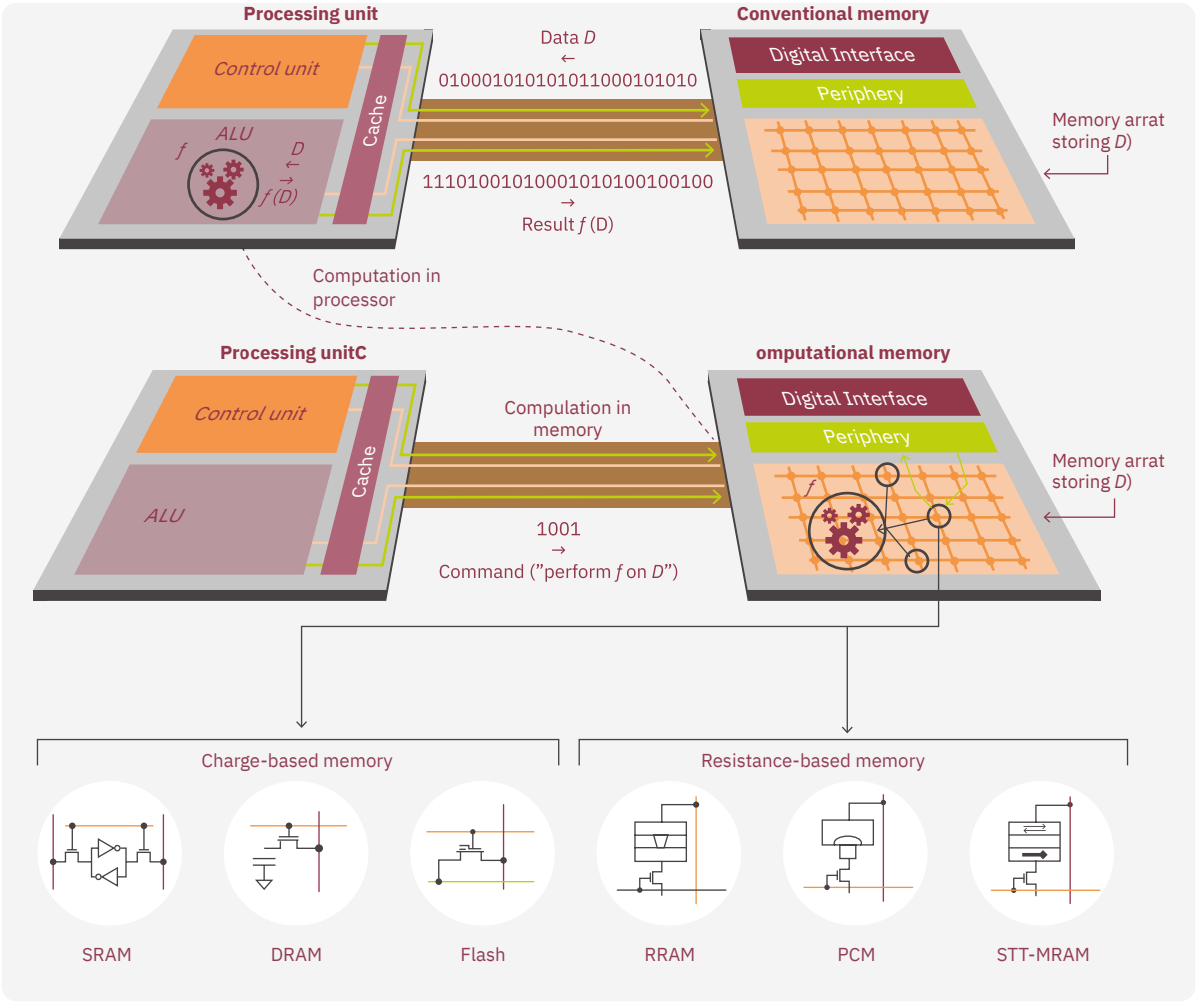
Axelera AI, a start-up from Eindhoven, aims to democratize AI at the edge by delivering a platform that combines multicore in-memory computing with a custom dataflow architecture. In the words of Fabrizio Del Maffeo, CEO and co-founder of Axelera AI: “We will give a powerful tool to entrepreneurs who want to develop a product that unleashes the power of AI in markets such as retail, smart cities, medicine, and Industry 4.0” . The company’s first product, Metis, delivers higher performance and efficiency (TOPS per $ and watt) than its competitors. It was designed for running complex computer vision applications on edge devices.

The chips of Axelera are designed in 12 nm CMOS.

