Sep 2020

Strategic Research and Innovation Agenda 2021-27

European Technology Platform NetWorld2020

“Smart Networks in the context of NGI”

2020

Table of Contents

Executive Summary

The preparation of Horizon Europe as part of the next Multiannual Financial Framework of the EU is progressing. The NetWorld2020 European Technology Platform (ETP), organisations representing more than 1000 entities, representing 5 % of European GDP, are contributing to the definition of research areas especially in the domain of communication systems and networks. This effort has also benefited from the notable help of the 5G Infrastructure Association (5G-IA), and from contributions from the Alliance for Internet of Things Innovation (AIoTI) and from the Networked European Software and Services Initiative (NESSI). Other external organizations provided inputs in different stages of the development of this document – the Strategic Research and Innovation Agenda (SRIA).

This document provides a summary of the key areas that the European R&D Community believes relevant for the future of communications technology in Europe and in the World. This analysis has been anchored in the challenges identified by the United Nations, the Sustainable Development Goals, and in the current policies inside the European Union, notably the European Green Deal. We identified research and development directions for the communications technologies, in order to realize these high-level societal objectives. The current changes society is facing as a consequence of the COVID-19 pandemic have not matured enough to significantly change the long-term research view for this release, but we look forward to reflect on these changes for the next release of the SRIA.

A major conclusion clear from the European view, is that the communications technology field is an increasingly multidisciplinary challenge. Future communications technologies will increasingly depend on evolutions in devices, in software developments, cybersecurity and artificial intelligence. In many aspects, future communications infrastructures will be the outcome of the intersection of developments in these fields, and an analysis of critical research areas in these domains, and what will be needed for the future communications system is present in different chapters.

The more traditional areas of research have been exhaustively covered, from service and system architecture, to radio and optical communications. The R&D analysis presented aims to lead for a sustainable evolution of communications, presenting a realisable path that may incrementally lead towards 2030 to future 6G systems. Nevertheless, some disruptive areas for the future (and some more challenging views on existing areas) have been briefly highlighted in the last chapter, which does not have such a clear realizability ambition.

1. Introduction

The preparation of Horizon Europe as part of the next Multiannual Financial Framework of the EU is progressing. The NetWorld2020 European Technology Platform (ETP) and 5G Infrastructure Association (5G-IA), organisations representing more than 1000 entities, representing 5 % of European GDP, are contributing to the definition of research areas especially in the domain of communication systems and networks.

ICT in general and networks (mobile and fixed) in particular is a fundamental enabler of a modern society. The Smart Networks of the future will be the nervous system of the Next Generation Internet and other commercial networks and are the platform for driving the digital transformation. Future communication systems and networks (Smart Networks) are the foundation of the Human Centric Internet. They provide the energy-efficient and high-performance infrastructure on which NGI (Next Generation Internet) and other digital services can be developed and deployed. Smart Networks will apply intelligent software (Artificial Intelligence and Machine Learning – AI/ML)) for decentralised and automated network management, data analytics and shared contexts and knowledge. Such infrastructures are the enabler for the future data economy. By virtualisation and strict policies, they will foster a free and fair flow of data, which can be shared whilst at the same time protecting the integrity and privacy of data, which is confidential or private: Users should be able to control their environment in the Internet and not to be controlled by the Internet.

The United Nations 2030 sustainable development goals [C1-1] require Smart Networks in many different domains using various appropriate communication technologies to support the digitalisation of society and economy in developing and developed countries. The United Nations Broadband Commission for Sustainable Development has set deployment targets for 2025 [C1-2] to underline the importance of communication systems and networks.

1.1 Global Megatrends – Societal Challenges

The beyond 5G research and development are unfolding against the background of global megatrends that will shape our world over the coming decades. Megatrends are change-related phenomena and transformative, global forces that define the future world. They have impact on businesses, societies, economies, cultures, and personal lives. Though the COVID-19 pandemic dominates the news, climate change is probably the most topical today, but phenomena including population growth and demographics, increasing environmental pollution or global competition for resources are also current megatrends [C1-3]. The United Nations in Geneve (UNOG), as well as the Government of Norway, have pointed out the challenge of Digital Divide as one of the global challenges [C1-4]. The governmental report on “Digital transformation and development policy” identified the four barriers to digitalization (i) access, (ii) regulation, (iii) digital competence and (iv) exclusion, as priorities for the government. In a similar matter, USAID has added “last mile connectivity” as the main priority for development policy. From the perspective of digitally empowering the users globally, it should be emphasized that digital inclusion is the catalyst for achieving the UN SDGs, implying that we also need to address how everyone in the society can benefit from basic information in the digital world available free on the web. Furthermore, certain UN SDGs require global cooperation to be successful leaving aside national interests, such as environment and ecological reconstruction, controlling spread of diseases [C1-5].

The trends identified in the 6G Flagship White Paper on Business of 6G [C1-6] and 6G Flagship White Paper on 6G Drivers and the UN SDGs [C1-7] were taken as the baseline and categorized according to the Quintuple Helix model [C1-8]. The Quintuple Helix aims to support a win-win situation between ecology, knowledge and innovation, creating synergies between economy, society, and democracy. Identified key megatrends in functional domain impacting Future Networks are summarized in the Fig. 1-1 utilizing PESTLE (Political, Economic, Social, Technological, Legal and Environmental) framework.

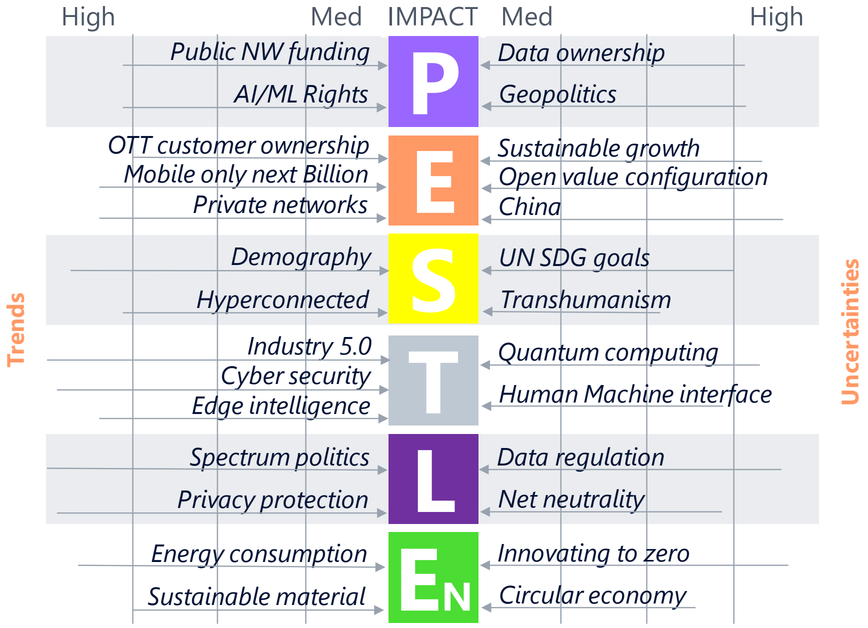


Figure 1-1. Identified megatrends impacting Future Networks, PESTLE (Political, Economic, Social, Technological, Legal and Environmental).

1.1.1 Trends related to the natural environment

Arguably the key trend influencing our future is the need for ecological reconstruction: how do we respond to climate change, decreasing biodiversity, the dwindling availability of resources and waste-related problems? Other trends need to be viewed against this backdrop as well. For example, the political system should be able to make decisions quickly – but are they made through centralization of power or inclusive decision-making or political correctness? Other trends in the natural environment system includes e.g. sustainable materials, which in turn contributes to Innovating to zero and circular economy trends. Towards 2030 companies will shift focus and develop products and technologies that innovate to zero, including zero-waste and zero-emission technologies. While technology offers new solutions for the production of energy it also simultaneously increases the demand for energy. This creates a conflict: to what extent does technology promote ecological reconstruction and to what extent does technology hinder it? 6G net positive impact and sustainability are expected to be through enabling increased efficiencies and improved environmental performance. Technologies of computing will be miniaturized to the extent that they sustain on the power generated through everyday human activity. The everyday activity of walking, jogging and everyday housework could produce energy to support the person's information devices which in turn would monitor person's vitals from time to time as well as cater to information and entertainment needs through over-the-top connectivity.

At the same time, environmental awareness among people and companies has increased, reflected in a growing number of people and communities changing their habits and companies taking corresponding actions to offer customer experience. Dissatisfaction with the current measures taken with respect to climate change and biodiversity has motivated a growing number of people to voice their opinions and participate in demonstrations.

Responding to the increased environmental awareness requires changes in culture and practices and has been accompanied by a polarization of views.

In the future, smart grids are extended to a variety of sectors including electricity, internet and healthcare. All of these are hyperconnected and completely automated. They serve as a middle layer between humans and natural environments enhancing the capabilities of both. These networks have been put together with public-private-personal ownership funding model with a view towards sustainable growth and usage of digital infrastructure.

1.1.2 Trends related to the political system

We are moving from multipolar world to a "poly-nodal" world, where strengthening of relational power is vital. In geopolitics the tension between the globalization, networking power and the urgency of the ecological reconstruction is linked to the balance between centralized decisions and the strengthening of inclusion and democracy. Towards 2030 power configuration is transforming from a multi-polarized world to a poly-nodal world where power will be determined in the economic, technological and cultural networks and interaction. Power is determined by relational influence and held not only by states but also by companies, regions, communities and transnational organizations. Societies are struggling to find a balance between quick-moving decision making, community engagement and the reasserting of democratic values and commitments.

Public network funding has traditionally been directed to unserved and underserved areas in terms of broadband access and coverage. Lately support for deployment programs has extended to policy areas, such as smart city community development, worksites and ecosystems (such as harbors and airports), advanced health services, logistics and transport, Smart Cities, public safety and critical infrastructure at length. Smart society builds dependable systems and communication where standardized data is utilized by walled garden platform monopolies across verticals. Smart city focus is extended to rural inclusion. 6G will transform urban and rural living existing at the intersection of geopolitics, growth of nationalism, rights to information transparency and information citizenry. Resource orchestration and configuration relates to the power over development and adoption of innovations and technology that ubiquitously embedded in society and our daily life. Technology is increasingly seen as a geopolitical issue of power and the questions of future resource orchestration power emerge: who will own the continuously accumulating data, who will get to decide on technology, and who sets the rules and regulation.

Access to data and data ownership are increasingly the major factors in value creation, and limiting such access is a means of control. Creating a system that transforms how data is collected, shared and analyzed in real time can create strong drivers for future value, introduce novel stakeholder roles, but may also lead to serious privacy and ethical concerns over the location and use of data. Privacy regulation is strongly linked with the rising trends of platform data economy, p2p sharing economy, intelligent assistants, connected living in smart cities, transhumanism and digital twins' reality at length. "I own my data" is expected to grow particularly in Europe based on GDPR evolution, though severe differences in the global data privacy laws are expected to be living on borrowed time. Artificial intelligence rights trends have contrary interpretations. Assuming the availability of appropriate data sets for training purposes, artificial intelligence has the ability to propose solutions to increasingly complex problems. AI can be the source of great economic growth, shared prosperity, and the fulfilment of all human rights. In the alternative future it drives inequality, inadvertently divides communities, and is even actively used to deny human rights.

Wireless spectrum politics and spectrum management in the 6G era will reveal a new level of complexity that stems from the variety of spectrum bands and spectrum access models with different levels of spectrum sharing. Local deployments of networks by a variety of stakeholders is expected to further grow in 6G. Furthermore, in national technology politics, spectrum regulation is used to gain competitive advantage. According to Net neutrality ruling Internet access providers should treat all traffic equally irrespective of sender, receiver, content, service, application or device in use. At the same time, the 5G evolution is already developing a network that can be extremely tailored to a use case intending to treat traffic differently for each use case. This legislation creates uncertainties by impacting companies' capabilities to create and capture value in virtualized network-based services between telecom operators and cloud providers.

1.1.3 Trends related to the education system

Diversification of the population: We see two trends, both ageing in well-established societies, as well as young generations in developing economies trying to catch up to the welfare of established economies. In well-established societies, the population is not only ageing but also becoming more diversified in terms of backgrounds, opportunities and habits [C1-3]. Highly functional educational system can help to provide equal opportunities for all. Continuous competence development is not required only from work force but all citizens. Learning capabilities will have a key role in the future, which creates pressure to re-evaluate educational structures and policies.

New technological innovations can revolutionize education in the future. Emerging opportunity of redefining human machine and brain-UI interface enables connecting people and the biological world in novel ways. Holopresence systems can project realistic, full-motion, real-time 3D digital twin images of distant people and objects into a room, along with real-time audio communication, with a level of realism rivaling physical presence. Images of remote people and surrounding objects are captured and transmitted over a 6G network and projected using laser beams in real time. The pervasive influence of artificial intelligence will not just reflect what something looks like but also its context, meaning and function, creating Internet of skills, Internet of senses and digital twins [C1-9].

However, not all have equal opportunities to access latest educational technology. Trends in mobile only next billion means, how ubiquitous cheap phones and increasingly affordable network connections in mega cities and rural are helping another billion users join the internet and access applications and digital content increasing at non-English speaking markets (www.basicinternet.org). For many, mobile is the primary or only channel for accessing the internet and services. With its unprecedented scale and growing impact on daily lives, mobile is a powerful tool also for achieving the SDGs and drive sustainable economic growth.

Academia is crucially important stakeholder not only when providing education but also fostering groundbreaking research. Alternative concepts and models for computing such as quantum computing that is at its best in sorting, finding prime numbers, simulating molecules, and optimization, and thus could disrupt segments like finance, intelligence, drug design and discovery, utilities, polymer design, AI and Big Data search, and digital manufacturing. For long the technology will be limited to selected industries, such as military, national laboratories, and aerospace agencies while other alternative approaches of compute to help handle greatly increasing level of parallelism in algorithms may be available more widely. Another aspect that needs to be considered is “ethics by design”, that will need to be considered, including the Responsible Research and Innovation (RRI) requirements defined by the EC [C1-22]. As services and applications are intermixed with the new technologies and networks and communications, the involvement of the stakeholders since the conception of the new service is a must which unavoidably include ethical considerations about the new technology, application or service.

1.1.4 Trends related to the economic system

The redefinition of the economy, is also linked to sustainable development: is the environment only to be regarded as a resource or should the economy aim to improve the state of the environment?

Urbanization will bring 5 billion people together to live in cities by 2030 occupying 3 per cent of the Earth's land but accounting for 80 per cent of energy consumption and 75 per cent of carbon emissions. 95 per cent of urban expansion in the next decades will take place in developing world where 883 million people live in slums today. Rapid urbanization is exerting pressure on fresh water supplies, sewage, the living environment, and public health. Cities are hungry, global economic engines and the economic powerhouses of the global economy. In 2015, 85% of global GDP was generated in cities. Cities are increasingly functioning as autonomous entities, setting social and economic standards. Urban identity will grow in importance compared to national identity.

Open value configuration and open source paradigm may provide a powerful avenue to reinvent civil society participatory process and regulatory capability. Utilizing sharing and circular economy trends co-creation utilize existing resources and processes to promote the stable interaction of mechanisms. As counterforces to the creation of platform monopolies decentralized platform co-operatives, the peer-to-peer economy and sharing economy models and the progress of a human-driven fair data economy has emerged. Towards 2030 platform ecosystems will provide an infrastructure on which innovation and transaction platforms are built. Crowdsourcing and crowdfunding are expanding the space for new forms of organization and innovation. Component prices have already decreased significantly, more sophisticated components will become cheaper, and new ones are invented. Low component prices provide possibilities for both local businesses as well as international ones. This encourages frugal innovations that are supported by the global do-it-yourself culture where sharing of blueprints and working processes is the norm to fight inequality in the world. In the heterogeneous society social networks and the trust and reciprocity they foster are highlighted from the perspective of well-being as well as working life. Novel decentralized business models do not necessitate a focal point but depicts the design of transaction content, structure, and governance to create value.

Via edge and extreme edge intelligence the proliferation of ever more powerful communication, computing and analytics resources at the edge of the network has turned the architecturally disaggregated 6G access networks into a rich service provisioning and access platform. Hyper-local services, such as augmented reality scenarios, do not require connectivity to distant service platform but, instead, perform better with local real time service access. Furthermore, the person will be supporting part of shared information processing and edge intelligence networks that address collective problems for humanity, such as genome sequencing (citizen science), through shared resources. The individual emerges as a node in the network of intelligence relations rooted in the local physical world while connected to the hyper-real 6G intelligence networks. This adopts a viewpoint of a public good through digital infrastructure of 6G supported through an ecology of information devices, products and services of IoT/IoE.

Over-the-top (OTT) companies will utilize their customer data, cloud infrastructure and AI/ML capabilities to challenge traditional operator's customer relationship ownership as users perceive connectivity as basic utility. In addition to the media space OTT players are offering basic telco services such as voice or messaging and are active in growth areas, such as cloud space and services, competing with telecom operators for clients and revenue. They are tying customers to their own ecosystems with carrier-neutral connectivity, while making reliance on traditional operators a thing of the past.

Towards 2030 industry 5.0 will allow collaborative human-machine interaction (HMI) across services and industries as human intelligence is in perfect harmony with advanced, cognitive computing. With real-time data, effective data monetization and digital automation of the manufacturing process, businesses will be able to shift focus towards generating higher revenues from servitization of products. Advanced manufacturing capabilities will help to overcome design complexities and with its ability to facilitate extreme long tail of mass customization and further return the control back to customers and in a haptic way. Private networks driven by industrial digital automation calls for standalone networks for high reliability, high performance in terms of both bandwidth and reliability, secure communications and data privacy, fulfilling business and mission critical needs.

1.1.5 Trends related to the media-based and culture-based public system

The technology is an embedded enabler in everything: in knowledge creation and circulation of knowledge. Technology is becoming a part of society and everyday life.

The sense of community created by 6G technology and the ability to directly collaborate with others enables humans to participate and act in society in an unprecedented way. Hyper Connected globe will continue to feel ever smaller in 2030: Globally 90% people will be able to read and have access to Internet and this trend is on the move. The aim is to recognize that 6G will transform urban and rural living existing at the intersection of geopolitics, growth of nationalism, rights to information transparency and information citizenry. Thus, once the infrastructure of 6G is in place, the content growth will be in terms of supporting multiple social and technological identities of people through a variety of media, creating new scenarios where privacy and human rights must be defined and preserved. This would require a mindful view on decision making and regulation of future, data, information, media and network usage in light of sustainable growth for economy. Thus, the human in 6G worlds will be increasingly sophisticated in terms of their media and service consumption while being rooted in their local economies. Connectivity is therefore not only virtual and digital, but also physical, such that the physical world will seamlessly meet the virtual world and, with novel ways of interacting with the human biological world.

When 6G technology has penetrated most parts of the world, IoT devices and sensors controlled by AI are an integrated part of environment. Automatic collection of different kinds of data (from humans: functioning of human body, biometrics, biosensors, etc. as well as from our environment etc.) and its analysis are used for highly sophisticated products and systems that make people’s lives easier and provide better user experience through convenience as everything is automated.

The need for cybersecurity and trust will be ubiquitous in the hyper connected world 2030. Even a temporary loss of technology may have, not only a productivity impact, but also a psychological impact on our lives. Furthermore, the subversion or corruption of our technology could result in the disastrous harm to our lives and businesses if, e.g., medical treatment devices deliver the wrong medication, educational systems teach propaganda, or home, on the autonomous move or work automation cause us injury or damage our products and businesses. In particular, expectations to protect and safeguard society and critical infrastructures from emergency situations by means of technological advancements are anticipated to grow.

Furthermore, with increasing polarization and the ageing and diversification of the population new tribes and communities will emerge around various imaginary groups representing wide variety of values, place of residence, political opinion, consumption choices or lifestyles. With weakened and fragmented future prospects, the absence of togetherness and the polarizing effect of social media have led to a rise in populism, skepticism towards changes in the environment and in the worst case, extreme attitudes. The consideration of different technologies and different types of services may foster undesired outcomes or limit them instead.

Transhumanism reflects the rise of technology-driven evolution at an unprecedented speed of change, propelling deeper questions into what it is to be human from biological, behavioral and human-machine evolutional perspectives. Cognitive intelligence revolution via ascendancy of sentient tools and further possibly human orchestrated self-directed selection in biological, neurological, and physical evolution.

1.2 Strong Contribution to the European Economy

In Europe, the entire ICT domain contributes significantly to the economy with about 5 % of total GDP, which corresponds to a market size of € 600 billion. This is also confirmed by findings of Eurostat. In particular, the communication systems and networks sector [C1-10] (manufacturing, including communication equipment and telecommunications) is critical in this market with a strong contribution to the European economy:

- About 27.2 % (1.74 million employees) of ICT employment [C1-10]
- 37 % (€ 234 billion) of ICT market size [C1-10]
- 47 % (€ 15 billion) of R&D expenditure in Europe [C1-10].

These numbers do not reflect the multiplication factor of advanced communications in the economy. The WorldBank has shown that the availability of broadband access increases economic growth and employment [C1-11]. Ecosystems connected to digital platforms and marketplaces create value for all members and have the potential to disrupt entire industries and show significant economic and social impact. For example, the automation achievable by the Internet of Things across a broad range of sectors will lead to a potential economic impact in the range of $4 trillion to $11 trillion by 2025 [C1-12]. A fully functional Digital Single Market could contribute €415 billion per year to the European GDP [C1-13]. Overall, the digitalisation of society is still in an early stage. For example, Europe's Digital Progress Report 2017 [C1-14]. points out that only 20 % of the companies in the EU28 countries are highly digitised and there are still many opportunities to be exploited especially by SMEs. According to an Accenture study [C1-15] the economic opportunity from digitalisation in Europe is over € 4 billion in value per day.

Smart Networks are of strategic importance as the enabler for basically all sectors in society and economy for jobs and economic growth.

1.3 Smart Networks Vision

5G as a network of networks is just the beginning of a new paradigm after the successful development of mobile communication systems such as GSM, UMTS, LTE and other networks in a complementary manner. Further development is absolutely crucial to address new challenges and requirements coming from many different sectors in society and industry. The smart network architecture will be software defined and provide features significantly going beyond connectivity: Multiservice and Mobile Edge Computing will allow to store and process data locally at the edges of the network to provide fast reactions and efficient use of network resources. Programmable aggregation and virtualisation functions as well as built-in security functions enabled (e.g. by the support of blockchains) will create a trusted environment for the Internet of smart things in which new applications and ideas can flourish. Future cost-effective communication systems and networks, both terrestrial and non-terrestrial, will increasingly be based on AI/ML and increased softwarisation in addition to requiring the continued development of classical communication technologies. Therefore, it is recommended to research future communication systems in close cooperation with these domains from an overall system perspective. The communication infrastructure will form the nervous system of the future Human Centric Internet and the digital transformation. It will intertwine distributed network, compute and storage resources to facilitate an agile composition of new services supporting a multitude of markets and industry sectors. From supercomputers and parallel computers, to data analytics, passing through cybersecurity, the Internet of Things (IoT), cooperative robots, or autonomous vehicles, it is universally agreed that every system and application must be interconnected to its peers, as well as to other related entities and systems. The interconnection of everything will be a distinguishable flavour of a competitive advanced society. Without network innovation the digital transformation is likely to fail.

Therefore, it is vital that the Smart Networks area is adequately represented in Horizon Europe.

The Smart Networks concept provides the necessary infrastructure and builds on scientific advances in the areas of physical and logical[1] sciences as well as key enabling technologies to provide a coherent framework supporting the future network designs. It is a combination of Smart Connectivity, Data Analytics (AI and ML), high performance distributed computing and Cybersecurity.

This Strategic Research and Innovation Agenda is summarising the different research domains to make the overall vision of Smart Networks happen.

Currently, there is a wide acceptance that 5G networks will have a significant impact in the worldwide economic development and revolutionize all aspects of everyday life. Even though 5G networks offer undeniable improvements over legacy networks, key findings suggest that European research activities in the communication networking sector need to continue at an increased pace. For this reason, we need to address several key challenges in a structured way so as to achievement maximum positive impact, see Figure 1-2.

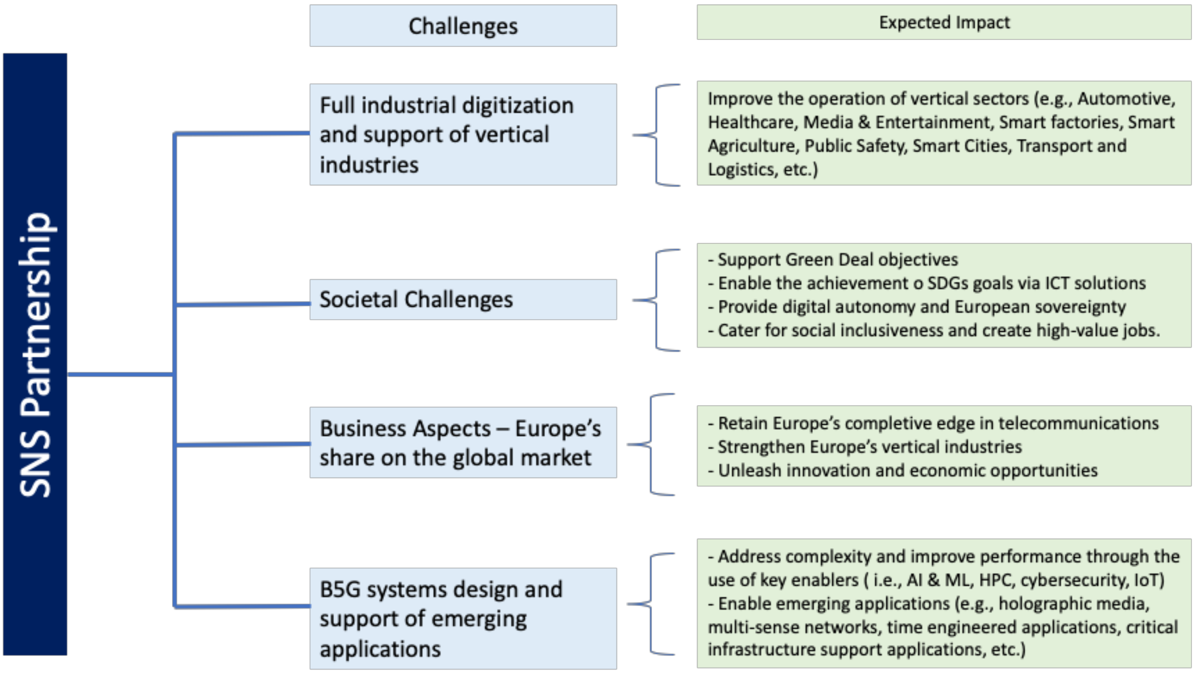


Figure 1-2. Challenging areas for the SNS partnership and expected impact, source: SNS partnership proposal [C1-16]

The common vision of the SNS Partnership envisages the integration of selected elements from new technologies such as Artificial Intelligence (AI) and Data Analytics; High Performance Distributed Computing, as well as Cybersecurity and Trust.

---

[1] Logical science means the specialised logic and mathematical development applied in ICT and Computer Science.

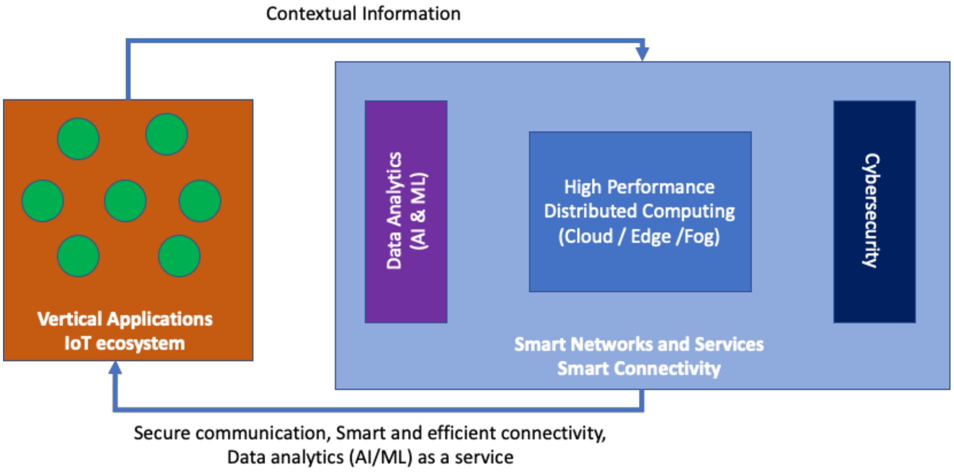


Figure 1-3. Integration of new technologies into the Smart Networks and Services vision [C1-16]**.**

Figure 1-3 above illustrates the interplay and integration of such new technologies whereas the rationale and main elements identified by the Partnership are described. Some of these enabling technologies are:

- Artificial Intelligence (AI) and Data Analytics**:** AI is perceived as a potential technology solution to cope with the increasing complexity of Smart Networks and Services system design and associated management, due to the extreme range of requirements for user experience, efficiency, and performance. Future **Smart Networks and Services will require robust intelligent algorithms to adapt network protocols and perform network and resource management tasks for different services in different scenarios.** For example, for network administration duties, AI will be mainly expected to make smart predictions, and to take improved decisions and actions that until now were mostly based on pre-defined rulesets obtained from experience accumulated by humans over the years. And, vice-versa, AI itself capitalises on digital technologies. For example, in their training/learning phase, AI algorithms rely on a significant amount of storage and computational resources [C1-17] so that huge amounts of training data can be fed into the algorithms. Mostly, the storage resources where data are originally residing and the computational resources where the algorithms are executed are not co-located. AI will also help telecom operators to optimise network deployment in real time according to user requirements. Accordingly, **Smart Networks and Services are needed to connect in an efficient and timely manner the data to the AI algorithms.**
- High Performance Distributed Computing (HPC): The computing paradigm has radically changed [C1-18]: in past decades, it was common to assemble the highest performing processors inside the same infrastructure equipment. However, the today's model is to design a data centre scale system with the entire network in mind. Thus, the trends that are most affecting High-Performance Computing (HPC) today are highly influenced by the need to have all aspects of the distributed computing infrastructure working in unison. Two of the most relevant aspects are the need for (i) a fast (and potentially massive) access to storage from anywhere in the network without disturbing the processing engines, which can only be accomplished with an intelligent network that can efficiently handle such tasks; and (ii) efficient strategies for network-wide off-loading, since nowadays CPUs are reaching their performance limits and, thus, the rest of the network must be better utilised to enable additional performance gains. Thus, current HPC architecture approaches depend heavily on offloading technologies that free the CPU from non-compute functions, and instead place those functions on the intelligent network. Thus, **Smart Networks and Services emerges as a key enabler for the transformation of the HPC domain.** Regarding the impact of HPC to Smart Networks, significant advances in components and techniques for wireless transmission are expected from continued research on quantum

processing and technologies in such a way that emerging **quantum technologies are expected to provide true breakthrough advances in, e.g., computing speeds within the next decade and, by doing so, pave the way for more secure communications** through potentially unbreakable cryptography.

- Cybersecurity and Trust: Starting from the NIS directive followed by strategic positions [C1-19] GDPR or continuing efforts on the Cyber Act [C1-20] the EU has recognised the fundamental role of cybersecurity in the digital world. This leads to at least two major priorities focused on (i) the **contribution of Smart Networks and Services to secure the digital world**, which includes the provision of data and information protection (e.g., location, data on the move) in all dimensions to protect citizens, enterprises and governments against malicious or outlaw usages with diverse security levels, as well as the detection of attacks and mitigation of their risks through the various components of digital services; and (ii) **the cybersecurity threats related to the Smart Networks and Services themselves**, namely, mitigating security threats related the various sensitive (new) components coming with 5G and beyond 5G technologies (e.g., identity and access management in mobility, virtualisation, softwarisation, over-the-air updates, AI technologies, etc.) as well as the monitoring and sustaining security levels for complex Smart Networks and Services.

Currently, many governments, globally, identify that world-class communications networks are, or will become, an intrinsic component of their critical national infrastructure and essential to ensuring that citizens can take full advantage of increasingly pervasive digital services across the plethora of existing and emerging applications, use cases and verticals, see e.g., [C1-21]. In particular, this can accelerate data-driven innovation, industrial automation, AI deployment and ensuing social and economic opportunities across economies, and increasingly between economies, for example the European Union’s developing Digital Single Market, or if considering how an autonomous system might operate across borders.

2. Policy Frameworks and Key Performance and Value Indicators towards 2030

2.1 Policy Objectives

One of the key objectives of Horizon Europe is to **ensure the competitive edge and sovereignty of EU's industry** [C1-16]. At the same time, the digital autonomy and European sovereignty have as a main objective to achieve a framework of alternative European offers so as to support freedom of decision making by keeping free trade in a global competitive economy. To achieve this, it is essential to support the research development and validation of B5G systems and building blocks as this technological area is extremely competitive at a global level. As mentioned above, Smart Networks and Services will be enhanced by innovative enablers (e.g., AI/ML, HPC, cybersecurity and IoT) and provide the means to support all vertical sectors and the emerging applications (e.g., holographic communications). Only if Europe is developing the state-of-the-art solutions it will be possible to retain its competitive edge and also boost those business sectors that are currently lagging behind (e.g., cloud, microelectronics, end devices). Obviously, the research activities have to be linked with the standardization activities so as to ensure that the key ideas arising from EU funded projects will have a global impact.

An important **Horizon Europe's policy objective is the digitization and respective transformation of the industry.** Europe has the competitive advantage of having a strong globally recognized vertical industry and telecommunications industry. Europe has also a top reputation for taking areas like privacy and security very seriously. The full digitization of the vertical industries, based on these principles, through the evolved communication networks will place Europe on the leading position.

Moreover, **Horizon Europe aims to create new sustainable and high-value jobs while at the same time contribute significantly at the social inclusiveness.** Research activities for smart networks and service should also take into consideration these aspects.

Finally, a key policy objective for the EU is the fostering of climate neutral, circular and clean industry. As will be discussed in the following section, telecommunication networks evolve, become more sophisticated but may also be more energy demanding. Smart networks should address energy efficiency in the telecommunications' sector and **contribute to the goals of the Green Deal** by making also vertical sectors more efficient. Moreover, by the full digitization of vertical industries should contribute to a number of goals of the **UN SDGs** as it will be discussed in more detail in the following sections. These points are further analysed in the following subsection.

2.1.1 Global view: UN Sustainable Development Goals

The 17 United Nations 2030 SDGs [C1-1] are a key driver for future developments to address societal challenges globally. The SDGs and the environmental sustainability challenges call for new and evolved Smart Networks and Services capabilities, high-level requirements and demands in capacities that must be carefully understood in a human-centric and societal context.

Smart cities and municipalities, smart mobility, smart eHealth and smart building solutions are example use case areas that can help to mitigate societal challenges around the ageing population and the increasing urbanisation. Smart manufacturing and communication services can enable more distributed and decentralised production of physical goods. Together with smart mobility solutions new possibilities will emerge that can positively influence and lower the urbanisation trend as well as result in reduced transportation and pollution.

There are challenges and opportunities within aquaculture, agriculture and waste handling for any vertical or segment. These can be addressed by supporting further advancements across various areas of IoT. New sensor technologies and new methods and techniques for life-cycle assessment can be enabled and supported to reduce the use of natural minerals and resources, reducing toxic waste and improve their handling, as well as reducing climate gas emissions.

Existing reports (e.g., [C2-1]) explain how mobile networks are contributing to the economic growth and are addressing social challenges.

ITU has summarised the contributions of the ICT sector to work on the UN SDGs [C2-2] as well as the investment in digital technology [C2-3]. These ITU guidelines will be used to design the future SNS platforms for the efficient support of key UN SDGs as explained in Annex 3 of the SNS partnership proposal [C1-16].

ITU has provided a methodology to identify which are the ICT building blocks for each SDG target group. In the SNS Partnership proposal [C1-16] there is a specific Annex on how this methodology can be used for Smart Networks and Services and is provided in Section 2.3 of this document.

The European answer to the SDGs has been organised around areas where missions [C2-4] will be developed. These missions will be an integral part of the Horizon Europe program and will link activities across different disciplines and different types of research and innovation.

Five areas have been identified : 1) cancer, 2) adaptation to climate change including societal transformation, 3) healthy oceans, seas coastal and inland waters, 4) climate-neutral and smart cities and 5) soil health and food.

Digital infrastructure and SNS should have a key role in this landscape and according to the missions described by the relevant boards, there will be a need to identify where SNS will be able to contribute and also what are the news requirements that SNS will need to take in consideration.

2.1.2 The Green Deal

The European Green Deal is effectively a response to challenges related to climate change [C2-5]. It is also focused on a *“new growth strategy that aims to transform the EU into a fair and prosperous society, with a modern, resource-efficient and competitive economy where there are no net emissions of greenhouse gases in 2050 and where economic growth is decoupled from resource use”.* Smart Networks and Services can contribute significantly to the multidimensional goals of the Green Deal.

The technological revolution enabled by a myriad of technologies opened gates of green revolution transversal to all the facets of the economy and vertical domains. It is recognised that *“new technologies, sustainable solutions and disruptive innovation are critical to achieve the objectives of the European Green Deal”.* Smart Network and Services will enable disruptive innovation within industries, across sectors and across the single market. This will result in building new innovative value chains while increasing significantly the large-scale

A Smart Networks and Services Roadmap will be created, as discussed in the SNS Partnership proposal [C1-16], that includes focus areas and synchronises with the Green deal timeline while anticipating and reacting towards upcoming Climate Pact and other upcoming directives.

The success of the Green Deal will be measured by its implementation milestones and the impact achieved by the related contributions to each focus area [C2-5]. During 2021-2027 time period, the contribution of Smart Networks and Services will be multidimensional and synchronised with the timeline of the critical milestones and the legislation rollouts where appropriate and possible. It will take all the elements of the value chains within the Partnership to make major parts of the Green Deal work.

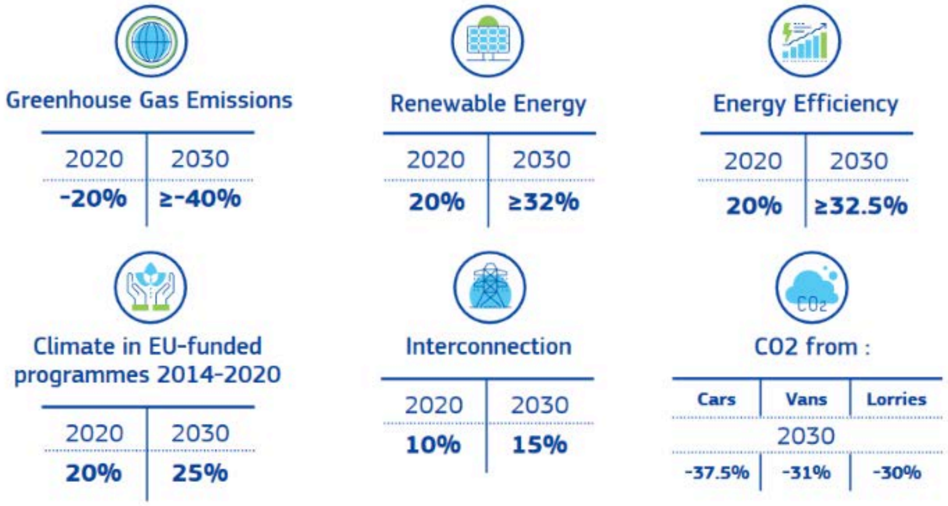


Figure 2-1. EU-Energy and Climate Targets 2018/2019 / Source: European Commission 2019 [C2-6].

2.1.3 Full industrial digitization and support of vertical industries

Full industrial digitization is needed to empower virtually all sectors of society and economy. For this it is necessary to integrate business and technological enablers to support the future vertical domain applications requirements (e.g., new types of IoT devices and components, large number of devices and components that need to be accessed simultaneously etc.). At the same time, independently of the addressed vertical sectors, the digital transformation is needed to open new business opportunities related to the provision of end-to-end cybersecurity and privacy services. Thus, in the next decade it will be decisive for Europe to develop lead markets to ensure its competitiveness at a global scale and keep the technology leadership.

Telecommunication networks are an essential pillar of modern infrastructure. Each successive generation of telecommunication networks goes beyond a simple increase in speed or performance, bringing unique new service capabilities. Especially the step from LTE to 5G is showing a paradigm shift by supporting vertical sectors. The full digitization of the industry, the necessity to integrate technological and business enablers, and mainly the need to address European and global challenges across the full value chain have created the basis for new research and innovation targets. Communication Service Providers (CSPs) will need to adapt rapidly to this trend, offering new services over digital channels in several vertical domains creating strategic alliances with vertical sectors to build and offer powerful and persuasive B2B2X propositions. The full digitization of the industry (a.k.a. vertical sectors) will bring significant financial benefits and ensure the competitive edge of EU industry.

Smart Networks and Services will empower several vertical domains. An indicative list includes the Automotive sector, Healthcare and Wellness, Media and Entertainment, Smart Manufacturing, Smart Agriculture and Agri-food, Transport and Logistics, Smart Cities and Communities, Utilities and Energy, Smart Buildings, Public Safety, Smart (Air)ports, Smart Water Management etc. Improving the operation of these sectors will assist in providing solutions for several Sustainable Development Goals (SDGs) as defined by the UN and further discussed in Section 2.3.

Currently, the **full digitisation of the industry is not yet achieved**. In standardisation activities most verticals have not been thoroughly examined to identify key and meaningful requirements. Even for those cases (e.g., Connected and Automated Mobility – CAM) where an extensive investigation has been made, the full support of the most advanced cases (i.e., implementation of critical services in an optimal autonomous driving mode in SAE levels 4 & 5) cannot be yet supported by existing network solutions without having a full integration of mobility data. Therefore**, the evolution of networks must take place hand in hand with the vertical industries**. The telecommunication sector has to cooperate with the IoT world to a) identify specialised devices for e.g., industry automation, b) define appropriate and real-life requirements and c) validate that future networks can meet them. Also, this cooperation is needed to define the network to applications interfaces in a secure and trusted way [C2-7]. Deploying and managing a large set of distributed devices with constrained capabilities is a complex task. Moreover, updating and maintaining devices deployed in the field is critical to keep the functionality and the security of the IoT systems. To achieve the full functionality expected of an IoT system, research should be done in advanced network reorganization and dynamic function reassignment. Research is needed for providing new IoT device management techniques that are adapted to the evolving distributed architectures for IoT systems based on an open device management ecosystem.

Components (micro-electronic components) and devices mainly for IoT and vertical sector applications are essential elements of future secure and trusted networks and to support the digital autonomy of Europe. With respect to the increasing demand and expectation of secure and trusted networks, especially for critical infrastructures, there should be European providers for such devices as an additional source to latest technologies to complement the European value chain and mitigate the existing gaps.

2.2 Societal, Economical and Business Drivers for 6G

5G was mainly targeted to address the productivity demand especially in vertical sectors and, to some extent, utilize new technology opportunities. With future Smart Networks and Services we need a substantially more holistic approach and should include a larger community into the definition of Smart Networks and Services to address the goals, trends and demands to avoid merely commercially driven system definition. Even though 5G development was characterized to be driven by the verticals (URLLC and mMTC scenarios), main emphasis was on deployments driven by mobile network operators (MNOs) for eMBB scenarios. On the other hand, Smart Networks and Services will introduce super-efficient short-range connectivity solutions that are driven by new players in the market resulting in new ecosystems.

This Strategic Research and Innovation Agenda is identifying future research topics in the mobile, wireless and general communications networking domain. In the community different terminology is used for future target systems:

- 5G evolution

- Beyond 5G
- 6G

In the coming years the actually developed and deployed 5G mobile communication system will be further developed in upcoming 3GPP Releases to provide more challenging capabilities beyond the today's 5G vision to support newly identified requirements, e.g. from vertical domains. This is an evolutionary development by taking into account today's investments.

Experience in the last decades has shown that developments in policy objectives, society, business drivers, technology and applications domains are continuing, which will result in and allow significantly more challenging requirements than can be supported by today's and evolved systems. This leads to systems beyond 5G. Disruptive technology breakthroughs such as a new paradigm of a triangle of the physical (hardware systems), digital (software systems) and biological (wetware systems) worlds for real-time applications will allow completely new services and applications which eventually leads to 6G. These are for the time being targets for research and innovation with a time horizon of about 10 years from now.

Therefore, there is always the duality of

- an evolutionary further development of today's systems (i.e. 5G), its standardization and deployment and
- in parallel revolutionary disruptive research towards systems beyond 5G / 6G with a ten years' time horizon.

It is essential to follow this dual approach to initiate research and innovation activities at an early stage in this globally very competitive environment to keep pace and shape future developments, generate essential IPRs and to be prepared for the next technology steps.

The different time horizons of the evolutionary path and the disruptive path need to be considered to put the overall development and investment cycles into perspective.

Societal and business drivers will increasingly shape the Smart Networks and Services development, including political, economic, social, technological, legal and environmental (PESTLE) drivers as summarized in Figure 2-2, whereas Figure 1-1 captures the functionalities behind the drivers. To ensure that the benefits of **smart city** services and urbanization are fully shared and inclusive, policies to manage urban growth need to ensure access to infrastructure and social services for all, focusing on the needs of the urban poor and other vulnerable groups for housing, education, health care, decent work and a safe environment. The rise of always-connected, omni-present systems, gadgets and sensors serving digital automation of critical processes will set high requirements for trustworthiness and **resilience**. The ubiquitous connectivity and contextual **awareness** of future Smart Networks and Services networks is expected to promote ICT accessibility and use for the social and economic development of people with specific needs, including indigenous people and people living in rural areas. Thus, future Smart Networks and Services architecture fosters digital **inclusion** and accessibility unlocking also **rural economic** values and opportunities.

High energy efficiency targeting reduced network energy consumption is a critical requirement for Smart Networks and Services. Combined with the optimization of choice, use, reuse and recycling of materials throughout the product life cycle enable reduced total cost of ownership, facilitates the extension of network connectivity to remote areas, and provides network access in a sustainable and more **resource-efficient** way. Extensive research has been conducted into possible health effects of exposure to many parts of the frequency spectrum including mobile phones and base stations. All reviews conducted so far have indicated that exposures below the limits recommended in the ICNIRP (2020) **EMF** guidelines [C2-8, C2-9], covering the full frequency range from 0-300 GHz, do not produce any known adverse health effect (UN WHO). Introduction of novel Smart Networks and Services technologies will initiate the need to review the status of the science and identify gaps in knowledge needing further research to make better health risk assessments.

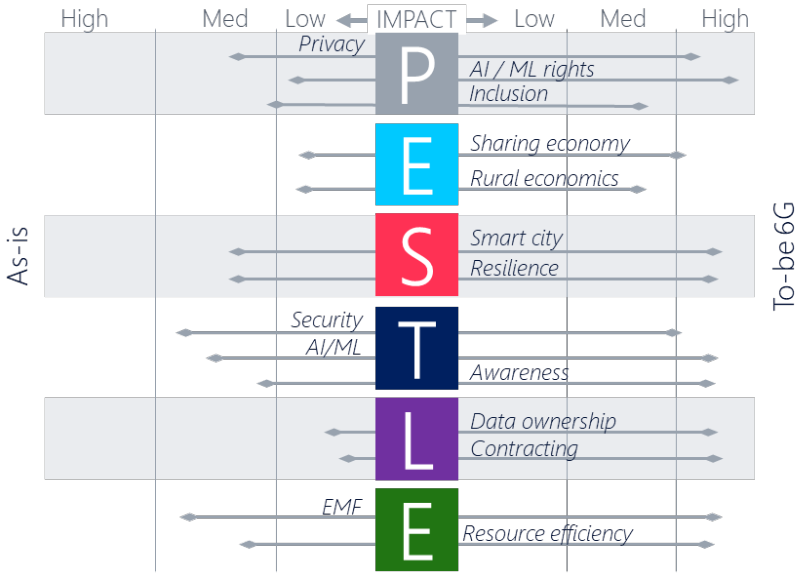


Figure 2-2. Future Networks PESTLE (Political, Economic, Social, Technological, Legal and Environmental) analysis results highlight inclusion, sustainability and transparency.

Data ownership is a source of value creation and control. Therefore, creating a system that transforms how data is collected, prioritized, and shared can create strong initial controversy, e.g., through raising serious **privacy** concerns over location and data. Furthermore, how to do business with data itself becomes a key question. The **contractual** policies between the actors will define the relative strengths of information and data ownership between parties, for example how the trust and ownership of information and data will be established in the autonomous smart device and service entities of the future. GDPR is one example of such policies.

While operations in the lower frequency bands (below 4 GHz) currently used for wide area mobile communication networks will remain rather stable with MNO market dominance due to long-term spectrum licenses, there will be new bands that target super-efficient short-range networks, especially indoors, but also in outdoor city spaces. These networks will target verticals with specialized demand and can be deployed by different stakeholders opening the market to new players, new investments and new ecosystems. Building several overlapping ultra-dense networks becomes infeasible and will lead to different stakeholders deploying a network within a facility to serve multiple user groups and services. Via softwarization and virtualization of network functions and opening of interfaces, **sharing economy** concepts will be utilized not only at high platform business layers but widely in network connectivity and data context layers. Changes in the ownership of spectrum access rights, networks, network resources, facilities and customers will result in several different combinations depending on the situation as different facilities have different requirements and infrastructures. New incentives will arise including functioning of the society.

As a sharing economy, Smart Networks and Services ecosystem will introduce new stakeholder roles and change the existing roles resulting in a complex network as outlined in Figure 2-3. While the MNO market dominance continues in 5G, future connectivity solutions can be driven by new players in the market. Stakeholders representing the different types of **demands and needs** originating from a variety of users that can be human or machine, and specific to the public sector or enterprises in different verticals are in the middle quadrangle in Figure 2-3. **Resources and assets** needed to meet the versatile needs are then provided by different stakeholder roles providing physical infrastructure (facilities, sites), equipment (devices, networks), data (content, context), under the regulatory framework set by the policy makers as depicted in the outer quadrangle. Demands and resources are brought together through the **matching/sharing** stakeholder roles including different kinds of operators (local or vertical-specific operators, fixed operators, mobile network operators, satellite operators), resource brokers, and various service/application providers such as trust/security providers.

Overall, the stakeholder roles in future networks are seen to change compared to the current mobile business ecosystem for public wide area networks and totally new roles will emerge. It is expected that especially the drivers listed in Figure 2-2 will fundamentally change the ecosystem and open new opportunities for different kind of stakeholders in 6G. The matching and sharing of resources to meet the demands will take place through new kind of activities to ensure inclusion, sustainability and transparency. Ultimately, the emergence and shape of the new 6G ecosystem are dependent on regulations which may promote or hinder the developments.

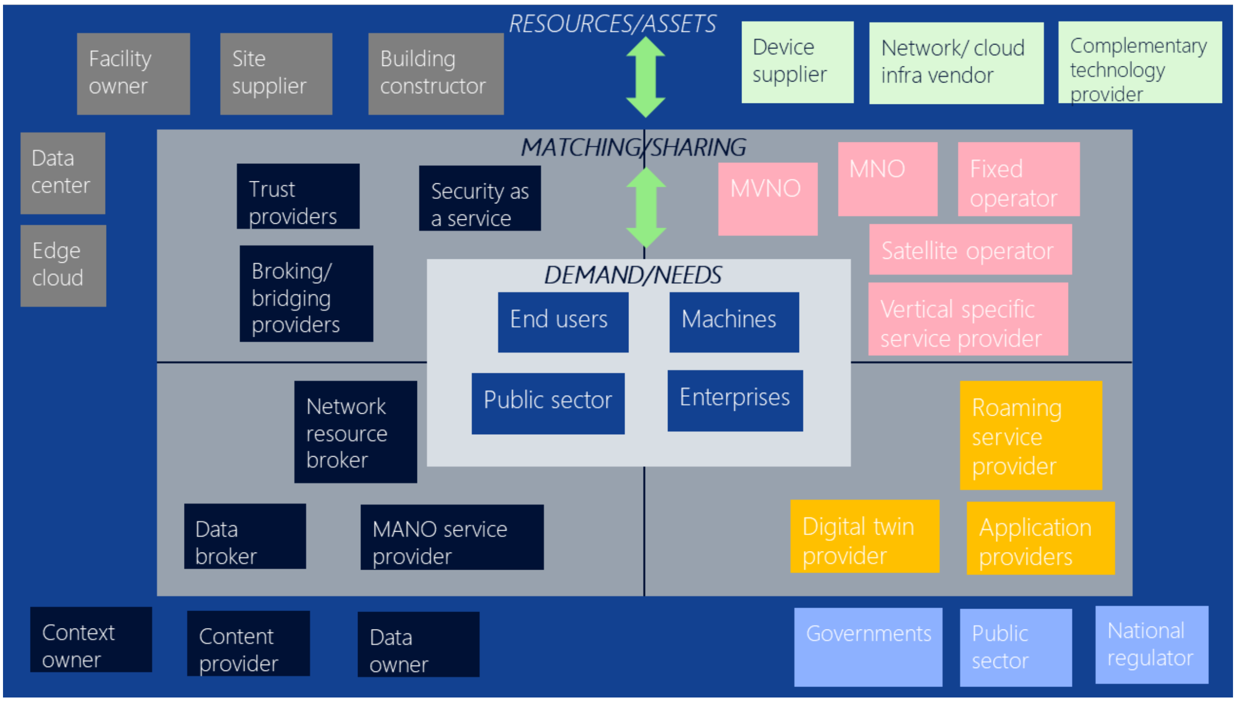


Figure 2-3. Stakeholders in future communications ecosystem.

The Smart Networks and Services vision results in a set of generic requirements to generate the envisioned societal and economic impact and to make the digital automation of everything happen. The following is a list of initial and indicative key requirements:

- Network operation will be automised allowing self-operating networks requiring the human operators to only validate the decisions of the system. Human intervention will be minimised, to correct/adjust the system-based decisions.
- Service deployment time will be reduced by a factor of 10 compared to similar tasks in 2020, based on slice creation and on the instantiation of the provider virtual machines where needed to deliver the service.
- Full integration of technical operations (routing, security, authentication, accounting, etc.) and business operations (invoicing, help centre, configuration, accounting, CRM, service contract), in a seamless digital infostructure.
- Slice creation on the fly with negligible time across the combined cloud, edge and fog infostructure, including data-centre computers, telco-based computer, and vertical user computers.
- Terabits per second will provide seemingly infinite network capacity and multi-core MEC servers will provide the required computing power for the future digital applications and services. This requires smart and effective alignment of application intent with network service offerings, enabling the concept of energy-efficient open multi-service Next General Internet (NGI).
- Application to application response time in the sub-millisecond range (latency) must be supported to enable a new class of highly responsive and interactive applications as well as a new level of industrial automation.
- Ubiquitous networks and services must be trusted, secure and dependable in the support of critical applications, also when the related infrastructure is deployed in a multitenant environment. Personalised and perpetual protection and privacy must be provided to cover the expanding threat surface due to the trillions of IoT devices and can deal with the growing number of threats triggered by the increasing value of data.
- Trillions of things and systems, including critical ones, need to be connected in a scalable and cost-efficient way.
- High efficiency in energy and natural resources usage is mandatory to limit the impact of ICT on climate change and sustain Earth resources. A significant energy reduction of network operation must be achieved.
- Combination of global reach, ubiquitous availability and optimised local service delivery capabilities, available on-demand, and enabling web-based software and IoT platforms.
- Spectrum efficiency above 256 bps/Hz will be achieved based on new waveforms and spatial reusability thanks to nanotechnology and massive MIMO techniques.
- Means for geographical and social inclusion must be provided in order to allow ubiquitous affordable Internet access.
- Infrastructure solutions should allow efficient deployment capital and operational expenditures, especially for very low-population density areas.
- Autonomous networks and systems based on AI and ML mechanisms combined with cyber physical security allowing to cope with the growing complexity and scale of operations. AI systems will enable intrusion/anomaly detection and effectively counteract distributed denial of service attacks coming from millions of hacked devices. Networks must be EMF-aware in order to allow deployment in areas with challenging EMF limits despite the operation in new additional spectrum bands and network densification.

2.2.1 Integrate new technologies and support emerging applications

The support of **emerging applications** (e.g., Internet of senses, holographic communications, full autonomous driving etc.) will require the improvement of the offered capabilities of B5G systems in terms of **Key Performance Indicators (KPIs) by at least an order of magnitude**.

New innovations are closely related to social inclusion and personal wellbeing, for example, as they are about the digital transformation of industries and businesses. These changes will **require a flexible and programmable architecture to satisfy the large diversity of use cases and applications**. In addition, the next generation of networks beyond 5G will go from software-centric towards the concept of human-centric: considering human skills, activities and behaviours first, and using automated functions to support them. The benefits can include reduced risk, higher rates of compliance, enhanced management support and improved interaction with users. These new functionalities have to come hand in hand with **advanced security and privacy schemes** to safeguard sensitive information for the users.

Additional modifications are also expected due to the requirements that will emerge through the **full digitization of the vertical industries.** Moreover, the operation of multiple logical networks (a.k.a. network slices) over the underlying network infrastructure supporting multiple vertical industries, will increase considerably the **complexity of the overall system**. **A new set of technological enablers has to be adopted by future communication**, such as advanced IoT solutions, Artificial Intelligence (AI) & Machine Learning, cloud & edge computing and cybersecurity. Meeting the performance KPIs, also requires a more extensive use of *high-performance distributed computing*. This way, data will be close enough to end devices to achieve almost zero latency. Moreover, advancements in electronics and specialised hardware is also needed to reduce energy consumption and help meeting strict KPIs in terms of delay, throughput, etc. These enablers are the most promising solutions to tackle network complexity and help us meet the desired KPIs.

Furthermore, it is expected that new emerging applications based on Tactile IoT, will be developed in the near future, see e.g., [C1-21]. Examples of Emerging applications using tactile IoT, as described in [C1-21] are:

- Holographic media applications: involve not only the local rendering of holograms but networking aspects, specifically the ability to transmit and stream holographic data from remote sites,
- Multi-Sense Networks: include emerging applications that involve not only optical (video, holograms) and acoustic (audio) senses, but as well smell and taste senses.
- Time Engineered Applications: use a communication system that can coordinate between different sources of information such that all the parties involved have synchronized view of the application.

  Critical Infrastructure support applications: support of critical infrastructures that refer to those essential assets that are considered vital to the continued smooth functioning of the society as an integrated entity.

2.3 Mapping of the UN SDGs to ICT development

The United Nations 2030 Sustainable Development Goals (SDGs) [C1-1] are a key driver for future developments to address societal challenges globally. The SDGs and the environmental sustainability challenges call for new and evolved Smart Networks and Services capabilities, high-level requirements and demands in capacities that must be carefully understood in a human-centric and societal context. Enabling European excellence in research, development, innovation and ubiquitous deployment of Smart Networks and Services will provide the tools to mitigate and tackle some of the major societal and environmental challenges of our times, such as ageing populations, increasing urbanisation, environmental protection and global warming. This must be achieved in a socio-technical and evolutionary context where the Partnership programme can adapt to changing needs and constantly evolving challenges and opportunities. By facilitating and enabling integrated and open ecosystem platforms with highly automated processes and service lifecycle support new tools will become available to empower societies, where the human-centric needs and the benefits to society will drive the developments.

Smart cities and municipalities, smart mobility, smart eHealth and smart building solutions are example use case areas that can help to mitigate societal challenges around the ageing populations and the increasing urbanisation. Smart manufacturing and communication services can enable more distributed and decentralised production of physical goods. Together with smart mobility solutions new possibilities will emerge that can positively influence and lower the urbanisation trend as well as result in reduced transportation.

There are challenges and opportunities within aquaculture, agriculture and waste handling for any vertical or segment. These can be addressed by supporting further advancements across various areas of IoT. New sensor technologies and new methods and techniques for life-cycle assessment can be enabled and supported to reduce the use of natural minerals and resources, reducing toxic waste and improve their handling, as well as reducing climate gas emissions.

While SNS will help other vertical sectors to become green and climate neutral, the telecom sector itself must continue to improve the efficient management of resources, the performance and coverage of the deployed infrastructure while lowering the climate and environmental footprint of the networks. With respect to coverage *“We need to balance the need for connectivity to become a utility with the cost of building the infrastructure to deliver it”* [C2-10]. In the future, as demanded by environmentally concerned customers, the sector must prepare for reporting upon own climate gas emission equivalents per type of subscription or even per service session use. This will enable the end-user to take informed choices, rooted in accurate climate and environmental footprint information. For instance, as new and high-fidelity live audio-visual person-to-person communication tools can be supported and become readily available, end-users will have a realistic alternative when there are needs for face-to-face communication. Thus, we can anticipate reduced time spend for travelling and a decrease of climate gas emissions.

Existing reports (e.g., [C2-1] [C2-2]) explain how mobile networks are contributing to the economic growth and are addressing social challenges. The United Nations Broadband Commission for Sustainable Development has set deployment targets for [C2-11] to underline the importance of communication systems and networks. The UN SDGs require the availability of ubiquitous and affordable communication networks to support the digitalisation of society and economy in developing and developed countries. ITU has summarised the contributions of the ICT sector to work on the UN SDGs (Figure 2-4) [C2-2] as well as the investment in digital technology [C2-3]. Mobile networks have a central role to play in this. Existing reports (e.g., [C2-12]) have presented, using a qualitative and quantitative analysis, the impact of mobile industry in all 17 SDGs (0). Obviously, not all SDGs are met equally well for a number of reasons (e.g., maturity of services, lack of required networking technological solutions etc.). However, these normalised scores should be improved significantly in the future. Beyond-5G networks will lead such efforts as they will affect a number of vertical industries that cover multiple sectors of everyday’s life. Continuous research activities will create a number of technological breakthroughs. These are needed for the efficient support of the diverse requirements of the verticals and the expected massive connectivity of end-devices. All these efforts require multi-disciplinary and cross organisational collaboration activities (public and private sector, regulatory bodies etc.). There is work to discover the linkage between future 6G and SDGs [8] where the role of 6G is seen as three-fold: 6G as 1) a provider of services to help steer and support communities and countries towards reaching the UN SDGs', 2) an enabler of measuring tool for data collection to help reporting of indicators with hyperlocal granularity, and 3) a reinforcer of new ecosystems based on 6G technology enablers and 6G network of networks to be developed in line with the UN SDGs.

Financial inclusion: Mobile access to financial services for the world's two billion unbanked

e-Agriculture: Access to market updates, and weather forecasts increases rural business productivity

e-Health: Be He@lthy, Be Mobile. Direct patient interaction, health informatice and telemedicine

e-Learning: Access to knowledge to all people no matter where they live or how much they earn

ICTs are an essential pathway to gender equality and empowerment

Smart water management systems, sanitation and hygiene

Energy efficiency, smart grids, green standards and technology for sustainable energy

Promoting the digital economy, e-commerce, tech-SMEs, entrepreneurship and cyber trust

Provide universal and affordable access to the Internet. ICTs are essential for a resilient 21st century infrastructure and access to services and applications

Narrow the digital and empower communities

Smart sustainable cities, intelligent transport systems, 5G and the Internet of Things

ICTs enable sustainable production and consumption through smart grids, smart metering and cloud computing

ICTs support greener lifestyles, climate monitoring, forecasting and early warning systems

Satellite oceanic observations and monitoring increases scientific knowledge of the ocean

Satellite observation of terrestrial ecosystems helo to protect biodiversity

Open data increases transparency, empowers citizens and drives economic growth

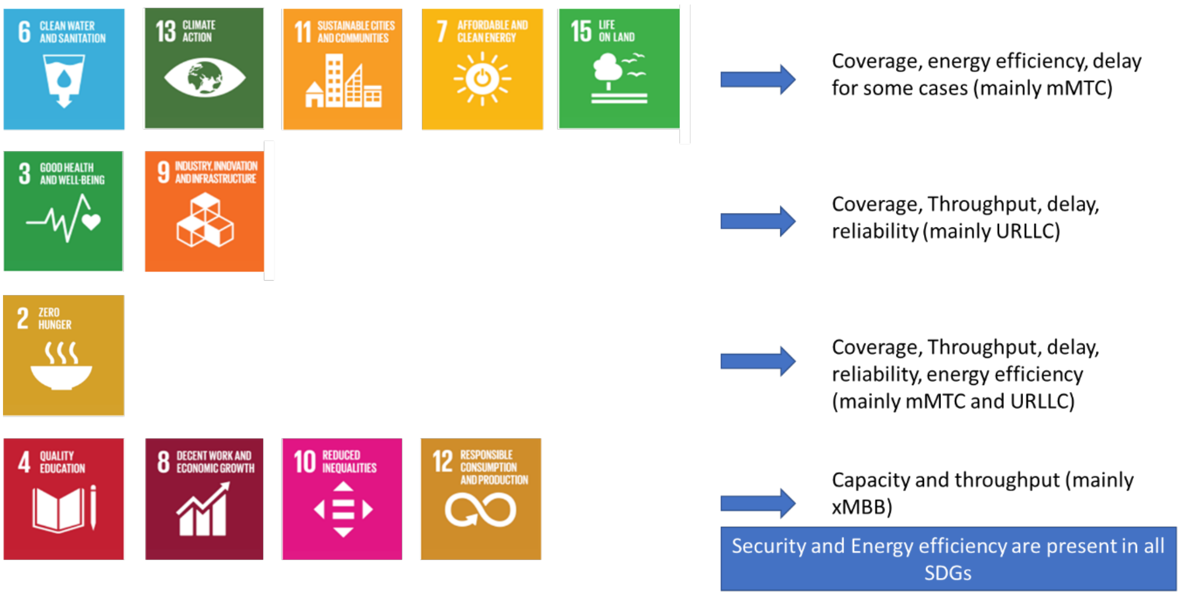
ICTs integrate and facilitate all SDGs through innovative collaboration and scaled up capacity building

Figure 2-4. ITU-R view in Sustainable Development goals [C1-1]

ITU has provided a methodology to identify which are the ICT building blocks for each SDG target group. For some of the SDG targets, this methodology is feasible since the related services of a vertical industry are quite mature. However, this is not the case for all SDGs, since some of these domains and processes are either in their infancy or they are currently evolving. Looking at the big picture though, one can easily identify what are the main technological areas that characterise each SDG.

Figure 2-6presents such an indicative listing where some SDGs (i.e., 6, 7, 11, 13, 15) require mainly ubiquitous availability, energy efficiency and massive IoT service management. These SDGs are related for example to the deployment of vast numbers of IoT devices that collect information and improve the everyday life of citizens (e.g., water, quality, smart cities, improved management of power and energy etc.). Other SDGs (i.e., 3, 9) require a virtually "infinite network capacity", high throughput, ultra low-latency and high reliability. Examples of related services to these SDGs are those used in the autonomous vehicles' domain. Note here that although significant work has been performed in this area (i.e., V2X communication), fullfilling the capacity and delay requirements for full autonomous driving (i.e., SAE level 5) is still not supported by 5G networks as additional technological breakthroughs are needed.

Figure 2-5. SDG impact scores [C1-1]



Coverage, energy efficiency, delay for some cases (mainly mMTC)

Coverage, Throughput, delay, reliability (mainly URLLC)

Coverage, Throughput, delay, reliability, energy efficiency (mainly mMTC and URLLC)

Capacity and throughput (mainly xMBB)

Security and Energy efficiency are present in all SDGs

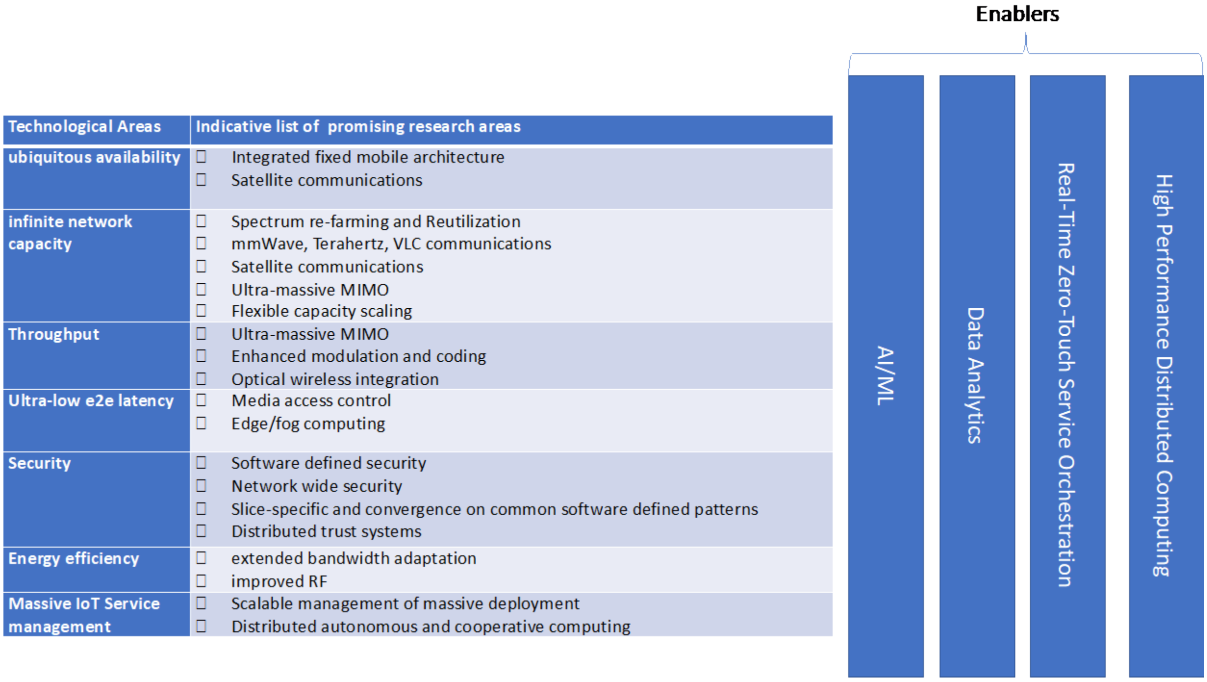
Figure 2-6. Indicative technological areas and their relation to SDGs

In relation to the SDG 2, an initial analysis indicates that it needs a combination of technological solutions discussed for the previous two SDG groups. For example, zero hunger can be addressed by increasing the food productivity (i.e., smart farming requiring a significant amount of IoT solutions) and efficient cultivation techniques (e.g., remote/autonomous driven tractors etc.). Finally, the last group of SDGs (i.e., 4, 8, 10 and 12) requires ubiquitous availabilty, infinite network capacity and mainly increased throughput to support advanced services (e.g., advanced collaboration using holograms, etc.). Note that security and privacy are present in all the technological areas.

To understand in detail how the SDGs can be met and which ICT building blocks need to be developed and used, one needs to follow a more formal approach. In [C2-3] a methodology is explained on how the SDG targets can be eventually mapped into the necessary building blocks. In Annex 1 of this document we briefly discuss this methodology and present two exemplary cases of such an analysis. A thorough analysis of all SDGs will take place during the first phase of the Smart Networks and Services Programme. This will allow the Programme to set solid and realistic goals until the end of the Programme.

Significant research efforts have to be undertaken during the next decade for all these technological areas. The following Figure 2-7 is an example on how the technological areas needed by the SDGs are mapped into research activities for B5G systems. Figure 2-7 also presents a list of enablers for these research areas.

The European Commission is committed to the abovementioned SDGs [C2-13]. In his State of the Union speech in 2017 the President of the EU Commission, Jean-Claude Juncker, formulated the vision of “***a Europe that protects, a Europe that empowers, a Europe that defends***” [C2-14] which guides to the proposed next Multiannual Financial Framework from 2021 to 2027 of the European Union and the proposed Horizon Europe programme [C2-15] Data security and citizen’s integrity are key European objectives. This is one of the key enablers for Europe’s strategic autonomy. The proposed Partnership in Horizon Europe contributes to this European vision as “***Smart Networks and Services empowers society and protects citizens***”.



Technological Areas	Indicative list of promising research areas
ubiquitous availability	☐ Integrated fixed mobile architecture ☐ Satellite communications
infinite network capacity	☐ Spectrum re-farming and Reutilization ☐ mmWave, Terahertz, VLC communications ☐ Satellite communications ☐ Ultra-massive MIMO ☐ Flexible capacity scaling
Throughput	☐ Ultra-massive MIMO ☐ Enhanced modulation and coding ☐ Optical wireless integration
Ultra-low e2e latency	☐ Media access control ☐ Edge/fog computing
Security	☐ Software defined security ☐ Network wide security ☐ Slice-specific and convergence on common software defined patterns ☐ Distributed trust systems
Energy efficiency	☐ extended bandwidth adaptation ☐ improved RF
Massive IoT Service management	☐ Scalable management of massive deployment ☐ Distributed autonomous and cooperative computing

Figure 2-7. Indicative list of research areas and technical enablers

2.4 Key Performance Indicators (Access, Network, Management)

There are several Key Performance Indicators (KPIs) that will be subject for enhancements in future Smart Network, and these span the device, the radio access, the core network, and the end-to-end system management. Examples of these KPIs include: Bandwidth, Capacity, Spectrum Efficiency, Peak Data Rate, User Data Rate, User-plane Latency, Control-plane latency, Jitter, Device Density, Area Traffic Capacity, Reliability, Availability, Energy Efficiency, EMF values, Coverage, Mobility, Positioning Accuracy, Service Deployment Time, Network Automation Metrics, Security Metrics, etc.

These technical KPIs and their targeted values in the evolution of 5G towards 6G are currently under discussion mainly in the scientific community with respect to the envisaged usage of future systems, cost implications, business cases and technical feasibility. For the time being no KPIs are formally agreed and set as requirements for research and innovation to be targeted in future wireless systems. With focus on the radio access, ITU-R WP5D has just recently (February 2020) initiated the development of a “Technology Trends Report”, which will lead to an updated vision document to agree technical KPIs on global level. In the coming years, associations in the commercial domain such as NGMN, GSMA, 5GAA, 5GACIA as well as regional associations, e.g. 5G IA and international counterparts will contribute to this discussion to achieve a global consensus.

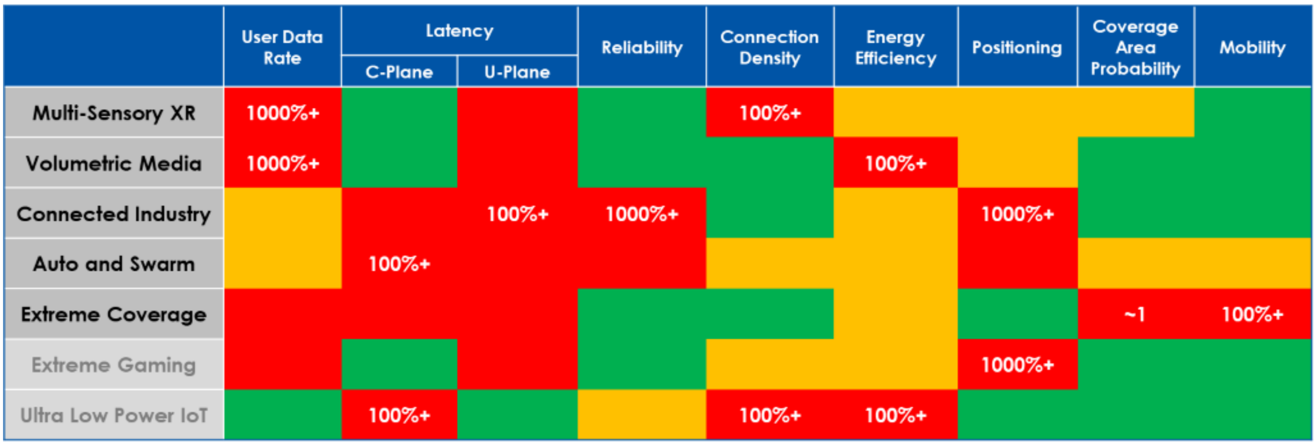
For instance, systematic EMF measurements for 5G networks and beyond, will either verify the compliance to the EMF limits of the existing networks, or will suggest better EMF values of the networks in the real-world combined networks. These measurements cover both health and safety reasons. For Societal reasons, the KPIs and EMF measurements should both protect the users in well covered areas, as well as identify underserved areas in which the users are “left out” creating a societal gap. A value of the network can be measured or appreciated only when it is available… Thus, the EMF measurements may become an indicator of satisfaction (or not) of the 5G network radio coverage.

The following Table 2-1 developed in the H2020 EMPOWER project2 shows an exemplary challenging sketch of potential future technical KPIs for the terrestrial radio access as target for research and innovation and an input to the upcoming global debate, which provides guidance for the discussion in the community of the expected research direction. Such KPIs are under investigation and need to fit to requirements of industry and will be regularly updated based on state-of-the art knowledge.

The KPIs in Table 2-1 have been derived as forecasts for the evolution of the KPIs targeted in 3GPP 5G NR for ITU-R IMT-2020 in the short, medium and long terms. These forecasts are based on an analysis of emerging use cases and their current and future requirements in the short, medium and long terms. Noticeably, seven use cases that embody the wireless challenges ahead have been considered in this analysis, namely:

1. Multi-Sensory Extreme Reality (XR) and Haptics
2. Volumetric Media Streaming
3. Connected Industries and Automation
4. Autonomous Vehicles and Swarm Systems
5. Extreme Coverage
6. Extreme Gaming
7. Ultra-Low Power IoT

The impact of the above extreme use cases on some of the current 5G NR KPIs is captured in the heat map illustrated in Figure 2-8.



