

European Core Technologies for future connectivity systems and components

Call/Topic: H2020 ICT-42-2020

Grant Agreement Number: 956830



Deliverable

D3.3 Initial COREnect industry roadmap

Deliverable type:	Report
WP number and title:	WP3 (Roadmap and recommendations)
Dissemination level:	Public
Due date:	30.06.2021
Document editor:	Björn Debaillie (IMEC)
Section editors:	Björn Debaillie (IMEC), Didier BELOT (CEA), Frederic Giancesello (ST), Gerhard Fettweis (TUD), Jochen Koszescha (IFAG), Marina Plietsch (IFAG), Patrick Cogez (AENEAS), Patrick Pype (NXP), Piet Wambacq (IMEC), Viktor Razilov (TUD)
Contributing partners:	TUD, 5G IA, AENEAS, BOSCH, CEA, EAB, IFAG, IIIV, IMEC, NXP, ST
Internal reviewers:	Mohand Achouche (IIIV/Nokia), Matthias Illing (BOSCH)

Document History

Version	Date	Author/Editor	Description
0.1	19.04.2021	IMEC	Outline
0.2	12.05.2021	All involved partners	Initial contributions
0.3	20.05.2021	All involved partners with support of experts	Advanced contributions
0.4	27.05.2021	All involved partners with support of experts	Completing most sections
0.5	03.06.2021	All involved partners with support of experts	Completing most sections
0.6	10.06.2021	All involved partners with support of experts	Feedback being processed
0.7	17.06.2021	All involved partners and internal reviewers	Partial internal review
0.8	29.06.2021	All involved partners and internal reviewers	Incorporation internal review feedback
0.9	05.07.2021	IMEC	Final consolidation
1.0	05.07.2021	IMEC	Final version (v1.0)

List of Abbreviations

Abbreviation	Denotation	Abbreviation	Denotation
4G	4th generation mobile communication	LoS	Line of sight
5G	5th generation mobile communication	LPWAN	Low power wireless access network
6G	6th generation mobile communication	MANET	Mobile ad hoc networks
ADAS	Advanced Driver Assistance Systems	MCM	Multi-chip modules
ADC	Analog to digital converter	MCU	Microcontroller unit
AE	Antenna elements	MEMS	Micro-electro-mechanical systems
AI	Artificial Intelligence	MIIT	Ministry of industry and information technology (China)
API	Application programming interface	MIMO	Multiple input multiple output
ARM	Advanced RISC machine	ML	Machine learning
ASIC	Application-specific integrated circuit	MLC	Multi-level cell
BAW	Bulk acoustic wave	mmW	Millimeter-wave
BiCMOS	Bipolar CMOS	MPSOC	Multiprocessor system on a chip
CAPEX	Capital expenditures	MRAM	Magneto resistive random-access memory
CAPI	Common ISDN application programming interface	MUX	Multiplexer
CCIX	Cache coherent interconnect	NB-IoT	Narrowband IoT
CMOS	Complementary metal oxide semiconductor	NIC	Network interface card
COVID-19	Coronavirus disease 2019	NVM	Non-volatile memory
CPO	Co-Packaged Optics	O/E	Optical-to-electronics
CPU	Central processing unit	OS	Operating system
CU	Central unit	OSI	Open Systems Interconnection
CXL	Compute express link	OXRAM	Metal oxide resistive memory
DAC	Digital to analog converter	PA	Power amplifier
DBT	Dynamic binary translation	PC	Personal computer
DeMUX	Demultiplexer	PCB	Printed circuit board
DPLL	Digital phase locked loop	PCM	Phase-change memory
DPU	Data processing unit	PCP	Programmable computing platforms
DSL	Domain-specific language	PhD	Doctor of philosophy
DSP	Digital signal processing	PoA	Power over Air
DU	Distributed unit	QoE	Quality of experience
E/O	Electronics-to-optical	QoS	Quality of service
EDA	Electronic design automation	R&I	Research and innovation
eDRAM	Embedded dynamic random-access memory	RADCOM	Radar and Communication
EG	Expert group	RAN	Radio access network

EIRP	Effective isotopically radiated power	RF	Radio frequency
eNVM	Embedded non-volatile memory	RFID	Radio-frequency identification
FBAR	Film bulk acoustic resonator	RISC	Reduced instruction set computer
FDSOI	Fully depleted silicon on insulator	RTO	Research and Technology Organization
FeFET	Ferroelectric field-effect transistor	RU	Radio unit
FEM	Front-end module	SAW	Surface acoustic wave
FinFET	Fin field-effect transistor	SiGe	Silicon germanium
FPGA	Field programmable gate array	SiP	Systems-in-package
FW	Firmware	SME	Small and medium-sized enterprise
GaN	Gallium nitride	SoC	System-on-chips
GPU	Graphical processing unit	SoC	Systems on chip
HARQ	Hybrid automatic repeat request	SOI	Silicon on insulator
h-BN	Hexagonal boron nitride	SRAM	Static random-access memory
HBT	Heterojunction bipolar transistor	STDP	Spike timing dependent plasticity
HEMT	High-electron-mobility transistor	SW	Software
HMI	Human-machine interface	TI	Tactile Internet
HPC	High performance computing	ToF	Time of Flight
HW	Hardware	TSN	Time-sensitive networking
I/O	Input /output	TSV	Through Silicon Via
IC	Integrated circuit	TTD	True-time-delays
IDM	Integrated device manufacturer	UE	User Equipment
IIoT	Industrial internet of things	ULP	Ultra-low power
InP	Indium phosphide	URLLC	Ultra-reliable low latency communications
IoT	Internet of things	USA	United States of America
IP	Intellectual property	UV	Ultraviolet lithography
ISA	Instruction set architecture	UWB	Ultra-wideband
JEDEC	Joint electron device engineering council	V2X	Vehicle to everything
LDMOS	Planar double diffused MOSFET	VLSI	Very large-scale integration
LNA	Low noise amplifier		

Disclaimer

This document has been prepared by the COREnect project. It reflects the collected views of all experts involved. It does not reflect the opinion of any single COREnect partner or any of the organizations with which the experts are affiliated. The COREnect project and its consortium partners are not liable for any consequence stemming from the reuse of this publication.

Table of contents

1.	Introduction	5
2.	Literature overview and highlights	6
3.	Strategic roadmap: challenges and approach.....	9
3.1.	Context and objectives	9
3.2.	Major challenges and opportunities related to future European connectivity systems and components	10
3.3.	Proposed Industry Roadmap Strategy: How should Europe address future connectivity technologies with a value-chain approach?	16
4.	First draft of strategic industry-relevant R&I roadmaps	17
4.1.	EG1: compute and store	20
4.2.	EG2: connect and communicate.....	31
4.3.	EG3: sense and power	43
4.4.	Common strategic actions across the expert groups	58
5.	Conclusion.....	60
6.	References	61
7.	Appendix.....	63

1. Introduction

One of the key COREnect objectives is to define a strategic R&I roadmap for future European connectivity systems and components, supporting European's strategic autonomy and sovereignty objectives. This roadmap is being defined based on input from all relevant stakeholders across different domains and communities (including SNS and KDT), covering the relevant actors from industry, research, academia, associations, policy analysis etc.

An end-to-end view of future connectivity systems is depicted in Figure 1. To achieve the upcoming needs from emerging applications in e.g., industrial automation, such connectivity systems need offer extreme high capacity, extreme coverage, extreme low latency and high reliability, all at low energy and cost.

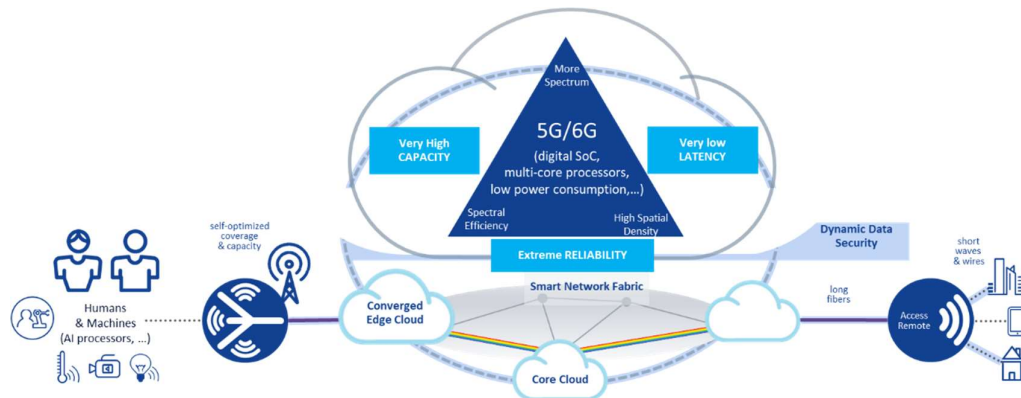


Figure 1: COREnect End-to-end System View.

The envisioned roadmap is created over three phases, as illustrated in Figure 2. During the first phase, the roadmap contours are defined, and potential roadmap directions are initially drafted. These roadmap directions are further explored and detailed during the second phase by incremental incorporation of recommendations and feedback. During the third and final phase, the roadmap definition is consolidated. The current deliverable (D3.3) reports the outcome of the first phase and describes the **initial** COREnect roadmap proposition. The second phase will be concluded with the delivery of D3.4, which will present the **intermediate** roadmap proposition. The **final** roadmap proposition will be described in D3.6.



Figure 2: COREnect roadmap building timeline.

The text below describes a roadmap with items for the short term, mid-term and long term. Basically, the short and mid-term roadmap banks on growth of existing strengths, while for the long-term items such as establishments of < 2 nm CMOS manufacturing are considered.

The outline of this document is as follows. Section 2 gives an overview of the key publications which act as a reference and baseline for the roadmap activities within COREnect. Section 3 gives an initial glance of the key challenges and possible approaches to define the envisioned strategic roadmap. Section 4 introduces the implemented structure of expert groups and describes the

initial roadmap ideas and potential directions of each expert group. This section is then concluded by the description of possible strategic actions across the different expert groups.

2. Literature overview and highlights

Europe is not the only region of the world where roadmaps for electronics components and systems for future networks are derived. Even in Europe, while COREnect is specifically focusing on developing a roadmap of core technologies for future connectivity, relevant roadmapping information can be found in documents issued by other groups, addressing a wider or different but related application scope. Table 1 gives a selected list of those documents.

Table 1: Other relevant roadmapping initiatives.

Document name	Related section	Date
ECS Strategic Research and Innovation Agenda ¹	Section 2.2 connectivity	01/2021
International Roadmap for Devices and Systems ²	Focus team "Outside System Connectivity"	2020
5G ACIA white paper ³		02/2019
Decadal Plan for Semiconductors - SRC ⁴	Chapter 3: New Trajectories for Communication	01/2021
Webinar "New Apps and New Possibilities: How 5G Will Dramatically Change the Semiconductor Industry" ⁵		07/2020
"Smart networks in the context of NGI" from NetworldEurope ETP ⁶		01/2021
NTT DOCOMO White Paper: "5G Evolution and 6G" ⁷		02/2021

In addition to these roadmapping activities, the COREnect consortium could evidence the intensity of the current debate regarding strategic digital autonomy in various regions of the world, translating into legislative actions and/or investment decisions. Below is a brief overview across several countries.

Legislative activities in the USA

Responding to concerns, the USA is increasingly reliant on imported microelectronics, Congress enacted the bipartisan *Creating Helpful Incentives for Producing Semiconductors (chips) for America Act*⁸ in December as part of the National Defence Authorization Act, its annual defence policy update. The legislation authorizes an array of R&D initiatives as well as a subsidy program for domestic semiconductor manufacturers. However, while early versions of the CHIPS for America Act envisioned spending more than ten billion dollars over five years, the enacted version makes no specific funding recommendations for either the overall initiative or its component elements. In any case, actual funding for them will have to be provided through separate spending legislation.

¹ <https://aeneas-office.org/pdf/sria-2021/>

² <https://irds.ieee.org/editions/2020>

³ https://5g-acia.org/wp-content/uploads/2021/04/WP_5G_for_Connected_Industries_and_Automation_Download_19.03.19.pdf

⁴ www.src.org/about/decadal-plan/

⁵ <https://bit.ly/3iXZSyP>

⁶ <https://bscw.5g-ppp.eu/pub/bscw.cgi/d392313/Annex%20v2.3%20-%20Public.pdf>

⁷ www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_v3.0.pdf

⁸ www.aip.org/sites/default/files/aipcorp/images/fvi/pdf/chips-for-america-act-final.pdf

On February 24, 2021, the President signed E.O. 14017, directing a whole-of-government approach to assessing vulnerabilities in, and strengthening the resilience of, critical supply chains. This resulted in the release, on June 8, 2021, of findings from this comprehensive 100-day supply chain assessments for four critical products: semiconductor manufacturing and advanced packaging; large capacity batteries; critical minerals and materials; and pharmaceuticals and active pharmaceutical ingredients. Of specific interest for the COREnect project are the key findings regarding semiconductor manufacturing and advanced packaging, namely:

- Promote investment, transparency, and collaboration in partnership with industry, to address the current shortage,
- Fully fund the chips for America provisions to promote long-term US leadership,
- Strengthen the domestic semiconductor manufacturing ecosystem,
- Support SMEs and disadvantaged firms along the supply chain to enhance innovation,
- Build a talent pipeline,
- Work with allies and partners to build resilience,
- Protect the US technological advantage.

In particular, the report states that *“as an initial step, Congress should fund the chips provisions with at least \$50 billion in funding”*. Details can be found in the document entitled *100-day supply-chain review report*⁹.

To decrease their dependency on foreign-based semiconductor production, the USA is also attracting foreign investment. A case in point is the Taiwanese company TSMC which broke ground in June 2021 on its \$12 billion semiconductor fab in Arizona¹⁰. Likewise, Samsung Foundry has filed documents with authorities in Arizona, New York, and Texas seeking to build a leading-edge semiconductor manufacturing facility in the USA. The potential fab near Austin, Texas, is expected to cost over \$17 billion and to create 1,800 jobs¹¹.

Finally, the USA is actively using export control laws to prevent mainland China from developing sub-10 nm node technology, for example by barring ASML to sell Extreme UV equipment to Chinese semiconductor manufacturers, and even trying to extend that ban to older, Deep UV equipment¹².

Mainland China

In 2015, mainland China released its “Made in China 2025” initiative, which included the ambitious goal of reaching 70 per cent self-sufficiency goal for semiconductor production. However, they are so far falling very short of being on a trajectory meeting that target: IC production in China, including output by both foreign and domestic players, only accounted for 15.7% of its \$125 billion chip market in 2019. If only companies with headquarters in China are considered, their production accounted for just 6.1% of China’s total IC market that year. At its current pace China will only achieve one third of its goal.

⁹ <https://www.whitehouse.gov/wp-content/uploads/2021/06/100-day-supply-chain-review-report.pdf>

¹⁰ <https://www.datacenterdynamics.com/en/news/tsmc-starts-work-on-12bn-arizona-semiconductor-fab-gets-funding-for-japanese-chip-rd/>

¹¹ <https://www.anandtech.com/show/16483/samsung-in-the-usa-a-17-billion-usd-fab-by-late-2023>

¹² <https://www.techzine.eu/news/infrastructure/56766/usa-tries-to-prevent-all-export-of-asml-machines-to-china/>

Nevertheless, mainland China main semiconductor manufacturer, SMIC, has been increasing its R&D expenditures rather rapidly in recent years. In 2014, the company spent \$189.7 million, or 9.5% of revenue, on research and development. Five years later, in 2019, the company spent \$629 million, or 20.7% of revenue on R&D. In parallel, its CAPEX was expected to reach \$4.3 billion in 2020¹³. On March 2021, SMIC announced 2020 sales of \$3.9 billion and gross profit of \$0.9 billion, while the debt-to-equity ratio “remained low”¹⁴. A large CAPEX can only be achieved via capital injection from the shareholders. Since by late 2018 the Chinese government controlled at least 46.36% of the company, this corresponds to a significant amount of public support.

Japan

Similar to the USA, Japan is also discussing with TSMC towards the building of a chip fab in Kumamoto, which would be TSMC first Japanese semiconductor factory¹⁵.

South Korea

Finally, the South Korean government announced on May 13th, 2021, a plan by companies to invest 510 trillion won (\$451 billion) throughout 2030 and beefed-up tax benefits to boost chipmakers' competitiveness amid a critical global shortage of the key components. *"Our government will unite with companies to form a semiconductor powerhouse. We will support companies concretely."* said President Moon Jae-in¹⁶.

As part of the effort, the Finance Ministry said it will raise the tax deduction ratio for semiconductor research and development investments by big companies to 40% from the current 30%, paving the way for Samsung and SK Hynix to benefit from the eased financial burden. The chipmakers also will enjoy higher deductions for investments in facilities, as the government is doubling that ratio to 6%, the ministry said.

¹³ <https://www.eetimes.com/smic-advanced-process-technologies-and-govt-funding-part-2>

¹⁴ https://www.smics.com/en/site/news_read/7809

¹⁵ <https://www.datacenterdynamics.com/en/news/tsmc-considers-chip-fab-in-kumamoto-its-first-japanese-semiconductor-factory/>

¹⁶ <https://asia.nikkei.com/Business/Tech/Semiconductors/South-Korea-plans-to-invest-450bn-to-become-chip-powerhouse>

3. Strategic roadmap: challenges and approach

3.1. Context and objectives

Considering the full value chain, COREnect's main objective is to:

Develop a high-level strategic roadmap of core technologies for future connectivity systems and components, targeting the next generation European telecommunications networks and services (5G and beyond).

This roadmap proposal is being developed while considering the following factors:

- **The end goal of this industrial strategy is to support Europe's twin transition towards a green and digital future:** enabling our society to embrace digitalization in a sustainable way (from cost and power efficiency point of view) is a key societal challenge.
- **The necessity to strengthen Europe's strategy to differentiate and lead in its most important value chains while enabling the European ecosystem to adapt to a profound on-going value chain transformation:** it is mandatory to continue to innovate on "more than Moore" technologies to maintain Europe's leading role on key verticals. Today, advanced connectivity solution does not necessarily need chips processed in ≤ 5 nm node, but they will do so in the future. Moreover, the European ecosystem must also adapt to the value chain transformation induced by digitalization which requires to develop the relevant skills and technologies.
- **Aim for a realistic strategy keeping in mind that available economical resources are limited:** connectivity systems require a broad range of technologies which cannot be completely mastered by a single geographic area. This implies to clearly define key priorities in agreement with existing strengths.
- **The societal impact of the COVID-19 pandemic:** the current pandemic has underlined the importance of connectivity infrastructure in the resilience of our society.
- **The COVID-19 pandemic impact on key technologies supply chain and importance to secure Europe's sovereignty** (which does not mean autarky but allows for own decisions in Europe): the current supply chain issues underline the importance of cooperation with like-minded partners to support open, fair and rules-based trade to reduce strategic dependencies.
- **The geopolitical trade tensions between the USA and China:** the export regulations on USA technologies have contributed to disruptions in the current supply chains. Moreover, Europe's current dependency on USA technologies hampers Europe to act independently and to take strong standpoints on business-related sovereignty topics.
- **The objective of the European Commission to enable Europe to produce 20% of the world's semiconductors by 2030 to meet future industry demand:** to secure European sovereignty, manufacturing of semiconductors in Europe should reflect the relative size of its domestic market and the strength of its industrial players on key verticals.
- **The necessity to reduce Europe's dependence on USA for EDA solutions, software and IP required to develop future connectivity systems:** the increased trade tensions between the USA and China highlighted Europe's vulnerability to USA technology; it indicated alarming sovereignty issues on connectivity technologies if extraterritorial rules are applied to European semiconductor players.
- **The necessity to bridge the current gap on advanced processor design to secure Europe's industrial strategy and digital sovereignty:** advanced computing is key to support the digitalization of our society. Today, Europe heavily relies on USA advanced computing technologies. This causes a critical dependency and might impact Europe's key ambition for open, ethical, trustworthy, and explainable AI and computing.

The European Commission's objective to enable a fabrication plant to produce leading edge technology (2 nm or even beyond) through selected partnerships to ensure security of supply, in the next 10 to 15 years: due to the importance of leading-edge semiconductor technologies to manufacture advanced computing chips, Europe has a role to play to enable a more diverse and consequently resilient supply chain to reduce its critical dependencies, while remaining open. It is quite likely that the advent of 6G will fire up the semiconductor market and its entire ecosystem. This is a strategic consideration but any decision on installing advanced CMOS manufacturing in Europe will need a detailed predictive market study.

Keeping in mind the previous context and objectives, we first propose a synthesis of the current position of Europe on the connectivity market to identify opportunities and gaps. We will then propose a global industry roadmap strategy which will be finally derived in concrete actions proposed by COREnect's expert groups.

3.2. Major challenges and opportunities related to future European connectivity systems and components

To identify Europe's major challenges or opportunities concerning connectivity technology, we start the discussion by a review of Europe's position in the overall value chain. As illustrated in Figure 3, Europe still holds a good share in materials and tools to produce electronic components. Europe's production share is, however, lower at levels such as electronic equipment, electronic boards, and electronic components.