[103] https://www.axelera.ai/technology#dimc

[104] https://www.imec-int.com/en/articles/start-story-axelera-ai

[105] https://vivatechnology.com/partners/axelera-ai



Conventional computing (a) and In-memory computing (b). Axelera uses SRAM to perform the matrix-vector multiplication.[106]

–

[106] https://www.nature.com/articles/s41565-020-0655-z

14

Conclusions and recommen-dations

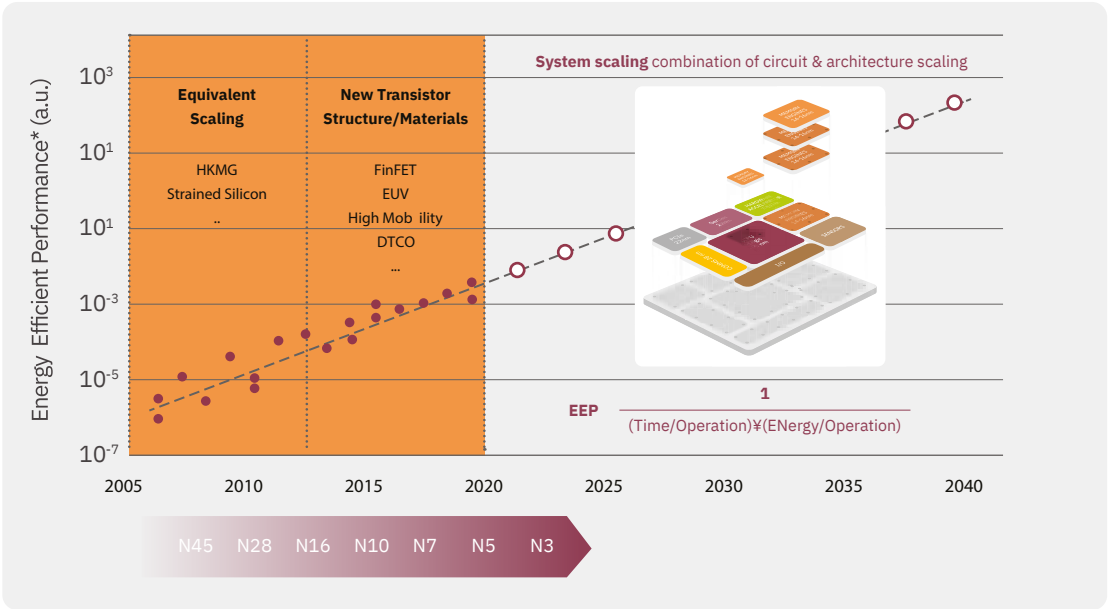
14. Conclusions and recommendations

“A new, more articulated and concerted approach is needed to boost the EU’s future competitiveness in the semicon sector.” That is one of the main conclusions of EU report “The future of European competitiveness”[107].

It is proposed to design the concerted approach around “chiplets”. A chiplet is a small, modular chip that performs a specific function very well. Multiple chiplets, for e.g. memory, logic and sensors, are integrated in a full functional chip. Stacked chiplets also have impact on power consumption by semiconductors. Data has to travel over a shorter distance compared to a homogeneous 2D-design. Therefore, less power is dissipated in the interconnect. This will be an important contribution to continue Moore’s Law with the metric of Energy Efficient Performance (EEP).

Heterogeneous integration of the chiplets require new metrology equipment to achieve high yield. This equipment can be developed by start-ups in the strong semicon equipment ecosystem of ASML, ASM and BESI.

IC design will change due to chiplets as new designs are made by combining various chiplets like advanced LEGO-bricks. This can be realized at lower design and manufacturing cost and shorter time to market. Opportunities will result for Dutch IC designer start-ups in proximity of mixed signal parties NXP and Nexperia. The IC designers create the new visionary applications that are the real drivers of Moore’s Law.



Use of chiplets will open opportunities in chip designs and metrology.[108]

–

[107] https://commission.europa.eu/document/download/ec1409c1-d4b4-4882-8bdd-3519f86bbb92_en?filename=The%20future%20of%20European%20competitiveness_%20In-depth%20analysis%20and%20recommendations_0.pdf

[108] https://www.techpowerup.com/327919/tsmc-cowos-capacity-doubles-for-two-years-still-insufficient-trendforce