	User Data Rate	Latency		Reliability	Connection Density	Energy Efficiency	Positioning	Coverage Area Probability	Mobility
		C-Plane	U-Plane						
Multi-Sensory XR	🟥 1000%+	🟩	🟥	🟩	🟥 100%+	🟨	🟨	🟨	🟩
Volumetric Media	🟥 1000%+	🟩	🟥	🟩	🟩	🟥 100%+	🟨	🟩	🟩
Connected Industry	🟨	🟥	🟥 100%+	🟥 1000%+	🟩	🟨	🟥 1000%+	🟩	🟩
Auto and Swarm	🟨	🟥 100%+	🟥	🟥	🟨	🟨	🟥	🟨	🟨
Extreme Coverage	🟥	🟥	🟥	🟩	🟩	🟨	🟩	🟥 ~1	🟥 100%+
Extreme Gaming	🟥	🟩	🟥	🟩	🟨	🟨	🟥 1000%+	🟩	🟩
Ultra Low Power IoT	🟩	🟥 100%+	🟩	🟨	🟥 100%+	🟥 100%+	🟩	🟩	🟩

Figure 2-8. Heat Map of the impact of example use cases on today's 5G NR KPIs.

In addition to the above heat map, and with the aim to keep the KPIs ambitious, especially as it is impossible to foresee all future use cases in a time horizon the spans the next decade, the KPIs' forecast in Table 1 also factor in assumptions on the pace of technological advances that would make it possible to meet or exceed these KPIs in the short, medium and long terms. Note that the table does not necessarily means that all these KPIs are realized at the same time for a specific user device/use case. They represent the collective most stringent values in a given performance domain, in order to support all possible use cases.

Table 2-1: Selected KPIs Forecast for Terrestrial Radio Communications during the short, medium, and long -term evolution of 5G NR.

Target KPI	**5G NR (Rel.16)** **2020**	**Short-term Evo.** ~2025	**Medium-term Evo** ~2028	**Long-term Evo.** ~2030
Spectrum	<52.6 GHz	<150 GHz	<300 GHz	<500 GHz
Bandwidth	<0.5 GHz	<2.5 GHz	<5 GHz	<10 GHz
Peak Data Rate	DL: >20 Gbps UL: >10 Gbps	DL: >100 Gbps UL: >50 Gbps	DL: >200 Gbps UL: >100 Gbps	DL: >400 Gbps UL: >200 Gbps
User Data Rate	DL: >100 Mbps UL: >50 Mbps	DL: >500 Mbps UL: >250 Mbps	DL: >1 Gbps UL: >0.5 Gbps	DL:>2 Gbps UL: >1 Gbps
Density	>1 device/sqm	>1.5 device/sqm	>2 device/sqm	>5 device/sqm
Reliability [BLER]	URLLC: $>1-10^{-5}$	$>1-10^{-6}$	$>1-10^{-7}$	$>1-10^{-8}$
U-Plane Latency	URLLC: <1 ms	<0.5 ms	<0.2 ms	<0.1 ms
C-Plane Latency	<20 ms	<10 ms	<4 ms	<2 ms
Energy Efficiency (Network/Terminal)	Qualitative	>30 % gain vs IMT-2020	>70 % gain vs IMT-2020	>100% gain vs IMT-2020
Mobility	<500 Km/h	<500 Km/h	<500 Km/h	<1000 Km/h
Positioning accuracy	NA (<1 m)	<30 cm	<10 cm	<1 cm

As for the satellite communications, satellite working group within Networld2020 is considering the following KPIs (Table 2-2).

Table 2-2: Selected KPIs Forecast for Satellite Radio Communications during the short, medium, and long -term evolution of 5G NR.

KPI	Short tTerm Evo.	Medium-Term Evo	Long-Term Evo
Minimization of unmet capacity[1]	<0.1.%	<0.05%	<0.01%
Maximization of satellite resource utilization[2]	>99%	>99.9%	>99.99%
Time to reallocate satellite resources[3]	<1 min	<5 sec	<1 sec
Solving and detecting time of satellite operation incidents	<10 min	<5min	< 1 min
Energy Reduction using adaptive intersegment links	>50%	>80%	>90%
Connectivity gain for converged satellite cloud scenarios[4]	>100%	>150%	>200%
Reduction of required manual intervention[5]	>50%	>80%	>90%
Widespread IoT coverage [6]	> 50%	>99%	> 99.9%
Reliability (perceived zero downtime)[7]	>50%	>99%	>99.9%
Experienced data rate (Broadband)	DL: >50 Mbit/s UL: >25 Mbit/s	DL: >500 Mbit/s UL: > 250 Mbit/s	DL: >1.0 Gbit/s UL: >0.5 Gbit/s
Area traffic capacity (Broadband)	DL: >75 Mbit/s/km2 UL: >37 Mbit/s/km2	DL: >750 Mbit/s/km2 UL: >370 Mbit/s/km2	DL: >1.5 Gbit/s/km2 UL: >0.75 Gbit/s/km2
Experienced data rate (NB-IoT)	DL: >2 Kbit/s UL: >10 Kbit/s	DL: >20 Kbit/s UL: >100 Kbit/s	DL: >40 Kbit/s UL: >200 Kbit/s
Area traffic capacity (NB-IoT)	DL: >8 Kbit/s UL: >40 Kbit/s	DL: >80 Kbit/s UL: >400 Kbit/s	DL: >160Kbit/s/km2 UL: >800Kbit/s/km2

[1] User demand that is not satisfied

[2] Used satellite resources such as power, bandwidth, etc

[3] Allocation of satellite resources such as power, spectrum, beampattern given a change in the demand

[4] Increase in successful connections

[5] Reduction with respect to today manual intervention

[6] Gain with respect to 2020 wireless area capacity

[7] % of total operation time

The optical community is proposing the following key performance indicators

	Target KPI	Current 2020	Short-term Evo ~2025	Mid-term Evo ~2028	Long-term Evo ~2030
Metro/Core	Spectrum[1]	5THz	15THz	30THz	50THz
	Port speed[2]	400Gb/s	1.6Tb/s	3.2Tb/s	6.4Tb/s
	Bandwidth[3]	<75GHz	<300GHz	<600GHz	<1200GHz
	Line capacity[4]	25Tb/s	200Tb/s	600Tb/s	1.5Pb/s
	Node capacity[5]	150Tb/s	1.2Pb/s	3.6Pb/s	9Pb/s
Access	PON speeds	10Gb/s	50Gb/s	100Gb/s	>200Gb/s
	User data rate[6] (consumer)	100Mb/s	~1Gb/s	>2.5Gb/s	>5Gb/s
	User data rate[6] (business)	1Gb/s	~10Gb/s	>25Gb/s	>50Gb/s
	Latency[7]	<1ms	<100µs	<10µs	<1µs
	Power consumption[8]	100% (baseline)	40%	30%	20%
	Service provisioning	Hour	Min	Second	Sub-second
	Network operations	Operator-controlled, reactive	Intent-based, proactive	Self-diagnosing	Self-optimizing

[1] 25% CAGR, in line with conservative traffic predictions

[2] Extrapolation of Ethernet roadmap

[3] Using 400G DP-16QAM as baseline

[4] 50% CAGR, in line with internet content provider traffic predictions. Assumes exploitation of frequency and space domain.

[5] Based on degree 4 node with 50% local add/drop

[6] 50% CAGR based on Nielsen's law

[7] Excluding propagation delay

[8] 15% reduction per Gb/s p.a., extrapolated from past transponder data

With respect to the **system architecture and networking** we have the proposed metrics below:

- Runtime Service Scheduling efficiency increase compared to overprovisioning (for a service requiring 99.999% or higher success rates and under typical traffic arrival conditions)

Short term	Medium term	Long term
2x in single tenant enviroments	10x in single tenant	At least 10x in multitenant environments

This includes aspects as path stretch ((ratio between the average control plane path and the average physical node distance) and resource overhead (services being provided by the network resources versus maximum capacity of those resources).

- Time required for runtime conflict resolution when applying resource efficiency methods, that is the increase in multiplexing desired when compared to independent exclusive allocations and the time that is required to settle all the conflicts that may exist.

Short term	Medium term	Long term
2x for multiple concurrent, overlapping allocations	10x for multiple concurrent, overlapping allocations	At least 10x with critical guarantees

- In terms of network-resources collection (network garbage collection), in the sense of recovering resources that are not being used anymore, we expect:

.Short term ~2025	Medium term ~2028	Long term ~2030
Feasible, additional recovery process off-line	Feasible, running with the resource allocation	Optimal, on resource allocation actions

- Features of the pervasive resource control, in terms of autonomic functions.

	.Short term ~2025	Medium term ~2028	Long term ~2030
Configuration	Only a minimal initial pre-configuration (only domain name + security association data, e.g. private/public key)	No human intervention	No human intervention across different domains
Scalability	High, large number of nodes	Very High, any number of nodes, densities	Very High, any number of nodes, densities and complexity
Bootstrapping	Reduced time to 70%	Reduced time to 40%	Reduced time to 10%
Convergence time of the control plane	Time reduced to 70%	Time reduced to 40%	Time reduced to 10%
Signalling overhead in reconfiguration	Reduced to 90%	Reduced to 75%	Reduced to 75% in multitenant environments

- In terms of network-suitable AI, we expect

.Short term ~2025	Medium term ~2028	Long term ~2030
Adaptation of current centric-implementation AI models	Fully distributed AI algorithms at the network	distributed AI supporting and serving several models at the same time

In **security domain**, being a mandatory condition for numerous objectives, security is de facto a pre-requisite for the ongoing Digitalization of our societies. Building trust is combination of awareness, understanding and obviously provision of the right solutions with the right level of security. The ambitious objectives listed below aims at being representative of this combination:

- Towards access to real time Cyber Threat Intelligence information (attacks/threats and vulnerabilities), risk Analysis tools and Services enabling 100% of awareness and level-based appropriate protection counter-measure deployment.

Short-term Evo. ~2025	**Medium-term Evo** ~2028	**Long-term Evo.** ~2030
Federated, consolidated, common basis across CERTs (CSIRT network, NIS directive application)	CTI platforms(including openCTI) and tools for State-of-The-Art sanitization	100% of qualified threats knowledge and appropriate counter measures made accessible

- Trust in ICT infrastructure through systematic Exposure of cybersecurity levels 100% compliant with European-legal basis (certification, Security Service Level attributes, GDPR/EU strategy for Data,...)

Short-term Evo. ~2025	**Medium-term Evo** ~2028	**Long-term Evo.** ~2030
5G systems & services certification frameworks, Basic security level exposure with generic security attributes defined	Methodologies and tools for composition and time evolution of certified perimeters (systems & services)	Evolutive approach for data and disruptive technologies

- Compliance with highly critical applications and essential services requirements leading to sovereign solutions able to provide 100% availability of services for verticals

Short-term Evo. ~2025	**Medium-term Evo** ~2028	**Long-term Evo.** ~2030
Local, private implementation for limited set of verticals	End-to-End hybrid implementation for most of verticals	High grade support with technology, system and solution independence

- Improve attack detection & response mean time of Cybersecurity incidents including zero % unprotected data leakage

Short-term Evo. ~2025	**Medium-term Evo** ~2028	**Long-term Evo.** ~2030
Benchmark strategy including data set and models	Monitoring and attack detection EU-wide strategy	Data protection strategy with response time and robustness outperforming attackers capabilities

2.5 Technical Standard Areas

Since architecture, standardization and interoperability will play a critical role in establishing a successful ecosystem based on smart networks, this section provides an overview of important topics and challenges in key standards areas.

Standardization activities for B5G Systems: 5G networks will evolve to meet the requirements from the full digitization and from the need to support highly demanding emerging applications. This evolution will require modification in all network domains (radio access network, backhaul, core network) and very likely in the wireless technologies to be used over the air interface. For this evolution to have global impact input to the dominant standardization bodies is required.

Enhancing Networks with key enabling technologies: B5G networks will require to improve considerably their performance as well as their ability to handle complexity. For this reason, it is necessary to adopt advanced solutions that will be based in AI/ML and that will be able run in distributed high-performance computing environments. Moreover, the support of all vertical technologies will require the adoption of end-to-end sophisticated cybersecurity solutions.

Artificial Intelligence and Edge Computing: Standardization support is needed in the areas where applications and services are enabled by IoT, 5G and AI at the edge, along with edge computing, and when considering that information and knowledge need to be exchanged between edge and data centre.

Semantic and Syntactic Interoperability: Semantic interoperability is achieved when interacting systems attribute the same meaning to an exchanged piece of data, ensuring consistency of the data across systems regardless of individual data format. Semantic interoperability must be supported in order to exchange not only data but also information and features related to the source of the information (e.g. location, status, technology associated) - thereby facilitating the disappearance of the vertical information silos of the heterogeneous platforms that current IoT data lakes represent.

IoT Relation and Impact on 5G: With the introduction of 5G, vertical industries will embrace digital transformation, to move beyond traditional service approaches, on an unprecedented scale. This will be a new engine for economic growth and social development. A core element of 5G IA/AIOTI cooperation will be identifying the key requirements imposed by vertical industry sectors to anticipate relevant trends in IoT use cases and apply the knowledge gained to define their impact on the 5G architecture and features.

Standardization of Combined/Integrated ICT and Operational Technology (OT): Enabling technologies like 5G and IoT, edge computing and artificial intelligence are needed to support vertical industry enabled smart networks. Currently, there is a strict division on standardization focusing on ICT (Information and Communication Technologies) and that focusing on OT. Maintaining this separation makes it challenging to provide integrated standards for the ICT/OT needed to support vertical industry enabled smart networks. The 5G IA/AIOTI cooperation should therefore investigate and promote combined/integrated standards for ICT and OT.

2.6 Key Value Indicators

In addition to the KPIs discussed in Section 2.4, the success of the future smart networks will increasingly be measured with indicators that assess how well the networks contribute to solving of the major societal challenges discussed in Sections 1.1 and 2.2. These indicators are more related to value provided by the use of the networks than to the performance of the networks themselves. A number of value dimensions need to be addressed including societal and sustainability related challenges, business aspects, ethics, trust, privacy and security, EMF awareness, and efficiency, among others. Proper indicators need to be defined for the value dimensions resulting in value related indicators for the future smart networks.

Societal Challenges will encompass broad and complex missions for technology-oriented research and innovation. They are characterized by open-endedness and undefined final outcome. Progress towards solving Societal Challenges with conventional metrics and indicators can turn to be a difficult task. It will be necessary to reflect on a new impact assessment paradigm, focused on value creation and sustained long-term benefits and effects, around identified domains.

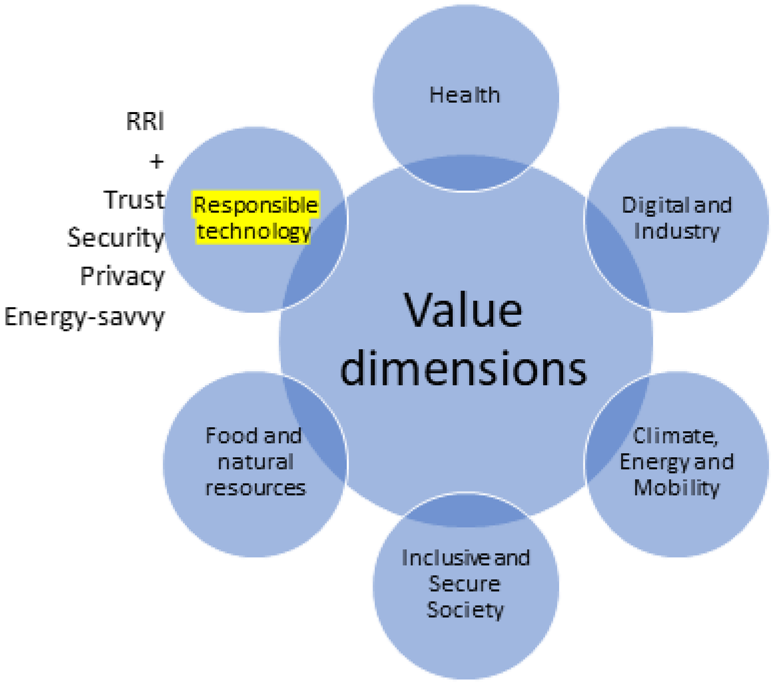


Figure 2-9. Value dimensioning of Smart Networks and Systems

It will be important to reflect on the values brought by technology research in achieving the societal goals. There should be possibilities for experimentation to better understand and assess the long-lasting impact of new technological solutions of society, environment and economy prior to their commercialization.

Responsible Research and Innovation [C1-22] is a promising approach for tackling grand societal challenges. The European Commission defines RRI as an approach that anticipates and assesses potential implications and societal expectations with regard to research and innovation, with the aim to foster the design of inclusive and sustainable research and innovation. It promotes fundamental conditions and ingredients for technology research to achieve the desired societal impact.

The four conditions for responsible research are anticipation, reflexivity, inclusion and responsiveness. Anticipation is about carefully examining both the intended and possible unintended consequences arising from research and innovation activities, including environmental, health-related, economic and social impacts. Reflexivity is about reflecting on and questioning underlying motivations and assumptions driving research and innovation. Inclusion is about involving relevant societal actors in research and innovation activities from early-on. Responsiveness is about aligning research and innovation activities with the new perspectives, insights and values emerging through anticipatory, reflexive and inclusion-based RRI processes.

The RRI ingredients comprise six keys - public engagement, open access, science education, gender, ethics and governance. Public engagement is about engaging a broad range of societal actors in the research and innovation process, including researchers, industry, policy makers and civil society actors. Open access is about making research and innovation activities more transparent and easily accessible to the public, e.g. through open data and free access to publications. Science education is about increasing society's general science literacy, and equipping civil society actors with the necessary skills to more actively take part in the research and innovation process. Gender is about promoting women's participation as researchers and integrating a gender dimension into research and innovation content. Ethics is about fostering research and innovation activities of high societal relevance, that comply to the highest ethical standards. Governance is about the legal and policy frameworks in place to support responsible research and innovation. In addition to these aspects of responsible research, trust-by-design, privacy-by-design, security-by-design and energy-savvy-by-design principles should become a benchmark for future smart network technology research and development. [C1-22] presents a potential set of metrics that could be used to assess these aspects.

Sustainability is well captured in the UN SDGs (see Section 2.3) that represent 169 detailed targets to be met by 2030 and are measured through 231 unique indicators. Although not explicitly mentioned in the goal setting, the achievement of the SDGs and their targets is heavily dependent on ICT and especially future smart networks. From the 231 indicators only seven are classified as to be related to ICT including proportion of schools with access to the Internet for pedagogical purposes, proportion of schools with access to computers for pedagogical purposes, proportion of youth/adults with ICT skills, proportion of individuals who own a mobile telephone, percentage of the population covered by a mobile network, fixed Internet broadband subscriptions, and proportion of individuals using the Internet. They are specific to a subset of the SDGs and do not capture the role of ICT in achievement of the goals. This calls for defining a new set of value related indicators to assess the impact of future networks in contributing to the SDGs and targets. Especially, measuring the benefits through the use of future networks in different verticals is critical in defining their value. The use of future networks as a smart measurement tool for collecting data on the impact is a new design criteria for the development of the future networks. Also, the development of future smart networks fully in line with SDGs requires that the additional burden they create is accurately measured and minimized in terms of e.g. energy and materials.

Concerns from the effect of electro-magnetic radiation on humans have increased among the general public which must be addressed by defining value indicators.

To conclude, at this stage it is too early to define the key value indicators, but it requires future work.

3. Human Centric and Vertical Services

3.1 Emerging applications and use cases

Every industry sector is undergoing profound changes in terms of digitalisation, which in turn will have positive socio-economic impact in the industries and the society. This digital transformation is acting as a strong driver for current and future developments in the area of Smart Networks and Services (SN&S). Networks become the key connectivity technology and infrastructure solution for enabling the Next Generation Internet (NGI) vision and adopting its Human Centric approach. Although Human Centric means empowering the individuals and facilitate the interaction with digital services (B2C) these principles also needs to be applied to the digital transformation of vertical industries (B2B). The Internet evolution must incorporate the EU values that represent our Society and respect the rights of EU citizens. Therefore, any technical solution must attain the necessary levels of trustworthiness, resilience, openness, transparency and dependability while inclusive and respectful with diversity [C3-41], and assuring societal privacy concerns. Telecommunication Networks are also critical and strategic infrastructures that need to be cybersecured with suitable tools [C3-43]. These concepts and values must be translated to technical requirements of future networks and services. Just by looking to what is happening across the different vertical sectors, it is easy to identify common digitalisation trends and emerging applications **in the short and medium term** that are demanding proper advances in the technological domains where SN&S can impact. *For illustration purposes*, five big trends and their impact in different industrial sectors and also in the citizens life, in a non-exhaustive manner, are briefly examined below.

- **Robotic automation**: advances in robotics and artificial intelligence are accelerating the adoption of interconnected and autonomous machines in numerous sectors.
    - In the farming sector, traditional machines like tractors and harvesters are evolving towards autonomous, driverless systems [C3-1]. There is also an increasing trend in smaller autonomous robots [C3-2] [C3-3] (both in the ground and in the air, such as UAVs), to make more efficient and safer all kind of harvesting tasks, analysis of the field parameters, application of phytosanitary products, etc.
    - In the manufacturing industry, robots are becoming both more autonomous [C3-4] thanks to increasing cognitive capabilities, and more cooperative (CoBots) [C3-5] thanks to enhanced human-robot interaction (HRI) capabilities by relying on smart sensing, specialised private networks and advanced AI, fulfilling at the same time strict safety requirements in unfenced environments.
    - Health and Social care is seeing an increasing interest in assistive robots for disabled and/or elder people [C3-6]. These will be autonomous and cooperative robots designed for assisting persons in their everyday life.
    - Future smart mobility services will be enabled by autonomous connected vehicles [C3-7] (both public and private) interacting and cooperating with their environment (V2X) by means of smart sensors and powerful AI systems, which will allow optimal decision taking optimizing the whole performance of the mobility system, e.g. reducing traffic congestion.
    - From automation of inventory tasks [C3-8] to autonomous delivery of goods [C3-9], the adoption of autonomous machines is expected to revolutionise retail and logistics operations in the near future.

  - Maintenance operations in utilities [C·-10] (e.g. power grids) and critical infrastructures [C3-11] are also moving towards automated means for efficiency and safety reasons.
  - Another category of relevant Industrial scenarios for Industrial IoT (IIoT) concerns connected environments, for instance, warehouses integrating Automated Guided Vehicles (AGVs) or Autonomous Mobile Robots (AMRs) and necessary sidelink communications for operating in cooperative and self-organizing way.

  Full deployment of autonomous robots in the sectors and use cases described above require, among others, ultra-reliable and low latency communications, embedded AI at the edge, autonomous decisions, accurate and real-time positioning (both indoor and outdoor), etc.

- **Massive monitoring and remote management**: fine-grained, continuous monitoring of processes is becoming the basis for productivity optimisation and higher efficiency.
  - The future smart grids (energy, water, gas) will need to optimise supply and demand in real time, managing energy flows between millions of nodes, dynamic pricing and transactions, anticipating weather and behaviours of consumers, in a sustainable way with a high level of integrity, security and resilience. This will be only be possible thanks to the capture of high-quality data in real time [C3-12]. In the case of freshwater, it is a vital resource which is becoming scarce, thus it becomes essential to optimise its consumption, detect leaks, and continuously monitor its quality [C3-13].
  - Precision farming demands fine (both in time and space) remote monitoring of crops and animals [C3-14] in order to optimise food production, making a more efficient use of resources and minimising the impact of pests.
  - There is an increasing interest in implementing effective remote healthcare and remote monitoring systems for patients. On one hand, this is motivated by the need monitoring long-term patients and those suffering chronic diseases, thus seeking more cost-effective health and social services, and also higher quality of life for the patients [C3-15]. On the other hand, continuous monitoring of our health, leveraging on wearable devices and AI technologies, makes possible to move towards healthcare prevention systems with early detection and diagnosis which will eventually lead to more effective and convenient healthcare [C3-16].
  - Industrial processes will be finely monitored in real time thanks to Industrial IoT (IIoT), CyberPhysical Systems (CPS) and remote Media-aided Systems, enabling higher optimization levels and predictive maintenance, reducing costs and increasing productivity [C3-17].

  The main requirements in terms of SNS include autonomous operation (ultra-low power), massive Machine-Type Communication (MTC), and large-scale orchestration of networks.

- **Digital twin (DT):** The massive, continuous monitoring opens the door to create virtual models capturing all the details of physical goods and processes, enabling real-time digital representations that allow data-centric management. 3D models and new media technologies (i.e. mixed reality) are also relevant in this ambit to enrich the interactivity. DTs can be applied is different scenarios to improve situational awareness, and as well enable better responses for physical asset optimisation and predictive maintenance.
  - DTs for physical infrastructures/products allow that closely reflect their whole lifecycle: buildings, cities, vehicles, industrial products (e.g. engines), smart grids, etc.

  - Evolving digital twins in energy sector with myriad of functionalities and purposes are paving way for open marketplaces and flexibility markets development, greater observability as well as flexibility of supply and demand.
  - DTs for processes are especially relevant for manufacturing in factories, food production, logistics, and retail.

  DTs open the doors to the full virtualization of the value chains and the creation of rich marketplaces. This is particularly relevant to sectors strongly structured around supply chains like manufacturing, food and retail. This dimension has gained importance due to the COVID crisis, where the disruption of the value chains is creating multiple supply problems.

- **Extreme pervasiveness of the smart mobile devices in Cities** together with the ubiquitous coverage of the mobile telecommunication systems may be employed to monitor large metropolitan areas. An interesting example is represented by the detection of critical urban anomalies, such as unexpected crowd gathering in metropolitan areas (e.g., concerts, football matches, traffic jam). This can be achieved through the collection of information that the different network elements (e.g., base stations, mobile terminals) are exchanging over time, in a totally anonymous fashion for the connected users [C3-44]. More accurate detection and sophisticated events may be considered when including heterogeneous sources of data, e.g., from sensors in the street furniture.

  Smart cities typically combine use cases from the Consumer, Enterprise and the Industrial domains. In the future, several smart city applications will mostly be cross-domain, meaning the data generated by a given deployment can be used multiple times by several applications. Advances in public safety also call for connecting public safety ICT infrastructure with smart city IoT infrastructure to automate emergency handling and enhance the safety of the citizens. Examples of use cases include:

  - Safety and security using video analytics at the edge: Video analytics will play an important role in several smart city use cases. These include crowd monitoring, intrusion detection, traffic monitoring, surveillance, tracking in urban spaces, etc. With an increasing number of video sources, processing the data in central clouds will not scale meaning the use of high-density edge cloud capabilities will be key to provide near real time video analytics addressing requirements from a multitude of applications.
  - Autonomous driving and cooperative ITS: The Society for Automotive Engineering (SAE) defined several levels leading to fully automated driving. There’s a common consensus that reaching the highest levels of automation will call for connectivity and a high degree of cooperation between vehicles on one side and a highly digitized transportation infrastructure (within smart cities), one that supports high throughput and low latency communications thanks to edge computing. Some advanced autonomous driving use cases may even require highly distributed edge computing capabilities.

- **Autonomous and Hyper-connected On-demand Urban Transportation** Millions of people move every day in big cities which are getting more and more crowded. That is why traffic congestion is becoming a critical problem in many of them, making people waste their time being behind the wheel of a car and being exposed to traffic security risks. Future cities will need to free up space for people and at the same time offer more agile on-demand mobility services in order to combat the traffic congestion and have sustainable cities.

On-demand smart mobility services can be achieved on one hand by providing a single platform for combining all mobility options and presenting them to the customers in a simple and integrated manner, that is, easy access to the most appropriate transport services being included in a bundle of flexible travel options, shifting from the provision of urban transport networks (i.e. buses, trams, trains) to what people require on-demand [C3-45]. That will be translated in door-to-door mobility by highly automated chains of different means of transport. However, according to TMForum [C3-46] on-demand services have not yet achieved a real-time 360-degree view of the customers. It is not only a matter of having enough data but there is the need to further leverage on intelligence in a way that not only provides actionable insights, but also complies with regulatory governance about security. If CSPs fail to build the necessary intelligence in their OSS/BSS, their networks will be unable to keep up with the speed and volume of changes required. By using AI, it is expected to unravel the complexity of travel patterns and identify how to reduce the social and environmental costs of transport systems.

On the other hand, traffic management is expected to be highly improved by means of connected vehicles interacting with their environment (V2X), in a higher degree as long as more sensors and intelligence are onboard connected smart public transportation vehicles, so it will be possible to collect more accurate real-time information, to know precisely which parts of the city are becoming congested. Each vehicle, e.g. shuttle service bus, would be acting as a smart object equipped with a powerful multi-sensor platform, computing units, IP-based connectivity, and big amount of data management capabilities [C3-48]. Applications onboard of connected public buses will have to manage real time traffic congestion information received from the numerous sensors in the environment and plan automatically in advance alternative routes, continuously updating information based on predictive data analytics tools. Also novel and scalable cyber-security solutions will be needed to manage millions of connected objects and automotive sensors [C3-49].

These hyper-connected autonomous vehicles will allow passengers to travel carelessly about the traffic (no need of consultation about traffic status or alternatives routes or their safety), which will have a direct impact in the increasing use of public transportation, and will contribute to the following key current societal challenges:

- decrease traffic congestion and passengers idling, which is translated into additional productivity and quality-of-life improvements.
- public vehicles will be transporting more people at a time, increasing road capacity, while providing significant energy savings.
- reduce time to reach travellers' destinations fomenting mobility which can turn in boosting social interactions and economic activities.

From the technology standpoint there is a playground that will combine advanced connectivity (beyond 5G/6G) with its capabilities and embedded intelligence, big data and AI mechanisms to enable future applications.

In the **longer term**, future emerging applications are currently being discussed in several communities. The ones that are described in this subsection are based on [C3-19]. Some examples of such future emerging applications are:

- Holographic media applications: involve not only the local rendering of holograms but networking aspects, specifically the ability to transmit and stream holographic data from remote sites,

- Multi-Sense Networks: include emerging applications that involve not only optical (video, holograms) and acoustic (audio) senses, but as well smell and taste senses.

- Time Engineered Applications: use a communication system that can coordinate between different sources of information such that all the parties involved have synchronized view of the application.
- Critical Infrastructure support applications: support of critical infrastructures that refer to those essential assets that are considered vital to the continued smooth functioning of the society as an integrated entity.

In addition to these emerging applications, [C3-18] and [C3-50] summarize seven representative use cases for ITU-T Network 2030: holographic type communications (HTC); tactile Internet for remote operations (TIRO); intelligent operation network (ION); network and computing convergence (NCC); space-terrestrial integrated network (STIN); industrial IoT (IIoT) with cloudification.

Holographic type communications (HTC)

The holographic display needs to satisfy all visual cues for the human observation of any 3D object, such that it can appear as natural as possible [C3-20]. According to [C3-20], holography is a method of producing a three-dimensional image of a physical object by recording, on a media photographic plate or film. In addition of being optical holograms can record the wavelength (colour) and intensity (amplitude) of light waves, as well as the phase of light waves (perception of depth). It is expected that in the next decade holographic displays will adopt the lenslet light-field 3D through the naked-eye or indeed through augmented reality (AR) and virtual reality (VR) via head-mounted display (HMD) media devices [C3-21]. Holographic type communications (HTC) are expected to digitally deliver 3D images from one or multiple sources to one or multiple destination nodes in an interactive manner. It is foreseen that fully immersive 3D imaging will impose great challenges on future networks.

Tactile Internet for remote operations (TIRO)

Tactile Internet is enabler the real-time control of remote infrastructure, creating a plethora of opportunities and opening new areas of applications within sectors such as Industry 4.0 or tele-medicine. It is expected, that immersive video streaming applications, such as the HTC 3D image streaming, will support real-time and immersive interaction between a human operator and remote machinery. Two typical use cases are described in [C3-18]. The first typical use case is remote industrial management that involves real-time monitoring and control of industrial infrastructure operations with augmented reality (AR) support. The second typical use case is remote robotic surgery. In this use case the surgeon gets a real-time audio-visual feed of the patient and operating room, using a human system interface (HSI), where a master console is installed.

These two use cases require the network to have very low (near zero) end-to-end network latency for real-time interaction, along with guaranteed high bandwidth to support the visual feed. They also necessitate strict synchronization between the various feeds to allow interactive control.

Intelligent operation network (ION)

This use case focuses on the need that future networks will have on intelligent operation capabilities in order to efficiently provide a variety of intelligent services and applications. Monitoring of the network for impairment (or potential impairment) is important, where the key performance indicator is 'network health'. For example, intelligent technologies such as artificial intelligence (AI) will play a key role in feature and pattern recognition [C3-22]. It is expected, that such technologies and models would be trained based upon historical data as well as learning (for example using neural network techniques) from live network operations and other statistics, to establish a human-brain like cognition to locate network malfunctions and faults accurately.

Network and computing convergence (NCC)

The recent advent of cloud computing and network functions virtualization has resulted in the emergence of network cloudification. It is expected that future networks may require multiple distributed network edge sites to interconnect and collaborate with each other. Orchestration capabilities that are computing aware may need to be supported. Furthermore, emerging edge applications (for example, AR rendering, autonomous driving and HTC) are characterized by high mobility and other time-varying features that will require one or more data centres to provide computing resources simultaneously in a coordinated way. Such applications will require the support of intelligent load-balancing among multiple edge sites. Therefore, future networks need to support computing-aware network capabilities, with unified management, control and operation in order to guarantee differentiated service experience with much higher granularity than in current networks.

Space-terrestrial integrated network (STIN)

In this use case [C3-50], a scenario is envisioned where a future seamlessly integrated terrestrial and non-terrestrial Internet framework is applied. The aim of this use case is to leverage the inter-connected multilayered non-terrestrial network (see Chapter 9) to build a parallel Internet network that can peer with its terrestrial counterpart. In this scenario, it is considered that one essential aspect is that future mobile devices (e.g. smart phones, tablets, etc.) will be able to directly communicate with the locally accessible NTN nodes, but without necessarily relying on traditional ground station infrastructures that are constrained by geographical distributions [C3-23].

Industrial Internet of Things with cloudification

Industrial networks enabled by the Industrial Internet of Things (IIoT) are connecting back offices to factory floors and provide the integration from device level all the way through to enterprise business systems, resulting in the automatic operation and control of industrial processes without significant human intervention. These enterprise networks need to deliver superior performance and mandate a real-time, secure, and reliable factory-wide connectivity, as well as inter-factory connectivity at large scales in the future. At the same time, operational technologies (OTs) and IT are converging, where control functions traditionally carried out by customized hardware platforms, such as programmable logic controllers (PLC), have been slowly virtualized and moved onto the edge or into the cloud.

Representations of the network requirements

These seven representative use cases have been further evaluated in [C3-18] according to five abstract network requirement dimensions, widely discussed in [C3-24], namely: **Bandwidth**, **Time**, **Security**, **Artificial Intelligence (AI)**, and **ManyNets, see Table 3-1.**

Table 3-1 Abstract dimensions with relevant network requirements, from [C3-18]

Abstracted dimensions	Relevant network requirements
Bandwidth	Bandwidth; capacity; QoE; QoS; flexibility; and adaptable transport
Time	Latency; synchronisation; jitter; accuracy; scheduling; coordination; and geolocation accuracy
Security	Security; privacy; reliability; trustworthiness; resilience; traceability; and lawful intercept
AI	Computation at edge; storage; modelling; collection and analytics for network programmability and management
ManyNets	Addressing; mobility; network interface; multiple RATs and heterogeneous network and computing convergence.

The relative scores for these five dimensions, ranging from 1 to 10, are shown in Figure 3-1, where all the scores are given according to the relative importance of a specific network requirement: 1 to 3 are for relatively LOW requirement; 4 to 6 are for MEDIUM requirement; 7 to 9 are for relatively HIGH requirement; and 10 means EXTREMELY demanding requirement. The detailed and absolute values of the requirements associated with each use case are provided in [C3-18].

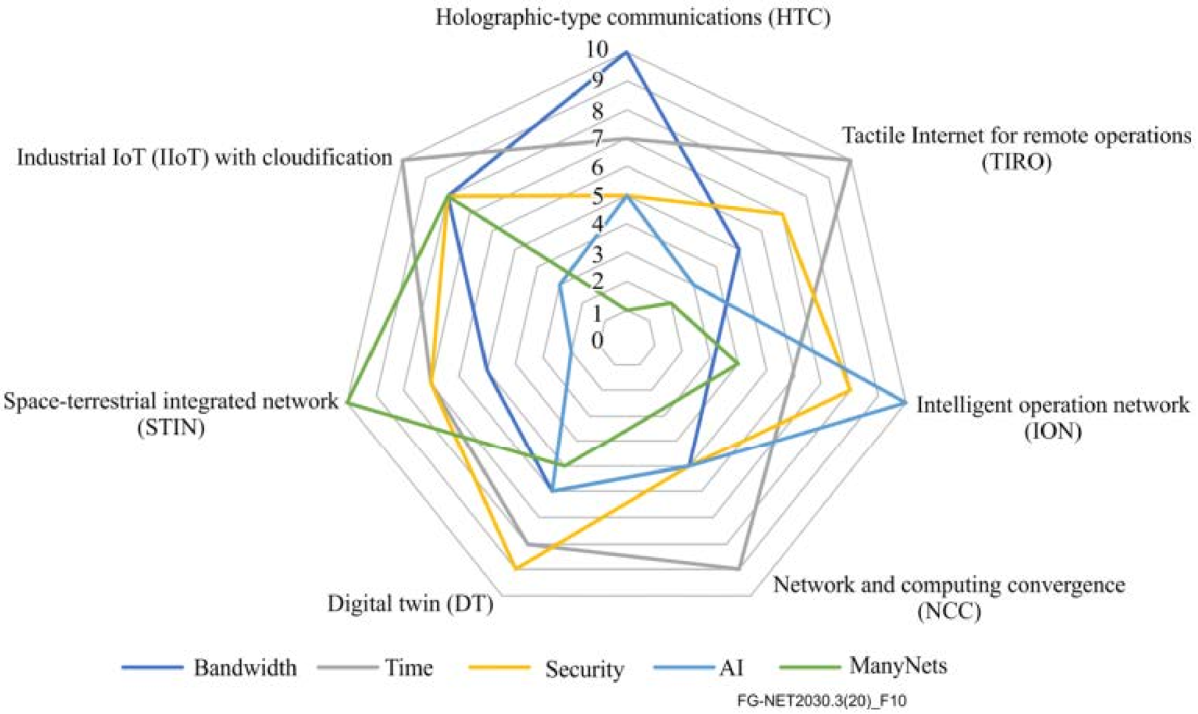


Figure 3-1 Relative network requirement scores for the seven representative use cases – reverse spider graph representation, copied from [C3-18]

3.2 Digital Service Transformation

Telecommunication networks have been regarded an essential pillar of any society's infrastructure to progress and sustain its economic growth. Millions of people rely on the diverse services offered by telecommunication networks, which is part of a real Digital

Economy. These services require underlying network technologies to support higher workloads and increasing traffic volume while reducing the overall network operation costs. Each new generation of mobile networks (GSM, 3G, 4G/LTE, 5G) goes beyond a simple increase in network speed or reliability, and instead it brings unique new service capabilities for people and vertical economic sectors. Disruptions driven by emerging and maturing technologies are impacting businesses and society with increasing pace and depth, and this trend will accelerate in the next years. Future networks, evolution of cloud computing, any type of connected object and the strategic use of data and analytics are the foundations of the digital disruption.

The envisaged trend is the convergence of these foundations and their complete fusion in an ICT continuum platform. We already perceive this trend. Public clouds have become mainstream today and cloud computing evolves towards hybrid models, combining private clouds and being extended towards the micro edge clouds with lighter virtualisation techniques. Network softwarisation also reinforces this convergence of networks and IT systems. At the same time the number of connected devices, tablets, wearables and IoT elements is growing to at least 50 billion by 2020 in the most conservative forecast. As this range of compute capability in the cloud and at the edge becomes increasingly connected though flexible networks, we see the emergence of an *ICT continuum*.

In the same way that we observe insect swarms (like bees) coordinating their interactions, we expect the clouds, networks, IoT and data to enable multitudes of entities and devices to combine to form dynamic and intelligent collectives [C3-25]. One feature of this will be localised and temporal interactions between compute nodes that combine resources to achieve a task greater than can be achieved with those nodes operating in isolation. The individual limited computing capacity of objects is complemented and supplemented by their connection to other objects in relevant communities. This new computing is emerging as a cooperative interaction between individual entities, each with their own autonomy, but working together for the benefit of the collective community. One example of this Swarm Computing would be autonomous vehicles, each capable of acting independently, but also interacting with connected objects around it and to centralised information and control centres to optimise traffic flow and improve safety. It will give rise to connected smart systems that are able to share information and autonomously regulate their performance in a concerted fashion, with the objective of optimising results, solving problems and mitigating detected risks. Depending on the specific business use cases, data will either be processed in a distributed manner by algorithms hosted in hybrid clouds or be processed locally using algorithms hosted within or close to the connected objects and connected robots. Real-time data analytics and AI/ML will have a major impact in the automation, optimisation and flexibility of connected ecosystems and the collectively offered services. This will be especially so in industrial supply chains as they become increasingly collaborative and responsive to market demands. This will be part of next industrial revolution to create intelligent networked enterprises and service ecosystems.

As the world becomes ever more digitally and globally connected, industries of any sector (Healthcare, Energy, Manufacturing, Telecom, etc.) are experiencing a digital transformation [C3-26]. Communication Service Providers (CSPs) who evolve quickest and are most able to adapt to this new digital economy, will be the ones to thrive. The traditional methods of revenue (voice and data services) are slowing and the price pressures over these services continue. CSP are suffering in both consumer and enterprise domain and are viewed as simple pipe providers. CSPs are seeing their profits stagnate, while watching the over-the-top (OTT) players erode their revenue [C3-27]. The changing trend of their users moving to alternative OTT services was one of the reasons for their decision to start on the journey to digital transformation. CSPs enter in this digital economy selling new services that can be delivered and management over digital channels in areas such as consumer entertainment, mobile banking services, autonomous transportation, etc. The business opportunities are huge and to reach a market outside the consumer's area, the CSPs need to create strategic alliances with the vertical stakeholders to build and offer B2B2X (Business-to-Business-to-X) proposition in vertical sectors.

However, the telecom community continues selling a lot of technology rather than solutions that enterprises want. Digital transformation is more than a technology drive, it encompasses end-to-end IT processes, automated operations, infrastructure management, human skills and corporate culture [C3-28]. CSP need to digitally transform the internal organisation – we have seen actions on this direction with flexible DevOps, Zero Touch management – and also adapt the external language to properly communicate to vertical sectors.

An opportunity for CSP is to offer their new networks as self-service platforms by using a high degree of operational automation and complete customisation. Service innovation and customer experience are the main drivers for their digital transformation. Future services for connected enterprises and people will be context-aware, immersive, omnipresent, intelligent and autonomous for real end-user experiences and will bring new technical challenges to ICT infrastructure. In order to support new types of services, operators need to upgrade their IT stacks and service operations. Over decades, CSP's top priorities are network's reliability and performance. Telecom networks must be always-on, and guarantee continuous services regardless used technology, since society and any type of business depend on reliable communications.

CSPs have built up many sophisticated systems and strict operational processes around telecom networks following rigorous standards for stability and quality to the point where they guarantee their high reliability. The problem is that to adapt this legacy to be ready for the digital economy and smoothly manage the end-to-end ICT continuum is a real challenge. Legacy BSS and OSS need to evolve to real-time, automated open platforms that are fully virtualised and use microservices to deliver this agile approach to provide service assurance and the way to do business [C3-29]. These end-to-end management platforms should be on one hand modular with a high level of resource abstraction so that can be based on multiple vendor combinations and on the other hand, also offer service capability exposure functions via open APIs to enable CSPs to partner with enterprises in vertical sectors. This will allow new digital services that combine CSPs offerings with partners' specialised domain capabilities (sophisticated content, IoT data, immersive technologies, etc.) to create innovative offerings. Bearing in mind the changes due to this ICT continuum and dynamic digital environment, there is a need for a new generation of service and network solutions for both operational and business management that enable the partnership of CSPs and vertical stakeholders while keeping an overall service assurance and carrier-grade reliability and performance.

3.3 From Software-Centric to Human-Centric Internet Services

The continuous technology evolution, the habits of the digital consumer and the volatile market, poses big challenges for the enterprise to become more competitive and profitable; this will not be possible if organisations do not bet for innovation. Human Centric innovation supports customers and communities in creating a prosperous and sustainable future where people are always connected. Human centric means empowering individuals and facilitate its interactions in digital services [C3-42].

The evolution of Internet pushed by Europe must incorporate the EU values that represent our Society and respect the rights of EU citizens. Any technical solutions must achieve the necessary levels of trustworthiness, resilience, openness, transparency and dependability while inclusive and respectful with diversity [C3-41]. Furthermore, GPDR concerns need to be addressed, guaranteeing privacy to citizens. These concepts must be translated to technical requirements of future networks, a task increasingly hard because of the increasing trend of cyber-physical social systems (CPSS), an area where we consider the impact of individual human beings that are an integral part of the systems – both as a user as well as a source of disruption. Today the concept of innovation only as the incorporation of a set of tools or management solutions is erroneous. The techno-social revolution should help customers to become more innovative and to help in the evolution of the digitalisation. As discussed in the previous section, we envisage a globally-connected continuum platform to enable new future digital services. This continuum must provide users with greater level of control, be more transparent in interactions with digital services (for instance search AI-based algorithms or data computes resulting in non-biased results presented to users) and even introduce some ethical values and behave with certain social capabilities. New innovations are about not only digital transformation of industries and businesses but also achieve a better social inclusion.

These new paradigms will lead to a flexible and programmable architecture based to satisfy the large diversity of use cases and applications the different vertical industries will demand. New and heterogeneous services as the autonomous cars, Industry 4.0, smart cities, immersive and augmented reality, etc. will require networks beyond-5G to quickly adapt to new demands and provide more control of the network services. The transformation from hardware-centric to software-centric networks is already in place. Networks Functions (NFs) are moving from monolithic equipment's into programs running in virtualised computational pool of resources. This will also mean that networks beyond 5G will be easier to upgrade to, reducing the expensive current ones based on the replacement of physical infrastructure. Software-centric networks may become more democratic networks thanks to the open source developments enabling multi-vendor interoperability.

In the same way, the future generation of networks (beyond 5G/6G) will go one step further from software-centric towards the concept of human-centric. Humans will be the central point and everything will revolve around humans. Human-centric business process management is an approach that considers human skills and activities first and uses automated functions to support them. Benefits can include reduced risks, higher rates of compliance, enhanced management support and improved interaction with users.

We envision that the future networks will be human-interfaced in the sense that people will interact with them as they currently do with other humans. It can be considered that the interface potential for network management, both from the perspectives of tech savvy network administrators or a non-expert end users – is mostly untapped. The network interfaces will recognise not only the order via commands or voice as nowadays but also gestures or moods and will be able to react based on them as it is done in a human-human communication. This extremely valuable information will be used to provide fully customised services as, for example, show a list of restaurants in the area the user is and based in his preferences and health status, thanks to the fact that the network interface can identify based in his vital signs, and the user behaviour while being stimulated by food smell. While several personal assistants are already widely popular, their usage still comes with significant frustrations and other blocking points. Their enhancement through continuous advances in areas such as NLP and intent-based networking, may unlock the ability to actually be able to have a "conversation" with a network management entity as if with a human, proving as de facto human-centric interface. Additionally, one must consider such potential across different specific domains or industries (e.g.sending of commands or orders to robots / Cobots in manufacturing, communication with personal assistants or robots for elder care, etc). Finally, this does not necessarily contracts the trend for zero touch automation / service automation, in the sense that humans will still have to provide input to the (increasingly intelligent) network system, just in a more intuitive way for increased user experience.

And this is only an example. Another service that goes beyond would be the "mind to mind communication" between human brains. In terms of services, right now the list of potential applications in beyond 5G networks fall in the field of science fiction, being one of the most present the "mind to mind communication".

This user centricity will be felt as well on the interfaces between human and machine (perceived as the interface to the intelligent network system). This interface will provide users (service designers) the ability to create new services (features) in a simple way, triggering changes on the whole network substrate.

Users will demand the future networks, beyond greater bandwidth or almost zero latency, global coverage and always-on connectivity. In this sense, we envision that humans will be part of the network, carrying microchips under the skin, so called e-skin, interfacing with multiple sensors applications as part of the service. The sensors will have the ability to compute data and perform cryptography in our bodies, or the ability to transmit and receive digital data and talk directly to machines in their digital language. There are already new companies [C3-30] targeting ultra-high bandwidth brain-machine interfaces to connect humans and computers.

Moreover, the limitation of the connectivity in certain areas, could be enhanced with the concept ”follow me network” identifying the user location, and adapting the pool of resources infrastructure improving network service, in terms of network coverage, whereby services are following users by creating ad-hoc networks with fixed and unmanned infrastructure. On the other hand, it will be needed to adapt the networks in terms of adaptability of resources to assure the service, but also identifying the limitation that the coverage might have, be able to deprive some service capabilities to the most basic services, to assure the quality of the network and the best use of resources.

3.4 Services Everywhere, Infrastructure No Limits

In today’s society, people are relying their entire daily living on the use of digital services and apps. Digital users are becoming increasingly dependent upon Internet services and moreover connectivity to those services – and this trend is clearly set to increase. The following Figure 3-2 depicts the global digital population (in millions) worldwide in January 2018.

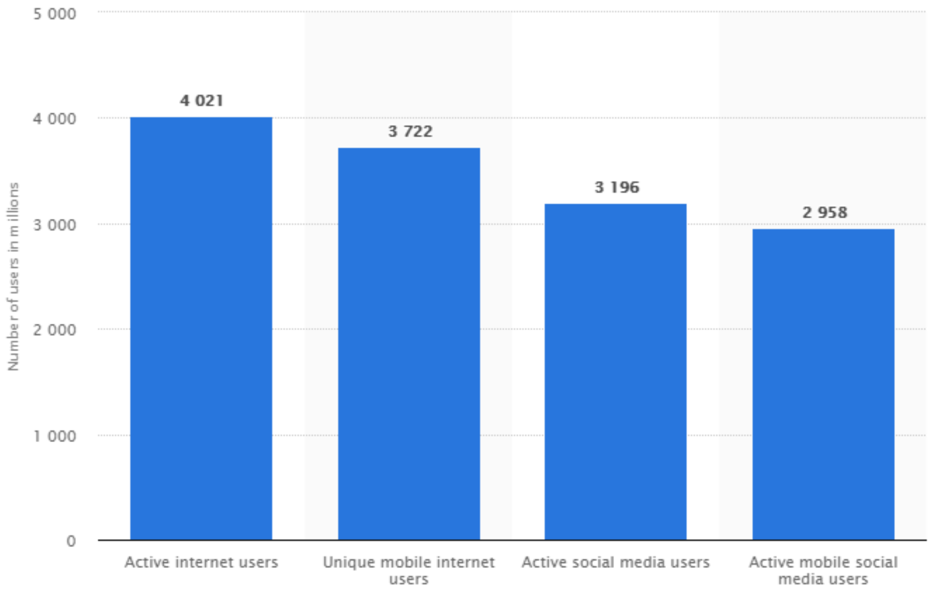


Figure 3-2 Global digital population [C3-31]

Moreover, we have entered in an era, that we are not conscious of the necessity until the industry creates a new product or service that enters in our day-to-day life and rapidly becomes totally necessary. The benefits to the society have become obvious during the last years, even the users that were not born into the digital world, and has ”survived” without the use of all these Internet services for their work and life, now we raised the question how did we previously survive without it? The main innovation is not just only to innovate with the youngest population, also to empower a more mature generation, and support them with easier use of the technology.

The potential and power of technology focus on the optimisation of the resources from time consuming, energy saving, reduce uncertainty, predict problems, generate experience or easier interaction, in order to provide a more productive environment. Around all these processes, technology is a must. This strategy has been possible with the exploitation of our personal data, and behaviour that we have assumed the loss of privacy and confidentiality when we accept the terms of privacy when selecting a new service, that will be used to create analytics to enhance decision and policy-making, and also to sell this information to their customers.

The digitalisation is making disruptive changes to traditional industries, products are offered on top of operators' networks. This necessity is increasing, with the digital connected society the market segment where any type of service is bringing a new level of technological sophistication to be able to access to all connected devices (robotics, immersive reality, IoT devices, autonomous vehicles), and the services acquired through the Internet: booking reservations (hotels, concerts, congress), financial, communication or social functions. The current IT infrastructure is not designed to support this intelligence everywhere; some architectural transformation and renewed focus on hardware and exploitation of communication regulations will empower companies to build the new future communication system.

At the same time, where connectivity brings convenience, it also brings challenges to cope with the increasing demand the infrastructure should cover. There are several emerging technologies that will become the backbone of the Internet, from the cloud to the edge, and everything in between. Currently there is a trend to get closer to the user, such edge architectures (MEC) will speed the maturity of technologies giving resources close to the user without relying on the core network, or personal data consolidator that will make easier the access to a user/client data. Customer hardware and hardware accelerators will meet the computing demands of intelligent environments. IT Infrastructures need to be developed to reach into the dynamic physical environments they want to serve – and it needs to happen now. Small Cells (SC) are playing a vital role in the future growth of private networks to better serve densely-populated areas or dedicated critical enterprise services, as part of the orchestration framework can reduce those costs by adding capacity to the network instead of new use of spectrum. The Distributed Antenna System (DAS) with SC can serve more than one mobile operator (neutral host model), so it can provide coverage and capacity for multiple carries, providing multi-tenancy. The new SC can increase the wireless backhaul, capacity and speed, locating the antennas close to the end-user and dynamically allocate the network capacity in the flexible C-RAN (Cloud-RAN) while delivering substantial cost savings in comparison to traditional distributed RAN, with the split in centralized units (CU), distributed unit (DU) and radio unit (RU). SC will be at a lower power and can eliminate the need of wire connection to many equipment. The use of shared and unlicensed spectrum lowers the cost per bit significantly, although it has the lack of covering large distance, so to provide well coverage carries need to rely on license spectrum to provide the needed range, or to be coverage by macro cells in the wireless backhaul. The combination of both will improve the network capacity and user experience, by relying in a SC controller-based architecture of those resources. To achieve the required network densification and coverage expansion to support no-limit connectivity scenarios, it is necessary to enable more affordable SC deployments of private networks by the usage of white-box approaches with open hardware and open source software platforms at RAN, Edge and Network Management. The rise of private and enterprise networks to support industrial and engineering use cases will be a key driver for telecom industry increasingly enabled by emerging shared spectrum. There will be not only the established CSPs controlling the massive deployment of SC but new actors such as private operators and neutral hosts will be also controlling many new cells for efficient indoor scenarios in Manufacturing, Retail, Transport, Smart Cities or Healthcare sectors.

3.5 Network-Unaware Vertical Services

5G networks were conceived to support advanced use cases from different vertical sectors with different network requirements by means of slicing the physical network into several logical networks or slices, each slice tailored for a different vertical use case. This enables offering a flexible solution that allows the optimal configuration of necessary resources to serve a customised service, empowering different verticals, such as high-quality multimedia on demand, health (remote surgery use cases), manufacturing (digital twins in Industry 4.0) or automotive industry (autonomous driving or connected cars).

However, higher levels of abstraction are envisioned to be in place towards the verticals to make this process for them in a fully automatic and network unaware mode. A vertical stakeholder, which could be, even more in the coming years than now, not necessarily a business entity, but an individual, will not want to speak the language to request concrete network requirements for each of their requesting network services for his/her vertical applications.

Those higher levels of abstraction will be part of a network agnostic automation process that will be needed to lower the barriers to satisfy coming business and users' necessities. That network agnostic process shall include also an automatic and transparent mapping to the network service consumer, who does not care about networking issues or requirements. The vertical applications (layer 7) will have to be totally network-unaware which turns out into a full automatic from human to network translation process. Future mobile intelligent applications will learn in the computing infrastructure and get balance deployments from the edge to core. The usage of AI/ML techniques will continuously improve application service delivery.

This will have to be accompanied consequently in the lower layers by higher degree of automation also in the orchestration processes that will need to combine vertical applications and network applications orchestration together as a whole. The future service architectures will have to comprise an intent-oriented service definition over abstracted infrastructure (advance models are needed), real-time telemetry of services and massive correlations, proactive adjustment of parameters to meet service intents. Literally, the network will be 'always-on' and automatically carry out 'follow-me' actions (as service motion concept, as pointed out in previous sections) to maintain QoE defined in composed SLAs. The whole architecture will embody a closed-loop structure for service life-cycle management. This continuum will be a self-driven platform and will also perform proactive business actions such as fault isolation, prevention across multi-layer and multi-vendor environments, fraud detections, deal with trust areas, hybrid orchestrations for a global optimisation of services at scale. The combination of an intent-driven approach and AI/ML techniques for managing both network and services will bring enormous gains in service efficiency (doing things in an optimal way, faster or better), in service effectiveness (doing the right task and achieving goals) and in functionality (doing new things previously not possible, business edge).

In addition, the future ICT continuum platform (compute and network) will intelligently learn the network environment and historic data, and dynamically adapt to a changing situation and enhance their own intelligence by learning from new data. This will become a big data problem that will need to be solved together with the associated services. It will learn and complete complicated tasks, such as redistribute workload, intelligent placement, traffic load balancing associated with link utilities, autonomic network operation, keeping dynamic flows in large-scale networks, etc. The platform will be able to even predict the future network situation for proactive controlling services performance and delivery reliability across multiple networks. This is a step towards self-driven networks for advanced network-unaware services.

3.6 Extreme Automation and Real-Time Zero-Touch Service Orchestration

Future networks envisage going more and more into the need to support a ”hyper-connected world” where more challenging performance requirements are expected beyond 5G ones towards: an always-on and ultra-fast connectivity, with full world coverage everywhere and massive machine-to-machine communications from all kind of different devices. This is accompanied with a complete digital transformation in all the society in which the globalisation, trust, sustainability and automation are presented as main pillars [C3-25].

There is a clear trend in the coming years towards the maximisation of automation in all processes and interactions (except those cases that could negatively affect customer experience). In a few years, social machines, smart contracts and other types of more advanced interaction will be a reality, in the way that some machines will be indistinguishable from people from the perspective of business process and interactions, with higher capacity of decisions, orchestrate common actions, make requests, etc.

This expected evolution should be taken inevitably also in the telco world as part of network operations in line to AIOps trend that offers huge opportunities for communications service providers (CSPs) to deliver better services and reduce Opex [C3-39]. Future networks will require higher demands on real-time network service management and a higher degree of automation, which will be crucial to increase cost-efficient operation.

The relevance of this research work needed for automation beyond 5G is also illustrated that ETSI NFV has started a specific group on ”Zero touch network and Service Management” (ETSI ZSM ISG) that plans to focus on some of the aforementioned challenges in relation to network and service management, in order to allow all the operational processes to be executed automatically [C3-32].

Several technical dimensions will contribute to increase the required level of automation towards extreme automation, which at the end will be driven by accomplishing real-time zero-touch orchestration including self-healing, this is, the capability of OSS remediation, and self-organised network services orchestration.

Future services of any vertical sector must interoperate with all management platform capabilities. Therefore, an open APIs should be offered to allow telecoms to partner with key enterprises in vertical sectors to properly address B2B2X scenarios. In addition, there is a need of a flexible and extensible universal template for service design, on-boarding and lifecycle management. The challenge is how to automate service behaviour updates, keep interoperability between diverse services without breaking overall end-user experience.

3.6.1 Enhanced policy management including huge data analytics

Greater policy-driven autonomic support will be required for network automatic self-healing and self-organisation [C3-33] with no (explicit) human intervention. Policy management will define operations that can be used autonomously by different network domain controllers providing activation and remediation functions, at the same time that cycle time and cost will be reduced. AI/ML is also needed to increase the degree of automation as an approach to enable the transformation from automatic functions (relying on handcrafted rules and hard coded logic) to autonomic functions that constantly and automatically learn from, and adapt to, rapidly evolving network behaviour. These data-driven functions eliminate the need for network tuning, detecting and resolving network incidents at much higher precision and significantly faster, in many cases even before they happen [C3-34], which enables a proactive assurance, based on a performance analysis, root-cause analysis, troubleshooting and fast resolution functions.

3.6.2 Artificial Intelligent driven orchestration and network management

Intelligent network management mechanisms using AI/ML combined with analytics at the service orchestration platform will further optimise the network operations experience towards automation and zero-touch orchestration. It will enable a much more informed elastic management and orchestration of the network, often allowing proactive resource allocation decisions based on heuristics rather than utilising reactive approaches due to changes in the load. This approach would be in line with the goal of ETSI ISG called Experiential Network Intelligence (ENI) [C3-35], which proposes an engine that adds closed-loop AI/ML mechanisms based on context-aware and metadata-driven policies for network management to more quickly recognise and incorporate new and changed knowledge, and hence, make automatically actionable orchestration decisions.

3.6.3 Cloud-native management and serverless approach applied to NFV orchestration

Cloud native refers to software built to change, scale, resilience and manageability. Cloud-native data is stored and structured in ways that encourage flexibility and gets comfortable with micro-databases that requires new levels of automation and self-service. Like cloud-native apps, cloud-native data platforms should scale up and scale out.

On the contrary, VNFs so far, have been implemented to be managed as monolithic applications by the current NFV MANO platform solutions. It is expected that the application of a cloud native orchestration approach to NFV MANO contributes greatly to increase that level of automation future networks is requiring towards real-time zero-touch orchestration or extreme automation [C3-36].

However, this is not a trivial process. Most of the existing NFV MANO orchestration solutions now follow the principles of ETSI NFV specification, which is focused on virtual appliances-based solutions, and lacks the information and guidelines to support the cloud-native architecture and environment. ETSI NFV ISG proposed some modifications in their proposed architecture towards the support of a cloud native approach [C3-37] and the working group about Software Networks within 5G-PPP discussed this thoroughly for multiple vertical services [C3-40].

The overall idea is that VNFs will have to be broken down into smaller microservices, and deployed as containers in the edge and both the public and private clouds. Leveraging Continuous Integration and Deployment (CI/CD), these microservices containers will be orchestrated and deployed with automation. In addition, Serverless computing as abstract execution model where the infrastructure provider runs the server, enables the connectivity and dynamically manages the allocation of resources is a trend to consider here. The independent software vendors who used to produce full-fledged network functions now become the vendors of smaller microservices. In this way, it is expected to achieve:

- **Auto-provisioning:** This is the management of resources automatically facilitating on-demand, self-service, programmatic provisioning, and releasing of resources. This will enable network services to run smoothly with on-demand allocation of resources directly from the VNF packages, and automatically handling the task of data analytics, and releasing the resources back to the pool when the service is finished.
- **Auto-redundancy:** This will enable to minimise failure risks automatically. The network services will be expected to be inherently resilient to failures. They automatically handle the outages and enable corrective actions. In the event of failure, the process instantly moves from one data centre to another without interrupting the service. This is executed so quickly that the service consumer does not even know. In case of the occurrence of a partial outage in one data centre, the VNFs/Network Services will continue running seamlessly.

3.7 Service Injection Loop

The use of services in the NFV ecosystem provides existing network functions as on-demand services for Enterprise applications hosted within the cloud. Those services are created as a dynamic path to access virtual resources created on-demand and with the flexibility to be deployed in different locations. The provision and configuration of services are managed by the NFV orchestration layer in real time, to set up (and turn down as needed) suites or catalogues of connected services that enable the use of a single network connection for many services, with different characteristics, based on the available VNF to form the SFC.

Architectural microservices provide modular, distributed software components that can be deployed in any environment with a standardised infrastructure, allowing distributed applications to be installed on a cloud infrastructure while maintaining maximum flexibility. Digital transformation is requiring significant changes to the network to be reliable and provide a high level of service to comply with Quality of Services (QoS) policies.

The creation of services should be reinvented for the new digital area; the way that services are provided today, should be flexible enough and tailored to customer needs to deliver a delightful customer experience. Using NFV, SDN and cloud computing create a virtual, cloud-based network that has the flexibility to automatically meet the service requirements of a rapidly growing set of network-based services.

Now, the innovation should be driven not only in the network transformation but also in the creation of new services, this will unlock the potential of digitalisation. Like the current trend on the DevOps approach and the agile development philosophy and operation of network services should be transferred to deliver the next generation of services. The use of predictive models is continuing to disrupt the way decision-making is occurring. This approach will create massive tailored service focused on the real need of the customer and specifically designed/customised for the user.

The main goal will be that services must interoperate with the platform capabilities, and be able to automatically adapt to the current needs and enhance the user experience. This approach can be followed by internally identifying and analysing the use of the service in order to recognise the limitations of the service. Human Centred Design (HCD) processes for interactive systems provides requirements and recommendations for design principles and activities throughout the life cycle of interactive systems.

Operators and infrastructure managers are already using AI/ML-based technologies to automate the use of the infrastructure (AIOps). The requirements must be evaluated against existing operations, maintenance processes and tools. The new services should be enhanced with AI/ML also with highly specific requirements as mentioned before such as unifying strategies including carrier-grade, performance and operational capabilities.

Current work is focusing on the challenge of integrating new services in the NFV platforms, to automatize interoperability without breaking the components, and make those services available as a ”service store or repository”. A potential approach could be achieved through a Metamodel [C3-38] for services, to model, testing and onboarding in a runtime environment to become self-service platforms. To link them on demand, requires a new way of describing the entire platform and all its constituent endpoints at the right level of abstraction. The metamodel should be a single point of connection to support end-to-end automation of testing, onboarding and the lifecycle management process that can dynamically ingest and combine a broad range of microservices to compose a larger and complex one.

The customisation of the services based on the previous description will provide several benefits:

- Better user experience (UX), thanks to the Human Computer Interaction, based on consistency of solutions across multiple interfaces (mobile, traditional systems and human interfaces).
- Support inter-data connectivity between verticals slices to define new knowledge for defining automatic new services requirements, imposed by the service limitations and user interaction.
- Automatic test and validation of requirements and functionalities in several domains (integration, deployment, on-boarding, lifecycle management).

For the onboarding phase, services will be complemented with testing files to be validated by the NFVO for a fully operational use, across multiple infrastructures. This approach will involve new business revenues to monetise the use of service by the user: pay-for-what-you-use services. It could also create models rather than sell a specific product: the most valuable part is going to be the output that the service offers.

4. System Architecture

4.1 Evolution of Networks and Services

Distributed computing has taken a significant step forward with the development and utilization of the Internet in many industries, pushing the digitization of processes and opening opportunities for creating or improving many business-to-business (B2B) and business-to-customer (B2C) processes. It does so, however, on the back of an Internet, whose core design is almost 50 years old, starting in the 1970s on very basic assumptions of an end-to-end connectivity between two remote machines, usually denoted as *client* and *server*. Inter-domain connectivity, enabled through the overall IP suite, allowed for reaching any machine through a multi-tier architecture of autonomous systems (ASs). This basic principle, unchanged to this day, had to shoulder the burden of *service routing*, i.e., associating a request to an instance of a service name. This had to be mapped to a combination of hostname and service path and, ultimately, a machine locator, i.e., IP address, bound to that service name. These limitations are currently being addressed in the evolution of the future of the IP protocols, with different protocol innovations being pursued in different frameworks (e.g. C4-10][C4-11][C4-12][C4-13][C4-14][C4-15])[2].

While unchanged in principle, many things have evolved from this basic picture of Internet connectivity. In the following, we differentiate three aspects, namely the *nature of communication* over the Internet, the *nature of services* (and their relation) and the *nature of provisioning* in the serving endpoints that are being reached via the Internet.

The *nature of communication* over the Internet has changed significantly from the single-client-single-server model. Today, many such servers are hosted in large-scale *data centres*, exposing services via a data centre's internal routing mechanisms to the wider Internet – here, the client communicates to the data centre (over the Internet) rather than the server directly, said data centre serving as a *point of presence* (PoP), enabling a service provider to host the service without having to own or operate their own resources. In recent years, those PoPs have been moved closer to end users in an attempt to reduce costs (e.g., for inter-domain transfer) as well as latency (by being closer located to the relevant users), particularly for services such as over-the-top (OTT) video or social media. This move has been driven by

---

[2] We expect that the increased impact of vertical (e.g. society) requirements will further constrain the evolutions on the Internet protocol.

large-scale service providers, such as Google and Facebook, but also by *content delivery networks* (CDNs). These companies have deployed their own PoPs and, by selling excess capacity, have established themselves as large cloud players. By pushing data centres towards the network edge, communication in the Internet has significantly concentrated on the customer access networks with, for instance, an estimated 61% of Asia Pacific Internet traffic expected to being served through CDNs alone by 2021 [C4-01]. Netflix's estimated 15% share of the Internet traffic is mostly served through localized PoPs [C4-02]. Extrapolating this to other content platforms (e.g., Amazon, Disney+, as well as country-specific platforms such as BBC iPlayer), we can project the amount of traffic originating and terminating in customer access networks to be easily around *90% of the overall generated traffic* downstream to end users. In essence, **the nature of communication has moved from servers towards services, the realization of which, in turn, moves closer to the end-user.**

When it comes to the *nature of services*, advances in software engineering broke up monolithic code blocks that served services with a single locus of consistency into smaller, independent pieces of cooperating *microservices*. Hence, the centralized client/server model has evolved into a *chains of (collaborative) transactions,* with typical challenges like *atomicity*, combined *resource management*, and *execution correctness* of the transactions. This, in turn, has created the desire to extend the basic DNS+IP service routing in place today by network support for such chaining, as witnessed by the ongoing Service Function Chaining (SFC) work in the IETF [C4-03]. This application-level trend goes hand-in-hand with the realization that a network cannot just limit itself to blindly forwarding packets; it needs to take an active role in, e.g., providing security (firewalls), assist in service routing (load balancing, redirecting), or traffic shaping. All this is, essentially, software that needs to operate on a stream of packets, just like many application services do. In consequence, this increasingly establishes application- and network-level services at an equal footing with utilizing the increasing *in-network processing & computation* capabilities. However, at present, a proper control framework for such in-network processing is still missing – while IETF ANIMA [C4-04] establishes a virtually separate control plane, it hides compute resources behind application functions. Some work has started, e.g., the recently established IRTF COIN (Computing In-Network) research group [C4-05] or IETF FORCES [C4-06] (separation of forwarding and control elements). Overall, **the nature of services has moved from monolithic services towards chains of collaborating microservices, at both application- and network-service level.**

Along with changes in the nature of services, the third aspect are changes in the *nature of service provisioning.* While microservices (networking or application-level) can be provisioned directly on bare metal, *virtualization* has opened up new opportunities. Since a long time, it has been driving the hosting model in clouds and PoPs; the evolution towards more lightweight virtualization approaches, e.g., through containers or unikernels, has increased the dynamicity of serving instances on a pool of available compute resources. Large-scale services, such as Gmail, YouTube and others, use this approach by dispatching service requests at the DC ingress to dynamically created micro-services, which in turn are based on container-based virtualization. The 5G community has realized the power of such flexibility and enabled its 5G Core specifications to use service-based architecture (SBA), which adopts the micro-service model for realizing vertical industry specific control planes over a cloud-native infrastructure, within a so-called *telco cloud*. *Service routing* becomes key here for the dispatching of service request, e.g., to establish a data traffic session quickly to the right service instance in the data centre of the mobile operator. Given proper service routing, the data centre can easily be distributed, giving mobile operators a decisive competitive advantage over conventional cloud operators in localizing services, as already observed above as a trend in the Internet. We observe that **the nature of service provisioning has changed towards virtualization, for both application services and network services.**

Many major Internet players, such as Google, have long recognized this trend and focused their attention on improving service access in the customer access network (to their POPs hosting their services). QUIC [C4-07], as an example, initially was implemented in the Chrome browser on top of UDP as a differentiator for Google services; standardization in the IETF only followed the initial deployment in millions of Chrome browsers. The intention here was clear, namely, to improve the invocation of services that support the (initially proprietary) extension, with the access network becoming even a dumber pipe and utilizing service end point instead for everything from name resolution to service invocation.

Complementing virtualization of service elements, *network programmability* has enabled programmatic changes of forwarding operations post-deployment. In consequence, programmability enables the functionality of all/some network elements, network functions and network services to be dynamically changed in all segments of the network infrastructure (i.e., wireless and wired access, core, edge and network cloud segments. Therefore, network programmability supports different and multiple execution environments at the forwarding plane level, those execution environments enabling the creation, composition, deployment, the actual execution and management of network services and/or network functions.

The *digitization of processes* has been proliferating in many industry branches, significantly diversifying the use cases for communication technologies beyond the often consumer-oriented focus of typical Internet services (such as social media or OTT video). Communication technologies have penetrated manufacturing, supply chains, vehicular engineering, health technologies and governmental services, among others. The Internet-of-Things (IoT) has created a vibrant industry sector with a plethora of service scenarios well beyond the consumer-oriented Internet. This has broadened the scope of services and both functional and extra-functional service requirements. The questions are a) if the existing networking model, with its one-size-fits-all approach, can support this mix of services, and b) whether custom-tailored, in-network service provisioned as in-network service chains are a superior model. These questions go well beyond the addition of a small set of QoS parameters to different data flows or the usage of network slices as isolated parts – it considers the whole set of resources and service semantics. As a trend, **new service types are realized by integrating application and network services and their provisioning, across all types of networks.**

Another key aspect is the assumed *service invocation model*. While we already discussed the transition from pure client-server to collaborative model, the ‘language’ chosen for the transactions performed in said collaborative chains also varies. Although arguments have been presented that HTTP/REST may be seen as the new waist of the Internet [C4-08], the reality of many service invocation frameworks and protocols persists. Those range from request-response models (such as in HTTP), over pub-sub models (with HTTP/2 enabling some functionality) and message passing abstractions to remote memory access models (to create the abstraction of a large yet distributed computer with shared local memory). Similarly, there is an abundance of service discovery protocols (Bonjour, UPnP, …), none of which are interoperable, and few of which are applicable outside very specific environments. We can observe from this situation that *distributed computing has not converged* onto a single universal invocation framework that can be used to connect to any other compute resource. Furthermore, each service invocation framework usually comes with its particular lower layer protocols onto which to map the service invocation itself (e.g., HTTP->TCP->IP), often leaving IP as the only common denominator. Therefore, **services choose the best means of interacting with each other, while relying on basic means to route service requests.**

A final aspect is the changing *nature of the relationships* between the entities providing these services. Currently, systems providing services are mostly assumed to be trusted (or not), and reliable (with occasional faults), but the overall trend we are witnessing is to an increasingly more complex environment, where multiple providers compete with different (albeit similar) offers, with not exactly the same levels of guarantees and trust. Hence, the overall system can only provide *trustworthy end-to-end services* by relying on high system dynamicity to adapt to variable trust relationships across the different system components. **A service environment**

of determined trustworthiness needs to be set up by dynamic and intelligent methods over subsystems or micro-services of variable trustworthiness.

The key takeaway from these trends is that collaborative services in the Internet have moved on significantly since devising the key fundamentals of network forwarding that underpin the transfer of bits over the Internet.

4.2 System Architecture Vision: Towards Smart Green Systems

With the general move towards collaborative services, the main problem is to overcome the traditional yet obsolete separation of the entire compute-and-communicate infrastructure into separate domains (logic: network vs. application; business: telcos vs. clouds; silos: automotive vertical vs. manufacturing vertical; …). Chiefly, if the original Internet was about inter-networking, i.e., bit transport between different networks, **future research must address the inter-computing**, i.e., service execution between different systems.

Like the Internet of today, the Internet of the future will be a complex planetary system made of myriads of physically interconnected elements, logically broken in separate islands, each possibly applying different security policies, routing mechanisms, access mode to application services. With more and more intelligence and computing power available per resource, resources will be configured and orchestrated dynamically (i.e., also reprogrammed in runtime), both to deploy/support new services and to better match the requirements of services running over the network. With this however, *unlike the Internet of today*, the Internet of the future will exhibit much higher dynamics, notably in its own topology, and will better support virtualization, paving a possible path for its own technology evolution, very difficult today at a global scale. With that *higher dynamics* and the *co-existence of virtual and physical entities*, the physically separate policy islands of today will often overlap in resources in the future.

A massive number of devices will be connected and will generate and exchange very large quantities of data. Useful insights will be generated based on the automatic analysis of all that data (e.g., using machine learning methods, ML). The infrastructure that supports society (IoT, cyber-physical systems) will be integrated with the Internet, which will help improve the effectiveness and efficiency of both, e.g., using adaptation through learning. *It is paramount to approach ML systemically to correctly assess the relevant trade-offs*: ML instrumentations per se require massive data transfers, are computation-intensive and, ultimately, consume massive amounts of energy. *Relying on siloed solutions and dedicated implementations limits the usefulness of ML, while it increases the cybersecurity risks (attack surface).*

These trends imply that the future network technology will have to support the Internet economy and the particular needs of the cyber-physical infrastructure alike. It will have to work with virtual objects and remote objects, the density, distribution, longevity and interconnection of which in any area can vary a lot (cmp. DC virtual machines against physical L3 routers). It will have to integrate local and remote objects and different connectivity modes seamlessly. It will have to handle its own constituting nodes and services of transient nature, which can disappear and reappear, possibly at a different location and in zero time, be multiplied and shrunk without notice, etc. At the same time, this future network will be expected to operate as a facility: it will be relied upon by private users, businesses, critical branches and governments. Therefore, it will have to be resilient to both failures and security threats, in a world, where autonomic operations for both services and infrastructures, and in particular AI/ML techniques, will be widely used. Open standards will be required, while governments will want to impose limits and regulations on the usage of all the data required to drive these new systems. In this context, overcoming the digital divide will be a key driver for technology evolution, and personal freedom and rights will need to be assured across all media.