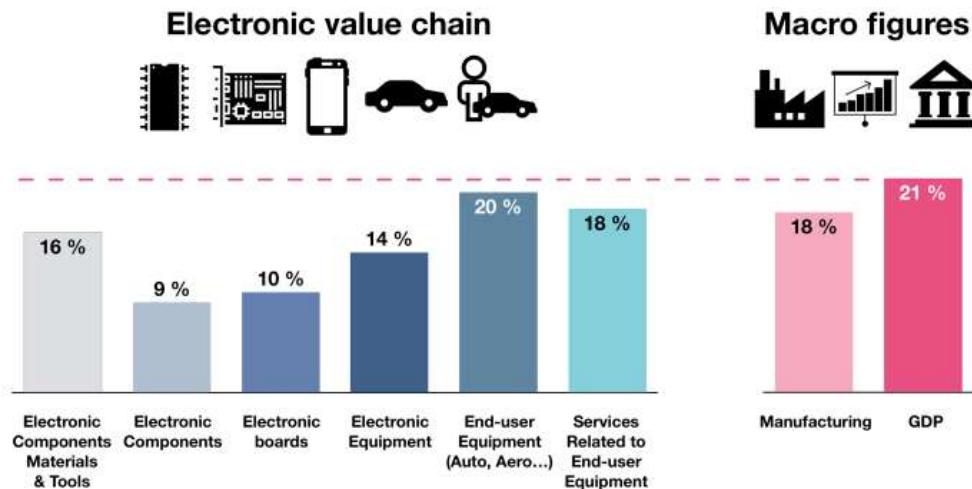


Figure 3: European share of the world production of the global electronic value chain [Dec20].

In Europe, the leading end-user segments are industrial electronics, aerospace defense and security, and automotive electronics. In the global electronics ecosystem, the leading segments are still the consumer mass markets (mobile phones, PCs). Consequently, Europe's share in the world production is also highest in those segments where Europe is strongest, as illustrated in Figure 4.

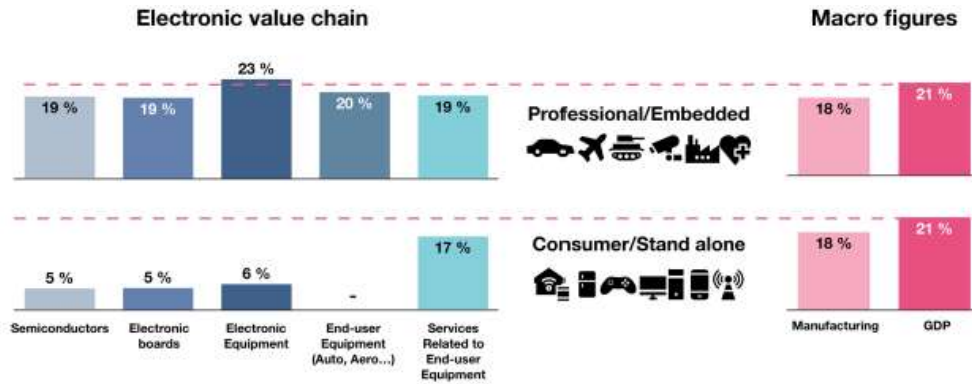


Figure 4: European share of the world production of the global focusing on professional/embedded and consumer electronic value chains [Dec20].

While Europe is just producing 9% of the overall electronic components (see Figure 3), its market share is 19% on the market it serves today (professional and embedded segments, the wireless infrastructure market being a good example with Ericsson and Nokia). This figure is in line with Europe's GDP. Since Europe hardly addresses the consumer market, the European ecosystem requires a moderated manufacturing capacity mainly focused on mature or derivate technology. For example: automotive represents today only about 10%, but this is expected to increase in the coming years. The installed European semiconductor manufacturing capability to address Europe's key verticals is sized accordingly. As illustrated in Figure 5, Europe has a strong presence on 200 mm facilities (with STMicroelectronics and Infineon among the top 5 leaders) which is in line with the technologies required by the European ecosystem and value chain.

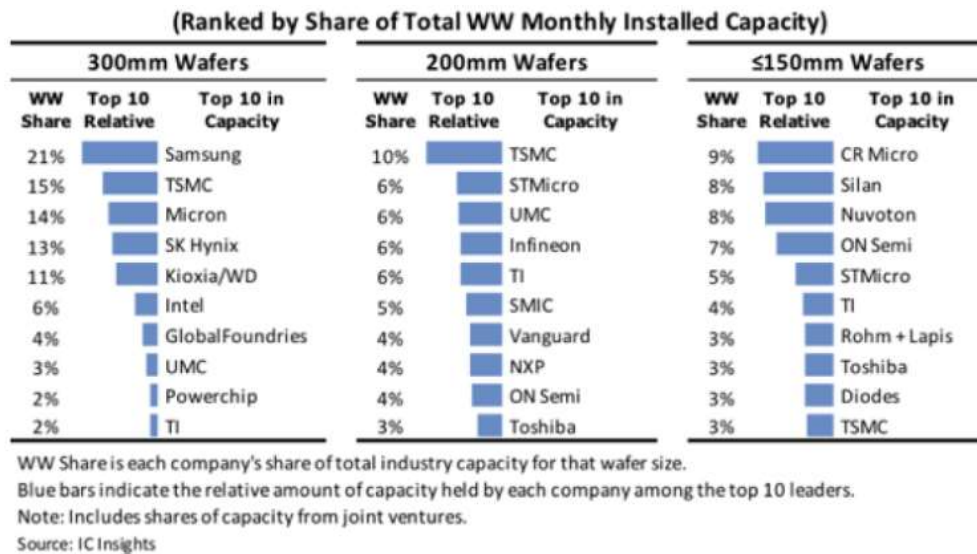


Figure 5: Installed capacity leaders in December 2020 by wafer size [EMS21].

The situation on 300 mm wafers manufacturing is completely different. On 300 mm, there is no European actor among the top 10 players. This is directly correlated with the European position on the market since the top 300 mm manufacturing players are addressing either memory (Samsung, Micron, SK Hynix, Kioxia) or advanced logic (Samsung, TSMC, Intel). These are two areas where Europe is hardly represented. As illustrated in Figure 6, the installed manufacturing

capability of a given region directly correlates with the technology nodes required by the targeted markets of the associated value chain and ecosystem.

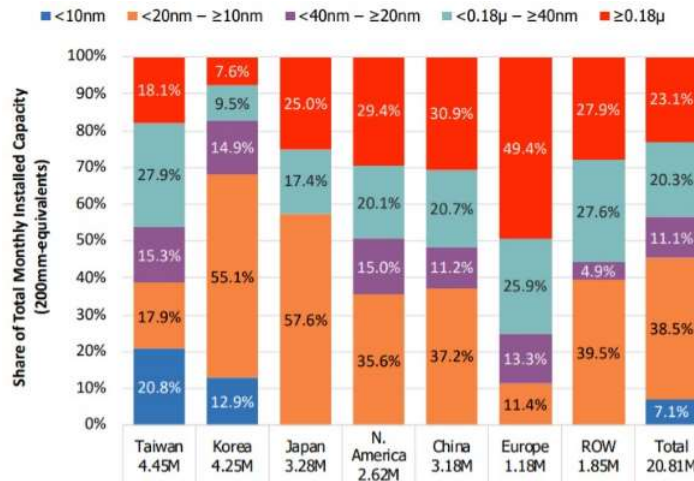


Figure 6: Monthly installed capacity for each geographic region in December 2020 [Nen21]. “ROW” means “rest of the world”.

China, Japan, USA, and Korea have most of their installed capacity for technology ranging from > 10 nm to < 20 nm, which serves their memory production. On the other hand, 75% of Europe’s installed capacity supports > 40 nm (50% for technologies > 180 nm) which serves its key verticals (automotive, industrial, health, ...). Taiwan has a more balanced situation because TSMC’s foundry business model clearly focusses on the most advanced nodes. TSMC’s 1Q21 revenue, depicted in Figure 7, shows that the smaller nodes are driven by the smartphone and HPC business.

1Q21 Revenue by Platform 1Q21 Revenue by Technology

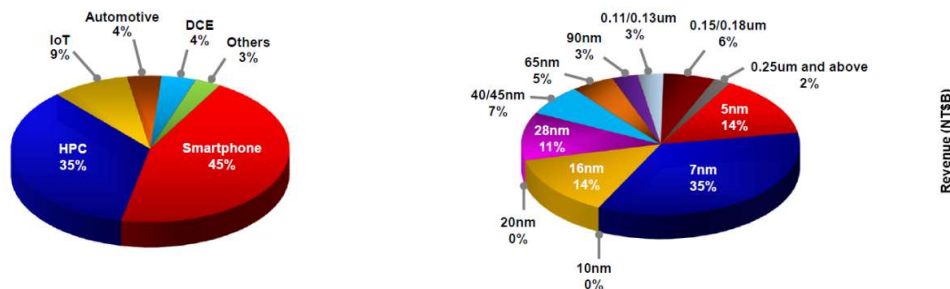


Figure 7: TSMC’s 1Q21 revenue split per application and technology [TSM21].

We can also note that, while 7 nm and 5 nm represent 20% of TSMC’s capacity, they generate ~50% of the revenue. This point is the foundation of TSMC’s high-end foundry business positioning: by maintaining its leadership on advanced nodes and being the first to deliver volume manufacturing, it captures most part of the market value. This enables them to support the high CAPEX required to develop the next nodes and install the necessary capacity (TSMC’s CAPEX in 2021 is set to 30 B\$). Consequently, even a company able to offer equivalent

technology and spending a large CAPEX such as SAMSUNG is today having a hard time to keep up with TSMC. From pure manufacturing side, the entry barrier is high and at short term, it may prove difficult for anyone to dispute the leadership of TSMC. Moreover, advanced manufacturing capability of 7 nm or beyond needs to come with a complex ecosystem that cannot be deployed on a very short term. Indeed, new Fab players will have to build first the design enablement ecosystem (IPs, CAD flow, ...) and prove to be a reliable partner able to deliver targeted performances in large volumes and in time.

While Huawei used to be TSMC's second largest customer just after Apple, the USA export restrictions have reshaped the landscape. Today, most TSMC key customers for 7 nm and 5 nm technologies are USA fabless companies. The only exceptions are Samsung and MediaTek. It illustrates a key weakness of the European fabless ecosystem. Since there are currently no large European fabless or system companies requiring high volumes in extremely scaled semiconductor technologies (< 7 nm node), the current industrial drive to develop such manufacturing capabilities shows to be rather limited. Moreover, given that China and USA are today leading in strategic topics such as AI, they are not expected to own the required manufacturing capability. As such, they are as dependent on the Taiwanese semiconductor technology as Europe.

However, the lack of < 20 nm node manufacturing capability in Europe does not mean that Europe refrains from this topic. Europe's strong position on semiconductor manufacturing equipment enables Europe to play a strategic role on the value chain. ASML is a good example since it is today the sole source of EUV lithographic scanner on the market. Figure 8 shows the EUV shipment forecast by customer and learns how leading foundries such as TSMC, Samsung and Intel are relying today on ASML (and consequently on European technology). This is a strong pledge for Europe.

Company	EUV shipments (unit)					
	2018	2019	2020	2021E	2022E	2023E
TSMC	7	16	18	28	31	33
Samsung	3	5	8	9	14	15
Intel	4	3	3	2	3	5
GlobalFoundries	1	0	0	0	0	0
Hynix	1	1	1	1	1	1
Micron	0	0	0	1	1	1
SMIC	0	0	0	0	0	0
Others	2	1	1	0	0	0
Total EUV shipments	18	26	31	41	50	55
EUV ASP (EUR mn)	105	109	145	145	153	163

Source: Mizuho Securities Equity Research Estimates

Figure 8: ASML EUV shipment forecast by customer.

The recent USA export restriction prevented Chinese companies to access to < 14 nm nodes by preventing USA vendors (such Applied Material, KLA, ...) and foundries to sell their USA-technology based products. This provides interesting perspectives concerning the position that Europe can adopt to safeguard its sovereignty and access to key technologies related to connectivity.

Consequently, even if Europe does not own today the complete connectivity value chain, it still has a leading position on key topics. These topics include semiconductor manufacturing equipment, manufacturing of differentiated technologies, leading position on the wireless

infrastructure market, ... This enables Europe to play a leading role in future connectivity technology development and secure its sovereignty by strengthening its partnership with other countries.

Worth mentioning is the leadership gained by the European RTOs in semiconductor science and engineering to seed the innovations in design architectures and manufacturing technologies. With the support of European funds incentives, basic research is stimulating and attracting industry R&D since their respective contributions are complimentary and not redundant. For example, the FinFET transistor architecture, which has replaced the classical planar architecture in the recent CMOS generations, is the result of several decades of R&D collaborations. Similarly, the technology used by ASML for its flagship EUV lithography started in the 1980s on the use of soft x-rays.

Moreover, from pure manufacturing point of view, the importance of leading-edge technology nodes versus legacy ones must be put into perspective. Figure 9 shows that the capacity in leading edge technology nodes will grow strongly in the coming 3 years while the legacy nodes will still represent a significant portion of the overall capacity (but also growing in absolute numbers).

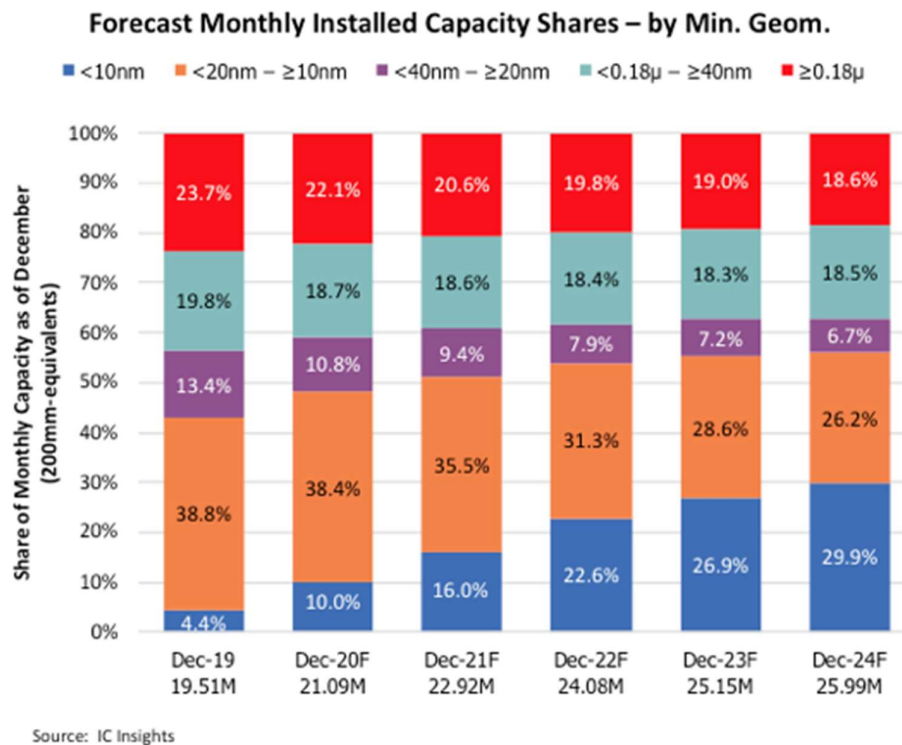


Figure 9: Wafer Capacity by Feature Size Shows [ICi20].

The growing trend of leading-edge technology capacity is mainly driven by consumer products such as smartphones. As illustrated in Figure 10, leading-edge technologies are used in 89 % of the smartphone's application processor and modem, whereas only 5% of the RF and connectivity chips use such technologies. Consequently, in the smartphone market, leading edge nodes represent 27% of the overall chip area (~7.5M 12" wafer count/year) while legacy nodes are addressing the remaining 73% (~47M 8" wafer count/year).

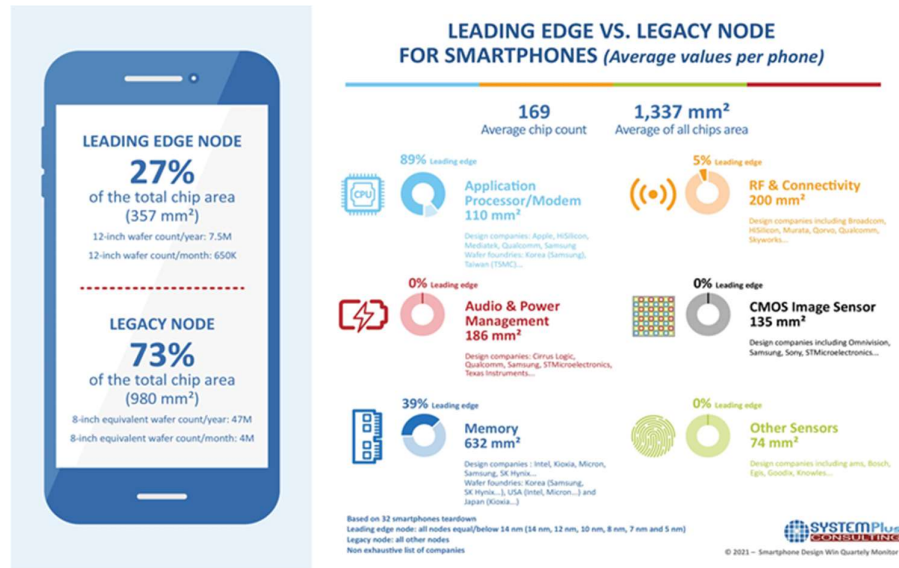


Figure 10: Leading edge versus legacy technology nodes used in smartphones [Sys21].

Given this persisting strong demand, the legacy technology node market will remain very active in the foreseeable future. The associated 200 mm installed capacity is expected to increase to record levels from 2020 to 2024, beating the last records seen in 2006 and 2007 (as illustrated in Figure 11). Moreover, one should also remember that European legacy technology node players are also transitioning to 300 mm fabs to increase even further installed capacity. Infineon's new 300 mm Fab in Villach in Austria, STMicroelectronics' new 300 mm fabs in Agrate in Italy, Bosch's new 300 mm facility in Dresden in Germany and STMicroelectronics' 300 mm fab extension in Crolles in France are good examples of such actions.

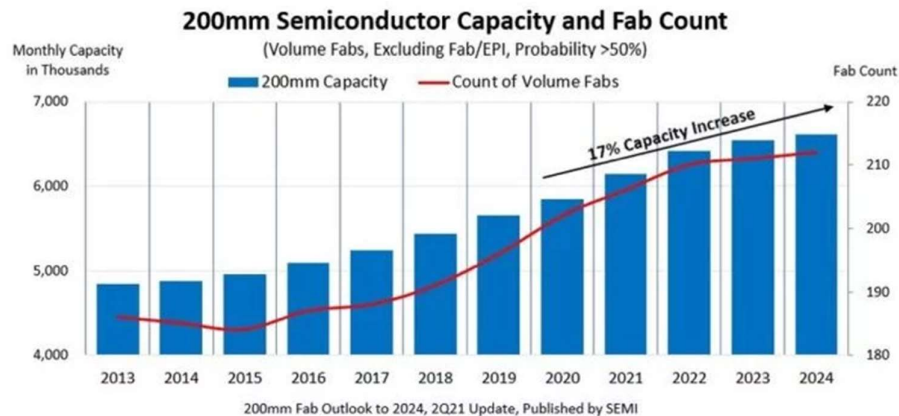


Figure 11: 200 mm semiconductor capacity and count of fabs [Fas21].

Put more simply, there is no opposition to be made between leading edge and legacy technology nodes. Strengthening key partnerships within Europe may secure the technology access and the European sovereignty. Still, not owning the complete technology portfolio does not prevent Europe to capture a significant part of the semiconductor manufacturing chain. Europe's developed technologies and associated installed might, however, remain closely linked to needs of Europe's key verticals and the European overall position in the value chain. Having such

technology and processing capacity, however, is likely to stimulate existing and to initiate new ecosystems in Europe.

3.3. Proposed Industry Roadmap Strategy: How should Europe address future connectivity technologies with a value-chain approach?

Based on the previous section, we can summarize European challenges and opportunities on connectivity technology to support its strategic agenda as follows:

Opportunities:

- Strong global position on wireless and wireline infrastructure markets and R&D
- Strategic link between core semiconductor technology capability and key verticals (automotive, industrial, space and defense, ...)
- Combining those two assets makes Europe a strong contender to take a leading role towards 6G and beyond, while covering the entire value chain
- Europe has strong research in heterogeneous integration for many applications (high-performance computing, photonics, RF)
- In the industrial data market, that is several orders of magnitude bigger than the personal data market, Europe should not leave the storage and handling of these data (both 'central' and 'edge') to non-European companies.

Challenges:

- Digitalization modifies the European value chain, in which Europe risks to reduce its technological sovereignty. The rise of AI, for example, is likely to increase our dependency on USA technologies.
- Geopolitical tensions and trade restrictions impose an increased risk of disruption of the European supply chain. It is key for Europe to mitigate this risk through a more diverse and resilient supply chain.