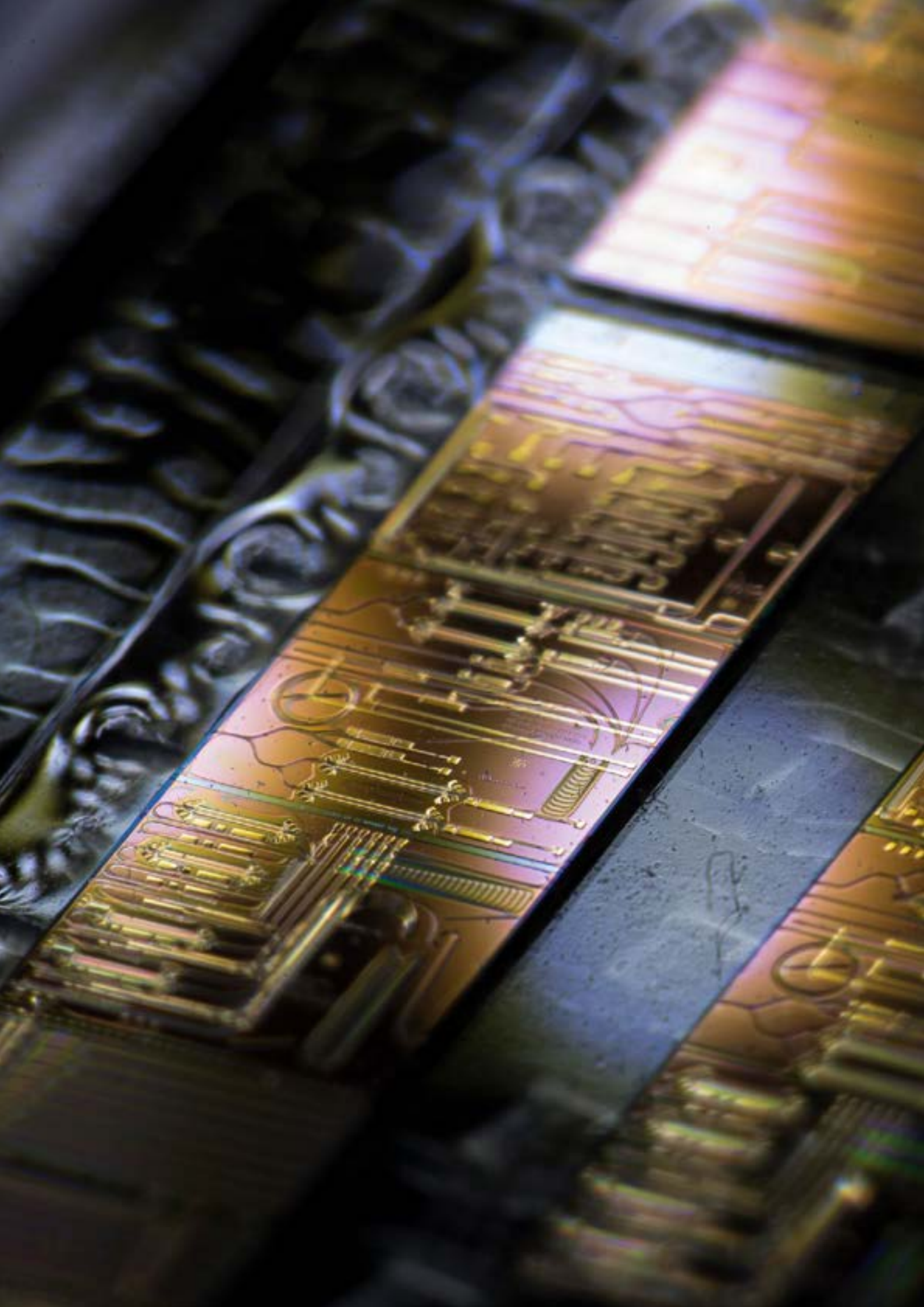
Chiplets can be a ‘blessing in disguise’ for Dutch and EU semicon industry. Let us face it: continuation of Moore’s Law is not an easy task for the worldwide (but mainly non-EU) semicon industry as huge technical barriers have to be become such as energy demand. Chiplets are crucial to overcome these barriers, and we are entering an era where EEP is enhanced not just by increasing the number of transistors per mm2, but by the ability to integrate advanced chiplets.

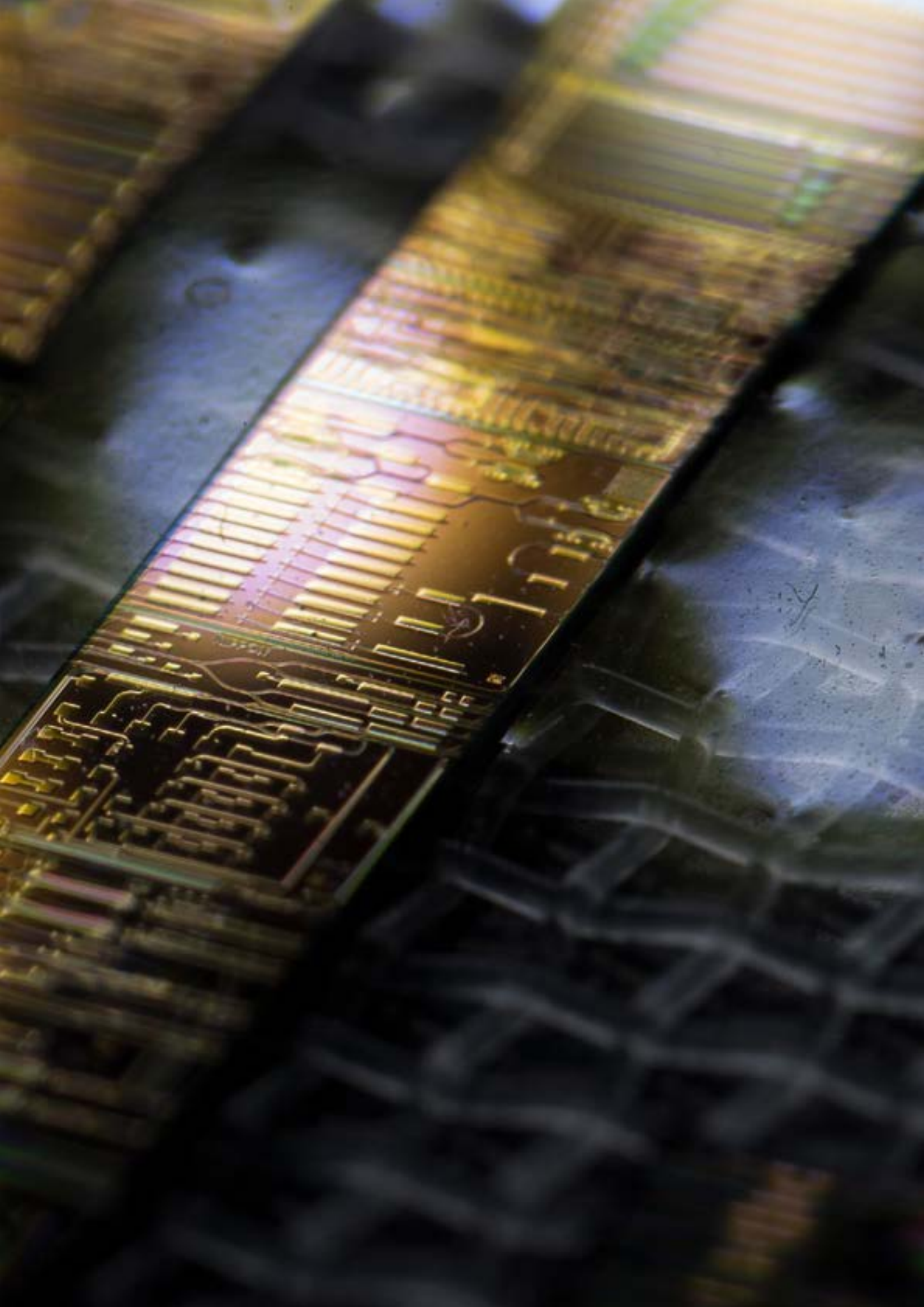
It all starts with solving a relevant problem, even if it has a specific application. And that has been the visionary approach of the semicon industry starting with the cartoon in Moore’s paper from 1965.