Here, **flexible provisioning and elastic execution on dynamic and changing resource pool emerge as key challenges for the future system architecture**. Flexible provisioning refers to the generality of the infrastructure and its capability to onboard and execute essentially any ICT service. The generality of the infrastructure, as opposed to the reliance on service-dedicated components, is important to increase *infrastructure sustainability* in time and *degrees of freedom for multiplexing gains*. Execution elasticity refers to an efficient adaptation during the execution, i.e., in runtime, and allows selection of best suitable links and components, to preserve the expected service properties while limiting overprovisioning. In particular, elasticity, as the capability of adjusting resources used in service execution, is key to enable **truly green networking**, as it allows to redirect requests to resources with better ecological sustainability and to limit the overall resource footprint while preserving the service throughput. Given the resource mix, we have to assume that elasticity and flexibility also apply to infrastructure resources. Hence, working with individual resources is limiting and not sustainable; rather, allocations and executions should refer to the resource pool as a whole. This in turn requires pervasive, resilient resource control.

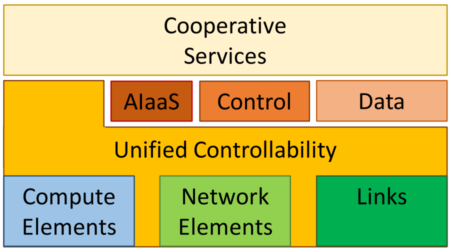


Figure 4-1 The Smart Green Networks Concept

We envision a Smart Green Network as a programmable system based on a single, unifying controllability framework spanning all resources a tenant is authorized to control, including from previously separate and heterogeneous domains, e.g., enterprise and telecom networks, virtual and physical, data centres and routers, satellites and terrestrial nodes, etc. The unifying controllability framework will glue the disparate resource islands to one system of the tenant supporting smart flexible instantiation and adaptive, elastic and correct execution of any service on the resources (Figure 4-1).

Hence, the key challenges that the Smart Green Network controllability layer must solve are: the aspects of control over multiple general-purpose, distributed, network control operating systems; the availability of powerful abstractions from resources to services; new naming schemes for virtualised resources; dynamic and automated discovery; structurally adaptive logical interconnection; multi-criteria routing in networks of different densities; intent-based open APIs and highly configurable policies to control the resource and service access as well as dynamics; isolation of application's execution environments and performances; efficient scheduling of requests to resources; a high degree of automation and support of self-* principles (*self-driving networks*); and distributed yet trustworthy ML instrumentations.

In addition to time-proven algorithm design approaches in order to provide provable and understandable behaviour, the Smart Green Networks concept will also use existing AI/ML algorithms as well as propose new, *network-suitable, distributed AI/ML*, to implement data-driven closed control loops that can enable cognitive and comprehensible system behaviour. The training and validation of such technologies require the availability of *cross-technology and cross-sectorial datasets* that do not exist yet. The networking research community needs to build those datasets, agreeing how they are generated, accepted and accessed.

Overall, it is imperative to:

- Allow dynamic pooling of resources from diverse participating systems, devices and objects;

- Integrate autonomics to enable self-organized, resilient programmability and elastic, correct service execution;
- Offer programmable analytics and cooperative machine learning to the service layer through open interfaces.

Keys to the realization of this vision are discussed in the next sections of this document:

- programmable infrastructures composed of versatile devices and subsystems (4.7),
- integration of AI/ML at the system level (4.6),
- efficient yet correct runtime resource allocations and their execution (4.5),
- extensible and flexible data plane protocols (4.4),
- and pervasive operational resource control (4.3).

Many of these aspects are currently being pursued in simpler forms inside telecom operators, which are pursuing network consolidation at the core (the integration of NSA, SA and future B5G networks in a single seamless solution, with fully unified control, rating and billing functions) along side the 5G deployment, given the expected gains in OPEX and improved network flexibility.

It should also be noticed that it is important that the directions captured in this Section 4 are **accompanied by the appropriate economic and policy work** in future research to make way for the envisioned new services that go beyond current 5G.

4.3 Virtualised Network Control for Increased Flexibility

4.3.1 Programmability is Control

Future infrastructures must be extremely flexible in operations and elastic in resource usage. Programmability of resources is the only way to achieve this. However, different from configuration management, **programmability requires runtime resource control**, i.e., a way for a program executed somewhere to receive some infrastructure event and to possibly tell to a given resource what to do, both proactively and reactively, including in runtime. The requirements on any control plane are classically intrinsically linked to the requirements on the data plane. Yet, with programmability (such as that being explored currently with P4), *any* data plane becomes possible, and hence, both functional and extra-functional requirements on the control plane are enormous. For a control plane used for software-defined infrastructure operations, network structure, the available functionality, transported payloads, data rates for the latter, the latencies of exchanges, the resilience and the security are difficult to predict.

A programmable system must provide an autonomic programmability after deployment. There are several pragmatic reasons for that: first, setting up such a versatile and resilient control plane manually is not a skill readily available in any environment; second, this approach would be delicate, as one would need to predict future needs correctly. The main reason however is fundamental: *autonomic organization is imperative to support infrastructure dynamics, which programmability as such creates*. Any programmability solution not able to self-organize or adapt is, therefore, incomplete. Network and system control cannot rely on rigid approaches, as any such approach would only be suitable for particular environments (e.g., centralistic control, particular hierarchies, etc). Instead, *novel solutions are required capable of organizing control flows and control-related processing dynamically among all controllable system elements*, i.e., across multiple domains, systems and layers. This includes initial self-organization, self-preservation during runtime facing external and internal events and *structural adaptation*. Modern ICT infrastructures need to provide dynamic resource management to fulfil different SLAs and to achieve E2E service assurance. Rigidity in any aspect limits the degrees of freedom and, hence, limits the optimality.

With infrastructure programmability (often referred to as “network virtualization” or “network slicing”, not to be confused with the “5G slicing” concept), the decoupling of the platform delivering the service and the service elements reaches a new level. While IP networking has decoupled services from network infrastructure by putting all services on the same technological foundation (the TCP/IP suite) and by pushing the service logic to the edge, network virtualization brings additional degrees of freedom in flow processing and combines edge and network in one logical entity: it is possible to have different flow processing logics active at the same time within the same physical infrastructure, usually in the form of software elements (different configurations, different active modules) deployed on top of more generically capable hardware resources. Whereas today's networks rely on specific flow processing machines (e.g., IP routers or Ethernet switches), whose flow processing capabilities are intrinsically linked to the purpose of the device, network virtualization breaks this barrier by allowing to define different flow treatments on the same network node and by concurrently reusing any given link for flows of different "slices" or services requiring different assurances. The same applies to the compute nodes.

4.3.2 Separation of control/controllability

This immediately raises a completely new question of a *service-independent control of resources per se*: as all infrastructure capacities are, in principle, service-independent, we need a novel means to make sure that the execution of any service-specific element on an infrastructure element is durably possible. In other words, while a router routes and a switch switches, and there is hardly anything to verify about that, programmability allows to tell a node to route, while this same node was not a router before, yet had other roles and tasks. It must be verified that it routes correctly over time despite possible task overlap. Classically, control was always integrated in a particular solution logic (on the respective OSI layer or abstraction level) and directly projected to resources dedicated to realize (a part of) that solution. Previously, as existence and function of a node used to be the same, so was their control. For example, network service errors can be traced down to network element errors, by using network service control means. With programmability however, this changes drastically. **We need to understand resource control as a new, paramount domain:** since node and links generally do not have single predefined functions, **there is a new requirement to allocate, monitor, migrate and execute/run several service elements on a shared, per se service-agnostic, infrastructure**. We call it *controllability.*

Additional complexity arises from the insight that, generally, an allocated function does not translate to a single infrastructure element, but can be sustained by resource capacities distributed over the infrastructure. Due to scalability and availability requirements, most network functions rely on hugely distributed realisations, causing the allocation, extension, monitoring or migration of a network function much more challenging than the question of copying a software state from one node to another.

4.3.3 Multi-Tenancy and Ownership

Network virtualization is resource sharing. Therefore, service footprints, projected to physical resources involved into the execution, are expected to overlap, constituting multi-tenancy in the overall system.

Multi-tenancy in management and control is generally hard, as it contributes to a so-called "split brain" problem: conflicts are likely to happen at the resource level, when several independent owners assign tasks to a shared resource (pool). Such conflicts can be in resource capacities (e.g., two tenants trying to book 2/3 of the resource each), or they can be of semantic nature (e.g., close port followed by open port). In control, multi-tenancy is harder to resolve, because of the potential time-criticality of the commands. **This calls for autonomic, system-integrated, runtime mechanisms for either conflict resolution or conflict avoidance, both in allocations and execution**. Candidate mechanisms per se should cater for multi-tenant operations and the expected system dependability and size. In particular, they cannot rely on single entities or centralistic approaches. This makes the design of such mechanisms generally harder and optimality as a goal questionable. Besides, while trying to provide service guarantees, such mechanisms should not sacrifice system availability and be aware of energy efficiency.

In spite of its expected pervasiveness, resource control solutions need to respect and maintain boundaries of the responsibilities, power and rights for each stakeholder in the ecosystem, as these are key for a secured, guaranteed SLA enforcement. The problem is that with network virtualization tenants can change their control scopes dynamically. Therefore, the classical notion of ownership is not well adapted to the problem space. Instead, the notion of **ownership through controllability** seems better suitable. This notion extends classical ownership through resources obtained through dynamic allocations, booking, and “leasing”. For instance, while resource limits of a virtual machine are up to the owner of the executing host, the definition of processes within the virtual machine is up to the owner of that virtual machine. Suitable control solutions should enforce this principle, also in the sense of (secure) isolation.

4.3.4 Known Unknowns

To support different realizations for semantically identical entities and to hide implementation complexity, a general key challenge is to separate enforcement (the ”how” part) from the decision (the “what” part). Given multi-tenancy and dynamicity, it is necessary to investigate the ways, in which the control boundary evolves between the objective (e.g., a number of decisions at a given point in time) and its realisation (e.g., considering the operational limits of realising any decision being made, the actually available resources, etc.).

Insisting on perfect knowledge in the described environment will often be in contradiction to the operational reality. Therefore, **solutions should be prepared to work with some degree of “fuzziness”**, i.e., with incomplete data, with data of different freshness, with unreliable postulates. That is why **adaptation is more important than optimality** in this regard. Generally, decision modules need intrinsic flexibility and call for software control elements, realising an adaptive control over the resources they manage. Changes in control objectives are reflected in the existing software, which, in turn, can establish additional software elements in order to react to changes in the control objectives. The enforcement, e.g., of flow handling or computation instalment, is realised by the resource owner, possibly self-constrained by objectives imposed by the physical infrastructure and its operational environment. With all this, the overall system will nevertheless need to fulfil the service requirements.

4.3.5 Self-Preservation

Given the importance of the controllability framework for the overall operations and its central position in the architecture, it is crucial to devise dependable, i.e., reliable and secure, solutions. In particular, the roles with respect to the programmability (controllability) and service operations (control) should be verifiable, and necessary protections must be applied to both control channels and control end-points, acknowledging decentralization, multi-tenancy and known unknowns, i.e., also dynamics in the overall span of the control plane and dynamics in the available infrastructure resources.

A running control framework must be able to adapt to such changes, e.g., include and remove resources, adjust its own resource usage yet still protect its own integrity. Besides, the execution of its constituting parts in possible remote, virtual objects on devices physically owned by other tenants calls for either trustworthiness verifications of such executing devices or for systemic approaches to mitigate dependency on any particular component.

The self-preservation solution must also counter so-called *self-inflicted errors* inherent to programmability: a running “program” of a tenant could have negative impact on the resource control framework per se. For instance, it could overload crucial control elements (e.g., putting controller under high load leading to timeouts), influence control transport channels (redirecting traffic) or the control plane structure (e.g., blocking control plane traffic to and from nodes and disconnecting controlees from controllers, etc). **Establishing system integrity and self-preservation in runtime for a distributed, dynamic resource control sub-system is one of the research challenges**.

4.3.6 Research Challenges

Challenges on resource control in Programmable Infrastructures include:

- Resource control emerges as an initial glue that first allows operators to program their infrastructures, i.e., as an initial new service that allows to allocate, monitor, execute and remove service elements on/from sets of nodes and links. To avoid vendor lock-in and to allow truly end-to-end slicing, it is exactly this glue that requires standardisation, and not any domain-specific management interface.
- Resource control must be able to reach out to all resources controllable by a tenant and be capable to check the states and operations of all service- or slice-specific elements on those resources. Besides, the realisation of the resource control itself should follow the insights from above, i.e., it must be distributed over all controllable nodes and must support elasticity of itself (reaching out to new elements, adaptability, including in structure, self-preservation, conflict awareness).
- Resource control needs to be able to handle questionable data quality, how to proceed when data is not good or biased, developing either fallbacks, or safe modes. Solutions need to be self-stabilizing, and able to address all these potential uncertainties.

Because of the novel degree of decoupling of service elements from the infrastructure, the central problem of programmability is not to make a blueprint, but to be able to execute any requested blueprint on top of a shared, distributed infrastructure composed of different capacities, occupied by loads from other executed services or slices. Such a distributed guaranteed execution under contention and with concurrency is extremely challenging and, currently, can only be solved on very small scales.

4.4 Re-Thinking the Data & Forwarding Planes

4.4.1 Design Considerations for an Evolved Data & Forwarding Plane

The original design of the IP-centric data plane of the Internet focused on three key fundamentals (i.e., principles – key design choices), namely ensuring *global reachability* through a *robust* packet forwarding mechanism that would provide a *best effort* service to higher layers [C4-09]. Those higher layers would complement the basic mechanisms through aspects of, e.g., reliability, error control, but also support for specific service invocation models.

From the discussion in Section 4.1, we derive a number of design considerations for data plane solutions that would ensure a continued support for the evolved services and interactions we have been seeing in the Internet, depicted in **Figure 4-2**.

Figure 4-2 Dataplane Evolution - Design Considerations

We exclude from our considerations the approach to *deployment of the solutions*, therefore not specifically addressing the possible *evolutionary* vs *clean slate* nature of re-thinking the data & forwarding plane in order not to constrain the research albeit pointing out that the *feasibility of solutions* will ultimately need to consider the evolutionary nature of any deployment in existing infrastructures. For instance, P4 is currently being explored as a tool for implementation of data plane programming, but this is effectively simply a technology tool.

It is important to note that evolved data & forwarding plane solutions do not need to necessarily address all considerations and we can already see examples for proposed solutions [C4-10][C4-11][C4-12][C4-13][C4-14][C4-15] considering certain aspects described here:

1. **Dynamicity:** As observed in Section 4.1, many relationships are bound to become ever shorter lived, driven by *virtualization* approaches, with the possibility of network resources to appear and disappear frequently. This introduces aspects of *dynamicity* into the relations that significantly depart from the long-lived locator concept that underpins IP, which assumed a long-lived relation between a client and a *portal* of information in the Internet. Instead, the assignment of forwarding relationships must align with the ability of the corresponding SW component to change relationships, or else the data plane will only inadequately support the advances we see in complex SW systems utilizing the Internet, e.g., through container-based micro-services.
2. **Green efficiency:** While we recognize that many of our considerations can be and partially have been realized through a myriad of *add-ons*, *extensions* to and *overlays* on top of Internet protocols, we strongly believe that *green efficiency* is a consideration that must be added to the design for an evolved data plane, even to the point where the selection of suitable mechanisms ought to include an *energy efficiency KPI* at the same level of today’s focus on performance KPIs such as throughput or delay. Overprovisioning and the aforementioned overlaying of solutions to improve on otherwise limited designs have played too long a role in communication networks for it to continue in the light of the increasing policy trends to fight against climate change, such as Europe’s Green Deal [C4-16]. While providing a flexibility in change (through yet another overlay), it has also led to complexity in management and the *inefficiencies* caused through indirections over many shim layers that make up the final communication relation. This not only stands in the way of achieving true high throughput and low latency communication, required by many emerging services, but also drives the ratio of ICT in the energy consumption [C4-17].
3. **Qualitative Communication:** Relationships will not only become more dynamic in nature but also more complex in terms of *inter-dependencies*. The current model in the Internet treats relationships at the application or session layer, realized through independent connections, managed through protocols like TCP and others, with separate resource management schemes. This leads to inefficiencies in cases where one sub-relationship is transferred well compared to the other, spending efforts on, e.g., error control, for a sub-relationship that is reduced in value due to reduced performance of another sub-relationship. The result is often overall loss of end user experience that ultimately decreases the value of the communication. This qualitative communication is crucial to be taken into account when designing data plane solutions in order to be able to optimize the use of resources spent on the overall relationship rather than the sub-parts of it. Leaving this handling purely to the application or session layer leads to inefficiencies of resource usage, which can be avoided through application awareness, e.g., additional in-packet metadata at a lower part of the data plane, expanding on existing concepts such as service function chaining (SFC) [C4-03] albeit for parallel not sequential transactions.
4. **Security** plays an important part in data plane mechanisms and the current Internet has well recognized this with security considerations having become essential in every protocol solution standardized, for instance, in the IETF. However, the fundamental of building *security on top of an otherwise unsecured packet forwarding* has not changed, therefore focussing efforts on end-to-end security of the application-level content, but not the *security nor the privacy of the packet forwarding operation* itself (who is talking to whom, compared to what is talked about). Consequently, this has enabled for long mechanisms

such as IP geo-tracing as well as enabling spoofing and therefore denial of service attacks. Mitigating methods deployed are add-ons to the otherwise unsecured IP, require extra effort rather than basing themselves on an *intrinsically secure* design per se where security of end points and networks alike is ensured together with the *privacy of the interaction* between communicating end points, striking the right balance between accountability and anonymity. Decoupling "security appliances" from the analysis of events and policy-based decision-making is another aspect to consider. Tiny security-handling functionalities embedded into virtual entities should monitor events, collect information and transfer it to suitable functions (possibly based on AI/ML) capable of more powerful analysis and anomaly detection, which in turn would enforce policy based-decisions back to the local actuators.

5. **Precision delivery:** the best effort nature of the current IP suite does not suffice for a number of the new emerging services, e.g., for Beyond 5G. Therefore, it will need to be extended in order to capture new demands for specific performance characteristics, such as *strict delay and latency bounds* for system control, human interaction and many other services as well as *on-time bounds*. This requires the control loops involved to ensure the specified performance requirements of various applications, particularly for access networks with widely varying performance characteristics such as wireless. Those control loops will also need to enable trading off latency at the control level against the necessary operations at the data plane.
6. **Diverse Addressing:** While the universality of higher layer service concepts over a single addressing scheme has been praised as key for the Internet protocol, we assert that the support for *diverse addressing* will need to replace this aspect of the current Internet in order to improve on efficiency when supporting the many new services, while still *ensuring the global reachability* that the current Internet has achieved. This could lead to solutions for *optimized Internet-of-Things* communication (with *smaller identifiers* being used for efficiency purposes), while preserving inter-domain access to the IoT resources. As another example, instead of relying on an interaction between DNS and IP routing, adding initial latency to the service exchange (and leading to problems in future service invocation if service relations might dynamically change), research in, e.g., routing on labels [C4-18], information-centric networking [C4-19] and solutions on *semantic addressing* [C4-20] have shown that those latencies can be significantly reduced through name-based addressing, pushing name information to the far edge of the network as a trade-off (which can be accommodated through increasing availability of storage, even in mobile devices), while still scaling to significant network sizes, particularly in the recognition that much Internet traffic is being localized, as discussed in Section 4.1. In addition, changes in named relations become merely an *ingress routing decision*, being removed as a burden from the DNS, for instance, therefore significantly *increasing flexibility* in routing when the service instance serving a named relation is changing in the light of virtualization of service endpoints, as discussed in Section 4.1.

The aforementioned considerations for *designing suitable packet delivery solutions* need to furthermore consider the following aspects when *being realized for and deployed* in the emerging communication infrastructure:

7. **Manageability:** All the above characteristics will require suitable instrumentation to monitor and validate the delivery of promised assurance levels. Furthermore, telemetry capabilities, i.e., the process of measuring, correlating and distributing network information, are required (and will need to be enabled at the data and forwarding plane level) to gain the visibility of network behaviour to improve operational performance over conventional network Operations, Administration, and Management (OAM) techniques to enable full network automation.
8. **Programmability:** As per Section 4.3, respective owners (e.g. service providers) will need to be provided with the methods to dynamically govern all resources incl. the forwarding plane in order to rapidly and easily introduce new network services or to adapt to new enhanced and modified contexts. A higher programmability of the forwarding plane could

be achieved, e.g., through insertion of programmable metadata in packet headers traversing the network. Such programmability particularly aims at providing the desired overall green efficiency by moving from HW to SW upgrades, including executable code injected into the execution environments of network elements in order to create the new functionality at runtime (in network compute) with the required characteristics (e.g. security). Furthermore, what is handled in-network or in the data place needs to be assessed.

9. **Slicing:** Resource management is discussed in details in Section 4.5. In the forwarding plane, it needs to promote easy and efficient execution of multiple and different types of delivery mechanisms, possibly each with different guarantees for KPIs/QoS/ stringent non-functional requirements of network services at a given time on the same infrastructure but across separate subsets of resources in the shared resource pool for realization of the desired functionality. Such "slices" may offer uniform capability interfaces to entities and network functions, abstracting the autonomous slice components, which may be loosely coupled, with different functional and non-functional behaviour. A challenge to address is the realization of large-scale and multi-domain data plane deployments in sliced environments, including aspects of identifying the participating resources being used.

4.4.2 Key Research Questions

The following research questions are not purely limited to the data and forwarding planes but address wider holistic systems aspects, leading to the following research challenges:

1. **Which layering in which part of the network?** To cater to the often starkly different 'scopes' of communication, ranging from localized sensor communication over POP-based access to OTT services to truly global communication, the question on layering is crucial in the light of an *efficient/green* implementation of the overall system. With the desire to support *diverse addressing* of the data plane, the question needs consideration as to *what layer best realizes the semantically different forwarding operation(s) most efficiently, taking into account not only the individual service itself but also the overall system efficiency from the perspective of resources that provide that service.* Note that this does not preclude combinations of different layers.
2. **What is the role of soft architecting?** With the proliferation of software-centric approaches to networking, allowing for a much higher degree of post-production as well as post-deployment programmability (cmp. Section 4.2), the question arises *what the deployed architecture really is* or *if everything manifests its own (soft) architecture?* Assuming such soft-architecting, as discussed in Section 4.3, the desire to agree on a common substrate, on top of which all such (soft) architectures reside, still remains, similar to the origin of the Internet protocol albeit with a possibly different answer. Instead of the commonality being that of a common postal system between locations, such *commonality* could be the interconnecting bus-like system between resources, where resource control becomes fundamental, while global transport and global routing degrade to applications, many of which can run in parallel. Any answer to that common substrate, however, should still provide the right set of fundamentals among those outlined in Section 4.4.1 that align with the services at hand. *In other words, while soft-architecting is a promising evolution path, ultimately, the considerations above need to be applied to and solved by the global "glue" at the resource layer, be it a control bus or the delivery system itself.* A possible advantage of a solution based on a resource control is a clear set of and a better understanding of the requirements of the latter.
3. **What are the tussle boundaries of the overall system?** Tussles [C4-21] are caused by interactions of players as defined through the interfaces of the overall systems, with each player often pursuing their individual interest. Understanding the boundaries of tussles, the mechanisms to express them and those to resolve them, is crucial for the overall working of the system. Much has been done to study the tussles of the Internet (and its main players) but postulating a system of high *dynamicity* also postulates one of changing relations, particularly when it comes to *trusted* relations. Enforcement through trusted third

party is often a mechanism that will not do in such often ad-hoc relationships and *solutions will need to realize more suitable, equally dynamic and ad-hoc mechanisms to ensure an otherwise trustworthy execution of the overall system, while also preserving the privacy and ensuring the security of the individual participants.*

4. **What (meta)data is required to make the data plane work (well)?** Any data plane solution, including existing ones, works on a set of metadata, such as identifiers, as well as state, such as link data. While much of this data is vital for the basic operations realized in the data plane itself, it is also required for *control plane* decisions (e.g., for load-depending resource allocations across the network) and for realizing *management goals* (e.g., matching long term demand to supply information). With this in mind, data plane solutions must not focus solely on hitting the key fundamentals outlined in Section 4.4.1 but also enabling a fruitful interaction with the corresponding parts of the overall system that ensure the working beyond the pure transport of relationship information.

4.4.3 Recommendations for Actions

The following list are suggestions for important actions towards realizing the research agenda for DP/FP evolution, not claiming to be exhaustive:

1. *Call for internationalized efforts*: given the challenge to evolve the data /forwarding planes, European efforts should liaise or even directly collaborate in internationalized research efforts, i.e., in the creation of solutions not just the exploitation in standards or OS communities. This could be realized through targeted **international calls** (e.g., EU-China, EU-US, …) on data/forwarding planes technologies as well as through the creation of **international expert groups**, e.g., in coordination and support actions.
2. *Call for experimentation*: although strong theoretical foundation is desired for any change of fundamental data/forwarding planes functionality, strong **experimental evidence** and **large-scale open testbeds** are crucial to show feasibility but also foster adoption through the operational community. This could be realized through an evolution of the original FIRE efforts or a similar trial phase as in 5GPPP. Open experimentation data/forwarding facilities are required for a large number of **third-party experimenters** of promising solutions and possibilities for looking, e.g., beyond 5G - an Internet of experiments (IoE).
3. *Call for data/forwarding planes research repository*: in order to foster the adoption of evolved data plane technologies, experimentation (see item 2) will need to ensure **replicability** in other, possibly pre-commercial or otherwise research, settings. This could be ensured through making evidence **data and code base availability** mandatory for certain aspects of data plane research (e.g., for certain TRLs upwards), including **migration solutions** that will allow legacy IP-based applications and IP-Services to be used with the new enabled forwarding plane capabilities.
4. *Call for clean slate research*: following the argumentation in other efforts, such as FP7 FI, NSF FIND, the evolution of core Internet technologies requires a combination of an **evolutionary and revolutionary** approach. This could be achieved through setting aside specific **clean slate** or greenfield funds for testing more revolutionary approaches to the data plane evolution.
5. *Call for funding data / forwarding planes research in solutions* along the considerations discussed in Section 4.4.1, such as those providing precision delivery in extension to existing best effort. Examples for such research aspects are
    a. Precision packet delivery (with QoS) to extend/complement best effort delivery;
    b. Intrinsically secure, i.e., authenticated and accountable, packet delivery;
    c. Semantic routing, extending current endpoint-based routing for lower latency and higher flexibility delivery of service requests;
    d. Deployment on tenant-specific (in-)network service functions;
    e. Inter-connection of compute/storage resources at Layer2, with focus on customer access networks while interconnecting to Internet-based clouds;
    f. Programmability of forwarding under control triggered by management.

4.5 Efficiency and Resource Management

Efficiency in terms of managing the resource pool of a communication system is essential for controlling costs and therefore OPEX in offering communication services. With Total Cost of Ownership (TCO) becoming a major design target, e.g., for 5G, and the push for *sustainability* of telecommunication infrastructures, the role of efficient resource management will increase significantly in future deployments. This translates to several new problem spaces, currently unaddressed, underestimated or completely overlooked in both the industry and academia.

4.5.1 Network Slicing versus Network Capacity Planning

As network slicing promises a sheer endless customisation of network-spread functionality, it becomes difficult to plan the capacity of network infrastructures in the same way as today. Whereas operators currently use their combined empirical knowledge regarding both infrastructure and the expected service (and its prices), network slicing turns this principle upside-down: while the infrastructure operator remains neutral to the service, the slice owner is expected to translate the *service to capacity requirements* onto the infrastructure capabilities, an exercise that lacks a reliable general methodology. Incapable of correctly translating service to capacity requirements, slice owners are likely to engage in a cloud-like operation model: start small, then expand or reduce contracts as you go. The *elasticity of the slice therefore is a central requirement*. This fact together with the required radical reduction of the service creation time (from 90 days to 90 minutes, as, e.g., per 5GPPP KPIs) underlines the upcoming shift from planning of the infrastructure to continuous (and likely dynamically adapting) runtime operations on the latter. In simple terms, network planning and network slicing are misaligned, as the former, driven by the presumed physical deployment, operates within completely different time frames than the latter, which exhibits on-demand elasticity.

Hence, **while the initial planning provides the larger operational bounds within which slicing can operate, it is the runtime (continuous, real-time, hot) management and control that determines the efficiency and therefore the costs of the sliced service**. If network slicing wants to succeed in the above sense, the employed technologies must embrace this change and provide mechanisms and practices that feed runtime control over a longer timeframe back into the planning and investment cycle for network infrastructure.

Independently of scale, slicing renders the infrastructure usage and occupation much more diverse and more dynamic. This emphasises the requirement for continuous operation of the real-time management or control, while infrastructure control and management are required to handle the dynamics in a new, currently unsupported manner. This includes handling node and service element loads, departures, additions, errors and the like.

Runtime management and control ultimately still drives the longer-term planning that we can see today in networks. Following our cloud analogy, the longer-term demand and supply pattern emerging from the many tenants of a data centre still drives the planning, and therefore investment patterns, for sufficient build-out of the cloud. Similar feedback must exist for slicing-based network infrastructure albeit situated in a many point-of-presence nature of resources, utilised over a possibly huge area of requirements on those resources.

4.5.2 Slicing Requires Conflict Resolution

To better support multi-tenancy and to allow efficient resource sharing, especially at bigger scales or facing known unknowns, **consistency and concurrency of slice allocations and slice executions should be addressed** at the systemic scale in runtime. Indeed, concurrent resource-competing or semantically contradictory requests at either allocation time or during (elastic) execution must be dealt with to avoid partial slice operation, generally being useless and, hence, waste of resources, while requiring **novel mechanisms for networked garbage collection** to free up any erroneous resource allocation during such conflict resolution.

While mechanisms exist for handling concurrency at individual component/node level, guaranteed slice allocations would require novel, system-wide mechanisms. Herein, fundamental systemic limits are to be properly addressed at large scale, since strong consistency of allocations (e.g., through consensus, atomic commit protocols with locking, etc.) might otherwise lead to a decrease of availability (starvation effects) and therefore reduce the supported dynamics in slice allocation and elasticity.

Inspired by distributed database management systems and distributed Internet services, **novel research should consider multi-level guarantees for services and service-level redundancy**. In spite of the similarity, the central insight here is the difference in the definition of consistency for databases and systemic allocations: while databases treat replica of the same object (which makes concurrent writes to replica R1 and R2 problematic), systems work with redundant, independent objects (e.g., concurrent allocations on two equivalent yet different paths are non-problematic). Given the observed increase in systemic redundancy (e.g., network density, trend to regional data centres), this insight promises better scalability of guaranteed allocations without sacrificing availability. Hence, **novel approaches could explore the suitability of concurrency-preserving schedules** (e.g., with commitment ordering) **for programmable networked IT systems.**

4.5.3 Elasticity: Slicing Efficiency Requires Runtime Scheduling

When addressing efficiency, *Total Cost of Ownership* KPI and *green ICT* become important aspects to consider. Given a slice blueprint, one must find suitable resources in the infrastructure and make a reasonable long-term allocation of the blueprint on the selected resources (as per slice lifecycle). This topic has received a considerable attention and is often referred to as *"virtual network embedding",* with both simplified greedy solutions and optimised heuristics (with tuneable sub-optimality bounds) being available. However, the overall resource allocation problem of network slicing is twofold, and the second part is unsolved, relating to the question of *elasticity of slices*. Indeed, to achieve slice properties not readily provided in the serving infrastructure (e.g., elasticity, but also availability, resilience, latency guarantees, etc.), slice embedding will be usually broader than the purely functional requirements of the blueprint. Therefore, for every entering flow, a simplified, yet more dynamic and online question of the resource allocation problem will arise: **which of the suitable function-equivalent infrastructure resources should be involved into the treatment of an incoming flow**? Note that this cannot be solved within the slice, if the infrastructure owner promises (and sells) extra-functional properties of the allocated slice; in other words, such provisioning will be done in the infrastructure, transparently to the slice owner.

The answer to this question of **runtime service scheduling** is paramount to address the TCO KPI, as solutions to this problem would allow to overprovision slices, without the need to overprovision the underlying infrastructure. The runtime service scheduling therefore is the answer to the questions of elastic and dynamic allocations, currently unsolved. Moreover, if an efficient solution to this problem can be found, network slices can and, for efficiency reasons, should be implemented as dynamically scheduled entities rather than exclusive reserved (and therefore possibly wasted) resource pools for tenant.

4.5.4 Towards Green ICT

In recent years, the *ecological conscience* has generally increased in Europe. Backed by political and economic initiatives both by the Commission (e.g., Renewable Energy Directive, Green Deal) and the Member States (e.g., German *Energiewende*), the main trend is *to reduce the dependency on conventional energy sources* (nuclear, fossil) *to the advantage of renewable energy supply* (wind, photovoltaics, hydroelectricity). Given the decreased flexibility in the energy production of the latter, *this shift must be accompanied by smart energy demand management functions*, resulting in a strong push for Smart Grids in the energy sector. That is where ICT is generally regarded as an important enabler (e.g., using 5G MTC and network slicing). However, swapping power sources does not address the power consumption of the consuming infrastructure as such.

Given the increased reliance of the society on ICT infrastructures, these have emerged as essential consumers. For instance, while 5G is 10 times more energy efficient than 4G in transmission, recent studies suggest that, by 2025, 5G alone can increase the anyhow growing energy demand in the data centres by up to 3,8 terawatt hours (TWh) in addition [C4-22]. Even though this effect is due to the increased "popularity" and not to a shortcoming of 5G per se, undeniably, **energy efficiency of data centres emerges as a central preoccupation for resource management**. While overprovisioning is a simple and popular method in networking (e.g., in fibre optics, it is a simple mechanism for both network development and service quality increase), overprovisioning is not a valid approach for the computing domain. Indeed, modern DCs reduce the required compute power for the same load, using DC-internal schedulers (e.g., Apache Mesos, Kubernetes K8).

Novel methods are required to overcome the limitation to a single DC and should embrace path and compute allocations together, in order to exploit infrastructure diversity. **Future research should explore and develop approaches to elastic resource management** in addition to the current trends limited to green energy power supply for data centres (using smart grid's demand management) and the "recycling" of waste heat from the DC cooling systems. Such novel approaches could generally rely on elasticity mechanisms, i.e., runtime redirection of incoming service requests to best suitable infrastructure components with the goal of increasing the throughput on the same resource footprint. Preferred redirection to eco-powered components can be integrated into runtime service scheduling.

This theme translates to the overall ICT sector and ICT infrastructures in that **green or sustainable ICT cannot be achieved without a profound consideration for resource management**. Given the steady increase in the dynamics and the diversity of services, pre-planning and fixed allocations of any kind (dedicated devices, pre-provisioning, long-term configurations, mapping to particular nodes, single points of failure) are doomed to overprovisioning, which, for the same service load, requires more resources to be deployed, maintained and powered up in the infrastructure. This wastes energy and is ultimately not sustainable.

4.5.5 Research Challenges

On the opportunity side, **programmable ICT infrastructures increase the degrees of freedom** in service-to-infrastructure mapping and, therefore, **could yield more sustainability both in time (flexibility) and in energy (elasticity)**. On the challenges side however, the mixed compute/ storage/ networking environments, even under the assumption of pervasive controllability, require suitable solutions with respect to resource management: the heterogeneity of resources makes it harder to rely on single mechanisms, as different domains apply their own approaches internally, and often do not exhibit this knowledge externally. Also, a given unique approach will likely not fit the requirements of different resource types. Besides, the scale of the overall infrastructure makes it hard to rely on any consistent, up-to-date picture of the current consumption vs. load, as described above.

Challenges in this area can be summarised in the following:

- The question of runtime service scheduling in programmable ICT systems is paramount, as it permits both to provide superior extra functional properties of the supported allocations ("slices") and to lower the Total Cost of Ownership. Indeed, the TCO of a slicing implementation using only fixed-quota assignments (meaning that the sum of the resources consumed by all slice instances will define the necessary infrastructure resource footprint) would be horrible, roughly comparable to hardware slicing. **The dynamic resource assignment problem,** as a quest for a more efficient infrastructure sharing, including computing, networking and energy resources, **is difficult because of heterogeneity, partial or outdated information, its runtime**

nature and the absence of any central party or mechanism (like ordering or synchronized clocks).

- The answer to the job scheduling in large networked systems requires a lot of fundamental research, such as leveraging existing solutions from data centre research and applying them at network scale with multitenancy and concurrency. Suitable **conflict handling mechanisms are required** here, especially if guaranteed execution is required. Utilizing insights from distributed systems research, **the major goal should not be optimality, but rather improved efficiency**: given the size of the infrastructure, *1 % efficiency increase might translate to hundreds of millions of Euros/Watts/additional users*/etc. Given the assumption of sub-optimality, novel **mechanisms for networked garbage collection can be considered**.
- The elasticity of slicing has to propagate towards subscriber level and even application level. For instance, an application could use different slices during its session in order to best utilise the network as well as to provide superior quality of service with respect to slice offerings. In the view similar to application-driven networking, an application could also explicitly ask for a "slice" suitable to its needs. This rules out any pre-provisioning and can only be reasonably implemented in public infrastructures like the telecommunication networks, if the provision of the slices is highly dynamic yet resilient. Thus, **application requirements need not only signalling but also suitable translation to constraints, under which the slicing control can operate to meet the applications' needs**.

4.6 AI/ML-based System Evolution

Utilizing knowledge gained over a longer time is well-established in the industry. OTT services have long been using AI/ML techniques, albeit operating largely on data sets derived from the services and their users directly. At the level of improving network operations, self-* solutions have advocated the use of operational insights to adapt network functionality without intervention from either human operators or users.

Given the vast amount of data available in complex network environments albeit in a distributed fashion, AI/ML is well suited to produce new insights into emerging behaviour patterns in such distributed environments. To this end, suitable AI/ML techniques are applied, provided as a service capability towards (a) *operations of networks* and (b) *improvements of service provisioning and functionality* itself. In other words, we see a strong evolution of future networks from a mere communication and computing infrastructure to an integral part of the overall knowledge pool that can be used to improve functioning of networks and services alike; *AI-as-a-Service (AIaaS)* provides this capability in a prosumer-centric notion.

4.6.1 Proliferation of AIaaS in Network Operations

We foresee AI/ML playing an increasingly important role in network management, with the aim of reducing costs, increasing productivity, deriving more value, and improving customer experience. A range of learning techniques can be used to predict the behaviour of the network and its users to better provision resources by avoiding today's typical over-dimensioning. In terms of OPEX optimization, where energy consumption is one of the major cost items for network operators, AI/ML will leverage ***"data lakes"*** to analyze performance and optimize energy consumption versus quality of service. We furthermore see a strong alignment with the move towards fully virtualized network functions, where AI/ML capabilities are utilized to ensure reliable controllability in a fully automated manner, specifically to:

- Instantiate a complete end-to-end network that includes, e.g., the RAN, mobile core, other forms of access networks (DSL, etc.), transport network, as well as the Data Network. This network may be logically separate and/or isolated for certain aspects

like services, users, etc.Extend such an instantiated network by computationally and storage resources where suitable, spanning from edge computing to backoffice data centres.
- Incrementally deploy network services in the operator's network (or elsewhere) in logically separated and/or isolated manner from the other already deployed services.
- Deploy and provide network services to other operators and/or service providers when requested, via open interfaces. This way, other operators and/or service providers can re-sell/extend the provided network services.
- Realize fast lifecycle management (LCM), automatically triggered based on vendor-independent FCAPS management.
- Instantiate new components into a live production network in a plug-and-play manner.
- Terminate one or more network slices or service(s).

AI/ML-based network control – as a way to implement fully automated Smart Networks – seem like a must for future networks rather than a nice-to-have. To wit, the scale of deployments made possible by function virtualisation, the extreme split in micro or atomic functions and the proliferation of more and more functions at the edge create network deployments of unprecedented complexity, challenging to manage and control with current decision support tools. Down the road, we see a need to overcome the current juxtaposition of conventional *model-based* approaches (who have, after all, driven the Internet for decades) with still untested but promising *data-driven* approaches and come up with integrated, hybrid solutions. Possibly, data-drivenness could compensate for fuzziness and uncertainty while model-driven approaches could provide a solid operational foundation.

- The system challenge here is to develop a future network with *Full Automation*, which reduces and tries to eliminate any human intervention. In principle, such automation can be achieved, once exact behaviour of all components is understood and expressed in a suitable model. In practice, however, for the highly complex and interwoven system outlined here, such a full-model description is not feasible, rendering *model-driven* automation and control impractical. For such situations, data-driven approaches leveraging powerful AI/ML systems might come to the rescue. One challenge here is to determine which data to use for what control aspect, using which AI algorithm. For example, there is a challenge that AI/ML is seamlessly applied to network control, to run automated operations of network functions, network slices, transport networks, in an end-to-end scope. As another example, we can consider the use of AI and machine learning for coverage hole detection and outage detection (AI and Machine learning can be used to predict the coverage hole based on MDT data and self-organise the networks, even e.g. triggering deployment of UAV base stations to improve reliability of the networks).

Moreover, a thoroughly integrated AI scheme would open up new venues, how to think about operating a network in general. For example, suppose good to very good predictions (load, failures) were available. Then, the possibility arises to implement predictive behaviours in the network, to make available a network control intelligence capable of mitigating failures, the usage load, etc. and quickly adapt network configuration to be always available at the target performance levels requested by the applications. Basically, we could switch from closed-loop control to open-loop control (or, at least, to receding horizon control – also sometimes termed “open-loop feedback control” – where the optimizing control strategy is recomputed in the light of new observations leading to new predictions over a forward shifting time horizon).

4.6.2 AIaaS Proliferation in Service Provisioning

Beyond the use of AI/ML for improving on network operations directly, AI/ML will enable innovative features when provisioning future digital services for homes, businesses, government, transport, manufacturing smart cities and other verticals. At the same time, we expect a significant increase in the amount of machine-to-machine (sensor) communication monitoring smart cities, Industry 4.0, smart energy, etc. These changed traffic patterns will drive the move of computational and memory/storage resources from huge data centres towards the edge of the network, therefore impacting network designs to support this move. New services powered by AI/ML may also bring significant socio-economic impacts together with improved sustainability models for Network Operators. For prosumers, we foresee the proliferation of *personal data platforms* that are tightly connected with network services, and the development of tools allowing Internet users to control their data.

Key to reaping these benefits lies in utilizing the knowledge derived from the vast pool of network data in the services provided over the future telecommunication infrastructure but also utilizing the *highly distributed processing capability* that an AIaaS offering would provide. It is crucial to also understand the impact of M2M traffic, generating e.g., smart city data, etc., which will shape system designs. Both aspects drive the *provisioning of data into the system* as well as *complementing processing capability* of the network with service-level ones. With this, we see AIaaS capabilities of the infrastructure merge with those capabilities at the data and processing level that vertical customers will bring to the table. As a consequence, we see an *emerging data marketplace* that goes beyond raw data (such as location traces) but is lifted to knowledge and insights provided by network operators to its service provider customers. For instance, radio measurements at the deep level of small-cell base stations can provide insights on physical objects that in turn can be utilized by service providers for consumer-facing services that would have otherwise required dedicated hardware deployments or other means of realization. However, key to making an AIaaS useful for service providers, clear and open interfaces, both for data provisioning but also the reasoning logic, are required. Furthermore, *control over the distribution of, and access rights* to, both data and processing is crucial for the alignment with privacy considerations that both network operators and service providers will adhere to and may contribute to future regulatory aspects.

4.6.3 Related Research Challenges

While such AI/ML-driven or self-driving networking can start using existing AI and ML protocols, algorithms and approaches, it will gradually require network-specific adaptations in several regards. Below are some of the challenges we can identify in pursuing an AIaaS vision:

- One aspect is the **availability of network-typical and network-characteristic datasets** for training and validation. There is no commonly agreed reference dataset to use in research or development to compare different approaches against each other, nor is there a good understanding which data is actually needed to drive an AI/ML scheme, which features need to be extracted from an operational network.
- Similarly, current experience shows that the **procedures to train and validate** AI/ML algorithms and the architectures they use are mostly focused on static pattern recognition (e.g., images, sounds, diagnostics of fixed analysis data…) and are therefore not well adapted to the nature of dynamic networks. We need schemes suitable for changing environments, changing number of users, changing topology, etc. – properties not typically found in popular ML algorithms.
- Even with suitable datasets and algorithms in place, **there is the need to extend the currently mostly centralized AI/ML algorithms to be distributed** to accommodate the distributed deployment in (often multi-domain and multi-technology) networks. This, in turn, will introduce challenges to ensure *scalability, consistency, consensus* and *convergence* of both data as well as reasoning in such distributed environment, providing auditable solutions that may foster future regulations. Complementing this need for supporting the distributed realization of AIaaS is the opportunity provided by the move towards *Edge and Fog Computing* that we can already see in 5G. This opens the opportunity to complement the resources of cloud computing data centres to analyse the expected vast amount of network data; it could even do so while better adhering to privacy demands through *localizing the processing* of raw data.In such a scenario, there are trade-offs between data volume to be transported vs. localized or distributed energy consumption and computational capacity; latency for training vs. latency for action; questions about ensemble learning when locally learned insights should be merged and

generalized. For both learning and control in ML, we need a *meta-control* that allows for deciding, which data is fed into a learning scheme, where and which learned models are distributed to which place in the network for taking control decisions. This is similar to provisioning micro-services in general. However, it might have quite different data-rate/computational/latency/resiliency requirements compared to an application-level microservice. In other words, *AIaaS will need its own control plane logic* built upon the control plane capabilities of the infrastructure itself.

- Meta-control immediately raises the question of self-application: **can ML be used to decide on ML**? This idea is currently gaining ground in the Auto-ML community, where ML is used to learn hyperparameters of ML. Here, we need ML to learn, how to apply ML to a network. Clearly, there is considerable risk of oscillations, feedback loops, etc.
- The **scope of AI/ML schemes will also need to be investigated.** One possible, perhaps naïve approach is to have one set of AI functions/data sets that is applied only to a segregated, intra-service based scheme ("*sliced AI*"), which is easy to realize and ensures data privacy, but squanders possible optimization potential. Removing redundancy and going to a cross-service, cross-network, integrated AI/ML ("*integrated AI*") scheme is promising, yet fraught with complex design choices.
- Given the increasing multi-domain and multi-technology deployment of infrastructure, AIaaS will require the capability for *multi-domain orchestration* of distributed processing, meaning end-to-end interoperability is a must (cmp. Sections 4.2, 4.3). This requires greater standardization efforts and further progress in the functional architectures.
- Furthermore, **aspects related to security** beyond the conventional application of AI as a tool, e.g., ensuring data flow provenance and distribution within the system, and dealing with AI-enhanced (-amplified or even -rooted) attacks are essential. With the emergence of high-bandwidth and low latency requirements of applications for Immersive User Interfaces such as Wearable Cognitive Assistance (e.g., Google Glass, Microsoft Hololens), private 5G networks, and IoT appliances, the edge clouds or cloudlets are becoming ubiquitous. The security and performance of such private cloud datacenters is of paramount importance. Development of automatic verification systems to assess the performance and security of edge clouds by leveraging Open Source solutions like Central Office Re-architected as Datacenters (CORD) and operationalizing the results is another interesting avenue for the researchers from academia and industry.

4.6.4 Recommendations for Future Actions

Based on the challenges above, we recommend research into the following aspects:

- making **available network-characteristic datasets** for training and validation;
- agreed procedures to **train and validate** the AI/ML algorithms;
- **distribution of AI/ML algorithms** instead of using centralized AI/ML algorithms, in order to apply AI/ML to a network, considering placement and distribution of AI/ML functions within a network;
- **meta-control procedures** applied for learning and control in AI/ML to decide which data is fed into a learning scheme;
- **integration with AI/ML features provided at the edge of the network**, e.g., when provisioning future digital services for homes, businesses, government, transport, manufacturing smart cities and other verticals;
- devise architectures, approaches and algorithms for **sliced vs. integrated AIaaS*;***
- **development of use cases** for new services powered by AI/ML at the network and service provider level;
- **development of network management techniques** embracing the AI/ML predictions;
- **support performance analysis and optimization** methods for energy consumption versus quality of service analysis, e.g., through an AI/ML enabled "data lake" approach;
- support for new AIaaS services and applications that require, e.g., **multi-domain orchestration** of distributed processing and end-to-end interoperability;
- address of **security and privacy challenges** and provide information for future regulation;

- usage of machine learning to learn **network anomaly/vulnerability patterns** and, thus, prove to useful in identifying persistent bugs/vulnerabilities.
- support the **provisioning of data** required for AI/ML learning phases, particularly from network infrastructure functions;
- address the **scaling requirements**, e.g., through partitioning mechanisms, to enable efficient AI/ML data processing to provide timely responses required by AIaaS solutions.

4.7 Deep Edge, Terminal and IoT Device Integration

Architecturally, the ‘deep edge’ with its IoT as well as end user or vertical industry devices well integrates into the vision of Section 4.2 by becoming part of the common resource pool, provided as a non-decomposable set of resources by some edge entity, such as an end user, industrial site owner, or a building owner. Following the ‘ownership through control’ mantra, described in Section 4.3.1, we therefore envision tenant-specific resource usage to expand into the deep edge with the same control and data plane considerations, as discussed in Sections 4.3 and 4.4 respectively, and resource management considerations, as discussed in Section 4.5, applying to all those resources. In other words, in principle, we see aspects of controllability of those edge resources to equally apply together with the general programmability for the realization of compute tasks as well as for data and forwarding plane operations through those resources.

However, some edge resources might not directly fit into this vision. For instance, IoT will introduce particular, service-dedicated, possibly intelligent yet resource-constrained components (micro-electronic, battery driven components), which will need a particular consideration for the integration with the rest of the system. Indeed, such IoT components and devices might impose additional requirements on, e.g., volatility and longevity, punctual presence at any moment, persistence, generality, capacities, connectivity, interfaces and APIs from/towards the system. Hence, they might not support direct integration and require particular solutions instead (e.g., gateways or subsystems).

Generally, edge resources often provide human- or generally task-centric input and output capabilities, expressed in a plethora of sensory capabilities, situational awareness, quality of experience perception, which make these resources very useful for integration into the overall vertical application. This yields a *richness* of resources that is challenging when being integrated into a common resource worldview. Unlike the emerging COTS (customer-off-the-shelf) platform basis in other parts of the communication system, e.g., in the core, the edge provides a more *diversified and heterogeneous environment* with many device platforms and their supported local connectivity technologies (e.g., WiFi, BT, LiFi, and others), all of which are provided through a plethora of programming environments. **Future research will need to develop a suitable common model of system-wide representation akin to ‘device drivers’ in existing computing platforms.**

This resource richness at the edge, however, often comes with a **limitation in capability**, e.g., in terms of available processing cores in smartphones that can be utilized in the common resource pool. Given that devices at the edge exhibit a high heterogeneity ranging from a simplistic sensor and IoT devices to edge data centres, other typical limitations include energy/battery, form factors, human-machine interface, storage, physical security. This stands in stark contrast to the perceived limitless resource capabilities in data centres as well as core networks and, therefore, impacts the *decomposition* of computational tasks over a resource pool that is geographically and physically limited. As a consequence, the aforementioned *controllability* will need to be ensured through the realization of a suitable *control agent* that integrates the (edge) resource pool into the larger system but also interfaces with the (edge) resource pool to adequately govern the resource usage in the light of the resource-specific characteristics in terms of constraints and dynamicity. Here, research into the *minimal requirements* in terms of processing and communication needs and the realization of those requirements as *novel control agent realizations* will need to ensure that integration into the overall control fabric of the larger system to align with our vision of a smart network as laid out in Section 4.2. Furthermore, *resource scheduling* requires extra consideration in the presence of potential resource scarcity, particularly when combining specific input/output capabilities into the scheduling decision. Scarcity may be increased when utilizing specialized resources, such as GPUs or NPUs, rather than general purposes ones. We may also find that *locality of the resources* becomes crucial when applying policies for, e.g., localized processing for privacy reasons. Scheduling solutions are required that provide suitable trade-offs between moving data to functions or vice versa, possibly under locality constraints. Ultimately, a scheduling decision in favour of one tenant may result in detrimental performance of another, calling for solutions to resource scheduling that likely extend beyond those operating on a large pool of resources with uniform capabilities. **Future research will need to address these edge-specific constraints through suitable scheduling mechanisms that take those constraints into account, while relying on edge-specific control agents enabling the enforcement of the policies underlying the scheduling solutions.**

The *dynamicity* of (edge) resources is another aspect to deal with as an edge-specific constraint. While edge infrastructure, such as in an industrial site, can obviously be very well managed and long-lived, we also foresee edge resources of a much higher *volatility*, particularly when considering end-user provided resources, therefore creating a *limitation in availability* in contrast to, e.g., long-lived data centres. Those resources could be switched off, temporarily disconnected or simply become unavailable, e.g., if linked to human behaviours or policies (such as “do not make my phone available, if battery drops below 15%”). From a control perspective, *maintaining the basic control fabric* needs to take such dynamicity into account, while the *scheduling* will need to react to disappearing and reappearing resources alike to operate at a defined optimum of resource usage. From a data plane perspective, volatile resources need consideration when *routing packets* but also when *establishing in-network state* for forwarding operations. While volatility of resources and dynamics are already covered by the controllability framework presented in Section 4.3, **future research will be required to delve into the systems of systems aspect of such controllability**, given that individual subsystems might not be fully independent.

Furthermore, *governance* of edge resources (and their provisioning through entities like individual users and localized industries) differs vastly from the often long-lived contractual relationships we can identify in the core network business. Instead, the addition and usage of resources with such volatile and temporary nature requires means for *contractual management*, including methods for billing, accounting as well as authorization of use that align with the dynamicity of the envisioned relationship. *Distributed ledger technologies* and *eContracts/ smart contracts* will likely lend themselves to being applied in this world of (possibly highly) ephemeral resource utilization with the appropriate means to keep the resource owner (e.g., the end user) in the loop in order to preserve *digital sovereignty* but also enable *participation in the digital market,* akin to the changes in the energy market but likely much more dynamic. An important challenge for entering contractual relations is the *advertisement of resource capabilities*. While today’s solutions are mainly focussed on the pure ability to communicate (e.g., through advertising a radio bearer), solutions are required that expand the negotiation towards clearly articulated *demands* beyond ‘just communication’ that can be dynamically matched against the *supply*. For instance, attaching to a WiFi access point is futile, if connectivity to particular backend services is not enabled at this edge resource. *Efficiency* is key here, avoiding unnecessary signalling between components. Particular consideration must also be given to *security*, both towards the tenant utilizing the resources and those providing them. With tenant-specific instructions eventually being executed on what are possibly end-user provided devices, *accountability* for this usage is key for accepting such usage in the first place, complementing (edge) platform capabilities such as secure enclaves to ensure trustworthy execution at the level of the computational instructions themselves. **Through research in this space, we foresee future solutions to enable an edge resource market that would allow for auctioning the availability of resources to tenants very**

much like the bidding for white space on a webpage as we know today, basing all interactions on a trusted, auditable, and accountable basis that caters to the dynamics experienced at the edge. For this edge resource market to emerge, policy descriptions with their rules and constraints will need to be specified in a form that can be enforced by the infrastructure on the services, since direct human oversight is not feasible at this scale. **This will require research into novel programming models and (e.g., policy) languages that not only support all of these services, applications and deployments but also cater to the expected dynamics of the market itself**. Deploying and managing a large set of distributed devices with constrained capabilities is a complex task. Moreover, updating and maintaining devices deployed in the field is critical to keep the functionality and the security of the IoT systems. To achieve the full functionality expected of an IoT system, research should be done in advanced network reorganization and dynamic function reassignment. **Research is needed for providing new IoT device management techniques that are adapted to the evolving distributed architectures for IoT systems based on an open device management ecosystem.**

The continued growth in video applications including augmented reality (AR) and virtual reality (VR) required by, among others, the emerging applications (cmp. Section 3.1), requires new architectural approaches and solutions. Surveillance and monitoring further complicate the space, as will the growth in real-time sensor data e.g., for industry and smart cities. The ongoing shift of TV distribution from broadcast to the Internet will accelerate, requiring at least a 10x increase in video traffic volume with increased performance and resolution. The implications on application level networking are tremendous: we will need to integrate video services with the web content framework, delivery model and APIs, with effective use of ultra-dense and diverse wired and wireless networks. Video provenance will become a key issue to combat ”fake news” and the effects of AI/ML-generated video that can subvert legitimate content. Strong security and integrity of applications, network transport and in-network processing will be required. A future key development in the system architecture can be the deep integration of application and service functionality pervasively within the network, as discussed in this document. To cope with that, this document introduces a highly dynamic system architecture (cmp. Sections 4.2-4.6).

This architecture will need to be supported by the nodes that constitute it (i.e. devices, elements, subsystems, etc. or whatever nature). Hence, at the node level, an active entity (e.g. an agent) becomes necessary, capable of a) offering runtime access to node-local resources and to all executed allocations and b) acting as part of a dynamic system, i.e. establishing and maintaining it. In its first controlee aspect, this entity is an entry point to the internal organization and realization of the node (e.g. of a whole subsystem). Exporting a common set of protocols and an API, it can hide the complexity of the internal organization through its own implementation and allow independent evolution of the node-internal and systemic organizations. In its second systemic aspect, this entity must autonomically and continuously construct, maintain and preserve the control plane considering the requirements in sections 4.3-4.6. In other words, beyond service provisioning, management and security, which are critical to effectively manage billions of devices, ensuring they are suitably configured, running appropriate software, kept up-to-date with security updates and patches, and run properly authenticated and authorized applications, this entity must ensure system integrity and resilience of the programmable environment per se, while taking into account the available resources of the node that it represents. Chiefly, the agent must assure that both intrinsic and situational capabilities of its node (e.g. secure boot, local secure hardware modules, secure enclaves; input/output capabilities e.g. positioning, sensing usable for discovery of other potential nodes; topological position of the node, e.g. its connectivity degrees and its centrality; but also the available generic compute and networking capacity) correspond to the role, tasks and the topological position in the overall system in both directions. Therefore, a balance between agent commitments towards the system vs. resources required by the agent itself is required.

Locally, the agent must consider additional considerations. For instance, in addition to classical contractual models, micropayments might become a key part of the system as the infrastructure to support in-network services and applications is not free. Privacy and data management, and the location of processing and data to match legal and moral restrictions on data distribution, access and processing, will be increasingly important. Many of the services and applications will operate on, process and deal with personal data that is increasingly (and rightly) subject to strict regulation, control and limitation. Strong tools do not exist to describe in human language, legal language or code how data can be processed, located and distributed. Policy descriptions, rules and constraints will need to be specified in a form that can be enforced by the infrastructure on the services, since direct human oversight is not feasible at this scale. In addition, novel programming models and languages are required to support all of these services, applications and deployments.

Research challenges in this area include:

- **delivery model and API**s, with effective use of ultra-dense and diverse wired and wireless networks (cmp. Sections 4.3 and 4.4);
- **effective management of billions of devices**, ensuring they are **suitably configured, running appropriate software, kept up-to-date with security updates and patches, and run only properly authenticated and authorized applications.**
- **privacy and data management,** and the location of processing and data to match legal and moral restrictions on data distribution, access and processing, will be increasingly important.
- **policy descriptions, rules and constraints** will need to be specified in a form that can be enforced by the infrastructure on the services.

5. Edge Computing and Meta-data

In this document, Chapter 4 (System Architecture) presents a vision, which specifically assumes that the overall infrastructure is composed of intertwined compute and network resources of both virtual and physical nature, including, in particular, the terminals. Chapter 4 therefore focusses on such mixed infrastructures, where in-network compute (highly distributed services) and in-compute networking (virtual networking) are just special views on the very same infrastructure. This vision is inspired by the overarching trend to "Everything-as-a-Service" (XaaS) and the corresponding realities of operators of such mixed virtual/physical ICT infrastructures. Compared to legacy infrastructures, such mixed infrastructures are characterized by higher dynamics and changing densities of both the constituting nodes and the system at large.

Yet, until today, there is a big difference between networks and service platforms. So far, network elements are not capable of compute, and networks only implement few network specific services, sometimes invisible to end-users. In contrast, non-network services are implemented by terminals and data centres, in an OTT fashion. The binding between both is generally a "grey area" and considered difficult. Overall, there is a seeming contradiction in the reigning network philosophies: while the Internet philosophy minimizes network functionality to a service-agnostic common transport (TCP/IP as a least common denominator), mobile networks feature dedicated, standardized, highly available network functions, which require non-negligible compute resources. In the Internet view, the required link between the service-agnostic network and the network-agnostic service is often established by dedicated means, e.g. by middleboxes trying to enforce service-expected properties, for instance, security in enterprise networks. In this view, mobile networks are special, large-scale "middleboxes" in the access, combining compute and network resources for some particular purpose. Vice versa, in the mobile network view, Internet access is just a particular service provided by the network.

When users require services with hard and challenging performance guarantees, both views are incomplete. To cope with this, our main vision (cmp. Chapter 4) seeks to overcome the current "compartmentalization" of the control systems and establishes pervasive resource control as a novel least common denominator for future ICT systems, over which fine-grained, resource-efficient service provisioning becomes possible. However, on the way to this reality, compute infrastructures must emerge and establish themselves. They currently strongly proliferate at the edge of global networks, most notably as part of the mobile systems (from telco cloud to user services) or within the infrastructures of different vertical industries (seamless service integration). This shorter time evolution, together with the corresponding relevant standardization and coordination activities is discussed in this Chapter.

5.1 Introduction

The edges of the Internet are quickly changing, moving rapidly towards the consumers and becoming even more heterogeneous, not just in terms of devices (personal devices; industrial IoT devices; autonomous vehicles; low-orbit smart satellites), but also in terms of services, data sources, and data volumes. The computation of large-scale heterogeneous and distributed sets of data, and the execution of highly distributed applications, bring in additional requirements that today's Cloud computing cannot fully serve. Specifically, while traditional

Cloud is already supporting the use of smart devices within the context of the Internet of Things, the interconnected number of active Internet devices, most of which mobile, is expected to rise until 75 billion[3] in 2025. The ever-increasing computational power of these devices provides a great potential to keep part of the data processing tasks running "on the Edge of the Internet". Gartner estimated that by 2022 more than 50% of enterprise-generated data will be created and processed outside the Cloud[4].

Additionally, scenarios such as Industrial IoT are bringing in larger and more complex sets of data that, if pushed to the Cloud, may lead to performance degradation, increased costs (financial, computational and in terms of resources such as memory, energy), management complexity and to other issues related to security, privacy, efficiency or reliability. The rise of IoT has also significant impact in data centers load, where it is estimated that barely a 10% of the data received is useful [C5-1].

Adding to this aspect is the application of virtualization techniques. There is a rise in the decentralisation of Internet services, which today are being offered under multiple models, e.g., *Software as a Service (SaaS)*, *Container as a Service (CaaS)* or *Function as a Service (FaaS)*. These flavors provide the support for decentralised applications and data centres, which in the future will be scattered across the Cloud and Edge. Note that this higher distribution and lower size will increase flexibility but it will also imply higher costs and lower efficiency, and it is estimated that Data Centres' energy requirements will increase from current 2% to up to an 8% of worldwide electricity consumption in 2030. Moreover, Fog/Edge Computing and Mobile Edge Computing are expected to provide a highly distributed computing environment where the applications, services, or data processing are executed in the close proximity to consumers, in order to reduce latency when accessing subscribed resources.

With the further decentralisation of Internet services, the notion of Edge is becoming more flexible. This trend is occurring on both consumer and industrial domains. For instance, in consumer domains one can consider the case of solutions that assist the end-user in her daily routine, e.g., vertical markets of Smart Home, Smart Health, Smart Cities, etc. In this context, applications and services consider more and more aspects derived from daily roaming or even from personal habits. It is feasible to consider, for instance, that personal assistant software in smartphones can be extended to be able to interact with other (allowed) applications, as well as with other devices along cities, countries crossed by users. In such case, vehicles themselves become part of the network, and functions such as traffic offloading can occur in a way that is derived from the integration of machine learning aspects to perform, for instance, mobility anticipation.

In industrial domains, Edge computing is also covering more flexible notions to reach a higher degree of automation to support decentralised computing across the Edge and the Cloud, for instance. Consequently, the underlying networks will have to deal with a huge amount of data and service component from the Edge generated by users, machines and sensors. Cisco estimates that 850 ZB of data are expected to be generated at the Edge by 2021 [C5-1]. The management of this data, based on traditional, centralized Cloud approaches, will create problems and increase costs and energy consumption [C5-2]. In fact, moving and processing all the data will require massive and highly performing infrastructures, e.g., high speed connections, or large data centres for storage and computation.

---

[3] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[4] https://www.gartner.com/en/newsroom/press-releases/2018-12-03-gartner-says-the-future-of-it-infrastructure-is-always-on-always-available-everywher

The booming of AI services, and of decentralised AI services, will further exacerbate the problem. With the recent advances in hardware that allow to run deep neural networks, deep learning-based solutions are expected to be massively adopted in many services, e.g., to enrich people's lifestyle awareness, to improve human productivity or to enhance social efficiency, among others. However, training a single complex AI model may lead to an equivalent carbon footprint of 284 tonnes, i.e., five times the lifetime emissions of an average car [C5-3]. Therefore, it will be important to concentrate the effort of both industry and academia to promote research and development of more computationally efficient algorithms. In this context, there are two main aspects to consider within the umbrella known as "*Edge AI*" or "*Edge Intelligence*" [C5-4]: i) AI for the Edge; ii) AI on the Edge. The physical proximity of such huge amount of data jointly with the availability of computing resources in the Edge, fosters the idea that AI will be possibly executed, mostly, directly in the Edge.

On the one hand, AI for the Edge concerns solutions for the optimization problems in Edge computing, recurring to effective AI methods. The power of AI can be used, also, for solving the resource optimization problems of Edge computing and of networking management in general.

On the other hand, AI on the Edge concerns the use of AI algorithms on Edge services, as it is the case with the widely and increasingly popular federated learning [C5-5]. Of relevance here is the fact that Edge AI needs to consider how to run AI models on constrained devices and how to integrate synergies from end-user/field-devices up to the Cloud. These and other solutions for performing distributed model training on the Edge should be further studied to evaluate their pros and cons for different service and network requirements. Examples are Knowledge Transfer Learning, Gradient Compression, DNN Splitting, and Gossip Training. Similarly, for behaviour inference, there are today several solutions available in related literature; however, it is not yet clear how to apply such solutions from an Edge computing ecosystem perspective. Specifically, which might be their application according to specific service requirements (which are usually constrained, for instance, by hardware and by network requirements, such as mobility management).

The capability to assist in performing more complex validation, storing, and forwarding, in a way that complements the Cloud operations, implies also that an evolutionary step has to be taken in regards to the infrastructure technology that supports the Edge-Cloud computing continuum. In this context, container solutions, such as Docker[5] and the orchestration counterpart, e.g., Kubernetes[6], bring in relevant advantages in regards to computation isolation and portability. Containers run on top of physical or virtual devices sharing the host kernel, thus leading to what is also known as "OS-level virtualization". In comparison to VMs, containers provide a lightweight and agile solution to hold the Edge workload. The containerized Edge is already here, in particular considering applications such as distributed AI analytics platforms, IoT edge solutions, and augmented reality.

There are, however, aspects that still need to be researched. A major gap in this context concerns an automated management of different clusters of containers across the Cloud and the Edge, i.e., to further enhance the management of CaaS solutions. Here, aspects such as the integration of AI to improve the degree of automation should be considered. Another gap concerns the need to further research the feasibility of deploying containers in embedded devices. For this case, there is the need to address both the container technology as well as the management of containers assuming constrained devices and mobility of devices. A third

---

[5] https://www.docker.com/resources/what-container.

[6] https://kubernetes.io/

gap to address in this aspect concerns a secure design for the containerized Edge, not just in terms of data privacy, but also in terms of channel security.

- Summarising, Edge Computing is essential to handle the high volumes of complex data being currently produced at the Edge, derived from IoT industrial and consumer environments, and which are today essential to provide adequate estimations of the different aspects of services and network operation. By keeping the data and its processing whenever feasible closer to the data sources, Edge Computing provides several benefits: lower computational, energy and memory requirements, as well as lower latency and improved security, derived from measures of proximity to data sources. It is envisioned that the data processing can be provided based on decentralised AI platforms that rely on resource-constrained hardware.

5.2 ETSI MEC evolution

ETSI MEC is the leading effort for the standardization of edge computing at a European level, with more than 100 members participating in the industry study group. ETSI MEC specifications are currently in a mature state, with release 2 having been recently published, and they include a reference architecture and a set of application-agnostic and vendor-neutral interfaces. The latter have been defined with a modern approach, following a REST paradigm and adhering to the OpenAPI specifications (https://www.openapis.org), which makes it possible to automatize code development and interoperability testing. Initially intended for mobile broadband systems only, the ETSI MEC is now multi-access and it also supports fixed access, WLANs, and IoT gateways.

It is expected that ETSI MEC will evolve in two orthogonal directions: internally, by increasing the number of MEC services supported (a MEC service is offered from the telco operator to the third-party applications hosted on edge nodes) and fostering a collaborative environment for the adoption of the standard (via hackathons, testing suites, open source emulators, etc.); and, externally by influencing the design and realization of edge computing solutions through liaisons and joint activities. The latter is considered of particular relevance, because the edge technologies today are highly fragmented, and it is unlikely that the telco industry will converge towards a single solution universally adopted. Noteworthy examples of open source frameworks with relationship with ETSI MEC are OpenNESS (https://www.openness.org) and Akraino Edge Stack (https://www.lfedge.org/projects/akraino), which are toolkits for the deployment of cloud services on edge premises.

Moreover, the evolution of ETSI MEC is tightly bound to advancements in other communication standards. The 3GPP 5G specifications introduced a Service-Based Architecture (SBA) [C5-12] which supports flexible procedures to offer and consume services, both third-party applications and network functions. As far as the latter are concerned, i.e., network functions, the transition of MNOs to Virtualized Network Functions (VNFs) is in a mature phase, and it is covered by dedicated standardization efforts, championed in Europe by ETSI NFV. Within ETSI MEC there are ongoing studies on the integration with both the 5G SBA (GR MEC 031) and ETSI NFV (GR MEC 017), hence we can expect in the next years a full convergence towards a modern architecture for both core and edge networks, relying on technologies that are already widely adopted in cloud and service infrastructures, e.g., from the Cloud Native Computing Foundation (https://www.cnfc.io/). Realizing this vision will require concerted efforts at R&D level in several areas, especially to support use cases with a very high economic impact but specific requirements, such as IoT [C5-13] or connected vehicle [C5-14] applications (e.g. vehicular micro-clouds, where cars can establish a dynamic cloud-like infrastructure that can act as a virtual edge server), and to facilitate a seamless integration with other emerging technology trends, such as network slicing [C5-15] containerization [C5-16] and serverless computing [C5-17]. Overall, an open European approach, founded on an open and collaborative ecosystem, can provide a realistic opportunity to compete with commercial offers for edge computing from US OTT players, e.g., Microsoft Azure Edge Zone or Amazon AWS Wavelength.

Another research branch works on extending mobile edge computing to using mobile ICT (computational, storage, etc) resources. One example, which is already quite mature is the concept of vehiuclar micro clouds. Here, cars establish a dynamic cloud-like infrastructure that can act as a virtual edge server [2]. Such virtual edge servers or micro clouds have been shown to be rather stable [3] and can be used for classical task offloading similar to classical MEC servers [4].

5.3 Activities on MEC in other Standardization Bodies

Several other standardization bodies and fora are addressing the ecolution of Edge Computing. A list of said standardization bodies and fora include:

- ETSI ISG MEC
- IIC (OpenFog)
- ONF CORD
- Telecom Infra Project — WGs on Edge Computing
- Linux Foundation Edge (Akraino, EdgeX Foundry)
- Open Edge Computing
- MobiledgeX

Overall, it appears that all these initiatives are approaching Edge Computing from different points of view and do not fully overlap, but in most cases, are complementary. There might also be overlap due to domain specific rather than competitive or supplementary nature of these activities. We recognize the different activities in Edge Computing domain but these are specific to certain use cases for domains within an industry which will allow Edge Computing to be an alternative to the old cloud computing model.

Global interoperability appears to be a “must” for enabling new services ecosystems, and this means that Industry needs to align on open and common APIs capable to ease Service and Apps Developers: in fact, this is crucial to promotes innovation and accelerate development by Third Parties applications and services, capable of enabling Network and Service Providers to capitalize on their investments on Edge Computing.

Moreover, it should be mentioned that some interfaces of the proposed architectural solutions (IaaS, PaaS or SaaS individually or in combination) may be open whereas in some other cases these can be proprietary solutions and are treated as black boxes. The open interfaces are standardised whereas the black box ones might not necessarily need standardisation.

5.4 NFV, SDN, orchestration

Several technological trends will affect protocol development. These include ultra-low latency end-to-end communication, capacity of access links and diversification of infrastructure for underlying 5G networks (e.g., Visible Light Communication links, mm-wave links, new WiFi standards). At the same time, Internet communication patterns are changing, and increased flexibility of in-network devices and networking software in end-hosts is becoming the new norm. Security, privacy and trust have also moved from being an afterthought in the design of new communication protocols to an absolute necessity in the face of an evolving cyber threat landscape. Several of these trends conflict with the traditional layering in the Internet and TCP/IP protocols. Some developments that partially address these issues have surfaced, such as Information Centric Networking (ICN). Mainly US industry has been developing methods to improve performance, security and flexibility of the Internet's transport layer but it is unclear whether these point solutions will satisfy the needs of upcoming and future applications, and be suitable for 5G network technologies and beyond. Greater flexibility in end systems and inside the network is a necessity, and Internet transport protocols will have to be exchangeable at run time. Improved interplay between applications and the underlying network will also be necessary. AI/ML and data analytics will be key drivers of self-adaptation and self-management, but such solutions are still in their infancy and require focused research efforts.

Energy consumption will likely return among the main concerns in this framework. The increasing use of general-purpose hardware to implement virtual network functions chains and of federated compute elements to perform complex computational tasks in the edge or in cooperation between edge and cloud needs energy efficiency to be accounted for in conjunction with performance optimization. Methodologies that need to be considered include hardware offloading, but also all the techniques for the control of energy-performance tradeoff, like power-aware resource allocation, power adaptation and low-power idle mechanisms [C5-18] [C5-19].

Furthermore, in order to achieve a tight proportionality between the dynamic workload produced by vertical applications and network slices on one side, and the run-time resource utilization footprint and energy consumption induced in the network, computing, and storage infrastructure, on the other, two main aspects should be considered:

- The shift from classical "cloud agility" schemes, where vertical applications/network services are deployed everywhere they should be available and scaled according to the incoming workload, to a novel "edge agility" paradigm, which will inject applications/services (or part of them) at run-time, only when, where and for the time needed by end-users and connected things, suitably and smartly adapting the virtual topologies and application/slice forwarding graphs;
- The introduction of a "hybrid multi-tier elasticity" paradigm to make all stakeholders, acting at any levels, to jointly, automatically, and rapidly provision, adapt, and de-provision network, computing, and storage resources/artefacts in order to opportunistically "move the consumption from connectivity to (edge) computing".

5.5 Computing platform technologies

Serverless computing is an emerging trend in cloud technologies, where applications are realized as a composition of short-lived and stateless function calls (called lambda functions) [C5-20]. Due to the absence of local state and in combination with a flexible virtualization infrastructure, typically based on containers, it is possible to up-/down-scale services in a fast and easy manner, also achieving fine-grained billing granularity. Serverless computing is becoming relevant also in edge domain, with a growing importance due the widespread adoption of its related Function-as-a-Service (FaaS) programming model among cloud developers and the clear benefit of flexibly deploying services to edge or cloud depending on run-time conditions [C5-21]. Moreover, due to the scalability promises and isolation guarantees offered, it is expected that serverless computing will also play a crucial role in NFV, especially for the implementation of large-scale SDN controllers and VNFs [C5-22]. However, significant research efforts will be required to unlock the full serverless benefits in edge computing and NFV, in particular major concerns include a high data-plane latency and inefficient handling or transfer of the application state [C5-23].

Microservice architecture is becoming the most popular architectural model to implement highly distributed and scalable services. The main idea has been in decomposing large monolithic systems or services into a set of loosely coupled microservices that usually take care of a single function or a set of closely related functions. This, in principle, allows for developing and deploying each microservice independently and thus makes it possible to compose services using modules from multiple vendors. As discussed in section 1.2.1, 3GPP has adopted microservice based SBA in 5G core and adoption in MEC is likely to follow.

The REST architecture is currently the dominating way to implement microservices and it is also selected for 5G SBA. However, REST has certain shortfalls in efficiency, scalability, reliability, and API evolution that makes it necessary to seek improvements especially for, e.g., safety critical applications. Partially, these problems are related to current implementations where REST relies on HTTP/1.1 transport and thus the obvious solution would be upgrading HTTP to v2 and later to HTTP/3. It is planned that the latter one will use QUIC instead of TCP and thus provide, e.g., multipath connectivity. There are also initiatives to replace REST with more efficient protocols like gRPC, which would allow for backward-compatible schema evolution that is necessary for long-live evolving systems like 5G networks. In any case, even if the REST based architecture is becoming the standardized solution for 5G, it is rather important to research if adopting more advanced concepts would be necessary it the future.

Container-based computing platform virtualization is currently the main trend in cloud computing and also in 5G edge computing. As explained in the next section (5.6), containers have many desirable features that make them quite ideal in implementing microservice architectures. However, there are always some security concerns about sufficient isolation and size of potential attack surface. While legacy hypervisor-based hardware virtualization systems running full OS images are considered to have unnecessarily large image sizes and long service start-up times, there are new microVM systems like Firecracker that can provide much more streamlined systems. Combined with unikernels such microVM systems are able to compete with, or even exceed, image sizes and start-up times of container-based solutions. While container-based solutions are the current main-stream and essential to 5G network infrastructure evolution, it is necessary to evaluate the alternative approaches in case that, e.g., better isolation between service instances is needed for critical services.