With financial resources being limited, COREnect is proposing an industry roadmap with a 10- to 15-year timeline. This roadmap addresses different timeframes with a changing focus in terms of strategic investments, markets, and technological development.

Short term (2 years from now):

- **Strengthen areas where Europe is leading (BiCMOS, III-V, RF, analog, mixed signal, photonics ...)** to secure its position and gain market share to ensure Europe to play a leading role for 6G: Europe needs to continue to innovate on differentiated technologies where it leads today to secure its leadership. Transitioning from 200 mm to 300 mm fab manufacturing differentiated technologies is a key industrial challenge.
- **Secure access to < 7 nm CMOS technology:** Digitalization requires the know-how to design dedicated advanced computing chips. Europe needs to secure both the design capability as well as the access to a diverse and trustworthy supply chain.
- **Strengthen the education on IC design (both in analog/RF and digital):** To enable an appropriate pool of experts able to address European industrial players' needs but also make Europe an appealing place to invest. Contribute on open-source initiatives in processor core design IP (RISC-V) to differentiate on what today is openly available.
- **Enable Europe to lead on future connectivity IPR generation, standardization actions while moving higher in the value chain:** Leveraging its current strength, Europe can play a leading role in the definition of 6G. Core technologies developed in Europe should also enable to develop more complex connectivity solutions (smart sensors enriched by AI features, more integrated 6G RF solutions, combination of radar and communication, ...) and then capture more value.

Mid-term (5 years from now):

- **Define an aggressive timeline for the deployment of 6G at the scale of all EU state members:** while EU is today playing catch up on 5G deployment, it would be necessary to well anticipate this topic for 6G. Learning from what China did through MIIT's deployment policy, an aggressive agenda for 6G deployment at the scale of all EU member states will greatly boost EU initiatives and secure EU's capability to leverage connectivity infrastructure asset to serve its key verticals.
- **Strongly support module integration technologies (both design and fabrication) to combine components from a wide range of technologies (advanced digital and memory chips, but also legacy CMOS, FD-SOI, SOI, BiCMOS, III-V, photonics, sensors):** such approach will enable Europe to capture a higher portion of the value chain by delivering systems instead of components. Moreover, it may also allow viable solutions in markets that are not big enough to support the high development cost required by the smallest CMOS nodes. Instead, functionality could be implemented with multiple modularly designed chips in legacy technologies. This can enable defense industry to produce entirely in Europe exploring available production technologies even leading-edge nodes are not available.
- **Strengthen the European position on EDA, IP & software:** Creating an open and more diverse ecosystem is key to ensure a resilient supply chain and enable Europe to mitigate sovereignty risks. Chip design, a field where USA leads, is expected to play a progressively larger role in driving performance improvements as transistor shrinkage slows.
- **Enable smartly positioned European fabless ecosystem:** To play a role in advanced computing chips (or modules) and to mitigate Europe's current dependency, a strong European fabless ecosystem is a mandatory starting point.

Long term (10 to 15 years from now):

- **Make Europe the IC design champion:** Europe must become the global IC design champion to ensure its position in connectivity system manufacturing and strong vertical sectors (application industry and domain knowledge as input for chip design) and to preserve their claimed value chain share.
- **Enable the establishment of < 2 nm CMOS manufacturing in Europe to support the created domestic market.** This can be implemented much sooner through partnerships with strong non-European players.

4. First draft of strategic industry-relevant R&I roadmaps

Based on initially identified COREnect end-to-end system view together with value chain consideration, COREnect organizes three COREnect Expert Groups to address the industry roadmaps in three strategic focus areas:

- Expert Group #1 Compute and Store
- Expert Group #2 Connect and Communicate
- Expert Group #3 Sense and Power

Today, the expert groups consist of 96 experts from industry, SME's, research institute, universities, and associations. 61 experts are consortium members, and 35 experts are from outside the consortium. COREnect continuously seeks to optimize the expert group's composition via open calls (<https://www.corenect.eu/news/call-for-experts>) to obtain an optimal scientific, organizational, and societal balance. For each expert group, a Chair and Vice-Chair have been selected by the COREnect consortium with the responsibility to coordinate

these groups. As a result of several public workshops and focused meetings and brainstorm discussions, the experts aligned their views on Europe’s major challenges and opportunities. Combining their specialist knowledge, COREnect benefits from broad insights into crucial fields that require our attention and, thus, address the upcoming challenges in Europe regarding 5G and beyond. Altogether, capturing and crystallizing the discussions, the three expert groups are cooperatively developing the COREnect industry roadmap.

Although each Expert Group is focusing on a defined strategic area, one can find many overlapping topics within this holistic approach. Security and energy efficiency are cross-cutting design considerations in the COREnect technical vision and inherently the common work of Expert Groups. Therefore, the sphere of **Energy Efficient, Green Communication Electronics** will be equally covered in the chapters dedicated to the expert groups. Therein, advanced computation and sustainable fabrication are going to be examined from the perspective of the device itself (EG1), the transceivers (EG2), as well as from the sensors and power management (EG3) point of view.

Regarding 5G/6G application areas, security and trustworthiness are both a focus for several stakeholders. Therefore, technologies to cope with untrusted third-party IP (EG1), telecommunication hardware and software (EG2) and failure detection and security technologies (EG3) are contributing to the question of how to derive **Electronics for Trustworthy Communication** in 6G and beyond. Regarding the technical aspects of the European sovereignty and its ecosystem, the key question is whether a whole European value chain including the tools and knowledge is feasible, required, and how it can be realized.

An integrated approach of **Future Core-Technologies and Integration** throughout all three expert group chapters is going to highlight the areas in packaging, materials, semiconductor, production, and assembly that can bring Europe forward towards a more inclusive and carbon-neutral society. For instance, multi-chip modules (MCM) and the integration of memory chipllets (EG1), semiconductor trade-offs for wireless and wireline transceivers (EG2), or heterogeneous integration and the semiconductor processing technology landscape (EG3) and its opportunities for Europe are being reflected on in each chapter. Social inclusion with new user interfaces, knowledge, education and job generation, and the general orientation towards megatrends (e.g., demographic change, global warming) will equally play a role, guiding the proposed strategies towards a common goal.

The specific objectives of the Expert Groups *Compute and Store* (EG1), *Connect and Communicate* (EG2), as well as *Sense and Power* (EG3) are described in the following subsections, along with the initially identified gaps and challenges. The key envisioned strategic actions of the individual expert groups and well as across the expert groups are given in Table 2, Table 3, and Table 4. Each table refers to a time perspective as described in section 3.3, i.e., short-, mid- and long-term. Note that these tables are not complete. They will be revised and consolidated over the upcoming deliverables D3.4 and D3.5

Table 2: Envisioned strategic actions on a short term for EG1, EG2, EG3 and across the expert groups.

Short term (upcoming 2 years)		
Strengthen areas where Europe is leading	Secure access to < 7 nm CMOS technology	Strengthen the education on IC design
RAN: broad rollout of 5G over the whole EU	Strengthen Europe on semiconductor manufacturing equipment	RAN: ensure enough IC design engineers are available
Consumer: focus on RF FEM market	Foster digitization of European industries	Increase collaboration between European entities
Consumer: RF Filtering (moved from RAN)		
Industrial: support the migration of EU 200 mm fabs to 300 mm		
Support EU microkernel OS research		
Develop an MPSoC meta-level description standard for integration of third-party IPs		

Table 3: Envisioned strategic actions on a medium term for EG1, EG2, EG3 and across the expert groups.

Mid-term (upcoming 5 years)		
Enable Europe to lead on future connectivity standards	Strengthen European position on EDA, IP & software	Enable smartly positioned European fabless ecosystem
RAN: RF-GaN	Datacenters: extend EU IP portfolio (beyond ARM) to new class of IPs (ADC, ...)	RAN: digitization of transceiver
RAN: mmW beamforming transceiver	x-EG: Strengthen Europe position on EDA solution market	Consumer: mmW transceiver design
RAN: RF-GaN mmW	Develop a modular secure operating system framework	Datacenters: enable low power CPU/GPU design house in EU (using ARM or RISC-V)
RAN: D Band transceiver		Strategic Infrastructure program
Consumer: RF filtering at mmW		2.5 and 3D packaging
Datacenters: enable EU optical module makers		Low power eNVM for AI
Support Europe contribution to standardization activities		
Strategic Infrastructure program		
Radio interface standards		

Table 4: Envisioned strategic actions on a long term for EG1, EG2, EG3 and across the expert groups.

Long-term (10 to 15 years from now)		
Make Europe the IC design champions	Enable < 7 nm CMOS manufacturing in Europe	
RAN: IEEE 802.15.3D transceivers		
Consumer: module for D band communication		
Consumer: radar functionality with D band transceiver		
Consumer: Low data rate wireless devices		

4.1. EG1: compute and store

4.1.1. Introduction

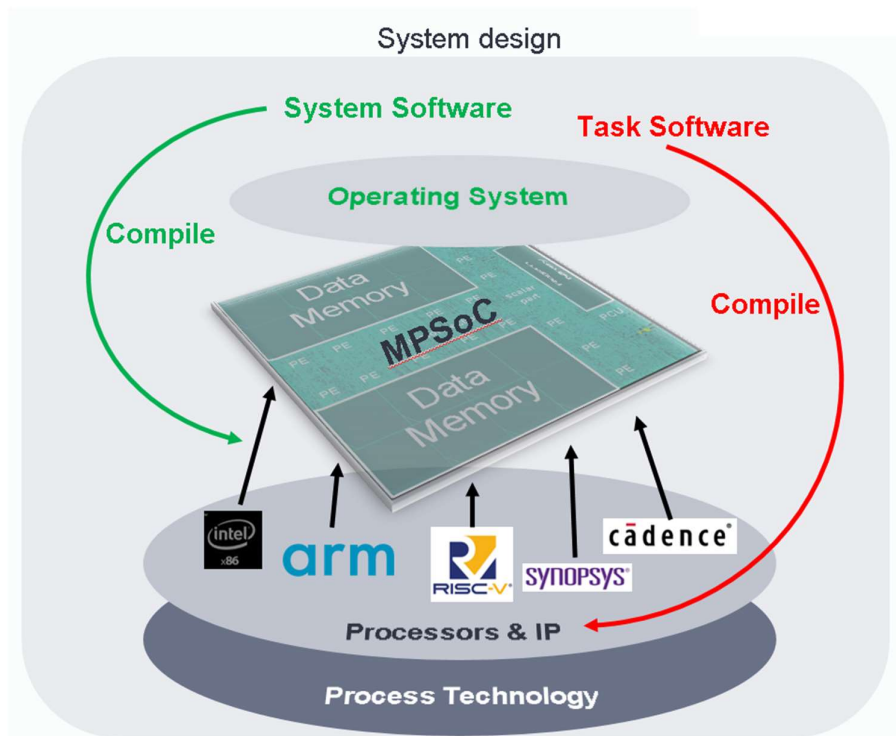


Figure 12: Common architecture of computing systems.

For 5G and beyond, ever more performant digital computing platforms and components will be required. However, Europe is becoming increasingly dependent on non-European supply for these components and systems. This makes the European supply chain highly susceptible to disruptions such as trade wars. Expert Group 1 (EG1) therefore investigates the role of computing and storage solutions for Europe's 5G and 6G sovereignty.

EG1 identifies key technologies that enable Europe to build trustworthy and competitive systems for communication (terminals and the RAN) and different applications (IoT, personal devices and personal mobile robotics). The focus will be on programmable computing platforms and the included storage components under the consideration of different operating constraints and the need to support legacy software.

In section 4.1.2, we will begin with a description of the system requirements for the RAN and the UE (User Equipment) use cases. From these requirements we derive the form needed to fulfill the specific functions at each layer of the computing platform from the bottom-up in this whole chapter (c.f. Figure 12). At the bottom there are the process technologies that are used for physical realization. Section 4.1.3. states some initial thoughts on the requirements which were left out in previous statements of EG1. Modern computing platforms contain at least one, more often multiple, cores whose software interface is defined by the instruction set architecture (ISA). For different situations, such as infrastructure equipment or edge devices, different ISAs may be used, as elaborated in section 4.1.4. The next section, section 4.1.5, analyzes the European position in memory and storage and how it can be improved. The

challenge of integrating multiple heterogeneous cores, memories, accelerators (c.f. EG3) and intellectual property (IP) blocks in a Multiprocessor system on a chip (MPSOC) is portrayed in section 4.1.6. Finally, section 4.1.7 proposes an operating system framework which is based on the principles of modularity and microkernel architecture.

4.1.2. 5G/6G System architecture

The 5G/6G System Architecture is discussed in two parts: first the Radio Access Network (RAN) and next the User Equipment (UE). Figure 13 shows a logical architecture of the RAN. It is a refinement of Fig. 6.1-1 of 3GPP TS 38.401. The additional interfaces (in green) are specific to the O-RAN Alliance proposal but can also be proprietary.

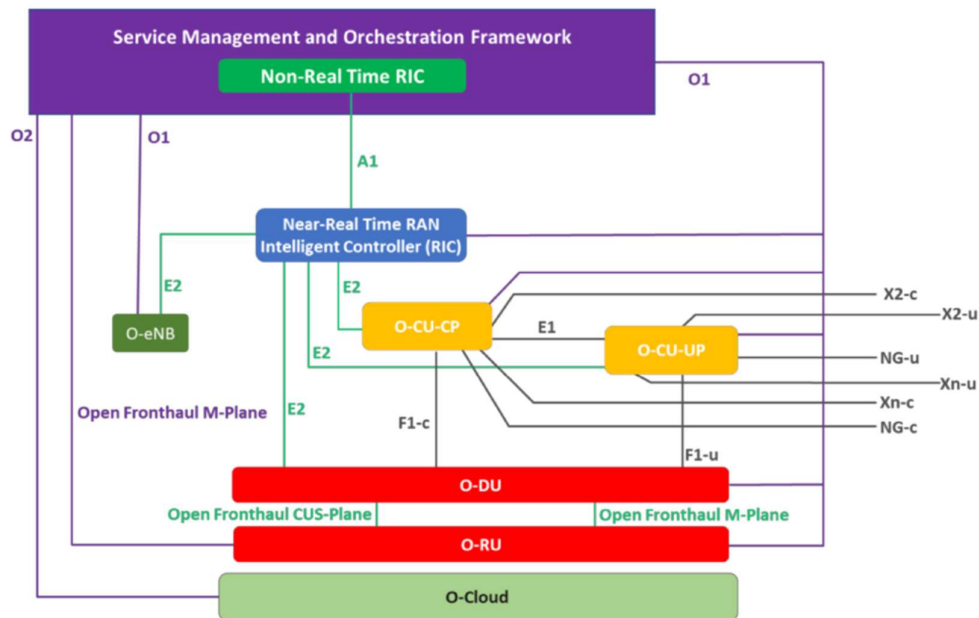


Figure 13: O-RAN overall logic architecture [ORA19].

4.1.2.1. RAN workload

Most of the RAN compute load is in the radio (RU), central (CU) and distributed units (DU), where (most of) the L1-L3 signal processing occurs. This workload:

- addresses multiple standards, including (3G), 4G, and 5G;
- is complex, highly dynamic, and involves a large number and large variety of different tasks;
- is measured in Peta-ops/sec;
- has to meet extreme latency constraints, measured in micro-seconds per task;
- must support so-called macro network blocks, including their activation and termination.

In addition, some application functions may migrate in part to the base stations (close to the O-CU and final IoT edge AI devices).

4.1.2.2. RAN compute platform

The compute platform likely is to comprise multiple System-on-Chips (SoC). These SoCs are realized in advanced CMOS, possibly including new nanoelectronics technologies such as MRAM, and are connected using novel packaging and interconnect technologies.

- Each SoC includes multiple (100s), diverse compute cores. The HW architecture is heterogeneous.
- Cost pressure and the need for software upgrades will push for programmable ISA-based cores, whereas power constraints will dictate a significant degree of specialization, using a variety of specialized accelerators.
- These cores must be real-time capable to support their dynamic allocation on μ sec-msec time scales. Only then can today's massive over-provisioning of hardware resources be avoided.
- The memory organization will be a key challenge: the need for low cost and high flexibility will push for memory unification and centralization, whereas the need for low power will push for memory specialization and distribution.
- The mix of cores and the memory organization must be such that the overall multi-MPSoC RAN can be scaled over a range of workload sizes.
- Also, new system architecture design and thermal-aware and energy-aware optimization methodologies need to be created to enable trade-offs between security, power and performance for the 5G/6G context.

4.1.2.3. RAN run-time mapping

Running the complex, highly dynamic petaflops workload on such a multi-SoC RAN computing platform *efficiently* is exceptionally challenging. It involves:

- Dynamic, real-time multi-tasking. *Virtualization* is an ultimate form of this, and it is seen as the holy grail.
- Adaptive resource management, meeting power and thermal constraints during operation.

4.1.2.4. UE

On the UE side, the main challenges arise from the huge *diversity* in 5G/6G modem requirements, from primitive IoT devices to high-end smartphones. This diversity is about:

- the range of standards (4G, 5G, Wi-Fi, Bluetooth, GPS, NFC, Ethernet...) to be supported,
- the ranges of required bitrates (kbps – Gbps), and latency requirements.

Accordingly, the main architectural challenge is building a scalable modem (as IP block, incl. HW/FW/SW), addressing multiple market segments?

Additional UE architecture challenges include:

- support of beamforming, including distributed (coordinated) beamforming by multiple IoT devices;
- additional flexibility on MPSoC to enable ISA extensions and use programmable or reconfigurable (coarse-grained or fine-grained) accelerators;
- adaptive resource management approaches at system level, including power and thermal constraints at run-time during operation;
- potential definition of dynamic, and real-time multi-tasking and architectural flexibility, where virtualization can be an ultimate form of such a flexibility.

4.1.3. Process technology

The future perspective of programmable computing platforms (PCP) for communications dictates the availability of semiconductor technologies with very specific characteristics. Due to, in many cases, contrary requirements of the components in such PCPs, we can already see today that those computing platforms will consist of components based on application-optimized

technologies. Within the digital domain, following criteria will dominate, based on PCP requirements:

- data rate → computational performance at low power
 - leading edge (7 nm-5 nm-3 nm)
 - thermally optimized 2.3D / 3D packaging technology
- real-time → latency → specialized hardware-based AI accelerators
 - low-power technologies for AI on value added technologies
 - low-power embedded Non-Volatile Memory (eNVM)
- cost → cost efficient and reliable 3D packaging technology for system-in-package
- trust → hardware (technology) based secure components (PUFs, secure NVM, etc.)
- bandwidth → RF semiconductor technologies with peak frequencies above 600 GHz

Logic/Foundry Process Roadmaps (for Volume Production)

	2016	2017	2018	2019	2020	2021	2022
Intel	14nm+	10nm (limited) 14nm++		10nm	10nm+	10nm++	7nm EUV
Samsung	10nm		8nm	7nm EUV 6nm EUV	18nm FDSOI 5nm	4nm	3nm GAA
TSMC	10nm	7nm 12nm		7nm+ EUV	5nm 6nm	5nm+	4nm 3nm
GlobalFoundries			22nm FDSOI 12nm finFET		12nm FDSOI	22nm+ FDSOI 12nm+ finFET	
SMIC				14nm finFET	12nm finFET		8-10nm finFET
UMC		14nm finFET			22nm planar		

Note: What defines a process "generation" and the start of "volume" production varies from company to company, and may be influenced by marketing embellishments, so these points of transition should only be seen as very general guidelines.

Sources: Companies, conference reports, IC Insights

Figure 14: Leading Edge foundry process roadmap [ICI21].

Moore's Law has always been about power, performance and cost. Continuously increasing amount of data being transferred and processed will require system components in a PCP which are built based on leading-edge digital technologies ("7 nm", "5 nm" down to "2 nm"). Only these technologies will be capable to resolve the trade-off between computational performance and power requirements (and cost). Figure 14 shows an overview of the logic/foundry process roadmaps as distributed by [ICI21].

It should be noted that Figure 14 depicts the leading-edge offerings only. Integration of NVM, analog, power and RF into CMOS-technologies can only be done on legacy technologies. This aspect is of high importance, as significant reductions in power, construction size and cost can only be achieved with value-added integration.

In addition to technology advancements, significant reductions in power consumption can be achieved using specialized AI components performing dedicated tasks within a PCP. These AI

accelerators not only contribute to the real-time capability of the entire system. Properly designed and using AI-optimized technology components, like low-power analogue calculators and low-power embedded NVM, these AI accelerators significantly reduce system power consumption. Neither the AI functions, nor the eNVM components require the most advanced technologies, but would like to use the functional integration of value-added technologies. Moreover, eNVM can only be offered in rather mature technologies, which are also available through European semiconductor manufacturers. So, the focus here must be placed more on the architecture and design of such AI components and the usage on trusted technologies from Europe in view to achieve increased sovereignty. In addition, the number of specialized AI accelerators in the PCP should be increased to a possible maximum to save as much power as possible.