Visionary applications are the real drivers as Gordon Moore already concluded in his paper from 1965. [109]

–

[109] https://hasler.ece.gatech.edu/Published_papers/Technology_overview/gordon_moore_1965_article.pdf





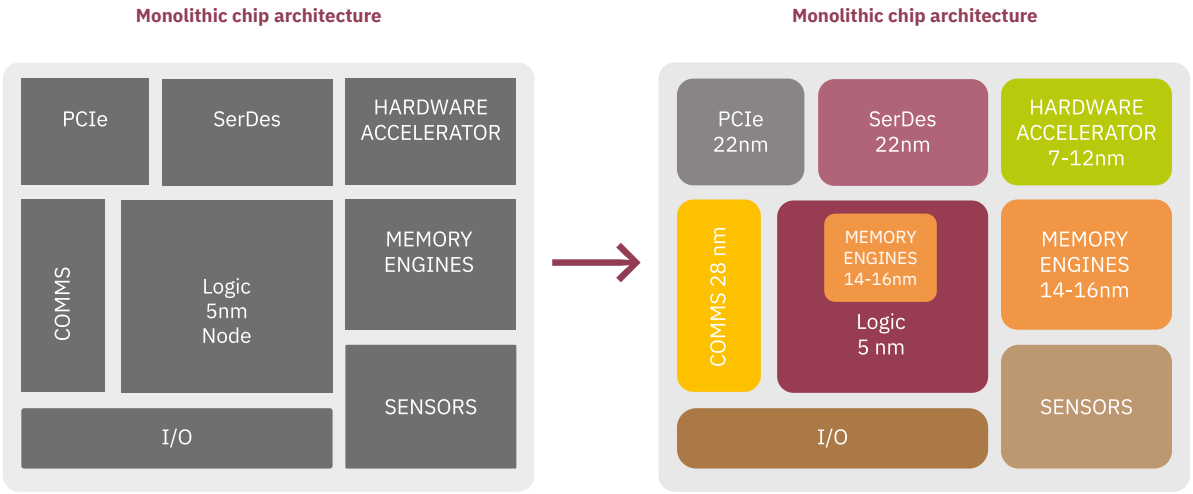
Annex. Cost of Ownership Monolithic versus Chiplet

In this section the prices are compared of a product made as a monolithic chip with the very same chip made chiplet based. For the numerical case a chip area of 300 mm2 is assumed which equals a state-of-art processor.

When such a processor is made monolithically, the entire area is processed at once so the most demanding section (i.e. CPU) drives which production node is used. With chiplets, this choice can be made individually per chiplet.

In the manufacturing process not all dies will perform to specifications as some show defects. It is known that the manufacturing yield of dies drops when they get larger, especially at high resolution. This is reflected in the cost of Known Good Dies (KGD) that are ready for bonding.

Bonding comes to a price per die that has to be bonded, and a percentage of the bonds made are not yielding. In the calculations below it is assumed that the entire package is discarded when one of the dies is not bonded properly. This is the worst case, assuming that no rework is possible.