In cloud computing and big data applications, stream processing has become a popular solution to manage and process vast streams of data and events. Stream processing systems like Kafka Streams, Flink, and Apache Samza are used either to consolidate data streams prior to, e.g., AI/ML model training, or to implement complex event processing services. Concepts like Reactive Streams are developing standardized frameworks for asynchronous stream processing. In one end, stream processing can be performed near real-time with high scalability. While stream processing has been mostly done in centralized cloud environments, the distribution of computation towards network edge makes it important to consider stream processing also in Edge computing. One obvious idea is to distribute stream processing to edge nodes, i.e., stream processing graphs should be embedded into distributed computing resources between far edge and central cloud. Besides providing services for data processing and real-time services, it should be studied if stream processing concepts could be utilized in implementing 5G network functions, e.g., for network monitoring.

5.6 Containers and container orchestration

Container-based solutions provide the grounds to bring novel, bandwidth and computationally intensive services to the Edge, an aspect which becomes particularly relevant in scenarios such as IoT, AI at the Edge, or AR services [C5-13]. Via container-based solutions and their clustering, it is feasible to implement decentralised applications closer to data sources and to address the migration of applications from the Cloud to the Edge in a way that takes into consideration for instance, context-awareness [C5-14] A second benefit of container technologies in the context of Edge-Cloud operation concerns is being able to run services and microservices in embedded devices, thus reducing costs and reducing energy consumption. The portability and lightweight design provided by containers are additional benefits to take into consideration.

In the context of 5G, containers are being considered to facilitate virtualization alongside with NFV [C5-15], via the concept of Cloud-native Network Functions (CNF)[7], derived from efforts from the Linux Foundation (LF) in the context of its group *Cloud-Native Computing Foundation (CNCF)*. CNF is addressing the possibility to transform network functions into microservices that provide networking services. These microservices are deployed in containers, e.g., Docker, and orchestrated with tools like Kubernetes. Such service-based approach is integrated into the 5G *Service-Based Architecture (SBA)* framework and therefore relevant for the context of network slicing, for instance. Also relevant is the fact that CNFs are expected, due to their integration into container-based solutions, to be runnable in bare-metal. Containers are therefore key in regards to deploying a 5G core that can operate in any Cloud-Edge environment, be it based on physical or virtual machines.

The key enablers for container technologies in 5G are:

- Open-source and user-friendly.
- Lightweight and fast deployment. A Docker container can be deployed in seconds, for instance.
- Portability and scalability. The isolation provided by container solutions make it a basis technology for deploying distributed applications in a modular way (where micro-services can be placed in individual containers). Containers can be added during operation, thus facilitating scalability aspects.

In terms of challenges, there are two main aspects for the full deployment of a 5G Edge-Cloud continuum: i) providing an adequate CNF specification and implementation[8]; ii) evolving orchestrators e.g., Kubernetes, towards adequate automation of clusters and federations of clusters, taking into consideration aspects such as security by design as well as mobility.

---

[7] https://github.com/cncf/cnf-testbed

[8] https://github.com/cncf/telecom-user-group/blob/master/whitepaper/cloud_native_thinking_for_telecommunications.md#1.5.4

5.7 Distributed services

There are several hard challenges in designing, implementing and operating distributed services in heterogeneous computing environments, especially with various constraints like end-point location and mobility, real-time requirements, and reliability. Furthermore, while the microservice architecture makes it easier to build efficient and reliable service agents, there will be a new layer of complexity that arises from connections and inter-dependencies between various service agents. Single service sessions are most often not very long-living which means that the state, e.g., resource requirements, is not static but sessions are leaving and entering the network all the time. This means an additional challenge to network and service optimization as new requests have to be embedded to computing resources without exactly knowing what will happen in the near future. In the network edge (especially far edge), this will be more challenging as fluctuations in service requirements will be much larger (in relative sense) and faster compared to centralized cloud.

Until now, this kind of resource allocation problems have been addressed in, for instance, transport networks by using specific path computation elements. Similar heuristics can be used as a starting point when methods for near-optimal service chain embedding are designed. Dependability issues with complex microservice architectures should be considered with great care – fatal error cascades in complex networks are not necessarily caused by simple design errors or software failures but they may originate from the normal engineering design process that aims towards optimal fault-tolerance (Highly Optimized Tolerance). Thus, the modelling and design methods for highly inter-connected microservices and their embedding to physical network structure should be one of the focus areas in developing federated edge computing environments.

5.8 Edge, Mobile Edge Computing and Processing

These approaches require responsive network connectivity to allow “things” and humans to touch, feel, manipulate and control objects in real or virtual environments. Edge processing in the architecture is essential for ultra-low latency and reliability, while the AI processing is transferred at the mobile/IoT device. Research challenges in this area cover open distributed edge computing architectures and implementations for IoT and integrated IoT distributed architectures for IT/OT integration, heterogeneous wireless communication and networking in edge computing for IoT, and orchestration techniques for providing compute resources in separate islands. In addition, built-in end-to-end distributed security, trustworthiness and privacy issues in edge computing for IoT are important, as well as federation and cross-platform service supply for IoT.

Similarly, distributed service provisioning will extend also even beyond the edge, i.e., to on-premises devices such as Industrial IoT devices, robots, AGVs, connected cars. Novel forms of dynamic resource discovery, management and orchestration are required, allowing service provisioning to exploit on-premises devices as “on-demand” extensions of resources provided from the core or the edge. In this framework, novel resource control schemes, balancing between autonomy of devices and the overall optimization and control of the network by the operator(s) will be required, thus innovating the existing collaboration models between different network service providers. This will also allow to take in better account users’ context, exploiting the typical co-location of users with on-premises devices and, sometimes, their very tight physical bound. In this sense, this approach will allow designing network services in a more user-centric way.

IoT Distributed and Federated Architectures Integrated with 5G architecture and AI: Further research is needed in novel IoT distributed architectures to address the convergence of (low latency) Tactile Internet, edge processing, AI and distributed security based on ledger or other technologies, and the use of multi-access edge computing. Research challenges include serving the specific architectural requirements for distributed intelligence and context awareness at the edge, integration with network architectures, forming a knowledge-centric network for IoT, cross-layer, serving many applications in a heterogeneous network (including non-functional aspects such as energy consumption) and adaptation of software defined radio and networking technologies in the IoT.

5.9 Edge AI

According to [C5-1], 850 ZB of data will be generated by people, machines and sensors at the network edge by 2021. The physical proximity between the data and the computational resources provided by the edge computing represents a promising marriage, the so-called edge intelligence or edge AI [C5-26]. Moreover, the recent booming of deep learning has been achieved thanks to the innovations in hardware, which allows to manage neural networks of many layers. However, these networks need more data in order to learn the huge number of parameters they are composed of. Moving these data toward a centralized cloud can be very inefficient in terms of delay, cost and energy. Therefore, in order to efficiently exploit data on the edge, the scope of edge AI is twofold: run AI models (inference) and train AI models (training).

For what concern the training, the main problem of a distributed solution is the convergence of a consensus, i.e., whether and how fast the training can be considered finalized. This problem is related on how the gradient is synchronized and updated. Several solutions have been proposed in this respect, the most promising one being represented by federated learning [C5-29]. In this solution, the server is in charge of combining the results of the training of a shared model. Specific gradient methods have to be used, like the Selective Stochastic Gradient Descent [C5-30], which however is not optimized for working with unbalanced and non i.i.d. (independent identical distribution) data. The frequency of the updates of the model at the central server is also an open issue. Too frequent updates allow to relax the hardware constraints of the edge, but imply more risks for the unreliable network communications. An interesting approach to overcome this issue is the Blockchain Federated Learning [C5-29], which allows to work without a central server by performing the updates via blockchain. Another interesting solution is the Knowledge Transfer Learning, where a teacher network is trained with general data and then student networks are retrained on a more specific local dataset. This allows to reduce the resource demand at the edge devices.

For what concern the inference, the main problem is the limited resources of the devices at the edge. In this case the solutions try to relax the computational requirements of the model when performing the inference. In model compression, some of the weights can be pruned according to a specific policy, e.g., their magnitude [C5-32], the energy [C5-31]. In model early-exit the inference is performed only with a subset of the network, according to the latency requirements. While, for reducing the computational complexity on the device model partition and input filtering represent interesting solutions, which rely on pre-processing the data on the device and perform the inference at the edge. When considering the processing of the original data, another technique that edge AI will need to investigate accurately on is data curation, which is the process of selecting the subset of data that is really valuable, especially when it comes from heterogeneous sources.

Particularly in the perspective of distributed services at the edge and beyond (see 5.8), edge AI is also a technique to keep data local to devices of their legitimate owners for privacy or ownership reasons such as, for example, data related to manufacturing processes in industrial environments. Moreover, keeping models “close” to edge devices might be the only viable solution to guarantee stringent time constraints. A research challenge is therefore how to guarantee accuracy and efficiency of both the training and inference phases given specific constraints in terms of where data can be moved inside the network.

AI will play an important role also for providing solution to the resource management problem in edge computing, the so-called AI for the edge, which is complementary to the problems above, where the issue is how to carry out the AI process on the edge (AI on the edge). Typical examples are radio resource management in wireless networking, computation offloading strategies and services placement and caching. In this case the challenges are on the model definition, which often has to be defined as a tractable Markov decision process, on the algorithm deployment, since it has to work on-line and, consequently, a trade-off between optimality and efficiency has to be found.

A possible distinction regarding in particular the application of AI parametric approximation models adopted for control and resource allocation purposes on the edge may be between “function approximation” and “parametrized infinite dimensional (or “functional”) optimization” [C5-24]. The (functional) solution of many complex control and decision problems can be approximated by families of fixed structure parametrized functions, where parameters also appear within the basis functions themselves (e.g., one- or multiple-hidden-layer networks). If a family of approximating functions can be found that allows avoiding the so-called “curse of dimensionality” (the growth in the dimension of the parametrization with increasing number of variables the function to be approximated depends on), the optimization problem might be solved “off-line” (e.g., in the background in the cloud), whereas the “local” implementation of the decision strategies can be performed at almost negligible computational cost at the edge, over time frames within which the parameters do not vary. However, a possible problem to consider in this case would still be the transfer of big amounts of data to be processed. In this respect, techniques of local data aggregation and pre-processing, redundancy reduction, importance sampling and the like are worth investigating in this context. On the other hand, distributed computational methods for the local coordinated execution of parametric optimization techniques are also of interest, to perform the strategy approximation over limited computational resources in the edge. It is also worth noting that parametric approximations of infinite dimensional (functional) optimization problems can be based on sound problem formulations, which can help understanding their algorithmic behaviour.

AI techniques and methods are necessary for IoT in an edge computing environment to provide advanced analytics and autonomous decision making. AI encompasses various, siloed technologies including Machine Learning, Deep Learning, Natural Language Processing, etc. In future IoT applications, AI techniques and methods will be increasingly embedded within several IoT architectural layers to strengthen security, safeguard assets and reduce fraud. Research challenges overlap with topics identified earlier in this document but it is worth mentioning AI-IoT integration subjects at the “edge” such as new energy- and resource-efficient methods for image recognition, edge computing implementations (neuromorphic, in-memory, distributed), distributed IoT end-to-end security, swarm intelligence algorithms, etc.

Finally, in the design of AI solutions it will be crucial to consider the energy consumed, as anticipated in the introduction. The high energy requirements of deep learning solutions suggest that both industry and academia promote the research of more energy efficient AI algorithms [C5-32]. Moreover, all the new proposed AI solutions should be presented with their training time and computational resources required, as well as model sensitivity to hyperparameters. Examples of such analysis are the characterization of tuning time, which could reveal inconsistencies in time spent tuning baseline models compared to proposed contributions. To this respect, tools like Machine Learning Emission Calculator [C5-33] and Green Algorithms [C5-34] should be used to analyse, audit and report the carbon footprint of novel solutions proposed.

6. Radio Technology and Signal Processing

Whereas the current 5th generation wireless system provides basic support for a multitude of services, incl. enhanced mobile broadband (eMBB), massive machine type communications (mMTC), ultra-reliable low-latency communication (URLLC), V2X, etc, the beyond 5G/6G system is expected to meet significantly higher and more stringent requirements, such as Tbps data throughput, sub-ms latency, extremely high reliability, everywhere mMTC, extreme energy efficiency, very high security, cm-level accuracy localization, etc [C6-1].

A straightforward way forward to deal with these challenges is to consider more and higher electromagnetic spectrum such as the sub-THz or THz spectrum, infrared and visible light spectrum, and the relevant/specific transmit and receive technologies. Besides, the centimetre and millimetre spectrum currently utilised for 5G and other legacy wireless systems need to be re-farmed and more efficiently reused, and co-existence issues to be carefully addressed. In addition to further enhancing the widely used technologies (such as waveform, modulation and coding, non-orthogonal multiple access, full-duplex, massive MIMO, etc) to approach the theoretic limits, e.g. in terms of spectral and energy efficiency, investigations are needed, e.g. for intelligent reflecting surfaces, integrated positioning, sensing and communications, random access for massive connections, wireless edge caching. Moreover, machine learning (ML) and artificial intelligence (AI) as a tool has been successfully applied in many applications. For the application in communication technologies and radio interface design, careful study is required.

This chapter aims to address the enabling technologies for the next generation radio interface, including

1) Spectrum re-farming and reutilisation, as well as co-existence;
2) Millimetre wave systems;
3) Optical wireless communications (OWC), especially VLC;
4) Terahertz communications including new materials (graphene);
5) Massive and ultra-massive MIMO including intelligent reflecting surface;
6) Waveform, non-orthogonal multiple access and full-duplex;
7) Enhanced modulation and coding;
8) Integrated positioning and sensing including radar;
9) Random access for massive connections;
10) Wireless edge caching for further increased spectrum and energy efficiency.

6.1 Spectrum Re-farming and Reutilisation

Allocated frequency spectrum is one of the main factors that determines the system capacity. However, radio spectrum is a very scarce resource. Especially the lower frequency bands are precious and tightly regulated. In order to satisfy the high bandwidth demands of upcoming generations of mobile systems, it is crucial to reutilise the existing spectrum resources. While the traditional approach allocates a dedicated spectrum to each radio access technology (RAT), spectrum reutilisation between RATs offers a more efficient and flexible utilisation of resources, e.g., for load-balancing. Spectrum reutilisation, also known as spectrum sharing, can be applied to both licensed and unlicensed bands.

A straightforward approach to inter-RAT spectrum reutilisation is *spectrum re-farming*. Re-farming performs static allocation of spectrum resources to different RATs. This method was already used to clear GSM spectrum to make it available for 3G. Because of its static nature, it has a poor spectrum utilisation.

A more efficient utilisation is achieved by dynamic inter-RAT resource scheduling with optimised multi-RAT handover and interference coordination. Preferably, this is based on a centralised multi-RAT radio resource management. The signalling overhead can be reduced by decentralised strategies.

For the joint utilisation of licensed and unlicensed spectrum, adaptive strategies are required such as cognitive radio concepts, in which spectrum awareness, e.g. based on a combination of advanced SIGINT (signal intelligence) techniques and AI (artificial intelligence) mechanisms, can be employed and co-existence with existing (e.g. analogue) services should be ensured. Such considerations and new concepts for spectrum licensing and reutilisation are particularly important in the context of new radio technologies such as the millimetre wave, optical wireless, and terahertz communications discussed below, which have a radically different interference footprint compared to the conventional sub-6GHz communications. Their highly directional links and susceptibility to blockage reduce interference, which significantly increases the potential gains of spectrum sharing and simplifies its use [C6-2].

Spectrum reutilisation is supported by multi-RAT connectivity, which allows the UE (user equipment) to choose the best RAT depending on the link qualities. This added diversity not only increases the performance due to better spectrum utilisation, it also makes the network more robust and resilient towards shadowing effects, hence improving the reliability and availability. On the other hand, spectrum awareness and reutilisation can help to increase security at radio level, e.g. through detection and countermeasure of threats such as RF jamming or spoofing.

Future networks will support different services, enabled by network slicing based on a multi-RAT radio access. Multi-RAT connectivity can also make flexible use of licensed and unlicensed bands. E.g., data and voice traffic can be offloaded to WiFi or LTE small cells operating in unlicensed bands as an enhanced mobility concept. Hence, utilising unlicensed bands is important and technologies to bring the quality to the licensed spectrum level are open to study. This not only increases the overall throughput but also enables low latency.

Network slicing and edge network function virtualization (NFV) also contemplate multi-RAT operating scenarios, based on highly reconfigurable software defined radio (SDR) hardware featuring heterogeneous processing resources (i.e., general purpose processing elements tightly coupled with hardware accelerators). The functionality of such agile SDR units could be updated at run-time according to traffic context, signal propagation conditions and required performance (e.g., in terms of throughput, latency, and resiliency). An efficient way to achieve field updates of this type is by jointly optimizing the multi-RAT radio and processing resources through suitably selected machine learning (ML) techniques.

To evaluate these complex multi-RAT scenarios, open source simulation models for 4G and 5G technologies from 3GPP releases and different IEEE standard amendments in multiple bands, are needed for an end-to-end and high-fidelity evaluation of smart solutions. The simulation models need to capture the wide range of spectrum considered for communication services, e.g., from 0.4 up to 71 GHz for 5G NR Rel-17, and consider the multiple heterogeneous spectrum paradigms like licensed, unlicensed, dedicated and shared, which are to be harmoniously used through intelligent frameworks in order to take the best advantage of spectrum resources.

Existing *short-range wireless communication* technologies, including WiFi, Bluetooth and Zigbee, share the same spectrum, e.g. in 2.4GHz. Co-existence of different wireless network technologies in/near such a carrier frequency may cause radio interference, which can lead to relatively high error rate in data transmission. This problem happens especially in unlicensed bands. How to efficiently share the spectrum and improve the co-existence needs careful considerations. Scalability and power efficiency are critical for success of a macro, micro, or pico network. Current short-range communication technology provides either high throughput with high power, or low throughput and low power consumption. Whereas IoT devices operate in a very low power mode in most of the time, they need to support a short-time high bandwidth transmission. Scalability is needed to support both short-time high throughput transmission and low power transmission. A unified and scalable architecture will be beneficial to support both low data rate (e.g. with Bluetooth, ZigBee, RFID, NFC, etc) and ultrahigh data rate (e.g. up to 100Gbps within 10m coverage). Further requirements to be considered include, e.g. scalable network topology supporting P2P (point-to-point), MP2MP (multipoint-to-multipoint), as well as the smart home and smart building coverage; more power/cost efficient designs, e.g. for zero-power consumption in some dedicated scenarios; and the capability of information and energy simultaneously transporting (IEST).

The wide mmWave spectrum region accounts for different access paradigms, including licensed (e.g., 28 GHz bands), unlicensed (e.g., 60 GHz bands) and shared (e.g., 37 GHz bands) for various applications such as vehicular and cellular. Co-existence of multiple technologies and standards like 5G NR-U (NR in unlicensed), NR V2X (vehicle-to-everything communications) and 802.11ad, 802.11ay, 802.11bd in different spectrum bands should be properly addressed considering various regulatory requirements and access mechanisms. Innovative solutions that increase spectral and energy efficiency need to be considered [C6-3].

6.2 Millimetre Wave System

Millimetre wave (mmWave) systems have attracted large research interest in recent years due to the huge available bandwidth required to fulfil the today's traffic demand. This is reflected in WLAN and WPAN standards: in the license free 60 GHz band, the IEEE802.11ad WLAN standard provides rates up to 8 Gbps and the upcoming IEEE802.11ay WLAN standard will provide rates up to 30 Gbps. The fifth generation (5G) wireless networks aim to use mmWave in mobile networks, where the transmitter/receiver nodes may be moving, channels may have a complicated structure, and the coordination among multiple nodes is difficult [C6-4]. The 2018 Winter Olympics in Korea already provided first glimpse at the 5G services powered by Korea Telecom with support from global equipment makers. This show case included a 28 GHz mmWave backhaul network for moving hotspots, such as buses. Additionally, the mmWave band in combination with mobile edge computing (MEC) is highly suitable for on-demand content (multi-media) delivery services, hence enabling the enhanced mobile broadband (eMBB). This combination of mmWave and MEC is the only way to satisfy both extreme communications requirements of ultra-high speed and low latency at the same time. Beyond 2020, MEC is expected to enable automated driving using mmWave based V2X/V2V links. This requires, however, cooperative perception and the exchange of HD dynamic map information between vehicles and radio units, to enhance the visibility area. The automated driving use case can be considered as the most important application of mmWave and MEC, which requires both ultra-high speed and low latency [C6-4]. The high channel resolution due to the large bandwidth also makes mmWave technology a prime candidate for joint communication and radar, which can complement or replace conventional radar and lidar systems in autonomous vehicles [C6-5]. Mobile virtual and augmented reality application represent a further future use case with similarly high requirements that may only be met by mmWave technology [C6-6].

Beyond 5G, it is expected that the data traffic due to mobile nodes (smart phone and tablets) will be more than 100 petabytes per month by 2023 [C6-7], which is 10 times of the traffic in 2017. In Western Europe alone, the data traffic is expected to be as high as 12 petabytes per month, which amounts to 56 terabytes per person per month [C6-8], hence, offering a huge potential to exploit mmWave bands and even the Terahertz frequency range (not considered so far by 5G). It is projected in [C6-9] that the volume of traffic generated from smart phones will be 86% of the global data traffic by 2021 and among this more than 50% data will be offloaded to the fixed networks by means of Wi-Fi devices and small cells each month, while the remaining traffic will be covered by the cellular networks. In order to achieve high data rates, one would require a large amount of contiguous bandwidth suitable for communications over short ranges, that is to be found beyond 100 GHz, for example around 140 GHz. The use of these frequency bands provides an excellent opportunity, since many antennas can be packed in a small area to direct a beam to the intended user.

An important business case for mmWave is in so-called ‘smart factories – Industry 4.0’. Due to its ability for spectrum re-use that enables multi-connectivity for high reliability, mmWave provides a complementary solution to low frequencies. Additionally, due to the high penetration losses, mmWave is inherently more secure against eaves-dropper and is a suitable candidate for industry environments. An additional feature of mmWave is sensing/positioning with high accuracy. This allows detection with higher spatial and velocity resolution that is suitable for both V2X and industry automation scenarios.

Small cells are to play a key role to cope with the increasing traffic demands in mobile network. These small cells connect to the core network via wired or wireless backhaul links. The dense deployment of small cells and a variety of services offered by the RAN having diverse requirements on throughput, latency and reliability, poses new challenges on backhaul links. One way to address these challenges is self-backhauling using mmWave, i.e., the access and backhaul share the same wireless channel. 3GPP stage 1 in its Release 15 [C6-10] outlines the requirements for the self-backhauling in 5G networks. Among these requirements are the flexible partitioning of resources, autonomous configuration, multi-hop wireless connectivity, topology adaptation, and redundant connectivity.

One of the main challenges will be to manage the different network features introduced in 5G and developed beyond the first release and optimise them collectively. Diverse network components need to be integrated, such as D2D, self-backhauling, and multi-casting/broadcasting. While these technologies will be already available in 5G, the new challenge consists of extending them by advanced mmWave massive MIMO techniques, which are dynamically coordinated, considering interference and mobility. As mmWave network device density as well as the number of antenna elements and RF chains of devices increase, channel estimation overhead becomes a major challenge and intelligent solutions for initial access and beam training based on context information will be needed [C6-11].

Due to high propagation loss and severe atmospheric effects, mmWave communication are more suitable for short range communication and small cells. However, the increase in the system capacity through densification is not linear and therefore potential benefits of using high frequency bands for longer ranges and macro cells need to be considered. Designing macro cells without appropriate planning and optimization may limit the potential gains of mmWave communication [C6-12].

6.3 Optical Wireless Communication

Despite the tremendous improvements due to the small cell concept and the allocation of new radio frequency (RF) spectrum, the continued exponential growth in mobile traffic [C6-13] means that inevitably the RF part of the electromagnetic spectrum will not be sufficient to be able to drive the 4th industry revolution which is centred around data-driven economies and data-driven societies [C6-14].

It is, therefore, natural to consider the infrared and visible light spectrum, both of which are part of the electromagnetic spectrum for future terrestrial wireless systems. In fact, wireless systems using these parts of the electromagnetic spectrum could be classified as nmWave wireless communications systems in relation to Section 6.2. Light based wireless communication systems will not be in competition with RF communications, but instead these systems follow a trend that has been witnessed in cellular communications by inspecting all the generations developed during the last 30 years. Light based wireless communications simply adds new capacity – the available spectrum is 2600 times larger than the entire RF spectrum.

An important advantage is that off-the-shelf optical devices can be used to harness these unregulated and free transmission resources. By using advanced devices, lab demonstrations showed 8 Gbps from single light emitting diodes (LEDs) and 17.6 Gbps using laser diodes (LEDs) [C6-15]. Recently, a record of received data rates of 500 Mbps by using a single solar cell has been demonstrated. The use of these types of 'data' detectors has the appealing advantage of achieving simultaneous energy harvesting and high-speed data communication – a feature that will become ever more important in mobile machine-type communication (MTC) [C6-16]. By 2026, it is expected that micro-LED technologies and spatial multiplexing techniques will be mature and cost effective such that white light based on different wavelengths will unlock throughput, leading to potentially 100 Gb/s plus for ultrahigh-data-rate VLC access points [C6-17].

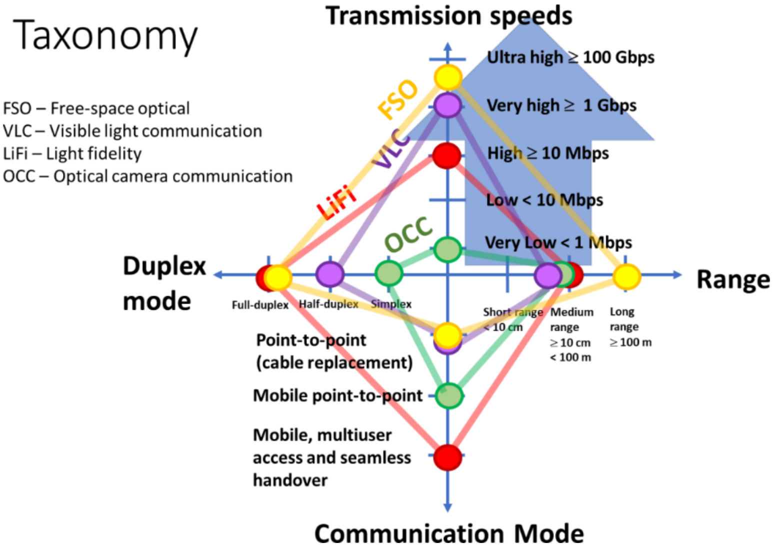


Figure 6-1 A taxonomy of emerging light communication technologies

Figure 6-1 shows a taxonomy of different light communication approaches. Free-space optical (FSO) is point-to-point long range optical wireless communications with target data rates of tens of Gbps primarily using laser diodes and coherent transmission. Visible light communication (VLC) has been used in the context of line-of-sight high-speed point-to-point communication, primarily using LEDs in conjunction with IM/DD. VLC systems are usually designed for ranges less than 100 m, as well as for bi-directional communication. Optical camera communication (OCC) in contrast is simplex communication using embedded CMOS camera sensors as data detectors. Due to the use of CMOS sensors, the achievable data rates are well below 1 Mbps. OCC is primarily used for indoor navigation, asset tracking and positioning. These applications assume some user mobility.

Cellular wireless networks which are based on VLC are referred to as LiFi (light fidelity) [C6-18]. LiFi enables bi-directional networked communication including multiuser access and handover. The blue arrow in Figure 6-1 indicates that the major research efforts in the last 15 years have been focused on enhancing link data rates of intensity modulated (IM) / direction detection (DD) optical wireless communication systems. With the advent of LiFi the research focus has begun to shift to challenges related to networking issues using light.

As in RF networks, there are issues surrounding interference management and interference mitigation in LiFi networks. However, since, for example, there is no multipath fading because the detector sizes are much larger than the wavelength, techniques developed for RF systems may only be sub-optimum. There are also fundamental differences as a result of IM/DD, in that signals can only be positive and real-valued. Consequently, new LiFi-bespoke wireless networking methods must be developed. Moreover, because light can be confined spatially by using very simple and inexpensive optical components, interference can be controlled much easier. This feature also allows step-change improvements of the small cell concept as single cells might cover sub-m^2 areas.

Furthermore, due to the extremely small wavelength, the active detector sizes are very small, and massive MIMO structures can be implemented at chip-level. This property can be used to develop unique and LiFi-bespoke MIMO systems, networked MIMO approaches, and new angular diversity techniques in conjunction with low computational complexity cooperative multipoint systems. Diversity techniques in LiFi systems are especially powerful to combat random blockages that naturally occur in a mobile scenario. Moreover, the spatial confinement of signals in LiFi enables the development of radically new physical layer security concepts.

LiFi is currently being standardised in a Task Group within IEEE 802.11. The new LiFi standard has received the following reference: IEEE 802.11bb. Similarly, VLC is being standardised in IEEE 802.15.13, while OCC has been standardised in IEEE 802.15.7r1.

Convergence with 3GPP access: LiFi communication is bi-directional. Due to the abundance of optical spectrum, typically the visible spectrum is used for the downlink by piggy-backing on lighting systems, while the infrared spectrum is used for uplink transmission. The simplicity of IM/DD in conjunction with advanced modulation techniques [C6-19] enable highly energy- and spectrum-efficient transmission systems suitable for the uplink. These modulation techniques are based on multicarrier approaches. Therefore, it could be argued that a *tight interaction between radio and optical components should be considered at the level of baseband processing*. Since OFDM transmission (e.g. 5G waveforms) is feasible on a free-space IM/DD optical link, it is definitely worth investigating the use of the same basic waveform and protocol stack for radio and LiFi systems. This would allow for a *common baseband processing platform* in both the small-cell transmitters and terminal receivers. Moreover, the 3GPP access-layer protocols are perfectly adapted to the use of downlink-only component carriers.

6.4 Terahertz Communication

Wireless data rates have doubled every eighteen months for the last three decades. Following this trend, Terabit-per-second (Tbps) links are expected to become a reality within the next five years. While mmWave communications are a step in the right direction, the total consecutive available bandwidth in such systems is less than 10 GHz. Consequently, supporting Tbps would require a physical layer efficiency of 100 bit/s/Hz, which is several times higher than the state of the art.

In this context, **Terahertz-band (0.1–10 THz)** communication is envisioned as a key technology to satisfy the need for much higher wireless data rates [C6-20]. This frequency band, which lies in between mmWave and the far infrared, supports huge transmission bandwidths: from almost 10 THz for distances below one meter, to multiple transmission windows, each tens to hundreds of GHz wide [C6-21], for distances beyond several tens of meters. However, this very large bandwidth comes at the cost of a very high propagation loss. Moreover, for many years, the lack of efficient ways to generate and detect THz signals has hampered the use of the THz-band in practical communication systems.

To date, different technologies are being considered to close the so-called THz gap. In an *electronic* approach, the limits of silicon CMOS technology [29], silicon-germanium BiCMOS technology [C6-23], and III-V semiconductor HEMT, mHEMT, HBT and Schottky diode technologies [C6-24] are being pushed to reach the 1 THz mark. In a *photonics* approach, uni-travelling carrier photodiodes [C6-25], photoconductive antennas [C6-26], optical down conversion systems [C6-27] or, more recently, quantum cascade lasers [C6-28] are being investigated for high-power THz systems. In both approaches, fundamental device limits are being reached, as the frequency is ”too high” for electronic devices and the photon energy is ”too low” for photonic devices to efficiently operate at *true* THz frequencies.

More recently, the use of **graphene to develop novel plasmonic devices** for THz communications has been proposed. Graphene is a two-dimensional (2D) carbon-based material that has excellent electrical conductivity, which makes it very well suited for propagating extremely-high-frequency electrical signals [C6-29]. Moreover, graphene supports the propagation of THz surface plasmon polariton (SPP) waves at room temperature. SPP waves are surfaced-confined electromagnetic waves generated by the global oscillation of electrons. By leveraging the properties of graphene, nano-transceivers [C6-30][C6-31] and nano-antennas [C6-32][C6-33] have been proposed and are being developed. These devices are intrinsically small, efficient to operate at THz frequencies, and can support very large modulation bandwidths. Moreover, graphene is ”just the first” of a new generation of 2D materials (such as MoS2 or Hb-N), which can be stacked to create new types of devices and leverage new physics. THz signal may also transmit through waveguide such as polymer microwave fibre (PMF) to provide high data rate in low cost short reach communication (e.g. datacenter interconnections).

In parallel to the development of new device technologies, there is a need to understand and model the THz-band channel. In the case of line-of-sight (LoS) propagation [C6-34], the main phenomena affecting the propagation of THz waves are the (i) spreading loss and the (ii) molecular absorption loss. The *spreading loss* accounts for the attenuation due to expansion of the wave as it propagates through the medium and is common to any wireless communication system. The *molecular absorption loss* accounts for the attenuation that a propagating wave suffers because a fraction of its energy is converted in vibrational kinetic energy in molecules (especially water vapour). In the case of non-line-of-sight (NLoS) propagation [C6-35], in addition to the two aforementioned phenomena, high reflection loss, diffused scattering and diffraction by obstacles need to be captured. Ultimately, stochastic multi-path channel models are needed to statistically characterise the channel. In addition, there is a need to understand the channel characteristics in mobile environments, see e.g. [C6-36].

In light of the capabilities of THz devices and the peculiarities of the THz-band channel, there is a need to develop new communication algorithms and networking protocols, tailored to THz communication systems. At the physical layer, new types of modulations are needed. For short-range communications (below one meter), the use of impulse-radio-like communication based on the transmission of one-hundred-femtosecond-long pulses following an on-off keying modulation spread in time has been proposed [C6-37]. Such very short pulses are already at the basis of many THz sensing systems and can be generated and detected with current technologies. For longer communication distances, new **dynamic bandwidth modulations** [C6-39] are needed to not only overcome but even leverage the unique distance-dependent bandwidth created by molecular absorption. Specialized (single-carrier) waveforms that are robust to phase noise, Doppler, and carrier frequency offset have been investigated as well [C6-38].

Independent of the modulation, and similar to any wired or wireless Tbps communication system, physical-layer synchronisation (both in time, frequency and phase) becomes a major challenge. The front-end non-idealities, e.g. non-linearity and phase noise, can severely impact the achievable throughput. Going to Tbps throughputs implies increasing the bandwidth to tens of GHz. This is another challenge for implementations for two reasons: first, ADCs and DACs in the tens of Gsamples/s are needed; second very wideband analogue baseband circuits are needed (half the RF bandwidth). The lack of digital-to-analogue and analogue-to-digital converters (DACs and ADCs, respectively) able to handle multi-GHz bandwidth signals, limits the application of traditional digital signal processing and motivates the research and development of new mixed (digital and analogue) techniques where some traditionally digital functions such as synchronisation or equalisation can be moved to the analogue domain. Additional challenges include new channel coding strategies, which leverage the uniqueness of the THz-channel, or physical layer security schemes for THz-signals. Very generally, efficiencies become dominant bottlenecks: at 1 Tbps, an efficiency of 1 pJ/bit (impossible today if we consider the whole PHY) translates into 1 Watt of power consumption; similar considerations about implementation efficiencies in silicon technology (area efficiency (bit/s/mm^2) and power density (W/mm^2)) show huge challenges at Tbps rates.

Similarly, many challenges arise in the higher layers of the protocol stack. At the link layer, novel **MAC protocols** are required for THz-band communication networks, since classical solutions do not capture the peculiarities of this band. The very large available bandwidth almost eliminates the need for nodes to contend for the channel. The transmission of very short signals also minimises the chances for collisions. All these come at the cost of more complex synchronisation schemes between devices. Ideas to be explored for new MAC protocols include, among others, the development of receiver-initiated transmission schemes to ensure that the transmitter does not waste resources when the receiver is not available, especially when highly directional systems are used. Additional challenges also include packet size optimisation and adaptive error control strategies.

At the network layer, new **routing mechanisms** could be developed that take into account the availability of both classical active relaying nodes as well as novel passive dielectric mirrors, which can direct the signal towards its final destination. In addition, new routing metrics that consider the channel molecular composition and its impact on the available distance-dependent bandwidth need to be explored. At the transport layer, as wireless multi-Gbps and Tbps links become a reality, the aggregated traffic flowing through the network will dramatically increase. These will introduce many challenges at the transport layer regarding **congestion control** as well as end-to-end reliable transport. For example, we expect that a revision of the TCP congestion control window mechanism will be necessary to cope with the traffic dynamics of THz-band communication networks.

For the validation and refinement of the developed solutions, new **experimental platforms** and integrated testbeds will be needed. For the time being, these are mainly focused in the sub-THz windows (300 GHz, 650 GHz), but systems at *true* THz frequencies will be required. Finally, in parallel to all the scientific developments, work needs to be done towards regulation and standardisation of the THz-band [C6-40].

Smart devices will be able to sense and monitor the environment based on the radiated communication signals. This will enable smart context-aware networking and new applications, which are based on the knowledge of certain features of objects in the surrounding, e.g. position, velocity, structure, or used frequency bands. Sensing not only includes the recognition of devices that radiate THz signals, but also passive imaging techniques can be integrated. Environment-aware communication will be a key component of next generation smart networking, which will push the integration of data analytics to a new level.

6.5 Massive and Ultra-Massive MIMO

The grand challenge for mmWave, THz-band and optical communications is posed by the very high and frequency-selective path loss, which easily exceeds 100 dB for distances over just a few meters in the presence of LoS (line-of-sight) and becomes even worse in NLoS (non-line-of-sight) conditions. As a result, high-gain directional antennas are needed to communicate over distances beyond a few meters.

Similarly, as in lower frequency communication systems, antenna arrays can be utilised to implement MIMO systems, which are able to increase either the communication distance by means of beamforming, or the achievable data rates by means of spatial multiplexing. In the last few years, the concept of **massive MIMO** (mMIMO) has been introduced and heavily studied in the context of 5G systems [C6-41][C6-42][C6-43]. In such schemes, very large antenna arrays with tens to hundreds of elements are utilised to increase the spectral efficiency to communicate over a large distance. In these arrays, it is important to take mutual coupling between the antenna elements into account in a physically consistent way [C6-44][C6-45]. Very large antenna arrays have been proved to be very useful for mmWave communication systems [C6-46][C6-47]. When moving to the THz-band, antennas become even smaller and more elements can be embedded in the same footprint. However, linearly increasing the number of antennas is not enough to overcome the much higher path loss in THz-band.

In this context, the concept of **ultra-massive MIMO (um-MIMO)** communications, enabled by very dense plasmonic nano-antenna arrays, has been recently introduced in [C6-48] and [C6-49]. Instead of relying on conventional metals, nanomaterials and metamaterials can be utilised to build plasmonic nano-antennas (see Section 6.4), which are much smaller than the wavelength corresponding to the frequency at which they are designed to operate. This property allows them to be integrated in very dense arrays with innovative architectures. For example, even when limiting the array footprint to 1 mm × 1 mm, a total of 1024 plasmonic nano-antennas designed to operate at 1 THz can be packed together, with an inter-element spacing of half plasmonic wavelength. Such plasmonic nano-antenna arrays can be utilised both at the transmitter and the receiver (1024×1024) to simultaneously overcome the spreading loss problem (by focusing the transmitted signal in space) and the molecular absorption loss problem (by focusing the spectrum of the transmitted signal in the absorption-free windows).

By properly feeding the antenna array elements, different operation modes can be adaptively generated. In **ultra-massive beamforming**, all the nano-antennas are fed with the same plasmonic signal, as in conventional beamforming. This mode can effectively overcome the very high attenuation at mmWave, the THz-band and optical frequencies and thus enhance the communication distance. Moreover, beamforming has the benefit of avoiding co-channel interference while exploiting the angle diversity by steering the narrow beam dynamically to the targeted angle directions. In **ultra-massive spatial multiplexing (um-SM)**, physically or virtually grouped array elements can be assigned to communicate with an individual user. This mode uses multiple streams on a single carrier to increase the capacity per user and can be most effective when radio links operate in a high SNR regime with a bandwidth limitation. This mode improves the network throughput by means of spatial multiplexing, given that the um-MIMO channel matrix is well-conditioned, or equivalently, provides sufficient diversity and rank. Obviously, any combination in between UM Beamforming and UM Spatial Multiplexing is possible.

In addition, to maximise the utilisation of the mmWave- and THz-channel and enable the targeted Tbps-links, more than one spectral window could be utilised at the same time. In this direction, **multi-band um-MIMO** enables the simultaneous utilisation of different frequency bands by leveraging the electrically tunable frequency response of graphene-based plasmonic nano-antennas. By tuning (virtually) grouped sub-arrays to different frequencies, a single um-MIMO system can simultaneously cover multiple transmission windows. One of the key advantages is that the multi-band approach allows the information to be processed over a much smaller bandwidth, thereby reducing overall design complexity as well as improving spectral flexibility. In this direction, advanced **space-time-frequency coding and modulation techniques** need to be developed for the um-MIMO systems to exploit all of the spatial, temporal and frequency diversities, and hence, promise to yield remarkable performance improvements.

Besides the challenges related to the plasmonic nano-antenna array technology, the realisation of um-MIMO communication requires the development of novel **accurate channel models** able to capture the impact of both plasmonic nano-antenna arrays in transmission and reception, as well as the behaviour of a very large number of parallel THz-waves propagating in space. Existing MIMO or mMIMO channel models for lower frequency bands (e.g. [C6-41][C6-42][C6-43][C6-50][C6-51][C6-52]) cannot be utilised because they do not capture the peculiarities of the THz-band channel, including the frequency-selective absorption loss or the very high reflection loss. Similarly, the few THz mMIMO channel models developed to date [C6-53][C6-54] do not take into account the capabilities of plasmonic nano-antenna arrays, such as the sub-wavelength size and separation, and the opportunities this brings. Therefore, a 3D um-MIMO channel model for ultra-broadband communications is needed.

A new and revolutionizing technique able to improve substantially the performance of wireless communication networks is smartly changing the propagation characteristics of the wireless channel through the use of **intelligent reflecting surfaces** (IRS), which are made of a large number of low cost passive reflecting elements able to independently change the amplitude and/or phase of the incident signal so as to achieve specific propagation effects [C6-55][C6-56]. The IRS can be used to improve the coverage, reduce interference levels, increase system capacity. Additionally, they can be employed to increase physical layer security and even support wireless power transference.

Large intelligent surfaces (LIS) [C6-57] can be an even powerful technique. They are also surfaces covered by a large number of antenna elements but, contrarily to IRS, where the elements are passive, the LIS elements are low complexity antennas connected to a processing unit. These elements are activated according to the user location and its transmission requirements. This allows unprecedented capacity gains [C6-57], as well as accurate positioning [C6-58]. LIS can also be used for wireless transference of power.

Although the LIS is made of antenna elements with very low complexity, its implementation can still involve considerable challenges due to the large number of antennas and the associated circuitry. A natural option to overcome these difficulties is to split the LIS surface in small panels (say with 10x10 antennas), each one able to separate a small number of users (say, 2 or 4 users), and only a fraction of the panels is activated. Radio stripes [C6-59] are another interesting variant of LIS, with the antennas placed over a stripe instead of a surface. As with LIS, radio stripes can enable huge capacity gains, enabling cell-free systems [C6-60], together with accurate positioning, having the additional advantage of a very simple and low cost and implementation.

The communication using LIS and/or radio stripes schemes will require advanced, low complexity techniques for the signal separation, as well as new resource allocation spacial aspects (i.e., which antennas/panels are activated for a given user). Highly efficient, low complexity amplifiers, namely switched amplifiers, will be required, especially for systems operating at higher frequency bands (e.g., mmWaves and THz). To further improve the performance, disruptive techniques that take advantage of hardware imperfections such as nonlinear and/or memory effects can be employed [C6-61][C6-62].

The large dimensions of LIS and radio stripes (several tens of meters), together with the relatively short communication ranges (tens of meters or even less), leads to a near field communication effects, with its inherent potential and challenges. The channel estimation can be a considerable challenge due to the large number of parameters to estimate (this also applies for IRS). To overcome these difficulties, parametric channel estimation and tracking techniques [C6-63] can be employed, eventually supported by positioning information.

Network densification, which increases the number of antennas per site and leads to smaller cells, is one of the solutions offered to achieve high data rates targeted for 5G [C6-60]. The antennas of such massive MIMO systems can be deployed either in a collocated fashion where a large array of antennas is mounted in a single location in a compact way or in a distributed fashion with antennas spread over the covered area. The former approach is known as the centralized mMIMO [C6-66] and the latter the distributed mMIMO [C6-67]. Distributed mMIMO can be implemented with a cell-based approach where the access points (APs) are divided into disjoint clusters and APs of each cluster cooperate to serve the user equipments (UEs) within the cell defined by the cluster. This scheme is called coordinated multi-point (CoMP) with joint transmission in 3GPP LTE [C6-68], but unfortunately it did not provide much practical gains [C6-69].

The cell-centric approach can be changed to the user-centric one where the cluster serving a particular UE can be determined dynamically by choosing the subset of APs closest to the UE. The combination of TDD and mMIMO operations with the dense distributed network topology and the user-centric approach leads to the concept of **cell-free massive MIMO** in which all APs are able to serve UEs cooperatively without any cell restrictions. The cooperation among the APs can be implemented via a fronthaul connection between each AP and CPU and a backhaul connection between CPUs. Compared to its cell-centric counterpart, the cell-free mMIMO is considered as a promising technology [C6-70] due to its improvements in terms of spectral and energy efficiency, as well as its ability to balance the centralized and distributed processing, especially for indoor and hot-spot coverage scenarios [C6-71]. Nevertheless, some crucial questions remain open, such as the relevant initial access, power control, distributed processing considering encoding/decoding, resource allocation, channel modelling, downlink channel estimation, compliance with existing cellular standards and prototype design.

In the different variations of massive and ultra-massive MIMO schemes, **machine learning** (ML) can be used in certain scenarios. One scenario is when the available mathematical model is flawed or is only a coarse approximation of the real underlying model. An example is the linear channel model which neglects non-linearities caused by hardware and other practical effects. ML can be used to improve the solutions that are based on the linear model approximation. ML can also be effective to define detection schemes that take advantage of nonlinear effects [C6-64][C6-65]. A further example is molecular communications discussed in Chapter 11, where an accurate model describing the molecular channel is yet to be developed. Another scenario is where the optimal solutions are computationally demanding and/or not possible for practical hardware architectures. In this context, ML can be used to approximate those optimal solutions with lower complexity, albeit clearly at a performance loss. Examples include maximum likelihood detection, channel estimation, etc.

6.6 Waveform, Multiple Access and Full-Duplex

Cyclic prefix orthogonal frequency division multiplexing (CP-OFDM) has been adopted in several wireline and wireless standards such as ADSL, Wi-Fi, LTE, and recently in 5G NR [C6-72]. CP-OFDM divides the bandwidth into several orthogonal subcarriers. The orthogonality is preserved as long as the transmitters are synchronised to each other. Fine time and frequency synchronisation is then required to maintain the subcarrier orthogonality. However, strict synchronisation is limiting in certain scenarios. For example, sporadic access in internet of things (IoT) and machine-type communications (MTC) requires relaxed synchronisation schemes, in order to limit the length of the signalling overhead [C6-73]. Ideally, the massive number of devices could just transmit their messages asynchronously; being only coarsely synchronised [C6-73]. This could also be advantageous for low-latency communications. However, in multi-user asynchronous access, the CP-OFDM subcarriers are no longer orthogonal, which introduces high inter-carrier interference [C6-74]. Therefore, CP-OFDM is no longer viable in such scenarios.

Several waveforms, e.g. filter bank multi-carrier (FBMC), generalised frequency division multiplexing (GFDM) and universal filtered multi-carrier (UFMC) may be more suitable since their subcarriers are better localised in the frequency domain, and therefore limit the inter-carrier interference. A good frequency localisation may also be beneficial due to other reasons, e.g. sensitivity to phase noise in mmWave, required accuracy of frequency-synchronisation, etc.

The waveforms differ in whether they are orthogonal, whether and how they employ a cyclic prefix, and how the subcarriers are filtered to make them well localised in the frequency domain [C6-75]. FBMC is quasi-orthogonal, performs a per sub-carrier filtering and eliminates the cyclic prefix, but care must be taken in the implementation since contrary to OFDM, GFDM and UFMC, it uses offset quadrature amplitude modulation (OQAM). GFDM also performs per-subcarrier filtering and reduces the overhead of the cyclic prefix by employing it for several symbols, instead of per symbol as in OFDM. However, its non-orthogonality introduces self-interference even if the transmitters are perfectly synchronised. This requires a more complex receiver using e.g. successive interference cancellation. UFMC eliminates the cyclic prefix and applies a filtering for a sub-band consisting of several subcarriers, where the subcarriers within a sub-band are orthogonal to each other but the sub-bands are non-orthogonal, introducing less inter-carrier interference compared to GFDM. Numerous comparisons between those waveforms have been made regarding implementation complexity, spectral efficiency, robustness towards multi-user interference (MUI) and resilience to power amplifier non-linearity etc., see e.g. [C6-76] and [C6-77].

There are further new waveforms, including orthogonal time frequency space (OTFS) modulation [C6-78], which can be considered as a special case of multicarrier code division multiple access (MC-CDMA). It uses long spreading sequences that are well localized in the delay-Doppler domain. This kind of spreading sequences have originally been designed for radar systems. Constant envelope OFDM (CE-OFDM) [C6-80] uses phase modulation to modulate an OFDM signal onto a carrier to reduce the peak to average power ratio (PAPR). A low PAPR is advantageous, as it enables more efficient power amplification, as the lower the PAPR is, the smaller the power backoff can be. CE-OFDM uses Hermitian-symmetric inputs to the IDFT, which leads to a real valued output used to modulate the phase. Low-complexity receivers for CE-OFDM have been studied in [C6-81].

Even if they have not yet been adopted in 3GPP, these post-OFDM waveforms are promising schemes, especially in asynchronous multiple access for massive IoT scenarios. Therefore, application-oriented research on algorithms and proof-of-concept implementations are needed to make them more mature.

Relaxing the orthogonality constraint generally leads to a more efficient and flexible use of the wireless channel. Non-orthogonal multiple access (NOMA) has attracted significant attention in recent years, as it does not only result in larger achievable rates for scheduled uplink and downlink transmissions, but also provide means to cope with packet collisions for MTC scenarios with grant-free access [C6-82]. Challenges for NOMA research include

- *User pairing:* Current research in NOMA mostly considers pairing of two users with the same resource due to high interference accumulation at receiver. However, some publications suggest that with a careful design, more than two users can be paired to use same resource [C6-83]. The application of dense users using same resource can be further extended by considering other candidates such as massive MIMO and mmWave.
- *Interference cancellation:* Several potential candidates for interference cancellation in NOMA have been discussed intensively. Successive interference cancellation (SIC), parallel interference cancellation (PIC) and hybrid interference cancellation are some possibilities [C6-84]. Yet the challenge to find the optimal one is still broadly open. The main focus is to find a balance between error rate performance, number of paired users, each user’s throughput and overall throughput.
- *Power control:* It is important to design an intelligent power control scheme since power is a crucial resource in NOMA. The design of power control in NOMA can affect other performances such as receiver interference level and throughput. E.g. the work in [C6-85], where the power constraint is jointly allocated in full-duplex NOMA, can be further extended to multi-cell scenario.
- *Physical layer security:* One of the earliest questions when NOMA was introduced is: Is NOMA safe? This question arose since in most NOMA cancellation techniques, one user can decode another user’s signal in its own device. Such an issue needs further investigations (see e.g. [C6-86]).

Furthermore, advanced self-interference cancellation techniques can potentially double the spectral efficiency, and enable in-band full-duplex (IBFD) transceivers that offer a wide range of benefits, e.g., for relay, bidirectional communication, cooperative transmission in heterogeneous networks, and cognitive radio applications [C6-87]. However, for the full-duplex technique to be successfully employed in next generation wireless systems, there exist challenges at all layers, ranging from antenna and circuit design (e.g. due to hardware imperfection and nonlinearity, non-ideal frequency response of the circuits, phase noise, etc, especially when taking MIMO and massive MIMO into account), to the development of theoretical foundations for wireless networks with IBFD terminals. Much work remains to be done, and an inter-disciplinary approach will be essential to meet the numerous challenges ahead [C6-88].

6.7 Coding and Modulation

Channel coding aims to correct errors to establish reliable communication and can be regarded as one of the most complex parts of the baseband transmission chain [C6-89]. For decades, researchers sought for channel codes with good error correction performance approaching Shannon's capacity limits with manageable complexity. Modern channel coding schemes such as Turbo, LDPC and Polar codes with excellent performance made their way into several communication standards after advancements in semiconductor technology. However, as the decoders for those codes are very complex, there will be implementation bottlenecks (w.r.t. computational complexity, algorithm parallelisation, chip area, energy efficiency, etc.) to be addressed for high throughput (e.g. when throughput is over multiple Gigabits per second) and/or low latency applications are targeted by future communication standards.

For Polar codes, the state of the art CRC aided successive cancellation list (CA-SCL) decoding doesn't scale up well with throughput due to its serial nature of the algorithm. Hence, iterative algorithms like multi-trellis BP (belief propagation) decoding [C6-90][C6-91] may be considered. Furthermore, modified polar code constructions can be adopted to improve the performance of iterative BP algorithms. Approaches like unfolding the iterative decoders using deep neural networks can be used to improve the latency and throughput of the decoders [C6-92][C6-93].

Even though these modern coding schemes show near-capacity error correction performance for many channels (e.g. binary input additive white Gaussian/BI-AWGN channels), their combination with higher order modulation schemes (such as QAM) can lead to a sub-optimal performance. One reason for this degradation is the so-called 'shaping loss' caused by the probability distribution of the transmitted symbols [C6-94]. In order to approach capacity, the transmitted symbols need to have a certain probability distribution (e.g. discrete Gaussian distribution is needed for the transmission over AWGN channels) and using uniformly distributed symbols results in a performance loss, which can be up to 1.53 dB on AWGN channels.

Several solutions for constellation shaping are proposed to compensate this loss. One option is to optimise the locations of the modulated symbols in the constellation diagram to obtain non-uniform constellations (NUC), as adopted in the ATSC3.0 standard [C6-95]. This scheme is also called geometric shaping and shows improvements compared to uniform signalling. Another approach is the so-called probabilistic shaping [C6-96][C6-97][C6-98], where a shaping encoder is employed to encode messages in a way that the transmitted codewords have a non-uniform probability distribution, resulting in a capacity achieving distribution when combined with simple QAM symbols. This approach is shown to perform close to channel capacity. Another feature of probabilistic shaping is that the probabilities of transmitted symbols can be changed to adapt the transmission rate without changing the FEC code. This is of particular importance since a single FEC code design is sufficient for rate-adaption. Considering the diverse requirements of future communications systems, several shaping encoders suitable for both high throughput and ultra-low latency (short blocks) have been proposed in the literature [C6-97] and [C6-99]. However, hardware implementation of efficient shaping encoders and decoders needs further investigations.

Constellation shaping provides significant improvements in terms of error correction performance. In general, signal shaping is a fundamental and important technology to further improve the spectral efficiency of wireless and wireline communication systems, as the shaping loss may be considered as one of the last gaps between Shannon's information theory and the practical communication systems to be bridged.

6.8 Positioning and Sensing

Especially in the massively connected world of the "Internet of Things" (IoT), it is getting more and more important to be aware of where all these "things" are located, e.g. via positioning and sensing. Mobile radio-based technologies envisaged for a future system will play an essential role in providing high accuracy positioning of the "things". State-of-the-art cellular systems like 4G LTE support positioning in a non-cooperative way, i.e. in the downlink, several base stations send reference signals and the UE measures them and, in the uplink, a UE sends reference signals and several base stations measure them. This is good enough to support the requirements imposed by the FCC for localisation of emergency calls (so-called E-911), where an accuracy in the order of 50 m can be required [C6-100]. There are use cases for future mobile communications, e.g. V2X, smart factory and others, however, where a higher localisation accuracy is required. For instance, for V2X vulnerable road user discovery, an accuracy as high as 10 cm may be required (see e.g. [C6-101]). Currently, 3GPP is considering a positioning accuracy of 50 cm for the future releases of the 5G NR standard, due to the **higher frequencies and large signal bandwidths**, **dense networks** as well as improved **device-to-device** communications available for 5G. In general, wireless sensing can be used to monitor and record the physical conditions of the environment, which can in turn be utilized for more power-efficient communications, such as in communication with integrated sensing. High accuracy sensing without weakening the wireless communication should be an indispensable feature for future networks, including short-range communication.

While it is by now known that MIMO systems improve spatial diversity and result in spatial multiplexing gains, their power in improving positioning accuracy has not yet been fully exploited. Large antenna arrays at the BS (base station) result in very fine angular sampling, which can be leveraged for positioning methods. Further, existing positioning methods only work well in strong LoS environments in general. Many environments, however, experience strong multipath which cause performance degradations and reduces position accuracy. For that reason, the existing methods need to be revised or new methods need to be developed to accommodate multipath propagation. Such methods can additionally leverage the presence of large antenna arrays at the BS [C6-102]. Clearly, having multiple antennas at the UE can improve positioning. In particular, the ability for a receiver to measure the **time-of-arrival, angle-of-arrival, and angle-of-departure** of distinct multipath components improves not only the ability of the UE to exploit the LoS path (including the possibility to determine the UE's orientation), but also its ability to **map the environment**, in order to determine the location and the extent of dominant reflectors. Such **radar-like abilities** can occur in either bistatic operation (piggybacking on standard positioning reference signals) [C6-104], or in monostatic operation (requiring full-duplex processing at the BS) [C6-103]. The price to pay is the complexity cost of the associated simultaneous localization and mapping algorithms [C6-104], which likely need to be solved through mobile edge computing. Moreover, fully harnessing these physical dimensions would require **novel signals in temporal spatial and frequency domain** [C6-105].

Once such radar-like abilities of communications systems are available, a convergence of radar and communications technologies is very likely. However, passive radar technologies depending on RF (communications) transmitters that are not under control of the locating entity cannot be used for critical applications where **service availability and reliability** is crucial. Active radar is necessary. The more active radar systems will be employed, the more interference will be experienced. Radio resource management is one means to cope with the interference. Well-known technologies from cellular communications can be employed using communications links for exchanging such control information. Ultimately, new waveforms can be deployed for combining radar and communication [C6-106][C6-107].

Cooperation can boost the positioning accuracy [C6-108][C6-109], especially in massively connected scenarios. In cooperative positioning, the user equipment (UEs) can send and receive signals and exchange their position relevant information. If the density of UEs is large, it is likely that there are line of sight (LOS) propagation conditions to each UE from several UEs, which is significantly increasing achievable localisation **accuracy and coverage**. There are two different approaches to position calculation, a centralised approach where a central entity calculates the position and a decentralised approach where UEs calculate their position based on the position estimates of the UEs in their vicinity. With side-link communication in 5G, new opportunities for localization and sensing arise, not only in signal design, but also in protocols and algorithms. Important use cases are in the vehicular and drones contexts, where relative location information from cooperative links can have direct implications for safety and global situational awareness.

Accurate positioning can be leveraged to enable **location-aware communications** [C6-110], e.g. design of narrow beams targeted towards the intended user in traditional cellular systems, facilitate autonomous driving, etc. These effects will become increasingly pronounced as communications systems shift to ever higher carrier frequencies (0.1 THz and beyond). Furthermore, accurate positioning is a prerequisite for emerging industrial and factory applications. Therefore, in contrast to legacy systems, positioning has a big impact on the operation of future communication systems. For these reasons, investigating new positioning paradigms, e.g. for joint communication and positioning, is essential, as it can further improve spectral efficiency, energy efficiency, and reduce latency. Similar to other applications such as ultra-massive MIMO, **machine learning** can be used in the case of positioning where complex propagation environments cannot be accurately modelled.

6.9 Massive Random Access

The future vision of IoT envisages a very large number of connected devices, generating and transmitting very sporadic data. The challenge here is how to coordinate such a network without consuming much of the network resources and node energy for protocol overhead. Modern information theoretic research has formalised this problem as follows: consider a number of nodes, each of which makes use exactly of the same code, which is hardwired into the device for system simplicity and cost reasons. These nodes access a common transmission resource at random in a very sporadic manner. The receiver (e.g., a base station) must decode the superposition of codewords without knowing a priori who is transmitting [C6-111]. After decoding the messages (payload), the ID of the transmitter can be found as part of the message, if necessary. For example, in some applications it is important to know the transmitter, but there are applications in which it is important to get the data and not the identity of the transmitter. The challenge now is to design such new random-access codes for which the superposition of up to K distinct codewords can still be uniquely decoded.

In such a framework, the design of channel access protocols departs from conventional approaches used for predictable, persistent, and synchronized data sources. This new random-access paradigm is inherently related to **group testing**: A set of statistical procedures for which it is possible to identify the presence of certain individual agents by sampling combinations thereof [C6-112] and [C6-113]. A related setting consists of coded slotted Aloha, where sparse codes with iterative message passing decoding are developed along multiple random transmissions, to effectively eliminate interference by a sort of low-complexity successive interference cancellation [C6-114]. The performance can be further improved using low-rate channel codes in combination with multi-user detection at the physical layer [C6-115]. While traditional access protocols were designed to avoid interference, the key idea of such innovative approaches lies on the ability to harness information from multi-user interference and constructively utilize it for contention resolution, in combination with advanced signal processing techniques at the receiver [C6-116].

A related problem consists of activity detection, e.g. using a receiver with a large antenna array: In this case, users are given unique signature sequences and transmit at random in a completely uncoordinated way. The base station has multiple antenna observations and must identify the “active set” of users that are transmitting. This problem is related to **compressed sensing** where the sparse vector to be estimated is the vector of 0s and 1s, denoting “absence” or “presence” of the transmitters. Modern techniques based on approximated message passing (AMP) can be used for this purpose [C6-117] and preliminary research results show the exact trade-off between the length of the signature sequences (protocol overhead) and the number of active users, such that the probability of identification error can be made as small as desired [C6-118] and [C6-119]. Compressed sensing-based multi-user detection may also be combined with coded random access schemes [C6-120].

Massive MIMO technology can be efficiently exploited in massive random access to improve the activity detection accuracy by leveraging the high spatial multiplexing gains. The combination of massive MIMO with non-orthogonal multiple-access (NOMA) techniques emerges as a promising area for the design of novel random access protocols. With the aid of multiple-measurement vector compressed sensing techniques [C6-121], the user detection error in grant-free random access can be driven to zero asymptotically in the limit as the number of antennas at the base station goes to infinity. Another approach of jointly addressing the problems of activity detection and collision resolution is the grant-based strongest-user collision resolution protocol, able to resolve collisions in a distributed and scalable manner by exploiting special properties of massive MIMO channels [C6-122].

In both cases the massive random-access and the activity detection problems, a significant research effort must be made in order to bring the abovementioned theoretical ideas to practice and to facilitate a solid system design. Furthermore, even the basic theory needs to be extended, for example, to encompass asynchronism and presence of unknown parameters, such as phase and frequency offsets, and random fading coefficients, for which the current theory has only partial answers.

In a second step, this line of research should consider waveforms adapted for low-latency sporadic access for the cyber-physical systems characteristic of the tactile Internet [C6-123]. Here, sub-ms latencies may be required in order to control moving or even flying objects (passenger drones) or other similar scenarios requiring the combination of ultra-reliable communication with centralised control systems. Similar mechanisms will also be required for evolved Industry 4.0 applications [C6-124]. It is envisaged that the physical-layer transport mechanisms will be associated with real-time cloud computing (mobile edge computing) in proximity to the radio network to implement the necessary control loops. This concerns primarily sub-6GHz access for the uplink and massive connectivity of objects to wireless infrastructure. The objective is to provide solutions for the evolution of cellular IoT uplink waveforms and protocols that scale to huge number of connected devices with stringent energy and potentially latency constraints.

Another promising research direction lies on the use of data-driven methods for the design of new generalized random-access protocols, where the receiver exploits certain side information about the (possibly correlated) activation patterns of the devices. In this context, AI/ML techniques have the potential to build on the availability of data and identify features that could enable the interaction with the underlying random access protocols, e.g., reduce connectivity overhead and prevent the under-utilization of the scarce radio resources [C6-125].

6.10 Wireless Edge Caching

Wireless communication networks have become an essential utility for citizens and businesses. Wireless data traffic is predicted to increase by 2 to 3 orders of magnitude over the next five years [C6-9][C6-126]. The implications of these trends are very significant: while continued evolution is to be expected, the maturity of current technology (e.g., LTE-Advanced for cellular and IEEE 802.11ac for WLAN) indicates that the required orders of magnitude throughput increase cannot be achieved by an incremental “more-of-the-same” approach. As far as wireless capacity is concerned, the forthcoming 5th Generation (5G) of standards and systems is focused to a certain extend on the traditional view of “increasing peak rates” [C6-127]. In contrast, it is widely recognised that a major driver of the wireless data traffic increase is on-demand access to multimedia content (Wireless Internet) [C6-9][C6-126]. Peak rates do not necessarily yield an improved user Quality of Experience (QoE). For example, typical video streaming requires rates ranging from ~400 kbps (standard quality) to ~2 Mbps (high quality). What really matters for the end user QoE is the availability and stability of such rates, so that a video can be played anywhere, at any time, and without interruptions. Also, we observe that the users’ content consumption pattern and the operators' data plans are dramatically mismatched. For instance, a standard monthly data plan in the EU includes ~3 Gbytes of LTE traffic at a cost ranging between 15 and 50 EUR, while a single movie requires ~1.5 Gbytes of data, such that the whole plan would be depleted by streaming ~2 movies.

In light of the above considerations, a novel content-aware approach to wireless network design is needed. Such novel approach should support the paradigmatic shift **“from Gigabits per second to a few Terabytes per month for all”**. More precisely, the special features of on-demand multimedia content can be leveraged in order to deliver a target of ~1 $TB/month of content data to each user in a scalable and cost-effective manner. This target is far more challenging than achieving Gbps peak rates, which have been already demonstrated by various “5G-ready” experimental platforms [C6-128][C6-129].

Meeting this challenge requires a **profound and non-incremental advance** in the information theoretic foundations, in the coding and signal processing algorithms, and in the wireless network architecture design, in order to exploit the potential gain of content-awareness.

Recent research in information theory and wireless communication has shown that content distribution over a wireless network (e.g., on-demand video streaming) can be made much more efficient than current state-of-the-art technology by caching content at the wireless edge [C6-130][C6-131][C6-132][C6-133]. This means pre-storing segments of the content files at the base stations, at dedicated “helper” nodes, and also in the user terminals.

Traditional caching (e.g., prefix caching) decreases the transmission load by the fraction of data already present (pre-cached) at the destination. With these novel modern techniques, based on extensive use of network coding, it is possible to show that a constant (non-vanishing) per-user throughput can be achieved while the number of users grows to infinity. We refer to this behaviour as “full throughput scalability” [C6-134]. For the sake of concreteness, consider the analogy with conventional TV broadcasting: in this case, leveraging the broadcast property of the wireless medium, an infinite number of users can be served with a finite transmission resource, i.e., a finite bandwidth and transmit power. For example, this approach is taken in the so-called enhanced multicast-broadcast multimedia service (eMBMS) in 4G networks. Now, the reason for which eMBMS turned out not to be a huge success is that users do not consume wireless multimedia as they used to consume traditional live TV: they wish “on-demand” services, to access what they want at the desired time and location, and not at the time decided by a TV broadcaster. With on-demand delivery, the broadcast nature of the wireless medium cannot be exploited in a direct and trivial manner. In fact, streaming services today treat the on-demand traffic as unicast individual traffic, as if the content was individual independent data. An important consideration here is security. The data can be stored on user’s local cache that depends on the demand of other users in the network. This leads to the possibility of spying and tampering. Authors in [C6-135] formulate a shared-link caching model with ‘private demands’ with the goal to design a two-phase private caching scheme with minimum load while preserving the privacy of the demands of each user with respect to other users.

Treating on-demand content as unicast traffic is highly inefficient, since it does not exploit the huge redundancy inherently contained in the users’ requests, which concentrate on a relatively small set of very popular files, especially in video-server services where the library of popular movies can be controlled by the service provider, and can be updated at a relatively slow pace (e.g. the library is refreshed every day/week/month). Such redundant requests arrive to the server in an asynchronous way, such that the probability that many users wish to stream the same file at the same time is basically zero. Coded caching techniques have the ability of turning the unicast traffic (on-demand streaming) into a coded multicast traffic, such that again the scalability of broadcasting a common message is recovered and full throughput scalability is achieved.

Beyond these very compelling theoretical results, a significant knowledge gap must be filled to make these ideal of practical value. Therefore, a significant research effort needs be made e.g. in the following areas:

- Coding (e.g., combining edge caching with modern multiuser MIMO physical layer schemes);
- Protocol architectures (e.g., combining edge caching with schemes for video quality adaptation such as Dynamic Adaptive Streaming over HTTP (DASH) [C6-136]);
- AI/ML based content popularity estimation and prediction, to efficiently update the cached content [C6-137].

7. Optical networks

Within the next decade, the world will go digital, improving our quality of life and boosting the industrial productivity. Artificial intelligence will free us up from routine tasks and unleash human creativity and product innovation. We will enter a new era in which billions of things, humans, and connected vehicles, robots and drones will generate Zettabytes of digital information. All this information needs to be transported, stored and processed in an efficient way.

Smart connectivity will be the foundation of this new digital world: Always available, intrinsically secure, and flexibly scaling. A programmable network infrastructure will be the nervous system that the digital society, industry and economy will heavily rely upon. Delivering the required performance, resilience and security levels, while satisfying cost, energy efficiency and technology constraints, presents a formidable research challenge for the next decade.

Overcoming the challenges in scaling electronic interconnect speeds, advanced electro-photonic integration will enable a new generation of optical networking and IT equipment. Combining the advantages of optics and electronics is the way forward to deliver unprecedented functionality, compactness and cost-effectiveness.

Optical networks have long been the solution of choice for submarine, long-haul, and metro applications, thanks to the unparalleled capacity, energy efficiency and reach of optical fibre transmission. In recent years, optical network technologies have conquered inter- and intra-data center networks and have created tremendous growth in this sector.

Building a mobile network will require optical connectivity to each radio antenna and a powerful optical network behind. Therefore, the mobile networking we all expect everywhere and which forms the basis for IoT or other smart networking applications will continue to rely on progress in the optical infrastructure to higher capacity, lower latency, increased programmability, enhanced reconfigurability, increased environmental hardening and significantly reduced power consumption.

Current optical access network solutions will evolve further to also fulfil requirements of future applications demanding ultra-high speed and low latency. While classical point-to-point fibre solutions can already provide high capacity and low latency, these are not scalable to support dense 5G deployments at a low-cost point. New architectures, derived from low-cost Fibre-to-the-Home solutions need to be developed so that the high performance can be delivered at a cost that is compatible with the revenue generated by smaller and smaller cells. Additional benefits such as their reliability and EMI (electro-magnetic interference) immunity make optical network technologies also attractive for applications such as critical infrastructures, factories of the future, private enterprise networks, and vehicular networks.

For all these networks from access through 5G any-haul, to critical infrastructure, factory, and enterprise network deployments, free space optical technologies such as Li-Fi should be investigated as complementary solutions to RF implementations as they promise highest data rates combined with added security gains due to EMI immunity and high directionality.

From ground-breaking discoveries such as new types of optical fibres and EDFAs over products such as WDM systems and 100 Gb/s transponders to global standards such as SDH and OTN, Europe has been at the forefront of optical communications R&D for many years.

Seven out of the top 20 network operators are headquartered in Europe while five out of the 10 largest optical equipment manufacturers have major R&D centres in Europe. By revenue, they represent more than 50% of the global optical equipment market. Two of the largest component manufacturers have operations in Europe and more than a hundred SMEs and universities provide complementary innovation on network, system, or component levels. Optical technologies leverage a telecommunication infrastructure market of 350 Billion EUR and impact more than 700,000 jobs in Europe [C7-1].

Yet, innovation cycles are fast and competition is fierce. New research challenges require a continued effort to defend and strengthen Europe's leading position.

7.1 Sustainable capacity scaling

The question of data traffic growth in optical networks comes up periodically in analyst discussions. A few years ago, it was claimed there was a slowdown in the pace, but these observations failed to notice that there was a massive transfer of data from public internet to the private intranets of cloud providers. Overall, the global data traffic has been doubling every 2-3 years over the past 15 years and it will continue to increase at an impressive rate. Projections of future traffic predict required data rates of 10Tb/s for opto-electronic interfaces and over 1 Pb/s for optical fibre systems by 2024 [2]. Networks also need to provide headroom for unexpected traffic increases, as recently observed in Europe due to a health emergency condition in several EU member states.

The network evolution stumbles upon the most fundamental limits of physics which are: the Moore’s law on Silicon integration and Shannon’s limit on fibre capacity. These limitations are already slowing capacity increase and will become gating items just a few years from now so urgent research efforts are necessary. There is a clear danger that a two-fold increase in the requested capacity will require doubling the amount of optical/electronic hardware. This will increase cost in a linear fashion and threaten future capacity growth. Obviously, disruptive approaches are now needed in optical networks in order to push Shannon’s and Moore’s limits out further.

First, recent successful innovations will be exploited far beyond current status. It can be predicted that optical communications are moving to coherent transmission everywhere. Once viewed as prohibitively expensive, coherent technologies will massively expand from long-haul systems into all fields of optical communications: to support the new generations of wireless systems beyond 5G, to offer enhanced broadband access, to cope with the growth of inter data center communications, to make edge cloud a reality, and even to allow a new breed of intra-data center networks. Coherent is the most promising technology to bridge the gap which is caused by the Shannon limit, leveraging “shaped” modulation formats, flexible rates, and increased density WDM. Symbol rates in excess of 100 GBd, although they are challenging from technology point of view, are crucial to maintain the transportation per bit cost at the necessary levels to stimulate innovation.

Contemporary digital coherent technologies were essentially born in Europe in 2010 and Europe is in a good position if it continues to innovate at fast pace. A mutualisation of research efforts to establish and drive standards development is however required. We argue that a new flagship initiative “Open Coherent Communication Everywhere” tackling reach, capacity and cost considerations across all applications based on a common technological approach would be key in federating Europe's strengths.

To expand network capacity beyond the Shannon's and Moore's limits, given by current fibre and integration technology, we need to exploit all dimensions in space and frequency, opening new optical wavelength bands and space division multiplexing.

The exploitation of new wavelength bands will require advances in a multitude of technologies ranging from optical amplifiers, tailored to these new bands, to a large variety of opto-electronics devices and sub-systems; namely, tuneable lasers, optical multiplexers, couplers, optical mixers, photodiodes, wavelength selective switches and other optical switching solution. System design guidelines will also have to be revised and updated taking into account the new physical impairments which will undoubtedly come up in the new bands. Intensive research efforts are necessary along these lines.

In parallel, space division multiplexing must be investigated. This approach can offer significant capacity increase, either by multiplying fibre count in cables, or by introducing multicore or multimode fibres. Here again, new node and system architectures, new digital signal processing, new space division multiplexers, new switches, new optical amplifiers are needed along with the new fibre types needed. For space division multiplexing to become a cost-effective reality, a change of scale in component count per square millimetre will be required.

Any improvement in fibre bandwidth, attenuation, and nonlinearity directly impacts the capacity of optical fibre systems, and the fibre plant itself is still amenable to such improvements. In particular, hollow-core fibres promise both larger bandwidth (up to several hundreds of nm) and much lower latency (30% lower) compared with standard fibres, at the cost of a much higher attenuation, which limits the transmission distance. The attenuation of those fibres is expected to decrease in the next decade. In parallel, improvement of the attenuation in the C-band is still possible, for instance using pure silica core fibres, showing that innovations are still possible in this field.

Last, capacity can also be gained through margin reduction. Recent publications show that a doubling of network capacity, or even more, is possible through careful margin reduction, which has been rendered possible by the availability of a wealth of monitoring data in new optical networks, as further explained in Section 4.5. However, one should be careful not to jeopardize reliability when reducing margins, as sufficient margin needs to be maintained for reliable network operation.

7.2 New switching paradigms

Future applications, such as autonomous driving, augmented/virtual reality and augmented workspace, will severely change architecture and dynamism in optical networks. Both new network architectures with edge clouds close to the end user and centralized clouds with flexible distribution of network and application functions will be required. Cloudlets at the edges can be viewed as "data centers in a box", that can be flexibly deployed at the network edge to meet the capacity or latency constrains required by the applications.

The optical transport network then can be seen as a programmable network fabric that dynamically provides slices of distributed network, compute and storage resources to applications and tenants. Of course, switching is a necessary ingredient. Switching can be accomplished on layer 0-3 depending on technology availability and service requirements and may cross multiple vendor or operator domains. With the trend towards disaggregated switch platforms, multi-layer switching functions often need to be orchestrated across multiple platforms. Technologies such as software-defined WAN (SD-WAN) move intelligence to the end-points of a connection and can use multiple transport options in-between to optimise cost, performance and survivability.

Flexible coherent pluggable technologies for point-to-point and point-to-multipoint high-speed transmission will enable a tight integration with the packet layer. Novel white box solutions encompassing programmable packet forwarding technologies (e.g., P4, and its expected evolutions) and based on open node/network operating systems are needed for elastic networking, measurability, reliability, in-network operations, and embedded security of the packet-optical networks.

Flexgrid technology on the optical layer allows the introduction of a spectrum-as-a-service model offering the opportunity for a flexible network slicing in the wavelength domain. An operation over multiple wavelength bands and spatial dimensions requires new switch and transponder architectures which have not been discussed in great detail yet. Some applications may require network resources only for a very short time. Consequently, approaches enabling a faster reconfiguration (<1ms) on the optical layer and taking into account concerns such as amplifier power transients need to be developed.