Replacing generic AI-accelerators with dedicated AI-accelerators on processes with high integration value, trustworthy and secure provision from Europe, will position us well for the future in the context of 6G.

Due to numerous functional components in a future PCP, special focus must be put on the system integration. 2.5D and 3D packaging technology is being extensively used to reduce parasitic limitations (I/O loadings) and thus increase system efficiency to both, power consumption and data throughput. On top, appropriate packaging technologies are capable to combine trusted components (AI, eNVM, RF) with those originating from untrusted manufacturing sources (digital “7 nm”, “5 nm” down to “2 nm”). Using appropriate components providing trust, untrusted system components can be used and “upgraded” to the required trust level.

Unfortunately, standard packaging technologies are not very well represented in the European technology landscape. However, there is the opportunity to gain a lead in advanced 2.5 & 3D integration. Special focus will have to be placed on thermal aspects and we will have to divert some of the gains of Moore’s law from “performance” to “lowering power”, more than we have done in the past.

Controlling 3D-Integration will provide us with a lead in designing security solutions and build trusted systems out of Europe independently.

4.1.4. Instruction set architecture (ISA)

The Instruction Set Architecture (ISA) and its realization as a microarchitecture can be an enabling factor for innovation as described in D3.1. In the context of 5G/6G applications, we need to consider solutions covering IoT and edge devices as well as infrastructure equipment. For infrastructure equipment standard ISAs provide a good solution, while for small IoT and edge devices more application-specific solutions based on non-standard ISAs/DSLs are needed. For both scenarios also, solutions based on application-specific adaptations of standard ISAs, such as RISC-V which is an open-source initiative for core processor architecture design, play an important role. We identified the following open challenges and topics for each of them.

Standard ISA for infrastructure equipment:

- **High Performance Chip Design:** High performance implementations in silicon are very challenging and require highly skilled CPU architects/designers (and possibly custom design at the gate level), which is currently a scarce resource in Europe.

- **Dynamic Binary Translation (DBT):** Another road to high-performance ISA implementation is using DBT to a simpler core, as done by NVIDIA with the Denver/Denver2 cores implementing ARMv8 ISA and by Apple with the Rosetta DBT. An extended RISC-V implementation could be the target of a DBT, both could possibly be in the EU skillset.
- **Adaptations of Standard ISAs (e.g., 64-bit RISC-V):** There are a few EU 64-bit RISC-V implementations: The Ariane-based from ETHZ (now on openhwgroup.org/cva6), the NOEL-V from Cobham Gaisler, the Avispado and Atrevido from SemiDynamics. It is not expected that EU-designed RISC-V processors become competitive with high-end X86 or ARMv8 implementations in USA or Asia (China, Japan, Korea) for latency-constrained applications. However, it is expected that these eventual performance limitations of EU RV64G cores will not be a major problem for accelerated computing scenarios at the edge, in particular for 5G/6G applications. On the contrary, the ISA adaptation opportunities of the RISC-V environment will enable more tuning for these accelerated computing scenarios, in particular in the privileged ISA (on the memory models).
- **Formal verification of ISA implementations:** Whether RISC-V related or not (OneSpin also did it on the Infineon TriCore2, on Bosch DSPs, etc.) are of high importance to ensure trustable and reliable realizations.

Non-standard ISAs / DSLs for edge devices:

- **Non-standard ISAs:** Can be hidden and easily integrated into the software stack if they operate beyond a DSL; this is already the case for AI in Multi-access Edge Computing and could be the case for the L1/L2 RAN processing.
- **Acceleration DSLs:** May lead to specific/relaxed requirements on the accelerator memory model and the way it is seen from the GPP. Having the accelerator to operate under CAPI, CCIX or CXL may be less important than in datacenters, as latency and energy-efficiency plays a crucial role in the edge.

4.1.5. Memory and Storage

In this section, we describe our vision of Europe's industry roadmap in the broad field of storage, including on-chip memory. First, we discuss the scope of this roadmap. Second, we perform a situation analysis by addressing the following questions:

- Why is it important for Europe to work on storage?
- In terms of storage, what are the key developments and trends inside and outside Europe?
- In terms of storage, what is Europe's wanted position in 10 years?

Third, we discuss industry targets for Europe in terms of storage during the next 10 years to be able to reach the wanted position. Fourth, we discuss research questions that Europe should address to be able to meet these industry targets. And finally, we discuss partners and competences that Europe would need to be able to conduct that research and reach the targets.

The scope of this roadmap can be described as follows. The main scope is on the architecture level of the on-chip memory hierarchy. For instance, the scope includes: 1) new packaging technologies (e.g., 3-D integration, chiplets, large-scale wafer), 2) on-chip memory interfaces and interconnects, 3) embedded non-volatile memory (eNVM) and especially low-power eNVM, 4) high-bandwidth DRAM (HBM DRAM), and 5) intellectual property (IP) for on-chip support of wireless storage (storage-over-radio-network) enabled by low-latency and high-bandwidth connectivity like 5G and 6G. The scope does not include, e.g.: 1) off-chip DRAM DIMMs and 2) memory-technology details like materials and device physics. We discuss embedded technology access only in cases where we can strongly motivate its strategic weight for Europe.

The first question of the situation analysis is: Why is it important for Europe to work on storage? Today there are no companies in Europe that work on storage, and this gap is a potential threat to Europe. By working on storage, Europe has a chance to master 3-D integration, which can be considered as a key technology that other countries might be lacking. Likewise, by working on storage, Europe can *understand* the design of memory IP and thus be able to do memory-aware system design (i.e., acquire competence to design systems-on-chip (SoCs) and systems-in-package (SiPs) that have efficient utilization of the memory hardware). Further, Europe could *master* the design of memory IP (including circuit design) and license that memory IP to the rest of the world.

The second question of the situation analysis is: What are the key developments and trends with respect to storage inside and outside Europe? We would like to point out that memory research from Europe gets productized outside Europe. At the same time, Europe imports memory components. Europe is strong in innovation but weak in selling its innovation. Given this, a question is which trends do we want to assume: will Europe become stronger in innovation? Will Europe become weaker in selling its innovation? [Kle21] describes that fabs are decreasing in their numbers and provide to external parties less and less access to design kits and technology details of their memory IPs. Thus, Europe needs to assess the potential threat of losing such access completely. Another trend is that technology giants outside Europe prefer to either master technology themselves or to acquire European companies that own the technology. This means that it might become more difficult for Europe to benefit from licensing memory IP. Another trend is that due to miniaturization and higher integration, memory and storage converge. This means that 3-D integration might be a key technology. At the same time, we know that data amounts grow faster than local (on-device) storage capacity, which increases the usage of wireless storage (e.g., storage-over-radio-network, such as cloud accessed via a cellular network).

The third question of the situation analysis is: What is Europe's wanted position with respect to storage in 10 years? We think that Europe might want to master 3-D integration – a technology on which other countries might depend. Europe should be able to affect standardization in councils like JEDEC, where Ericsson and Nokia are members. Europe should have strong bonds with strategic partners cleared by governments to minimize the risk of supply chain disruptions due to factors like trade wars. Since Europe has Ericsson and Nokia, that can enable next-generation storage-over-radio-network, Europe should use this momentum to further increase the use of such wireless storage. We think that Europe should be able to research and innovate (R&I) *at least* on the architecture level and to productize systems (i.e., to be able to design SoCs that employ innovative memory IP and then productize such SoCs). A higher level of ambition for Europe would be to also R&I on the memory-component design level and to productize components. The highest level of ambition would be to also R&I on the embedded technology level and to productize technology. Like we mentioned in the description of the scope of this roadmap, embedded technology access requires strong motivation, and here we merely describe it as the highest ambition level for Europe's wanted position.

Regarding industry targets for Europe with respect to storage during the next 10 years, we have the following recommendations. In general, we think that Europe needs to increase the amount of collaboration among European entities (currently, there are more collaborations with entities outside Europe). Such collaborations include knowledge exchange between academia, research institutions, and corporations. To increase such collaborations, Europe should reconsider the

collaboration and patent strategies in European academia, research institutions, and corporations. To keep momentum in standardization, Europe should set appropriate targets in increasing the European portfolio of IP (e.g., patents) in the broad field of storage, even if European companies do not productize such IPs.

In terms of work on disruptive technologies, Europe could consider facilitating the transition from the classic memory hierarchy used in von-Neumann computers to memory used in emerging non-von-Neumann computers. In terms of securing leadership, Europe could focus on 3-D integration, as the technology is still relatively new, yet it has a potential of becoming a future dominator technology.

Regarding research questions that Europe should address to reach its storage-related industry targets during the next 10 years, we again recommend choosing an appropriate level of ambition (architecture-level, circuit design-level, or technology-level). On the architecture level, we consider questions about SoC design, chiplet library, 3-D integration, integration of eNVM for high-density on-chip storage, tight coupling of SRAM and eNVM, hybrid memories like eDRAM & eNVM, and efficient integration of storage-over-radio-network into the memory hierarchy. We would like to point out that 3-D integration technology might be driven by storage (e.g., due to high market volumes), and at the same time storage might be driven by 3-D integration (e.g., due to supporting higher storage densities (capacities per area)). On the circuit design level, we consider questions about memory throughput (e.g., pipelining), memory interfaces (e.g., vectorization), latency / bandwidth / energy optimizations of components used in storage-over-radio-network, and the support of multi-level cell (MLC) operation (e.g., programming and sensing techniques). On the technology level, we consider questions about eNVM improvements, e.g., how to increase endurance, how to reduce cell variance and thus enable MLC with greater numbers of logical levels, how to reduce operational latencies, how to reduce power, how to reduce the process node and thus mitigate the dependency that eNVM might limit the process node for the rest of the chip. Likewise, on the technology level, we consider questions about SRAM improvements. This, in turn, raises a strategic question whether Europe wants to be independent in terms of CMOS technology. The CMOS race is already lost for Europe against giant fabs like TSMC or Samsung (for instance, today Europe still cannot fabricate FinFET, while the giants have been fabricating it for multiple years and continue to increase their momentum with investments like the recently announced fabrication capacity expansion plan exceeding of \$100 billion). However, to be able to fabricate advanced FinFET technology nodes like 7 nm (but not the most advanced ones like 2 nm) still might be a strategic opportunity for Europe – a way to become less dependent and to avoid being at the end of this long supply chain. Moreover, a research question that spans from the architecture level down to the technology level is about memory security features.

Finally, regarding partners and competences that Europe would need to be able to work on the research questions and achieve its industry targets, we would like to state again a potential threat that today Europe has no companies with memory / storage products. Inside Europe, potential partners on the architecture level are: Bosch, Dolphin Design, Ericsson, Infineon, NXP, ST; on the design level: Fraunhofer, imec, Mentor (Siemens); on the technology level: ST. Outside Europe, potential partners on the architecture level are: Apple, Huawei, Intel, Kioxia, Micron, Qualcomm, Renesas, Samsung, SK Hynix, Western Digital; on the design level: ARM, Intel, Synopsys, TSMC; on the technology level: GF, Intel, Samsung, SMIC, TSMC.

4.1.6. Multiprocessor system on a chip (MPSoC)

Besides considering the initial requirements regarding trustable blocks, and security indicated in D3.1, which all are key research questions for the MPSoC level in the three scenarios shown in section 4.1.2, the core concept that needs to be developed at the MPSoC level is the definition of a new meta-level description standard. This new standard needs to provide common interfacing between IP blocks and units of the MPSoC to operate in real-time during run-time operation. Moreover, this standard must guarantee real-time scheduling (microsecond level) in close coordination with the OS level.

As a result, we have identified the following five core challenges and topics to cover in the design of MPSoC for the 5G/6G context to develop the aforementioned core concept:

- New MPSoC design and optimization methodologies need to be created to enable trade-offs between security, power, performance, and time-predictability for the 5G/6G context. MPSoC design is not currently real-time capable (i.e., it can fulfill hard-real time of ten μ sec to few msec latency). Therefore, to accommodate this requirement today, MPSoCs used in telecom infrastructure are massively over-provisioning the memory hierarchy (power and area wastage).
- Additional flexibility on MPSoC to enable ISA extensions and use programmable or reconfigurable (coarse-grained or fine-grained) accelerators for the compute-intensive parts of 5G/6G layers. Currently, the key elements or blocks used in this context by industrial designs are hard-coded IP hardware accelerators or FPGA devices. However, the latest communication standards need to be more flexible to enable scalable MPSoCs to increase system performance by including more specialized processors, AI chips to increase automation and heavily heterogeneous designs.
- Hardware/Software co-design of MPSoCs for the 5G/6G context requires a new higher-level software stack definition. This part must include the OS and development frameworks to exploit further application-level knowledge to understand the performance bottlenecks of different 5G/6G services and effects on the MPSoC architectural blocks. In this context, the classic use of coordination languages based on execution models, such as dataflows, is not enough. These languages are too rigid to adapt to the time-varying nature of telecommunications workloads.
- A new generation of flexible interconnects for MPSoCs needs to be added to system components and IP cores that support the real-time capabilities. In particular, these new interconnects should enable terminating and activating the macro network interfaces, such as synchronized or time-sensitive Ethernet, in the elements of the MPSoC architecture. In this context, this new generation of MPSoC interconnects has to ensure the quality of services, including guaranteed latencies and throughput.
- Definition of how to interconnect multiple MPSoCs as beamforming and other operations of 5G/6G require multiple MPSoC instances to cooperate and operate synchronously closely. Therefore, new connection standards needed for 5G/6G at the MPSoC level need to go beyond near-memory standards or classical networking interfaces to get more efficiency. In addition, at the MPSoC level, it is necessary to control functional isolation and security for the distributed infrastructure.

4.1.7. Operating System Framework

The Operating System Framework is the common ground for the base-level software platform that facilitates all use-cases targeted by COREnect, and beyond. All use-cases have in common that any network-connected system needs a sound base-level software platform that can fulfil indisputable requirements of security, safety, and trustworthiness, together with versatility, real-time, virtualization and applicability to run multiple distrusting applications with diverse

and potentially contradicting requirements on a single platform. The framework shall also closely follow and collaborate with hardware development for an efficient interaction between software and hardware.

COREnect addresses the areas of mobile devices, infrastructure, e.g., cloud and HPC, as well as severely constrained devices used in sensors and IoT applications, which is a wide range of the hardware and compute scale. Further, ongoing development in the hardware area pushes the increase of compute capacity and overall resources by adding more cores, increasing interconnect bandwidth, and adding diverse compute units (commonly called accelerators). All this is driven by the mobile and cloud markets that have a high volume and consequently resources to invest into the evolution of IT hardware architectures. On the software-level, this evolution cannot be observed. Currently, there is no base-level operating system software implementation that can address the increasing demand on the hardware and its diversification due to accelerators and the difficulties to continue with Moore's Law.

Given the wide range of hardware and different applications that shall be covered, there is no "one size fits all" approach to design an operating (OS) system. The OS rather needs to be a modular system that can provide the required functionality in both features and properties such as security and safety and fully exploit the functionality of modern and future MPSoCs. Modularity enables that only functionality is included that is required, leaving out unnecessary software complexity, and allows to cluster functionality. Thereby the system can be built with multiple isolated domains. A generic OS is therefore actually a framework of building blocks, allowing to build application-specific systems and subsystems within one system such that all requirements, especially for safety and security, can be implemented. The key ability is to minimize dependencies as much as possible such that only required functionality is within the dependency of a feature. Eventually the framework allows to build, together with virtualization technology, a hierarchy of operating systems running on the base-level OS.

Modularity is also becoming a cornerstone to the base-level software as architectures are getting more distributed by placing compute units across a platform. Distinct computing systems are not only running in the general-purpose CPUs but also in peripherals and devices, e.g., in Smart-NICs/DPUs, as well as in computing systems that are connected to the same memory or are connected via fast coherent or non-coherent connections or networks, such as CXL, CCIX, or Gen-Z.

COREnect therefore recommends that an operating system framework follows an architecture that is based on an open microkernel design, which is founded on modularity from the ground up and passes the following distinct criteria:

- A modular, microkernel-based system that focuses on security, safety, reliability, and trustworthiness: Only with small software modules it is possible to implement convincing security, safety, and real-time properties that can eventually withstand evaluation and certification as required for many important use-cases.
- Modularity also brings flexibility in system design and consequently portability and adaptability to different use-cases and upcoming hardware evolutions.
- Openness is required to establish trustworthiness of the system, allowing an independent evaluation of it. The best approach to providing the required openness is an open-source system.

- The OS software layer needs to be efficient to provide applications as much compute resources and thus minimize power consumption. Providing necessary means for controlling hardware regarding power consumption is a must.
- For a broad applicability, the OS needs a decent level of documentation and accompanying documentation such that it is generally usable.
- The system needs to provide a decent level of compatibility to be able to host existing software and applications which is typically implemented through virtualization techniques and providing common APIs for applications.

Europe is in a distinct position for an open and versatile operating system framework with prior research and development that went into secure, safe, real-time, and open-source operating system designs and implementations, funded by EC and national funding bodies. Ideas and implementations have been brought out of academia and are used in industry for selected use-cases today. Building upon those initial research and funding investments as well as industry experience, we propose that an operating system framework shall be developed openly and collaboratively for the benefit of providing an open and accessible OS framework. This shall avoid parallel development efforts and multiple, possibly incompatible, implementations for the benefit of all.

Providing an open, secure, and safe implementation of a versatile operating system framework on common hardware platforms, developed together with ongoing hardware activities, requires support and resources. Besides the core development and maintenance, especially providing the necessary means for qualified software demanded by many markets, needs support and resources as the required efforts are substantial for achieving the high level of confidence required for software.

The operating system framework must be open and freely accessible to provide the service of improving the overall system security in the IT landscape and providing a benefit for users and system integrators compared to other existing solutions that are used in the market today.

4.2. EG2: connect and communicate

6G wireless communication will make use of frequency bands from 5G and its predecessors but in addition use extra bands. A clear extension of the spectrum beyond 5G is the use of carrier frequencies above 100 GHz, which we address here as (sub-)THz communication. Here, the D-band will be used initially. There are two definitions of the spectrum of the D-band: the waveguide D-band is from 110 GHz to 170 GHz, while the ETSI Industry Specification Group (ISG) on mm-wave transmission (mWT) defines the D-band from 130 GHz to 174.8 GHz. In a later phase, even higher frequencies than the D-band can be addressed (e.g., the first standardization efforts have resulted in IEEE 802.15.3d, which targets the spectrum 253–322 GHz, which is the higher part of the G-band). However, a lot of communication in 6G, if not the majority, will happen below 100 GHz, using existing frequency bands or gaps in the microwave and mm-wave spectrum. In this section, we refer to the microwave frequency region as the region of carrier frequencies roughly below 20 GHz, whereas we call the mm-wave region the frequency band above 20 GHz up to 100 GHz. Above 100 GHz we address the spectrum as the sub-THz band.

Designing transceivers that use carrier frequencies in the sub-THz band is challenging for the active devices. Whereas CMOS downscaling has enabled a massive deployment of CMOS in microwave and mm-wave and high-speed wireline applications, the speed of a single transistor, expressed in terms of the maximum frequency oscillation f_{MAX} , at which power gain has dropped to 0 dB, is more or less saturating to a value below 400 GHz for downscaling beyond the 28 nm generation. This means, for example, that in the D-band (roughly at one third of this 400 GHz) the maximum power gain per stage is less than 3, but it is always lower due to losses in passive components. Further downscaling of CMOS does not boost f_{MAX} anymore, it only leads to a reduction of the size of digital standard cells, still giving a performance improvement for digital signal processing but not anymore for analog signal processing at high frequencies.

Higher f_{MAX} values than the best values for CMOS can be obtained with silicon bipolar transistors, which can be combined with CMOS in a BiCMOS process. Even higher f_{MAX} values are obtained with high-mobility III-V materials such as InP. Such material even has the advantage over silicon that it can still operate at higher voltages, such that InP is better suited for power generation in the sub-THz region.

With an f_{MAX} value for CMOS of almost 400 GHz, CMOS can be used for wireless transceiver design from the low-GHz range up to the D-band, as evidenced both by prototypes described in publications and by products. Only for the generation of power, CMOS falls short, both for cellular applications in the low-GHz range and for the D-band. The consequence is that for transceiver design, except for the power amplifier, technology choice is more based on economical or strategic considerations than on performance.

For the roadmap discussion in this section, we divide the field of “connect and communicate” into the following 4 domains:

- Radio access networks (RAN)
- Consumer grade connectivity
- Industrial grade connectivity
- Datacenters

The generic block diagram of a wireless or wireline transceiver is shown in Figure 15. In between the transmit part and a receiver (from another transceiver elsewhere), there is a medium, which can be the air, a cable, a PCB track, ... The interface from the electronics to the medium is either an antenna or antenna array for wireless communication, a photodiode for optical communication and a copper wire for other wireline communication. The analog electronics make the connection between the digital part of a transceiver and the interface to the outside world. In between the analog electronics and the digital part, analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) perform the conversion from the analog to the digital domain or vice versa. The functionality of the analog electronics depends on how much functionality of a classical analog front-end is implemented in the digital domain. There is a shift towards the digital domain of functionality that is classically performed in the analog domain. Examples are seen in digital transmitters for wireless communication and in highly digital wireline transceivers.