Monolithic chip	Parameter	Chiplet-based
400 mm^2	Chip area	440 mm^2 10% extra for interconnects
N5: 400 mm^2	Chip area per node	N5: 105 mm^2 (24%) N7: 50 mm^2 (11%) N14: 60 mm^2 (14%) N22: 130 mm^2 (30%) N28: 95 mm^2 (21%) (% reflects fraction of total chip area)
$ 68.5	Bare die price total [$]	$ 38.5 43,7% lower price than monolithically
$ 100.9	Known Good Dies price total [$]	$42.4 58% lower price than monolithically
$ 106.2	Known Good Bonded Dies price total [$]	$ 63.9 40% lower price than monolithically
$ 109.2	Packaged chip price total [$]	$ 87.9 19.5% lower price than monolithically

Annex. Cost of Ownership Monolithic versus Chiplet

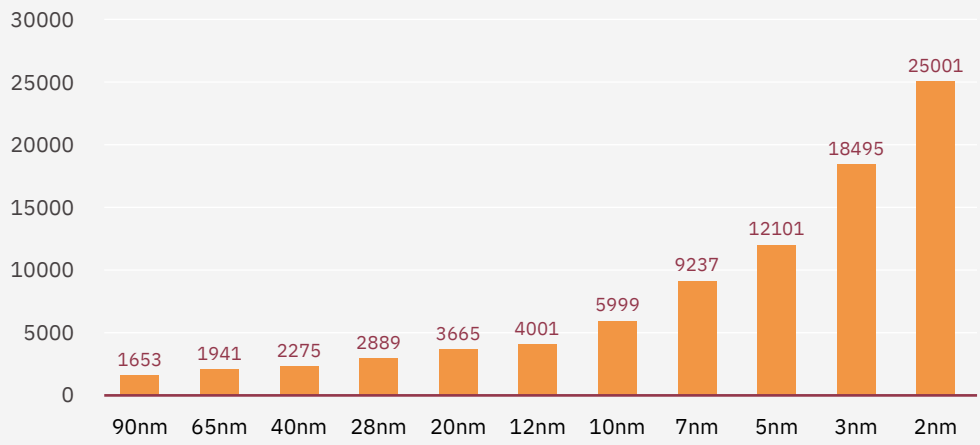
Wafer cost per node[110]

Wafer cost is based on 4 main factors:
- Wafer technology node, e.g. 5nm, 65nm, 130nm, etc.
- Wafer options/features, e.g. all the options required on top of the plain vanilla, e.g. mim cap, flash, high voltage, etc.
- Wafer volume, e.g. how many wafers will your produce, e.g. obviously, the more wafers the lower the price.
- Wafer yield is an important factor that often neglected but need to be considered from the start.

Foundries calculate wafer price based on the technology node complexity. Some of the factors related to wafer price are: fab upfront investment in building the facility, installing the right tools and instruments and operation cost. It's known that the labour cost of a fab is less than 10%.

Below please find a graph that shows the average wafer price of different technology nodes.

Estimated Wafer Price Per Node (USD)



Wafer cost[111]

Monolithic chip	Parameter	Chiplet-based
N5: $ 68.5	Bare die price per node [$]	N5: $ 18.0 N7: $ 6.5 N14: $ 3.4 N22: $ 6.7 N28: $ 3.9
$ 51.4	Bare die price total [$]	$ 38.5 43.7% lower price than monolithically

–

[110] https://anysilicon.com/silicon-wafer-cost/

[111] https://anysilicon.com/silicon-wafer-cost/

Annex. Cost of Ownership Monolithic versus Chiplet

Die yield[112]

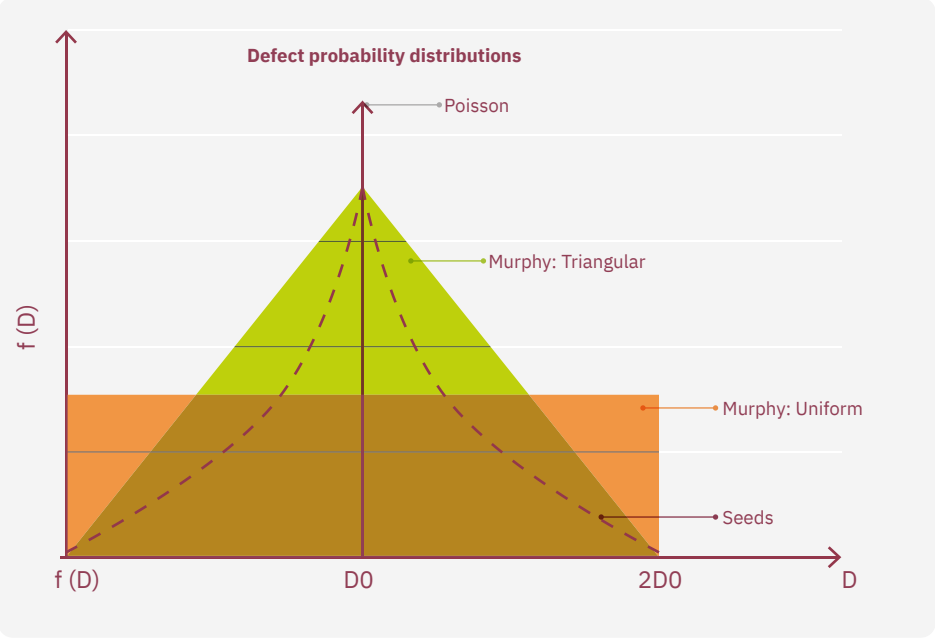
Semiconductor foundries, like most manufacturers, are unable to get every chip to work due to varying tolerances, faults and physical limitations in real-world contexts. A semiconductor foundry, such as Intel, Samsung or TSMC, invests a significant amount of money and effort to boost process yields, which are often targeted to be greater than 90%. Given the huge volume of chips manufactured in foundries, die yield is a critical profitability parameter that attracts much deserved attention.