While optical switching in commercial applications has been limited to circuit switching so far, advances in photonics integration could allow optical flow or packet switching approaches to become practical, which were previously considered too costly or complex to implement. This opens up a range of new applications and use cases.

Research is required to investigate the benefits and drawbacks of different switch architectures on the network and application layers, to develop novel switching architectures and new routing protocols within, and to develop new semantic description and information models allowing the control of new devices by an SDN controller platform.

7.3 Deterministic networking

While the Internet mostly relies on a best effort traffic paradigm, the digitalization brings a multitude of applications in which reliability, latency and signal quality need to be guaranteed (e.g. uRLLC = ultra-reliable low latency communication in 5G). Examples range from mobile fronthaul traffic over critical control applications in the vehicular and industrial space to high-resolution machine vision or augmented/virtual reality applications. Some of the most challenging requirements discussed today are <100µs latency (including fibre transmission which adds 5µs/km), <8ns timing error, and several tens of Gb/s throughput.

Networking assuring deterministic end-to-end performance is in the centre of attention that includes, but is not limited to, controlled physical layer performance, guaranteed throughput of high-priority services, and upper bounds in QoS parameters such as latency and jitter. Typically, a precision timing solution is required to provide a time reference to all network nodes with sufficient accuracy. A central traffic management is desired to avoid an overbooking of the network and to be able to provide service-level assurance for the services running over the network. In addition, deterministic networking should also be revisited in the context of network and central office (CO) virtualisation. Research on how to achieve deterministic QoS targets while using function chaining over shared compute and network resources needs to be addressed. This includes study on hybrid use of electronic and optical switching as well as studies on the scalability of guaranteeing deterministic QoS for large number of flows/applications. In such cases, also the physical layer requirements may have to be adapted, creating the need for the development of ultra-high speed, low latency, and low jitter physical layer implementations. This is especially challenging in shared bandwidth applications, such as access networks, where changes to scheduling, signalling, and synchronization may be required to fulfil the QoS requirements. A redundancy concept should ensure continued operation in case of an equipment or link failure. Some of the applications will have to run over a public network. In other scenarios private network builds are also possible or sometimes even required.

Multiple options and technologies are being debated in standards bodies such as ITU-T, IEEE, and IETF and range from OTN over FlexE to time-sensitive Ethernet or IP approaches. However, the picture is not clear yet and the achievable performance of different methods need further investigation. Further research is required to determine the optimal solution set for a diverse range of applications, to develop the necessary planning and provisioning tools as well as the means for service assurance and performance verification.

7.4 Optical wireless integration

5G is deeply transforming the underlying transport network, due to several concurrent causes: the end-user capacity is increasing, the coverage of the mobile network becomes more dense, and a split architecture in the radio access network is introduced, where different functional splits between baseband and radio units are supported by the same transport network. The latter means that the traditional distinction between fronthaul and backhaul networks is blurred. Since it is impossible, for cost and operational reasons, to design dedicated networks for the considerable number of heterogeneous last-drop technologies, the adoption of a shared network infrastructure that makes use of common transmission and switching platforms is unavoidable. This is not a trivial task and will be the big network design challenge for several years from now. The ambition is to open new opportunities by comprehensively applying fibre technology to various scenarios, in a sense making the paradigm of Fibre to the Home become a Fibre to Everything. [C7-4].

In principle, several candidate technologies exist for enabling the coexistence of fronthaul and backhaul networks. In practice, all require a redesign and a redefinition of their application space. For example, packet switching with new packet friendly fronthaul interfaces is likely to be implemented in scenarios where many users generate a low amount of traffic data each. However, the need to meet tight latency constraints (which can be as low as of 100 µs) and deterministic delays across several packet switches may require the development of new framing, multiplexing and synchronization techniques that can guarantee the requirements of time-sensitive networking without losing the advantages of statistical multiplexing. The convergence of radio and fixed access, the necessity to support dedicated enterprise connections, and the transport of high bit-rate fronthaul signals corresponding to the low layer split options defined by the 3GPP will foster the adoption of interchangeable multi-layer switching platforms supporting various switching granularities that range from packet to time-slot and wavelength channel level. This will lead to a further level of convergence, where a layered control plane offers the ability to set up services in short time to operators, while being unaware and decoupled from the underlying transmission and switching technology. However, the current transport networks based on OTN and DWDM are, in some respects, inadequate to support the requirements posed by the new generation mobile systems.

In the emerging new “mobile transport network”, where it is of particular concern to lower costs as well as to mitigate the jitter effects caused by complex justification mechanisms, the adoption of a single time-division multiplexing hierarchy across all network segments (access, aggregation and the core) would be highly desirable. Moreover, 25 and 50 Gb/s channels, corresponding to low layer radio functional split interfaces, should be multiplexed and transported in an efficient (i.e. with minimal overhead) and in a scalable way. At a physical layer, DWDM already has the aggregate capacity capable to support 5G broadband services in densely populated areas, scaling up to several hundreds of Gb/s/km^2. However, breakthroughs in technology are still necessary to bring down the total cost of ownership, something that is seen as the necessary condition to justify the large-scale deployment of DWDM in x-haul (i.e. the network segment where mobile front- and backhaul as well as fixed-access transport converge). Examples of enabling technologies here are: cost-effective 50Gb/s (and beyond), optical interfaces with direct detection receivers capable of reaching distances up to 40km in the 1550nm window, possibly employing new modulation formats; tunable photonic integrated devices for compensating the chromatic dispersion; "lite" coherent transceivers with low DSP complexity as well as new technologies for developing cost-effective and compact optical amplifiers. A new generation of coherent transceivers and optical amplifiers would also allow to implement fronthaul networks exploiting broadcast-and-select DWDM Passive Optical Networks (PONs) architectures. Coherent technologies are key to compensate the attenuation of passive splitters in the optical distribution node, which is a key issue for bit-rates 25Gb/s and above. This scenario is of considerable importance as it enables the convergence of fixed access and fronthaul on PON fibre infrastructure which is capillary deployed, at least in big cities. For optical-wireless integration, however, additional modifications and developments on the physical layer in the PON might be necessary to achieve the high speed, low latency, and low jitter required for radio applications.

Another technological breakthrough would be the design of “fully colorless” DWDM networks, namely networks based on port-agnostic devices which offer the advantage to greatly simplify network provisioning, with lower installation and operational costs. Fully colorless networks require both tunable transmitters and reconfigurable optical add-drop multiplexers (ROADMs). Although ROADMs are widely used in wide area DWDM networks, current Micro Electro-Mechanical Systems (MEMS) or Liquid Crystal on Silicon (LCoS) technologies can hardly scale down their cost enough to be applicable to an access network. Novel system-on-chip devices based on silicon photonics could potentially reduce the cost by two orders of magnitude. This would allow ROADM technology to spread beyond core and metro networks, also penetrating highly cost-sensitive markets like access, data center, and campus networks. Here, also power consumption will play an important role, requiring increased research efforts.

All aforementioned transport technologies deal with digital radio split interfaces. Analogue Radio over Fiber (A-RoF) is a well-known alternative technology for the distribution of wireless signals: ideally, an A-RoF system acts as a mere medium converter, creating in optical fibre an exact copy of the radio signal on air, without further processing, with obvious benefits in terms of hardware complexity and power consumption. In practice, many performance issues remain to be solved before A-RoF can be deployed to the variety of scenarios where digital units are used today. Examples of challenges to be addressed by A-RoF systems are: noise mitigation techniques that compensate for the absence of equalization and forward error correction mechanisms implemented in digital systems; and linear modulation and photodetection devices to decrease the effects of high peak-to-average power ratios and inter-modulation in multi-carrier wireless systems. Those two aspects are especially critical in 5G, due to the high order modulation formats (256-QAM) and the high number of subcarriers (2048 OFDM) used in wireless systems. Furthermore, the introduction of millimetre waves necessitates the development of linear devices with high signal bandwidth, up to 100 GHz.

Centralized architectures, where the baseband processing unit is shared by a certain number of remotely placed radio units, meet operators' plans for a reduction of the number of network nodes and consequent saving of operational costs, but pose further challenges due to the increased link budget and the accumulated chromatic dispersion caused by the higher distance between antenna unit and central office.

Finally, it is important to define an end-to-end control system, able to manage the interface with the packet switched core network and to monitor the quality of service for the whole link, encompassing A-RoF and packet switches. For example, one of the issues brought about by massive densification of cells is the high traffic fluctuation they experience due to the lower number of users served. As new fronthaul technologies capable of adapting the transport data rate to the variable cell traffic(e.g., through functional split or variable-rate fronthaul) become available, the control plane needs to coordinate statistical multiplexing of the allocated resources across the wireless, access transport and central office domains, merging priority and best effort applications over the same shared infrastructure.

7.5 Optical network automation

Technological advances in optical networks, including an increased programmability and remote configurability at the device level, also require advances in network control, automation, and autonomicity. Software-defined optical networks (SDON) based on logically centralized control and management provide the basic capabilities for new operational paradigms. A centralized controller supports flexible multi-layer, multi-domain optical networks with tuneable wavelengths and variable modulation schemes. It also simplifies multi-vendor integration. Yet, autonomic transmission and networking area still in its infancy and need further investigation.

SDN principles such as centralized deployments, unified data modeling frameworks, and open Application Programming Interfaces (APIs) allow an easier integration with Operation/Business Support Systems (OSS/BSS). They also facilitate an improved connection to billing systems and flexible pricing schemes offering new revenue streams for network operators. Smart contracts may then pave the way to dynamic, programmable end-to-end connectivity on an international scale.

The industry is slowly converging on a common approach to device information and data modelling. First initiatives mainly cover low-level optical systems and devices such as transceivers, open line systems and ROADMs. They aim at vendor-independent models easing network control particularly in so-called disaggregated deployments, where individual network components can stem from different vendors. Activities should cover other aspects related to network operation as well, topology and inventory management or service description for example. This requires a gap analysis of current modeling languages, an identification of shortcomings and limitations, as well as extensions where necessary. Any research and development should be carried out with subsequent standardization in mind, avoiding the complexity of having multiple, often overlapping and incompatible models, frameworks and languages. It also involves finding a balance between de facto and de jure standards, combining the outcomes of Standards Developing Organizations (SDOs) and Open Source projects (OSPs). SDOs follow a pre-defined multi-year workplan, whereas OSPs typically adopt an agile software development approach that is based on frequent releases and a continuous integration/continuous deployment (CI/CD) model.

Control and management solutions have often been monolithic, vendor-dependent and strongly coupled to the underlying infrastructure. Evolving operator requirements, market pressure, and the general trend to softwarization drive disaggregation also at the software level. Monolithic software is being replaced by a micro-services architecture, in which service chains can be composed of reusable functional elements with standardized interfaces and allocated depending on the actual needs.

Smart connectivity in the context of NGI requires additional solutions and innovations for optical network automation beyond simple programmability. Global reach and optimized local service delivery capabilities need to be combined in highly flexible and granular ways and should be available on-demand for integration with web-based software and IoT platforms.

Intent-based zero-touch provisioning of network services is a precondition for smart connectivity. It is also an enabler for increased network resilience and higher levels of cybersecurity. Intelligence in optical network nodes will *i*) facilitate a faster detection of degradations and security issues; *ii*) create a hierarchal incident management system that is far more robust that its centralized counterpart; and *iii*) foster autonomous network control based on advanced Artificial intelligence (AI) and Machine Learning (ML) techniques. All this together can potentially bring network resilience to unprecedented levels.

The advances in optical transmission and switching technology create new challenges for network automation and autonomicity. Spatial division multiplex (SDM) introduces an additional dimension of flexibility in routing and spectrum allocation. Margin-optimized transmission to maximize the fibre capacity requires dynamic re-routing and network re-optimization in case of performance degradation, e.g. due to aging effects or additional loss after fibre repairs.

The complexity of the underlying optical technology, resulting in a large number of interdependent configuration parameters, requires cognitive networks powered by streaming telemetry, real-time network measurements, AI and ML. Optical devices can potentially produce a huge amount of operational and monitoring data. For example, coherent transponders inherently provide monitoring and sensing capabilities. Operators are eager to process and use such data in search for efficiency gains in network operation and automation. The realization requires flexible architectures and protocols for streaming telemetry, as it is commonly accepted that current methods are limited in encoding efficiency, overhead and sampling rate. Consensus needs to be reached on which parameters are key indicators, so that common procedures and algorithms can be applied in heterogeneous scenarios. It is also worth mentioning that telemetry does not only rely on retrieving data from one (or multiple) monitoring points, but also involves measurements at network level, potentially requiring active network probes and joint processing of data measured at different places and times (spatial and temporal diversity). The design of such probes can also be a challenge, since it is well known that it must not be service affecting or consuming an excessive amount of resources. As mentioned in Section 4.1, monitoring information can be leveraged to reduce network margins, especially during operation, through continuous re-optimization. However, just as active probes should not disturb existing services, margin reduction and continuous network re-optimization should not jeopardize network stability. Continuous monitoring may help here.

Coherent transponders provide inherent monitoring and sensing capabilities on a per-connection level. Since there are multiple sources of transmission impairments in an optical network, there is a need to advance the development of cost-effective optical performance monitoring probes that i) can be widely deployed; ii) are agnostic to optical signal waveforms; iii) provide suitable figures of merit (e.g. OSNR, power density, spectral occupancy); and iv) are non-intrusive. Different transmission parameters can then be monitored in real-time and combined to assess the performance of each signal/data stream and its evolution along a given optical network path.

Streaming telemetry and continuous performance monitoring provide the data for AI and ML. Using model-based approaches and statistical analytics, SDN controller can observe the network behavior, analyze the performance, detect anomalies, and autonomously act and re-configure and optimize the network. However, the simplistic approach of “throwing” some machine learning algorithms on a large set of data will not achieve this goal. To fully utilize the expertise from big data analytics generally available, dedicated research activities for data curation and analytics, machine learning and artificial intelligence in optical networks are required.

Given the support that the optical infrastructure provides to 5G-based services, optical performance monitoring and analytics based on such data are critical to understand the base-line performance of such upper layer services.

7.6 Security for mission critical services

The ever-increasing interconnectedness not only of people but also of devices starting from huge power plants down to billions of IoT devices like sensors or appliances does not only increase the dependence on the network infrastructure but also expand the threat surface and therefore the vulnerability of every individual and of the society as a whole. Important threats do not only include hacking and espionage, but also network outages due to natural catastrophes as well as terrorism and sabotage targeting critical infrastructure. Therefore, it is getting more important to better safeguard our network infrastructure against data leakage and unexpected service outages.

A signal on an optical fibre can be tapped once the physical access to the fibre is available. At this point, the data of millions of users and billions of applications is exposed to theft and manipulation. Therefore, authenticity, privacy and data integrity is essential and needs to be kept at a level playing field with increasing threat scenarios, e.g. by allowing for crypto-agility. Improvements need to consider quantum-safe solutions for authenticating the communication partners, for protecting the data against tampering and for exchanging secret keys by employing post-quantum cryptography or secure quantum communication, e.g. quantum key distribution. Also, novel research directions like physical layer security for optical networks should be explored.

Adding redundancy is the conventional, but also expensive way to improve the reliability and resilience of networks. Solutions for low cost and low power implementation of redundancy solutions (e.g. high radix optical switches) should be studied. Alternative concepts, that are high on the research agenda today, are increased flexibility, massive monitoring and software control of optical networks. Especially the monitoring of optical distribution networks like passive optical networks should be improved, which is especially difficult due to their passive implementation. Based on the data generated by the monitoring solutions, data processing (e.g. by means of ML methods) can help to detect upcoming problems early and counteract them in advance with the available flexibility. It should be possible to employ this functionality beyond the borders of a single networking domain.

The higher flexibility of optical networks, enabled through software controlled network elements (software defined networking, SDN), also increases the vulnerability of such networks to various kind of attacks and therefore security and resilience aspects need to be part of the concepts from the beginning (including both the hardware and software layers of the network). More generally, the design of network equipment needs to employ modern security and reliability paradigms (security by design) and apply modern software technology to foster efficient and secure implementation of increasingly complex network elements.

7.7 Ultra-high energy efficiency

With data center traffic consuming nearly 2% of all electricity used today and the share of communications technology in overall world energy consumption steadily growing over the last decade, there is an urgent need for a paradigm shift to greener ICT technology.

Increasing use of optical technologies within the IT and communications industries is one key opportunity to limit the increasing energy consumption against the massive growth of overall data capacity that networks and datacenters are handling. Since light can travel vast distances through fibres, fibre optics consumes only a fraction of the energy used by conventional technology that transports electrons via copper wires.

This means that higher reach and higher capacity interfaces can significantly reduce the energy consumption of networks as they reduce the number of optical interfaces and regenerations needed.

Also, the inherently low power consumption of optics can be directly used to reduce power consumption overall, if optical functions replace more power-hungry electronics, e.g.:

- Functions may be turned off or switched into a low power mode if not in use
- Electronic processing can be bypassed more frequently through new control mechanisms that optimize traffic flows across network layers, particularly if combined with optical space and wavelength switching.
- Optical switching, in particular circuit switching, can replace and/or complement electronic switch functions and save power by eliminating O/E/O conversion points, provided enough link budget is available.
- Electronic interfaces of modern communications ICs are a strong driver of power consumption. Ways to replace those interfaces by lower power optical interfaces that are integrated or co-packaged with those ICs can reduce power consumption substantially.
- The further drive to higher capacities and interface speeds per electronic chip will make optical communication within chips as well as chip-to-chip a future necessity.

7.8 Optical integration 2.0

Two developments require progress in the field of optical and electronic integration: Advances in electronic integration follow Moore’s law and, on one side, lead to increased throughput requirements of electronic ICs on or between printed circuit boards (PCBs). Packaging and I/O limitations will require a transition from electronic to optical chip interconnects when further scaling up electronic processing capabilities. Silicon-compatible, compact and low power datacom transceivers are required to facilitate an integration into next-generation multi-chip switch and processor modules. On the other side, increased data traffic in optical transport networks will require more and more high-speed optical interfaces, when exploiting higher network capacities enabled by wavelength and spatial multiplexing. To avoid scaling of cost and power consumption with the exponentially increasing data traffic, the development of standardized components for future spectral and spatial unit cells are required [C7-2].

Both use cases point to Silicon Photonics (SiPh) as base platform, enabling to adopt Silicon mass-manufacturing processes for optical applications. Also, standard electronic packaging technologies can be leveraged, eliminating the need for expensive gold boxes and special assembly processes. Known as a low-cost platform for non-hermetic and high-temperature operation, Silicon provides good passive optical properties for routing, modulation and detection of light. It is also a natural fit for integrating RF electronics into this platform to drive and controlling the optical SiPh functions. It needs to be mentioned, however, that while Moore’s scaling of electronic memory and processors yields ever smaller structures in Silicon, this miniaturization is not feasible for optical components, where the telecommunication wavelengths on the order of a micro meter pose a limit on the structure sizes. Further integration of photonic and digital processing functions will require scale adaptation. What is more, the level of photonic integration needs to be scaled up substantially to meet the demand of next-generation optical networking equipment.

Work is also required to enable active functionality, like lasing or amplification, into SiPh platforms. Initial steps have been taken to integrate and structure III-V materials (e.g. Indium-Phosphide) into Silicon substrate as well as other passive materials (e.g. Silicon Nitride). Yet, it is crucial that such integration can be accomplished at wafer scale and without sacrificing the benefits of the silicon platform which are coolerless, non-hermetic and high temperature operation.

Moving to higher channel bandwidths, further performance gains are necessary and can be achieved by adding organic materials into the Silicon platform, providing potentially very high optical coefficients and reducing the required driving power [C7-3]. This will ultimately be beneficial for the integration of optical functionalities on every size scale.

Advances in photonic integration will pave the way for a raft of new IT and networking devices in which optical, RF and digital electronic functions can be combined, e.g. in multi-chip modules (MCM) comprising highly integrated CMOS dies and high-speed optical engine chiplets on the same package substrate. Appropriate package design will be required to allow a reflow soldering of these components under standard process conditions.

8. Network and Service Security

The Digital Infrastructure, where Smart Networks and Services plays a fundamental role, is one of the key areas mandating Cybersecurity. Each and every citizen, enterprise or governmental application is requiring protection. The critical assets and the experience of the users depends on the capabilities of the infrastructure but will be subject to global services. Far from being homogeneous the expected grades of security and privacy will be as diverse as one can find in terms of more traditional Quality of Service. But while attacks and risks are becoming numerous, more sophisticated, the trend will remain to require more and higher grades of security.

As a result, the future of ICT development in Europe is dependent on our capabilities to ensure trust and provide appropriate cybersecurity solutions. This is a major challenge, particularly in Europe, where the aim at being forerunners on IPR and privacy protection should push us to take the opportunity to reach global leadership in securing the digitalisation of our society. High expectations in security and privacy should be taken as enablement for added value in IoT, networks, Cloud and services segments.

The Network and Information System Security benefits from a long European story [C8-1] [C8-2]. Building on strong cybersecurity and Networks research capabilities, focus should be given to innovation, standards, certification in order to reach full matching between services expected and appropriate security and privacy levels. Following early initiatives on 5G deployment and its Cybersecurity [C8-3], and considering among others the need to serve up to highly critical services, the infrastructure protection capability should benefit from available open solutions but should also take care of maintaining autonomy and sovereignty based at least partially on an EU-based supply chain.

8.1 Rationales for Security Transformation

Targeting the objectives to fulfil expectations in terms of security and privacy mandates to recognize the game changing factors that require transformation of the security itself. This starts with the deep architectural, behavioural, semantic nature of the systems and services drawing a new picture of risks, then means to mitigate them. Second, whatever disruptive technology or architecture one wants to introduce, it comes with faces, on the one hand securing this new technology but potentially use it for the sake of better security. Digitalization means new usages, new and higher expectations, so that security considerations and developments need to reach up to mission critical supports, it thus conducts to significant impact on the infrastructure. Last but not least, the change is also impacting the stakeholders and their business models. Security as managed services delivered by pure players, role of ethical hackers, move of liabilities perimeters among service providers and tenants, the way to ensure security is also moving…

8.1.1 Change in system nature

8.1.1.1 Security must evolve as per the system evolution

As a matter of fact, cybersecurity solutions are depending on the nature of the target system. As an immediate consequence, the cybersecurity which is intrinsically is intrusive, embedded must evolve as the system does.

New dynamics

The current and expected network models' evolution is thus a key rational for Security transformation. From System or even System of Systems point of view, we have simultaneous disruptive changes both in time and space leading to so-called "metamorphic" properties. One can easily conclude that applying ole paradigms such as static security solutions to a problem where dynamics are becoming predominant will fail. Moving from servers to services and rich interactions between micro-services is leading to new time distribution in provisioning. From data plane towards service (and their usages) plane, security functions deployment must adapt to the temporal properties.

New distribution

Historical basic security paradigm, which was mainly a perimetric approach is becoming inefficient with the dispersed capabilities across a hybrid computing/network model. By nature, the security policy is global and related to the expectations and needs of the users and applications. It must be now derived into more atomic but interacting sticky policies attached to the data, functions, services and composed in time and space.

Matching policy diversity and distribution/composition

Another major area advocating for security transformation is linked to old issues of distributed systems. The overall architecture will rely on the composition of modular systems and services, potentially mission-oriented. As a direct consequence, the scheduling of resources must be mission-aware but also security-policy aware. Security here is raising numerous challenges, the first set is related to policies themselves: how to declare those policies, how to compute compliant resources and services with the policy constraint, the second part is related to the security enforcement deployment: how to schedule (and guarantee) security functions and associated resources.

Multi-tenancy over multi-service architecture: As diversity of requirements is implying differentiated solutions, latency, throughput and other traditional QoS metrics, security mandates also to step away from "one size fits all" solution. Security is not anymore limited to a de facto property of a perimetric system but provisioned service attributes for each tenant over an architecture allowing a range of security levels and services.

World is data centric: The system evolution, its smart control are driven through Artificial Intelligence capabilities, in turn depending on data quality and availability. Either for Users' data or System data, cybersecurity becomes a major concern so that data protection, data sharing in confidentiality, data integrity become mandatory. Data centric security technologies such as Homomorphic encryption or Multi-Party Computation may be introduced for the sake of system protection and protection of its users.

Data is key but darker:

As the sustainability of ICT mandates new technology introduction, it will translate in focused security challenges that should be considered from start. Considering addressing, routing area, although combined with ciphering spread everywhere, a major issue is the ability to authenticate packets (fields), flows, applications, services, users, etc. as it should be. Data and more generally information specifically used for system operation and awareness are becoming less accessible as the one of the first protections is to encrypt it for protection purpose. This is happening, among other, to protect for instance slice integrity and isolation across multi-owned infrastructure segments. This is even more sensitive looking at control, management, service planes exchanges, leading to a potential specific dedicated approach such as a super secure slice. Finally, this should be most of the time a default situation aiming at protecting privacy and confidentiality, eventually conforming specific regulation. Smart solutions, AI-based, depends critically to the access of data set and information extracted from the infrastructure and flows. Encrypted data may lead to more sophisticated application of AI, rather based on behavioural analysis, inference of states, …than assuming free access to explicit information.

Unprecedented complexity:

Our society is more and more depending on the communications and ICT infrastructure. As a result of the ever-growing demand from people, but also Business-to-Business or Machine-to-Machine, the complexity/manageability issues are becoming predominant. Moreover, introducing any innovative or disruptive technologies, potentially not fully known in terms of risks. Cybersecurity often conservative and requiring a significant level of situation awareness is very sensitive to those complexity and novelty dimensions.

New security paradigms applied:

By analogy with SDN, an obvious fundamental dimension of the security transformation may be described as Security Orchestration (aka Software Defined Security). This means that flexibility, spatial/temporal distribution of functions, adaptive capabilities, models and abstractions, etc. apply to the security domain. Among the ICT challenges, security is a fundamental cornerstone, but security must in turn transform towards renewed paradigms. Complexity is expected to be much higher than in the past and control/management technologies should remain in a converged framework, optimising re-use of reduce skills set, complexity and OPEX.

The expected concomitant evolution disrupts traditional approaches where security concerns often came afterwards (or too late), it even goes beyond the" by-design" paradigm with functions and services integrated as intrinsic component of the concept.

Fundamental science advances whether physical or logical will equally impact traditional network functions and security functions. Security will then have in addition to provide solutions to secure the functions themselves. Taking the example of quantum technologies (true Quantum infrastructure not limited to QKD), the security architecture, process, workflows, protocols may be subject to entirely new paradigms taking benefits of the fundamentals of the physics and properties of the Qubits.

Deception and Moving Target Defense (MTD) are also promising candidates and hot spots to influence dramatically the way security is delivered and operated. One can draw a direct link with system flexibility capabilities as this can be envisaged only thanks to the system evolution.

While the need of transformation is considered, it comes with a major challenge of definition of integrated renewed security framework, processes, engineering able to provide relevant answers to the ongoing disruption of architectures and technologies.

Last but not least, security will remain a competition between attack and defence. In the end, the next expected result should be that networks are easier and cheaper to operate for service providers. Digital identities will become a much more widespread concept, and will be able to represent multiple entities (e.g. a citizen, or a customer). These digital Identities must be verifiable, must have a legal meaning while respecting privacy, must ensure ethical usage of itself, should be non-traceable and non-trackable for security or privacy purposes such as hackers, adversaries, or unwanted advertisements. Regular users should get access transparently and their services must be available with appropriate quality. Attackers should have a much higher cost of attack as well as a much higher probability of being caught. Deterrence is thus becoming an entire part of the security strategy.

8.1.1.2 New risks, new attack surface

The purpose here is not to make a full risk assessment of all system evolution, but to point out two essential properties of the future systems. The first one is linked to the IoT which raise issues relative to the scalability, security elements distribution changing by far the attack surface. The second one is linked to the Software Life Cycle application in critical infrastructure which is also feeding the attack graph and potential attacks.

Internet of Things and Industrial Internet of Things (IoT&IIoT)

The nature of the termination points of the system is dramatically changing considering the IoT. This lead at least to two major evolutions of the system impacting the attack surface thus the security. The first one is the scalability just by the order of magnitude of devices, data, usage (including sharing) required, and the second is the intrinsic security capabilities of the objects that are expected to remain quite poor among other reasons for cost constraints considerations. IoT, in particular IIoT is thus coming with high grade requirements but a new attack surface leading to re-think the way risks are mitigated. This may encompass the balance between Hardware and Software, the distribution of the root of trust and numerous control and management issues starting with enrolment of the objects in the system.

Software life cycle

Another fundamental change is that the system is becoming software predominant. What the system does, what the service delivers is depending on the code. Beyond safe code, which is an entire cybersecurity area, the whole Software life cycle is driving the security conditions. Updates, upgrades, exchanges, workflow, privilege management...the path towards system control is much more at risk than previously with quasi-unique path from OSS to EMS. Looking at the emergence of 5G, first security analysis such as the one done by ENISA [C8-4] shown the tremendous importance of the software and its life cycle. Next generation and evolution are clearly re-enforcing the trend and the needs. Software Life Cycle is a security priority following system evolution.

8.1.2 Disruptive Technologies integration

Security transformation is a mandatory companion to most of (if not all) disruption dimensions expected during the next period. This goes into two directions. On the one hand, disruptive technologies must be secured, as on the other hand technologies may be used for cybersecurity purpose or to reach consistency with systems properties.

8.1.2.1 Virtualization

From security point of view, software-based and virtualized functions have been considered weaker than those based on hardware. Nevertheless, it would make no sense to block all the system flexibility without mitigating the issues raised by virtualization. Security is thus purpose of evolutions in virtualization areas:

- secured “hyperware”: Operating Systems, binding with Hardware platforms, Hypervisors and containers
- virtualization of the security functions for various platforms from objects, terminals towards cloud servers. This includes a large diversity of functions from classical firewalls, towards Early Detection & Response systems (EDR) with potentially sophisticated smart protection and detection inside. Nevertheless, some hardware-based Trusted Elements may still be necessary to manage the root of trust and the secret elements through the systems.
- The virtualization belonging to the SNS vision, including various flavours of slicing is raising multiple issues in the security axis: Confidentiality, Integrity and Availability. Identity and rights management on virtual perimeters, Isolation (data flows, activity confidentiality but also resource sharing), control and management workflows, horizontal and vertical interactions, liabilities,…

8.1.2.2 Softwarization

Falling into this type of disruption, one can consider orchestration and chaining at various levels (functions, services). Orchestration is intrinsically linked with dynamic composition of (micro)services which raise the need among others of:

- smart algorithms to compute policy-based compliant and optimized deployments of protection and detection functions. It should be noticed that this is adding even more complexity to already difficult multi-criteria QoS issues.
- potential disruptive security paradigm taking benefits of flexibility such as Deception or MTD.
- Verification/continuous assessment of security conditions and conformity.
- Secured workflows and authority as well as Software life cycle management.

8.1.2.3 Cloudification

In order to reach the required scalability of security applications and align to the best Threat Intelligence updates (vulnerabilities, attacks) security should be as much as possible following the cloud “as a service”. This may still be considered as disruptive for some security functions which are intrusive in the systems.

8.1.2.4 Data-centric

As the entire Digital world, SNS will have massive usage of data and information retrieved from data. From its own data (monitoring, logs,…) to payload from data flows, usage and expectations are set very high but threats are almost proportional. Hence disruptive technologies preserving confidentiality and integrity while allowing some level of sharing are required, for example:

- Full Homomorphic encryption
- Multiparty computation
- Zero Knowledge Proof
- Control of fake/biased data, counter good/bad AI
- Anonymisation/pseudonymisation

8.1.2.5 Quantum related technologies

When it comes to security specifics, the first set of question will be linked to encryption, then on sensing information.

- Post quantum encryption
- Quantum Key Distribution (QKD)
- True Quantum Communication Infrastructure (including Quantum crypto)
- Quantum Sensor/communication interfacing/fusion which may raise or on the contrary solve integrity issues. It should be noticed that in case of proliferation of (quantum) sensors and SNS providing the required hyper-connectivity, a large amount of private information will be potentially used for malicious purpose.

8.1.2.6 Distributed Ledgers Technologies (DLT)

Made famous with the blockchain and crypto-currencies, DLT have particularly interesting properties for distributed systems without “a priori trust”. When DLT will bring value to SNS, most of the time, it will be by its security capabilities. Secured exchanges were following in the past the dominant exchange pattern (client/server), DLT has potentials to participate to security transformation with better fit, in particular with the (re-) distribution at the edge.

8.1.2.7 Artificial Intelligence (AI)

The age of Artificial Intelligence is impacting security of digital infrastructure as plenty of other sectors. From rule-based, statistical learning towards Contextual reasoning, AI is the purpose of many applications in security such as attack detection (pattern matching or anomaly detection) or mitigation of wireless jamming attacks. Nevertheless, the dependency on trustable, unbiased training material or Explainable AI issues should be solved before allowing AI to control autonomous systems which may cause dramatic damages.

8.1.3 Change in Security grade expectation

Time is over for solutions not considering security & resilience. Application to critical sectors, strategic economy, risk awareness, autonomy and capabilities to respond to a large diversity of attacks and attackers require both security upgrade (up to high grade) and adaptation to heterogeneous needs and expectations.

8.1.3.1 Security grade awareness

During the first ages of ICT, security was aften forgotten or considered costly and painful. Nowadays, those who were agnostic are much more aware of the risks and the build of trust requires simple and understandable exposure of grade of security associated to systems and services. Certification is a cornerstone in this landscape and have been already properly identified for the current development and usage of 5G [C8-4]. Beyond this initial effort, the road towards full trust is requiring research and innovation in process, methodologies, and technologies in order to determine security attributes for complex systems and services.

8.1.3.2 Service differentiation

As for QoS metrics or other attributes of the services, security and resilience profiles will have to match a wide diversity of needs. From basic protection to high grade requirements, SNS will need to adapt and satisfy very different demands. Not all usages mandates 1+1 protection or Traffic Flow Confidentiality (in the sense of making activity detection impossible by means of constant traffic), beside prohibitive costs there is no need for that. One size fits all cannot answer the question. Thus, defining concepts of Quality of Security (QoSec) and/or Security Service Level Agreement (SSLA) is a direction to be developed for the sake of matching diversity of demands and usages.

8.1.3.3 Autonomy & Sovereignty

Sustainability of many sectors, including critical ones shall as much as possible avoid dependency on any third-party able to prevent security measures application. In addition to certification programs and from security research point of view, it drives into diverse topics including, safe code, supply chain traceability, secured control & management planes, forensic, Advanced Persistent Treat detections, Cyber Threat Intelligence sharing, etc.

8.1.3.4 Regulations

SNS architecture & technologies, as began with 5G will be at the crossroad of the ICT industry and the vertical sectors specific security (and safety) frameworks. Mutualized usage of a common infrastructure or components, or algorithms will then have to be complaint with various regulations (thousands of standards worldwide). As formal proof is fortunately not mandatory everywhere and similar to the advocacy for service differentiation, security transformation will have to take into account polymorphic adaptation to the regulation frameworks.

8.1.4 Change in scope

The last set of rationales leading to the need of security transformation for SNS, is the scope of the architectures and game changing among stakeholders

8.1.4.1 Security perimeters

With both the (relevant) extend to digital infrastructure encompassing networks, IoT and cloud and the convergence with services and applications, distribution of authority, liability perimeters are impacted and foster much more interoperability and global vision than before. Security is the result of the composition of all segments and none of those segments can discard their involvements in the big picture. This will for instance be obvious for collaborative detection or response to incident or even investigation for identification of attackers.

8.1.4.2 New stakeholders' game

Multiple-providers End-to-End from physical infrastructure to applications will be I charge of protecting users data, enable lawful interception, ensure continuity of services, etc. Distribution of roles between manufacturers, tenants, virtual providers, managed security providers, service provider of all kinds is the purpose of evolution. Despite uncertainty of this evolution and from technical point of view, global semantic of security will impose those actors define ways to exchange and cooperate against threats. Another side of the stakeholder picture is the hackers, on the one hand they are relying on structured, powerful criminal organization using sophisticated technologies (even supported by some states), on the other hand ethical actors and their role in developing bug bounties for instance is of prime importance to consider in scope for maintenance of secured networks and services.

8.2 System-wide Security challenges

This section will first draw security challenges sorted along security phases, some of them being already implicitly addressed in the transformation discussion of the previous section. Then more focused areas will be depicted in relation with the specific architecture directions proposed in the framework of SNS.

8.2.1 Further Security challenges in phases

8.2.1.1 Identify

The recent trend in network attacks has demonstrated that the network as a whole is both a target and a vector of attacks. The Mirai botnet has demonstrated the capability to knock out even large organizations of the Internet. The Wannacry worm has demonstrated the reach and extend of networking connectivity, connecting every device to the Internet and exposing its vulnerabilities to the whole world. Malicious applications are increasingly present in application stores, providing computing power, storage and over all access to sensitive information to attackers, free of charge. In addition to topics already mentioned in other sections, it is worth to mention some specific challenges to AI usage in this phase. The first one is related to cyber ranges, where AI should help to model systems and services as well as attack/response schemes in order to train SNS actors to relevant strategies for SNS protection. The second is also massive usage of Ai in testing, evaluation and certification methods and tools together with Digital Twins. Another important one, is the Cyber Threat Intelligence (CTI) area based on Opensource Intelligence OSINT and Dark Web crawling allowing to predict signs of malicious activities.

8.2.1.2 Protect

Here, specific challenges are related to the protection of the systems and Services at various dimensions, from components to systems and services including processes and organization of the providers. Therefore, the first challenge of next generation networks is to include protection and resilience by default, deeply rooted in their architectures, so that attacks become harder to carry out, less effective impact-wise, easier to attribute. This requires that trust anchors are put in place and resilient configuration patterns deployed, so that networks or virtualized network overlays can resist attacks, and that it becomes impossible for attackers to abuse the network to inject malicious management traffic.

- **Physical layers:** protection against eavesdropping (IMSI-catchers,…) and jamming are two areas of renewed challenges
- **Data Confidentiality:** including control, monitoring, management data handled by the systems. Although, issues related to Lawful Interception remain in the scope of SNS and may be antagonist to the measures for data confidentiality.
- **Identity and Access**: challenges related to the transitivity of rights across domains and services. Including performances issues raised by high speed mobility.
- **Smart orchestration**: challenges related to the smart time and space distribution of security functions across the systems and means to ensure compliance with security policies.
- **Security strategies**: Risk assessment, and risk management strategies, potentially involving AI/reasoning taking into account limited visibility of the global system.

8.2.1.3 Detect

Whatever the quality of protection measures, attacks will be there. Worse attacks will be more sophisticated and with potential large-scale impacts. Detection must be timely (in the sense that attacks should not have created damage by the time the alert is handled by the security operating centre) and accurate (provide limitations on false positives and false negatives).

- **Distributed Denial of Service**: mainly due to the low level of security and control over termination points with no hopes of significant enhancements. Local and distributed (correlated, cooperative) detection is still a challenge as well as appropriate counter measures.
- **Smart Orchestration**: As for protection, there are open challenges to optimize the time and space distribution of EDR, probes and means to detect known attacks or anomalies in case of zero days attacks.

8.2.1.4 Respond

For a long time, response to incident were not considered with a lot of developments with often the fear that the mitigation may be worse than the problem. Nowadays, the flexibility gained in the orchestration of the system enables powerful strategies including planned safe configuration and/or contextual/reasoning dynamic capabilities.

Respond mechanisms such as communications, planning, mitigations,…or any means involved of the respond must be protected themselves against attacks and should not be another vector of attacks.

8.2.1.5 Recover

It should also be actionable, leading to either the elimination of the source of the attack, or to a limitation of its effects.

Learning from experience takes longer time scale. Recover should be the matter of studies based on learning (including AI) of vulnerabilities and sequences of attacks. This experience should in turn used to make a loop feeding the previous phases.

8.2.2 Specific challenges as per SNS architecture

In this section, in addition to challenges already listed above, we give a focus on security challenges related to the intrinsic architecture of the SNS vision.

8.2.2.1 Control and Management security:

Programmability of security: keeping the systems into security policies boundaries mandates to use convergent tools and technologies similar to pure networking functions. Orchestration, chaining, will be enablers of concomitant deployment of security. One of the major challenges is that security should mandates periodic evaluation/verification operations without disrupting the services running and performances of the systems.

Secured programmability: the expected programmability of the infrastructure shall be handled extremely carefully from the security point of view as consequences of vulnerabilities in this area may result in massive outages, data leakage or large impacts on economy and society. The control/management infrastructure shall become more resilient to attacks, against configuration protocols, routing and naming.

- Highly secured Identity and Rights management over control means.
- Safe and secured workflows (protocols, APIs,...)
- Ensure quality and integrity of data used to fuel the AI
- Master the entire software life cycle, from safe code to updates over the air

Another challenge is the capability to better understand attackers, to enhance detection. Beyond honeynets and honeypots, there is a definite need to improve the traceability of activity occurring in the network. This requires the ability both to generate better logs, to make it more difficult for the attacker to hide inside the flow of ”normal” traffic, but also the ability to include tripping points in the network, that will signal anomalous activity. This can be extended to the security at RF level where a complete spectrum and contextual awareness will help to detect and countermeasure different Access Network threats like radio jamming, signal spoofing, fake access network nodes and user equipment or illegal use of spectrum resources. This will support the development of better detection and deterrence methods, reducing the attacker’s gain and increasing its risk. From a non-technical standpoint, it also means that it should become more expensive for attackers to attack, and more expensive for developers to create vulnerable applications than safe ones.

8.2.2.2 Multi-tenancy, multi-party

Systems and services targeted by the SRIA result in much more complex relationship mesh than managing default access to a best effort network.

Multi-tenancy enabled by virtualization of resources, eventually provided as slices is raising issues related to the re-distribution of authority perimeters, isolation, regulation (lawful Interception) or forensic/liabilities.

Multi-party is raising more than ever the issues of End-to-End security as a composition of segment/layers. Considering individual perimeters without concerns for end user resulting security is not an option anymore.

8.2.2.3 Network virtualization and slices

The network virtualization and associated slicing concept are major aspects of foreseen architecture, combining Software Defined Security and Security as a Service promising set of solutions. The network-slicing concept implies de-facto some sort of sharing of the control/management with the underlying infrastructure. The immediate obvious security issue is related to the isolation between slices but already in [C8-5] and [C8-6] NGMN listed a set of key security issues beyond basic isolation concerns. Thus, as of today, the 5G slicing security is one security area among many other specific issues that are already identified. As a continuity of 5G, Next Generation Internet will encompass non-3GPP domains, extended roaming procedures, Service Based Architecture (SAB, including Security as a Service) and will certainly bring unprecedented architectures, services or business models. Considering the 5G architecture and procedures introduced in [C8-7] and [C8-8], security areas are more specifically addressed in [C8-9] and [C8-10]. Going far beyond 4G complexity, 5G and beyond is imposing to reconsider many security aspects such as Authentication and Authorization, RAN (multi-access) security, User Equipment (including IoT), confidentiality and key management, etc. and finally Network Slicing security. An overall vision of this security landscape is given in [C8-11] where one can find references to groups and bodies active on 5G security such as ETSI – in particular working groups dealing with Network Function Virtualization (NFV) and management issues (MANO), IETF, IMT-2020, etc.

The nature of 5G and beyond components, systems and services lead to an unprecedented combination of specific software-based vulnerabilities, function distribution, boundaries variations in time and space. Moreover, the multiplicity of stakeholders and authorities in the case of network slicing raises serious challenges which are exacerbated by the so-called Mission-Critical support. Through technological but also architectural and business aspects, or even regulation, the slicing shows a novel attack surface but also great opportunities to deliver the relevant level of cybersecurity. The slices are software-based and as such, inherit among others from SDN and NFV security issues and solutions. Many threat Intelligence aspects are already addressed in the literature including a comprehensive survey [C8-12] or, for instance for SDN-NFV components, available as Common Vulnerabilities and Exposure (CVE) list

When considering knowledge of slice assets and its consequences in terms of security issues, a specific problem occurs by nature as the slices are both:

- an abstraction/composition of the actual systems and services delivered by third parties and,
- flexible, dynamic or even adaptive to satisfy the varying needs through multi-party complex business environment.

Ensuring trust and consistent security policies/governance between tenants and providers point of views is a remaining challenge. This may be considered as a pre-requisite when applied to vertical sectors under stringent security and resilience requirements, beyond existing standards and certification schemes often applied to limited perimeters. There is thus a need to evaluate and expose security attributes of subnets, systems and services involved in the composition of the 5G slices.

8.2.2.4 Edge and IoT

The current IoT vision is grounded in the belief that the steady advances in microelectronics, communications and information technology witnessed in recent years will continue into the foreseeable future. However, technical flaws and threats of intrusions might significantly lower the benefits of the new developments. Traditional protection techniques are insufficient to guarantee users' security and privacy within the future unlimited interconnection. There is a widely acknowledged need to guarantee both technically and regulatory the neutrality of the future Internet. Moreover, all aspects of security and privacy of the user data must be under the control of their original owner by means of as simple and efficient technical solutions as possible. Security and privacy aspects need to be taken into account also in Sections 4.6 Security for Mission-critical Services, 5.3 Massive IoT Services and 9.2.2 Impact of IoT on the Network. Behind the tremendous diversity of applications and deployment constraints, IoT relies on common requirements requiring a cross-sector standardization approach. The first wave of IoT adoption, driven by ICT companies, focused mainly on business / consumer oriented upstream data gathering for centralized cloud processing. Regulations protecting individual privacy are a predominant concern for implementing security in such applications. Industrial IoT, on the other hand, requires closed loop connections between sensors and actuators with real-time constraints, requiring fully distributed processing and filtering of information exchanges on open networks. Moreover, in several sectors, security breach can result in critical safety hazards. While ensuring cybersecurity in the internet world of powerful mainframes computers connected to main power supply with near-permanent high bandwidth connections and close human control is already a challenge, IoT cybersecurity has to adapt to reduced computing resources, sporadic low bandwidth connections, battery and constrained devices, long lifetime, limited human involvement and other operational constraints.

A proper balance needs to be established between the traditional internet approach of addressing dynamically discovered security breaches through reactive software updates, with the real world paradigm of ensuring safety by careful initial physical design. Security is a moving target and this is particularly true with the Internet of Things. The role of AI in constantly keeping up with security threats is not to be under estimated. Artificial intelligence plays a vital role in detecting intrusions and widely considered as the better way in adapting and building intrusion detection systems. Neural network algorithms are can be applied to real-time security problems

Scalability: Here the main intrinsic system characteristics are related to the scale of devices and components involved, most probably not coming with strong individual security capabilities. Therefore, under constraint of limited CPU, network resources availability, secured hardware and software platforms, challenges will encompass:

- efficient and optimized root of trust distribution,
- crypto light adapted to the context such as Size Weight and Power (SWaP) constraints.
- relevant security protection and detection capabilities
- smart deployment under security policies
- collaborative composition and response to security threats

Micro service interconnect/mesh: the evolution of paradigms leads to modified patterns of exchange (less point to point) and is raising some new or renewed challenges:

- protocols such as QUIC need to be considered for more flexibility in secured exchanges

- numerous service components will be used to achieve overall services delivery, this lead to challenges about composition respecting numerous potentially orthogonal criterias (security green, mix criticality,…)
- overall constancy is involving

In-network processing and convergence with applications: The last challenge is the coupling of network and application security. While network security is handled independently of application and services security, there is a high likelihood that certain network attacks occur because of application or service vulnerabilities, or that network insecurity impact services. This challenge deals with the coupling of network and application modelling, to support more effectively anomaly detection than has been possible before. Only by getting more and better data as input for activity models will we be able to leverage AI/ML techniques to build efficient anomaly detectors, which are able to include contextual information in alerts to support threat mitigation. The smart binding between what occurs at application level and zero trust approaches are actually not in competition as the second address quite exclusively, confidentiality issues and not availability for instance.

8.3 Operational Security Research directions for System & Services

In the following section, we explore research directions aiming at solving at least part of the challenges and issues described in the previous sections. Central,--if not a pre-requisite, the capability to quantify the various levels of security is proposed in relation with operational integration at various stage of the systems and services. This last direction is then complemented with a companion in terms of Security as a Service, in turned followed by some of the promising introduction of some deep tech such as DLT and AI.

8.3.1 Security quantification

The complexity and fragmentation of the system & services landscape, although coming with large openness are mandating the usage of mechanisms to identify, evaluate, certify the level of security. Based on such quantification, trust can be given to providers and services enabling the growth of the sector by answering the diversity of demand from mass market to specific B2B verticals.

8.3.1.1 Evaluation, certification

Current state of the art in terms of certification methodologies have two main drawbacks:

- This is based on Long, costly and complex procedures, based on numerous national and international frameworks/standards but finally for limited product perimeters (piece of codes, chipset,…) and not complex systems or services.
- The variation in time of the security level due to change in configuration, usage, updates, upgrades, … is very rarely taken into considerations whereas real life is intrinsically generating dynamics un usages.

Starting with composition and incremental certification, solutions should come out from this necessary research direction.

8.3.1.2 QoSec

Similar to QoS, security Quality of Security (QoSec) should be defined together with operational feasibility studies demonstrating the value created by matching actual user needs and QoSec delivered by the providers. Robustness to various type of attacks, level of encryption, service disruption time, ratio of anomaly detection,…QoSec should match two major constraints: making sense from user requirements point of view and being measurable.

8.3.1.3 SSLA

Except if the don't need or don't care about security of ICT services they consume/buy, the users should be able to know and monitor the attributes of security attached to the service they are using. Security Service Level Attributes is thus an enabler of security consideration in a service-based architecture. This is applicable to security products and services themselves but needed for most of digital services.

8.3.1.4 Continuous monitoring

As aforementioned, knowledge of slice security conditions is not straightforward considering the multiplicity of authority perimeters and the complexity of dependencies between sub-systems, services etc. Slice security mandates continuous monitoring tracking events or anomalies end-to-end. This dynamic assessment may require both specific advanced tools but will have to face the boundaries of respective stakeholder perimeters. Monitoring or reporting security data is an open field. A particular case is the response to incident which basically require sharing of information from detection towards tenants or adjacent party interconnected.

8.3.2 Green Security

Security functions are not agnostic to the sustainability issues and to relation with power consumption in general. First, energy consumption is known as an attack vector, but what we would like to point here is that, as other functions of the digital infrastructure, smart distribution, smart orchestration and frugal usage of security functions have the potential to considerable savings in a landscape where speed of development has been often equal to greedy approaches. This should not be detrimental to the level of protection, but just to give an example, applying cryptography at optical transport layer or at application layer doesn't have the same foot print…and doing both may be useless.

8.3.3 Security as a Service

Not all tenants will have internal up-to-date expertise available to manage all security aspects. It is thus expected that a wide range of security will be delivered by managed security service providers. Beyond the slicing case, Security-as-a-Service may be considered as one of the very few directions to scale the security needs across the ICT infrastructure and services.

One associated challenge is to manage exposure, negotiation, monitoring, billing of service attributes, and in particular for security services allowing relevant usage of services. Similar to QoS issues, the qualification, evaluation, exposure of those attributes is a difficult problem considering variations of the systems and diversity of risk-based policies. As an example, Identity and Access Management (IAM) as a Service will be key to distribute and control respective authorities and access across multiple 5G systems stakeholders. Many other security aspects (Key Management, Intrusion detection...) will be handled this way participating to the slice security.

Scalability: with ten's of million of enterprise, it is an utopia to believe that skilled people, aware of latest and full Cyber threat Intelligence will be present all the time to protect enterprises or even governmental assets. Even with efforts to increase awareness of citizens, not everybody can be expert in security. Thus, in order to reach the right scale of security delivery, it shall be done as much as possible with Security as a Service provided by trusted sources. Most of the time, the provider will need to come as Managed Security Service Provider (MSSP) in addition to the other parties (connectivity, computing, storage,…)

8.3.4 Security orchestration

Orchestration has been extensively addressed in this chapter as well as the rest of the document. Let's just remind here that from security point of view:

- orchestration will remain a tool to enforce the policies defined by users expectations
- interoperability and combination of the various authority perimeters will be required to ensure global security
- Smart protection deployment, smart detection deployment, smart remediation strategies will greatly benefit from the flexibility offered by orchestration
- Crossing another key area for research, orchestration will be smart and hopefully simple thanks to massive usage of Artificial Intelligence technologies

8.3.5 Disruptive Security Strategies

With the unprecedented dynamics and growing complexity of the Digital landscape, cybersecurity strategies become more and more open to new concepts and somehow disruptions. Beyond traditional perimetric, defense in-depth, protection by-design, one can observe emergence of promising strategies taking benefits of systems properties or new technologies. Research priorities should encompass at least:

- **Deception**: The control and flexibility of the software defined security is an enabler for innovative defence strategies. Among virtualized systems such as 5G slicing related security concepts, two categories are emerging, in both cases using the ability to automate the manipulation of the system morphology: • Micro-segmentation can provide fine grain isolation, specific access control and sticky security policies. • Deception, overcoming historical honeypots, is assuming that advanced (often unknown/zero days) attacks will be defeated by deceiving the attackers with enabled dynamic/smart proactive security.

- **Moving Target Defense (MTD)**: what if your critical assets were in continuous unpredictable (for the attackers) movement? The capabilities offered by virtualization and orchestration enable new approaches and is a candidate for a solution set making life of attackers more difficult.

8.3.6 Distributed Ledger Technologies

The evolution of the ICT infrastructure is coming after years of hyper-centralization, future opportunities and use cases take the assumption of hyper-connectivity for distributed applications. By nature, the widening of scope with edge computing, including mobility or IoT is re-introducing issues of distribution of trust across segments and stakeholders.

The role of the ICT infrastructure, beyond its own security, is to provide a trustworthy platform allowing development and innovation on top of it. Various verticals and applications will benefit from the critical infrastructure with inheritance of the security properties.

With both intrinsic distribution issues and usage by verticals, one of the main challenges will be to manage the component of trust and their distribution, deployment and usage across the highly diverse and dynamic infrastructure. As mentioned before, the application of the perimetric security paradigm is not well suited for an architecture with billions of objects, sporadic activities, and mobility. Novel approaches will have to be developed matching the

”metamorphic” properties, thus delivering smart tools for trust distribution and sharing within the unprecedented system complexity.

Distributed ledgers, or blockchains are promising technologies expected to play an important role where the distribution aspects are predominant. Beyond the known basics of those technologies, issues related to the ”evolutivity”, the interoperability, the sharing between numerous private/public initiatives are specifics of the ICT application.

Further research is required on DLTs integrated with distributed architecture and edge processing to support the regulatory requirements of business/industrial IoT applications and moving from a centralized to decentralized transaction model. Adoption of these technologies faces challenges across scalability, performance and storage.

The research challenges include techniques for increased scalability especially regarding consensus mechanisms (DLTs and blockchains do not scale as required by IoT applications for use in distributed systems), solutions for dealing with processing power constraints against requirements, simplifying implementations, and interoperability between DLTs. Different participation modes are envisioned for DLTs, ranging from fully non-permissioned approaches to various forms of permissioned DLTs. If non-permissioned DLTs are desired, strategies have to be devised to reduce the size of on-chain data and computation. One possible solution is to devise suitable sharding and anchoring mechanisms. Hence, a large-scale DLT could be partitioned into smaller and independent DLTs, which are then bridged together to achieve target performance while ensuring data integrity and isolation.

Another key challenge is how to adjust the operation of a DLT to the different target performance of IoT-based use cases, including lower latency for real-time monitoring applications, or the dynamism of the IoT network where new IoT devices can join, and existing IoT devices can leave at any time. Some non-permissioned blockchains support so called light clients (e.g., Ethereum) which are better suited for IoT devices with limited computational capabilities. The light nodes have to communicate with so called full nodes (holding full state of blockchain) to find the desired blocks and, thus, it might be necessary to place full nodes close to the network edge in order to minimize traffic load and delays. One proposed solution to DLT scaling issues is to use Directed Acyclic Graphs (DAG) instead of blockchain. The main benefits of DAG based DLTs is that miners are not needed and thus DAGs can process larger amounts of transactions with shorter delays and much lower energy consumption. However, there are some concerns about security and resilience against malicious users and thus DAG-based solutions might be better suited for permissioned DLTs. There are already some real-life demonstrations for DAG-based IoT systems and micropayments (e.g., IOTA Tangle) and, thus, supporting such systems at network edge should be considered. Extending multivendor and multioperator federated edge computing all the way to far edge devices will mean that there will be a huge set of computing nodes, attached devices and various versions of software and firmware installations.

Furthermore, at far edge, the physical security of the devices can be much easier compromised compared to centralized cloud infrastructure. Thus, various security measures are needed to enable various involved parties to ensure that the network infrastructure is not compromised. Permissioned DLTs can provide suitable mechanisms for this kind of needs. As the participants to permissioned DLTs can be trusted, there is no need for energy and time-consuming consensus methods, and more efficient methods like Paxos or Raft can be used. Smart contracts can be utilized to create automated mechanisms for leasing edge computation capacity. DLTs can be also used as an immutable log for tracking the lifetime of edge computing equipment, sensors and IoT devices from manufacturing to installation and operation, and finally decommissioning.

8.3.7 Artificial Intelligence

The increased application of AI technologies, such as (but not limited to) Machine Learning and other data analytics technologies, raise the question about their robustness and inherent vulnerabilities. These technologies are a source for new attack vectors. On the other side applying these technologies in the security domain enable more intelligent security solutions. Both aspects of AI and security, the security of these new technologies itself (how to make them more robust and secure) and applying these technologies to create more intelligent security solutions should be addressed in security research with a focus on research areas like:

Dependencies between physical and cyber world aspects in connected cyber physical systems and their impact on security: Today safety analysis methodologies need to be enhanced to also consider physical vulnerabilities. Faked reality attacks (manipulating how sensors see the physical world) will become a new main threat vector by injecting manipulated data that finally provokes malicious responses of intelligent autonomously acting systems. On the other side, security monitoring can be improved by integrating information and models about the physical world (e.g. ignoring anomalies caused by a known physical component fault).

Security risks inherent to AI systems: We need a better understanding of the robustness of AI-based systems. These systems should be aware of the limitations of their models and know the assumptions underlying it. A main security threat of AI/ML systems are adversarial learning attacks, the manipulation of the learning sample set, which in the worst case allows to create hidden backdoors.

xAI systems that can explain their response: In particular, deep learning technology needs to be enhanced by integrating and creating human understandable models and reasoning. This is relevant for human interaction with such systems and the acceptance of semi-autonomously acting systems.

Security in black: Security analytics is facing encrypted data. Today security monitoring and anomaly detection very much depend on data that is not available in case of encrypted traffic. This data needs to be substituted by sampling and analysing new types of data and traffic characteristics that are available also in case of encrypted data and allow similar effective security analytics.

Smart autonomous orchestration: beyond statistical learning technologies, AI scope should encompass, part of control theory in order to enable appropriate response to the ever-growing catalog of attacks. From fingerprinting of radios equipment, jamming source detection,… to global coordinated response to incident, numerous AI-based schemes are relevant in the various security phases.

CTI: Automated threat Intelligence creation and management: This will be a differentiating factor in future security business and pervasive use of AI is the means to achieve this. The final goal is to create even some predictive capabilities. Challenges are to automatically collect and analyse examples from deployed security appliances, create new learning examples, measuring the effectiveness and efficiency of currently deployed models, and incrementally optimise the model learning. Another challenge is utilizing a diversity of other Threat Intelligence resources (structured and unstructured) and being able to exchange and combine models.

8.3.8 Human-centric privacy

In future networks, the Artificial Intelligent (AI) and the Internet of Things (IoT) provide an enormous amount of data (Big Data) realizing the possibility to obtain, to store, to process, and to deliver a diversified high volume data that may refer to human sensitive information - and could be acquired even without the awareness of the interested subjects. Automatic profiling (i.e. profiling without any human intervention) of personal data, automatic facial recognition (and in the foreseeable near future even the analysis of individual pheromones[9]) are impressive examples of personal sensitive data processing that will occur in the NGI.

The European Union (EU) since 2012 and stated, among many other principles:

- The IoT shall not violate human identity, human integrity, human rights, privacy or individual or public liberties
- Individuals shall remain in control of their personal data generated or processed within the IoT, except where this would conflict with the previous principle." [C8-14]

5G networks, AI and IoT raise security and privacy issues that could neither be addressed only by GDPR, even assuming its complete compliance by the service providers, nor by a Distributed Ledger Technology. This latter technology guarantees only the a posteriori control of the user data. In addition and in synergy with GDPR rules and Distributed Ledger Technology, innovative scientific and technical solutions must guarantee the users the a priori complete control to the access and the use of their personal data. Furthermore, the recent pressure on Ethical AI, is another aspect that will be indissociably connected to privacy guarantees. In accordance with the above EU statement a new paradigm of “individual a priori data usage control” is required, defined as:

“except in cases of force majeure or emergency, any use in any form and for any purpose of personal data must be authorized always in advance and explicitly by its owner, correctly informed of the purpose of use”.

This highly challenging objective must be achieved synergizing the innovative and revolutionary GDPR directives and new efficient technological tools specifically dealing with the a priori user direct control of her/his data by means of technical solutions and technological implementations sufficiently simple for the use by the common citizens in the NGI [C8-15].

---

[9] https://iapp.org/resources/article/privacy-2030/

9. Satellite Communications Technologies

9.1 Introduction

Today's non-terrestrial networks are based on a communication model in which, from an end-to-end perspective, satellites provide point-to-point forwarding support toward a complex ground infrastructure consisting of the interconnection of heterogeneous networks. In this model, satellites are flying in different orbits, i.e., Geostationary orbit (GEO), and non-geostationary orbit (NGSO) such as Medium Earth Orbit (MEO), Low Earth Orbit (LEO), and very Low Earth Orbit (vLEO), and can be interconnected by inter-satellite links, i.e., communication links between satellites flying in the same orbits. Satellites in different orbits essentially operate as separate systems providing different type of services and applications' support, from broadcasting to broadband, from critical infrastructure interconnection to Internet of Things (IoT). Integration of these separate systems in the overall telecommunication infrastructure has been mainly based on proprietary and custom solutions. Moreover, satellite networks have been so far considered completely independent of terrestrial networks and mainly used to provide backhauling solution for the terrestrial infrastructure. However, it is now well understood that future networks and systems will be able to provide access to all essential services everywhere, anytime, at any device only through a shift of paradigm in which heterogeneous networks are integrated into a single network of networks.

The current Non-Terrestrial Network (NTN) architecture shall therefore evolve toward a flexible, and yet scalable and cost-efficient, hierarchical architecture [C9-13, C9-14] seamlessly integrated with the terrestrial telecommunication infrastructure and able to support space-borne nodes, e.g., satellites in different orbits, i.e., GEO and NGSO, and air-borne nodes, i.e., High Altitude Platforms (HAPs) and Unmanned Air Vehicles (UAVs). Flying nodes are possibly interconnected through inter-satellite, or inter-node, links providing support for communications both intra-segment, i.e., among nodes of the same constellations, or inter-segment, i.e., among nodes of different constellations. This creates a three-dimension multi-layered NTN architecture fully supporting future networks, systems, and services.

The research and innovation challenges that shall be addressed in the next decades to fully enable this architecture and thus exploit the full potentiality of the NTN component are presented in this chapter along with their expected impact.

9.2 System architectures

Different main factors are contributing to the evolution of the satellite system architecture in the next decades:

- growth of data rates and throughput thanks to the use of wideband transponders, higher spectral efficiency figures (about 7 bps/Hz) and the adoption of new frequencies, e.g., Q/V band and even optical communications offering larger bandwidths for feeder-links and intersatellite links;
- advances in Software-Defined Radio (SDR) allowing for flexible, reconfigurable and cost-efficient payloads; the integrated use of different orbits, GEO, MEO, LEO and vLEO;
- the design and deployment of new constellations such as, mega-constellations with hundreds or thousands of satellites or incomplete constellations with few tens of satellites interconnected through inter-segment links;
- reduction of the size of the satellite hardware components, both the primary ones, such as engine, attitude control, battery, antennas, and the payload ones;

- development of new generation of micro, nano, and pico satellites providing lower costs, low communication latencies, low energy consumptions, and high fault tolerance
- softwarization of the ground and space segment aiming at a fully satellite network virtualization based on software defined ground and space segments; integration in the telecommunication infrastructure through orchestration and coordination of the virtualized NTN objects.

In the following the main research and innovation challenges related to the evolution of the system architecture are described

HTS broadband GEO

In the last two years, the next growth axis of the SatCom market has taken shape with the launch of high throughput satellites (HTS) and payloads worldwide enabling efficient highspeed internet-by-satellite. Achieving large capacity is the next milestone for the satellite industry in Europe, to offer enterprises and citizens with an even wider gamma of services provided with unprecedented quality of service. To this end, implementing the concept of Tbps capacity can be sustainable by migrating the feeder links to Q/V or optical bands and delegating the Ka frequency band to the user links. To this end, suitable RF/optical conversion strategies implemented on-board the satellite (and on-ground) are desirable to implement microwave photonic on-board processing functionalities. On the other hand, GEO satellites will be also crucial in providing coverage with IoT direct access for non-delay critical applications with ultra-low data rate requirements [C9-15]

HTS broadband MEO

MEO satellites (e.g., O3b with a constellation of MEO satellites) is addressing emerging and insufficiently connected markets mostly in Latin America, Africa, the Middle East, Asia, and the Pacific while offering a low latency approach. One of the main advantages is to provide data access at a reasonable latency given the lower altitude of MEO satellites with respect to the GEO counterparts. However, a more advanced design of antennas is necessary to boost the performance.

LEO constellations

The very next frontier will be represented by broadband mega-constellations equipped with optical inter-satellite links. On the one hand, the large number of satellites will help achieve high-granularity coverage, so that higher capacity will be made available from the entire satellite system. On the other hand, the availability of optical inter-node links will drastically reduce the latency of re-routing operations in space, taking also into account that the speed of light in free space is higher than in terrestrial optic fibres. Moreover, the recent improvement in small satellite technologies is making the employment of small-satellite-based solutions appealing in different use cases including the IoT and in particular in LEO satellite constellations as testified by the many industrial initiatives aimed at offering global IoT coverage thus enabling IoT devices to transmit data irrespective of their location and form factor.

vLEO Constellations

Very Low Earth Orbit Constellations, i.e., at an altitude lower that 300 km, are expected to play a fundamental role in future networks thanks to several benefits encountered at those altitudes, e.g., deorbiting time, lower radiation effects, etc., and to the increasing availability of low-cost launchers. vLEO will be of particular interest in support of IoT services by means of improved on-board computation capabilities. Their function to extend the terrestrial network capacity will be carried out by means of nano- and pico-satellites, which together with CubeSats, will make use of high data rate links built on Ka-frequency band or even optical links.

Hierarchical Aerial Networks (Satellites + HAPS + UAVs)

UAVs and HAPs provide low-cost support for a large class of applications and can be used both individually or as a swarm. Using multiple UAVs and/or HAPs needs coordination and data exchange services among them, leading to a multi-layer hierarchical infrastructure known as Space Information Network (SIN). SINs are complex network infrastructures relying on heterogeneous network segments implemented by space platforms, such as satellites, UAVs, HAPs, and airships. They can play a key role in many different applications: connectivity for otherwise disconnected areas, emergency communications, environmental monitoring, mMTC, IoT, to cite a few. For instance, in rapid deployments, drones can be considered as backhaul hubs to integrate small cell fronthauls where fibre connectivity is not available or limited due to NLOS wireless fronthaul.

Fractionated and cooperative constellations

New concepts of satellite constellations will also include the cases of incomplete and cooperative constellations, i.e., the satellite coverage is not continuous and the connection is therefore intermittent, where satellites pertaining to different constellations, e.g., vLEO and GEO, communicate through inter-node links and, if on board computation capabilities are available, also cooperates.

Three-dimension multi-layered architecture

Extension of the above constellations into the three-dimension multi-layered architecture integrated architecture is one of the future challenges to achieve the so-called integrated space data highway, consisting in integrating different classes of satellites into different tiers of the same hierarchical satellite architecture. In other words, the integration of GEO, MEO, LEO, HAP and UAV in a unique mesh network in the sky will help achieve large capacity, resilience, flexibility and cost-efficiency to be offered to end users.

9.2.1 Expected Impact

The potential impact of satellite networks in this described ecosystem composed of new applications developed in a plethora of scenarios is linked to their intrinsic ubiquity and broadcasting capabilities. Satellites can act as a main single backhaul segment for rural areas, aircraft, vessels, and trains; as additional backhaul means to opportunistically provide additional connectivity/bandwidth resources, also improving service continuity; or as a pure transport subnetwork. Associated outcomes are in the field of Smart Cities, Smart Industry, and Smart Farm.

9.3 Evolution of Networking Architectures

The appearance of smart miniaturized satellites has facilitated the deployment of satellite constellations in a large number (e.g., hundreds of satellites). These deployments, albeit complex, introduce the possibility to explore satellite-to-satellite communication in an opportunistic way, to assist in the support of the most varied Internet scenarios, and particularly covering wide area regions. At the same time, the Internet fringes are expanding with the integration of the most varied IoT devices. The Internet fringes are rapidly expanding and also embracing the most varied IoT devices in addition to classical users' equipment, which interconnect to the internet in large-scale and over large distances [C9-1]. This added complexity brings in the need to rethink the underlying support for end-to-end data exchange on satellites. The design of underlying network architectures, where the majority of services are IP-based, needs to take into consideration new challenges such as the large volume of mobile devices involved in data exchange; the increasing number of technologies to support; the possibility to explore direct data exchange between satellites, Moreover, it is essential to consider satellites as smart in the view of their onboard processing and computing capabilities, hence making them Edge devices.

The design needs therefore to evolve to accommodate, firstly, a significant growth of end-user and IoT devices. To better support the resulting connectivity demands, it is firstly necessary to enable opportunistic cellular (D2D)/wireless data exchange by ensuring, for instance, a smooth integration of Delay Tolerant Networking (DTN)/Information-centric Networking (ICN) paradigms into elements that are today the basis of the IoT data exchange, e.g., gateways, or specific Edge devices.

Secondly, the current terrestrial-airborne networking architectures need to evolve by accommodating smart satellites as new Edge elements. It is imperative to consider ways to support global interoperability in a way that is interoperable with today's networking architectures, which consider TCP/IP based services and consider new environments, such as Industrial IoT, where legacy non-IP devices abound. It is essential to address multi-provider network management and ways to automate the management of data exchange, also considering how to assist direct data exchange across satellite constellation. It is also relevant to address a more dynamic routing behavior on selected pools of ground stations, which may support dynamic, secure IP routing (e.g., based on context-awareness).

In this framework the following challenges must be addressed.

Design Smart Satellites as Edge Nodes.

From a networking perspective, this requires a better interoperability definition on how to best handle requirements coming from different heterogeneous networks (e.g., ULL, LTE/5G, WiFi 6/WiFi 7) with some degree of automation, and securely [C9-3]. For instance, which functions need to go into a gNB to support DTN/ICN, or whether a smart satellite should behave as a gNB. Also, how to support translation between IP-enabled field-level devices or terminals (e.g., accommodating an automated translation between different IoT protocols such as OPC-UA, MQTT and DTN/ICN). Such design also requires research on how to handle ways to integrate dynamic routing and context-awareness into ground stations (DTN/ICN support, for instance); how and whether to integrate simplified forms of routing into smart satellites.

Dynamic network management across multi-provider, terrestrial/airborne environments.

An evolution of the terrestrial/airborne network architecture needs to assume the deployment of large constellations with thousands of smart, IP-enabled satellites in multiple orbits, managed by a single provider, and the interconnection to ground stations managed by another single provider, as well as the offer of multiple services by different providers. To enable a rich and living ecosystem, openness for operators, vendors and service providers on different stages is mandatory. Aspects that should be addressed today to assist this support relate to identifying a best content-centric type of global routing, for instance, a form of source-based routing. It is also relevant to address which simplifications would have to be done to meet small satellite constellation requirements, e.g., low-energy routing). Other aspects concern also the support for inter-domain SLAs, routing, handovers, security, and other topics dealing with interactions of providers and network management.

Assist the integration of direct communication across satellite constellations.

DTN technology, currently being applied to deep space environments as well as to remote terrestrial areas to cope with tolerate intermittent connectivity (see Incomplete and Cooperative Constellation in section 9.2) and large delays is attractive also for satellite-to-satellite communications, as it relies on the store-and-forward principle and can handle well network partitions. DTN pushes data to be transmitted into ”Bundles”, which are then transmitted on every possible opportunity between networking nodes. Networking nodes store persistently the received data. Nodes can then send their information without a full knowledge of the topology, and to act in the case of intermittent connectivity, which can be opportunistic or scheduled, thus making DTN relevant also in the context of future smart satellite constellation communication. The ICN paradigm can provide also support for intermittent connectivity and brings in semantics relevant to security and to mobility. Moreover, ICN principles such as the Pub/Sub receiver-driven model are today highly relevant to assist data exchange in environments such as IoT, where an N-to-M asynchronous data exchange is required. The name-based routing of ICNs is a relevant aspect regarding mobile environments. Coupled with DTNs, the ICN paradigm brings in the possibility to further explore asynchronous communication between earth and space.

9.3.1 Expected impact

The Internet is evolving towards service decentralization, having data and application centers scattered across Cloud/Edge. Moreover, the increasing number of IoT and Industrial IoT use cases will lead to a tremendous increase in devices and increasing heterogeneity in communication relations between users, data sources, data sinks, back-ends, and Clouds/Edge-Clouds. Smart satellites and Cloud computing are 2 separate worlds, even though there are some initial efforts to bring these words together, as occurs, for instance with some providers such as Amazon[10] which starts to develop some form of Edge transition. At the same time, there is a new generation of more complex Internet services to support, e.g., decentralized AI; Augmented Reality; Mobile Crowd Sensing.

The evolution of the underlying network architectures, addressing the aforementioned challenges in a way that considers that smart satellites are Edge nodes with the capability to store and forward and to pre-process large sets of information will ignite novel business opportunities and allow satellite systems to evolve beyond the usual forwarding aspects. Satellite-based data centers (such as SpaceBelt, 2019) are highly appealing from a business perspective, as the storage possibilities increase, and the cost lowers. Adding to this, the possibility to store highly sensitive information in areas, which today are not (easily) hackable, are an additional selling point.

9.4 Hybrid infrastructures: Broadcast/Multicast/Unicast/Storage – EdgeCasting

It is only by combining different networks and technologies that we can have a sustainable communication infrastructure: using Broadcast and Multicast with local storage capabilities is an essential element for a reliable and sustainable delivery infrastructure reaching at any time 100% of the population on 100% of the territory.

Combining unicast IP with IP broadcast and multicast push services, using Satellites + Terrestrial Towers + Cellular to deliver: events interesting large number of users and entire territories (broadcast), non-real time multimedia contents (push multicast to local storage), one to one personalized contents (unicast). Contents will encompass information, entertainment and educational but also other public service contents (e.g. live traffic/alerts, navigation corrections and emergency information) and more in general software and information distribution to large population of users with a zero-marginal cost per additional user. The same contents delivered to mobiles/vehicles can be received and managed at the very edge of the network (end devices) and at the level of any edge server in general, using a local storage to maximize efficiency and economical sustainability. Content and services of media organizations, in particular public service media (PSM) organizations, have evolved from a limited number of linear channels into a rich and differentiated offering across digital distribution platforms including IP-based services which, often through media players, now also provide nonlinear and catch-up content – content and services which are increasingly popular with audiences.

Audiences now widely consume media content and services (audio and video) on conventional and smart TV sets, portable and in-car entertainment systems, smartphones, tablets, personal computers, smart speakers, etc. Much of this consumption happens in the home, as well as on the move.

Conventional broadcast technologies (e.g. DVB-T/T2) could deliver linear services to portable and mobile devices. However, these standards are not widely supported in such devices. Typical HPHT broadcast networks for TV are in general not designed to support mobile use cases and will need to be adapted for such use cases, e.g. for public transport in cities.

---

[10] https://aws.amazon.com/ground-station/?nc1=h_ls

Additionally, conventional broadcasting standards are not capable of delivering nonlinear services without a complementary IP connection.

In principle it is possible to deliver linear and nonlinear media content and services to mobile and portable devices with conventional mobile broadband (e.g. unicast IP over mobile networks), however questions remain about this distribution mode, such as:

- Whether mobile networks could provide adequate quality of service to large audiences by means of unicast connections.
- Whether unicast would be affordable for consumers given data volumes consumed by media content.
- Whether distribution over mobile networks is affordable for service providers - and is compatible for example with the remit of PSM organisations.

5G has the potential to provide answers to some of the questions above, particularly for linear services which can be delivered with 5G's broadcast functionality. This may enable meeting the media industry's technical, commercial and regulatory requirements at the same time. The 5G Broadcasting solution provides for the development of a cooperative network that combines satellite, terrestrial and cellular infrastructure in an intelligent way. By combining those technologies, the potential cost savings are relevant. Popular media content currently delivered via unicast can be shifted to push multicast delivery. When combined with the use of storage at the edge of the network or in the device itself, it becomes an even more attractive opportunity. For the distribution of live content such as sports and news events to mass audiences, broadcast will continue to make the most sense. Satellite & Terrestrial networks can cover 100% of territories & population.

9.4.1 Expected Impact

Broadcasters - Under huge pressure from increased global competition and shifting consumption patterns, broadcasters and other content providers need to adapt their distribution models to meet user expectations. The development of hybrid services, combining linear and time-shifted elements, along with personalized on-demand services, using a combination of broadcast and multicast content delivery, and caching paradigms represents a cost effective and sustainable solution. A win for the media industry will see broadcasting reinvented to use collaborative infrastructure that combines the reach and efficiencies of terrestrial and satellite, broadcast and multicast, and caching at the edge with the high throughput and personalized delivery mechanisms of mobile networks.