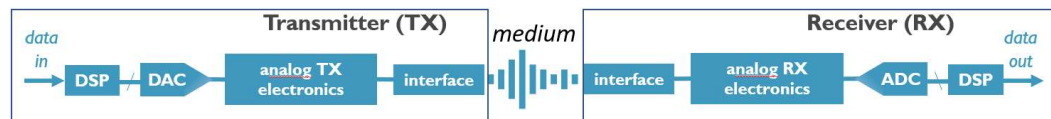


Figure 15: Generic block diagram of a transceiver.

4.2.1. Radio access networks and base stations

Base stations will cover a wide range from outdoor to indoor, from large cell to small cells. As already pointed out in Section 3.3, Europe has a strong position in the wireless infrastructure market and this position should not get lost.

The new RAN will heavily rely on software-defined solutions, meaning that virtualization of the RAN will be essential for new 5G architectures, where the use of software-defined solutions will bring flexibility to the architecture in both terms of functionality (software) and openness (multi-vendor/hardware). Within this future scenario Open RAN aims to improve interoperability in the future network architectures and deployments, in particular when referring to disaggregated multi-vendor solutions. The next section discusses Open RAN more specifically. After these sections, the more hardware-related items are discussed.

4.2.1.1. Open RAN

Open RAN aims to create a multi-vendor RAN solution to improve network flexibility, competition and costs. Such solution will enable the functional disaggregation between hardware and software, offering open interfaces via software solutions that controls the network. The current initial deployments based on Open RAN, are exploring –and validating– the impact on performance KPIs (e.g., latency) due to the disaggregation, as well the impact on security. The O-RAN alliance is defining the specifications for the software-based RAN considering the specifications of 3GPP. The O-RAN alliance is an Operator and Vendor oriented organization which aims to provide the specifications and recommendations to enable interoperability among different vendors and to enable new functionality with the support of AI/ML for intelligent control and aims to provide efficient management and orchestration of the RAN.

As mentioned above, there is a strong software component in the new RAN, where hardware needs to be aligned to open RAN to be ‘pluggable’ in the network and easily orchestrated via the common interfaces. Therefore, hardware needs to be capable to host the flexible open RAN software. In other terms, hardware which is locked to a closed solution will be no longer interesting for the use of networks operators.

Europe requires to have strong players in the Open RAN ecosystem. The O-RAN alliance comprises the big players (operators and vendors) from the USA, Asia and Europe. However, there is sufficient room for smaller players (vendors and integrators) that can innovate in the different RAN components from HW and/or SW perspective. European players could strengthen their position in the Open RAN arena from an integrator perspective, where their expertise on the open SW of open RAN is demanded by different worldwide vendors. Moreover, the emergence of the players may be associated not only to open RAN but also to companies with strong hardware (vendors) or software (vendor/integrator) expertise.

As open RAN evolves, other capabilities are expected to emerge and to further develop its capabilities related to intelligent and programmable RAN. This may pave the way for new actors like startups to emerge and take relevant roles. The European Commission, through the launched study on 5G supply markets and Open RAN¹⁷, certainly has included Open RAN in the road map of the Digital Strategy for Europe, considering Open RAN as a key technology to be addressed and to be prioritized by European companies (both small/medium and large companies). In the long run, operators must be able to offer connectivity based on flexible, multi-vendor, disaggregated architectures without compromising security, reliability, availability, QoS, and energy efficiency.

4.2.1.2. Short-term hardware needs

The increase of wireless data communication, be it at microwave, mm-wave or sub-THz frequencies, will inevitably lead to higher complexity in the processing (e.g., implementation of beamforming at mm-wave or sub-THz) at the base station side. This complexity would justify the use of highly downscaled CMOS. The exact choice of the CMOS generation requires attention as the non-recurring engineering cost increases with downscaling, while volumes for base stations are small compared to user equipment.

The signal processing at base stations will increase in complexity and AI will play an important role here. The design of complex base station chips in Europe requires that enough engineers should be available for digital design in advanced CMOS technologies and that European universities and research centers provide sufficient education and training in this domain. In addition, the data converters (analog-to-digital and digital-to-analog) most often reside on the same chip as the digital part, and hence, sufficient design skills for data converters need to be present. Time and cost design including testing is still limiting the development of Analog/Mixed signal ICs even for most digital SoCs. Improvement are still required for Analog/Mixed signal design productivity and predictability.

¹⁷ <https://digital-strategy.ec.europa.eu/en/news/european-commission-launches-study-5g-supply-markets-and-open-ran>

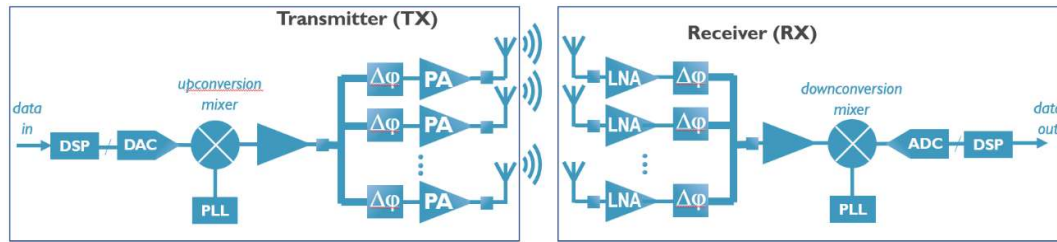


Figure 16: Beamforming in a wireless transceiver (PLL: phase-locked loop, LNA: low-noise amplifier, PA: power amplifier, ADC: analog-to-digital converter, DAC: digital-to-analog converter, DSP: digital signal processing, $\Delta\phi$: phase shifter).

Further, at the strategic side, a quick and broad rollout of 5G over the whole of Europe is advised to set the good initial conditions for the European players in 6G hardware and for the European market to efficiently absorb 6G products in a timely manner.

4.2.1.3. Mid-term hardware needs

4.2.1.3.1. Towards more digitization in transceivers

There is a trend to more digitization in transceivers for base stations. Designing digital transmitters and applying digital receiver techniques require specific training, which could be provided by research centers or universities at the PhD level.

Next to an increase of the complexity of the digital signal processing mentioned above, the transceiver part also becomes more complex due to increased usage of MIMO and beamforming. When a transceiver has many antenna paths, single-chip solutions therefore become less feasible. Instead, multiple chips will be combined. In this way, the packaging level will receive more attention: the different chips will be linked to the antennas and packaged via heterogeneous 2.5D or 3D integration. As several European research groups, among which the large RTOs (imec, CEA-Leti and Fraunhofer), are strong in this domain, a transfer of this know-how to the industry is an opportunity for Europe to play a stronger and critical role at a higher level in the supply chain than solely at the level of chip design and processing.

4.2.1.3.2. RF GaN

In wireless communication, generation of transmit power with a sufficiently high efficiency is always a challenge. Base stations today still use silicon-based technologies (BiCMOS, LDMOS) but it is seen that RF GaN is gaining traction here for example for E-band applications. Europe has various RF GaN foundry players, both at industrial level and at research level. Whereas (RF) GaN on SiC is the most widely used material scheme, certainly for base stations today, GaN on Si is a cheaper alternative that could be used for smaller-cell base stations or in general for applications where GaN on SiC is an overkill in terms of performance. RF GaN devices (HEMTs, high-electron-mobility transistors) have the potential of generating higher output powers than their silicon counterparts although their nonlinear behavior might limit the spectral efficiency of wireless transceivers. A breakthrough of RF GaN therefore should come with efficient predistortion solutions. Know-how on predistortion techniques therefore is a specialty that should be present in universities and research centers to deliver sufficient engineers to the involved industrial players in Europe.

4.2.1.3.3. RF filtering

Another challenge in wireless communication is filtering: especially in the low-GHz range, the spectrum is crowded such that a wireless receiver can pick up interferers and jammers, which need to be suppressed early enough in the receive signal path such that these unwanted signals, which can be several orders of magnitude stronger than the wanted signal, do not drive the active circuitry in the receiver into saturation. RF filtering nowadays makes use of off-chip passive components like surface acoustic wave (SAW) filters, bulk acoustic wave (BAW) filters such as FBAR filters. Such filters typically use piezoelectric materials which are often not suitable for manufacturing in a CMOS-compatible processing environment. Recent breakthroughs in Europe, both at industrial level (see [Soi20]) and at research level (see [PiT20]), provide technologies that enable RF filter design and production on a large scale with the prospect of compact heterogeneous integration with active components. Banking on these results is again a means for Europe to play a role in a field where today Far East and USA players are leading.

4.2.1.3.4. mm-wave beamforming transceivers

With the advent of 5G, the lower part of the mm-wave spectrum is being/will be used for wireless communication both for the consumer market and for wireless backhaul. Indeed, the so-called FR2 band (parts of 24.25-52.6 GHz) has been allocated for 5G communication. At these frequencies, the path loss, which is the loss between the transmitter and the receiver is considerable. This is compensated by the use of beamforming (see Figure 16). At the transmit side, beamforming is a means to reduce the transmit power per power amplifier for a required total transmit power. Therefore, existing silicon-based technologies can be used for many base stations.

4.2.1.3.5. GaN at mm-wave frequencies

Just as for the microwave frequency region, RF GaN is considered as a technology to generate higher transmit powers beyond the capabilities of silicon technologies. However, this requires downscaling of the channel length of GaN HEMTs to achieve the speeds required for mm-wave operation, while reliability still needs to be maintained. Further, the nonlinear behavior of GaN HEMTs, that has already been mentioned above, might again call for predistortion, which is more challenging than in the microwave region as the symbol rate at mm-wave frequencies is higher.

4.2.1.3.6. D-band transceivers

To cope with the insatiable demand for more data, backhaul is challenged to accommodate ever increasing data rates. The adoption of wireless backhaul avoids drastic infrastructure works that are typically needed to install wired backhaul. The high data rates can be obtained by using the bandwidths that are available in the D-band. The combination of wider channels and spectrum efficient methods can be used for ultra-high-capacity backhauling and Fixed Wireless Access (FWA). For base stations that use carrier frequencies in the sub-THz frequency region, beamforming will be used with even more antennas than in base stations operating below 100 GHz. The challenge for the transceiver electronics that will be needed here, is that the power gain of active devices is not high anymore. But this is augmented by the large improvement, by N^2 , in the effective isotropic radiated power (EIRP) in an N-antenna phased-array (although in a real active phased-array implementations, the spatial-combining gains for a larger N are partially offset by on-chip distribution losses). However, several silicon technologies, including CMOS

ones with an f_{MAX} just below 400 GHz, can be used to design most transceiver functions except for the power amplifier (PA). For the latter, HBTs can be used from SiGe-BiCMOS or, for even higher powers, high-mobility materials such as InP that combine a high f_{MAX} with a supply voltage well above 1 Volt.

For SiGe-BiCMOS, continuous improvement of the SiGe HBT f_T and f_{MAX} is fundamental in next generation BiCMOS process platform, to get the needed design margin for product performance robustness and to well exploit the unique feature to integrate together the digital and analog functionalities locally in each active channel of an antenna array.

Concerning InP, Europe has several foundries and research centers with know-how in fabrication of high-mobility III-V devices. However, these technologies today are still niche technologies, making use of small wafers (e.g., 100 mm). Here is a great opportunity for Europe to deploy InP to a level that can serve a mass market: elements to consider are the use of larger wafer sizes and Cu metallization, improvement of yield, reliability, compact modelling, ... just as with GaN, where silicon wafers are already used as a low-cost alternative to SiC wafers, one could also consider 200 mm or 300 mm silicon wafers as a starting material and grow the III-V material onto silicon instead of starting from a native, smaller InP wafer.

4.2.1.4. Long term hardware needs

4.2.1.4.1. Transceivers for IEEE 802.15.3d

On a long term, the G-band will be used for wireless communication at extremely high data rates, theoretically up to 315 Gbps in the IEEE 802.15.3d standard [Pet20]. Research has already published silicon-based transceivers for this frequency band, but these are based on harmonic generation as the operating frequency is very close to or even above f_{MAX} . This leads to very low efficiencies. A more efficient approach would be to rely on fundamental frequency operation of the transistors instead of at harmonics. This is only possible if f_{MAX} is sufficiently high above the carrier frequency, which could be accomplished with InP [Urt16]. Further, co-design of the chips and package that includes the antenna array will be even more critical than in the D-band. Here again, Europe should bank on know-how that is being developed in European research on packaging and heterogeneous integration (see e.g. [Hei21]).

Instead of pushing the speed of silicon or III-V devices, the so-called 2D materials (graphene, hexagonal boron nitride (h-BN), transition metal dichalcogenides (TMDs), silicene and phosphorene) are today an alternative, other device types could be considered such as resonant tunnelling diodes. Further, instead of generating waves electronically, low-cost approaches to efficiently generate waves with photonic components should be considered.

4.2.2. Consumer grade connectivity

User equipment (UE) is by far the largest market for wireless communication devices. UE will evolve from cell phones and smart watches today to devices that incorporate - on top of today's functionality – technologies such as virtual and augmented reality, various sensors to monitor the user and his/her health, ... For the communications part, 6G user equipment will make use of legacy frequency bands from 5G and its predecessors, in addition to several frequency gaps in the microwave and mm-wave frequency range that are not yet addressed but might be freed for 6G. Also, it is expected that for very high data rates, carrier frequencies above 100 GHz will be used in user equipment.

The difference in technical requirements with transceivers for base stations is – apart from a large difference in market size - that the pressure on cost, form factor and power consumption is much stronger. Further, the required transmit power for the uplink is much smaller than for a base station due to lower uplink data rates compared to the aggregated downlink data rate. These boundary conditions have several consequences. First, given the large market size, economics of scale will drive the big IC design players to transceiver design in recent CMOS nodes, combining the analog/RF transceiver functionality with the complex digital modem. For microwave frequencies, where beamforming is not applied, silicon-based transceivers are aided by III-V-based power amplifiers (today, GaAs is widely used). For the lower mm-wave frequencies (28 GHz, 39 GHz) beamforming with a limited number of antennas (< 10) is used in UE. Thanks to this beamforming, the required transmit power per PA is limited and it is in reach of CMOS, such that heavy-duty III-V-based PAs can be avoided, as it is reflected in the first wave of 5G cell phones available at this time of writing. The filtering problem at mm-wave frequencies is less severe than at microwave frequencies due to the spatial filtering by the beamforming and the higher path loss at mm-wave frequencies. For the higher mm-wave frequencies and the D-band, beamforming is also used, but the limited power gain of silicon-based devices might necessitate the use of III-V-based power amplifiers.

4.2.2.1. Short term

The 4G market is a well-established, consolidated market. The role of Europe today here is limited to delivery of a few parts (such as accelerometer and gyroscope sensors from STMicroelectronics). Clearly, this position should be maintained.

Regarding transceiver chips in the microwave region, these have almost become a commodity for sub-6 GHz 4G LTE and are designed in advanced sub-20 nm CMOS generations by Far East and USA companies. For 5G user equipment a similar scenario is about to happen – both for FR1 (410 MHz – 7.125 GHz) and FR2 (24.25 GHz – 52.6 GHz) – by lack of a big European player. Hence, intruding into this market for 5G and beyond looks very difficult on the short and mid-term. Front-end modules for UE constitute the only space where Europe can still play a significant role on the consumer connectivity market leveraging the derivative technologies developed by major European IDMs.

4.2.2.2. Mid-term

4.2.2.2.1. RF filtering at microwave frequencies

Just as for wireless infrastructure (see Section 4.2.1.3.3), RF filtering at microwave frequencies, as part of the front-end module is a domain in which European industry can capitalize on recent results.

4.2.2.2.2. mm-wave transceiver IC design

For frequencies above the FR2 band of 5G and for the sub-THz band, the number of antennas that will be used for beamforming will be so large that the transceiver functionality will be divided over multiple chips such that the drive for single-chip modems is heavily attenuated. Instead, an assembly containing multiple chips and antennas is a realistic scenario. Also, here EU industry could capitalize on the know-how on heterogeneous integration and advanced packaging technology research in Europe. When the analog/RF transceiver functionality no longer resides on the same chip as the digital modem, cheaper technologies than recent CMOS

node technologies can be considered. Several of these technologies from European foundries could be used here. Also, the market of mm-wave transceivers operating above FR2 frequencies is still small and European players could consider stepping into this market.

4.2.2.3. Long term

4.2.2.3.1. Modules for D-band communication

Augmented and virtual reality will ask for wireless data rates of several tens of Gbit per second, requiring large RF bandwidths, which can be made available in the spectrum above 100 GHz. Just as for the mm-wave frequencies below 100 GHz, the transceiver functionality will be split over multiple chips as beamforming will be used. If for the D-band the same footprint for an antenna array in user equipment is used as today for 5G at 28 GHz, then the array for the D-band contains 25 times more antennas. Further, due to the limited power gain and the limited power supply voltage of silicon technologies compared to InP, the best implementation technology for the power amplifiers is InP, yielding considerably higher output powers and efficiencies in the D-band. Here then come some opportunities for Europe:

- InP technology should be brought to a level where circuits like PAs can be fabricated in a cost-effective way, such that the 6G UE market can be served.
- Heterogeneous integration, combining InP with silicon-based technologies as well as antennas and heatsinks in a cost-effective way (including high reliability) is an activity that can back on several research in Europe about InP technology and heterogeneous integration.
- Europe should make sure that sufficient design engineers are trained for the design of the analog/RF transceiver functions.

4.2.2.3.2. Radar functionality with D-band transceivers

D-band transceivers will use channel bandwidths well above 1 GHz. Such large bandwidth can be exploited for high-resolution radar.

4.2.2.3.3. Low data rate wireless devices

With the advent of 6G, a new, potentially large market might appear for devices that wirelessly connect a huge number of various appliances (whitegoods, gardening and household tools, different kinds of sensors, cameras, ingestibles, skin patches, wearables, ...) to user equipment or to small base stations. In many cases, the data rate for information exchange is very low. Also, complexity of the edge computing in many of these devices might be relatively low, such that there is no need to use recent node CMOS. This market, which might become an important one, will not be challenging for the IC technologies. The challenge will be more on the power consumption and the form factor of the module. Hence this is an opportunity that Europe could decide to take with low-power transceiver design and advanced module and packaging development.

4.2.3. Industrial grade connectivity (including connected cars)

We focus here on key connectivity technologies (Low Power Wide Area, Bluetooth, WiFi, Ethernet, 5G) and microcontrollers required by industrial applications (basically industrial IoT, health and automotive). The strong European automotive ecosystem has provided a natural business opportunity for connectivity solutions developed in Europe. This comment can also be applied to other sectors, especially on the industrial market. However, another explanation also

lies in the low cost and low power requirements of industrial applications which then call for the use of mature technologies. Most of the current MCU solutions are manufactured today in 90 nm CMOS with a slow transition to the 40 nm node (and below). Consequently, those markets are perfectly fitted for European IDM semiconductor manufacturers. This could also explain why Europe is having a key role on the MCU market.

Industrial production is a major economic factor in Europe, accounting for around 25% of Europe's Gross Domestic Product. There is enormous innovation potential of IoT technologies when fully adopted not just in the production of physical goods, but in all activities performed by Manufacturing Industries, including pre-production (ideation, design, prototyping, 3D printing) and in the post-production (sales, training, maintenance, recycling) phases. Therefore, one of the major objectives of several European initiatives (e.g., Industry 4.0, Smart Factories) is to bring IoT paradigms to industry, production, and logistics [Ver15] demanding, amongst others, industrial grade connectivity.

From the end user perspective, i.e., production facilities taking up advanced connectivity solutions, this implies the availability of solutions that are tailored to meet demanding industrial requirements such as bounded latency and reliability, rather than mere capacity. This not only implies the availability of chipsets, but the incorporation (or even customization) of such chipsets into industrial systems and their smooth and gradual integration into existing production environments (with a plethora of other standards and protocols), step-by-step and in a modular way. From a technology perspective, this opens opportunities to innovate across the entire value chain, from the smallest components needed to enable such connectivity up to the realization and integration of complete end-to-end solutions, possibly combining different communication systems. Further, the integration of communications and sensing may transform the way we perceive industrial grade connectivity. In this aspect, devices will sense, then locally process to some extent the sensed information and communicate.

4.2.3.1. Short term

Europe has a clear leadership and can take advantage of a sizeable indigenous market thanks to strong industrial and automotive actors in Europe. Integration is also happening here, the capacity to design highly integrated solutions using more advanced nodes will have to be secured to maintain leadership. IoT will be a key driver for this market, but without any leadership on the cloud side Europe may capture a very limited part of the global value.