Die yield loss is related to flaws in the wafer manufacturing process. A defect is any flaw in the physical structure of a die that causes the circuit to fail. Anomalies take many forms, including shorts and opens, particle contamination, spatters, flakes, pinholes, scratches, and so on.

For larger die sizes, the Murphy model with triangular distribution is most commonly used. The defect density (D) can be expressed as a distribution f(D), with a mean value of D0. Based on this B. T. Murphy of Bell Labs proposed die yield calculation as:

$$DY = \int_0^\infty e^{-DA} f(D) dD$$

The figure below shows a few forms that f(D) can take depending on how defects are distributed on the wafer.



How Foundries calculate die yield.[113]

–

[112] https://www.viksnewsletter.com/p/how-foundries-calculate-die-yield

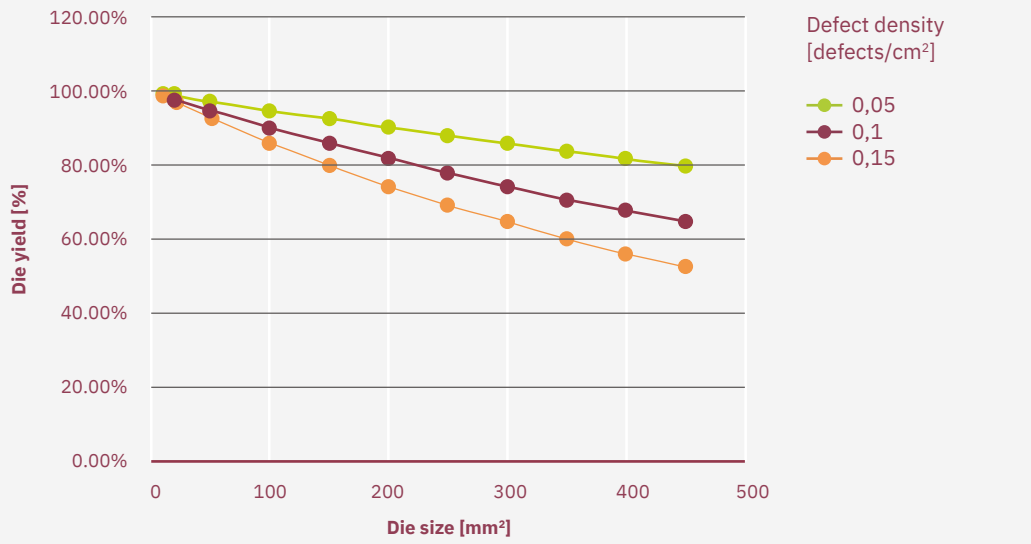
[113] https://www.viksnewsletter.com/p/how-foundries-calculate-die-yield

If f(D) approximates a Gaussian distribution with a triangular function, die yield after integration is given by

$$DY = \left[\frac{1 - e^{-D_0 A}}{D_0 A}\right]^2$$

The triangular distribution and resulting die yield is called the Murphy Model.

A defect density of D0 <0.1 defects/cm2 results in a die yield of more than 90% for a die of 1 cm2, which indicates a mature, manufacturable technology node.



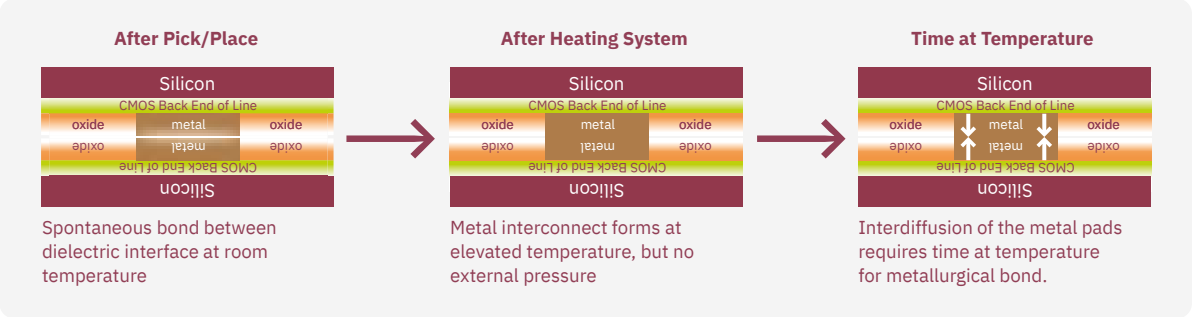
Monolithic chip	Parameter	Chiplet-based
N5: $ 100.9 [67.9%]	KGD price per node [$] [die yield]	N5: $20.0 [90.1%] N7: $ 6.9 [95.1%] N14: $ 3.6 [94.2%] N22: $ 7.7 [87.9%] N28: $ 4.3 [91.0%]
$ 100.9	Known Good Dies price total [$]	$ 42.4 58% lower price than monolithically