MNOs -The MNOs are in a strong position, with a well-established business built on direct relationships with end users and strong leverage over device manufacturers. But they need to greatly expand their media content offer which – in their current model - comes with a heavy investment burden. A win for the MNOs will involve enhancing their media offer to both mobile devices and cars via 5G broadcast and multicast modes, as a powerful means of efficiently and cost-effectively using available network resources, such as those offered by a seamlessly and fully integrated NTN component (see section 9-14). Efficiencies can be translated also in huge cost savings as speculated in the market assessments.

Automotive - As our transport infrastructure becomes ever smarter, we will require networks that can meet the need for entertainment, navigation, safety and software updates. Such networks need to cover 100% of the territory and 100% of the population with guaranteed quality of service (QoS) at a sustainable cost/user. Currently, no single infrastructure can achieve this. A win for the automotive industry will involve equipping connected cars and networks with the intelligence to use broadcast, multicast and unicast in an efficient and reliable configuration that enables the full potential of smarter, safer and – eventually – self-driving cars.

Emergency Services - Alerting the public in emergencies, whether natural or man-made, is an essential element of public safety systems. Both broadcasters and telecoms operators have a regulatory obligation to build the necessary infrastructure and put it at the disposal of local, regional or national authorities. This is technically challenging and costly. A win for the national authorities, broadcasters and telecoms operators, and ultimately the citizens, will be a reliable, cost-effective, advanced network infrastructure with near-universal coverage of the population and territory that can support public warnings in emergency situations.

9.5 Smart Satellite Networking

To really allow the satellite systems to be a component of the 5G network and to be considered one of the possible solutions to guarantee connectivity and best performance, as well as other technologies, it is strictly mandatory to push softwarization, implying to develop all the functions necessary respecting the NFV/SDN paradigm and 5G standard, including the functions implemented in the modem. Then, some functions shall leverage on cloud capabilities and on AI/ML, making SatCom fully compliant with terrestrial technology.

The adoption of a NFV framework standardizes the reference points along which the components interact among one another. It natively supports service chains, providing the basis for decomposition and provisioning of complex network services via a series of interconnected network functions, where each of which could reside in a different (virtual or physical) machine.

The research and innovation challenges are to design and implement a number of functions realizing also a service chaining by constructing a forwarding graph interconnecting network functions in the NFV layer (Virtual Network Function Forwarding Graph - VNFFG). A detailed, but non-exhaustive, list of virtualized functions to be developed to make the satellite network fully compliant with 5G architectures is: Transport Layer Optimization VNF, Deep Packet Inspection VNF (DPI-VNF), Adaptive modem and CODEC box, multi beam multi gateway, Bonding proxy, Virtual Cache, Virtual PEP. Some of the listed activities may imply the redesign of the radio access and physical layer to be suitable with these paradigms. More ambitious virtualization research could address the SatCom mission segment functionalities enabling: higher degrees of automation, the instantiation of satellite resources, multi-tenancy schemes and secure SatCom infrastructure “virtual” fragmentation, enhanced and more accessible SatCom resource allocation and usage, beam-hopping payload capability. The need to cope also with mobility shall imply to develop efficient and seamless handover at beam, gateway, and satellite level, considering also the advent of mega-constellations. Therefore, suitable orchestration schemes, fully centralized or distributed to better cope with the increase of processing delay, shall be developed. The heterogeneity of the technologies involved in the setup of complex architectures requires the definition of new network management models, unifying the different models, each one referred to a single technology, to get a flexible and consistent network management plane, with consequent implications also at system orchestration and security level.

9.5.1 Expected impact

The outcomes of these activities are expected to be the availability of a full set of virtualized functions to be utilized by service providers and virtual telecom operators (agnostic from technological point of view but wishing to achieve the best performance satisfying user requirements) installed on servers along the network infrastructure, ready to be instantiated when needed. The network management will be particularly impacted in terms of resource allocation and dynamicity. The impact will also be an increased use of the satellite segment because no longer depending on the capability to convince content providers, OTT, telecom operators and verticals that they need the satellite but on the optimization of network management algorithms which will utilize the satellite when this really improves performance and increases efficiency. In terms of performance in strict sense, an additional contribution to overall latency introduced by cloud computation shall be considered.

The most promising service scenario in which the utilization increase can be more meaningful is CDN management. Both satellite terminals and (more futuristic) payload may be equipped with caching capabilities but, mainly, taking advantage of the intrinsic best cost/effectiveness of satellite-based architectures to set up multicast connectivity, the feed of already deployed CDNs directly through a satellite link has been demonstrated to be optimum and can represent an interesting market perspective.

The development and deployment of networks based on virtualized functions, which will include the satellite segment among the possible technologies to be used to provide efficient services due to the introduced flexibility, will also unleash new business cases for virtual operators. In fact, it will be possible to set up complex networks optimized on specific services satisfying the relative requirements and QoS transparently for the final user who may not be aware to use the satellite but will just enjoy the performance improvement.

9.6 Optical based Satellite Communications

In future High Throughput NTN networks, the use of multibeam coverage, through multibeam satellite or distributed antenna systems (see the Antenna Section) will require the feeder to transport enormous data rate. Innovative approaches shall be therefore researched and developed. Amongst the different solutions proposed, using optical wireless technology in the feeder link is certainly particularly appealing for the following reasons:

- Wider EM spectrum. The achievable data rate in a point-to-point optical wireless link can be few orders of magnitude larger than with an RF wireless link.
- License-exempt. Optical wireless systems do not require a license for operation.
- Interference management. The propagation conditions of the optical wireless channel, as well as the high directivity gain of the optical wireless transmitter and receiver telescopes, mitigates notably the interference that is generated in ultra-dense deployments.

In the design of the optical feeder link the research and innovation challenges to be addressed depend on the payload characteristics.

In the *“ideal”* case of a **fully regenerative payload**, the optical feeder link terminates in the satellite; therefore, robust Forward Error Correction (FEC) can be used to correct the long burst of bit errors that the turbulent optical satellite channel introduces. Moreover, this approach enables the implementation of advanced signal processing mechanisms on-board the satellite, such as beam precoding, enabling higher spectral/energy efficiency. In contrast, the fully regenerative payload demands the highest processing power on-board the satellite and gives limited flexibility to adapt to modifications on the communication standard during the lifespan of the HTS system. Therefore, transparent non-regenerative *“bent-pipe”* solutions should be initially implemented regardless if they are analog or digital transparent.

In the **analog transparent** payload, the instantaneous value of the RF signal modulates the Intensity of the feeder link Laser Diode (LD). The most practical approach to implement the analog transparent optical feeder link consists in multiplexing the RF signals intended to the spot-beams of the HTS on a different wavelength in the infrared spectral window (1064/1550 nm). This Wavelength Division Multiplexing (WDM) solution enables high flexibility due to its channel scalability. For example, typical dense WDM systems in the C-band (1525-1565 nm) can accommodate 40 (80) optical channels at 100 (50) GHz spacing. For ultra-dense WDM solutions, 25/12.5 GHz spacing is also possible. The random turbulence that exist in the atmosphere leads to wave-front distortion and scintillation. For the uplink, the smaller aperture size of the receiver telescope leads to deep turbulence-induced fading events. For the downlink, the spatial dissemination of the received beam creates a similar effect after coupling the received optical signal into a Single-Mode Fibre (SMF). Different techniques exist to mitigate the effect of turbulence, such as 1) transmit micro-diversity schemes, adaptive Optics, and macro-diversity schemes.

On the other hand, in the **digital transparent** payload, I/Q components of the baseband radio signal are oversampled and quantized to guarantee an acceptable Signal-to-Distortion Ratio (SDR). After that, the resulting sequence of bits modulates (digitally) the optical carrier of the feeder link. In the digital transparent payload, only limited signal processing capabilities are required on-board the satellite to reconstruct the complex-valued I/Q baseband signal; moreover, Forward Error Correction (FEC) can be used to protect the transmission against the impairments of the optical turbulent channel. Unfortunately, the digital transparent approach limits the aggregate data rate that the HTS system can support, as the digitalization process expands *many times* the original bandwidth of the complex I/Q baseband signal that wants to be transported.

9.6.1 Expected Impact

The main expected outcome is to identify a unified technology for the implementation of the optical feeder link. So far, mostly theoretical studies have been reported on the literature on the different implementation approaches. Nevertheless, if a common standardized solution is found in a reasonable time window, large-scale experimental trials could be later organized, paving the way to a full implementation of the technology in an operative HTS system. This way, a giant leap will be taken to integrate the satellite segment into the Beyond 5G landscape.

Once the implementation approach of the optical feeder link is defined, different signal processing methods to mitigate the impact of the impairments introduced by the optical components of the feeder link, the electrical components of the access link, and the different effects that the atmosphere creates on the optical signal mainly (*i.e.*, turbulence, particles, etc.), should be also taken into account. For this purpose, a consensus on the requirements and implementation limitations should be clearly defined, including the use-cases in which the HTS systems will be involved and the KPIs that should be maximized.

9.7 Software Defined Payloads

Payload softwarization is the key development that facilitates the system reconfiguration to accommodate dynamically changing demands, optimizes cost and processes, and enables self-management and automation of operations. The latter combined with a full digital implementation has multiple advantages, namely beam forming allowing any beam layouts (or their combination), full routing allowing any network topology, gradual deployments, dynamic power and bandwidth allocation and interference mitigation techniques. In particular, the new on-board processing capabilities combined with the emerging role of active antenna systems, open the door to advanced resource management techniques capable of maximizing the satellite resource utilization while maintaining QoS guarantees, and dynamically matching the distribution of the satellite capacity on ground to the geographic distribution of the traffic demand and following its variations in time. To this end, the deployment of active antenna systems on-board represent an additional degree of freedom allowing beam reconfiguration over time.

In a mid-term future, regenerative payloads (demodulation) are foreseen to enhance signal and remove impairments affecting the uplink. Moreover, with the aim of reducing latency and the launching costs, lower orbit constellations are gaining momentum by providing satellite-based data services with a superior end-user experience. The coordination of these small satellites is tightly connected to the resource allocation strategies, not only for maximizing the performance of the forward link, but also to optimize the ground segment (e.g. number of gateways and feeder link allocation/diversity). Inter-satellite, or inter-node, links additionally provide solutions to drive these developments.

9.7.1 Expected Impact

Software-based satellite systems bring important improvements from a network management point of view, by allowing a better orchestration of the satellite resources. Unavoidably, “softwarization” will expand to the whole satellite ecosystem, replacing the custom hardware solutions, resulting in a more flexible and dynamic system with overall better performance and efficiency. Footprints, frequencies and power can be dynamically changed over time to follow the business evolution with the potential to adapt continuously to optimize usage. However, standardization and production in series are key to achieving CAPEX reduction.

9.8 Radio Access Network beyond 5G and 6G

9.8.1 Rationale

In recent years, Radio Access Network aspects have been deeply studied and discussed, in the framework of the NTN integration into the 5G ecosystems. Studies have clearly had an impact in 3GPP and shown the feasibility of such integration and the suitability of the terrestrial solutions to the characteristics of the non-terrestrial scenarios.

It has been shown that, with limited and acceptable modifications, the NR air interface can be adapted to the NTN case, thus introducing a completely new dimension in the overall 5G ecosystem. At the same time, it is also clear that the full potentiality of the future NTN component can be only empowered by research and innovation activities addressing those RAN aspects that can benefit from a design that takes into consideration the future NTN strengths and peculiarities, from the very beginning.

The future NTN component calls for a change in the air interface design paradigm: flexibility, adaptability, and overall efficiency of the RAN design shall become predominant with respect to the classical approach of a link spectral efficiency maximization. The research and development activities shall focus on the design of techniques able to flexibly and dynamically adapt to different requirements, scenarios, and conditions by maintaining at the same time extremely high spectral and energy efficiencies. In particular, the introduction of GSO and NGSO constellations, active antenna payloads, aggressive frequency reuse requires a RAN design aware of the peculiarities of the NTN component so as to satisfy the overall system KPIs.

Also, for those scenarios where the system requirements require a shift from a bent pipe to a reconfigurable and regenerative payload approach, it is of a paramount importance to address the design of highly reconfigurable RAN techniques that, thanks to Software Defined Payloads (see the Software defined payloads section), allow the NTN component to adapt to the ever-changing communication scenarios without the need of costly updates of the flying nodes.

The support of multiple logical networks, and even multi radio access, increases the complexity and requires enables as AI/ML in cooperation with data mining solutions (see the Machine Learning for SatCom section).

9.8.2 Expected Impact

The expected outcome is the design of a unified air-interface that can provide connectivity in an heterogeneous environment, where users may access the network through terrestrial and satellite links, and a reconfigurable radio access network that can be dynamically adjusted to changing conditions and requirements and ease the co-existence of different services. The expected outcome related to radio access network opens new research avenues. As it is summarized in [C9-12], the research and development activities addressing the Radio Access Network of the Beyond 5G and 6G future systems shall provide support for:

- Space-borne and air-borne integration;
- GSO and highly NGSO constellations (down to vLEO for the space-borne platforms) ;

- Multilayer integration and handover;
- Multi NTN system integration;
- Highly flexible and adaptable air interface for reconfigurable, regenerative, and software defined payloads;
- Advanced Radio Resource Management algorithms to allow satellite beams to overlap terrestrial cells and also a seamless terrestrial-satellite integration.

9.9 Antennas

The next generation of satellite communication systems requires a paradigm shift in the field of antenna technologies. Whereas current solutions are mainly based on relatively small reflector antennas with fixed beam patterns, future applications will necessitate the use of larger and/or more flexible antennas able to support higher frequency bands (from Ka-band up to optical bands). These innovative technologies will especially enable the generation of narrower beams for a more efficient use of power and spectral resources. The steerability of the beams will also play a key role to cope with the dynamicity of the traffic demands and avoid interference between different systems (GEO/NGSO, satellite/terrestrial). Advances in antenna design will thus be of paramount importance to make satellite systems ready for very high throughput requirements and fulfil the challenges of multi-orbit and mobility scenarios.

Active antenna arrays

Active phased arrays and metasurface antennas offer the flexibility required for future systems. Whereas phased arrays rely on many low gain active elements (including especially power amplifiers and phase shifters), metasurface antennas exploit artificial materials whose properties can be modified to produce a desired radiation pattern. An advantage of metasurfaces over typical phased arrays is their lower power consumption and cost. It should however be mentioned that advanced approaches for the design of phased arrays (e.g. sparse arrays) have been proposed to improve the power efficiency and reduce their cost. Low-profile electronically steerable antennas are advantageous for the ground segment design since they can be integrated in the structure of mobile platforms (cars, planes, etc...) contrary to complex mechanically steerable reflectors. This will enable to efficiently establish connectivity with different satellite systems operating in various orbits. Active antenna technologies will also impact fully flexible GEO and NGSO satellite payloads with the ability to create a given beam response and support null-steering in any direction and at any time. Full use of the Ka-band along with tight tracking beam footprints, spectrum reuse and adaptive resource control will support the growth of the satellite industry in the coming years. For an optimal compromise between the flexibility and the power/mass requirements of the payloads, hybrid analog/digital beamforming architectures will play here a central role.

Large apertures

Some use cases in next-generation systems will also require the generation of very narrow beams (e.g. to form extremely small main beam footprints with a high power density from a GEO very high throughput satellite). Here, large deployable reflector antennas (~5-30m diameter) represent an appropriate solution. Moreover, the structure of such large antennas will allow a reconfiguration of their shape to adapt the coverage during the operational lifetime of the satellite. Since the maximum diameter of large deployable antennas will be limited by size and weight constraints, solutions such as sparse arrays will be useful. A different approach to obtain footprints of only a few kilometers is to form a larger effective aperture through a distributed and cooperative approach. They consist in the deployment of a large number (>100) of small spacecrafts flying in formation, and possibly interconnected through intra-segment links, to obtain a coherent antenna array of up to several kilometers in diameter. CubeSat swarms are in this case a promising and cost-wise viable approach. A distributed array is also advantageous in terms of maintainability, scalability, resilience and flexibility as elements can be independently added or removed without significantly affecting the system operation. The principle of distributed coherent arrays will also be of interest on ground to build feeder link gateway stations with small aperture antennas instead of using an expensive single large aperture. Meanwhile, due to the challenges still encountered in the design of large-scale distributed arrays (e.g. time and phase synchronization), technological maturity will only be reached within the next decade.

Multiple antenna technology

Since future systems will mainly use spatially separated antenna elements on Earth (gateway stations, users in a multibeam coverage) and in space (antennas on one or several payloads), advanced spatial multiplexing approaches will be used to improve spectral efficiency. The approach which is known as the multiple-input-multiple-output (MIMO) scheme enables to transmit independent signals in the same time and frequency resources and is already widely used in terrestrial networks. The use of multiple antennas can moreover provide additional gains and features for transmission and communications security on the physical layer.

9.9.1 Expected Impact

The development of new antenna designs will be a key enabler for the seamless integration of satellites in existing and future wireless networks. In a digital world where connectivity anytime, anywhere will be required, flexible beam steering will allow the operators to effectively respond to the traffic demands whereas end users will be able to benefit from high speed and lower latency connections with NGSO satellite constellations. Innovative design methods and mass production are expected to significantly reduce the costs of adaptive antennas. On the other hand, on-going activities in the domain of large deployable reflector antennas have also brought advances which will impact in the coming years the design of GEO ultra-high throughput satellite systems for which high power density and/or very small beams will be required. To further increase the effective aperture size, cubesat swarms also appear as another promising solution. Their technical maturity will however only be attained on a longer-term perspective.

9.10 Spectrum usage

The spectrum allocation discussions for satellite and terrestrial networks remains as a very timely topic due to coming 5G rollouts and New Space networks including NGSO constellations. Studies are on-going in re-allocating the used spectrum e.g. from satellites to 5G terrestrial networks also in higher frequency bands, supported by regulatory conclusions in World Radio Conferences. Conventionally each radio system has its own operational band and the world is shifting more and more to flexible spectrum sharing. Satellite industry has good arguments to defend the situation in order to provide ubiquitous critical services as well as being part of development of future spectrum decisions.

In the three-dimensional multi-layered NTN architecture multiple links coexist, i.e., feeder and user links, as well as inter-segment and intra-segment links. Most of these links are implemented with RF technologies, thus requiring RF spectrum. Simultaneously there is a trend to go to higher frequencies both in terrestrial and satellite systems, mainly driven by increased capacity and bandwidth needs that cannot be fulfilled with currently used frequencies. Very high throughput satellite systems aiming to achieve terabits per second connectivity need to use large bandwidths that are only available in Q/V band (30-50 GHz) and higher. It is therefore evident that there is a need to study both how new frequency bands could be used to provide services to end users as well as studying how the existing and new bands could be most efficiently exploited and shared. The overall challenge is to develop new technologies for dynamic and flexible spectrum use and practical licensing models, while being able to protect satellite usage.

The three-dimensional multi-layered NTN system could either be integrated into a single system controlled by the same core network or not so tightly integrated system where e.g.

systems in different layers may coexist in the same area, possibly using same frequencies. Thus, the spectrum use research regarding the multi-layered architecture can be in high level divided into two topics: 1) Going to higher frequencies and 2) Spectrum sharing. Both are enablers for meeting the increasing capacity requirements.

Spectrum sharing

Dynamic spectrum sharing techniques are essential in the envisioned multi-layered architectures to increase bandwidth and enable more services to the users. While previous research has looked at many possible technologies, new challenges arise regarding the new frequency bands and new sharing scenarios. This can consider 1) Sharing between satellite systems, 2) Sharing between satellite and terrestrial systems and 3) Applying AI techniques to enable sharing and predictive routing in a mobile 3D networks considering movement patterns of different aerial and space platforms.

Coordinated ways to share spectrum with guaranteed QoS such as licensed spectrum access (LSA) can be used to provide predictable access to spectrum to sharing parties (such as secondary operators). Predictable quality of service is essential for the operators in order to be able to invest in the infrastructure and the development of services. Many techniques such as power control, beamforming, and carrier aggregation can be used in conjunction with the licensed sharing. Particularly interesting are techniques allowing seamless integration of networks, e.g. NFV/SDN based negotiation and handling of spectrum resources, will make use at convenience of advanced ground-based and space-based radio access and management techniques to ensure systems coexistence.

Alongside spectrum sharing techniques, new technologies are needed to accurately assess spectrum use and avoid harmful interference situations according to the increased complexity of the upcoming spectrum picture, characterized by a significantly higher number of actors and inter-dependencies among them. The proposed spectrum sharing techniques need to be regulatory compatible, economically attractive and viable, and technically efficient.

Higher frequencies and channel modelling

In order to support ever-increasing need of capacity, spectrum sharing can provide some help. However, one needs to look new, higher frequency bands such as W band for new services. In this case, channel characterization and measurements from different orbits and altitudes are required to provide accurate model to be used in the system design and management. Also, spectrum sharing studies with Q and W bands will be needed. Antenna developments for space and ground segments are expected as described in earlier sections.

9.10.1 Expected Impact

The expected outcome of the spectrum related activities will benefit the whole wireless community by enabling more users to access limited spectrum resources and exploit higher frequency band for new services. Particularly this will lead securing proper spectrum resources for satellite communication missions. A flexible spectrum allocation approach, especially coordinated spectrum sharing for non-tightly integrated systems in the multi-layered architecture and sharing fostered by international and national administrations is foreseen.

Operators can use spectrum sharing to decrease costs and increase efficiency of the resource use. On the other hand, spectrum sharing will enable innovation and can be used in providing complementary services in the same spectrum. Licensed spectrum sharing system allows only licensed secondary users to use the spectrum. The model is a promising concept for both terrestrial and satellite bands aiming to ensure predictable QoS for all the holders of the spectrum use rights. It is foreseen that coordinated spectrum sharing approaches will enable new business models and promote collaboration between satellite and terrestrial players.

The new class of spectrum management procedures is expected to take into use latest advancements in radio access and management techniques of satellite and terrestrial domains. This will enable successful deployment of these techniques in practical systems to come.

9.11 Artificial Intelligence for SatCom

The NewSpace economy is primarily dominated by the necessity of substantially reducing the operational costs of satellite systems. This fact involves rethinking the entire set of space operations from both ground and space segment. Similarly, to other industries (e.g. automotive), satellite communication stakeholders are targeting a digitalization and automation of many functionalities currently managed by engineering experts. For instance, managing unintentional interferences constitutes one of the major time consuming and; thus, expensive roles of the satellite control centre [C9-5]. Reducing the time-to-react of interference events by automating its detection and classification represents a relevant cost reduction and quality-of-service increase to the satellite customers.

New satellite technologies will also require their proper automation with the aim of being competitive in the global internet access. In particular, the deployment and management of mega-constellations presents a challenging multiagent problem. These systems do not only present issues on the possible collision of different space elements, but also its radiofrequency emissions which may cause interference to other GEO or NGSO systems. In this context, automating system level decisions is a must since human intervention might be very difficult. Artificial intelligence allows fast decision-making even in millisecond timescales in order to meet full potential of integrated systems consisting of space-borne, air-borne and terrestrial components. This could include predictive spectrum allocations, predictive routing, and even autonomous replanning at the satellite without the delays introduced by the decision-making loops on ground.

The integration of satellite and terrestrial networks obviously calls for adequate resource allocation schemes, considering traffic variations across the entire network as well as the characteristics of the various links over which data delivery should be carried out. Given the high heterogeneity of networks in play, an efficient resource allocation concept must be implemented at the different layers of the protocol stack. In a first instance, the simultaneous use of different frequency bands as well as the opportunistic access of the same one will have to be properly coordinated by means of adequate cognitive systems. As such, efficient spectrum management will be of paramount importance to mitigate and where possible avoid cross-link interference. In turn, the capacity request from different vertical services will have to properly map on different network slices, to properly met QoS guarantees across the entire network. To this end, efficient joint resource allocation schemes will have to be developed, depending on the different traffic flavor and network condition variations over time. Therefore, the use of AI/ML concepts will play an important role, especially in the short-term prediction of operative network conditions, as it will be instrumental to efficient network selection and therefore the consequent resource allocation.

Considering the above discussion, we think that satellite communication industry is dealing with a short-term automation process of already launched systems and a medium/long term technology development of forthcoming systems based on GEO and NGSO satellites. Although some of the functionalities of the mentioned systems could be tackled by rule-based tools (e.g. if A then do B else do C), other mechanisms would require the use of machine learning techniques. In the following, we provide some examples of potential usages of machine learning in SatCom.

While spectrum sensing signal processing approaches have been studied and deployed during decades, certain interference detection aspects that appear in satellite operations would require additional approaches. For instance, in band interference or intra-system interference which consists of having the same signal duplicated in the same frequency bin. For these spectrum monitoring scenarios, the use of deep learning could potentially allow the detection of a variety of spectrum events [C9-6]. Once the interference is detected, satellites being launched this year will have the capability of reacting to this situation in order to meet the customer service level agreements. This involves to judiciously change the frequency allocation, transmit power and beamforming of the satellite payload. Although an operator could try to provide a new configuration based on the location of the undesired interference and the affected bands, an automated system able to configure the payload will reduce the time-to-react. This configuration update can be based on genetic algorithms [C9-7], [C9-8] with a potential support of deep learning regression techniques [C9-9].

Other promising applications of AI/ML in the SatCom field are those related to Physical Layer aspects where traditional signal processing algorithms can be enhanced, such as channel estimation in non-linear systems [C9-10] or signal classification [C9-11].

Finally, mega-constellation systems offering capacity to mobile users such as vessels and airplanes must tackle a two-level dynamic environment: the satellite and the user terminals. This varying system must be able to deal with potential hotspots beams (i.e. a coverage area which a high demand) that change over time and satellite. Although a classical data fusion between satellite and user terminals trajectories may help in detecting these hotspots, demands and motion present a stochastic nature which could be tackled via data-driven algorithms. In other words, the past detection of hotspots may support the inference of future ones.

Future satellites serving multiple, very different user applications may want to support AI/ML directly within the satellite and offer these AI as a service for end user applications. Recent advances in embedded AI accelerators promise significant performance and lower energy consumption so the integration into a flying edge cloud is possible. This can enable a multitude of new applications, i.e. pre-processing image data or meeting intelligent control decisions.

9.11.1 Expected Impact

As a general statement, the adoption of machine learning mechanisms by the SatCom industry aims to increase the efficiency of current systems by reducing the operational costs and the processing time of certain operations. At the same time, in the future systems it is envisioned that some functions, due to its complexity, will not rely on classical operations but on data-driven techniques. The digital transformation of the satellite industry will allow to discover new potential opportunities. Having an easy access to data sources will promote the creation of innovative SMEs and innovation departments in big companies able to provide data-based solutions to unexplored SatCom aspects.

9.12 Security

There is a potential for running large number of applications and services over a hybrid satellite and mobile networks. Therefore, it is not enough to just provide end-to-end security at the application layer (such as encrypted video stream at the application layer). In 3GPP 5G standards, new network security schemes are incorporated and as we look more towards integrated systems with satellite being used to roll out 5G security systems must operate over mixed delivery systems. There is a challenge here to involve the special characteristics of satellite systems. There is need to address security problems for the future hybrid satellite and mobile networks and the new services applications. In addition, more emphasis is being placed on interplanetary systems as we focus on lunar and then mars missions where again security in these environments needs to be visited.

There is a potential for new security topic in future satellite Communication such as:

Satellite and future mobile networks security integration:

An example of future mobile networks is the 3GPP 5G system. There is a need to address the security interworking between the two systems (satellite and future mobile networks). 3GPP defines two interfaces for native 5G and non 5G access networks:

- 5G native interface with 5G-AKA (Authentication and Key Agreement) procedures: This means tighter satellite integration into the 5G security system.
- Non-3GPP Interworking Unit (N3IWF): Satellite will use a recently defined EAP-AKA procedures. This implies looser security integration than 5G-AKA.

There is a need for detailed security analysis of the two scenarios above, supported by simulation/emulation or real-life demonstrators.

Blockchain Technology (BCT) for secure SatComs

Most of the previous security studies focused on key management authentications and routing protocols. BCT is a decentralized and distributed database invented by Satoshi Nakamoto in 2008 and used widely in digital currency. The main property of BCT is, it provides and authenticated record of history of changes such as configuration and re-configuration history of the satellite and space information network. Therefore, there is need to explore the advantages BCT in future satellite networks such as resilience to DoS, DDoS and insider attacks. Also BCT challenges should be examined as well, such as public and private BCT, the BCT database storage and distribution for all satellite nodes in the network.

Quantum Key Distribution (QKD) and key management over satellite

QKD is now a mature quantum communication protocol which allows the verifiably secure sharing of encryption keys between two communicating parties. There has been a series of proof-of-principle demonstrations of satellite-based quantum key distribution in 2017. There is a need now to expand the QKD key-management aspects such as access control, primary and secondary communication security key derivations and key lifecycle management.

Secure multicasting over satellite:

Multicasting is an important use case in satellite and future mobile networks such as using Multi-access Edge Computing (MEC) platform for Content Delivery Network (CDN) integration with efficient edge content delivery. Assuming that the current multicasting technical challenges in mobile networks are resolved such as MBMS, then secure multicasting should also be analyzed in detail, such as group key management and access control. Therefore, the secure group communications should be re-examined in the context of multicast key management architecture and security interworking in hybrid satellite and future mobile networks environment.

New mechanisms of security at RF level:

With the expected emergence of new SatCom architectures, supporting GSO and NGSO satellites combined with other aerial networks such as HAPS and UAVs, the need to protect and secure communications from threats such as unintentional interferences, intentional jamming attacks or signal spoofing will significantly increase. To countermeasure these threats, new mechanisms of security at radio level should be examined, such as the combination of advanced SIGINT (Signal Intelligence), AI (Artificial Intelligence) techniques and location of interference or attack sources.

Secure multiple access

Multiple access techniques are becoming more and more sophisticated in order to achieve higher spectral efficiencies. However, cyber security threats to multiple access systems can still be very effective regardless of the access technique, including jamming, interruption or endangering of the overall satellite communication system. Hence, security enhancements or protocols are needed at the lower layers of return channels.

9.12.1 Expected Impact

Good security solutions will be essential and a strong enabler for the integration of satellites for the new 5G applications such as smart cities Intelligent Transport Systems, eHealth, smart content distribution and Industry 4.0 (including Industrial IoT, IIoT) as well as extensions of space networks to the planets.

9.13 Communication, Computation and Storage

In the future NTN architecture, the flying nodes can be given different functions/roles facilitating control and operability and opening new business models.

A fundamental functional distinction in upcoming applications and services is communication and computation (e.g. as required for AI and/or in-orbit cloud services), both requiring storage at different levels. With this distinction of communication, computation and storage, it is possible to foresee different orbital planes/layers of nodes providing services to other planes/layers of nodes. As an example, while data gathering is clearly more efficiently performed by LEOs, the large amount of required data storage can be serviced by a GEO layer, which can both perform computational services on the data and/or download to Earth also more efficiently.

A structured dense space network enabling efficient services among different orbital planes/layers of nodes requires innovation and novel research topics.

An important consequence of different orbital planes/layers of nodes providing services to other planes/layers of nodes is that links need to be tailored to the functional roles. For example, links for data transfer when approaching the storage and/or computation nodes should be adaptive with performance targets different to those for communication.

Coding for dynamic multi-source transmission

It is clear that the large amount of data that will need to be handled will require not only optimized adaptive links but also novel concepts of transmission and coding. Specifically, energy-efficiency will be essential for achievability of competitive costs per bit/channel use. Hence, simultaneous transmission from different satellites to another one needs to be optimized for example using coding across sources taking advantage of geometrical properties.

Coding for secure distributed storage dynamic systems

While several fundamental results already exist for secure distributed storage in terrestrial systems, these are not directly applicable to non-terrestrial systems due to the orbital dynamics and diversity of the satellite hardware limitations. However (efficient) secure distributed storage is essential in space due to energy, hardware limitations and proximity of satellites in a dense space network.

Peer-to-peer protocols for data access

Data storage and/or computing may take place at different orbital layer, hence, access protocols to stored data at any time need to be developed.

Orbital topologies and reprogrammable satellites for computation and storage

Current orbit and constellation topologies are predominantly optimized for communication with satellites with communication payload for typical channels ground-to-satellite, satellite-to-ground or inter-satellite. However, new orbital geometries and re-programmable satellites will facilitate secure and efficient in-orbit cloud services for storage and/or computation.

9.13.1 Expected Impact

- Substantial reduction of cost per bit/channel use due to the correct distinction between in-orbit resources and roles for communication, computation and storage.
- Further structure in the non-terrestrial network, which allows better control and more agile design of new services.
- Facilitates the functional-based re-programmable design of space nodes considering the communication, computation or storage services that the node can obtain from other nodes.

- Facilitates new business models where in-orbit nodes (e.g. data gathering or security) are serviced by other in-orbit nodes (e.g. storage nodes collect the data gathered or offer Blockchain Technology (BCT) nodes).

9.14 Plug and Play Integrated Satellite and Terrestrial Networks

The role of NTN systems in future networks will be drastically changed with respect to the 5G and IoT sectors and plug and play satellite integration into 5G and beyond is the core of this fundamental change. It is therefore crucial to foster the development of an attractive plug-and-play satellite solution for 5G and beyond. That enables terrestrial operators and network vendors to accelerate 5G deployment and creates new and growing market opportunities for the satellite industry. Thus, significant efforts are required to:

- Design SatCom solutions, targeting integrated satellite / terrestrial 5G architectures adopting and integrating 5G key features;
- Exploit SatCom capabilities (e.g. broadcast, ubiquity and reliability) while mitigating its inherent constraints (e.g. propagation latency) in standalone or multi-link network topology;
- Ensure seamless integration of SatCom at orchestration and core levels;
- Foster satellite inclusion in the 5G ecosystem as a key access network technology, to fulfil 5G implementation in the society.

Innovation of technologies for communication systems, specifically in satellite domain, cloud technologies and 5G terrestrial systems is reaching a point of convergence, which promises a new communication paradigm. It enables a new range of features such as agile service provisioning, multi- tenancy, software controlled and dynamic management, on- demand service-oriented resource allocation, universal multi- access, and ubiquitous connectivity. Standardization bodies as 3GPP and ETSI recognize and promote terrestrial and satellite interworking. The complete integration can be achieved with the combination of radio networks, including core and access, and the satellite systems, with computational resources expanded from the core to the network's edge. On this way several challenges have to be addresses, such as:

<u>5G Core Enhancement</u>

The 5G core and edge networking enhancements to support satellite and terrestrial access via the multi-layered terrestrial and non-terrestrial architecture. In particular, capability of discontinuous operation of the satellite RAN and the 5G core network over satellite systems. This includes the definition of new splits between the RAN and core network functions installed in the satellite payload and in the ground segment and the design of delay tolerant transport protocols that enable discontinuous connectivity of the feeder link. It enables the provisioning of NB-IoT services on LEO constellations using regenerative satellite architectures. The main challenge to address is the capability of completing 3GPP procedures even when a LEO satellite does not have simultaneous coverage with the sensor and the ground station.

<u>Satellite Resource Management and Service level APIs</u>

To support the new dynamic satellite use cases, driven by 5G and other verticals, the satellite ground segment should push more for the resource softwarization and introduces efficient satellite resource and service manager elements. It will help to make the ground segment more dynamically configurable by external third-party resource and service orchestration systems. It promotes features like unified management and control with vendor-neutral interfaces; CAPEX and OPEX reduction; multi-tenancy; reduction of time-to-market for new appliances and protocols; promotion of market openness; as well as radical enhancement of the telco service portfolio with novel network services based on virtual appliances.

Cross-domain Coordination Layer

As a key technical enabler for the realization of the smart networking using the next generation integrated satellite and terrestrial systems is Cross-domain Coordination layer. Such solution provides end-to-end service federation by orchestrating in a coherent fashion the satellite domain and third-party domains such as terrestrial 5G networks or Edge/Cloud domains. This solution should enable agile service provisioning (in order of few minutes) as well as the optimization of network capabilities, performance and behavior to support in the best way the dynamic nature of the applications and services that are required at any given time by the users. This technology should also allow the network to serve more users and more applications given specific infrastructure and thus provide more revenues for the service providers.

9.14.1 Expected Impact

Novel business models and economically viable operational collaborations that integrate the satellite and terrestrial stakeholders in a win-win situation can be only possible if effective technical solution for Plug and Play integrated satellite and terrestrial networks become a reality. The impact of this step forward for the European space industry is the provisioning of a globally competitive and ubiquitous 5G network in matter of minutes, offering the end user at least 50 Mbps 5G broadband and optimizing satellite communications within the 5G network infrastructure to provide services in the following applications: media and entertainment, transportation, health, logistics and agriculture.

Development of a cost-effective and efficient satellite-enabled eMBB solution for 5G and beyond to allow telecom and network operators to accelerate the deployment of 5G across all geographies while creating new and growing market opportunities to position Europe as a key player in the global competitive market of satellite and 5G. It is one of the main goals of the New Space environment, where can be also expected to NB-IoT services defined within 3GPP.

10. Opportunities for Devices and Components

Progresses in all aspects of the wireless network are highly dependent on electronic technologies, components and devices that are used for implementation. This chapter gives an overview of the expectations towards the component and device researchers, designers and manufacturers to support the requirements of wireless networks up to the end of this decade. This includes the whole range of components such as processors, memories, analog, RF, converters, antennas, packaging and optical components.

10.1 Sub-10GHz RF

The market of sub-10GHz has been dominated by cellular networks based on 3GPP standardized radio access and Wi-Fi local area access of 802.11 family. Mobile connectivity has utilized even larger portions of the spectrum at frequencies up to 6GHz. LTE (4G) and advanced Wi-Fi features will be complemented by ISM band applications from Bluetooth to home automation, NFC and IoT with narrowed spectrum allocations as well as satellite-based positioning systems.

The trend has been and will be for a more efficient use of spectrum at the range that is the most suitable for compact and highly integrated electronics, i.e. RFICs with efficient DSP in terms of form factor, cost and power consumption in battery operated devices. Although technologies for RFICs and other components may sound mature and speed of any transistor is not a bottleneck, complexity of the designs has become enormous and on the other hand new data intensive applications require enhanced broadband operation. In addition, limited and scattered spectrum availability will lead to increasing parallelism of signal paths from antennas through RF to DSP. Both carrier aggregation to enhance data rate over several bands and massive MIMO to use spectrum more efficiently at those bands multiply the parallel active RF signal paths.

The challenges are obvious but hard to tackle including increased power consumption and interference between simultaneously operating radio paths (co-existence). For those, the solutions cannot be overcome solely by bulky filter banks at the front-ends but require increasing signal purity at transmitters and improving linearity and internal filtering capability in the receivers. This cannot be simply solved by increasing digital content due to bottlenecks in digital conversion. Even if higher resolution ADCs and DACs can shift part of the processing to digital that is in many ways highly beneficial, the new requirements mandate similar or even faster development of RF circuitry including antennas, external filters and switches as well as RFICs to achieve the goals.

Densifying networks is a must also at lower frequencies and not only starting from mmW region due to better range and frequency utilization including coming cell-free MIMO approaches. Co-optimization from antennas to digital over different technologies and techniques is a core competence in this field in addition to squeeze out the best possible performance from the technology. As RF integration is always balancing between speed of the transistor for digital and optimal performance of RF for power amplifiers, highly linear receivers and the best possible RF filtering contradictory requirements determine case-by-case the system partitioning of the functionality to components and IC’s. Large SoC’s and multi-chip RFIC and modem combos have their specific purposes also in the future.

To name the key opportunities in components, programmability and flexibility even beyond well-established topologies is a must. That is not anymore only cleverly placing digital switches inside RFICs, but also techniques that can better jointly optimize antenna-filter-RFIC combination in terms of performance and flexible spectrum use. Holistic view on the system performance gives still many opportunities to boost system performance and minimize cost. Digital content approaches are also needed in classical RF signal processing blocks including digital PLL's, transmitters and to certain extend also receivers with minimal RF content keeping in mind that dynamic range is the key to solve any near-far problem especially in cellular transceivers and co-existence scenarios between systems. In addition, solutions for simultaneous transmission and reception in the same channel i.e. in-band full duplex are still lacking proper commercialization even for a single band. Multi-band and MIMO pose huge challenges to in-band full duplex. Similar challenges apply both for RF transceivers in mobile solutions as well as in infrastructure with different trade-off in required performance vs battery lifetime.

Finally, opportunities below 10GHz are not only limited to more efficient use of spectrum but serving different kind of applications from narrow-band IoT to radar. These two are among examples that set very specific requirements for the circuits. In some cases, they can be seen as individual problems for specific devices like temperature sensors or hart beat monitoring of elderly people. However, the opportunity to utilize the same, extremely programmable circuitry to achieve multiple goals could enable a new set of new devices. The search for optimal combinations or to design more optimal circuits to serve different combinations is an emerging challenge and opportunity for various wireless systems effectively utilizing spectrum and hardware at frequencies below 10GHz.

10.2 Millimeter-wave and TeraHertz

10.2.1 THz Communications:

With an important amount of unused spectrum, the sub-TeraHertz (sub-THz) frequencies between 90 and 300 GHz are potential candidates to achieve high-data rate wireless and wireline communications and hence to fulfil the requirements of the next-generation of wireless networks and wireline (e.g. data center) networks [C10-1], [C10-2]. THz signals may also be carried over low cost waveguide structures such as polymer microwave fibre (PMF).

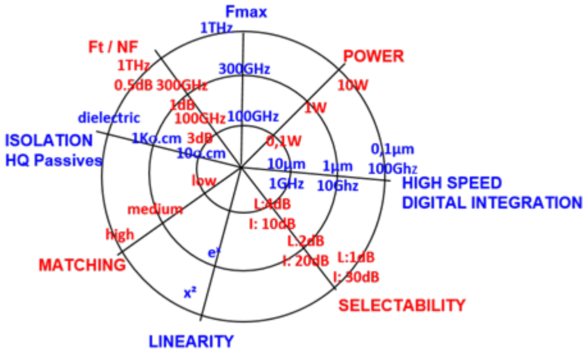
However, before making the deployment of system in the sub-THz band, many challenges need to be addressed. First, the free-space propagation losses increase with the square of the frequency. These losses have to be compensated by using high-gain antennas, which entails severe constraints on antenna directivity and alignment. The design of electronically controlled steerable beam remains today an open issue. In addition, sub-THz systems could suffer from strong phase impairments due to the poor performance of high-frequency oscillators [C10-3]. Therefore, the study of new digital transmission schemes optimized to mitigate the impact of RF impairments such as phase noise (PN), or strong group delay distortion and polarization rotation in PMF, are essential to guarantee good performance [C10-4].

10.2.2 Solid-state technologies for THz applications:

Nowadays, silicon-based technologies offer low-cost solutions for RF and millimetre-wave applications combined with a high complexity in the digital domain. CMOS, however, has its limits in speed and power generation, which become apparent at operation above 100 GHz. This is evidenced in the on-line survey of power amplifiers, maintained by the Georgia Tech

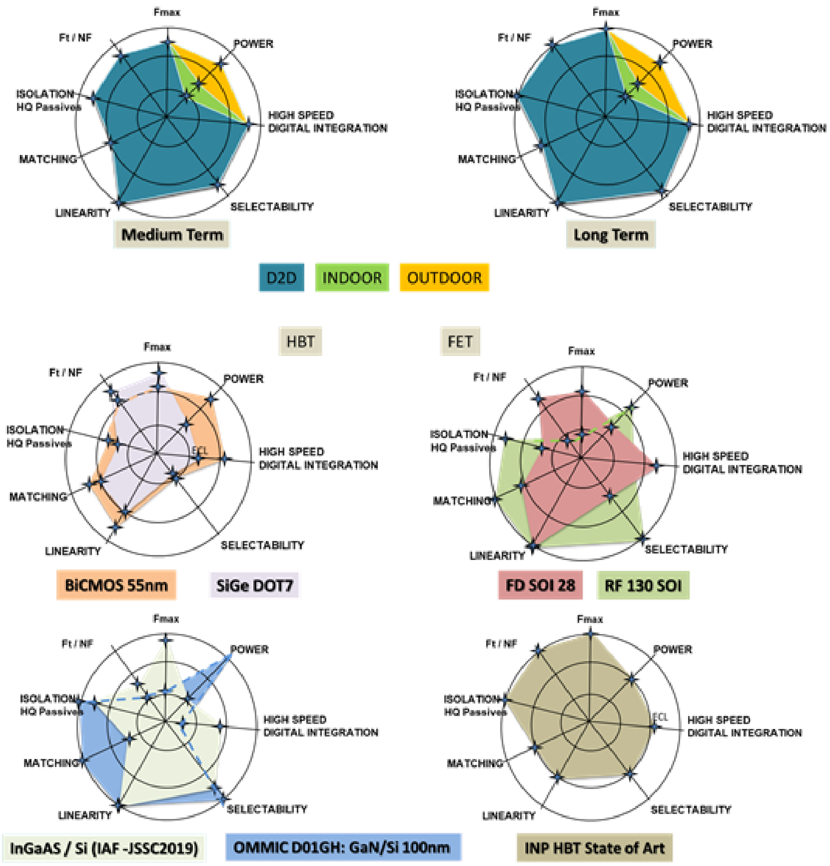
University [C10-9]. Hence, the very high-speed part of a THz communication transceiver will need a different technology. Application of this “non-CMOS technology” will be limited to the transceiver functions that cannot be implemented in CMOS: transceiver architectures will be designed such that most functionality can be done in CMOS, keeping the non-CMOS part as limited as possible.

In the H2020 NEREID CSA project [C10-7], IC technologies are compared based on wireless transceivers and their most critical/representative RF building blocks: a power amplifier (PA), a low noise amplifier (LNA), and a voltage-controlled oscillator (VCO). Based on the performance of these circuits a spider diagram can be constructed (see figure below), which uses eight criteria:



The **Power** capability depends on the maximum bias voltages that can be applied over a transistor and the maximum current capability. The **High-Speed Digital Integration** capability is a function of inverter size and efficiency (transit time / current). Here, again, a disclaimer needs to be made that digital functionality will be adapted to enable CMOS integration, even if other technologies could offer better digital circuit performance. Next, the **selectability** is the ability to switch RF signals with high isolation (low loss, high isolation). **The linearity** of transistors is given by their input-output relationship (e.g. x^2 for a quadratic device or e^x for a bipolar transistor). Further, **matching** between 2 minimum-size transistors is an important criterion in differential circuits which are widely used, from low frequencies up to mm-wave frequencies.

Isolation and HQ Passives depend on the substrate resistivity and on the availability of a thick metal. This gives the ability to limit the effect of pulling of a VCO by a PA, and to have high-Q passives in dielectric conditions. The couple **f_T/NF** is the property given by the cutoff frequency f_T and the minimum noise figure NF_{min} of the technology. The cutoff frequency indicates the potential for high-speed digital applications and NF_{min} for the lowest achievable receiver noise. Further, **f_{max}** is the frequency where power gain has dropped to 0 dB. Typically, the maximum frequency of operation for RF amplifiers (which need to able to make a minimum amount of power gain) is limited to about $f_{max}/3$.



A connectivity roadmap has been defined in [C10-7] for three main application domains, namely D2D (device to device communication), indoor and outdoor applications. An outlook is given for the medium term (5 years outlook) and for the long term (10 years). Based on this analysis, the connectivity requirements for each process technology can be extracted. This is illustrated in the next figure. The main difference between the technologies is the output power. At low mm-wave frequencies, GaN is the champion here. However, in outdoor applications and also in most indoor ones, beamforming will be used. Then the required transmit power per PA is drastically lowered compared to one single antenna path, making many technologies suitable for the entire front-end including the transmit part. Further, for some D2D applications the linearity constraint can be relaxed. Whatever the application, increasing the operating frequency will impose strong specifications on f_{max} and NF_{min}, while increasing the data rate will require a higher f_T. NEREID’s forecast for these 3 device parameters is 1THz for both f_{max} and f_T and 0.5dB for NF_{min}.

In the spider diagrams we consider different (non-CMOS) technologies that are available at this time of writing: silicon-based ones (RF-SOI, FD-SOI, SiGe BiCMOS,,) III-V on silicon substrates (GaAs/Si, GaN/Si) and III-V on native substrates (InP). The benefit of BiCMOS and FD-SOI is the ability to combine mm-wave circuits with complex digital circuits, although to a lesser extent than ultra-downscaled CMOS beyond the 10nm generation. RF-SOI and new GaN/Si bring RF power and selectability. For generation of power in the D-band and in higher frequency bands, the survey of [C10-10] indicates that the best performance is obtained nowadays with InP. A deployment of InP on a very large scale is hindered today by the small number of metal levels that are typically available in commercial foundries and by the small wafer sizes. Integration of high-mobility III-V devices on 300mm silicon wafers as in [C10-13] and going further to co-integration with CMOS is currently investigated. Another route is a further evolution of silicon bipolar transistors, for which cutoff frequencies above 1 THz are predicted [C10-14]. BiCMOS has the advantage of being a silicon technology, with a larger ecosystem than e.g. InP but still, the product of mobility and breakdown voltage of III-V devices is higher. Finally, GaN HEMT devices are also subject to downscaling and might become a strong candidate for D-band operation when gate lengths well below 100 nm can be realized. Operation at 100 GHz with 100nm GaN on SiC devices with gold metallization has already been demonstrated but still with moderate efficiencies for a PA [C10-15]. Also here, a wide deployment of GaN devices might benefit from integration on 300 mm wafers and with many Cu metal levels as in [C10-15].

Finally, THz communication will use very wide bandwidths to accommodate high data rates. As a result, the bandwidth that needs to be handled by baseband (both analog and digital) circuitry is huge, compared to the early days of wireless communication. This is a challenge for the design of analog-to-digital converters. There will be a need for ADCs with clock rates beyond 10 GHz. The ADC typically resides on the same chip as the digital functionality of a transceiver, which, for mass-market applications, is expected to further follow the CMOS downscaling trend. It still remains an open question how the performance of extremely high-speed ADCs will evolve when logic devices will transition from a finFET device architecture to a gate-all-around structure or even forksheet devices [C10-16].

10.2.3 Passive THz Imaging:

THZ Imaging state of the art (SoA) reports two main competing categories of 2D-array image sensors:

1. The above-IC bolometer-based THz image sensors based on a classical infrared (IR) sensor that offer a high sensitivity and currently a good maturity [C10-8]. but, using two different circuits, it is an expensive solution.
2. The monolithic CMOS-based THz imagers have recently emerged as low-cost competitors [C10-9, 10]. Even with their current poor sensitivity (1000 times less than bolometer-based sensor), these CMOS-based THz image sensors have proven to be a viable cost-effective alternative to bolometer-based imagers.

Passive THz Imaging has applications in digital health technologies, passive, continuous, home-based monitoring of biochemical markers in biofluids, such as sweat, tears, saliva, peripheral blood and interstitial fluid.

10.2.4 Active mm-wave and THz radar imaging:

Active radar imaging makes it possible to add the range and even Doppler dimensions to the image (3D or 4D imaging). On the lower edge of the spectrum, in the mm-wave and low THz bands, radar imaging is evolving fast to satisfy the requirements of ADAS and autonomous driving. The trend there is to resort to MIMO techniques whereby a virtual antenna array is created with a size equal to the product of the number of TX and RX antennas. 79GHz radar imaging with wide field-of-view, resolutions of 1 degree by 1 degree and cm-scale range resolution is experimentally feasible today and radars with wide field-of-view and LiDAR-like resolutions is an active field of applied research. Using higher carrier frequencies such as 140 or 300 GHz is a longer-term trend, resulting in smaller form factor or better angular resolution as well as better range resolution, thanks to the wider bandwidths. Some experimental radar chips show already the potential of CMOS in the low-THz regime (140 GHz) [C10-9]. These highly miniaturized radars will enable new applications, such as intruder detection, gesture and activity recognition, and heart rate and respiration rate monitoring, among many others.

10.3 Ultra-low Power Wireless

It is expected the number of IoT nodes will continue to grow to 100 billion by 2030, and ultra-low power wireless connectivity will be the key enabler. However, the existing wireless connectivity has limitation to support such large number of nodes. For example, battery replacement of billions of sensor nodes will not be feasible. Additionally, the trend is to spatial awareness to the IoT end-nodes using the front-end already used for radio communication. Using the sensing capabilities, channel state information can be collected. This allows for new types of applications like presence detection or localization. These functions will pose additional requirements on the radio front-end and thus the design choices. To further scale up the number of IoT nodes, several important challenges need to be overcome.

10.3.1 Battery-free operation

Batteries will be the primary limitation of IoT nodes. Manually battery replacement of 100's of billions IoT devices will be too expensive, and the disposed batteries will be a serious environmental issue. To support a sustainable growth of the IoT devices, battery-free operation will be a key solution.

Most of the existing battery-free wireless communications adopt simple modulations (e.g., backscattering) and protocols. However, they will not be able to scale up to network with large number of nodes. One potential solution as demonstrated in [C10-17] is to adopt a “back-channel compatible” wake-up receiver which monitors the energy profile of the signals sending from the central hub. This wake-up receiver consumes very low power consumption, so it is compatible with battery-free operation. It only activates the main transceiver for sending sensor data only if certain energy profile is detected.

10.3.2 Spatial Awareness

For spatial radios, we can differentiate between active and device-free localization. In active localization, two (or more) IoT nodes are ability to measure the distance between them, using channel state information. For device free localization, time variation of the channel state information is used to detect changes in the propagation environment, e.g. due to human movement [C10-18].

Currently, channel state information is often based on received signal strength (RSS), mostly because it is easy to implement. However, in multipath fading environments and increasing distance the accuracy is rather poor. To improve robustness against multipath fading, it is well-known that a large radio signal bandwidth is required. This will increase spatial resolution, beneficial to both active and device-free localization.

Using the concept of phase-based ranging [C10-19], a wideband view on the radio channel can obtained. By sounding the radio channel in a sequential manner over individual narrowband channels using half-duplex bi-direction signals, a wideband measurement of the radio channel is realized. For each individual measurement, only narrowband signals are used, making it suitable for radio front-end used for e.g. Zigbee of Bluetooth [C10-20], but also Wi-Fi [C10-21].

Aside a modification of the radio protocol to incorporate such measurements, also the frond-ends will be impacted, most considerably the Local Oscillator (LO) Generation/Phase Locked Loop. For accurate distance measurements, also the phase of the radio channel should be measured. This means that the generated LO should be continuously, when switching from

TX to RX and vice-versa and from channel to channel. This leads to a whole new set of requirements and challenges for PLL design [C10-22].

10.3.3 Degradable Devices

One alarming trend in IoT, is the increasing number of disposable devices that are designed to fail and become e-waste once they run out of battery [C10-52]. To solve this problem, we need energy autonomous devices that uses ultra-low power (ULP) radios and harvest the energy they need. However, in order to eliminate the e-waste problem, research is also needed to develop ULP radios that could be manufactured by printing using biodegradable substrates and renewable materials. This starts to emerge in the RFID domain (e.g. [C10-53]) but would have to be adopted more widely to the Internet of Everything applications in order to avoid environmental problems.

10.4 Antenna and Packages

10.4.1 On-chip antennas, lens-integrated antennas, antenna MIMO arrays

Packaging of mmWave/THz chips for low-cost consumer electronic applications requires low-cost packaging solutions. Conventional low-cost silicon packaging technologies, however, exhibit a typical 1nH/mm lead and wirebond inductances, which are prohibitively high at mmWave/THz frequencies and plastic packaging materials and encapsulants are quite lossy. In addition, even expensive high-performance coaxial cables and connectors have significant losses at mmWave/THz frequencies. As a result of this, future THz packaging technologies must avoid interconnects as much as possible and antennas need to be integrated into the chip package or even on chip. Fortunately, the radiator size scales down at higher frequency and this makes compact and integrated antenna solutions feasible. However, the free-space propagation loss at THz frequencies becomes very high (80 dB for 1 m at 240 GHz) and this loss needs to be compensated with an appropriately high directivity of the antenna system in order to provide sufficient link budget. Due to their large silicon area, however, high directivity antenna arrays are costly on chip. Future solutions, therefore, include alternative lens on-chip assemblies which exhibit a better cost performance ratio.

On-chip antennas: For efficient and low-cost THz signal escape from the chip level, appropriate on-chip antenna systems need to be developed. On-chip antennas embedded in the BEOL stack of a lossy silicon chip [C10-23] are very challenging because of potential multi-mode propagation issues (e.g. surface waves) within the volume of an electrically large and thick silicon die leading to very inefficient radiation with very poor control of radiation patterns. Because of very high carrier frequencies with large fractional RF bandwidth, standard design techniques relying on narrowband matching become less efficient and will result in limited circuit performance. Depending on the application, antennas should support very wide operation bandwidth with minimum group delay distortion for high-speed modulation and stable phase characteristics for precise location of the focal point position in an imaging system across the bandwidth of interest. Furthermore, for sufficiently high frequencies, classical buffering circuits become unfeasible and true antenna-circuit co-design at multiple harmonics simultaneously is necessary for high-fidelity system operation.

Lens-integrated antennas (chip-on-lens assembly): Further research is required on new ultra-wideband silicon lens-coupled antenna system allowing efficient coupling of THz radiation into the intrinsic device without classical matching structures. Antenna may further provide dual-polarization functionality with two transmitter/receiver paths connected to each orthogonal polarization.

MIMO arrays: Highly directive terahertz antennas can minimize interference between adjacent channels, and frequencies can be reused more frequently, thereby improving spectral efficiency and signal quality. This enables higher channel isolation in a MIMO (Multiple Input Multiple Output) network topology. At THz future MIMO networks could reach data rates of up to one Tbit/s easily. Future MIMO arrays need to support not only faster links, but also real-time operation by rapid channel switching and/or beam-steering/tracking at a very low latency.

10.4.2 Metamaterials and metasurfaces

The development of metamaterials (MM) is another promising technology for beyond-5G networks and services scenarios: one remarkable use case, for instance, is the exploitation smart radio environments (both indoor and outdoor) with ultra-massive MIMO and Artificial Intelligence (AI).

MM are materials which contain inclusions (e.g., metallic or dielectric of various shapes and dimensions) designed and engineered to manipulate electromagnetic (EM) waves. Examples of inclusions embedded into a host metamaterial include EM scattering element and nano-resonators. These properties, for instance, could be used for developing smart antenna and EM processing functions, including methods for AI.

Metasurfaces (MS) are 2D metamaterials (MM). By modifying the structure and spatial distribution of those sub-wavelength reconfigurable passive elements in the metasurface, the electromagnetic characteristics of the elements can be changed and independent phase shifts are added by different MS elements to incident signals without using any power amplifier, complicated coding and RF processing. In this way, passive reflection, passive absorption, passive scattering can be realized which may even not exist in Nature (e.g., zero or negative refraction) [C10-24], allowing a wide range of EM processing functions and pushing the physical environment to change towards intelligent and interactive.

MS can be seen as arrays of nano-antennas: by shifting the resonant frequency, through the nanoantenna designs, it is possible to effectively control the amount of the phase shifted in the scattered signal. [C10-26] describes a prototype of an information metasurface controlled by a field-programmable gate array, which implements the harmonic beam steering via an optimized space-time coding sequence.

The main advantages of MS include: completely passive and low power consumption, supporting free-duplex and full-band transmission, requiring no high-cost components, being able to be deployed densely and expandable and reconstructing electromagnetic waves at any point on its continuous surface.

It is expected that MS will have several possible applications, such as:

i) radio coverage in areas not well covered by installation of base stations, and face NLoS limitation [C10-51
ii) smart radio environments (indoor and outdoor): being combined with AI, IoT and edge computing to enhance performance in smart cities, smart homes, health monitoring, and safety inspection,
iii) to serve cell edge users, relief multi-cell co-channel interference, expand coverage, reduce electromagnetic pollution, and implement dynamic mobile user tracking,
iv) automotive applications, vehicle and air networks and high-speed railway scenarios
v) running quantum algorithms directly with EM waves or in optics (e.g., in transformation optics [C10-25], quantum radio-optics devices, ultra-fast switching devices, detecting and recognizing images, holographic applications, etc.)

Furthermore, the possibility of coating surfaces in building or kiosks with intelligent (AI-based) MS will allow creating smart radio environments capable of radio waves propagations by introducing, in a software-controlled way, localized and location-dependent gradient phase shifts onto the signals impinging upon the MS. As a brand-new material, MS can be combined with antenna technology, massive MIMO, millimeter wave, terahertz communication, D2D and other technologies to form a controllable smart radio environment.

10.5 High-speed Transceivers, Wireline and Optical

10.5.1 Radio-over-fibre communication, sub-systems and components for B5G and 6G networks

In the near future, we can expect significant overhaul of the mobile network, targeting the use of mmWave frequency bands to deliver much higher capacities over the air. Mobile networks will use advanced radio transmission concepts such as coordinated beamforming, coordinated multipoint and massive MIMO (multiple input, multiple out) as well as pico-, femto- and even attocells. It has been long recognized that it is better to centralize (C-RAN, centralized radio access network) the digital signal processing (DSP) required for modulation and demodulation of the RF carriers. Advanced signal processing is now centralized in the baseband units. It is expected that for future mmWave networks, this fronthauling (bringing the data from the antennas to the centralized or cloudified baseband units) will be done through optical fibre given the high capacity and/or frequency of the signals that need to be transported.

While today this fronthauling is built upon standards such as CPRI (common public radio interface) or OBSAI (open base station architecture initiative) in which the digitized IQ samples themselves are transported, for future mobile networks the amount of traffic that will need to be transported will explode. For example, assuming 2GHz modulation bandwidth, 4 carriers, 3 sectors each with 32 antennas, digitization at 8bits, 8B/10B encoding and 10% overhead for control messages, then a total sustained throughput of 25Tb/s will be required in the fronthaul link. To overcome this problem alternative fronthauling techniques will be required:

- Analog radio-over-fibre, in which the RF signals are directly modulated onto optical carriers. This will require the development of highly linear optical modulators, which today form the biggest hurdle in the deployment of analog radio-over-fibre systems for mobile network applications.
- More efficient digitization of the RF signals as opposed to directly transporting IQ samples: one example is sigma-delta modulation in which the RF carrier is oversampled and the resulting digital signal is transported over the fibre using conventional low-cost optics (as opposed to likely more expensive analog radio-over-fibre) [C10-27].

10.5.2 Terabaud capable opto-electronic transceivers

Over the past decade we have witnessed the rapid emergence of large-scale data centers which now underpin most cloud-based software applications. The adoption of AI and machine learning is even further accelerating the deployment of data center infrastructure. The rise of these huge data centers and growing capacity demands in telecommunications are driving the following trends in the industry:

- **Need for new generations of optical transceivers with ever higher capacity**, within approximately the same physical footprint and power envelope as the current 400G generation (e.g., CFP8, 40x9.5x102mm, 24 Watts and 15 Watts for even smaller QSPD-DD and OSFP form factors). Industry is now seriously considering 800G and 1.6T capable transceivers. If this trend continues, even 3.2T or 6.4T generations may be required in the next decade.

- **Deployment of optical links at ever shorter distances:** in the past 5 years, we have witnessed how optical links have now displaced the copper-based interconnect at distances even as short as a few meters. This so-called intra data center optical interconnect (short-reach optical interconnect) covers distances of up to e.g. 100m (OM4 multimode fibre) and up to 2km (single-mode fibre).
- **More pervasive use of coherent transceiver technologies:** Originally set out to increase capacity in long-haul and subsea networks, coherent transceiver technologies are now becoming commonplace in the metro domain also. New developments will continue to increase data rates in these applications but will also focus on adaptations to make technologies fit metro-DCI, access and aggregation networks. Even for future use inside data centers, coherent transceiver technologies are being debated.
- **Advent of electro-photonic Systems-in-Package and co-packaged optics:** High-capacity switch ICs double their capacities every 2-3 years and processors and FPGAs rapidly scale up in capacity as well. Optical transceiver chiplets are expected to replace electrical I/Os in these applications in the next 5-10 years and will need to be closely integrated with the CMOS data processing blocks inside the same package to minimize I/O power consumption and maximize throughput.

This need for optical transceivers with still exponential growing capacity is driving following trends:

- The most effective way to increase capacity is to **increase the channel symbol rates**. Today, for short-reach applications typical symbol rates are 53Gbaud, which are then combined into 4-channel modules to create a 400G optical transceiver (full duplex). It is anticipated that within the next 2 to 5 years there will be a need for >100Gbaud symbol rates, to create 800G and 1.6T transceivers.
- The second way to increase capacity is to increase the number of parallel lanes within a cable. Lanes can be wavelengths or parallel fibres within a ribbon cable.
- To more effectively use the available analog bandwidth of optics and front-end electronics, **modulation formats with higher spectral efficiency** are being used also for short-reach applications (where the use of 4-level pulse amplitude modulation is now widespread). It is expected that this trend to even more complex, i.e. coherent-based modulation formats (already widely used for longer reaches) will continue.
- Denser integration, through the use of e.g. 3D integration methods, to fit the transceivers into small form factor modules.

10.5.2.1 Broadband signal generation and detection beyond CMOS

The need for optical transceivers at higher symbol rates will soon hit the limitations of the most advanced CMOS nodes. Indeed, the transition frequency fT (measure of speed of a CMOS process) has saturated around 300GHz for the last five years. Hence novel transceiver architectures will be required in which functionality is shifted from the CMOS ASIC to chips (or chiplets) realized in processes that have higher speed capabilities such as SiGe BiCMOS or InP. First such devices include >100GBaud capable analog multiplexers and demultiplexers recently demonstrated by NTT, Nokia and imec [C10-28]. Such heterogeneous approaches can be expected to further gain importance as symbol rates scale well beyond the 100Gbaud region in the near future.

10.5.2.2 Ultra broadband optical modulators and detectors

Today, the dominant processes for densely integrated optics use InP and Silicon Photonic technologies. To scale well beyond 100Gbaud novel materials will be required. A range of such materials that can be envisioned are:

- Organic Hybrid, with the most promising approach using plasmonic organic hybrid materials, with modulation bandwidths as high as 500GHz demonstrated [C10-29],
- Ferroelectric materials such as BTO, which hold the promise of integration onto a Silicon Photonic platform of pure phase modulators with good modulation efficiency [C10-30],
- Lithium Niobate (LNB) under the form of thin-film LiNbO3 integrated onto Silicon Photonic platforms, offering again a route to broadband (>100GHz) modulators without residual amplitude modulation [C10-31].

10.5.2.3 Monolithically integrated optics and electronics

For future developments, monolithic integration of photonic and electronic components would be desirable as this significantly reduces parasitics and maximizes overall system efficiency. Such a monolithic integration can be achieved in several ways. Some examples demonstrated so far are:

- Integration of plasmonic organic hybrid modulators on a metal layer on an electronics ASIC has been demonstrated [C10-32], this kind of technology is expected to achieve highest symbol rates per modulator.
- Monolithic integration of silicon modulators and detectors in silicon ASICs as has been demonstrated here [C10-33]. Through massively parallel implementations, Multi-terabit/s data rates are expected per optical waveguide.

10.5.2.4 Optically assisted analog-to-digital and digital-to-analog conversion

Scaling far beyond 100Gbaud may require further shifting of functionality currently performed in the electronic (digital) domain to the optical domain. A first example already demonstrated for some time is the generation of 4-level pulse amplitude modulation (PAM-4) in the optical domain using segmented Mach-Zehnder modulators [C10-34]. This simplifies the electronics (now requiring only simple binary driver electronics) and shifts the complexity of generating high symbol rate multilevel signals to the optical domain. It has been shown that such approach can result in better transmission performance (better error rate at same optical modulation amplitude) compared to conventional solutions [C10-34].

At the receiver side, the need for ever higher conversion rates will eventually be limited by the jitter of electronic oscillators and the resolution achievable at such high rates. One route to overcome this is the use of optical assisted analog-to-digital conversion, in which the low phase jitter of e.g. mode locked fibre lasers can be used to improve the achievable effective number of bits (ENOB) compared to conventional electronic ADCs [C10-35].

10.5.3 Ultra low-cost and low-power coherent “lite” transceivers

Scaling to higher symbol rates will at some stage require a transition to the use of coherent optical transceivers at a given fibre length. Indeed, higher symbol rates typically go hand in hand with reduced optical budget (due to worsening signal-to-noise ratio), which can be alleviated through the use of coherent detection in which the incoming optical signal is mixed with a local oscillator laser. The use of coherent transceivers further allows higher spectral efficiency (as apart from amplitude, also phase can be modulated, and this on two orthogonal polarizations). Finally, as the full optical field can be reconstructed through coherent detection (hence both polarizations, phase and amplitude), the linear degradation due to dispersion, can be fully compensated, which would typically limit direct detection.

The downside of using coherent detection is the complexity of both optics (the need for an external cavity low linewidth, tuneable laser) and electronics (extensive digital signal processing to assist the carrier recovery).

Reduction of the complexity and cost of coherent detection is a crucial step to enable their use for higher volume applications such as intra data center applications. Today, coherent detection relies extensively on digital signal processing techniques using analog-to-digital conversion operating at least at symbol rates (but typically higher). Optical and analog techniques can be developed to reduce the burden on this DSP or to simplify it by shifting some parts to the optical domain. Several options are outlined hereafter.

10.5.3.1 Integrated narrow linewidth laser sources

Today, coherent receivers rely on feedforward carrier recovery in the digital domain due to the difficulty of detecting optical phase directly at the optical carrier frequency in the region of 193THz (typical optical carrier frequency). A first step to avoid these techniques is the realization of narrow linewidth laser sources such as external cavity laser sources. Narrow linewidth (e.g. much smaller than 100kHz) requires optical platforms with ultra-low losses, possibly combined with electronic techniques to suppress linewidth broadening.

10.5.3.2 Integrated optical phase locked loops

A next step in the avoidance of digital feedforward carrier recovery is the realization of optical phase locked loops (OPLLs). In such loops the phase of the optical local oscillator laser is directly locked to the incoming optical signal, using e.g. optical Costas loops. Such OPLLs have been demonstrated already however further close integration of the optics (laser and detector parts) and required feedback electronics is required to lower power consumption and improve performance.

10.5.3.3 Novel equalization approaches relying on co-developed opto-electronics

Equalization is now extensively used to overcome bandwidth limitations of EO (electrical-to-optical) and OE (optical-to-electrical) components, as well as chromatic and polarization mode dispersion from the fibre plant. Typically, these are all implemented in the digital domain at both transmitter and receiver side. If some of these functions can be shifted to the optical domain under the form of passive filter structures, significant electrical power can be saved by moving them out of the digital domain. In addition to still requiring large electrical and opto-electronic bandwidths, this will require passive filter structures with very low insertion loss as otherwise any gains will be offset by the need for higher laser power.

10.5.4 Optically assisted wireless subsystems

As explained before, new generations of B5G and 6G mobile wireless transmission systems will rely extensively on advanced radio transmission concepts such as beamforming (requiring true time delaying of RF signals), or operate at very high carrier frequencies (100s of GHz, which can be generated by beating lasers on photodetectors spaced apart by the required carrier frequency). Such microwave photonic techniques can play an increasingly important role at these high frequencies.

10.6 Baseband Modems

Figure 10-1 shows the range of the processing options that have been explored over the years for baseband modems [C10-36]. Traditionally, most of the heavy lifting was carried out by various application-specific integrated circuits (ASICs) that had moderate programmability. ASICs were necessary because processor performance was limited by transistor count. More recently, flexible solutions that use a reconfigurable processing element, such as field-programmable gate arrays (FPGAs) instead of ASICs, have been studied. Unlike fixed-function ASICs, FPGAs can be reprogrammed dynamically, although the development effort is still high and significantly higher than writing new software. Therefore, there has also been significant interest in truly programmable baseband processors.

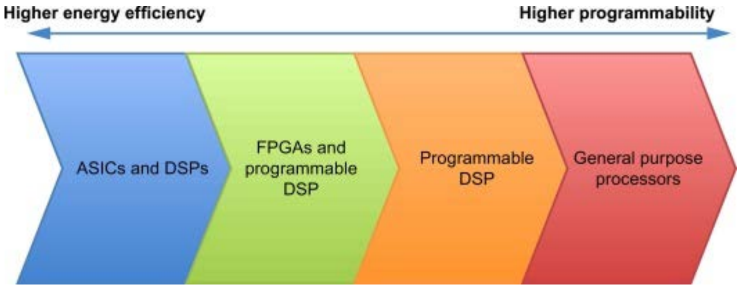


Figure 10-1. Baseband processing options [C10-36].

One style of programmable processors integrates the functionality of FPGAs or ASICs into enhanced digital signal processors (DSPs). These designs not only exploit the data level parallelism inherent to baseband processing workloads but also include domain-specific features that are tuned for baseband processing, such as specialized shuffle networks and arithmetic units. A more radical departure from specialized processors and the adoption of very general fully programmable hardware is an attempt to use off-the-shelf CPUs to process all the tasks of the physical layer processing. Such solutions potentially enable wireless operators to further reduce the cost to build and upgrade RAN infrastructure with commodity off-the-shelf CPUs. Current CPUs have a large, and growing, number of cores and integrate single-instruction multiple-data (SIMD) units within each core. With this high level of parallelism, commodity CPUs can now meet the performance demands of even advanced physical-layer processing.

While there is a large spectrum for possible processing options, the fundamental trade-offs remain the same. Programmable and reconfigurable processing elements are more flexible in that they can work with different signal frequencies, modulations, and coding schemes, and even completely different channel access methods and processing pipelines. This allows wireless operators to reuse hardware resources when migrating to new wireless technologies. Consolidating the functionality of ASICs into fewer processing elements also greatly reduces the cost of both hardware and software development. Finally, flexible equipment can enable even better resource utilization through a more sophisticated resource scheduling strategy such as dynamic resource allocation between different wireless communication technologies. However, these benefits come at the price of energy efficiency and performance because fewer opportunities for low-level specialization and hardware tuning are available with commodity parts than with specialized fixed-function accelerators. Previous work suggests that the performance and efficiency gap can be 10× to 100× between ASICs and general-purpose processors. With the expected slowdown of device scaling and the benefits it provides for performance and energy-efficiency, the trade-off between energy efficiency and programmability in baseband processing hardware is becoming more important than ever.

A recent, and potentially disruptive development is the application of deep learning for the physical layer. By interpreting a communications system as an autoencoder, several groups are developing a fundamental new way to think about communications system's baseband design as an end-to-end reconstruction task that seeks to jointly optimize transmitter and receiver components in a single process [C10-37], [C10-38]. Compared to traditional baseband architectures with a multiple-block structure, the DL based AE provides a new PHY paradigm with a pure data-driven and end-to-end learning-based solution.

10.7 Processors for Cloud-AI, Edge-AI and on-device-AI

The requirements of AI applications are driving the development of dedicated hardware architectures at a rapid pace. CPUs and GPUs are being refined with the purpose of increasing the energy efficiency and reducing the latency. New technological solutions are being leveraged for enabling in-memory compute (i.e. using non-volatile memory technology), multiple chip integration (i.e. chiplets, interposers …), sensor integration.

The rapid pace of adoption of new technologies and ASICs opens up new application segments, since the more processing power is available, the harder the problem addressed. The application space is therefore very broad today but can be split into two main categories: applications that rely on cloud-based solutions and application that run at the edge.

Requirements for cloud-based processors are very specific. First, the cloud is still the workhorse for the learning phase, handling ever larger databases and complex learning algorithms. The compute load must be balanced over many processing units. The first challenge is thus to ensure scalability up to large scales: the associated research areas deal with the interconnect and the memory hierarchy (RDMA over Converged Ethernet being today the solution). Secondly, the cloud must ensure low latency to inference tasks, which are too computation- or memory-intensive to be handled at the edge. The second challenge is thus to provide accelerators optimized for being efficient when handling low batch sizes (typically a size of 1): the research area is the one of data flow and systolic architectures. Finally, there is also a need for energy efficiency, since the datacentres are a large and growing contributor to greenhouse gases emissions. For that, apart from the classical Moore's law pursuit, work is for example being done on data encoding: this has led to the development of the BF16 (Brain Float 16b) representation, which helps save energy and die area compared to the FP32 representation, at almost no accuracy penalty. The research work must be pursued on dynamic encoding.

AI techniques and methods are necessary for IoT in an on-device or edge computing environment to provide advanced analytics and autonomous decision making, impose additional computation requirements on the hardware architectures.

In particular, for applications that run at the edge or on-device the first and foremost key parameters of interest are the energy dissipation and the memory footprint. Both can be addressed thanks to extreme weight quantization, down to binary synapses. This eases analogue in-memory compute, using non-volatile memory technology. The challenge, in this case, is one of learning algorithm: several tricks have to be employed to keep the impact on classification accuracy low. It remains to be seen whether extreme weight quantization is the solution for future applications needs. Indeed, the trend is to have edge platforms or endpoints exhibiting unsupervised or lifelong learning abilities, for applications such as predictive maintenance or adaptation to the environment. The weights accuracy must therefore be higher for the learning algorithm to converge and the on-chip memory larger for storing all the intermediate results. The challenge is to design very dense, local, memory with a low energy access cost. Furthermore, edge or endpoints devices will require sensors for interfacing with the physical world. The difficulty will lie in sensor integration and fusion, with algorithms enabling the use of multimodality (i.e. different input types such as image, sound, vibration). Moreover, research might be needed on flexible on-device operating systems able to cope with open device management ecosystems and AI-based dedicated hardware architectures.