Compared to wireless transceivers for UE or RANs or transceivers in datacenters, transceivers operating in an industrial environment often need to function in harsh environments. Key requirements for industrial communication are reliability, bounded latency, robustness, etc. rather than capacity. 5G aims to target these requirements under the umbrella of URLLC, but the entire feature set (slicing, integration with Time-Sensitive Networking, etc.) will only come with R17, targeted for standardization in 2022 with devices 1-1.5 years later. This gives a window of opportunity of several years to focus more strongly on the design and realization of industry-graded 5G chips and complete end-to-end systems, including 5G-enabled industrial components that differ significantly from consumer devices. The more control over the lower-level building blocks and the less outsourcing outside Europe, the more room for innovation and end-to-end solutions that stand out.

Along with this, more uniformity in spectrum availability must be strived for to avoid fragmented markets. Further steps must be taken to step away from spectrum that is reserved at a large geographical scale and enabling spectrum allocations for local private industrial networks, as a stimulus for industrial competitiveness.

Further, as cellular technologies come with their own peculiarities (e.g., cost, complexity, etc.), the opportunities of the newly available unlicensed 6 GHz spectrum (or others for e.g. industrial use: e.g., 3.7 – 3.8 GHz in Germany for campus networks) must not be ignored. With its large bandwidth, interesting propagation properties and anticipated Wi-Fi 6E chips, its promotion should lead to parallel high-impact innovations for industrial connectivity that stand closer to current industrial communication systems, can also serve many indoor and short-range scenarios and a lower environmental footprint. To facilitate transitions in connectivity, technology interoperability must be given attention.

Accessible testbeds and trials are required to bring concepts to reality and validate whether solutions that have been driven by standardization in closed bodies dominated by limited big stakeholders can actually live up to the requirements of players within the industrial ecosystems, including the many SMEs in Europe.

4.2.3.2. Mid-term

On the mid-term, additional steps must be taken for Europe gaining more control over industrial connectivity solutions and systems. Within the context of 5G URLCC, openness in terms of O-RAN is simply not sufficient. Adopters of the technology are still bound to the interfaces provided by component providers, depend on the willingness of these providers to get more access or richer APIs, and are restricted with respect to end device innovations. Standardized open interfaces solutions can avoid such lock in situations compared to proprietary solutions.

To obtain trust in components and enable customization in a market that cannot have a one-size-fits all solution, approaches such as O-RAN must be further pushed down to the individual components, end devices and even chips. In January 2020, the possibility of the open-source movement entering and threatening the chip industry was discussed in [Kin20]. From economic reasons, economy of scale is necessary. This can only be achieved by standardized solutions and to ensure interoperability between components from different vendors.

Apart from potentially disrupting the chip industry, more open systems also have a role to play in educating students and could help to realize a better interplay between standardization and running code/HW. Due to the huge outsourcing efforts during the past decades, lots of expertise has left Europe and needs to be re-established to get back control on end-to-end solutions. Hence, efforts at the educational level must be foreseen to appropriately train engineering students, giving them more cross-disciplinary and applied hands-on training skills besides theoretical & high-level python design skills. By filling the gap between great analog design skills at one end and strong high-level software skills on the other hand, i.e., addressing real-time embedded and digital processing skills, engineers are formed that master the skills to understand complete systems.

In parallel to 5G's evolution in serving industrial communication, one should stimulate parallel innovations in communication, performed in collaboration with large industrial players within EU, and taking advantage of new unlicensed spectrum such as 6 GHz or local licensed spectrum. For instance, in [Pan17] the concept of Wireless High-Performance Communications was raised

for building tailored industrial solutions, not built on general-purpose chips and having significantly lower architectural complexity.

Naturally, the quest to achieve ever higher data rates, but combined with industry-grade robustness, must be pursued for advanced, bandwidth-hungry industrial use cases. This requires components and solutions operating in the higher frequency bands (mm-wave and higher).

4.2.3.3. Long term

Push new radio architectures (like smart surfaces and cell-free massive MIMO) from the conceptual level to validated and affordable solutions beyond theoretical studies and simulations, considering hardware limitation and integration complexities (HW-SW integration, front/back/mid hauling). This requires a lot of technology innovation across multiple disciplines, such as meta-materials, programmable intelligent surfaces and electromagnetics, on the one side and information theorists and signal processing experts on the other.

The higher frequency bands with unfavorable propagation conditions may support the bandwidth-hungry applications. Still more efficient spectrum usage and spectrum reuse in lower frequency bands with excellent propagation properties can also be considered by using smart radio architectures. Cell-free massive MIMO, for example is capable to concentrate the wireless signals by minimizing interference and maximizing spectral reuse.

Sustainable radio communication infrastructure: reducing overall energy consumption per useful bit for data traffic (not only in end devices, but also in access, core and transport networks, (edge) clouds, also including energy consumption in control plane and traffic due to signaling and protocol overhead)

4.2.4. Datacenters

The quest for more data is not saturating in the foreseeable future. This will result in a growth of data centers, of the number of the data centers and of the baud rate that the front-ends will have to process. Today, commercially available optical transceivers used inside datacenters obtain speeds around 56-75 Gbaud per lane, but next generations will require 120...180 Gbaud and even more. Roadmapping in this domain must consider that the market of datacenters is smaller than the automotive market and certainly compared to the mobile handset market. However, DC market is largely dominated by webscale companies (e.g., Facebook, Amazon, Microsoft, Google, and Apple) that even implement their own equipment. The short-term challenge will be in increasing data traffic to 51.2 Tb/s while drastically limiting power consumption thanks to Co-Packaged Optics (CPO), a technology that co-integrates optics and ASICs. Microsoft and Facebook have launched the CPO initiative in 2019 to enable the development of common design elements that will guide technology vendors (like Broadcom). Electrical and photonic IC design and manufacturing activities in Europe for this market have been reduced over the years, one of the reasons being the limited market size. However, Europe still holds several low-volume open platforms both in industrial and research labs (imec, CEA-Leti, IHP, FhG, LIGENEC, VTT, Smart photonics, HHI, Jeppix). Moreover, next to optical communication, photonics is present in many other domains. A large growth in these domains could boost the growth of the photonics industry, from which the optical communication market could benefit. An example of such potential 'trigger application' is the photonics-based glucose sensor in the Apple Watch or the Lidar technology for the autonomous cars.

4.2.4.1. Short term

Scaling of optical transceivers to higher frequencies comes with extra complexity. Product prototypes using sub-10 nm CMOS have been demonstrated by several companies, many outside Europe. Few counterexamples exist such as Nokia who is designing its own 7nm DSP for 90GBaud/400G and 600G coherent systems. From there we see a trend towards more digitization, requiring very high-speed data converters. It shows to be challenging for European companies to increase their market share; creating design IP in such scaled technologies implies to compete with large established players in the field.

4.2.4.2. Mid-term

On the mid-term, the saturation of the speed of CMOS devices, as discussed in the introduction of Section 4.2, can become a showstopper to extend a further increase of the baud rate per lane in future transceivers for application beyond 1Tb/s (>200 Gbaud). Here is an opportunity for non-CMOS technologies that feature devices that are faster than CMOS ones, such as BiCMOS and InP, which are already present in Europe. InP technology should, however, be upscaled to a higher level of maturity (cost, throughput, yield). Worth mentioning is the new architecture design of very high-speed data converters (>100 GS/s) conducted by leading research teams in Europe (imec, University of Stuttgart and Saarland, III-V Lab, etc.) on analog multiplexing and demultiplexing that could offer new opportunities to further extend transceivers speed.

At a higher level in the supply chain, the packaging of transceiver modules will become more complex as the increase of the data throughput not only comes from an increase of the baud rate per lane but also in an increase of the number of wavelengths, requiring more lasers. Photonics packaging is an activity already present in European research centers (Tyndall, imec, ...) and a transfer of this know-how to industry could be a means to cultivate European players in this domain.

4.2.4.3. Long term

The challenge to solve for the longer term is how to achieve > 100 Gbaud in a cost-effective, densely integrated and energy-efficient way with the constraint of speed limitations and/or high losses of current electronics and optics. What are cost-effective and industrially scalable material systems offering O/E (optical-to-electronics) and E/O (electronics-to-optical) bandwidths well above 100 GHz, while simultaneously offering monolithic integration of passive components such as wavelength MUX/DeMUX, polarization handling, fiber coupling? What is a cost-effective and industrially scalable method to integrate the laser? InP might be a suitable material here, again implying the need for upscaling InP technology, not only for transistors, as mentioned in Section 4.2.2.3.1, but also for optical components like lasers. Further, new materials could be investigated to address these questions.

4.3. EG3: sense and power

In addition to computing and communications core technologies, future networks require the development of supplementary core technologies that support the design of computing components (EG1) and communications components (EG2) and are essential for controlling the telecommunication and vertical value chain. The goal of Expert Group 3 (EG3) is to identify key challenges for innovations and sovereignty in the areas of sensing and power for future network. This will encompass the following key domains to focus upon:

- Sensor Processing in 5G and beyond
- Power Management
- Core Process Technologies
- System and Component Architectures

This expert group explores the challenges and opportunities that are going to confront not only the telecommunications and microelectronics industry, but the whole value chain and society. Overall, EG3 aims to define the overall European strengths for maintaining and developing our strongholds in the domain of sensing and power. It targets to address selected weak points that are critical yet can be realistically achieved.

4.3.1. Sensor Processing in 5G and beyond

4.3.1.1. Introduction to Connected Sensors

Connected sensors will be key for Europe to achieve its industrial, social and environmental goals. The massive deployment of such connected sensors will be made possible by the wide installations of 5G and beyond Telecom Infrastructure. Sensors are at the heart of Enterprise digital transformation and Industry 4.0. They are required to contribute to European Industrial leadership, increase the safety on the roads, optimize use of energies, and to enjoy a better air quality of our environment and in public or commercial buildings. This pervasion of connected sensors will also contribute to improve our safety as well as environment at home and at work, our health and to achieve transparency in the logistic of our food and medicines.

With billions of sensors connected, tendency rising, there will be numerous challenges to be addressed, ranging from RF bandwidth availability, optimization of energy usage, reliability and maintenance, security of transmitted data, as well as utilization and analysis of the data collected from the sensors. This chapter is going to explore the challenges and opportunities, that are going to confront not only the telecommunications and microelectronics industry, but the whole value chain and society.

4.3.1.2. System and Architecture for Sensor Processing

Sensors can vary from simple ones using MEMS for accelerometers, pressure, microphones for noise recognition, or time-of-flight sensors based on ultra-sound or light. More complex 2D,3D, 4D sensors can be using Lidar, radar, CMOS Image sensors, thermal imaging, etc., or a combination of several sensors. Therefore, the amount of AI and processing required greatly depends on the type of sensor and nature of the data necessary for the application. Discrete sensors such as the ones used in metering or agriculture require to be as autonomous as possible. They can be connected to cellular LPWAN infrastructure and will require very simple embedded processing. A key challenge is to provide energy autonomy to such sensor. This is

why they are associated to extremely low power MCUs and can greatly benefit from energy harvesting techniques.

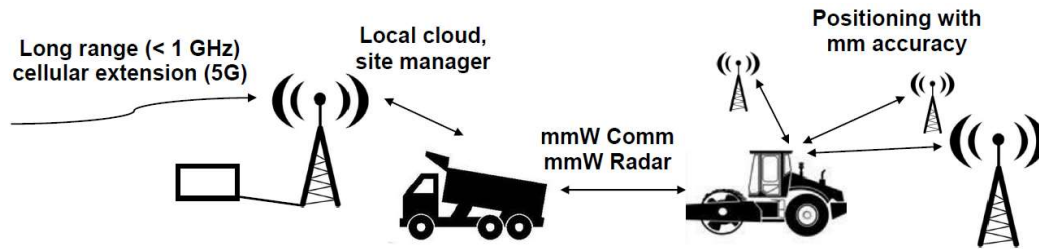


Figure 17: Sensor and network application within the 5G network in agriculture and construction sector. Source: AMMCOA project, BMBF Germany.

Sensors embedded in cars (e.g. radars) or industrial machine vision sensors (Figure 17) will generate data that will transit to local or remote servers where they will be processed in real time for immediate action or can be stored for further analytics. This creates important challenges related to the data itself:

- Usage of crowded bandwidth to move data from sensor to cloud;
- Power dissipated by the wireless transmissions and receptions;
- Security of the data during transmission time.

This raises the question about the reduction of the data to transmit, limited to the minimal useful set. Else, it also legitimates to properly balance where the processing will be made at the sensor node (deep edge), in a central server located on the sensor network (edge) or in the cloud. Processing or pre-processing at the edge is a way to reduce the amount of data transit, thus limiting RF bandwidth, power consumption, and security of data.

Edge and Cloud Computing

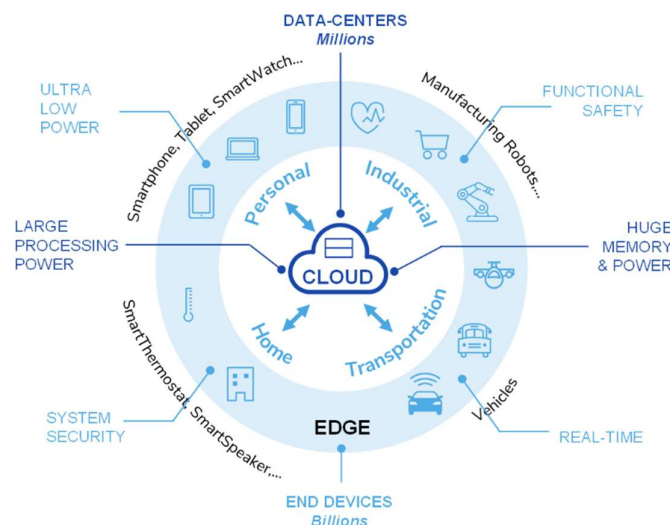


Figure 18: Data collection, processing, and decisions taken at the edge, and edge devices securely connecting to the cloud.

It is expected that the number of connected devices will grow to 125 billion mainly thanks to the 5G and future 6G deployment. As of today, a typical device senses its environment and transmits

to the cloud a large amount of data that are computed in the cloud where decisions are taken and transmitted back to the device (Figure 18). This process will not be sustainable with such a high number of devices. Moreover, many applications are constrained by latency or privacy which are not consistent with heavy use of radio transmission. Therefore, it is highly desirable to embed enough computational capacity in the sensor node to take decision.

Most applications are based on classification of events so that it will be necessary to develop power efficient AI accelerators and processors. Here, we refer to “Edge” as enabling technologies allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services and upstream data on behalf of IoT services. The rationale of edge computing is that computing should happen at the proximity of data sources as described in [Shi16]. As of today, Edge AI is limited to inferences while training task is left to the cloud. Edge AI processor units are shared between CPU, GPU, DSP and Deep Learning accelerators. CPUs afford the maximum flexibility at the expense of the lowest efficiency at the other side of the ladder, accelerators afford the maximum efficiency but are application specific. For computation and power efficiency purposes, fixed point DSP with bit-width of 8 bits or lower are chosen [Lee18]. As of today, the AI approach is limited to edge inferences. Nevertheless, edge training will be necessary to bring adaptability to sensor nodes. Furthermore, the current tools to implement and train neural networks on microcontrollers are provided by non-European companies. Therefore, to increase European security and sovereignty, the development of European tools in this field are needed to train and program sensor nodes processors. With the drastic increase of data traffic as a result of all the new possibilities that are given to the industry and the consumer with 5G deployment, there will be several challenges. Edge computing with processing of information at the edge before moving it to the core processing unit or the cloud can become the key to handle the data mass. So, the motivations for edge computing are the **reduction of data traffic, the reduction of power consumption and cost reduction and the increase of security.**

Edge computing (deep edge) requires MCU processing capabilities aligned with application requirements. The advantages of processing at the sensor node are:

- lower power consumption and cost reductions thanks to reduced data transmission,
- increased security of the data,
- enablement of sensor fusion,
- better MCU capabilities with embedded AI,
- good for ultra-low latency applications required in Industry 4.0.

Today sensors network architectures are very centralized: sensors transmit their data to a single storage & presentation platform, regardless of their actual needs, which then act as a proxy. Distributing this process over the network, with several instance located at different layers / areas of the network will likely help to reduce unneeded data transfers while saving data integrity.

Its goal is to move from big data handling to efficient data handling. Thereby, pre-processing of information to hide part of the information to the upper layer or to normalize data between different network elements will be of great benefit to reduce the data traffic.

There, protocol overhead can be high, especially for sensors transmitting rather small packet. In this area, different optimization, from header compression to the specification of new protocol for small packet are already on the way and should start to create benefit within the next

upcoming years. Also, ultrawideband sensor networks are gaining importance, since it allows to achieve two contradictory goals: fast establishment of links between network nodes, and realization of energy saving modes as described in [Dmi18].

An alternative of today's database architecture could be stream processing, which is used to query continuous data stream and detect conditions, quickly, within a small period from the time of receiving the data. The detection time varies from few milliseconds to minutes.

Sensor Fusion

In sensor fusion, several sensors are part of the same system. Each sensor sends its data, eventually pre-processed in real time, to a central processor or server that performs the analytics. This central processor can be on the same chip as the sensors or remote. An example of areas where sensor fusion is applied is in Advanced Driver Assistance Systems (ADAS), for which cameras, sonars, radars, and lidars each send data of the surrounding environment simultaneously to a sensor fusion processor that processes the data in real time. The advantage of sensor fusion is to acquire robust data through usage of multiple sensors that provide redundant or complementary information simultaneously.

Starting with ADAS level 3 and to achieve full vehicle autonomy, the latency to acquire, transmit, process the information, and send an order to control the vehicle needs to be extremely low and requires real time processing of complex 3D data. To complement the data sent by on-board sensors, real-time map-based localization stored in the cloud will also be accessed to add to the robustness of the overall solution through V2X or other communication standards. Such sensor fusion is very complex both in terms of design architecture, AI and software. Low latency computing and high number of operations require the most advanced CMOS process nodes. But sensor fusion is not limited to ADAS or autonomous vehicles and applies to other applications like medicals in which the data of several sensors can be combined to assess the condition of a person. In that case, constraints on processing and latencies are much more relaxed than for cars. AI inference from an MCU located in the sensor can be sufficient to detect abnormalities and send data to the cloud for further analytics.

Privacy and Security

Regarding the increasing demand for security and data protection, mobile ad hoc networks (MANET) in all forms with increasing privacy sensitivity could become a competitor for the cloud. They consist of a set of mobile nodes connected wirelessly in a self-configured, self-healing network without having a fixed infrastructure. MANET nodes are free to move randomly as the network topology changes frequently. Each node behaves as a router as they forward traffic to other specified nodes in the network. Downsides are, that their resources are yet limited due to various constraints like noise, interference conditions, etc. Also, a lack of authorization facilities as well as the vulnerability due to limited physical security still form a problem. On the contrary, with deployment of MANET, following benefits can be achieved:

- Separation from central network administration.
- Each node can play both the roles of router and host showing autonomous nature.
- Self-configuring and self-healing nodes do not require human intervention [Faz16].

Data reliability over a (wireless) network

Reliability is essential for enabling safe and secure communications. However, providing reliability in the communication implies adding overhead to the transmitted data, which

conflicts with the low latency requirements in several industrial use cases, like machine remote control, safety assurance, discrete processes control, and many other applications of sensor networks in different verticals.

The low-level retransmission procedure – HARQ (Hybrid Automatic Repeat reQuest) is an instance of a widely deployed reliability mechanism in LTE networks, but it introduces latencies up to 8 ms, so cannot guarantee latency < 1ms, as demanded in a variety of industrial applications. In the same context, LTE profiles like Narrowband IoT have included enhanced robustness (e.g., using double HARQ) for monitoring applications not requiring time critical response like lighting control, smart metering or sensing in agriculture. Research initiatives funded by the German government, like HiFlecs or ParSec, have proposed low-latency and high-reliability communication solutions that might fulfill those requirements in mobile robotics use cases, but power consumption becomes a limitation regarding wireless sensor networks.

Today, two techniques are used to increase reliability, which are either retransmission of the entire packet of information, or redundancy, where some coding is added. But with the increase of connected sensor devices, new challenges for its reliability arise. Therefore, the gaps are:

- to evaluate sensor data reliability with a confidence level so as to identify a defective device into a subset of devices.
- to track performance thanks to rapid monitoring and feedback systems, even allowing remediation actions to enable self-healing wherever possible (e.g., software updates).