Annex. Cost of Ownership Monolithic versus Chiplet

Bonding yield
Bonding is used to connect the chiplets together. High performance can be achieved with hybrid bonding. Hybrid bonding, as a low-temperature direct copper bonding technology, in particular, will become widely adopted for a broad range of high-performance semiconductor devices in the years to come. It will enable the continued growth of economically viable power-efficient computing, for example, utilizing chiplet-based heterogeneous integrated packaging technologies.

Hybrid bonding technology is an all-solid-state joining technology that forms a spontaneous dielectric-to-dielectric bond at room temperature and then establishes a metal-to-metal connection (usually Cu-to-Cu bond) by a low-temperature batch annealing process (150 – 300°C), potentially enabling sub-micrometer bonding pitch.

The yield of the hybrid bonding process is key as rework is not possible. This means that an error in bonding will result in loss of at least two expensive dies or even more when the package contains more chiplets. The following table shows the bonding yield as function of the yield of an individual bond and the number of chiplets in the package as basis for our calculations



Hybrid bonding interconnect formation.[114]

–

[114] https://ewh.ieee.org/soc/cpmt/presentations/eps2202a.pdf

%

# Chiplets in package									
1	2	3	4	5	6	7	8	9	10
99%	98%	97%	96%	95%	94%	93%	92%	91%	90%
98%	96%	94%	92%	90%	89%	87%	85%	83%	82%
97%	94%	91%	89%	86%	83%	81%	78%	76%	74%
96%	92%	88%	85%	82%	78%	75%	72%	69%	66%
95%	90%	86%	81%	77%	74%	70%	66%	63%	60%
94%	88%	83%	78%	73%	69%	65%	61%	57%	54%
93%	86%	80%	75%	70%	65%	60%	56%	52%	48%
92%	85%	78%	72%	66%	61%	56%	51%	47%	43%
91%	83%	75%	69%	62%	57%	52%	47%	43%	39%
90%	81%	73%	66%	59%	53%	48%	43%	39%	35%

Monolithic chip	Parameter	Chiplet-based
1 95%	# dies in package Bonding yield of package at 95% single die bonding yield	8 66.3%
$ 106.2	Known Good Bonded Dies price total [$]	$ 63.9 40% lower price than monolithically

Bonding direct cost
Bonding comes with additional cost for equipment and processes. This is assumed to be $ 3 per bond. It can be enhanced by higher throughput and/or lower equipment cost.

Monolithic chip	Parameter	Chiplet-based
$ 109.2	Packaged chip price total [$]	$ 87.9 19.5% lower price than monolithically

Glossary

Abbreviations and terms used in this report are described below[115]

ADAS	Advanced driving assistance systems
ADC	Analog-to-digital converter
AI	Artificial Intelligence
ALD	Atomic Layer Deposition
APU	Application processor unit
ASIC	Application-specific integrated chip
BEOL	Back end of the line consists of depositing metal interconnect layers onto a wafer already patterned with devices.
BPDN	Backside power delivery network
CD	Critical Dimension
CFET	Complementary field-effect transistor (FET)
Chiplets	Chiplets are small, modular integrated circuits that can be combined to create a more complex system-on-chip (SoC) or multi-die design.
CIC	Commodity integrated circuits
CMOS	Complementary metal oxide semiconductor
CMP	Chemical Mechanical Polishing
CoWoS	TSMC chip-on-wafer-on-substrate
CPU	Central Processing Unit
Cu damascene	Copper interconnects are made with Damascene process: pattern is made in di-electric material, overfilled with deposited copper and CMP is used to remove excess copper.

[115] Descriptions from https://www.wikipedia.org/ and https://en.wikichip.org/wiki

CVD	Chemical Vapour Deposition
DAC	Digital-to-analog converter
Dennard scaling	Dennard scaling refers to the process of reducing the supply voltage of more advanced technologies in order to operate more transistors with the same power as the older technology
Diodes	A diode is a two-terminal electronic component that conducts current primarily in one direction
DRAM	Type of volatile memory chip
DUV	Deep UltraViolet Lithography with 248 nm and 193 nm light
D2W bonding	Die to wafer bonding
EDA	Electronic Design Automation
EEP	Energy Efficienct Performance
EMIB	Intel's Embedded Multi-Die Interconnect Bridge
EUV	Extreme UltraViolet Lithography with 13.5 nm light
FEOL	Front end of line (FEOL) is the first portion of IC f abrication where the individual components (transistors, capacitors, resistors, etc.) are patterned in a semiconductor substrate.
FinFET	Type of field-effect transistor (FET) that has a thin vertical fin
FLOPS	Floating point operations per second
Foundy	Company that makes chips for others in their “fabs”
Foveros	Intel name of chiplet integration
FPGA	Field programmable gate array
GAA	Type of transistor: Gate All Around
Gbps	Giga bits per second