10.8 Memories

10.8.1 Memory technologies towards 2030

10.8.1.1 Entering the zettabyte and yottabyte eras

The amount of data produced in the world will soon exceed 100 zettabyte, with an annual growth rate of 1.2 to 1.4x. The IP traffic is expected to be about 0.25 zettabyte in 2020 and has a similar growth rate. The yottabyte is the order of magnitude for 2030. This huge amount of data and traffic are partly generated through well-known applications such as Amazon, YouTube, Facebook or Netflix. But emerging IoT applications will make a significant contribution as well, such as autonomous cars, smart buildings, smart city, e-health, etc.. Huge amounts of bandwidth are required to transport all this data – from the application to an edge node, then to a base station, and then to a data center – a challenge that will be tackled by 6G and optical networks. Throughout this data flow, stringent requirements will be imposed on memory and storage – in terms of density, bandwidth, cost and energy.

10.8.1.2 Clever data mining, and reduced energy consumption

At some point in the flow of data transport, the generated data will need to be analyzed and converted into knowledge and wisdom by means of machine learning techniques. The exact point at which this will happen, will significantly impact the requirements on memory and storage. For example, if machine learning can be applied just after data generation, it can help relax the requirements. If, on the other hand, data is turned into wisdom later in the process, more raw data will need to be stored throughout the whole process.

The zettabyte and yottabyte eras will also challenge the power that is consumed by the growing amount of data centers, for processing, transporting and storing all the data. Without energy consumption optimization, the energy consumption for these operations, data centers worldwide may use almost 8000 terawatt-hours by 2030. (source: https://www.labs.hpe.com/next-next/energy).

10.8.1.3 The slowdown of today's memory roadmap

Let us have a closer look at the memories used in a typical laptop. Close to the central processing unit (CPU), fast, volatile embedded static random-access memories (SRAMs) are the dominant memories. Also, on-chip are the higher cache memories, mostly made in SRAM or embedded dynamic random-access memory (DRAM) technologies. Off-chip, further away from the CPU, mainly DRAM chips are used for the working memory and non-volatile Flash NAND memory chips for storage. In general, memories located further away from the CPU are cheaper, slower, denser and less volatile.

For half a century, Moore's Law has driven cost improvement of memory technologies, and this has translated into a continuous increase of the memory density. However, despite large improvements in memory density, only storage density (Flash NAND devices) has truly kept pace with the data growth rate. With the transition from NAND to 3D-NAND devices, density improvement for this storage class is however expected to slow down as well and go below the data growth rate soon.

To meet the memory requirements of the zettabyte and yottabyte eras (i.e., improved density and speed, and reduced energy consumption), multiple emerging memory options must be explored for standalone as well as for embedded applications. Options range from MRAM technologies for cache level applications, new ways for improving DRAM devices, emerging storage class memories to fill the gap between DRAM and NAND technologies, solutions for improving 3D-NAND storage devices, and a revolutionary solution for archival type of applications.

10.8.1.4 MRAM technologies for embedded cache level applications

Spin transfer torque MRAM (STT-MRAM) technology [C10-39], [C10-40] has emerged as a candidate technology for replacing L3 cache embedded SRAM memories. It offers non-volatility, high density, high speed and low switching current. The core element of an STT-MRAM device is a magnetic tunnel junction in which a thin dielectric layer is sandwiched between a magnetic fixed layer and a magnetic free layer. Writing of the memory cell is performed by switching the magnetization for the free magnetic layer, by means of a current that is injected perpendicular into the magnetic tunnel junction. Because of speed limitations, STT-MRAM are limited to L3 cache.

An MRAM variant, the spin orbit torque MRAM (SOT-MRAM) [C10-41], can potentially replace the faster L1 and L2 cache. In these devices, switching the free magnetic layer is done by injecting an in-plane current in an adjacent SOT layer, as such de-coupling the read and write path and improving the device endurance and stability.

VCMA-based (Voltage Control of Magnetic Anisotropy) MRAM [C10-42] is another interesting emerging option offering low power, high performance and high density non-volatile memory solution.

10.8.1.5 DRAM scaling

DRAM is structurally a very simple type of memory. A DRAM memory cell consists of one transistor and one capacitor, that can be either charged or discharged. To downscale the structure, the aspect ratio of the structure must be increased. Another concept could be to place the peripherical logic directly under the array of capacitors and transistors. This logic circuitry controls how data is moved to and from the memory chip, and typically consumes considerable area. Today, the transistor of the DRAM memory cell is however built on silicon. To be able to move the peri logic underneath the DRAM array, we need to replace this transistor with a non-Si transistor that is back-end compatible. 3D DRAM integration is yet another improvement path.

10.8.1.6 Storage class memory

Storage class memory has been introduced to fill the gap between DRAM and NAND Flash memories in terms of latency, density, cost and performance. This new memory class should allow massive amounts of data to be accessed in a very short time. Most probably, more than one novel memory technology will be required to span the entire gap. Candidate technologies include various cross-point-based architectures for the memory element, such as phase-change-RAM (PC-RAM), vacancy-modulated conductive oxide (VMCO), conductive bridging RAM (CB-RAM) and oxide RAM (OxRAM).

10.8.1.7 3D NAND... and beyond?

Since its introduction several years ago, 3D NAND [C10-43] has become a mainstream storage technology because of its ability to significantly increase bit density. This is enabled by transitioning from 3 bits per cell to 4 bits per cell. And, instead of traditional x-y scaling in a horizontal plane, 3D NAND scales in the z direction by stacking multiple layers of NAND gates vertically. Today, stacking over 60 layers has become possible but the density improvement of 3D NAND is expected to slow down and will soon not be able to follow the data growth rate [C10-44].

10.8.1.8 DNA storage: the holy grail of archival storage?

DNA storage promises storage densities orders of magnitude higher than semiconductor memories. DNA can be kept stable for millions of years. DNA as a medium for storage is also extremely dense and compact. Writing can be performed by encoding binary data onto strands of DNA through the process of DNA synthesis. The DNA strand can be built up with the base pairs representing a specific letter sequence, through a series of deprotection and protection reactions. As from the read side, there is an enormous technology push to sequence DNA faster and faster and at lower cost. Progress in DNA sequencing has been amazing, even outpacing Moore’s law. But researchers still have a long way to go before reasonable targets (1Gb/s) can be reached. To realize this, faster fluidics, faster chemical reactions and much higher parallelism are needed than what is possible today.

10.8.1.9 Conclusion

It is clear that the classical memory roadmap cannot handle the zettabyte and yottabyte world in terms of energy, density, speed and cost. It will be crucial to improve and develop new memory and storage technologies.

And, lastly, sustainability brings another aspect of the zettabyte and yottabyte eras forward: recycling. To be able to process and store all the data, massive amounts of devices will be produced. The advent of emerging technologies will also bring in new materials, which today are hardly recycled. The semiconductor industry should therefore also find ways to improve the recyclability of all these materials.

10.8.2 Compute-in-Memory

The Figure 10-3 provides a classification of computer architectures and highlights the differences [C10-45]. Depending on where the result of the computation is produced, four possibilities can be identified; they are indicated with four circled numbers and can be grouped into two classes: Computation-outside-Memory (COM) and Computation-In-Memory (CIM). In COM the computing takes places outside the memory core, hence the need of data movement; it has two flavours. COM-Far refers to the traditional architectures such as CPU (circle 4 in Figure 10-2) and CIM-Near refers to architectures that include computation units with the memory core(s) to form an SiP such as Hybrid Memory Cubes (circle 3). In CIM (based on memristive OR devices) the computing result is produced *within* the memory core, and consists also of two flavours; CIM-A in which the result is produced within the *array* such as IMPLY [C10-46] (circle 1), and CIM-P where the result is produced in the memory *peripheral* circuits such as Scouting Logic [C10-47] (circle 2). Note that CIM architecture have relatively low amount of data movement outside the memory core and may exploit the maximum bandwidth (as operations happen inside the memory array). However, CIM requires more design effort to make the computing feasible (e.g., complex read-out circuits); this may result in large complexity which could limit the scalability. Moreover, as CIM performs computations directly on the data residing inside the memory, the robustness and performance are heavily impacted by data misalignment.

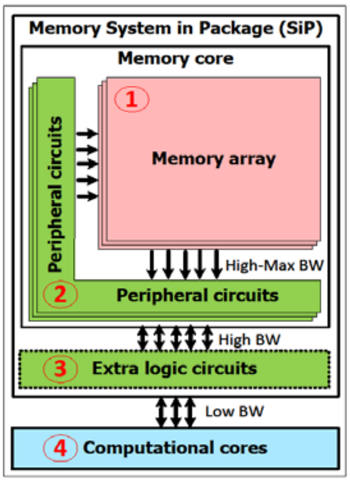


Figure 10-3- Memories in computer architecture

If successful, CIM will be able to significantly reduce the power consumption and enable massive parallelism; hence, increase computing energy and area efficiency by orders of magnitudes. This may enable new power-constrained computing paradigms at the edge such as Neuromorphic computing, Artificial neural networks, Bio-inspired neural networks, etc.

Hence, a lot of application domains can strongly benefit from this computation; examples are IoT devices, wearable devices, wireless sensors, automotive, etc. However, research on CIM (based on memsristive device) is still in its infancy stage, and the challenges are substantial at all levels, including material/technology, circuit and architecture, and tools and compilers.

- *Materials/Technology*: there are still many open questions and aspects related to the technology which help in making memristive device-based computing a reality. Examples are device endurance, high resistance ratio between the off and on state of the devices, multi-level storage, precision of analog weight representation, resistance drift, inherent device-to-device and cycle-to-cycle variations, yield issues, exploring 3D chip integration, etc.
- *Circuit/Architecture/communications*: Analog CIM comes with new challenges to realize (ultra) low power and simple designs of the array structure, peripheral circuits and the communication infrastructure within the CIM and to the I/O interface. Examples are high precision programming of memory elements, relatively stochastic process of analog programming, complexity of signal conversion circuit (digital to analog and analog-to-digital converters), accuracy of measuring (e.g., the current as a metric of the output), scalability of the crossbars and their impact on the accuracy of computing, the partitioning across crossbars and the corresponding intra- and inter-communication under various constrains such as latency, bandwidth and power, etc.
- *Tools/Compilers*: Profiling, simulation and design tools can help the user not only to identify the kernels that can be best accelerated on CIM and estimate the benefit, but also perform design exploration to better guide optimal designs and automatic integration techniques for CMOS and emerging memrisive devices (e.g., monolithic stacking).

As of today, most of the work in the public domain is based on simulations and/or small circuit designs. It is not clear yet when the technology will be mature enough to start commercialization for the first killing applications. Nevertheless, some start-ups on memristor technologies and their applications are already emerging; examples are Crossbar, KNOWM, BioInspired, and GrAI One.

10.9 Hardware for Security

Due to their cost efficiency and promised performance improvements, decentralized deployments are experiencing an important raise of interest in the industrial community, especially in high-risk environments like production sites. Consequently, two main features related to them, namely security and privacy, are getting more and more implemented as hardware (HW) features. In parallel, one can easily witness how all aspect of nowadays life are increasingly supported by HW with extended lifetimes, which forms the core of the so-called Internet of Everything. Today, all vendors compete towards rapid and widescale deployments of billions of devices in the most diverse fields like autonomous vehicles, smart cities, smart homes, and industrial automation [C10-48]. Unfortunately, while a long lifetime of the HW is required, today's state of the art is unfit to ensure such needed-for long-term security feature. *Sustainable security* is therefore a major concern in the industrial ecosystem, as without it billions of vulnerable but active devices will pose a substantial and increasing security risk to the broader society. Therefore there is the need for research actions into sustainable security and privacy, which will shape trustworthy devices that can maintain well-defined guarantees (security, privacy, safety) of critical services over extended life-times (e.g. 20+ years) at affordable cost [C10-49].

Today, a constant stream of risks from many sources (SW, HW, Crypto, Infrastructure, …) renders devices vulnerable and enables mass-scale attacks such as the Mirai Botnet [C10-50]. Devices can only remain secure under active and costly maintenance (vulnerability management, patching, update), which requires a dedicated development team per vendor supporting legacy devices. In practice, three approaches to long-term security are predominant – none of them satisfactory:

- *No or Time-Limited Maintenance*: The most common approach is to only provide limited-time maintenance and accept the fact that devices remain in operation while security rapidly degrades. This creates a substantial risk to society and to users of critical services.
- *Limit the Device Lifetime*: Vendors sell devices with a limited life-time (e.g. limited by warranty). Some vendors use remote update to render devices unusable afterwards. This is not satisfying for users and not environmentally sustainable.
- *Continuous Maintenance and Service Contracts*: For some segments, vendors can offer “devices as a service”, by which vendors are paid for continuous maintenance including security. While this costly approach works for some industrial settings, it will not be realistic for the majority of the existing scenarios.

To solve the problem of sustainable security and privacy, we believe that multiple research areas must be pursued in parallel to mitigate risks to long-term security:

1. *Long-term Security Maintenance*. Smart systems are increasingly deployed in systems that have a long lifetime. Examples include smart cities, smart infrastructure, industrial, and vehicles. Today, each individual system requires costly maintenance (vulnerability scanning, bug fixing, patching, …). This will create a maintenance nightmare for systems that live 20+ years. *We suggest pursuing research on how to build systems that self-maintain their security for 20+ years with minimal maintenance cost.*
2. *Fail-Security + Survivability under Major Attacks.* Even though everyone would agree that designing secure systems is an indispensable feature, in reality the currently deployed systems are far from perfect: if a system was successfully attacked, security can no longer be guaranteed at all, and systems need most of the time to be manually restored, cleaned, and patched. *We therefore suggest exploring new HW mechanisms that allow graceful degradation under attacks while supporting automated recovery of security while the system maintains its critical services*.
3. *HW Security Roadmap towards Post-Quantum Secure Systems:* We believe that quantum computing can break today’s HW implemented security mechanisms. *Since there is no one-size fits all for post quantum security, it is important to analyze a wide range of usages and make appropriate recommendations how to mitigate this risk*.

10.10 Opportunities for IoT Components and Devices

Deploying and managing a large set of distributed devices with constrained capabilities is a complex task. Moreover, updating and maintaining devices deployed in the field is critical to keep the functionality and the security of the IoT systems. To achieve the full functionality expected of an IoT system, research should be done in advanced network reorganization and dynamic function reassignment. Research is needed for providing new IoT device management techniques that are adapted to the evolving distributed architectures for IoT systems based on an open device management ecosystem.

Components (micro-electronic components) and devices mainly for IoT and vertical sector applications are essential elements of future secure and trusted networks and to support the digital autonomy of Europe. With respect to the increasing demand and expectation of secure and trusted networks, especially for critical infrastructures, there should be European providers for such devices as an additional source to latest technologies to complement the European value chain and mitigate the existing gaps.

10.10.1 Approach for components

European semiconductor players are stronger in IoT and secured solutions, while mass-market oriented market are dominated by US or Asian players. For European industry to capture new business opportunities associated with our connected world, it is crucial to support European technological leadership in connectivity supporting digitisation based on IoT and Systems of Systems technologies.

Increasingly, software applications will run as services on distributed systems of systems involving networks with a diversity of resource restrictions.

It is important to create the conditions to enable the ecosystem required to develop an innovative connectivity system leveraging both heterogeneous integration schemes (such as servers, edge device) and derivative semiconductor processes already available in Europe.

Smart services, enabled by smart devices themselves enabled by components introducing an increasing level of “smartness”, will be used in a variety of application fields, being more user-friendly, interacting with each other as well as with the outside world and being reliable, robust and secure, miniaturised, networked, predictive, able to learn and often autonomous. They will be integrated with existing equipment and infrastructure - often by retrofit.

Enabling factors will be: Interoperability with existing systems, self- and re-configurability, scalability, ease of deployment, sustainability, and reliability, will be customised to the application scenario.

Related to technological game changers in 5G network infrastructure, Europe strengths are RF SOI and BICMOS technologies for cost-effective GaAs replacement, FD-SOI for integrated mixed signal System on Chip.

The 5G technologies and beyond utilise the sub-6 GHz band and the spectrum above 24 GHz heading to millimetre-wave technology moving towards 300 GHz and Terahertz frequencies for 6G technologies. The design of electronic components and systems to provide the 5G and beyond connectivity have to take into account the new semiconductor processes for high-speed, high-efficiency compound semiconductor devices considering the significant increases in the density of wireless base stations, wireless backhaul at millimetre wave frequencies, increased transport data rates on wired networks, millimetre wave radios in 5G equipment and multi-frequency/multi-protocol IoT intelligent nodes to support higher data rates, more devices on the network, steerable beams resulting from massive MIMO antennas, low power consumption and high energy efficiency. It is expected that the mobile and intelligent IoT devices to provide edge computing capabilities and intelligent connectivity using multi-frequency/multi-protocol communications technologies. Cellular IoT devices covering higher frequencies need to integrate microwave and analogue front-end technology and millimetre wave monolithic integrated circuits (MMIC).The development of 5G technologies and beyond requires semiconductor technologies that are used for RF devices, base stations, pico-cells, power amplifiers to cover the full range of frequencies required. The new Horizon Europe SNS and KDT Partnerships have to address the development of III-V semiconductors-based GaAs, GaN, InGaAs, SiC semiconductor technologies to implement new components, devices and systems to have the edge in efficiency and power usage needed for base stations.

The new devices for 5G technologies and beyond need to combine RF, low operating power, thermally and energy-efficient, small form factor and heterogeneous integration of different functions. These new requirements push for creating new components based on multi-chip modules and Silicon in Package (SiP) and various technologies that combine the capabilities of silicon CMOS with III-V semiconductors. The focus for new 5G and beyond connectivity IoT devices is on providing new components including hybrid electronic circuits able to operate with better stability, less noise, providing increase functionality, complexity, and performance. The new functionalities include stronger security mechanisms and algorithms integrated into the devices and components and designed for easy implementation of end-to-end security at the application level.

Activities need to be aligned with the KDT Partnership to develop 150 mm and beyond wafers for III-V semiconductors on Silicon to provide the components for 5G and beyond wireless cellular networks and devices for providing optimum use of available bandwidth for millimetre-wave and higher frequencies.

The proposed Smart Networks and Services Partnership will not directly be involved in component research, development and design. However, the research and development in Smart Networks and Services will enable other initiatives to provide the know-how and later the design and production of communication and computing components.

These activities will help to facilitate the re-launch of the micro-electronics industry in the ICT domain in Europe by means of cooperation with the ECSEL JU and/or the proposed Key Digital Technologies Partnership by promoting the development of European added value embedded solutions for innovative and secure applications. Smart Networks and Services will develop the communication know-how and IPRs and will provide algorithms to the micro-electronics industry, which will be dealing with the design and production. With this approach ongoing activities in the ECSEL JU and/or the proposed Key Digital Technologies Partnership can be leveraged. From the Smart Networks and Services perspective that could be a fabless approach. A joint effort of different Partnerships under Horizon Europe will involve the appropriate expertise from different communities.

10.10.2 Approach for devices

Devices and especially end devices for IoT and vertical applications including critical infrastructures are an essential part of future networks. In addition to components they also must fulfil a high security level. The Smart Networks and Services will enable and validate, among others, specialised devices for IoT and sensor systems especially for vertical sectors by leveraging system on chip activities and specifying the way they communicate in the network/systems as well as controlling them and integrating them in their operational systems in vertical (and as well cross- vertical) application domains by means of cooperation with the ECSEL JU and/or the proposed Key Digital Technologies Partnership and leveraging AIOTI activities. System on chip activities can be leveraged for such industrial device activities. The close cooperation between vertical sectors and the ICT industry in Europe will support the development of entire communication and networking solutions in Europe. These activities offer opportunities for start-ups to design communication modem chips and other components devised for many vertical applications.

10.10.3 Requirements for IoT devices

Devices with IoT gateway capabilities in support of different IoT connectivity modes, both at local and public network level. In particular, for each supported vertical industrial domain and as well cross vertical industry domains:

- requirements will be derived on which software and hardware capabilities and characteristics these multi-modal IoT devices and network elements should support, when integrated and used into the 5G and beyond 5G network infrastructures. Considering that these IoT devices support e.g., wireless technologies that are non-5G and beyond 5G radio technologies, such as Bluetooth, Wi-Fi, ZigBee, LoRa, Sigfox
- integration and evaluation activities of these multi-modal IoT devices and network elements in the 5G and beyond 5G network infrastructures will be planned and executed.
- Hardware requirements for IoT Devices:
    - Requirements applied for each supported vertical industry domain and as well cross vertical industry domains when integrated and used into the 5G and beyond 5G network infrastructures.
    - At least three different frequency bands for sub-1 GHz (700 MHz), 1 - 6 GHz (3.4 - 3.8 GHz), and millimetre-wave (above 24 GHz) and integrate multiple protocols in addition to cellular ones.
    - Functional and non-functional requirements, such as high data capacity, highest levels of reliability (connectivity), fast reaction times (low latency), sensing/actuating, processing and storage capabilities; low power consumption.

11. Emerging Technologies and Challenging Trends

The Internet is arguably the most complex infrastructure created by mankind. It is constantly and rapidly evolving to satisfy increasingly important and diverse requirements. Its underlying network infrastructure is in the process of changing from a transport-only, data-less, dumb infrastructure to a multifaceted and distributed system mimicking a living being and consisting of a stratum of fluidified networking and computing resources, dynamically organised and managed by more and more intelligent and autonomous algorithms, which generate and exploit increasing quantities of data, and provide customised services and applications alike everywhere.

Physical connectivity supporting a transparent transfer of information is not anymore the only functionality required to the network. Intelligent algorithms are needed to make the network able to adapt and evolve to meet changing requirements and scenarios and to provide tailored services to users.

Data generated by the network and by the users need to be used by the network itself and to be exploited outside the network. Applications will be more deeply rooted within the network, to provide adaptive features, tailored to user needs, capable to better exploit network-generated data and functionality and to be (dynamically) instantiated (close to) where they are needed. Vertical industries stakeholders will be more and more involved in the communication network value chain, as integrated and distributed applications pose to the network diverse and specific demands.

The network will become even more pervasive and more integrated, further absorbing residual conceptual differences, e.g., between telephone/cellular and Internet/data networks, and imitating the structure of a living being, composed of a Physical stratum + Algorithms + Data + Applications.

Figure 11-1 cursorily summarises the evolution process of the telecommunications networks, showing:

i) How networks developed and evolved (left side of the figure);
ii) Past and current main trends behind such progresses (right side of the figure);
iii) Traditional and new requirements (upper side of the figure) that should be satisfied by the future instance of the network, which we call *x*G[11]. Note that softwarised network will evolve more rapidly and incrementally with new software releases, rather than with major generational leaps.

As shown in Figure 11-1 (left-hand side), networks have been traditionally designed and standardised as independent silos.

Wired Wide Area Networks (WANs) started as the plain old telephone system, evolved into the Integrated Digital Network (substituting analogue with digital technologies), tried to integrate services (ISDN) and improved performance (B-ISDN), but failed, due to the perhaps not fully expected deployment of the Internet in the early nineties; the aftermath is that technologies such as the Asynchronous Transfer Mode (ATM), meant to become ”universal”, became just another component of the growing Internet before disappearing.

---

[11] For the sake of having a neutral future-proof placeholder name.

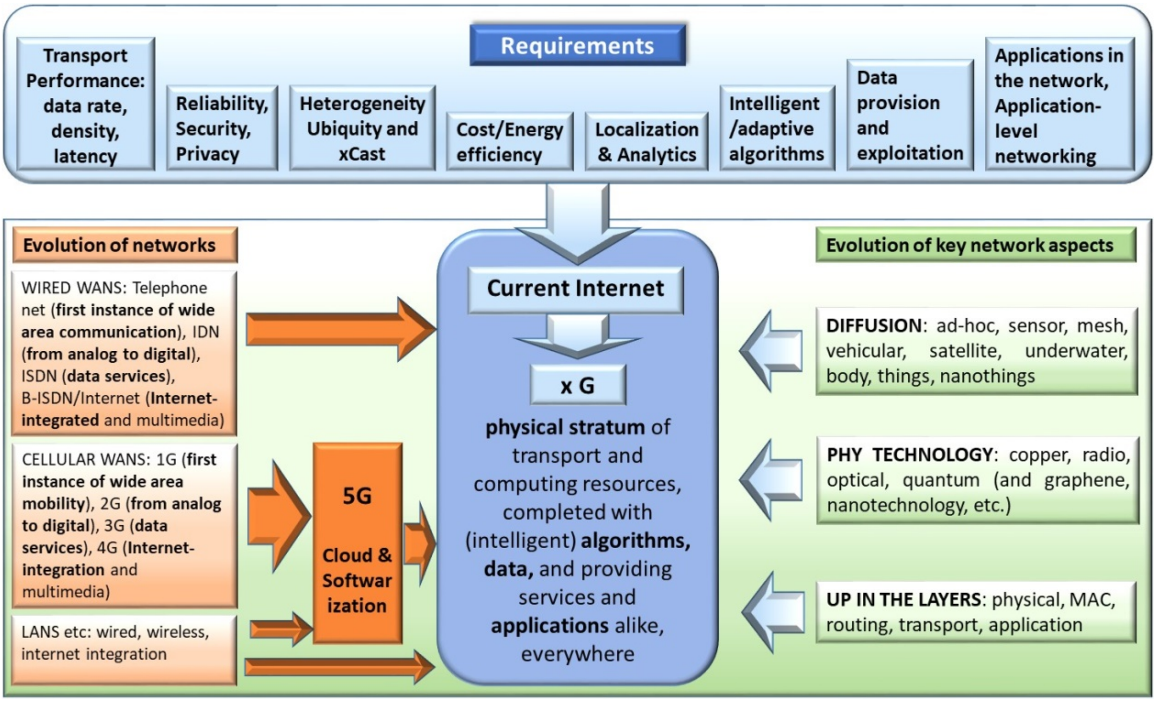


Figure 11-1. Overall picture of the evolution of networks

Cellular WANs took a completely independent evolution path. The first analogue generation (1G) made a first giant leap when it was replaced by the digital GSM cellular network (2G) coming along also with the first data services (SMS, GPRS, WAP). Then 3G brought about new radio technologies, better security and full data services, and 4G made a further step towards the confluence within the Internet, which will be a major unifying feature of the emerging 5G.

In parallel, Local/Metropolitan Area Network technologies started from wired, proprietary, low capacity systems and later extended their realm in the transport network backbones. Meanwhile, the end of the nineties gave the start to an impressive deployment of wireless LANs, which, starting from little more than a low-rate cable replacement, have now almost topped the 10 Gbps speed (e.g., the incoming 802.11ax technology), have become the most common Internet access technology in the home/office/campus, and are being integrated in the 5G network.

Meanwhile, also the networking research trends have evolved along different dimensions and key aspects (Figure 11-1 right-hand side). A first trend regards the diffusion of the networks both in space and in type, with specific infrastructures expanding both the reach and the functionality of networks (ad-hoc, sensor, mesh, vehicular, satellite, underwater, body, things, nano-things, etc.). A second, obvious, trend regards the improvement of physical systems, increasing the performance of copper, radio, optical, and now quantum communications.

The third trend highlighted in Figure 11-1 is probably subtler: networking research in the early days mostly focused on lower layers (MAC protocols, routing); more recently it shifted towards higher layers. At the same time, also the functionality implemented in the network according to the ISO/OSI stack followed the same process: increasingly, higher layer functions are executed also within the network, rather than in end-systems only.

At this point, the reader may wonder why Figure 11-1 does not explicitly lists an ”Internet” box below the evolution of the networks column (and it is mentioned in the middle convergence box). The reason is obvious: The Internet was in fact not a network technology itself (in contrast with LANs, WANs, or cellular technologies or systems), but was focusing on the inter-networking among such underlay technologies, to provide end-to-end services.

However, is this approach still valid today? In the past Internet was one among the services and a marginal one if compared to the circuit-switched voice service, today the traditional differences between Telecom/communication and Internet/computer science communities is blurring (or has blurred), together with the integration of telephone/cellular and Internet/data networks. The TCP/IP paradigm started as an overlay of network technologies (the IP-over-anything frenzy), then with TCP/IP deployed in each technology, then today with TCP/IP having integrated all pre-existent network infrastructures and starting to include in the network also transport and application functionality.

In the meantime, 5G is threatening this model, being more ambitious than previous cellular generations and aspiring to play a bigger role. In fact, 5G is being designed by including also (part of) the wired/core section of the network, as well as LANs (as a matter of fact, important characteristics of 5G would be meaningless if confined to the cellular section proper, e.g., the slice concept), architecturally integrates cloud/fog systems, and is natively a software network, providing differentiated service support. Thus, the current Internet and 5G are almost on a par with each other and converging to an integrated fully inter-operable network.

In fact, 5G is not just an evolution of 4G in terms of performance but creates a breaking point with respect to previous generations, with an end-to-end architecture which blurs and crosses many boundaries with respect to our historical definition of Internet and inter-networking. Moreover, the focus on simultaneously supporting (eventually via slicing) widely different services and ”vertical” applications will make of 5G a much larger ecosystem, including more stakeholders than in the past, with more complex relationships, more heterogeneity and more dynamicity. Last but not least, availability of (big) data, cloud integration and softwarisation will provide a viable technological infrastructure and an in-network ”knowledge plane” – quoting Clark et. al’s vision [C11-1] – which, if today is ”just” used for cost saving and flexibility, may turn into an unprecedented and pragmatic ”arm” for realising the ultimate promise of a truly cognitive network.

To complete the picture, the upper part of Figure 11-1 reports traditional and new requirements that should be satisfied by *x*G networks. From left to right, we start with traditional needs expressed in terms of *performance of the network*, continuing with *ubiquity and xcast (broadcast is an important requirement for 5G)* and the more and more felt *reliability/security/privacy*, and *cost/energy efficiency*, which are cornerstones of the 5G network. Then we enter in more advanced and recent requests made to the network, which include *localisation and physical analytics*, *intelligent/adaptive functionality*, *data provision and exploitation* and finally the need of a tighter *coupling and interactions between telecommunication services provided by the network and applications* and the opportunity arising from directly providing applications as a network service.

This process leads quite naturally to the emergence of *Living and Fluid Networks* [C11-2] namely network architectures holding the ability to autonomously change to best adapt to context. In *Living and Fluid networks*, algorithms and protocols will be capable of understanding what they are used for and will tailor their algorithms and parameters to best suit the requirements of the different applications, appearing in different ways to different packet flows, behaving differently for each one of them, and providing the desired performance to each. Unlike the early cognitive networking insights, we believe that the time for such networks has now come, due to the reached maturity of their two fundamental and complementary aspects:

- the unprecedented modularisation and ”fluidification” of computing and networking building blocks brought about by modern programmable networking trends (i.e. the ”fluid” aspect), and
- the ability to take autonomous (or semi-automated) cognitive orientation strategies and reconfiguration decisions, as provided by the extremely effective and increasingly pervasive modern AI/ML approaches (i.e. the ”living” aspect), and the intertwining of data and functions inside the network.

However, the high softwarization that ICT technologies will experience has a counterpart in terms of energy consumption. In this regard, data centres can increase their energy consumption from the actual 2% of worldwide electricity, to up the 8% by 2030 [C11-3]. Moreover, also autonomous cognitive strategies will further contribute to the energy budget, since AI based solutions are energy hungry. In fact, training a single AI is equivalent to 284 tonnes of carbon dioxide equivalent, i.e., five times the lifetime emissions of an average car [C11-4.]

Key to this evolution is the availability of: i) better underlying technologies, drastically improving communication and computing performance; ii) new techniques for network softwarisation and related primitives and interfaces; iii) intelligent and autonomous algorithms; iv) data; v) applications integrated with the network, performing in part also networking functionality.

This chapter provides a view of the main future technologies and trends behind the evolution of the network as sketched above and is correspondingly structured in three subsections: the physical stratum, algorithms and data, applications.

11.1 The Physical Stratum: Communication and Computing Resources

11.1.1 Nano- and Bio-Nano Things

The many ”Things” we are interconnecting in the Internet are progressively extending to the micro-things, i.e. those computational and service elements that run on small/tiny and non-intrusive things. Nano-communications are emerging to extend the reach of smart control to the level of molecules and cells, with unprecedented impact in medicine and material manufacturing [C11-5]. Combating diseases via autonomous nano-machines, ultra-fast degradation or toxic waste, self-healing and self-monitoring materials constitute a few of the most visionary applications. Materials with software-defined electromagnetic behaviour constitute applications presently under development, paving the way for programmable wireless environments [C11-6].

Recent research on nanomaterials and nano-network architecture components (nodes, controllers, gateways) are opening new prospects of usage of nano-scale things. At the PHY Layer, graphene antennas enable nano-communication within the 0.1 − 10 THz spectral window, which promises unprecedented communication data rates despite the nano-scale. At the MAC Layer, pivotal protocols have been studied to target mainly Body Area Network (BAN) applications [C11-7] and self-monitoring and adapting industrial [C11-8].

Despite these initial promising results, there is a need to provide a more in-depth view and modelling of the network architecture and communication mechanisms in this field, which needs to address various challenges like channel modelling, information encoding and more efficient protocols which allow energy-optimised nano-networks communications.

Critical research challenges to be addressed in the area of nano-networks include:

- Classifying nano-communication paradigms per application scenario. A generalised and unified nano-node architecture is difficult to be obtained. Specific hardware and protocol designs must be produced for each envisioned application, to ensure that the limited nano-node size is optimally exploited in terms of the specifically required functionalities.
- Experimentally validated, application-specific communication channels. Pivotal studies in nano-networks have showcased the workflow to be followed for deriving communication channel models. Such models are important for crafting efficient, application-specific hardware and software for nano-nodes. Thus, it is important to systematically repeat and expand these studies in a wide set of targeted application areas, ranging from solids to biological material.

- Solving the power supply problem. Studies must address the power supply problem in the case of all nano-network types incorporating electronic modules. To date, state-of-the-art autonomic power supplies are sizeable (approximately mm-scale) and of low-capacity. Wireless power transfer and carrier-powered approaches can abolish the need for batteries, under the condition that they are properly designed for their environment and application scenario.
- Solving the battery problem with nano-electronics, fast charging, explosion proof batteries are a must for the future.
- Cost-efficient, massive nano-node integration and production. Separate nano-node components have already been manufactured and tested. The next step is to fully assemble nano-node prototypes, including all separate components. In the case of electronic manufacturing, the related processes exist but yield a high cost. New approaches are required to produce massive numbers of nano-nodes at a low cost and in rapid design/prototyping cycles.
- Hardware-software co-design. Presently, different protocols and software stacks are being developed, mostly from an exploratory angle, given the immaturity of the underlying nano-node hardware. The general consensus is that the software and protocols for nano-networks will indeed face severe limitations in terms of complexity, varying per application scenario. Thus, nano-networks call for co-designed hardware and software right from the start. In other words, the usual workflow of creating general-purpose hardware first and then developing the required software is most likely not feasible in the case of nano-networks.
- Security and Safety. The design of nano-nodes must ensure compliance with authentication and privacy standards, whose severity and criticality vary per application. For instance, consider a biomedical scenario, with a swarm of nano-nodes within a patient’s body. Finally, regarding safety, the presence of nano-nodes within an environment must be well-studied, to ensure that it does not upset its function in an undesired manner.

The artificial nature of IoNT devices can be detrimental for some environments such as inside the body or in natural ecosystems, where the deployment of nano-things and the electromagnetic communication could result in undesirable effects in health or pollution. The novel paradigm of Internet of Bio-Nano-Things (IoBNT) is introduced here by stemming from synthetic biology and nano-technology tools that allow the engineering of biological embedded computing devices. Bio-Nano-Things are defined as uniquely identifiable basic structural and functional units that operate and interact within the biological environment. An analogy can be drawn between a biological cell, which is the basic unit of life, and a typical IoT device since both can perform tasks and functionalities such as sensing, processing, actuation, and interaction with each other. Thanks to the advancements in synthetic biology, it is possible to control, reuse, modify and reengineer the cells’ structure and functions which enables effective use of biological cells as programmable substrates to realise BNT’s as biological embedded computing devices. In particular, the engineering of biological circuits through genetic code manipulation allows specifically designed functions to be performed by the cells such as AND and OR logic gates, switches and counters. Furthermore, artificial cells assembled from bottom up with minimal structural components and functions compared to natural cells can be ideal substrates for synthetic biology with a more predictable behaviour. Although very promising, the aforementioned technologies should provide solutions to major challenges in biotechnology. Reliable mathematical models and computer simulations need to be developed to capture peculiarities of underlying biological processes with intrinsic non-linearity and noise. Also, reproduction and mutation pose extra challenges.

Due to the very small size, previously described nano-things can perform meaningful operations when they communicate and coordinate with each other. As the design of BNT’s, the communication among BNT’s is inspired by nature where the exchange of information between cells is based on the synthesis, emission, propagation, and reception of molecules, called molecular communication. In MC literature, many systems have been proposed such as calcium signalling based on Ca+2 exchange among neighbouring cells in muscle or heart tissues for short range, bacterial chemotaxis and conjugation where bacteria is loaded with information encoded in the genetic material by conjugation and sent to swim to the receiver by chemotaxis for medium range and endocrine communication inside the body among the cells of distant organs by the propagation of hormones through the circulatory systems. Main challenges in communication of BNT's lie in the mapping MC into the classical communication systems, and in the use of tools from systems and information theory with the final goals of modelling and analysing the main telecommunication characteristics and performance, such as range, delay (latency), capacity, and bit error rate.

Bio-Nano-Things are expected to not only communicate with each other, but also interact into networks (typically body networks), which will ultimately interface with the Internet [C11-35. To this end, the definition of network architectures and protocols on top of the aforementioned MC systems is an essential step for IoBNT development. A further challenge for the IoBNT is the interconnection of heterogeneous networks, i.e. composed of different types of Bio-Nano-Things and based on different MC systems. A solution might come from the natural way our body manages and fuses several types of information to maintain a stable, healthy status, or homeostasis. Calcium signalling within a cell can trigger release of hormones to the circulatory systems which in turn control processes such as blood pressure, growth on the distant receiving cells. Biological circuits based on these processes could effectively provide a set of genetic instructions that mimic the classical gateways between different subnets on the Internet.

Finally, the realisation of interfaces between the electrical domain of the Internet and the biochemical domain of the IoBNT networks will be the ultimate frontier to create a seamless interconnection between today's cyber-world and the biological environment. A main challenge is to accurately read the molecular characteristics where information is encoded and translate them into the modulation of electromagnetic waves. This can be achieved by novel nanoscale chemical and biological sensors composed of materials characterised by electrical or electromagnetic properties that can be altered by the presence of specific molecules. Electronic tattoos or artificial cells encapsulating electromagnetic antennas can also be considered as potential bio-cyber interfaces.

IoBNT can revolutionise biomedical technologies and improve human health and quality of life. A possible application scenario is using IoBNT for continuous monitoring and early detection of infections earlier than regular methods relying lab culture. To accomplish this system, a tiny implantable BNT composed of electronic circuits and genetically engineered cell-based biosensors that eavesdrop the quorum sensing (QS) communication of bacteria inside the body can be designed. A gateway BNT will transfer the collected info about the infection from the other BNT's and relay it to a wearable hub for the disposal of healthcare professionals.

With the aim of realising minimally invasive, heterogeneous and externally accessible electrical/molecular communication channels between implantable or wearable electrical and biological IoBNT devices, microbiome-gutbrain axis (MGBA) can be exploited. The molecular information exchanged among bacteria inside the gut is translated into electrical signals by the enteric nervous system and transported to the brain and other IoBNT devices inside the body connected to the nervous system. Hence, MGBA can be considered as a IoBNT communication network infrastructure.

One final general comment on these technologies is related with their potential as enablers for Brain Machine Interfaces. The development of bio/nano networks and devices, and the system ability to obtain a more complete sensory context from the user, will enable the development of increasingly better brain-machine interfaces, which can radically develop new use cases through wireless brain-driven information transmission / reception.

11.1.2 Quantum Networking

In a global scenario where innovation will continue to drive the global economy into the next decade, Quantum Optical Communications are likely to disrupt Information and Communications Technology (ICT) and Telecommunications. Future application scenarios where Quantum Optical Communications technologies are exploited will contribute providing innovative services such as: ultra-massive scale communications for ambient intelligence and connected spaces, holographic telepresence, tactile Internet, new paradigms of brain computer interactions, and new forms of communications. These scenarios are expected to come to fruition in the timescale of this SRIA.

This new generation of systems based on the intrinsic quantum properties of the physical stratum has been progressively made available during the last 20 years. Such systems consist of a classical interface communicating with the standard equipment on one side, and a quantum sub-system sufficiently decoupled from the environment that is constituted by individual atoms, photons and charge carriers (either electrons or holes), controlled by the former. Encoding the information by quantum objects enables the implementation of quantum computing algorithms and communication protocols. Such algorithms rely on the principles of quantum mechanics. The features of the physical stratum at the quantum limit allow encoding information by novel degrees of freedom such as the spin, to exploit quantum superposition of states, entanglement and to rely on the impossibility of cloning states. Quantum systems for networking are generally subdivided between quantum communication systems and quantum computing systems.

Quantum communication will play a central role in the creation of the next generation of secure telecommunication networks. Quantum communication relies on the use of quantum resources to achieve tasks that cannot be reproduced with classical theory. Because quantum communication involves numerous technologies, platforms and application, recommendations on protocols, components and infrastructures require continuous update.

Quantum key distribution (QKD) based on either single or entangled photons relies on the capability of detecting an intruder because of its effect on the quantum states used to encode the secure one-time-pad key. The physical stratum for quantum communication consists of single photon and entangled photon sources, and single photon detectors. Several physical implementations of various protocols have been developed by disparate materials, operating at different wavelengths and consequently capable of peak performances at some temperature, ranging from cryogenic to quasi-room temperature. Nevertheless, true single photon sources operating at room temperature still lack, so low cost solutions can be deployed by using weak coherent lasers for which the communication protocols (such as BB84) are adapted by more sophisticated ones (such as decoy-states protocol). Single-photon detectors that combine high performance (high detection efficiency and low dark counts rate), with low cost possibly by integration in the silicon photonics platform still lack. In order to spread the development and the employment of quantum communication, the implementation of quantum sources and detectors by integrated photonics operating at room temperature is mandatory. Such hardware is complemented by quantum random number generators, currently targeting 1 Gbps, based on intrinsic randomness of quantum processes, ranging from single-charge related electronic noise to photon arrival time. The most successful system is currently based on the conversion into a stream of bits of the arrival time of photons emitted by a weak source as detected by an array of 1024 SPADs in silicon in 45/65 nm technology node.

Quantum Key Distribution protocols are currently the most advanced among the secure quantum communication protocols. However, to be competitive with existing security technologies two main areas of development are identified.

The first area concerns security. Various demonstrations of quantum key distribution systems have been made over the last years. Mainly, the security proofs of all these demonstrations still rely on that the communicating parties have full control of their local ideal devices and therefore there is no information leakage except that needed for the protocol. For quantum encryption to fully guarantee security, the protocols used must be independent of the type of preparation or measurement performed, or the internal workings of the devices used to make the measurement. Hence realistic device independent protocols [C11-9] should be developed and implemented. Short term (3 years) and medium term (6 years) targets are > 10 Mbps and > 100 Mbps at metropolitan distance.

The second area concerns the performance and the application. Device independent protocols address security loopholes issues introduced by imperfect or untrusted systems, as both location and the number of nodes that can be used are an important challenge to be solved. Parallel to the development of these protocols, proposition or implementation of protocols for new applications or improvement of secure transmission performances have to be proposed. Among the potential method allowing the improvement of the secure transmission, high dimensional protocols [C11-10] is a way to increase the capacity and to enhance the robustness against eavesdropping. New challenges are to: i) propose different implementations, in particular different from those using orbital angular momentum of light to make the method compliant with existing single mode fibre networks, ii) Extend the transmission up to 10 km in fibre links.

Apart from the Quantum Key Distribution, other quantum secure communication protocols should be extended beyond laboratory proof-of-principle demonstrations. Here a non-exhaustive list is presented: quantum multiparty communication protocols [C11-11], quantum public-key cryptography [C11-12] quantum secure direct communication [C11-13] quantum digital signatures [C11-14].

Free-space and optical fibre are the most commonly used transmission channels of QKD systems. Numerous proof of concept experiments has been performed demonstrating long distance transmission both in fibre [C11-15] and in free-space links [C11-16].

Free-space QKD can address the increasing demand for security in handled devices for short distance applications, e.g. secure transmission to ATM terminals, or can allow the development of global scale QKD network using satellite communications. The development of on-chip and relatively low-cost devices operating at 810 nm for short distances, or 1550 nm for satellite communication (immunity against the daylight) would be a key enabler for the realistic implementation of QKD for free space applications.

In fibre networks, extend the transmission distance beyond 400 km remains a challenge due to the intrinsic fibre loss. Two main strategies can circumvent this limitation. The first strategy would be the development of trusted backbones, which allow meshing the long distances. This strategy entails also the development of efficient, low cost, on chip sources, detectors and manipulating components operating at 1550 nm. The second strategy consists in implementing quantum repeaters, which is still challenging.

Alternatively, quantum communication based on continuous variables relies on the encoding by modulating the signal below the quantum noise level. The recipient of the communication splits a beam into idler and signal, so to be the only one capable to cancel the noise on the returning beam modulated by the transmitter.

Quantum networking for quantum key distribution deals with the emission of photons and their detection, so the communication ends by destroying the quantum state by the measurement action. Another option consists of enabling communication of quantum states between quantum processors distributed at the nodes of the network, so to create the equivalent of a computer cluster connected through a quantum Internet. Such distributed quantum computing is also called networked quantum computing.

A distributed quantum computer based on quantum networking requires quantum processors including one or more qubits, communication lines for photon transmission, optical switches and quantum repeaters for transport of qubits along distances, because of lack of amplification of quantum states at a fundamental level. The quantum processors may consist of arrays of quantum logic gates involving either defect centres in semiconductors from cryogenic to room temperature depending on the defect and the host material [C11-17] or cryogenic ion traps or cavity quantum electrodynamics. The communication lines by either optical fibres or free space have been already discussed. Quantum switches may rely on the matter-radiation interaction such as single atom-photon based switches, capable for instance to switch the phase of the quantum state [C11-18] In order to extend the communication range, which is limited by both decoherence and losses, quantum repeaters based on cryogenic rare earths-based components are required. Alternatively, a suitable protocol scheme applicable at room temperature has also been proposed, with the advantage of not involving cryogenic equipment [C11-19]. Another direction towards quantum simulation is the Ising-Hamiltonian model (a mathematical model of ferromagnetism in statistical mechanics) that is represented by room temperature networks of optical parametric oscillators as coherent Ising-machines [C11-20]. An new direction that can also be related to the AI-Edge is research on how can classical machine learning (ML) and AI algorithms be enhanced by interaction with quantum computation, leading to a mixed quantum/classical ML/AI to speed up searches in a higher dimensional spaces where mixed/latent information are more separable.

Quantum systems have been demonstrated by several materials at cryogenic temperature. In order to spread their employment in standard telecommunication systems for a novel level of quantum-based security, an effort is required to integrate such quantum systems into silicon photonics and integrated photonic platforms from quasi-room to room temperature.

11.1.3 AI/ML for the Physical Layer

The ever increasing exigence of higher throughput, lower latency and extremely high density of connections is taking the Physical Layer to approach its ultimate performance limits. This implies dealing with channel and transceiver impairments that were usually discarded in the interest of understandable and tractable models. AI and ML can be very helpful tools, also for the design of new transmission and reception techniques, when we cannot rely on the classical models and optimization approaches.

In this regard, there are essentially two kinds of problems where AI/ML may help:

- Offline design of elements or algorithms that may be possible to generalize to a wide variety of scenarios and used afterwards as building blocks of the transmit-receive chain.
- Online optimization of some elements or even the whole transmit-receive approach, that may be able to adapt to the changing (and difficult to model) characteristics of the channel and interference.

Obviously the requirements for both are quite different. While offline optimization requires - but also withstands - training with a large amount of diverse-enough data, the convergence time and complexity are key requirements for the online optimization in real time. These may be achieved by reinforcement learning (RL), a method that interacts with a dynamic environment by producing a series of actions and receives rewards according to the performance of such action with respect to the environment situation.

An example of the offline approach is the use of modern evolutionary computation techniques as solvers of several complex optimization problems. Evolutionary computation [C11-21] is a subfield of AI and soft computing, which is composed of global optimization techniques based on mimicking biological evolution. In [C11-22] it is used to propose novel constellation designs for non-coherent massive MIMO. A different approach is used in [C11-23] where the learned constellation with unsupervised ML mitigates nonlinear effects of the optical fibre channel.

End-to-end learning aims to learn transmitter and receiver implementations optimized for a specific performance metric and channel model. It was first presented in [C11-24. The idea is to interpret the whole communication chain as an “autoencoder”, an unsupervised learning technique. Here the communications system design is conceived as an end-to-end reconstruction task that seeks to jointly optimize transmitter and receiver components in a single process.

ML may also allow us to understand and model channels for new applications, such as molecular communications. In [C11-25] the diffusion-based molecular MIMO channel is modeled with an artificial neural network (ANN).

Learning from the physical layer may also bring us a better knowledge and use of the channel state information, which may be helpful, for example, to locate people or detect their movement, gesture or activity.

Immediate problems to solve are the need for algorithms that have speed and complexity compatible with the real time data transmission. Also, the availability of large and representative data sets for training and benchmarks are less developed than in other current applications of AI/ML.

11.1.4 DSL

The ideal of offering ubiquitous broadband for all requires a high capillarity of high speed wireline access, even to facilitate wireless connections. Today DSL is an alternative to the deployment of optical fibre. Will we be able to count on the already deployed copper wire infrastructure to support 5G and future networks? The possibility of reaching data rates in DSL comparable to fibre would be instrumental to providing universal broadband connectivity in a cost-effective and quick way. However, current DSL technologies seem to have achieved their maximum potentiality with downstream transmission rates of up to 100 Megabits per second at a range of 500 meters, and more than 1 Gigabit per second at shorter distances.

In [C11-26] the channel properties of a 200 GHz signal transmitted through a waveguide structure that is designed to approximately emulate the type of paired phone cable typically used for DSL transmissions are investigated. They find that aggregate data rates on the order of Terabits per second are feasible over short distances.

Transmitting Terabits per second through a copper pair’s sub-millimetre waveguide modes is an emerging technique that would allow us to continue leveraging the copper infrastructure for wireline access and also contribute to accelerate the deployment of small cells with reduced infrastructure cost. Future research needs to bring this idea from the lab to practical applications, solving issues such as how to extend the system to a larger range by reducing the amount of energy lost due to resistance.

11.1.5 The Air Mobility Network

The new Air Mobility Network is the new network serving all the “things” between the ground and 20 Km height. It provides control and communication services for the drone, urban air mobility (future urban transportation systems that move people by air), balloon, aircraft, etc.

Besides traditional Manned Vehicles, the number of flying things has (and will continue to) skyrocketed with the increasing adoption of Unmanned Aerial Vehicles (UAVs), and it is envisioned that this space (20 Km above the ground) will be filled with flying things, which require to be connected in multiple scenarios. The flying things can be categorized into multiple categories, using different classifications, according to the speed of the flying thing, the altitude where they fly, the inter-distance expected between them, or the type of requested services and communications, just to name some examples. However, here, we can categorize the flying things into two main classes: i) aircrafts (airplanes), and ii) UAVs.

Aircrafts have existed for multiple decades now. Airplanes fly at very high altitude being at the extreme limit of considered elevation, with high speed (around 900Km/h), also at the highest end of the considered speeds of civilian flying things on Earth. UAVs reside at the other edge of the spectrum, when it comes to flying altitude and speeds, but with wider variation, when it comes to speeds and altitude, since UAVs vary widely in their speed and altitude capabilities. Note that popular expectations on urban mobility suggest private helicopters, or flying taxis/dispatching units. Nevertheless, these flying units will be essentially self-flying, and will be on the flight altitudes associated to these UAVs, so from the communications challenges perspective, they can be considered in the same class as “UAVs”.

The two classes differ much in their properties, when it comes to speed and altitude, but they still have common properties when it comes to their required communication capabilities. In both classes, the control service of each flying “thing” is critical and mandatory, while the data communication maybe more flexible, depending on the service requirements (Surveillance? Transportation? Logistics? Robotics?). Future communications and networking research should be able to provide the required reliable communications for both categories (control and data), in the most efficient way, while considering cost-effectiveness as a design criterion/requirement.

A) The control service (which can also be referred to as control and nonpayload communication - CNPC - services) of airplanes is critical and mandatory, which defines the basic rules, forms the virtual paths in the air, avoids collision, and reserves passage for airplane/UAVs take-off and landing. The control service may be based on satellite (low earth orbit in most cases), ground base stations, and even among flying “things”, with dynamic multi or single connection, to guarantee the “control” with very high reliability. Passive radio sensing detection is also necessary for the “control”, since it is very dangerous to rely on the positioning reported by the flying “things” themselves in congested airspaces.

B) Data communications (which are used for non-control necessities, including entertainment for passengers) maybe rather flexible, depending on the service requirements, and it could be supported by somewhat traditional communication (for example, connect with a ground base station), or probabilistic communication (carry out data communication within a defined period of time, while each data transmission complies with a certain success probability) among the flying vehicles.

Airplanes: In recent years, we started seeing Internet services onboard of airplanes, and even live TV, with minimal quality. Future networking architectures are required to provide improved reliable broadband communications capabilities to aircrafts. The first type of communication services requires high-reliability, very low-latency, and mostly low data rates. On the other hand, the reliability of the second type may be a bit lower, but still depends on the type of service. We may have variable communication requirements, even ultra-low latency, if for instance gaming, video-conferencing, etc. services are to be offered onboard. However, it is important to differentiate between the criticality of both types of services. In other words, control services have to be guaranteed, and provide high-reliability, to secure the safety of the aircraft itself, while data communication may need the same requirements; however, losing such communications would not jeopardize the safety of the aircrafts.

Communications to aircrafts, flying at very high altitude with ultra-high speed, face three major challenges, to meet their requirements: i) continuous (reliable) communications, even at high altitude; ii) low latency; and iii) high bitrates.

For Air mobility new network, the wireless network architecture could be completely different from the “flat” ground cellular network. The network could be based in clusters of three-dimensional spheres, without interference between clusters if dedicated spectrum could be allocated. How to share spectrum between all these clusters, ground and satellite is a challenge.

One main direction will be how to design an integrated ecosystem, encompassing the multiple networking paradigms, including mobile cellular networks, satellite systems, in addition to even the very high frequency (VHF) spectrum. The future architecture will need to propose interference management and mitigation schemes, between the different systems. For this requirement, channel modelling of cellular networks, towards such high altitude users, needs to be achieved. How cellular networking can provide such seamless continuous connectivity will be another challenge, required to be answered by the future networking generation.

Finally, another dimension for communications will be ad hoc communications between multiple aircrafts, again for safety reasons, to avoid collisions between aircrafts. Ad hoc communications may require completely different networking protocol and architecture.

Unmanned Aerial Vehicles: UAVs are becoming critical these days, and is foreseen to be an essential part of our future ecosystem, especially in the era of IoT and smart city. A study by PricewaterhouseCoopers (PwC) estimates the business market of UAVs to be valued at over $127 billion. This huge market has opened new challenges for the communications.

The first challenge is to guarantee reliable communications for all the flying “Things”. Most UAVs require continuous reliable communications, for the safe travel (flying). The challenge is how to provide such communications, within the random, high variable, high mobile environment of UAVs. It is clear that we will be entering a new domain, different from the plane cellular on the ground, but here, maybe with a huge number of things coexisting in a close environment.

From this new perspective, new research challenges arise:

- How to share the spectrum between different networks, and avoid interference to maximize resource utilization and increase quality and reliability?
- What would be the real-life propagation models in the new 3D environment?
- Study of the effect of low/high mobility of drones on the UAVs communication and how to design communication networks to provide reliable communications in such challenging environment
- Energy efficiency will be another challenge. UAVs are power-limited; hence, minimizing of energy consumption is a priority, to allow extended periods of flying, hence service, specially on swarm scenarios.

The second challenge is to answer the question of how the UAVs can be integrated in the overall networking ecosystem. The use of UAVs within the mobile cellular ecosystem has attracted lots of attention from the academic and industrial entities: The use of UAVs as relays (in particular as small cells), to provide cellular coverage in certain situation has been seen as a high beneficial solution, in the future networks. UAVs can provide networking in emergency situation; to remote areas; temporarily during events with expected large number of devices, etc.

The concept of UAVs as an integral part of the mobile cellular network opens a plethora of challenges for the future networks:

- What frequency should UAVs use? For fronthaul towards the UAVs? For cellular service provisioning to end-users?

- How to coordinate between UAVs, legacy BSs, small cells (which are randomly setup) etc.? Dynamic interference management in such scenario will be highly challenging
- Management of attached, detachment and re-attachment of UAVs will also be challenging, especially with mobility (if allowed in case of UAVs as small cells or relays).
- Energy efficiency will definitely be a high priority, due to the limited power capacity of UAVs.

Similar to how 5G networks revolutionized how communications were handled, so will the next generations further push the envelope beyond just performance enhancements solely based on radio and core architecture evolutions. Whole system rethinking allied to the integration of new enablers (e.g., as SDN, NFV, SBA, Cloud-native were thought for 5G), will transform and create new opportunities for Air Mobility New Network. As the amount of devices with link capabilities increases, due to the decrease in electronics size and price, more connected flying entities will occupy our skies and space. This means that a greater amount of simultaneous connectivity opportunities will present themselves to each individual connected flying thing. By coupling them with multiple network interfaces, air multihoming scenarios will surface, allowing such devices to pick the best connectivity path according to the specific needs of the application (or network control) services traversing those air link channels. This mandates a closer integration with Machine Learning techniques in order to optimize link connectivity choice, as communications will not only scale laterally but also upwards, and become 3D. This will also give birth to new kinds of delay-tolerant / relaying operations and protocols, where the dynamic connectivity opportunities need to be anticipated by learning algorithms, in order to minimize losses or misguided overutilization of air links that do not successfully deliver the information towards the intended target (as needed for reliable communications mandated by verticals or response teams in serious situations). As autonomy is a serious matter in unmanned air vehicles, the added complexity of this type of per-flow optimization (where each flying node can have different applicational flows going through different links in the sky and/or ground) demands greater computational power, new offloading mechanisms towards the edge of the ground network (or flying nodes with greater flight autonomy and even remote battery charge sharing between nodes), will allow such processing to become fully distributed. Obviously, the operational capability for the abovementioned scenarios emphasizes the need for thorough secure operations, deployed both at the control and data paths.

11.2 Protocols, Algorithms and Data

Softwarisation/cloud and security concepts are assumed as already included in the current network architecture.

11.2.1 Impact of AI/ML on the Network

During the last years, the use of AI/ML solutions has reached a great popularity, attracting several innovation activities and growing investments. As a matter of fact, since a few years we have been witnessing that AI/ML is one of the key enabling technologies capable of paving the way to the Digital Transformation of Telecommunications. In fact, AI/ML is impacting the three major techno-economic challenges that Operators are facing: simplifying the networks architectures (to provide any sort of digital services, with shorter time to market and better QoS); cloudifying/edgefying the virtual network functions and services; optimising and automating OSS/BSS processes to mitigate the increasing ”complexity”, dynamisms and pervasivity of the infrastructures.

On the service side, from autonomous driving to speech recognition, a plethora of functional applications have appeared in completely different business areas e.g., Internet of Things, Tactile Internet, Immersive Communications, Automotive, Industry 4.0, Smart Agriculture, Omics and E-Health, etc. The use of the huge data lake generated by the infrastructure will allow automating processes by introducing cognitive capabilities at various levels. Examples of cognitive capabilities include: understanding application needs and automating the dynamic provisioning of services; monitoring and maintaining network state; dynamically allocating virtual network resources and services; ensuring network reliability and enforcing security policies.

For example, services such as autonomous cars have latencies requirements that are so strict (e.g., order of ms) that it is not possible "to close the loop" executing them with cloud computing solutions. The deployment of local processing and MEC (Multi-Access Edge) solutions can help mitigating this problem, but it requires management/control and orchestration capabilities capable of integrating on-device, edge-based and cloud-based AI/ML-systems. The intelligence will progressively migrate towards the edge of the network, requiring research that goes beyond traditional approaches and moves towards novel federated learning solutions, to enable multiple edge devices to build a common, robust ML model without sharing data; multi-task learning, to learn from multiple related tasks, simultaneously; deep transfer learning to solve a particular task using a pre-trained model on a different task; multi-agent reinforcement learning, to distributively learn decision making policies, facing the game theoretic dynamics that appear due to the non-stationarity of the environment.

On the other hand, the management complexity of such future infrastructures (e.g., for FCAPS (Fault, Configuration, Accounting, Performance, Security) and orchestration of virtual resources and services will overwhelm human-made operations, thus posing urgent needs of designing and deploying OSS/BSS with AI/ML features. At the same time, the use of AI for the network will reduce the amount of person-power needed to deploy and operate the infrastructure, thus reducing the operational costs. Moreover, AI/ML can fuel the generation of new services that may lead to improved sustainability models for the network operators.

Nevertheless, the massive adoption of AI tools will exacerbate the problem of energy consumption of the ICT infrastructure. As introduced before, training a single AI is equivalent to 284 tonnes of carbon dioxide equivalent, i.e., five times the lifetime emissions of an average car [C5-3]. Therefore, it will be crucial to devise energy efficient architectures and computation algorithms in order to have energetically sustainable communication and computing paradigms for future mobile networks. For doing so, networks may exploit distributed energy generation and energy harvesting/ storage hardware which have to be included in the design of the network solutions in order to efficiently operate.

Changing Network Design

From the today's perspective AI/ML will enable innovative features when provisioning future digital cognitive services for homes, businesses, transportation, manufacturing, and other industry verticals, including the smart cities. In future scenarios, the increasing usage of End-Users' devices (i.e., smart-phones or tablets) together with the centralised and distributed computational resources will encourage the move of the computational and memory/storage resources from huge data centres towards the edge of the network (e.g., MEC).

Moreover, the huge amount of data sent by new AI/ML applications will lead to hybrid architectures where the data may be partially analysed/compressed in the edge of the network to speed up the whole process and save network resources. In particular, we will see applications able to execute the first layers of a deep neural network locally or in the edge to finish the execution in powerful data centres.

Furthermore, we expect a significant increase in the amount of machine-to-machine communications and of correspondent data to be processed with an increasing number of sensors and other IoT devices continuously monitoring smart cities, Industry 4.0, smart energy, etc.

Automated Operations and Network Intelligence

Today, many Telecom Operators are still relying on manual management processes, but there is a clear awareness of the potential for using AI-powered solutions for automation thus reducing costs, increasing productivity, and driving more value. The rationale is to use AI for automating the operations processes based on collection and elaboration in (almost) real time of data about states and level of performances of nodes/systems and logical/virtualised resources etc. For example, AI/ML can automate the management, control and orchestration (e.g. MCO) processes of physical pieces of equipment, which today are mostly carried out by humans, introducing control loops acting on virtual/logical entities (e.g., Virtual Machines, Containers, appliances etc.). In this direction, AI/ML promises to deliver scalable OSS/BSS functions based on AI/ML models capable of seeing and interpreting the state of millions of network entities via the analysis of huge data streams. Moreover, network and service computational intelligence (e.g., in the Radio Access Networks and in the Core), based on data about Customers' service patterns and traffic would allow improving the quality of the customer experience whilst optimising the use of resources.

One of the main challenges of AI/ML applied to networking is generalisation. State-of-the-art AI/ML solutions are unable to understand topologies not seen during training, since standard feed-forward neural networks are unable to work with graphs. As a result, AI/ML solutions need to be trained for each of the network where it will operate. Although transfer learning may alliviate this issue, the increased cost strongly limits commercialisation, and technologies (such as Graph Neural Networks) need to be developed for bridging these challenges.

Reducing Network Costs and Smart OPEX

In general, the use of AI/ML methods and algorithms will decrease both the costs of deployment and the costs of operations of the network in the following years. This technology will learn the correct network behaviour, being able, in a first step, to help understand possible problems and anomalies, and finally, autonomously acting over the network to correct those problems. For example, to reduce the cost of deployment, the AI/ML will be able to offer zero-touch network configuration for the most common network deployments. This will reduce both the time needed for new deployments and the manpower needed to achieve a proper configuration.

Moreover, different learning techniques will be used to predict the behaviour of the network. This will lead to better provisioning of resources in the network, avoiding the nowadays-typical situation where the networks are over-dimensioned. For example, AI/ML will also enable the adoption of ”QoE” models and indicators to support investment and design processes based on a data-drive approach (e.g., selection of deployment regions, strategic priorities, etc).

Eventually, regarding OPEX optimisation, it is well known that energy consumption is one of the major cost items for Network Operators: AI/ML methods and systems will allow using the data lake for implementing performance analysis and optimisation methods for energy consumption versus quality of service.

Creating new Services Using Network Data

It is likely that the appearance of new services powered by AI/ML will bring significant socio-economic impacts, together with improved sustainability models for Network Operators. Among these services, those ones able to improve mobility, privacy and security levels will be of great importance.

The appearance of personal data platform (tightly connected with the network service) is also expected that will allow Internet users to control their data. To this end, solutions will appear to analyse the network traffic in a privacy preserving and controlled way.

Mobile networks can be used as additional sensing platforms. Indeed, the extreme pervasiveness of the mobile telecommunication sector within the urban population together with its ubiquitous coverage may be exploited to monitor large metropolitan areas. The services that can be generated are manifold (housing, transportation, energy systems, education and health care). For instance, the detection of critical anomalies caused by unexpected crowd gathering in metropolitan areas (e.g., concerts, football matches) can be achieved through the collection of information that the different network elements (e.g., base stations, mobile terminals) are exchanging over time through the control channel [C11-27].

Cybersecurity

Future Networks and 5G will have to face all the security challenges typical of today's telecommunication infrastructures, but with a new and IT-oriented perspective brought by SDN and NFV. Nevertheless, these same enabling technologies, integrated with AI/ML will provide new instruments to mitigate such risks. To mention some examples: inferring proactive actions (even based on early-warning signals of attacks) allowed by AI/ML; adoption of flexible and automatic features for fast traffic steering (e.g., quarantine, honey pots, slicing segregations); automatic configuration of security virtual appliances to be added into the service chains.

Conclusions

The effective applications of AI/ML methods and systems in future 5G scenarios are likely to require multi-domain orchestration of distributed processing in the terminals/devices (could be e.g., Fog Computing), at the edge of the network (e.g., MEC) and in the cloud computing facilities.

In this direction, the end-to-end interoperability is a must and it requires more standardisation efforts and further achievements. First of all, it is necessary to consider the impact of current, and future, AI/ML systems and methods in the functional architecture of 5G. This means understanding which and how architectural functional blocks will be impacted, and what will be the related standardised interfaces. In this direction a global effort is still required from both hardware and software vendors to participate in standardisation bodies, including collaborations with Open Source communities (e.g., Linux Foundation, ONF, OCP-TIP).

11.2.2 Impact of IoT on the Network

The realisation of the Internet of Things vision has already gone through several profound transformations in recent years. In this context, we can clearly identify:

- a first generation of communication and network architectures and protocols, such as those proposed in the context of EPCglobal (https://www.gs1.org/epcglobal), mostly aimed at supporting the exchange of data produced by RFID systems;
- a second generation of networking solutions (6LoWPAN – for „IPv6 over low power Wireless Personal Area Network and CoAP), mostly aimed at making *things* equipped with low capability devices reachable through the Internet and enabling web programming in the resulting environment; and
- a third (the current) generation of solutions (e.g., NB-IoT, LoRaWAN, virtualisation technologies) aimed at supporting the interactions between *things* and some service running in the cloud within silo-ed *platforms*.

We can easily foresee that most of the effort in the next decade will be devoted to the development of solutions aimed at supporting the seamless integration of the above platforms and then at going beyond the *Internet of Platforms* model. Therefore, we will analyse here the impact of IoT on the network based on what we expect it is happening and is going to happen shortly.

Seamless integration of existing platforms requires applications to access IoT resources through some identifier, independently of their native platform, their hardware characteristics, and the protocols executed to interact with them. The corresponding services are today demanded to the application layer, but we can expect that they will become major components of the network itself. Such services should be distributed, should not be under the control of a single (or a few) player(s), should support resource discovery enriched with means for reputation management.

We expect that such solutions will start from the work carried out within EPCglobal in the context of the so-called *Object Name Service* and within the IRTF group ”ICNRG” for what concerns the application of Information Centric Networking techniques to the IoT. Also, concepts will be exploited introduced in the context of peer-to-peer systems, for what concerns the creation of a distributed catalogue of existing IoT resources, and in the context of the Social Internet of Things, for what concerns the creation service discovery and reputation management.

Also, the above solutions will be designed so that they are ready to support the next expected leap forward in IoT evolution, which envisions that individual IoT resources are not bounded to a specific, isolated platform. In other terms it is necessary that they support the case in which IoT resources owned by individual users are used by third party applications. In this way users become *prosumers* of IoT services which requires appropriate new authentication and accounting solutions. Starting point in this context will be the ongoing activities within the IETF Authentication and Authorisation for Constrained Environments (ACE) WG which is working on authenticated authorisation mechanisms for accessing resources hosted on servers in constrained environments and has completed a comprehensive use case document (RFC 7744).

In any case, the major feature of such context will be heterogeneity along several dimensions: access technology, identification/naming/addressing scheme, traffic patterns, deployment extension, device capabilities, etc. Such heterogeneity calls for a network which is highly flexible in all its segments**,** well beyond what is possible to achieve with current software defined networking and network function virtualisation technologies. In fact, while SDN/NFV mostly focus on the programmability of the behaviour of the network infrastructure, in IoT it is crucial to make the protocol stack of end devices programmable as well, so that they can react promptly to changes in the working environment. Also, slicing, which is one of the major concepts exploited to support several logical networks with heterogeneous behaviours on top of the same physical infrastructure, needs to be profoundly revised in the IoT contexts. In fact, in several IoT scenarios the same piece of information transported by the same packet can be of interest of several applications with very different QoS requirements. Such a frequent case cannot be supported by the current implementation of slices that are partitions of the packet space.

Also, the amount of data generated by the IoT is expected to increase with the number of devices at a pace that is orders of magnitude higher than the available data rates. This trend is not sustainable unless radical changes in the Internet infrastructure are introduced. This means that the Internet, which is mostly a communication infrastructure today, must turn into a *computing and communication* infrastructure capable of executing data processing and fusion in any of its components.

The raise of interest towards edge cloud and edge computing goes in this direction; however, the process must go well further and should impact the Internet architecture in its fundamentals. In fact, by turning all network switches/routers into computing nodes the Internet will become a huge and pervasive network of *middleboxes* and several assumptions that are at the very basis of the TCP/IP protocol stack will not be valid anymore.

Finally, it is clear that the true IoT revolution will happen only if a reasonable level of security can be guaranteed. In this context, work is needed to go beyond the work carried out by the ”*DTLS In Constrained Environment”* (DICE) WG that has produced a TLS/DTLS profile that is suitable for constrained devices. In fact, a recent Internet Draft has been produced by the IRTF ”Thing to Thing Research Group” (T2TRG) which provides an overview of open security issues in the IoT domain.

Emerging industrial IoT applications,

It is expected that new emerging applications based on Tactile IoT/IIoT, will be developed in the near future, see e.g., [C11-28]. Examples of Emerging applications using tactile IoT/IIoT, described in [C11-28] are:

- Tactile IoT: Research priorities in this area focus on real-time sensing/actuating using haptic interaction with visual feedback, and the integration of IoT systems supporting not only audio-visual interactions but also involving robotic systems to be controlled with a real-time response.
- Holographic media applications: involve not only the local rendering of holograms but networking aspects, specifically the ability to transmit and stream holographic data from remote sites,
- Multi-Sense Networks: include emerging applications that involve not only optical (video, holograms) and acoustic (audio) senses, but as well smell and taste senses.
- Time Engineered Applications: uses a communication system that can coordinate between different sources of information such that all the parties involved have synchronized view of the application.
- Critical Infrastructure support applications: support of critical infrastructures that refer to those essential assets that are considered vital to the continued smooth functioning of the society as an integrated entity.

Emerging industrial IoT applications, Tactile Internet and autonomous/robotic systems solutions will require far faster reactivity at the edges of the networks as it becomes increasingly inefficient to extract insights from the cloud with growing numbers of IoT devices. Research priorities include the development of new open integrated horizontal platforms for mobile edge computing and edge analytics solutions.

Extreme Automation and Real-Time Zero-Touch - Service Orchestration: In a few years, social machines, smart contracts and other types of more advanced interaction will be a reality. Some machines will be indistinguishable from people from the perspective of business processes and interactions, with higher capacity of decisions, the ability to orchestrate common actions, make requests, etc. Future networks will therefore require higher demands on real-time network service management and a high degree of automation. The challenge is how, without breaking overall end user experience. Several topics are to be addressed, including:

- Enhanced policy management including huge data analytics
- Artificial intelligence driven orchestration
- Cloud-native management applied to Network Function Virtualization orchestration

Service Injection Loop: The creation of services should be reinvented for the new digital area. Architectural micro services provide modular, distributed software components that can be deployed in any environment with a standardized infrastructure, allowing distributed applications to be installed on a cloud infrastructure while maintaining maximum flexibility. Research is required into new ways of describing the entire platform in metamodels. This innovation should be driven not only in the network transformation but also in the creation of a catalogue of new services. These services interoperate with platform capabilities and can automatically adapt to the needs of the user, and will involve new business models in pay-for-what-you-use services.

Digital Twins for IoT: Digital twins are virtual representations of material assets. While current solutions for IoT platforms have mainly been for the representation of physical objects, features such as simulation, manipulation and optimization are missing. Digital twins can be used, for example, to trigger and simulate threat scenarios, and help to optimize the security strategy to handle such scenarios if they occurred in the real world. Research is needed to address the Integration of IoT digital twins into IoT/5G industrial platforms.

11.2.3 Impact of Blockchain Technologies on the Network

We posit that the integration of Blockchain technologies in the Internet infrastructure itself, opposed to application-specific add-ons, will emerge as one of the major and most impactful innovation trends in the Future Internet. As discussed in the following, our belief is that *permissioned* blockchains will gradually extend beyond the very specific single-application realm of most of the today's use cases and will hold the promise to emerge as an open large-scale trust infrastructure, duly controlled and regulated unlike the current massively deployed permission less technologies laying at the foundation of today's crypto currencies. Such a trust infrastructure, while complementing the Internet's connectivity and data distribution services, will likely shape as a federation of independent (and mutually untrusting) providers. Current distributed ledgers' anarchy will most likely be replaced by a form of control and coordination loosely mimicking the way in which multi-domain/multi-country Internet regulation bodies and authorities are today governing and steering the operation of competing Internet Service Providers and autonomous systems. The range of potential use cases for future networks is huge: capacity sharing in distributed networks, spectrum, resource and infrastructure management and sharing, energy trading, resource and service federation in virtualized networks, etc. Facilitating all these scenarios Blockchain technologies would also further contribute to cut down the operational costs of running the network. Additional interdisciplinary interactions with distributed ML/AI solutions and the corresponding game theoretical effects are also to be explored to fulfil the view of Next generation self-organized networks.

The Dawn of Blockchains: The Era of the ”Wild”

Even if the three underlying baseline technology dimensions inside Blockchains root back to works carried out many years before (hash chains and Merkle Trees in the 70ies, consensus protocols in the 80ies, and smart contracts [C11-29] in the 90ies), Blockchains – as we know them today – emerged only in 2008, as the technical foundation and enabler of Bitcoin [C11-30] the first fully decentralised (peer-to-peer) digital/virtual currency. The massive interest in Bitcoin emerged because of its ability to permit transactions without any trusted financial institution intermediaries managing them. Indeed, the Bitcoin's Blockchain, as well as any other Blockchain technology behind the subsequent crypto currencies (Ethereum, Ripple, Litecoin, Cardano, Iota, etc – 1583 at the time of writing) is *completely wild*. With this terminology, we mean that anyone willing to deploy time and resources (e.g. computing power in the case of Bitcoin's Proof-of-Work), not only can participate in building – mining – the relevant blocks but might in principle even try to bias or change its operation. In fact, in Bitcoin, a new crypto currency can be deployed by ”just” convincing a critical mass of block miners to adopt different rules – see the many ”hard forks” popped up just in 2017 (Bitcoin Cash, Gold, Diamond, Private, etc. – we leave the reader to judge which of these initiatives were really necessary in solving real problems).

Even more interesting is the case of Ethereum: the utter flexibility of the relevant scripting language (a Turing-complete language called Solidity) permits anyone to easily create new applications on top of its blockchain by simply programming a ”smart contract”. Despite the hype, and the huge perceived potential in fields also outside crypto currencies, it is fair to say that such flexibility does not nearly come along without concerns[12], and has to date mainly used to launch new coins, often of questionable value – see https://uetoken.com for a very ironic Initial Coin Offer (ICO) which explicitly names itself ”Useless Ethereum Token” and self-describes it as (verbatim quote): ”*the world's first 100 % honest Ethereum ICO: you're going*

---

[12] Among the many disasters, see for instance the catastrophic Ethereum's DAO hack in June 2016 or the case of the Parity wallets, severely hacked as much as twice in one single year, 2017.

to give some random person on the Internet money, and they're going to take it and go buy stuff with it". Perhaps not so unsurprisingly, given the current level of hype, even such a clearly fake ICO (Initial Coin Offering) ended up in being traded for real, gathering as much as 310.445,00 ETH (Ethereum)!

The Emergence of Permissioned Blockchains

Even if emerged in the above discussed *wild* context of crypto currencies, most of the industry is nowadays understanding that blockchains may bring a significant value also in many concrete application domains, as a shared "database" replacement. In this direction, great business attention is currently posed on the so-called "permissioned" blockchain technologies (e.g., Multichain, Hyperledger, etc.), whose somewhat controlled/federated trust model permits them to circumvent the scalability issues and resource consumption (e.g. energy) which affects their public counterparts.