4.3.1.3. Hardware for Sensor Processing

FPGAs, ASICs

As state of the art, the electrical ASIC and FPGA components are electrical, but technological and societal needs exceed the forecast of Moore's law of technological advancements. By overcoming Moore's law, faster processing and less power consumption can be achieved. Also, the long-term trend to move hardwired implementation into more programmable, flexible and evolutive component started decades ago but reaches a limit. Thus, it turns out that generic CPU will not be able to efficiently handle the compute heavy tasks and the very high volume of data to proceed that are typically associated for example with Layer 1 / 2 and sometime 3 of the OSI network stack. To mitigate this issue while saving the benefit of software defined function (flexibility, evolution capability), the envisioned solution is to dynamically implement the required specialized processing in a dedicated programmable hardware that is typically an FPGA. For obvious physical reasons, it will not be as efficient as a dedicated ASIC but will be able to greatly reduce CPU load and power consumption. Implementing FPGA acceleration into a software process running on generic CPUs is currently feasible and requires specific expertise for the accelerator design and to manage FPGA configuration. The remaining challenges will be:

- To seamlessly integrate the use of FPGA resources in a code flow. At first with generic / standardized FPGA IP Library and a global framework to help to its integration. At a longer term, we can imagine that a compiler could become able to identify by itself what part of the code need to be executed on the FPGA and on the CPU and generate the appropriate configuration file.
- To lower the cost of high-performance FPGA, as is can remain a showstopper for some application.
- To shift towards optical FPGA and ASIC components, which is particularly suited for artificial intelligence where computing is based on analog signal processing.

- To carry analog operations with the signals to be processed by optical interference. Long-term goals are the evolvement of optical ASIC and FPGA components, where the integration density is improved.

It should be noted that actors of the CPU industry are getting closer and closer: Altera is an intel subsidiary since a few years and AMD has just concluded the acquisition of Xilinx. This will likely help AMD and Intel to integrate more deeply FPGA based accelerator with their more classical CPU and give them a significant advantage to support heavy processing tasks usually associated with AI, Machine learning & Network virtualization.

On the other side of the Atlantic Ocean, Europe is more or less out of the FPGA field. For years, Xilinx and Altera have dominated the market and currently own 90% of it. The remaining 10% is shared between several smaller manufacturers who have developed specific FPGA products to address specific needs like ultra-low power FPGA (Lattice) or Space and Defense (MicroSemi). All of them are based in the United States. In addition, the main editors (Cadence, Synopsys) of software tools used for FPGA design are also largely from north America. The only exception is Mentor Graphic who still mainly operate from US soil but get acquired by Siemens in 2016.

Consequently, if a European actor on FPGA should be developed, it will have only very limited existing ground to build on.

Neuromorphic computing

In neuromorphic computing, electronic circuits mimic the operation of neurons and their architectures to enhance the computational efficiency and reduce the power consumption. The circuits can be digital, analog or mixed, and combined with memory cells. An alternative way is to exploit high performance computing (HPC) with algorithms that emulate, for example, human information processing. Typically, the focus is on specific problems such as pattern recognition like image, speech, text, handwriting, as well as data mining. For example, text recognition at 20 pages per second has been demonstrated by a 500 TFLOPS HPC system.

Neuromorphic circuits use non-von Neumann architecture and can be realized by integrating a non-volatile memory cell as the synapse with CMOS circuitry representing the neuron. Recently photonic synapses based on phase-change materials were also demonstrated. The circuits can exhibit machine learning to perform, e.g., pattern recognition. Supervised learning is based on repetition of numerous samples, and unsupervised learning needs a few presentations to master the function. The latter is in many cases based on spike timing dependent plasticity (STDP) of the artificial synapses. The hardware realizations make use of combining high-end VLSI CMOS technology and multistate probabilistic memristor synapse systems to create efficient neuromorphic architectures. The unsupervised neural circuits can also be connected to living entities and, in the future, help to construct more efficient, partially artificial, implants. The human brain consumes power in the range of 10-20 W, which is orders of magnitudes less than the power consumption projected for an equivalent system using current technology. However, sub-fJ operation per bit has been predicted in phase change material synapses integrated with carbon nanotubes, showing that there is room to improve the energy efficiency.

An example of self-learning neuromorphic chip, based on OxRAM technology, having the ability to compose music by learning the rules of musical composition by detecting patterns in the songs, has also recently been presented. On the other hand, large European research projects with implications for neuromorphic engineering are NeurONN (www.neuronn.eu/) or the

Human Brain Project (www.humanbrainproject.eu/). Their attempt is to simulate a complete human brain in a supercomputer using biological data and to use that knowledge to build new computing technologies. The EU-funded TEMPO project (<https://tempo-ecsel.eu/>) uses emerging storage technology to design new innovative technological solutions. The neuromorphic algorithms of reduced-core computer operating systems will serve as demonstration models. These neuromorphic technologies could have important medium- and long-term applications.

4.3.1.4. Application

Radar

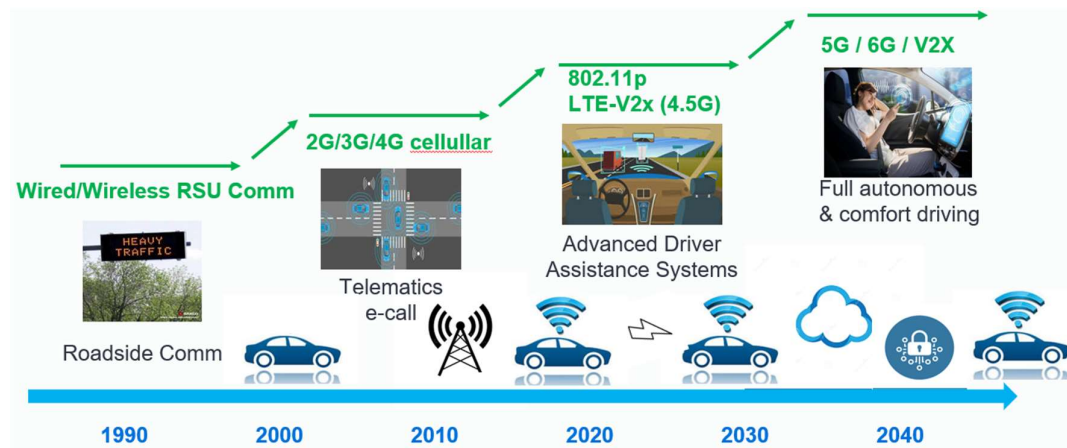


Figure 19: The evolution of autonomous driving with each communication technology.

Radar is one of the most important sensors of application domains like autonomous driving and industrial IoT, which will have a break-through with the evolving communication technologies 5G, 6G, and beyond (Figure 19). Investigations of 3D Radar with 250 virtual antennas, which currently meet a Time of Flight (ToF) of 30k Pixel, will have to evolve to a 4D radar within the next few years to guarantee safety and reliability for the vehicle on the road. The radar technology needs to be able to determine the location of an object in terms of range, azimuth, altitude, and relative velocity to provide detailed information about the vehicle environment. Therefore, the sensors will need to have the capability to detect an immense variety of hazardous situations, from traffic jams in approaching tunnels to cyclists or autonomous emergency braking for crossing traffic at intersections.

While the Integration of Radar and Communication (RADCOM) currently only exists as an idea with a specific application, it is crucial in a variety of future domains. In a concept of Satellites as a Service there is interest to have payloads offering 5G/6G Connectivity and EO SAR to optimize resources aboard and greener technology. Therefore, having the RADCOM technology validated and demonstrated in relevant environments (TRL 5-6) is a short-term challenge. For 2030, a complete and qualified system which is proven in operational environment (TRL 8-9) should be developed to guarantee the functionality of 6G applications.

User Interface

Haptic communication and control: Human/robotic communication information such as audio (hearing) and visual (sight) or a combination thereof audiovisual are transferred over

communication networks. Additionally, interacting sense of touch (haptic) and particularly the kinaesthetic (muscular/robot arm movement) component has much stricter end-to-end latency communication requirements between tactile ends [Aij18]. To enable bi-directional haptic control, indeed follow the widely accepted understanding that edge computing is a key driver behind Tactile Internet aiming to bring control and user plane services closer to where they are needed.

Typically, haptic information is composed of two distinct types of feedback: kinaesthetic feedback (providing information of force, torque, position, velocity, etc.) and tactile feedback (providing information of surface texture, friction, etc.). The former is perceived by the muscles, joints, and tendons of the body. The tactile feedback should not be confused with the Tactile Internet (TI) [Aij18] whereas the latter is consumed by the mechanoreceptors of the human skin. While the exchange of kinaesthetic information closes a global control loop with stringent latency constraints, this is typically not the case with the delivery of tactile impressions. In case of non-haptic control, the feedback is audio/visual and there is no notion of a closed control loop. In addition to enabling haptic/non-haptic control/data, the TI aims to enable an overlay low latency network platform, providing interoperability over different communication technologies, e.g., over 5G Ultra-Reliable Low-Latency Communication (URLLC), Time Sensitive Networking (TSN), etc.

The nature of human brain response time to light and audio makes the use of haptic feedback to alert faster and more reliable. Human brain response to the sense of touch in range of 1 ms where the response to audio and video is in 100s of milliseconds which makes it more responsive for alerts and/or navigation.

Application AR/MR/VR

AR cloud as a 3D digital copy of the physical world, an AR cloud will connect billions of sensors that will constantly and in real time feed that digital twin of our world.

Extremely low latency will be paramount to transfer the data and enable a seamless augmented and mixed reality experience. Transferring the minimum set of data (ex. Point cloud plus geo-localization) will require AI on the sensor, including machine learning (tinyML). One big advantage for the most intensive processing is to occur in the cloud, thus reducing the cost and processing power on the client devices. Power consumption from the data transmissions is minimized since only the relevant data is sent from the cloud to the end device. An example of such a cloud is the Road Experience Management (REM) by MobileEye, which will allow to maintain in real-time a replicate of all the roads in the world, by gathering sensor data from cars equipped with the MobileEye system.

Use cases for AR/MR/VR devices are: Augmented, Mixed and virtual reality are key to improve the productivity of our workers by supporting training, products assembly and quality insurance or providing effective remote support to customers. Health and retail are other sectors of applicability. The sensors embedded on AR/MR devices will be used both to enrich the AR cloud and to allow interactions of the users within the XR environment.

4.3.2. Power Management

4.3.2.1. Wireless Energy Transfer

To improve wireless energy transfer, low-cost and energy efficient approaches for beam re-focusing based on reconfigurable meta-surfaces in mm-wave and THz bands will be used to increase coverage of future 6G networks. Additionally, to further increase the number of sensors, novel mm-wave battery-less and chip-less IoT solutions will be considered to perform measurements and send back information using energy generated by interrogating mm-wave base stations able to focus antenna beam (Power over Air – PoA).

4.3.2.2. Heterogeneous energy sources & transfer

Large deployment of 5G network implies an increase in power demand for the network equipment. This increase can have severe consequences on the generation of the electrical power as well as on the power distribution grid with negative impact on the environment. National and international policies drive a shift from fossil fuel to renewable energy sources. Most of these renewable sources have intermittent behavior and availability: for example, solar energy is available only during daytime, wind and tide energy depend on environmental variability. On the other side, radio network must be operative at all times. For this reason, it is necessary to introduce algorithms, models, and converters to optimally exploit these heterogeneous sources and, in combination with suitable energy storage, make power available as necessary.

On the distribution side, integration with energy storage elements is fundamental to guarantee stability of the power grid. In this case AI-based scheduling algorithms can be employed to optimize exploitation of the different power sources based on availability and demand and cope with local failures or critical conditions. Multi-input, high-efficiency and high-density converters need to be developed for the user side, to allow the user equipment to seamlessly switch from one source to the other.

4.3.2.3. Energy Harvesting Systems

As the communicating systems market is booming, the role of energy harvesting (EH) will be growing. Indeed, the number of connected devices is planned to increase by a huge factor of 200. Connected devices are going to be used more and more in several fields such as healthcare, wearable, home automation, etc. The Internet of Things (IoT) market grows considerably leading also to the boom of the connected devices, and so highlighting the importance of energy needed to supply them in view of the limitations of current battery technology. In this particular case, we are focusing on small, connected devices with low power consumption below a few mW (or even a few tens of μ W). The objective here is that local energy harvesting will substitute battery-powered devices and eliminate the high demand in energy for battery manufacturing and distribution logistics. Therefore, self-powering systems for small IoT nodes must be developed.

Different wasted energy sources can be exploited and converted into electricity: sun or artificial light, heat, RF power (either intentionally transferred), mechanical movements and vibrations, etc. Moreover, this converted energy needs to be used and transferred wisely to sensors, microcontrollers or other electronic components included in the system. Thus, power management circuits and energy storage devices also become an essential element. In this roadmap report, we have assessed several promising technologies for EH including photovoltaic

cells for outdoor/indoor light EH, thermal energy harvesting, mechanical EH based on three concepts: piezoelectric materials, electrostatic and electromagnetic energy conversion, and RF energy harvesting/wireless power transfer. Targeting EH technologies with low fabrication cost, with high efficiency, and without toxic/rare materials is the main challenge. Adding flexibility and/or transparency is also an increasing demand for compatibility with wearables applications.

For the semiconductor companies, the interest is also to develop new devices compatible with Silicon technologies. The fabrication of components dedicated to energy harvesting and in particular to thermal energy is of high interest as no solution based on silicon technologies is available for implementation as of today.

Mechanical EH rely typically on input vibrations, and one of the main challenges is the compatibility with low frequency vibration source (most of applications use frequency vibrations < 100 Hz) and the increase of the operational frequency bandwidth. Although photovoltaic EH technology is mature (silicon based) for both outdoor/indoor applications, emerging materials are promising for their potential to add flexibility and low weight (thin films) at reduced cost and high efficiency (organics, dye sensitized, perovskites, etc.). RF harvesting systems (referred as rectennas, rectifying antennas) have been widely used and studied in recent years due to the vast presence of RF sources in humanized environments. However, too low power densities have usually been experienced because of RF regulations and RF power transfer “on demand” solutions are currently preferred and are exploited in a large number of passive RF system applications, such as RFID, wearable or implantable devices, realized using eco-compatible materials.

4.3.2.4. Energy-efficient Components and Communication

It will be mandatory to afford solutions to limit the power consumption of sensor nodes as their number will increase exponentially in the next decade. With a sensor node power consumption of about 1 W when in on state, the main solution is to put the sensor node in a sleep mode as long as possible. Therefore, it will be necessary to provide ultra-low power wake-up circuits as Real Time Clock, wake-up radio and wake-up sensors. From these techniques, the wake-up sensor is the technique that permits the longest sleeping time with no information loss contrary to RTC or WU radios techniques. A wake-up sensor embeds a feature extractor and a low complexity low power classifier and thus could be seen as an ULP Edge AI sensor device. As of today, the power consumption of industrial wake-up sensors lies in the $10\mu\text{W}$ to $100\mu\text{W}$ range [Lee18], [Rum20], while academic solution ranges between 1 and $6\mu\text{W}$ [Tou20], [Yan18]. $1\mu\text{W}$ industrial solution should be targeted in 2025 and sub μW solution in 2030.

To further improve energy efficient methods, low-cost and energy efficient approaches to beam steering will be developed to unleash beamsteering potential not only in base stations but also on the edge, especially as small inexpensive and energy efficient IoT sensors are considered. To this end new pattern reconfigurable antenna concepts will be developed, in which variable loads attached to passive beam-shaping elements will be used together with energy-efficient fast FET switches (e.g., in FDSOI technology).

4.3.3. Core Process Technologies

4.3.3.1. Heterogeneous Integration

In the past decades the advancement of microelectronics was mainly driven by “Moore’s Law” which is based on the scaling of CMOS technologies. Currently this trend seems to saturate as feature sizes of modern devices are approaching the range of atomic scales. Modern communication systems not only rely on an advanced CMOS node, but they also rely on the performance and physical principles of other semiconductor technologies such as III/V semiconductors for laser sources in optical communication systems. Rather than monolithically integrating this functionality into a silicon-based semiconductor which is a great effort and only feasible in few cases, heterogeneous integration is a promising way to realize next generations highly integrated leading-edge systems at low cost and attractive form factor.

The key to heterogeneous integration is a reliable combination of various semiconductor materials and metal systems in a 2.5D / 3D way. To achieve this goal, significant R&D effort on various technological fields has to be done. On one hand, next generation interposer technologies must further reduce pitch and bump size to allow an ultra-high routing complexity, but on the other hand, a significantly improved RF performance for possible applications in the Sub-THz range is required. 3D integration requires the further advancement of Through Silicon Vias (TSVs) that are used for RF and mm-wave interconnections and bump-less chip-to-wafer or wafer-to wafer bonding capabilities.

Another promising approach which still requires extensive R&D work for heterogeneous integration is in the field of active layer transfer and printing. As an example, this technology can be used to locally attach a III/V semiconductor layer stack into a Si-based environment which is patterned during the succeeding standard CMOS process. Using this approach, III/V devices such as laser sources or high-performance (high-speed or high power or very low noise) transistors can be realized boosting the overall systems performance.

Heterogeneous integration cannot be regarded from the technological point of view only. As several semiconductor and other technologies are combined within one closely interacting system a seamless design methodology beyond technology borders needs to be available. This also includes the thermal engineering and thermo-mechanical properties of the system. In the past, quite some effort has been done in the field of heterogeneous system design methodology but as these systems advance significant and continuous R&D effort in this field is required.

4.3.3.2. Low-power, no-charge based memory

The stand-alone market is currently dominated and driven by 3D NAND-Flash. This seems to be the main path on the roadmap for the next > 5 years, with very limited global actors sharing market quotas (Samsung and Intel/Micron). It is more fruitful to focus on emerging non-volatile memories (eNVM) that can become the next product for specific applications, i.e., SCM/Storage Class Memory and embedded memories for IoT, automotive, low power markets. Also, neuromorphic applications are seen as promising on a longer time scale as new potential applications are arising. ENVM are substantially represented by the non-charge-based technologies, PCRAM, OxRAM, CBRAM, MRAM and FeFET. All these concepts struggle today to show a convincing potential for a product. The issues are quite similar for all of them, and can be summarized into three points: variability, retention at temperature and density. Various

proofs of concept have been shown for each of those, and even a few products. From a European perspective the stand-alone market looks far away and running at an increasing pace. The embedded market seems to be the only field where a fair competition is still possible. Automotive-driven applications and non-mainstream technologies as FD-SOI can create the substrate where emerging eNVM can grow.

Currently, the need for new memory concepts is clear and strong, driving considerable efforts in research. Emerging eNVM however are far from being used in application, and the number of new concepts can lead to a risky energy-dispersive research work. In the short horizon, the main view needs to focus on clarify which technology has most potentials for success. It can be agreed that each eNVM needs to be refined with respect to given specification, but it is also clear that, for the time being, the lack of specifications is not the development bottleneck. Rather, emerging technologies are struggling to match in a strong and reliable way any specification. As the need for scalable embedded memory is quickly rising (nodes ≤ 28 nm) within 5 years we must expect to see a reliable demonstrator.

4.3.3.3. Silicon-on-Insulators

Fully Depleted Silicon-On-Insulator, or FD-SOI, is a planar process technology that relies on two primary innovations. First, an ultra-thin layer of insulator, called the buried oxide, is positioned on top of the base silicon. Then, a very thin silicon film implements the transistor channel. Thanks to its thinness, there is no need to dope the channel, thus making the transistor Fully Depleted. The combination of these two innovations is called “ultra-thin body and buried oxide Fully Depleted SOI” or UTBB-FD-SOI. FD-SOI technology enables control of the behavior of transistors not only through the gate, but also by polarizing the substrate underneath the device, similarly to the body bias available in Bulk technology.

As the slowing of Moore’s law signals the beginning of “Smart Everything”, each FDSOI process node can be a long-lasting technology with differentiated options (RF, Mixed signal, Ultra Low Power, Embedded Memories, sensors, etc.). The challenges and needs will be mainly focused on their ultra-low power possibilities to reduce as possible the supply voltage and improve the energy efficiency.

FDSOI MOSFETs can be considered as the best FETs planar devices for low power applications, harsh environments (radiation, temperature), and they are also very interesting for analog and RF applications. FDSOI is more dedicated to consumer applications, mid-range smartphones, IoT, automotive, wearables and sensors, whereas FinFET also covers high performance computing and high-end application processors.

From the technological point of view, performances can be boosted also with breakthrough approaches like SOI advancements, monolithic 3D integration, or new materials. From the device architecture point of view, FDSOI technology can suit sub-20 nm.

Other interesting applications of FD SOI are the following: Integration of FDSOI with NVM (like PCM, OXRAM, FeFET...), development of ULP design ($V_d < 0.4$ V) for IoT market (wearable, medical...), 3D integration and FDSOI can respond to future neuromorphic and quantum computing approaches.

4.3.3.4. RF CMOS

There are 2 key aspects for RF CMOS, namely cost reduction and increased integration:

Cost reduction

RF CMOS has the biggest potential to achieve lowest cost positions for mainstream applications (i.e., products in large quantities).

- Economy of scale: High production capacities available worldwide.
- Sufficient RF-performance for applications up to 100 GHz. Today the available performance of bulk Si (28/22 nm) and of FinFET with f_t , $f_{max} > 250$ GHz is sufficiently high. Similarly, for analogue applications ADC/DAC/DPLL with bandwidth requirements up to ~1 GHz the performance of today's RF CMOS technologies suffices.
- Lower supply voltages result in a lower power consumption, which leads to lower system cost as a collateral.