Glossary

GPU	Graphics Processing Unit
HBM	High-bandwidth memory
HPC	High-performance computing
Huang's Law	An observation in computer science and engineering that advancements in graphics processing units (GPUs) are growing at a rate much faster than with traditional central processing units (CPUs)
Hybrid bonding	Hybrid Bonding Interconnect (HB or HBI) or Direct Bond Interconnect (DBI) is a high-performance high-density vertical die-to-die interconnect technology used to transmit signal and power between multiple stacked dies through the direct attachment of two homogeneous surfaces forming strong covalent bonds.
IC	Integrated Circuit
IDM	Integrated Device Manufacturers
IMC	In-memory computing
IoT	Internet of Things
IP	Intellectual Property
IRDS	International Roadmap for Devices and Systems
KGD	Known Good Die
MEMS	Micro-electromechanical systems
MOSFET	Metal–oxide–semiconductor field-effect transistor (MOSFET, MOS-FET, MOS FET, or MOS transistor) is a type of field-effect transistor (FET), most commonly fabricated by the controlled oxidation of silicon.
MZI	The Mach–Zehnder interferometer is an optical device used to determine the relative phase shift variations between two collimated beams derived by splitting light from a single source.
NA	Numerical Aperture: quality of the optics system in lithography

NAND	Type of non-volatile memory chip
Nanosheet FET	A nanosheet is a two-dimensional nanostructure with thickness in a scale ranging from 1 to 100 nm.
Nm	Nanometre is a unit of length in the International System of Units (SI), equal to one billionth or one thousand million of a meter (0.000000001 m).
NPU	Neural processing unit
OEM	Original Equipment Manufacturers
OSAT	Outsourced Semiconductor Assembly and Test
PC	Personal Computer
PCB	Printed circuit board
PPAC	Performance, Power, Area and Cost
PVD	Physical Vapour Deposition
Qubit	A qubit is a two-state (or two-level) quantum-mechanical system, one of the simplest quantum systems displaying the peculiarity of quantum mechanics.
RDL	Re-distribution layer
RF	Radio frequency chips
Scatterometry	A scatterometer is a scientific instrument to measure the return of a beam of light or radar waves scattered by diffusion in a medium.
SOC	System on Chip
SoIC	TSMC System On Integrated Chips
SRAM	Type of non-volatile memory chip
Strained silicon	Strained silicon is a layer of silicon in which the silicon atoms are stretched beyond their normal interatomic distance.

Glossary

Technology node	The technology node refers to a specific semiconductor manufacturing process and its design rules. Different nodes often imply different circuit generations and architectures. Generally, the smaller the technology node means the smaller the feature size, producing smaller transistors which are both faster and more power efficient.
TIM	Thermal interface material
TOPS	Trillions or Tera Operations per Second
TSV	Through Silicon Via
W2W bonding	Wafer to wafer bonding
WFE	Wafer Fab Equipment

Patrick de Jager
BIRDS Creative Innovation
patrick.de.jager@birdsci.nl

De Jager has worked for many years in business development and research in the semiconductor industry. Last year he started BIRDS Creative Innovation to nurture innovative solutions in semicon and energy transition.

He started his carrier in 1997 at TNO Applied Physics Department working for ASML. In these days ASML was challenged with a potential end of Moore's Law, as traditional optical lithography was close to its diffraction limits. About 100 initiatives started worldwide to extend Moore's Law by means of novel lithography concepts based on electrons, ions, X-ray and EUV. A project started with the aim to "narrow down the options" as the semicon industry realized that the cost of developing multiple solutions was a new roadblock. De Jager was involved in this activity by investigating the role of e.g. shot noise and Coulomb Interaction in electron beam lithography. This contributed to the choice of ASML and the semicon industry for EUV lithography, which is the basis of ASML's success today.

Later on he started the BlueBird initiative of TNO to connect Dutch semicon companies around stacking chips by means of Through Silicon Via's. This vertical integration is now reaching High Volume Manufacturing as it is crucial for Artificial Intelligence processors.

In 2010 de Jager moved to ASML's Strategic Business Development where he investigated many threats and opportunities for the future of ASML and the semicon industry. In the period that EUV source power was a real challenge, he was involved in the development of a Free Electron Laser (FEL) as high-power source for EUV light. That resulted in the idea of de Jager to use FEL technology for clean production of medical radio-isotopes. De Jager and ASML received the status of National Icon of Innovation from the Dutch government for this concept called LightHouse.

As hydrogen is used to mitigate contamination in an EUV tool, de Jager also investigated innovation in the production of green hydrogen by means of electrolysis. It inspired him to create value in sectors such as the energy transition by addressing challenges in these sectors in the same way as in semicon.

Copyright & disclaimer

This research report is intended for informational purposes only.

Although the report has been prepared with the utmost care and accuracy, the use of the information in this report is at your own risk. We accept no liability for any damage, loss, or inconvenience that may result from the use of this information.

Invest-NL Deep Tech Fund
www.invest-nl.nl
Kingsfordweg 43-117
1043 GP Amsterdam