Increased integration

The integration of RF and analogue components with micro controllers

- Integrating sensors with logic circuitry helps to optimize complex systems, which in turn lowers system costs and system power consumption at a high performance. With the integration of microcontrollers, many system parameters can be processed and made available at a low latency, which lays the foundation for the optimization of RF CMOS circuits.
- Built-in self-calibration and self-test, and the reduction of number of interfaces enable additional optimizations and benefits.
- Immediate cost reduction arises thru omission of separate packages, saw-, and pad frames.

As a side effect, higher degrees of integration lead to a shift in the value chain and require the extension of expertise from component focus to the system focus. Consequently, the European Union shall reclaim a leading role for this technology development and production, as today most of the state-of-the-art (RF-)CMOS production takes place outside the EU.

4.3.3.5. RF BiCMOS

Even though Silicon is not the best semiconductor material if for example comparing the carrier mobility to III/V semiconductors, the matureness of manufacturing and the presence of a natural oxide make it predominant for the semiconductor industry.

SiGe based BiCMOS technologies combine two worlds - medium scaled CMOS and high-performance bipolar transistors - which support a RF and mm-wave applications for existing and next generation communication and sensing systems.

RF and mm-wave circuits crucially rely on the performance of the active devices. Properties such as the transit frequency f_t , breakdown voltage V_{CE0} , transconductance g_m and flicker noise corner frequency f_c are among the most important device parameters for the design of RF and mm-wave circuits.

Modern commercially available SiGe based BiCMOS processes offer bipolar transistors with transit frequencies (f_t) > 300 GHz and transconductance values of more than $200 \text{ mS}/\mu\text{m}^2$ and CMOS nodes down to 55 nm which allows to create complex mixed signal RF systems. These technologies already open the way towards applications in the sub-THz range.

Highly scaled CMOS devices still do not fully achieve the RF performance of state-of-the-art SiGe devices with respect to transit frequency, breakdown voltage (RF power handling capabilities) and noise properties. As an alternative BiCMOS technologies which are very well suited for

medium chip complexity due to their CMOS nodes down to 5 nm has superior RF performance at relatively low mask cost. For this reason, BiCMOS still is and will be in the future a leading technology when it comes to mm-wave and sub-THz range communication and sensing front-end systems.

As Si based power generation in the sub-THz still is still mainly based on RF multiplier and power combining architectures which offer limited efficiency and require substantial die area, the increase of the SiGe devices transit frequency f_t at reasonable breakdown voltages around 1.5 V is an important precondition to truly push silicon-based systems into the THz range.

Within the TARANTO project, funded by the European Commission and executed by leading semiconductor vendors and research institutes in Europe the way was successfully paved towards true silicon-based sub-THz systems achieving SiGe HBT transit frequencies of above 500 GHz.

These devices will unfold their full potential if integrated in a complete BiCMOS environment allowing the design of next generation communication and sensing mixed signal ASICs. Though this approach still requires substantial R&D effort.

Additionally, combining this new generation of SiGe HBTs with photonic components into an advanced monolithic EPIC technology including high performance optical modulators and photo detectors will further boost next generation high speed electro-optical communication and signal processing systems.

Finally heterogeneous integration technologies as described in chapter 4.3.3.1 are an important key to fully utilize the performance of state of the art and next generation SiGe based and EPIC BiCMOS technologies. When combined with highly scaled CMOS devices and III/V semiconductors for RF power generation and photonic functionality powerful functionality and performance can be realized either by 3D stacking and Systems in packages.

4.3.4. System and Component Architectures

4.3.4.1. Beamforming for 5G

Massive MIMO and Beamforming are two distinct techniques, that rely on the use of multiple antennas and a combination of these antenna signals in different ways. We distinguish exploitation of antenna arrays of multiple antennas for use of spatial-diversity or combination of multipath radio signals (including multipath reflections) to increase capacity and enhance the user experience (QoE), and the use of actual beamforming, beam pointing and beam steering using phased array antennas (that have $\lambda/2$ AE-spacing) as a Line-of-Sight approach (LOS).

A distinct form of beamforming, using a parabolic reflector, is a method where (part of) the radiation of an isotropic source is transformed into a parallel beam. This dish-solution is widely used for backhauling, satellite TV (receive), and radio astronomy. Mechanical (automated) steering and adjustment of the dish are used to configure for optimum signal reception. Phased-array antennas use a “non-mechanical” steering mechanism to control, point, and steer the beam and the beam shape.

The number of Antenna Elements (AE) of the Antenna Array will ultimately define the minimum achievable size of the pencil beam (Beam Width or First Null Beam Width) through the Array Factor, which is for broad sight view in first-order estimation inverse proportional to $360^\circ/N$, hence a 16-element linear antenna array could provide a 22.5° pencil beam. Detailed beam

pattern simulations for broad sight angled directions should be simulated per use case, including the detailed properties of the Antenna Array mentioned earlier and including intentional compensation for individual AE-gain, to obtain detailed insight in effective coverage, bandwidth, and gain for the desired configuration. Both, Pencil beam and multibeam (NxM) architectures are considered to enhance flexibility and optimize coverage, capacity and QoS. Beamforming, using phased array antennas, has been identified as a key technology for B5G and 6G applications to overcome the increased path loss at mm-wave frequencies and to increase the possible rate of frequency reuse by focusing the emitted energy in a confined area. However, electrical (RF) beamformers face challenges with regards to energy consumption, footprint, and heat dissipation, causing multi-beam transmission with continuous steering of the beam to be a highly difficult task. Optical beamforming, on the other hand, applies phase shifts, true-time-delays (TTD) in the photonic domain and allows the compact integration of entire beamforming networks, due to the large wavelength difference between optical and RF.

4.3.4.2. Open Standardization

Although cellular connectivity standards are being addressed by 3GPP, there is no similar standardization effort for sensing. A key example is related to Automotive radar for which there is no mechanism in place to avoid interferences that could, in the worst-case cause fatalities. For Automotive, frequencies are regulated, and radars operate in the 76-81 GHz Spectrum. But nothing prevents signals to interfere. Using coherent radars while ensuring that two cars in the same area cannot use the same waveform, would solve the issue. One way could be the usage of V2V or V2X to dynamically allocate and broadcast waveform parameters to cars in an area where their radar signals can interfere before they reach that area.

4.4. Common strategic actions across the expert groups

Beyond the technical actions proposed by each expert group, some transversal actions can be defined as key enabler to secure and strengthen the proposed strategy. We try to summarize in this section some actions proposal that can be beneficial to the full connectivity value chain:

4.4.1. Maintain & strengthen Europe on semiconductor manufacturing equipment to have some leverage on its capability to access key non-European technologies

As mentioned previously, semiconductor manufacturing equipment have played a key role in the ability of USA to prevent China to access to key semiconductor technologies (while not owning the associated manufacturing capability). To avoid Europe to encounter similar issues, it is key that Europe increasingly protects and strengthens its assets. As such, Europe has some leverage if global trade tensions escalate further. It would also enable Europe to speak on its own rather than being forced to peak side. A dedicated R&D program supported by the European Commission could be dedicated to support the Europe semiconductor ecosystem to maintain its position and build strong collaborations with other countries (for example Japan and South Korea).

4.4.2. Strengthen Europe's position on EDA solution market

EDA vendor are also playing a key role in the current trade tension and strategy deployed by USA. Today, the USA is the key semiconductor EDA solution provider (Cadence and Synopsys) but Europe has the potential to have a strong contender. Consequently, it is vital to strengthen this asset to secure Europe's sovereignty to access key technologies and to prevent European industrial players to be locked in a monopoly. Dedicated actions targeting to revitalize Europe's start up ecosystem on EDA solution could be one approach.

4.4.3. Support Europe's contribution to standardization activities

Standardization is playing a key role on the capability of industrial players to capture new market opportunity and to operate on wide and open markets. The way Chinese players have been pulling resources on 5G standardization and the impact it had on the value captured by Chinese players (such as Huawei) on the 5G market is a very good example. Notably, the first response of China's Ministry of Industry and Information Technology to increasing trade tension with USA was to announce the creation of a semiconductor standards committee composed of 90 leading Chinese companies. It is advisable that Europe acts diligently by, for example, providing tax incentives to companies allocating financial resources and dedicated staff to standardization activities. This could help to ensure that Europe plays a sufficient role on this topic. Another means are tax benefits for research in Europe to develop necessary standards essential IPRs, which will be implemented in global standards.

In the framework of COREnect it is important to stimulate research programs that accelerate the knowledge transfer from universities and institutes to the European semiconductor industry (as well as joint research) that generate technologies and innovations implementing the 3GPP-standardized URLLC features in the short term. Wireless sensor networks should not rely only on 5G-RAN to reach highly reliable low-latency capabilities. Also, the focus of governmental efforts and research institutes should be further set on URLLC over other RAN, e.g., WiFi or UWB, and help to close the standardization gaps. Consequently, in the future, the different technologies can be integrated in a wide interoperable network that allows seamless, secure,

low-latency and reliable interaction of sensors and IoT devices, independent from the RAN technology.

The vision for 2030 must be the European semiconductors industry and public institutions enabling a non-centralized communication network, in which every node and end devices (e.g., sensors, controllers, actuators, HMIs and the like) can guarantee reliable and secure communications according to the QoS demanded by individual markets (e.g. automotive and autonomous driving, industrial IoT and TSN), RAN-technology agnostic, and always oriented to satisfy end user and environmental requirements.

4.4.4. Strategic Infrastructure program lead by state members and the commission

Europe has wireless infrastructure players (Ericsson and Nokia) able to compete with Chinese ones (Huawei and ZTE). Still, in 5G deployment and associated value creation through new services, Europe is today lagging behind. In the NB-IoT market, for example, China accounts for 92% of the global NB-IoT connections. This is the result of the clear and aggressive deployment objective which was set by China's Ministry of Industry and Information Technology concerning NB-IoT. It enabled Chinese IoT hardware vendors to leverage their domestic market to develop their chipset solution. Cost effective solution and deployed infrastructures being available, key verticals have then leveraged this asset to create high value services addressing key societal challenges. As such, a clear and ambitious deployment strategy for an investment-friendly environment of 5G and 6G networks at the scale of Europe, as well as state funded infrastructure deployment program especially in areas which may not be served from purely economic reasons (as it has been done in France for optical fiber access), would clearly help to make sure that Europe can build its future leadership on connectivity technology on a solid foundation. States can also play the role of early adopters for public services to stimulate investment in public infrastructure for citizens. This suggestion also applies to cloud infrastructure to strengthen Europe's position on this market and enable a European ecosystem to emerge.

5. Conclusion

This report drafts the first contours of a strategic R&I roadmap for future European connectivity systems and components. These contours have been defined based on input from relevant stakeholders by the installation of expert groups. These expert groups have been divided over three strategic focus areas (Compute and Store, Connect and Communicate, and (Sense and Power). Across these expert groups, common challenges and roadmap approaches have been defined, and common strategic actions have distilled. This report advises to address three different timeframes to achieve enhanced focus in terms of strategic investments, markets, and technological development. The envisioned timeframes are short-, mid-, and long-term (respectively ~2, ~5, >10 years from now). Based on the defined roadmap directions, the currently defined strategic actions for Europe focus on strengthening its semiconductor manufacturing equipment, strengthening its position in the EDA solution market and fabless actors, enhance its contribution in standardization and maintain its support to strategic infrastructure programs.

This report is the first in a series of three. The upcoming work and deliverables will mature the strategic roadmap definition based on progressive insight and enhanced stakeholder consultation and interaction. The current report will be used as a platform for further discussion with the experts. The outcome of these discussions will result in the intermediate strategic roadmap report D3.4. That report will then be used for public consultation and final consolidation toward the final strategic roadmap report D3.5. Given this buildup, it should be clear to the reader that the current report (D3.3) is the first step in defining the envisioned strategic roadmap.

6. References

- [Aij18] Aijaz A. et al., "Towards a Tactile Internet Reference Architecture: Vision and Progress of the IEEE P1918.1 Standard", July 2018, <https://arxiv.org/pdf/1807.11915.pdf>
- [Dec20] Decision Etudes & Conseil, "Study on the electronic ecosystem. Overview, developments and Europe's position in the world", A study prepared for the European Commission DG Communications Networks, Content & Technology, 2020, online: http://www.decision.eu/wp-content/uploads/2020/02/DECISION_Study_Electronics_Ecosystem.pdf.
- [Dmi18] A. S. Dmitriev, L. V. Kuzmin, A. I. Ryshov, Y. V. Andreyev and M. G. Popov, "Ultrawideband Wireless Sensor Network for Real Time Process Monitoring", Engineering and Telecommunication (EnT-MIPT), pp. 76-80, 2018.
- [ECS21] ECS-SRIA, Electronic Components & Systems (ECS) Strategic Research and Innovation Agenda (SRIA), <https://aeneas-office.org/pdf/sria-2021/>.
- [EMS21] EMS NOW, "TSMC Ranks in Top-10 For Capacity in Three Wafer Size Categories", 2021, online: <https://emsnow.com/tsmc-ranks-in-top-10-for-capacity-in-three-wafer-size-categories/>.
- [Fas21] Frédéric Fassot, "Fabs 200 mm: vers une augmentation des capacités de 17%... en 2024", vipress.net, 2021, online: <https://vipress.net/fabs-200-mm-vers-une-augmentation-des-capacites-de-17-en-2024/>.
- [Faz16] E. Fazeldehkordi, I. Sadegh Amiri, O. Ayodeji Akanbi, "Chapter 2 - Literature Review ,A Study of Black Hole Attack Solutions", Syngress, pp. 7-57, 2016, ISBN 9780128053676, <https://doi.org/10.1016/B978-0-12-805367-6.00002-8>.
- [Hei21] W. Heinrich et al., "Connecting Chips With More Than 100 GHz Bandwidth", IEEE Journal of Microwaves, vol. 1, no. 1, pp. 364-373, 2021.
- [ICI21] IC Insights, "Revenue per Wafer Climbs As Demand Surges for 5nm/7nm IC Processes", March 4, 2021, online: <https://www.icinsights.com/news/bulletins/Revenue-Per-Wafer-Climbs-As-Demand-Surges-For-5nm7nm-IC-Processes/>.
- [Kin20] Ian King, "Intel and Softbank Beware. Open Source Is Coming to the Chip Business", 2020, <https://www.bloomberg.com/news/articles/2020-01-22/open-source-transformed-software-the-chip-industry-is-next>.
- [Kle21] Jan-Peter Kleinhans "The lack of semiconductor manufacturing in Europe", SNV, March 2021.
- [Lee19] Y. Lee, P.Tsung, M. Wu, "Technology trend of edge AI", IEEE International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2018.
- [Nen21] Daniel Nenni, "Global Wafer Capacity 2021-2025", SemiWiki.com, 2021, online: <https://semiwiki.com/forum/index.php?threads/global-wafer-capacity-2021-2025.13705/>.
- [ORA19] O-RAN Alliance, "O-RAN Architecture Overview", 2019, online; <https://docs.o-ran-sc.org/en/latest/architecture/architecture.html>.
- [Pan17] Z. Pang, M. Luvisotto and D. Dzung, "Wireless High-Performance Communications: The Challenges and Opportunities of a New Target", IEEE Industrial Electronics Magazine, Vol. 11, No. 3, pp. 20-25, Sept. 2017, doi:10.1109/MIE.2017.2703603.
- [Pet20] V. Petrov, T. Kuerner and I. Hosako, "IEEE 802.15.3d: First Standardization Efforts for Sub-Terahertz B", and Communications toward 6G", IEEE Communications Magazine, Vol. 58, No. 11, pp. 28-33, Nov. 2020.

- [PiT20] PiTrans - Fraunhofer IAF - <https://www.iaf.fraunhofer.de/en/media-library/press-releases/PiTrans.html>, press release March 16, 2020.
- [Rum20] Brandon Rumberg, "Analog ML is Relevant – Because Most Sensor Content isn't", tinyML talks, www.tinyML.org, 2020.
- [Shi16] W. Shi et al., "Edge Computing: Vision and Challenges", IEEE Internet of Things Journal Vol. 3, No.5, pp. 637-646, 2016.
- [Soi20] Soitec announces POI substrates business agreement with Qualcomm Technologies for 5G RF filters - <https://www.soitec.com/en/press-releases/soitec-announces-poi-substrates-business-agreement-with-qualcomm-technologies-for-5g-rf-filters>, Press Release July 2020.
- [Tou20] Baptist Tournoud, "Why VAD and what solution to choose", Dolphin Design Application Note, www.dolphin-design.fr, 2020.
- [TSM21] TSMC 2021 First Quarter Earnings Conference, April 15, 2021.
- [Urt16] M. Urteaga et al., "A 130 nm InP HBT integrated circuit technology for THz electronics", IEEE International Electron Devices Meeting (IEDM), pp. 29.2.1-29.2.4, 2016.
- [Ver15] O. Vermesan and P. Friess (book editors), "Building the Hyperconnected Society - IoT Research and Innovation Value Chains, Ecosystems and Markets", River Publishers, 2015.
- [Yan18] M. Yang, CH, Yeh et al., "A 1μW Voice Activity Detector Using Analog Feature Extraction and Digital Deep Neural Network", ISSCC, pp 346 – 347, 2018.

7. Appendix

The COREnect roadmap activities strongly rely on interactions with experts in the field. Some of the current members of the COREnect expert groups, as described in Section 4, are listed below.

Experts from within the consortium (non-exhaustive list):

Barkhausen Institut GmbH	Michael Roitzsch, Sebastian Haas
CEA	Bastien Giraud, Didier Belot, Dominique Moche, Emilio Calvanese-Strinati, José-Luis Gonzalez, Franck Badets
Ericsson AB	Dmitry Knyagin, Antonio D'Errico, Lars Sundström, Leif Wilhelmsson
IIIV/Nokia	Mohand Achouche
Imec	André Bourdoux, Björn Debaillie, Dimitrios Velenis, Eli De Poorter, Ingrid Moerman, Jan Craninckx, Jeroen Hoebeke, Johann Marquez-Barja, Mamoun Guenach, Michael Peeters, Nadine Collaert, Peter Ossieur, Piet Wambacq, Xiao Sun, Ilja Ocket
Infineon Technologies	Franz Dielacher, Giuseppe Bernacchia, Jochen Koszescha, Marina Plietsch, Siegfried Krainer
Nokia Bell Labs	Volker Ziegler, Wolfgang Templ, Patricia Layec
NXP	Frans Widdershoven, Cedric Cassan, Domine Leenaerts, Jan van Sinderen, Javier Velasquez Gomez, Jean-Claude Loirat, Patrick Pype
Robert Bosch Stiftung GmbH	Andre Guntoro, Andreas Schaller, Frank Hofmann
STMicroelectronics	Andrea Pallotta, Andreia Cathelin, Daniel Gloria, Frederic Ganesello, Pascal Chevalier, Pierre Busson, Raphael Bingert
Technische Universität Dresden	Diana Göhringer, Frank Fitzek, Gerhard Fettweis, Hermann Härtig, Viktor Razilov

Experts from outside the consortium (non-exhaustive list):

AT&S AG	Alterkawi Ahmad
Bergische Universität Wuppertal	Ullrich Pfeiffer
CTTC	Carles Anton Haro
Cyberus Technology GmbH	Werner Haas
Eindhoven University of Technology	Kees van Berkel
EPFL	David Atienza
ESA	Maria Guta
Ferdinand Braun Institute Berlin	Wolfgang Heinrich
Fraunhofer HHI	Martin Schell
Fraunhofer IAF	Thomas Merkle
Friedrich-Alexander-Universität Erlangen-Nürnberg	Robert Weigel
Gdansk University of Technology	Lukasz Kulas

Globalfoundries Dresden	Maciej Wiatr
Grenoble-Alpes University	Francis Balestra
Holistic innovation slu	Julián Seseña
ihp Microelectronics	Gerhard Kahmen
IMS Laboratory - Bordeaux	Yann Deval
InterDigital	Mona Ghassemian
Kalray	Benoît Dupont De Dinechin
Kernkonzept GmbH	Adam Lackorzynski
KU Leuven	Patrick Reynaert
LioniX International	Paul van Dijk
NaN	Werner Mohr
National and Kapodistrian University of Athens	Dimitris Syvridis
Orange	Jean Schwoerer
Politecnico di Milano	Salvatore Levantino
Racyics GmbH	Holger Eisenreich
Renesas	Marta Martinez-Vazquez
Silicon Austria Labs	Gernot Hueber
SINTEF	Ovidiu Vermesan
Soitec	Christophe Figuet
T3 Technologies	Gerd Teepe
Twente University	Eric Klumperink
United Monolithics Semi	Didier Floriot
Università di Pisa	Luca Fanucci
Université Nice Sophia Antipolis	Cyril Luxey
University of Oulu	Aarno Pärssinen
University of Pavia	Andrea Mazzanti, Danilo Manstretta
University of Piraeus	Angeliki Alexiou
University of Stuttgart	Markus Grözing
University of the Peloponnese	George Tsouslos
University of Valencia	Jose F. Monserrat