But what is a "permissioned" blockchain? Quoting a crystal-clear explanation by Gideon Greenspan, leader and developer of a permissioned Blockchain technology called Multichain, "*the core value of a blockchain is to enable a database or ledger to be directly shared across boundaries of trust, without putting any single party in charge. A blockchain lets a group of actors achieve real-time reconciliation of validated, authenticated and timestamped transactions, without the cost, hassle and risk of relying on a trusted intermediary*" [C11-31]. In other words, blockchains are clearly pointless in contexts where there is a trusted intermediary which guarantees that what you read from its database is "true". But they do unleash their full value when you need a shared (append-only) database, with multiple writers which do NOT trust each other, and without any trusted intermediary which may validate (and hence guarantee) that what writers are registering in the shared database is truthful.

This latter point – explicit and upfront validation of every transaction prior to storing it in the ledger – is what makes blockchains very different from ordinary databases. Indeed, the trustworthiness of the information contained in a blockchain is accomplished by the *joint* involvement of three complementary techniques:

i) a way to make sure that a transaction recorded at a given time cannot be modified in the future – i.e. the hash-pointer block structure which guarantees storage integrity;
ii) a way to resolve differences among different replicas of the blocks – this is accomplished by a suitable consensus protocol, and
iii) a way to explicitly verify that a transaction being stored is valid, via a suitable formal script associated to the transaction and "executed" prior to adding a transaction to the ledger.

The key advantage of permissioned blockchains with respect to their "wild" public counterparts is that not everybody can create blocks and add them to the chain, but only the subset of parties that have been granted an explicit *permission* to do. This fact completely changes many underlying technical requirements, and permits to significantly widen (and make explicit) the consensus protocols employed [C11-32], improve scalability, guarantee fork-less operation (e.g. with signature-based consensus), improve timestamping and time necessary until a transaction is guaranteed to be registered of orders of magnitudes with respect to the today's Bitcoin hours. Most notably, an upfront fixed number of "miners" permits to get rid of the need to defend against Sybil attacks, the primary reason which mandates the impressive waste of energy in the Bitcoin's Proof-of-Work.

Blockchains as Internet Infrastructure Extension?

A further advantage of a properly implemented permissioned blockchain also resides in the possibility to further control *who, specifically,* can create a smart contract (in other words, an application on top of the blockchain), and how.

Many readers might of course strongly complain that the presence of *controlled* parties which manage the chain, along with the possible restrictions set forth in terms of smart contracts' deployment (or even permissions to transact on the chain) are in sheer contrast with the original decentralisation reasons that have led to the invention and emergence of the Bitcoin's blockchain. While in principle we highly value full decentralisation and freedom, it is a matter of fact that lack of any form of control may easily yield abuses, scams, and fakes. The previously mentioned Useless Ethereum Token is a blatant example of how users, lacking the ability and the instruments to thoroughly vet ICOs, may fall into false ones. And, arguably well beyond the discussion carried out in this section, but still related to trust, the problem of data quality and fake information circulating over the Internet is arguably one of the most challenging and widely open Internet threats.

The point we wish to make here is that public, large-scale, infrastructure variants of permissioned blockchains, extending beyond the realm of a specific application, and rather providing a platform, managed by a controlled multiplicity of non-mutually-trusting ”trust providers”, may permit to share explicitly validated information across boundaries of trust. To remark that a large scale permissioned blockchain governed by agreements between independent countries and relevant authorities might not be impossible, it is worth to note that a controlled set of multiple competing providers is exactly the model at the basis of the today's Internet! Such a large-scale trust infrastructure may come along with a Copernican revolution and turn the burden of verifying the validity of a claim from the end user to the infrastructure itself. In other words, a data or transaction is valid when it is recorded in the chain, thus relieving the verification burden from the end layman's user. And validity is clearly specified by a validation script (a smart contract) deployed following a clearly specified governance model enforced by the permissioned blockchain infrastructure itself.

Before concluding, to bring evidence that our thesis might not be too far to come, we remark that initial steps in this direction have been already made. For instance, the ”Certificate Transparency” initiative [C11-32] launched in 2014 by Google gave end users and domain owners the possibility to transparently verify that a formally valid certificate (i.e. correctly signed by a certification authority) was really issued to the domain owner, thus mitigating the problem of fake TLS (Transport Layer Security) certificates. While Google's massive-scale block-based data structure leverages Merkle Trees and closely reminds a ledger, it is still purpose-specific (tailored to the very special case of TLS Certificates), is not meant to support decentralisation and shared ownership (via consensus), and – most notably – lacks any formal validation of the data inserted, which only a scripting language may provide.

With the growing understanding and maturity of permissioned blockchain technologies, with the support of policy makers for identifying the appropriate governance models loosely mimicking the way in which the multi-domain Internet is today controlled by multi-country Internet regulation bodies and authorities, and with the help of technicians for identifying the necessary extensions in the technological platform, such future is not too far to come.

11.2.4 Evolution of Protocols

Several technological trends will affect protocol development in the following years. These include:

1. Achieving ultra-low latency end-to-end communication is now recognised as the most important goal for many applications that will become ubiquitous in the years to come (e.g., networked virtual and augmented reality, automation, etc.). Moreover, for some of these applications (e.g., Augmented Reality (AR) / Virtual Reality (VR)) both ultra-low latency and very high data rates are required, so the traditional latency-throughput trade-off will not be longer applicable.
2. The capacity of access links rapidly increases, especially in the wireless domain. Additionally, hosts can now efficiently use multiple interfaces as if they were a single resource. Users make good use of the higher total capacity, as consumption and

production of high-bandwidth video have also been rapidly increasing. The net effect is that, after more than a decade of almost-certainty, it is today much less clear that congestion always appears in access links. Measurements have shown that core peering interconnections can also be throughput bottlenecks for traffic on end-to-end Internet paths.

3. At the same time, the infrastructure underlying 5G networks exposes increasingly diverse characteristics with e.g. Visible Light Communication links, millimetre-wave links and modern WiFi standards. All these access technologies have in common that they no longer emulate the behaviour of a static-capacity bottleneck. The increasing dynamicity of the exposed behaviour is further intensified by a shift towards greater mobility of both humans and machines (increasing usage of cellular networks, and intrinsically mobile usage scenarios such as Vehicular Networking or Unmanned Aerial Vehicle (UAV) networks).
4. Internet communication patterns have changed, in the sense that connecting to nearby Content Distribution Network (CDN) servers has become the most common way of consuming / using popular Internet services and content, instead of connecting to servers that are far away. Trends such as fog computing will increase the ”locality” of communication for many users (both humans and machines) and applications.
5. Increased flexibility in both in-network devices and networking software in end-hosts is becoming the new norm. The former has become malleable as they are changing from a static hardware design to software-based designs (Software Defined Networking (SDN), Network Function Virtualisation (NFV)). For the latter, developments such as the Internet Engineering Task Force (IETF)’s work on Transport Services (TAPS) and user-level protocol stacks are paving the way for avoiding ossification and making networking stacks more adaptive and future-proof.
6. Finally, security, privacy and trust have moved from being an afterthought in the design of new communication protocols, to an absolute necessity in the face of a growing and ever-evolving threat landscape.

Several of these trends conflict with the traditional layering in the Internet, where TCP/IP protocols interconnect applications across any underlying link layer technologies, and transport-layer congestion control optimises the sending rate. For example, TCP cannot handle quickly changing bottlenecks well and assumes a static bottleneck capacity (conflicting with trend #3) and causes delay by filling buffers (the ”bufferbloat” phenomenon, conflicting with trend #1). TCP is also ”blind” to the underlying technologies, even when there may only be a few hops across one or two types of link layers between a CDN server and an end user (trend #4 is an unused opportunity), and such a short path might be swiftly adapted and controlled in software (trend #5 is an unused opportunity).

Some developments that partially address these trends have surfaced: for example, Information Centric Networking (ICN) focuses almost exclusively on content distribution. The (mostly US-American) industry has been developing methods to improve the performance of the Internet’s transport layer, as well as making it more secure and more flexible; examples of such developments include Multipath TCP (MPTCP), new Active Queue Management (AQM) algorithms, the QUIC transport protocol, novelties in Explicit Congestion Notification (ECN) usage, and new types of congestion control. However, it is unclear whether these point solutions will be flexible and robust enough to both satisfy the needs of upcoming and future applications and be suitable for 5G network technologies and beyond.

The increasing heterogeneity and dynamics of the underlying infrastructure will necessitate greater flexibility, both in end systems and inside the network. Internet transport protocols will have to be exchangeable at run time. Also, better interplay between applications and the underlying network will be necessary. This will enable dynamically mapping the service needs of applications to the current network infrastructure. Inside the network, long-term traffic engineering by deploying hardware will of course prevail, but traffic engineering controlled by humans using software will be replaced by automation and new protocols that learn both from historical data and traffic conditions in real time. AI/ML techniques and data analytics will be key drivers of self-adaptation and self-management, both in network nodes and in the protocol stacks of end hosts. However, all these solutions are still in their infancy – at best – and will require important research efforts before they can be widely used and deployed.

Even more drastic solutions could try to address all the problems related to the trends above by challenging the traditional role of protocol layering. A possible step in this direction are recursive network architectures, as they allow to react faster by tightening the control loop, thus solving problems closer to where they occur. Could we envision a future where TCP/IP would only be used as a rendezvous protocol in the common case, and all communication would switch over to an entirely different technology when this different technology is found to work for a (typically short) end-to-end path?

11.2.5 Smart Living Environments

Smart Living Environments (SLE) encompass smart homes, smart city and smart transportations that can support vulnerable people, such as, but not limited to disabled and elderly people. SLE independent and as a part of institutional care settings, also lead to reduced costs for care systems and better quality of life for vulnerable categories of citizens. Most cities are currently completely unfriendly to people with disabilities. By 2050, the number of disabled or with limited-access people worldwide will reach an estimated 940 million people, or 15% of what will be roughly 6.25 billion total urban dwellers, calling an urgency to the UN's declaration that poor accessibility "presents a major challenge". Initiatives, such as Smart City for All, aim to help adopting a city for special needs by providing a toolkit that contains four tools to help Smart Cities worldwide include a focus on ICT accessibility and the digital inclusion of persons with disabilities and older persons. Recently Google provided Google Maps with "wheelchair accessible" routes in its navigation to make getting around easier for those with mobility needs. Many other steps are performed by governments, corporations and non-for-profit organizations to fulfill smooth operation of SLEs for humans, especially for those with limited abilities.

Recent developments towards so-called Human 2.0 or Augmented Human (AH) focus on the creation of cognitive and physical improvements as an integral part of the human body. These improvements are enabled by specially designed devices, such as leg or hand prosthesis, implants, artificial vision connected to the neural system of an organism, augmented reality glasses, hearing aids, insulin pumps, etc. It is highly anticipated that artificially recreated or extended abilities may improve quality of life and even give some competitive advantages for humans, especially vulnerable ones. Smooth and well-orchestrated communication between devices of AH itself, and AH and SLE becoming an exceptionally challenging due to multidisciplinary nature of a problem that includes implanted devices that are expected to communicate to external devices and be a part of services that can be of a critical nature, such as health-related for instance. A large number of devices deploying in high and ultra-high density and operating in different frequencies post another challenge when for example a signal from a prothesis will not be received due to e.g. a signal of a wheel-chair high-frequency radar coupled with interference from some other SLE high frequency signal, such one from a drone suddenly coming into vicinity.

The recently introduced emerging paradigms of the Internet of Nano Things (IoNT) and the Internet of Bio-Nano Things (IoBNT) for cyber-physical environment and contribution towards AH have already attracted significant attention within the research community and are currently ramping up on the commercial interest. Accordingly, the well-known concept of the Internet of Things (IoT) has been decisively pushed forward in light of novel research advances made in the field of nanotechnology, nanobioscience, and communications engineering, which enable the construction of networks of embedded computing devices, named nanothings. These devices are constructed from novel nanomaterials such as graphene or metamaterials having the scales that range from one to several hundred nanometers. The IoNT was introduced very recently and is based on the nanoscale devices (nanothings) that owing to their very small sizes can be easily concealed, implanted, and scattered across the environment, wherein they can cooperatively perform sensing, actuation, processing, and networking. While the use of nanothings has the potential to deliver the engineering of devices and systems to unprecedented environments and scales – similar to other devices – they still have the artificial nature, since they are based on synthesized materials, electronic circuits, and interact through electromagnetic (EM) radio communication.

The IoBNT is based on an analogy between a biological cell and a typical IoT-type embedded computing device. This paradigm shift is a novel research direction in the field of molecular biology and allows for engineering of nanoscale devices and systems. This is done by combining nanotechnology with tools coming from synthetic biology to control, reuse, modify, and reengineer biological cells. Both of the above concepts were considered as the basis of numerous future applications, such as military, healthcare, and security, thus enabling fundamentally new opportunities in the design of embedded systems.

A nanothing, whether biological or digital, out-of-body (OOB) or inside-the-body (ITB), aims to perform simple tasks of sensing, computation, and actuation. These tasks include e.g., generation of motion, sensing the environmental characteristics, performing chemical reactions, etc. Beyond communicating with each other and forming a sensor network, nanothings are expected to interact with other existing biological and cyber-physical systems to offer a set of new functions that will include medical (e.g., nanomedicine, targeted drug delivery, in vivo sensor and actuator networks, and tissue engineering), environmental (e.g., quality control), information and communications technology (e.g., implantable bio-sensor and actuator networks), and defense systems (e.g., biochemical sensing for anti-bioterrorism) applications.

Hence, all the nanoscale devices will need to ultimately interface with the Internet and operate together with macroscale networks and systems. To this end, the definition of future network architectures and protocols on top of such integrated systems is a crucial next step in their development. A key new challenge lies in the interconnection of heterogeneous networks, i.e., those composed of different types of biological and artificial nanothings. Ultimately, the realization of interfaces between the electrical domain of the Internet and its biochemical domain will become an unprecedented novel frontier. It will help create a seamless interconnection between today’s cyber-world and the biological environment thus approaching AH and further integrating with SLE.

As a result, nanoscale communications is a decisive paradigm shift that aims to allow the nanothings to communicate over either electromagnetic waves, similar to existing wireless technologies, or molecular and chemical communications. Although using electromagnetic waves is attractive, existing wireless technology must be adopted to the nanoscale environment due to constraints on the device size, its required energy, and the frequency bands that may be utilized for nanoscale communication. Recent works have demonstrated that it is indeed possible to design nanoantennas of a few hundred nanometers in size, which would operate in the Terahertz (THz) band. Real-world experimental antennas have thus been implemented by using graphene and carbon nanotubes, but groundwork is needed to identify the exact operating frequency, radiation efficiency, and coverage.

This research agenda concentrates on the nanonetworks formed by digital ITB and OOB nanothings based on EM communications, thus playing a significant role in materializing the IoNT vision. The state-of-the-art in EM communications technology currently limits the processing capabilities of the prospective nanothings. As a result, all of the analysis of the collected data needs to be performed on a macro device. Therefore, to effectively enable the envisioned applications of nanonetworks, their seamless integration into the existing networking infrastructure is in prompt demand. In this proposed research agenda, the interoperability between the already deployed macro networks and the emerging nanonetworks will be carefully investigated. Any solutions for this problem are non-trivial, since the existing macro wireless networks primarily employ carrier-based EM communications, whereas nanothings need to rely on the ultra-low-power pulse-based EM radiation or inherently mobile objects as its information carriers.

Therefore, the much-needed direct interaction between macro and nanonetworks currently remains infeasible. On top of this, contemporary solutions for nanocommunications have to be rapidly augmented to enable construction of large-scale networks on top of the existing link-level techniques. Numerous important questions are hence to be addressed to achieve this goal, ranging from the design of dual antenna systems that allow for interactions on both scales, adequate modulation and coding techniques to efficiently mitigate noise and interference effects, etc. The rapid advancements in IoNT networks and their information theoretic foundations fuel the progress in THz communications.

11.3 Applications

11.3.1 Application Level Networking

The continued growth in the video space and the push for increased quality, interactivity, and personalisation of media content, coupled with the widespread introduction of augmented and virtual reality, and increasingly heterogeneous and mobile platforms, challenges network performance, and will require new approaches and solutions. Surveillance and monitoring, whether fixed feeds or drone-based on-demand monitoring for disaster management, event security, etc., will further complicate the space; as will the growth in real-time sensor data distributed via machine-to-machine networks for control and management of industrial facilities and smart cities. Virtualisation of applications and their supporting services, enabling ubiquitous deployment via cloud and fog computing services, poses novel infrastructure management challenges. And the need to support innovation and overcome the performance limitations of edge devices requires new APIs and programming models.

The ongoing shift of TV distribution from broadcast onto the Internet will accelerate, driven by the need to transition spectrum to interactive services, cost constraints for all but the most popular content, and the desire to personalise and customise content to suit user interests, to support targeted advertising, and to match device capabilities. The complete transition of such content onto the Internet will involve at least a 10x increase in video traffic volume, yet video already comprises > 75 % of Internet traffic (Cisco Visual Networking Index, 2017). The ongoing transition towards 4k and 8k video, high dynamic range colour, and higher frame rates will further drive the traffic load. The desire for interactivity (e.g., dynamic viewpoint selection, augmented- and virtual-reality) further impacts load and introduces strict latency bounds for an effective user experience. The implications on application level networking are tremendous: the existing protocol stack cannot meet the needs of such applications and must evolve. It is necessary to move away from video as a specialist service, and rather integrate video content, live or pre-recorded, within the web infrastructure and content model. This does not mean abandoning quality of experience or quality of service guarantees – such will become ever more critical – but rather integrating those with the web content framework, delivery model, and APIs, to make video becomes addressable, accessible, and embeddable, and a fundamental part of the web experience.

The initial steps in this process are visible in the WebRTC standards, developed by W3C and IETF, that began the integration of real-time content into browsers – exposing novel APIs for capture, playback, and processing of real-time audio-visual content in web applications. The process is set to continue, with deployment of HTTP/2, QUIC, and future versions of the MPEG DASH and CMAF standards enabling convergence of real-time media and the web. But this is only the start – deep integration of multipath, to make effective use of ultra-dense and diverse wireless networks, is essential, as is effective multicast and multiparty distribution. Both require transport protocols and web infrastructure evolution, since they change the delivery and security models, and require effective trust delegation and novel security mechanisms.

Increasing network capacity and quality of service, deployments of ultra-dense wireless, and related 5G technologies, will make live upload increasingly possible and relevant for breaking news, live sporting and entertainment events, to augment and replace traditional broadcast coverage. It will require and enable live contribution feeds, editing, and content composition. Raw video content will increasingly be available to augment and supplement professional content, and will be edited and processed live or near-live. This will push requirements for contribution bandwidth and quality of service, edge storage and compute, and edge processing for high quality and capacity video content. As with professional video production, user video contribution will increasingly transition from being a pseudo-isochronous feed to be a contribution of tagged-frames, with rich metadata including geolocation and social context, integrated instantly into the web infrastructure for viewing, processing, and redistribution.

Video provenance will become a key issue, to combat ”fake news” and the effects of AI/ML-generated video that attempts to subvert legitimate content; these pose strong risks to the integrity of political and societal discourse, news, and the reputations of public figures, organisation, and events. While solving this is primarily a societal, political, and legal problem, the wide deployment of data provenance and signing infrastructure, to ensure the veracity of content before reputable organisations will distribute it, can support solutions in this space. Strong, vendor and government neutral, approaches are needed here, that must be multinational and verifiably outside the control of any single operator, vendor, or government to limit accusations of censorship and bias. This plays into the security and integrity of applications, network transport, and in-network processing.

To support these developments, the network must evolve to support highly distributed content, stored, processed, and delivered from a pervasive fog computing infrastructure, with effective quality of experience management. The security challenges are immense: how to ensure integrity and provenance of data, through multiple layers of caching, processing, and distribution, while maintaining privacy. Similarly, for meta-data management, quality of experience, and quality of service for media delivery, transport, and processing.

User device performance is strongly limited by thermal and battery constraints, despite the impressive growth in mobile compute performance. Edge compute, in the form of fog- and cloud-computing, virtualised infrastructure, offers an impressively scalable and flexible platform for off-loading processing, provided such processing scales in a parallel form. We need new APIs, replacing the venerable Berkeley Sockets, to address edge compute limitations, and user-space and kernel bypass networking protocols are increasingly needed to match application performance to the performance of the network. There are challenges in supporting the range of applications and protocols: more flexible APIs are needed, to align with application uses of the network, to support the increasing range of transport services (security, reliability, timeliness, quality of service, quality of experience, application intents, offload) needed by modern applications, and to support innovation by democratising network application development – excessive specialist knowledge is required to develop effective applications in this space.

Precursor projects such as NEAT and Post Sockets set the direction for novel APIs and are beginning to set the direction for future API standards but are just the starting point for the evolution in this space. Applications must express their needs for (partial) reliability, timeliness, robustness, security, quality of service, etc., in an abstract manner – requirements are mapped to underlying network capabilities, taking account network control information, load, wireless infrastructure capacity, and the need to co-exist with other virtualised applications. The policy framework below such an abstract API will increasing take in account not just user, application, and system policy, but also the state of the entire network, interacting with the SDN control place, NFV services, and virtualised fog compute resources to determine what network paths, features, protocols, and resources are available for the application to use. The traditional network API provides no support – the network of the future must do so in a manner that doesn't overwhelm applications, developers, operators, or network managers. We must abstract the complexity and enable intelligent control, while supporting policy choice and application needs.

11.3.2 Applications (Components) in the Network

One of the key developments in the network architecture is the deep integration of application and service functionality pervasively within the network. This is not just data centre and cloud computing resources, but the integration of programmable processing resources throughout the network: in the core, the edges, and pervasively. The concepts of fog computing apply, but also software-defined networks, network function chaining, virtualisation, and container provision. There are numerous challenges developing this vision.

Service discovery is essential. Existing mechanisms rely on a combination of anycast routing, domain name system (DNS)-based identifier-to-location mapping, and application-specific directories to locate services. Anycast routing abuses the Internet routing system to route to the nearest replica of a service but scaling it to large numbers of services bloats the Internet routing tables and is not sustainable. Parallel trends, including the transition to IPv6 and the subsequent use of IPv6 addresses as content identifiers – e.g., the Glass-to-Glass Internet Ecosystem proposal to give each frame of video content a unique IPv6 address, and similar approaches to service and container identification – already push routing scalability to its limits. Alternative routing algorithms, e.g., practical Compact Routing algorithms, or clean slate content centric networking architectures may help scale the routing infrastructure, but there are many open questions on how these will work. Directory services also have limitations around ensuring consistency, update performance, and scaling. The architecture will have to become much more dynamic, since the network of the future will no longer be addressing O(1000s) of proxies for a small number of centralised sites, but O(billions) of sophisticated data management and processing services within the network.

Service provisioning, management, and security are critical. A pervasive service platform is essential to supporting the applications of the future, but if implemented and architected incorrectly has the potential to be a significant platform for malware, surveillance, and denial of service. We must learn how to effectively manage billions of devices, ensuring that they are suitably configured, running appropriate software, kept up-to-date with security updates and patches, and run only properly authenticated and authorised applications. We have some solutions that approximate this in the cloud computing world, for managing large scale data centres, but these are much smaller scope than will be needed to support the services in the future – by several orders of magnitude – and make assumptions about the homogeneity of the user base, services, applications, platforms, and infrastructure that will not hold as the network scales. The network will become increasingly heterogeneous, devices will not be directly accessible from a centralised management node, will be subject to differing access and security policies, and will have strong and varying restrictions on usage. Management tools will have to adapt to support this heterogeneity, and differing policies, usages, and requirements. Existing cloud computing infrastructure, homogeneous and managed by, and in support of, a single organisation is not suitable, nor is it a good model going forward.

Security models must evolve. Some infrastructure will continue to be owned and operated by enterprises, network operators, and application/service providers, with controlled access to the data centres where it is hosted, but there will be increasing use of untrusted or partially trusted physically insecure infrastructure located in residential properties, public locations, or end-user devices. Tools for secure boot, code signing, and cryptographic verification of the execution environment will become critical. As will tools to manage and control data access, management, and provenance. Techniques such as homomorphic encryption, that allow devices to process data without having access to the data they are processing, have potential in this space, but are currently too slow and limited to be realistic – development is needed.

Authentication of services and service providers, while accounting for resource usage, is an essential part of the economics of the network of the future. Micropayments will become a key part of the system. The infrastructure to support in-network services and applications is not free – the CAPEX to deploy the underlying network and computational platforms is great, as are the OPEX to manage, power, provision, and support these services. Billing and accounting models for centralised services are already complex – how can we support, manage, and control costs for services scaling to billions of nodes, within a heterogeneous infrastructure into which the service provider has little-or-no visibility?

Privacy and data management, location of processing and data to match legal and moral restrictions on data distribution, access, and processing become increasingly important. Many of the services and applications envisaged operate on, process, and deal with personal data, that is increasingly – and rightly – subject to strict regulation, control, and limitation. We do not have good tools to reason about, describe, and discuss, how data can be processed, where it is to be located, and how it can be distributed – not in human language, legal language, or code. Policy descriptions, rules, and constraints will need to be specified in a form that can be enforced by the infrastructure on the services, since direct human oversight is not feasible at the scales considered. We have no good models, languages, or tools in this space, yet they must urgently be developed if the service-based model of applications is to scale while retaining any user trust.

Finally, novel programming models and languages will be needed to support these services, applications, and deployments. We accept that initial deployments will be based on existing infrastructure – Linux containers, virtual machines, and traditional network programming models and protocols – but these are clearly insufficient for the network envisaged for the future. The Linux container model is not secure, is not portable, and offers a programming interface that is too broad an interface to be made secure, and insufficiently expressive to meet future needs. Many existing programming languages and APIs are not type or memory safe, making services difficult to reason about, monitor, and control. Virtual machines with well-defined semantics offer one approach, as do novel APIs and programming models such as CloudABI and Capicsum, that constrain and control what operations a service can perform. Languages like Erlang offer compelling concurrency and fault tolerance mechanisms, but are weak at service orchestration, management, discovery, control, and security. The application of programming research via languages like Rust or Idris offers potential to go further, as do platforms such as Singularity, but there is still a long way to do before we can have confidence that services and applications perform as intended, and respect data protection regulations and user privacy.

11.3.3 Applications Making Specific Demands to the Network

The final interface to consider is that between applications and the intelligent, pervasive, and service oriented network of the future. It is clear this interface must broaden, and offer more compelling, easier to use, services and features to meet the needs of future applications. The traditional networking API – the Berkeley Sockets API – is not fit for purpose. It is too low-level, too limited, and does not expose the dynamic, changing, nature of the network, nor the high-level services and features needed to support modern applications.

At the transport layer, the network must evolve to allow applications to make effective use of the resources offered by the network. The increasingly heterogeneous and dynamic nature of the network is not exposed by traditional APIs. We must make applications aware that the network is not constant, and that the services and features offered depend on the location in which the application terminal is located. Available services, and their accessibility, vary from location to location, as does name resolution, routing, transport performance, and policy and privacy constraints. Network functionality varies – yet this is not exposed to the applications or higher layer protocols. At minimum, we must expose this heterogeneity to the applications: making the very different network functions and features visible, supporting multi-path and multicast services where offered, and allowing applications to understand their network environment, probe and use appropriate transport protocols, and make best use of the features and services offered by the network in which they find themselves. The Transport Services and Post Sockets work in the IETF Internet standards community is a first step in this direction, moving beyond the limitations of Berkeley Sockets and starting to embed a policy-compliant intelligence into the stack to help applications make best use of the network, but it is only a start and more needs to be done.

Higher-layer protocols must be enabled. The network API of the future will not be merely a transport API. Rather, the goal is to instantiate local support services and network functions to support the application and move appropriate data to the user location – subject to privacy and policy constraints – such that processing can happen nearby, meeting quality of experience latency bounds. The move is away from networking APIs, towards pervasive distributed systems that can run irrespective of the underlying physical infrastructure, and that are location, policy, and regulatory environment aware.

There are numerous systems that require such support. Simplest are perhaps video content distribution applications, delivering increasingly interactive audio-visual content from service and content providers to users, but also uploading user-generated content to viewers. More latency and performance sensitive applications will follow: augmented and virtual reality, gaming, business support, etc., that require local data and processing to meet latency bounds and service the needs of the application, tracking user behaviour and supporting interactive applications that must predict, and respond to, user needs in real time. More demanding still are applications such as teleoperation of remote devices, surgery, healthcare, autonomous systems, conversational interfaces, and other interactive and pervasive services. These have strict latency and quality-to-service bounds to meet user expectations, and will increasingly rely on sophisticated, multipath, transport services, novel and secure transport protocols, and the ability to spawn local data stores and computation in support of the applications.

At present, we have no effective APIs, protocols, or interaction models to instantiate such services. The cloud infrastructure we offer is low-level, starting virtual machines or containers that run on specific operating systems and hardware infrastructure, with little in the way of support services – the Linux/Unix model is not suitable for the applications of the 21st century, but we have yet to settle of a type-safe, memory-safe, secure programming environment for the future, nor have we begun to develop the APIs and services that applications can use to understand the network, service, and regulatory environment in which they find themselves, or the data that they must manage, have available, or can distribute. The challenges in offering this are immense and rely on the effective integration of network transport services, pervasive computing infrastructure, policy, and data management.

In this context, location is a key requirement common to many applications. Location-Based Services (LBS) and Real-Time Location Systems (RTLS) market is significantly growing, stimulated by the various networked applications offered to the users by the current networks. Nevertheless, localisation aspects (and especially the business exploitation of both localisation information and derived knowledge) have never been considered thoroughly in the network evolution, but have rather been addressed as a valuable, but still aside, add-on to the main communication services that networks are called to provide.

We call for the ambitious challenge of realising Smart Networks to incorporate by design technologies and APIs to enable location/context-based services and powerful business analytics on top of them as a way to fully respond to the needs of the vertical applications implementing new personalised services for the end-users.

Key challenges in the area of network based localisation include:

- **Terminal localisation with sub-meter accuracy**. This precision could be required by applications like personal security, infrastructural monitoring (e.g. structural monitoring of buildings, roads, bridges, etc.), etc. it is critical to consolidate the integration of localisation technologies designed into specific subsystems (Wi-Fi, GNSS, Bluetooth, visible light, inertial, etc.) and to enable the collection, interfacing, and fusion of location-based information coming from heterogeneous technologies and subsystems.

- **Device-free localisation**. The challenge here is to properly design and use a network of sensor radars which are coupled with functions of spatiotemporal analysis of signals backscattered by single and multiple device-free targets (persons, things, and vehicles) and can allow to derive the position information (localisation and tracking) of the target. The work to be done is not only in the integration and processing of the various signals, but also in waveform design to properly obtain localisation accuracy in a given context of propagation, bandwidth, and application. It would be useful to consider mmWave technology to assess the achievable gain in tracking accuracy, as well as to develop innovative algorithms for single and multiple target tracking which make use of signals of opportunity, both radio (such as LTE, DVB, and DAB) and non-radio (i.e. acoustic and visible lights), massive MIMO, etc.
- **Spatiotemporal analytics**. Analytics are key to provide verticals with elaborate knowledge learned from localisation data. Such analytics will primarily leverage basic spatiotemporal features of individuals or crowds such as presence, position, heading, velocity or trajectory. It is needed to develop analytics that take full advantage of the localisation accuracy and precision to derive useful information on the physical behaviour of individuals and connected objects to support business intelligence, smart intuitive buildings, intelligent transportation, smart management of the parking spaces, or network demand prediction.
- **Multi-modal Analytics**. In many domains where localisation is a driving technology, the individuals to be localised are associated with a multitude of data (e.g., accelerometer data, mobile application usage, imaging information activity patterns from the network such as HTTP(S) request sequences, etc.). The availability of additional data sources is an important opportunity to complement and enrich analytics, developing more comprehensive AI/ML models. There is the need to develop novel AI/ML models to combine the various data sources, build efficient representation models, and thus discover/detect collective anomalies. Hierarchical architectures for these analytics efficiently splitting the data engines between the core and edge of the network are key to guarantee low-latency, computationally efficient and scalable analytics processes.
- **Proximity Tracing**. A recent trend, associated with aspects of tracing, is coming from the need to have simple ways of trace contacts. This recent pressure, associated with the current covid19 crisis, will most probably remain for the future, as a tool for managing future societal crisis, and hopefully without depending on user interaction, but instead on network capabilities. This is a completely different set of requirements to the network (or end devices – both Android and iOS are addressing this challenge) that need to be fulfilled while assuring adequate citizen protection in privacy and security aspects.

This last example is an excellent example of the multi-dimensional challenges that future networks will face, and a fitting case to conclude the document.

Applications may benefit from the network infrastructure to obtain rich and detailed information about the location and movement of the mobile devices. This may be essential for some applications and services (e.g. contact tracing applications), but it also poses worrying concerns about how this information may be used (as seen in the society discussions in multiple countries). Contact tracing proposals arguing for the need of getting the geolocation information of the user’s devices and storing it in real-time is a clear example how technology may help to resolve ethical dilemmas. Society concerns about deep surveillance and tracing as a clear violation of individuals privacy, the challenges to perverse the security of the information collected and of the access rights found solutions coming from the technology community. Nowadays most of the countries are following a solution using random identifiers (not using geolocation at all) and a complete distributed approach that is basically respectful of privacy and human rights aspects. Societal aspects must be introduced in the basic technological innovation processes so that considering the increasing ethical dilemmas may be softened by technological alternatives.

Annex 1: SDG Evaluation Examples

Example 1: SDG 2 – Zero Hunger

From the description of SDG 2 the most relevant goals to the SNS Programme are the following:

- **Goal 2.3:** By 2030, double the agricultural productivity and incomes of small-scale food producers, especially women, indigenous peoples, family farmers, pastoralists and fishers, including through secure and equal access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment.
- **Goal 2.C:** Adopt measures to ensure the proper functioning of food commodity markets and their derivatives and facilitate timely access to market information, including on food reserves, in order to help limit extreme food price volatility.

Indicative use cases are:

1. Improve productivity and decrease cost through innovative production processes (e.g., smart farming, remote machinery operation).
2. Provide access to specialised information for some cases in real time (e.g., health of crops using AR/VR systems).
3. Provide access to markets and information in a globalised environment.
4. Provide the means for end-to-end product traceability information.

The corresponding workflow examples are:

1. Easy connection and management of massive IoT devices.
2. Support the automation of the farming activities (autonomic and remotely operated farming machinery).
3. Collection and management of vast amounts of information of localised significance.
4. Access to AI based systems to receive specialised information.
5. Global coverage and always on connectivity in an affordable way.

Finally, the needed ICT building blocks to support the abovementioned workflow are:

1. Support of novel functional architecture (e.g., slicing, advanced ad-hoc mode of operation, support mMTC and URLLC networks, satellite communications etc.).
2. Self-reacting core-orchestrators.
3. Provide efficient and affordable means for coverage and capacity in telecommunication systems (NFV/SDN, advanced edge computing and meta-data, energy efficient mechanisms).
4. Software defined security, distributed trust systems.
5. Scalable management of massive deployment distributed autonomous and cooperative computing, nanothings networking.
6. Energy efficiency through improved RF components and systems.
7. Swarm computing.

Example 2: SDG 3 – Good Health and well-being

From the description of SDG 2 the most relevant goals to the SNS Programme are the following:

- **Goal 3.6:** By 2020, halve the number of global deaths and injuries from road traffic accidents.
- **Goal 3.8:** Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all.

Indicative use cases are:

1. Provide improved road safety solution for vehicles and pedestrians.
2. Provide access to advanced health services (e.g., remote surgery).
3. Provide support for advanced warning system.

The corresponding workflow examples are:

1. Support V2X communication systems for gradually introduce new levels of autonomous driving.
2. Connection to real time and non-real time medical services and doctors even to remote places.
3. Access in a secure way to AI based systems to receive specialized medical information.

Finally, the needed ICT building blocks to support the abovementioned workflow are:

1. Support of novel functional architecture (e.g., slicing, support URLLC networks, integrated fixed-mobile architectures, satellite communications, intent-driven and AI/ML management of network etc.).
2. Seamless fog/edge/cloud orchestration to support V2X Application servers.
3. Provide efficient and affordable means for coverage and capacity in telecommunication systems (spectrum, bandwidth adaptation, advanced edge computing, improved radio technology and signal processing, e.g., Terahertz, mmWave, VLC, capacity scaling).
4. Network wide security and trust on an end-to-end basis.
5. Native integration of AI and ML solutions for both the network and the applications.

References

[C1-1] United Nations: Sustainable Development Goals, August 12, 2015, http://www.un.org/sustainabledevelopment/sustainable-development-goals/.

[C1-2] United Nations – Broadband Commission for Sustainable Development 2025 Targets: ”Connecting the Other Half”. http://www.broadbandcommission.org/Documents/publications/wef2018.pdf.

[C1-3] Dufva, M. Megatrends 2020. Sitra studies 162, Finland

[C1-4] "Digital Transformation and Development Policy", Norwegian Government, Message to the Parliament, Meld. St. 11 (2019-2020), Apr2020, https://www.regjeringen.no/no/dokumenter/meldst11_summary/id2699502/?ch=1

[C1-5] UN SDG report 2019 (https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf)

[C1-6] S. Yrjölä et al. 6G Flagship, “White paper on Business of 6G”, 2020 (to appear)

[C1-7] M. Matinmikko-Blue et al. ”White Paper on 6G Drivers and the UN SDGs”, 2020, arXiv:2004.14695 [eess.SP]

[C1-8] Carayannis, E. G., Barth, T. D., & Campbell, D. F. (2012). The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. Journal of innovation and entrepreneurship, 1(1), 1-12.

[C1-9] Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence. M. Latva-aho & K. Leppänen (Eds). 6G Flagship White Paper, September 2019. http://jultika.oulu.fi/files/isbn9789526223544.pdf

[C1-10] European Commission: Digital Agenda Scoreboard – The EU ICT Sector and its R&D Performance. 2018, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=52246.

[C1-11] World Bank: Exploring the Relationship Between Broadband and Economic Growth. Michael Minges, World Development Report, 2016, http://documents.worldbank.org/curated/en/178701467988875888/pdf/102955-WP-Box394845B-PUBLIC-WDR16-BP-Exploring-the-Relationship-between-Broadband-and-Economic-Growth-Minges.pdf.

[C1-12] McKinsey & Company: The Internet of Things: Mapping the value beyond the hype. McKinsey Global Institute, June 2015, https://www.mckinsey.de/files/unlocking_the_potential_of_the_internet_of_things_full_report.pdf.

[C1-13] EU Commission: Digital Single Market. Making the most of the digital opportunities in Europe. 2017, https://ec.europa.eu/digital-single-market/en/policies/shaping-digital-single-market.

[C1-14] EU Commission: Europe's Digital Progress Report 2017. Commission Staff Working Document, SWD (2017) 160 final, https://ec.europa.eu/transparency/regdoc/rep/10102/2017/EN/SWD-2017-160-F1-EN-MAIN-PART-27.PDF.

[C1-15] ETNO: Accenture Study ”Lead or Lose – A Vision for Europe’s digital future”. https://etno.eu/digital2030/people-planet-prosperity.

[C1-16] SNS proposal, private document between European Telecom associations and the European Community.

[C1-17] Rongpeng Li, Zhifeng Zhao, Xuan Zhou, Guoru Ding, Yan Chen, Zhongyao Wang, and Honggang Zhang: Intelligent 5G: When Cellular Networks Meet Artificial Intelligence – Section on “Challenges”. IEEE Wireless Communications, Vol. 24(5), October 2017; DOI: 10.1109/MWC.2017.1600304WC, https://ieeexplore.ieee.org/document/7886994?ALU=LU1043073.

[C1-18] Kagan, M.: Networking Trends in High-Performance Computing. CIOReview, https://high-performance-computing.cioreview.com/cxoinsight/networking-trends-in-highperformance-computing--nid-12770-cid-84.html

[C1-19] European Commission: Resilience, Deterrence and Defence: Building strong cybersecurity for the EU. Joint communication to the European Parliament and the Council. Brussels, 13.9.2017, JOIN(2017) 450 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017JC0450&from=en.

[C1-20] European Commission: Impact assessment accompanying the document Proposal for a Regulation of the European Parliament and of the Council on ENISA, the "EU Cybersecurity Agency", and repealing Regulation (EU) 526/2013, and on Information and Communication Technology cybersecurity certification ("Cybersecurity Act"), Brussels, 13.9.2017, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017SC0500.

[C1-21] Network 2030: "A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond", FG-NET-2030 ITU-T, May 2019

[C1-22] European Commission. Directorate-General for Research and Innovation: ""Indicators for promoting and monitoring Responsible Research and Innovation" Report from the Expert Group on Policy Indicators for Responsible Research and Innovation.", June 2015. EUR 26866 EN. ISBN 978-92-79-43169-29

[C2-1] GSMA: The Mobile Economy 2018. 2018, https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/05/The-Mobile-Economy-2018.pdf.

[C2-2] ITU: ITU News, October 31, 2018, https://www.itu.int/en/itunews/Documents/2018/2018-04/2018_ITUNews04-en.pdf.

[C2-3] ITU: SDG Digital Investment Framework - A Whole-of-Government Approach to Investing in Digital Technologies to Achieve the SDGs. ITU, 2019, https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-DIGITAL.02-2019-PDF-E.pdf

[C2-4] https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme/missions-horizon-europe_en

[C2-5] European Commission: The European Green Deal. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, 11.12.2019 COM(2019) 640 final, https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf.

[C2-6] "Fourth Report on the State of the Energy Union", European Commission, 9 April, 2019, to be retrieved via: https://ec.europa.eu/commission/publications/4th-state-energy-union_en

[C2-7] The Economist Intelligence Unit: Next Generation Connectivity. October 2018, http://www.osborneclarke.com/insights/next-generation-connectivity-jeremy-kingsley-the-economist-intelligence-unit/

[C2-8] ICNIRP Guidelines for Limiting Exposure to Electromagnetic Fields (100 KHz to 300 GHz)

[C2-9] International Commission on Non-Ionizing Radiation Protection (ICNIRP): March 2020, https://www.icnirp.org/en/differences.html. « Differences between the ICNIRP (2020) and previous guidelines »

[C2-10] Viola, R.: Blogpost, https://ec.europa.eu/digital-single-market/en/blog/speech-roberto-viola-director-general-dg-connect-european-commission.

[C2-11] United Nations – Broadband Commission for Sustainable Development 2025 Targets: “Connecting the Other Half”. http://www.broadbandcommission.org/Documents/publications/wef2018.pdf.

[C2-12] GSMA: 2018 Mobile Industry Impact Report: Sustainable Development Goals. 2018

[C2-13] EU Commission: The Sustainable Development Goals. https://ec.europa.eu/europeaid/policies/sustainable-development-goals_en.

[C2-14] Juncker, J.-C.: State f the Union speech. 2017, http://europa.eu/rapid/attachment/SPEECH-17-3165/en/EN-FR-DE-Speech.pdf.

[C2-15] European Commission: EU Budget for the Future. May 2018 https://ec.europa.eu/commission/sites/beta-political/files/communication-modern-budget-may_2018_en.pdf.

[C3-1] J. Si, Y. Niu, J. Lu and H. Zhang, "High-Precision Estimation of Steering Angle of Agricultural Tractors Using GPS and Low-Accuracy MEMS," in IEEE Transactions on Vehicular Technology, vol. 68, no. 12, pp. 11738-11745, Dec. 2019.

[C3-2] T. C. Thayer, S. Vougioukas, K. Goldberg and S. Carpin, "Multirobot Routing Algorithms for Robots Operating in Vineyards," in IEEE Transactions on Automation Science and Engineering.

[C3-3] B. R. Jerosheja and C. Mythili, "Solar Powered Automated Multi-Tasking Agricultural Robot," 2020 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2020, pp. 1-5.

[C3-4] V. Krueger et al., "A Vertical and Cyber–Physical Integration of Cognitive Robots in Manufacturing," in Proceedings of the IEEE, vol. 104, no. 5, pp. 1114-1127, May 2016.

[C3-5] A. Schwung, D. Schwung and M. S. Abdul Hameed, "Cooperative Robot Control in Flexible Manufacturing Cells: Centralized vs. Distributed Approaches," 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki, Finland, 2019, pp. 233-238.

[C3-6] Kachouie, Reza & Sedighadeli, & Khosla, Rajiv & Chu,. (2014). Socially Assistive Robots in Elderly Care: A MixedMethod Systematic Literature Review. International Journal of Human-Computer Interaction. 30. 10.1080/10447318.2013.873278.

[C3-7] R. Hussain and S. Zeadally, "Autonomous Cars: Research Results, Issues, and Future Challenges," in IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1275-1313, Secondquarter 2019.

[C3-8] M. Beul, D. Droeschel, M. Nieuwenhuisen, J. Quenzel, S. Houben and S. Behnke, "Fast Autonomous Flight in Warehouses for Inventory Applications," in IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 3121-3128, Oct. 2018.

[C3-9] S. Iranmanesh, R. Raad, M. S. Raheel, F. Tubbal and T. Jan, "Novel DTN Mobility-Driven Routing in Autonomous Drone Logistics Networks," in IEEE Access, vol. 8, pp. 13661-13673, 2020.

[C3-10] S. F. Rafique, M. K. Bodla, Z. Ahmed, U. Nasir, A. Zaidi and M. Saleem, "Design and implementation of a UAV for power system utility inspection," 2014 16th

International Power Electronics and Motion Control Conference and Exposition, Antalya, 2014, pp. 1146-1150.

[C3-11] Z. Mao, Y. Yan, J. Wu, J. F. Hajjar and T. Padlr, "Towards Automated Post-Disaster Damage Assessment of Critical Infrastructure with Small Unmanned Aircraft Systems," 2018 IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, 2018, pp. 1-6.

[C3-12] M. P. Kokare and S. H. Pawar, "Energy Monitoring System In Electric Grids: The Role Of Advanced Intelligent And Iot For Future Electric Grid," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-4.

[C3-13] N. Mhaisen, O. Abazeed, Y. A. Hariri, A. Alsalemi and O. Halabi, "Self-Powered IoT-Enabled Water Monitoring System," 2018 International Conference on Computer and Applications (ICCA), Beirut, 2018, pp. 41-45.

[C3-14] T. Wiangtong and P. Sirisuk, "IoT-based Versatile Platform for Precision Farming," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 438-441.

[C3-15] H. T. Yew, M. F. Ng, S. Z. Ping, S. K. Chung, A. Chekima and J. A. Dargham, "IoT Based Real-Time Remote Patient Monitoring System," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 2020, pp. 176-179.

[C3-16] J. Chen, A. Valehi and A. Razi, "Smart Heart Monitoring: Early Prediction of Heart Problems Through Predictive Analysis of ECG Signals," in IEEE Access, vol. 7, pp. 120831-120839, 2019.

[C3-17] A. W. Colombo, S. Karnouskos, O. Kaynak, Y. Shi and S. Yin, "Industrial Cyberphysical Systems: A Backbone of the Fourth Industrial Revolution," in IEEE Industrial Electronics Magazine, vol. 11, no. 1, pp. 6-16, March 2017.

[C3-18] FG-NET2030 – Focus Group on Technologies for Network 2030 "Representative use cases and key network requirements for Network 2030", ITU-T Technical report, 16 January 2020, to be retrieved via: https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-NET2030-2020-SUB.G1-PDF-E.pdf

[C3-19] "Network 2030: A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond", FG-NET-2030 ITU-T, May 2019.

[C3-20] A Survey on Holography: http://drrajivdesaimd.com/2017/09/17/hologram

[C3-21] Holoportation, Microsoft Research, https://www.microsoft.com/en-us/research/project/holoportation-3

[C3-22] Mazin Gilbert, Artificial Intelligence for Autonomous Networks (Book, CRC Press, 2018)

[C3-23] The Space-Terrestrial Integrated Network (STIN): An Overview, Haipeng Yao, Luyao Wang, Xiaodong Wang, Zhou Lu, and Yunjie Liu, IEEE Communications Magazine, 2018.

[C3-24] Page 6: Commented [GK34] Georgios Karagiannis 19-Apr-20 10:21:00 PM FG Network 2030 Sub-Group 1's interim activities: <https://extranet.itu.int/sites/itu-t/focusgroups/net-2030/SitePages/SubGroup%201.aspx>.

[C3-25] ATOS: Journey2022: Resolving Digital Dilemmas 2019, https://atos.net/content/mini-sites/journey-2022/index.html.

[C3-26] Ericsson: 5G Business Potential. 2017, https://www.ericsson.com/assets/local/news-and-events/events/2017/mwcs-2017/topic-3_ericsson_5g_business_potential.pdf.

[C3-27] Huawei: Next generation assurance. November 2017, https://www.sdxcentral.com/articles/featured/huawei-next-generation-assurance/2017/11.

[C3-28] ABIResearch: Mobile World Congress Highlights the Challenges Facing Mobile Service Providers. https://www.abiresearch.com/pages/future-mobile-industry.

[C3-29] OPENET: Digital Evolution: Cut BSS/ OSS Costs, Reduce Delivery Timescales. 2018, https://www.openet.com/doc-redirect/index_form2.php?docid=724.

[C3-30] Neuralink: Ultra-high bandwidth brain-machine interfaces to connect humans and computers. 2018, https://www.neuralink.com.

[C3-31] statista – The Statistics Portal: Global digital population as of January 2018 (in millions). 2018, https://www.statista.com/statistics/617136/digital-population-worldwide.

[C3-32] ETSI: ETSI NFV ”Zero touch network and Service Management” (ETSI ZSM ISG). December 2017, http://www.etsi.org/technologies-clusters/technologies/zero-touch-network-service-management.

[C3-33] TMForum whitepaper: ”OSS of the Future”. 2017, https://www.tmforum.org/resources/whitepapers/oss-of-the-future.

[C3-34] Ericsson: Zero-touch networks with cloud-optimized Network Applications. Ericsson whitepaper, 2017, https://archive.ericsson.net/service/internet/picov/get?DocNo=4/28701-FGB1010325&Lang=EN&HighestFree=Y.

[C3-35] ETSI: ETSI ISG Experiential Network Intelligence Whitepaper. October 2017, http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp22_ENI_FINAL.pdf.

[C3-36] YuLing Chen, Alon Bernstein: Bridging the Gap Between ETSI-NFV and Cloud Native Architecture. Cisco Systems Inc., October 2017, https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwiss4DOr63aAhUojK0KHQYuAsEQFgguMAE&url=https%3A%2F%2Fwww.nctatechnicalpapers.com%2FPaper%2F2017%2F2017-bridging-the-gap-between-etsi-nfv-and-cloud-native-architecture%2Fdownload&usg=AOvVaw3f517ofWFJMSssNLiEALDX.

[C3-37] ETSI: Network Functions Virtualisation (NFV); Architecture; Report on the Enhancements of the NFV architecture towards ”Cloud-native” and ”PaaS” Release 3. Draft version. January 2018, https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjLosm2sa3aAhVNmK0KHZGzDIgQFggnMAA&url=https%3A%2F%2Fdocbox.etsi.org%2FISG%2FNFV%2FOpen%2FDrafts%2FIFA029_Arch_enhancement_for_Cloud-native_%2526_PaaS%2FNFV-IFA029v001.docx&usg=AOvVaw3eJ1RrMe6CbOHH7tEQFLFD.

[C3-38] Analysis Mason: Flexible, extensible, adaptable: towards a universal template for VNFs onboarding and lifecycle management. Analysis Mason Limited, 2017, http://www.analysysmason.com/Research/Content/Reports/flexible-extensible-adaptable-towards-a-universal-template-for-vnf-onboarding-and-lifecycle-management-white-paper.

[C3-39] AI for IT & Network Operations (AIOps) 2019 https://www.tmforum.org/collaboration/catalyst-program/artificial-intelligence-operations-aiops/

[C3-40] Sayadi et al, 5G-PPP 2019 https://5g-ppp.eu/wp-content/uploads/2019/09/5GPPP-Software-Network-WG-White-Paper-2019_FINAL.pdf

[C3-41] EuroDIG 2018 https://www.ngi.eu/news/2018/06/05/a-human-centric-internet-trustworthy-resilient-sustainable-and-inclusive/

[C3-42] Principles in a data economy. EU2019.FI https://dataprinciples2019.fi/wp-content/uploads/2019/09/Dataprinciples_web_1.0.pdf

[C3-43] EC Communication to the EU Parliament Secure 5G deployment in the EU - Implementing the EU toolbox. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64481

[C3-44] H. D. Trinh, E. Zeydan, L. Giupponi, P. Dini, Detecting Mobile Traffic Anomalies Through Physical Control Channel Fingerprinting: A Deep Semi-Supervised Approach, IEEE Access, Vol. 7, pp. 152187-152201, October 2019.

[C3-45] The real-time city: Unlocking the potential of smart mobility. Hussein Dia. Centre of Sustainable Infrastructure-Swinburne University of Technology Melbourne. Nov'16.

[C3-46] 5G: Evolution and Revolution. TMForum. April'19

[C3-47] 5G and connected vehicles: communication and cooperation. Orange – Hello Future research blog 2019

[C3-48] Impact-Mobility, Report: Mobility Data Marketplaces and Integrating Mobility into the Data Economy. Jun 2019

[C3-49] Atos secures communications in connected vehicles. Jan 2019. https://erticonetwork.com/atos-secures-communications-in-connected-vehicles/

[C3-50] Networld2020 "SatCom Resources for Smart and Sustainable Networks and Services" Nov. 2019, available at https://www.networld2020.eu/wp-content/uploads/2019/11/satcom-ss-whitepaper_v1_final.pdf [C5-1] N. Rajatheva et al: "White paper on broadband connectivity in 6G," arXiv preprint: https://arxiv.org/abs/2004.14247, 2020.

[C4-01] "62% of all Internet traffic in Asia Pacific will cross Content Delivery Networks (CDNs) by 2021", online: https://www.prnewswire.com/news-releases/62-of-all-internet-traffic-in-asia-pacific-will-cross-content-delivery-networks-cdns-by-20211-300528121.html. Last accessed May 20, 2020.

[C4-02] "Netflix Eats Up 15% of All Internet Downstream Traffic Worldwide", Available online: https://variety.com/2018/digital/news/netflix-15-percent-internet-bandwidth-worldwide-study-1202963207/. Last accessed May 20, 2020.

[C4-03] IETF SFC WG. Available Online: https://datatracker.ietf.org/wg/sfc/about/. Last accessed May 20, 2020.

[C4-04] IETF ANIMA WG. Available Online: https://datatracker.ietf.org/wg/anima/about/. Last accessed May 20, 2020.

[C4-05] IETF COIN RG. Available Online: https://irtf.org/coinrg. Last accessed May 20, 2020.

[C4-06] IETF FORCES WG. Available Online: https://datatracker.ietf.org/wg/forces/about/. Last accessed May 20, 2020.

[C4-07] IETF QUIC WG. Online: https://datatracker.ietf.org/wg/quic/about/. Last accessed May 20, 2020.

[C4-08] L. Popa, A. Ghodsi, I. Stoica, 'HTTP as the narrow waist of the future internet', Hotnets-IX: Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks October 2010.

[C4-09] D. Clark, “The Design Philosophy of the DARPA Internet Protocols”, Proc. ACM SIGCOMM ‘88, Computer Communication Review Vol. 18, No. 4, August 1988, pp. 106–114.

[C4-10] S. Ren et al., “Routing and Addressing with Length Variable IP Address”, Proc. of ACM SIGCOMM NEAT workshop, 2019.

[C4-11] R. Li et al., “A Framework for Qualitative Communications using Big Packet Protocol”, Proc. of ACM SIGCOMM NEAT workshop, 2019.

[C4-12] R. Li, “Network 2030 and New IP”, 15th International Conference on Network and Service Management, IEEE NOMS 2020. Online: http://www.cnsm-conf.org/2019/files/slides-Richard.pdf. Last accessed May 20, 2020.

[C4-13] Pouzin Society, “RINA: Building a Better Network”, available online: http://pouzinsociety.org/education/highlights, Last accessed May 20, 2020.

[C4-14] D. Trossen et al., “Service-based Routing at the Edge”, available online: https://arxiv.org/abs/1907.01293v1. Last accessed May 20, 2020.

[C4-15] SCION Internet Architecture, online: https://www.scion-architecture.net/. Last accessed May 20, 2020.

[C4-16] European Commission, “The European Green Deal”, Available online: https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf. Last accessed May 20, 2020.

[C4-17] “Between 10 and 20% of electricity consumption from the ICT* sector in 2030?”, online: https://www.enerdata.net/publications/executive-briefing/expected-world-energy-consumption-increase-from-digitalization.html. Last accessed May 20, 2020.

[C4-18] M. Caesar et al., “ROFL: routing on flat labels”, ACM SIGCOMM Computer Communication Review, August 2006, https://doi.org/10.1145/1151659.1159955.

[C4-19] G. Xylomenos et al, “A Survey of Information-Centric Networking Research”, IEEE Communications Surveys & Tutorials (Volume: 16, Issue: 2, Second Quarter 2014).

[C4-20] Trossen et al., “Name-Based Service Function Forwarder (nSFF) Component within a Service Function Chaining (SFC) Framework”, Internet Society, RFC 8677, online: https://datatracker.ietf.org/doc/rfc8677/. Last accessed May 20, 2020.

[C4-21] D. Clark, J. Wroclawski, K. Sollins, R. Braden, “Tussle in Cyberspace: Defining Tomorrow’s Internet”, Proc. ACM SIGCOMM 2002. Available Online: http://conferences.sigcomm.org/sigcomm/2002/papers/tussle.pdf. Last accessed May 20, 2020.

[C4-22] Tim Höfer, Sebastian Bierwirth und Reinhard Madlener, „C15 - Energie Mehrverbrauch in Rechenzentren bei Einführung des 5G Standards“, available online: https://www.eon.com/en/about-us/green-internet.html. Last accessed May 20, 2020.

[C5-1] Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper

[C5-2] Andrae, A.S.G.; Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. Challenges 2015, 6, 117-157.

[C5-3] E. Strubell, A. Ganesh and A. McCallum, “Energy and Policy Considerations for Deep Learning in NLP”, Annual Meeting of the Association for Computational Linguistics (ACL short). Florence, Italy. July 2019.

[C5-4] Deng, S., Zhao, H., Yin, J., Dustdar, S., & Zomaya, A. Y. (2019). Edge Intelligence: the Confluence of Edge Computing and Artificial Intelligence. arXiv preprint arXiv:1909.00560.

[C5-5] Wang, Xiaofei, et al. "In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning." IEEE Network 33.5 (2019): 156-165.

[C5-6] Bellavista, Paolo, and Alessandro Zanni. "Feasibility of fog computing deployment based on docker containerization over raspberrypi." Proceedings of the 18th international conference on distributed computing and networking. 2017.

[C5-7] IEEE 802 Nendica Report: Flexible Factory IoT: Use Cases and Communication Requirements for Wired and Wireless Bridged Networks," in IEEE 802 Nendica Report: Flexible Factory IoT: Use Cases and Communication Requirements for Wired and Wireless Bridged Networks, vol., no., pp.1-48, 17 April 2020.

[C5-8] Chun, Byonggon, et al. "Kubernetes Enhancement for 5G NFV Infrastructure." 2019 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2019.

[C5-9] Ismail, B. I., Goortani, E. M., Ab Karim, M. B., Tat, W. M., Setapa, S., Luke, J. Y., & Hoe, O. H. (2015, August). Evaluation of docker as edge computing platform. In 2015 IEEE Conference on Open Systems (ICOS) (pp. 130-135). IEEE.

[C5-10] Ogbuachi, M. C., Gore, C., Reale, A., Suskovics, P., & Kovács, B. (2019, September). Context-Aware K8S Scheduler for Real Time Distributed 5G Edge Computing Applications. In 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-6). IEEE.

[C5-11] H. D. Trinh, E. Zeydan, L. Giupponi, P. Dini, Detecting Mobile Traffic Anomalies Through Physical Control Channel Fingerprinting: A Deep Semi-Supervised Approach, IEEE Access, Vol. 7, pp. 152187-152201, October 2019.

[C5-12] H. C. Rudolph, A. Kunz, L. L. Iacono and H. V. Nguyen, "Security Challenges of the 3GPP 5G Service Based Architecture," in IEEE Communications Standards Magazine, vol. 3, no. 1, pp. 60-65, March 2019.

[C5-13] Zanzi, L., Cirillo, F., Sciancalepore, V., Giust, F., Costa-Perez, X., Mangiante, S., & Klas, G. (2019). Evolving Multi-Access Edge Computing to Support Enhanced IoT Deployments. IEEE Communications Standards Magazine, 3(2), 26–34.

[C5-14] F. Giust et al., "Multi-Access Edge Computing: The Driver Behind the Wheel of 5G-Connected Cars," in IEEE Communications Standards Magazine, vol. 2, no. 3, pp. 66-73, SEPTEMBER 2018.

[C5-15] A. Ksentini and P. A. Frangoudis, "Toward Slicing-Enabled Multi-Access Edge Computing in 5G," in IEEE Network, vol. 34, no. 2, pp. 99-105, March/April 2020.

[C5-16] C. Campolo, A. Iera, A. Molinaro and G. Ruggeri, "MEC Support for 5G-V2X Use Cases through Docker Containers," 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 2019, pp. 1-6.

[C5-17] Cicconetti, C., Conti, M., Passarella, A., & Sabella, D. (2020). Toward Distributed Computing Environments with Serverless Solutions in Edge Systems. IEEE Communications Magazine, 58(3), 40–46.

[C5-18] A. Varastehy, M. De Andrade, C. Mas Machuca, L. Wosinska, W. Kellerer, “Power-Aware Virtual Network Function Placement and Routing using an Abstraction Technique,” Proc. 2018 IEEE Global Communications Conference (GLOBECOM 2018), Abu Dhabi, UAE, Dec. 2018.

[C5-19] R. Bolla, R. Bruschi, F. Davoli, J. F. Pajo, “A Model-Based Approach Towards Real-Time Analytics in NFV Infrastructures,” IEEE Transactions on Green Communications and Networking, published online 20 Dec. 2019; DOI: 10.1109/TGCN.2019.2961192.

[C5-20] Castro, P., Ishakian, V., Muthusamy, V., & Slominski, A. (2019). The Rise of Serverless Computing. Commun. ACM, 62(12), 44–54.

[C5-21] L. Baresi, D. F. Mendonça, M. Garriga, S. Guinea, and G. Quattrocchi. 2019. A Unified Model for the Mobile-Edge-Cloud Continuum. ACM Trans. Internet Technol. 19, 2, Article 29 (March 2019), 21 pages.

[C5-22] Aditya, P., Akkus, I. E., Beck, A., Chen, R., Hilt, V., Rimac, I., … Stein, M. (2019). Will Serverless Computing Revolutionize NFV? Proceedings of the IEEE, 107(4), 667–678.

[C5-23] Hellerstein, J. M., Faleiro, J., Gonzalez, J. E., Schleier-Smith, J., Sreekanti, V., Tumanov, A., & Wu, C. (2019). Serverless computing: One step forward, two steps back. CIDR 2019 – 9th Biennial Conference on Innovative Data Systems Research, 3.

[C5-24] R. Zoppoli, M. Sanguineti, G. Gnecco and T. Parisini, “Neural Approximations for Optimal Control and Decision”, Springer, 2019.

[C5-25] Cisco Global Cloud Index: Forecast and Methodology, 20162021 White Paper. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html

[C5-26] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1738-1762, Aug. 2019.

[C5-27] Q. Chen, z. Zheng, C. Hu, D. Wang, and F. Liu, “Data-driven task allocation for multi-task transfer learning on the edge,” in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2019.

[C5-28] R. Shokri and V. Shmatikov, “Privacy-preserving deep learning,” in Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. ACM, 2015, pp. 1310–1321.

[C5-29] H. Kim, J. Park, M. Bennis, and S.-L. Kim, “On-device federated learning via blockchain and its latency analysis,” arXiv:1808.03949, 2018.

[C5-30] S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” in Advances in neural information processing systems, 2015, pp. 1135–1143.

[C5-31] T.-J. Yang, Y.-H. Chen, and V. Sze, “Designing energy-efficient convolutional neural networks using energy-aware pruning,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017.

[C5-32] E. Strubell, A. Ganesh and A. McCallum, “Energy and Policy Considerations for Deep Learning in NLP”, Annual Meeting of the Association for Computational Linguistics (ACL short). Florence, Italy. July 2019.

[C5-33] A. Lacoste, A. Luccioni, V. Schmidt, Victor and T. Dandres, "Quantifying the Carbon Emissions of Machine Learning", arXiv preprint arXiv:1910.09700, 2019

[C5-34] http://www.green-algorithms.org/

[C5-35] Florian Hagenauer, Takamasa Higuchi, Onur Altintas and Falko Dressler, "Efficient Data Handling in Vehicular Micro Clouds," Elsevier Ad Hoc Networks, vol. 91, pp. 101871, August 2019.

[C6-1] N. Rajatheva et al: “White paper on broadband connectivity in 6G,” arXiv preprint: https://arxiv.org/abs/2004.14247, 2020.

[C6-2] H. Shokri-Ghadikolaei, F. Boccardi, C. Fischione, G. Fodor, and M. Zorzi: “Spectrum sharing in mmWave cellular networks via cell association, coordination, and beamforming,” IEEE J. Sel. Areas Commun., Nov. 2016.

[C6-3] S. Lagen, L. Giupponi, S. Goyal, N. Patriciello, B. Bojovic, A. Demir, and M. Beluri: “New radio beam-based access to unlicensed spectrum: Design challenges and solutions,” IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 8-37, March 2020.

[C6-4] K. Sakaguch et al: “Where, when, and how mmWave is used in 5G and beyond,” IEICE Transactions on Electronics, vol. E100-C, no.10, pp.790-808, 2017.

[C6-5] P. Kumari, J. Choi, N. González-Prelcic, R.W. Heath: “IEEE 802.11 ad-based radar: An approach to joint vehicular communication-radar system,” IEEE Transactions on Vehicular Technology, vol. 67, pp. 3012-3027, 2017.

[C6-6] M.S. Elbamby, C. Perfecto, M. Bennis and K. Doppler: “Toward low-latency and ultra-reliable virtual reality,” IEEE Network, vol. 32, pp.78-84, 2018.

[C6-7] Ericsson: Traffic Exploration. https://www.ericsson.com/TET/trafficView/loadBasicEditor.ericsson.

[C6-8] WorldoMeters: Western Europe Population. http://www.worldometers.info/world-population/western-europe-population/.

[C6-9] Cisco: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016 - 2021 (White Paper), [Online]. Available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[C6-10] 3GPP: 3GPP TS 22.261, Service Requirements for the 5G System (Release 15), 2017.

[C6-11] C. Fiandrino, H. Assasa, P. Casari and J. Widmer: “Scaling millimeter-wave networks to high density deployments and dynamic environments,” Proceedings of the IEEE, 2019.

[C6-12] M.U. Sheikh and J. Lempiäinen: “Macro cellular network transition from traditional frequency range to 28 GHz millimeter wave frequency band,” Wireless Networks, 24(3), pp. 943-953, 2018.

[C6-13] M.S. Islim, R.X. Ferreira, X. He, E. Xie, S. Videv, S. Viola, S. Watson, N. Bamiedakis, R. V. Penty, I.H. White, A.E. Kelly, E. Gu, H. Haas and M. D. Dawson: “Towards 10 Gb/s OFDM-based visible light communication using a GaN violet micro-LED”, Photonics Research, vol. 2, no. 5, pp.: A35-A48, 2017.

[C6-14] T. Cogalan and H. Haas: “Why would 5G need optical wireless communications?” in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1-6, 2017.

[C6-15] Y. F. Huang et al: “17.6-Gbps universal filtered multi-carrier encoding of GaN blue LD for visible light communication,” in Proc. of the *Conference on Lasers and Electro-Optics (CLEO)*, pp. 1-2, 2017.

[C6-16] J. Fakidis, S. Videv, H. Helmers and H. Haas: “0.5-Gb/s OFDM-based laser data and power transfer using a GaAs photovoltaic cell,” *IEEE Photonics Technology Letters*, 2018.

[C6-17] E. Calvanese Strinati et al: “6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication,” IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 42-50, Sept. 2019.

[C6-18] H. Haas, L. Yin, Y. Wang and C. Chen: “What is LiFi?” *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1533-1544, March, 2016.

[C6-19] D. Tsonev, S. Videv and H. Haas: “Unlocking spectral efficiency in intensity modulation and direct detection systems,” IEEE Journal on Selected Areas in Communications, vol. 33, no. 9, pp. 1758-1770, Sept. 2015.

[C6-20] I. F. Akyildiz, J. M. Jornet and C. Han: “Terahertz band: Next frontier for wireless communications,” Physical Communication (Elsevier) Journal, vol. 12, pp. 16–32, 2014.

[C6-21] S. Priebe, D. M. Britz, M. Jacob, S. Sarkozy, K. M. Leong, J. E. Logan, B. S. Gorospe, and T. Kurner, “Interference investigations of active communications and passive earth exploration services in the Thz frequency range,” IEEE Trans. THz Sci. and Technol., vol. 2, no. 5, pp. 525–537, 2012.

[C6-22] A. Nikpaik, A. H. M. Shirazi, A. Nabavi, S. Mirabbasi and S. Shekhar: “A 219-to-231 GHz frequency-multiplier-based VCO with 3% peak DC-to-RF efficiency in 65-nm CMOS,” IEEE Journal of Solid-State Circuits, vol. 53, no. 2, pp. 389–403, 2018.

[C6-23] H. Aghasi, A. Cathelin and E. Afshari: “A 0.92-THz SiGe power radiator based on a nonlinear theory for harmonic generation,” IEEE Journal of Solid-State Circuits, vol. 52, no. 2, pp. 406–422, 2017.

[C6-24] W. R. Deal, K. Leong, A. Zamora, B. Gorospe, K. Nguyen and X. B. Mei: “A 660 GHz up-converter for THz communications,” in Proc. of the IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS), pp. 1–4, 2017.

[C6-25] H.-J. Song, K. Ajito, Y. Muramoto, A. Wakatsuki, T. Nagatsuma and N. Kukutsu: “Uni-travelling-carrier photodiode module generating 300 GHz power greater than 1 mW,” IEEE Microwave and Wireless Components Letters, vol. 22, no. 7, pp. 363–365, 2012.

[C6-26] S.-W. Huang, J. Yang, S.-H. Yang, M. Yu, D.-L. Kwong, T. Zelevinsky, M. Jarrahi and C. W. Wong: “Globally stable microresonator turing pattern formation for coherent high-power THz radiation on-chip,” Physical Review X, vol. 7, no. 4, p. 041002, 2017.

[C6-27] T. Nagatsuma, G. Ducournau and C. C. Renaud: “Advances in terahertz communications accelerated by photonics,” Nature Photonics, vol. 10, no. 6, p. 371, 2016.

[C6-28] Q. Lu, D. Wu, S. Sengupta, S. Slivken and M. Razeghi: “Room temperature continuous wave, monolithic tunable THz sources based on highly efficient mid-infrared quantum cascade lasers,” Scientific reports, vol. 6, 2016.

[C6-29] A. C. Ferrari, et al: “Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems,” Nanoscale, vol. 7, no. 11, pp. 4598–4810, 2015.

[C6-30] I. F. Akyildiz and J. M. Jornet: “Graphene-based plasmonic nano-transceiver employing HEMT for terahertz band communication,” U.S. Patent No. 9,397,758 issued on July 19, 2016 (Priority Date: December 6, 2013).

[C6-31] J. M. Jornet and I. F. Akyildiz: “Graphene-based plasmonic nano-transceiver for terahertz band communication,” in Proc. of the 8th European Conference on Antennas and Propagation (EuCAP), The Hague, The Netherlands, April 2014.

[C6-32] I. F. Akyildiz and J. M. Jornet: “Graphene-based plasmonic nano-antenna for terahertz band communication,” U.S. Patent No. 9,643,841, issued on May 9, 2017 (Priority Date: April 17, 2013).

[C6-33] J. M. Jornet and I. F. Akyildiz: “Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks,” IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Emerging Technologies in Communications, vol. 31, no. 12, pp. 685-694, Dec. 2013.

[C6-34] J. M. Jornet and I. F. Akyildiz: “Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band,” IEEE Transactions on Wireless Communications, vol. 10, no. 10, pp. 3211-3221, October 2011.

[C6-35] C. Han, A.O. Bicen and I. Akyildiz: “Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band,” IEEE Transactions on Wireless Communications, vol. 14, no. 5, pp. 2402–2412, 2015.

[C6-36] K. Guan et al: “On millimeter wave and THz mobile radio channel for smart rail mobility,” IEEE Transactions on Vehicular Technology, vol. 66, no. 7, pp. 5658–5674, 2017.

[C6-37] J. M. Jornet and I. F. Akyildiz: “Femtosecond-long pulse-based modulation for terahertz band communication in nanonetworks,” IEEE Transactions on Communications, vol. 62, no. 5, pp. 1742-1754, May 2014.

[C6-38] J. Dore et al: “Above-90 GHz spectrum and single-carrier waveform as enablers for efficient Tbit/s wireless communications,” 25th International Conference on Telecommunications (ICT), St. Malo, 2018.

[C6-39] C. Han and I. F. Akyildiz: “Distance-aware bandwidth-adaptive resource allocation for wireless systems in the THz band,” IEEE Transactions on Terahertz Science and Technology, vol. 64, no. 5, pp. 2130-2142, May 2016.

[C6-40] T. Kürner and S. Priebe: “Towards THz communications - Status in research, standardization and regulation,” Journal of Infrared, Millimeter, and Terahertz Waves, vol. 35, no. 1, pp. 53–62, 2014.

[C6-41] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath: “Spatially sparse precoding in millimeter wave MIMO systems,” IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499–1513, 2014.

[C6-42] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors and F. Tufvesson: “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40–60, 2013.

[C6-43] E. Torkildson, U. Madhow and M. Rodwell: “Indoor millimeter wave MIMO: Feasibility and performance,” IEEE Transactions on Wireless Communications, vol. 10, no. 12, pp. 4150–4160, 2011.

[C6-44] M. T. Ivrlač and J. A. Nossek: “Toward a circuit theory of communication,” IEEE Transactions on Circuits and Systems I, vol. 57, no. 7, pp. 1663-1683, July 2010.

[C6-45] T. Laas, J. A. Nossek, S. Bazzi and W. Xu: “On reciprocity in physically consistent TDD systems with coupled antennas,” arXiv:1907.10562v1, 2019.

[C6-46] E. Bjornson, E. G. Larsson and T. L. Marzetta: “Massive MIMO: Ten myths and one critical question,” IEEE Communications Magazine, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[C6-47] A. L. Swindlehurst, E. Ayanoglu, P. Heydari and F. Capolino: “Millimeter-wave massive MIMO: the next wireless revolution?” IEEE Communications Magazine, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[C6-48] I. F. Akyildiz and J. M. Jornet: “Realizing ultra-massive MIMO communication in the (0.06-10) terahertz band,” Nano Communication Networks (Elsevier) Journal, vol. 8, pp. 46-54, March 2016.

[C6-49] I. F. Akyildiz and J. M. Jornet: “Ultra massive MIMO communication in the terahertz band,” U.S. Patent No. 9,825,712, issued on November 21, 2017 (Priority Date: December 6, 2013).

[C6-50] O. El Ayach, R. W. Heath, S. Rajagopal and Z. Pi: “Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays,” in IEEE Globecom, pp. 3476–3480, 2013.

[C6-51] X. Huang, Y. J. Guo and J. D. Bunton: “A hybrid adaptive antenna array,” IEEE Transactions on Wireless Communications, vol. 9, no. 5, pp. 1770-1779, 2010.

[C6-52] S.-H. Wu, L.-K. Chiu, K.-Y. Lin and T.-H. Chang: “Robust hybrid beamforming with phased antenna arrays for downlink SDMA in indoor 60 GHz channels,” IEEE Transactions on Wireless Communications, vol. 12, no. 9, pp. 4542–4557, 2013.

[C6-53] C. Lin and G. Li, “Indoor terahertz communications: How many antenna arrays are needed?” IEEE Transactions on Wireless Communications, vol. 14, no. 6, pp. 3097–3107, 2015.

[C6-54] C. Lin and G. Li: “Adaptive beamforming with resource allocation for distance-aware multi-user indoor terahertz communications,” IEEE Transactions on Communications, vol. 63, no. 8, pp. 2985–2995, Aug. 2015.

[C6-55] Q. Wu and R. Zhang: “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming,” IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.

[C6-56] C. Huang, A. Zappone, M. Debbah and C. Yuen: “Achievable rate maximization by passive intelligent mirrors,” in Proc. IEEE ICASSP, 2018.

[C6-57] S. Hu, F. Rusek and O. Edfors: “Beyond massive MIMO: The potential of data transmission with large intelligent surfaces,” IEEE Transactions on Signal Processing, vol. 66, no. 10, pp. 2746–2758, May 2018.

[C6-58] S. Hu, F. Rusek and O. Edfors: “Beyond massive MIMO: The potential of positioning with large intelligent surfaces,” IEEE Transactions on Signal Processing, vol. 66, no. 7, pp. 1761–1774, April 2018.

[C6-59] P. Frenger, J. Hederen, M. Hessler and G. Interdonato: “Improved antenna arrangement for distributed massive MIMO,” Patent application WO2018103897, 2017.

[C6-60] G. Interdonato, E. Björnson, H. Ngo, P. Frenger and E. Larsson: “Ubiquitous cell-free massive MIMO communications,” EURASIP, 2019.

[C6-61] J. Guerreiro, R. Dinis and P. Carvalho: “On the optimum multicarrier performance with memoryless nonlinearities,” IEEE Transactions on Communications, Vol. 63, No. 2, pp. 498 - 509, February 2015.

[C6-62] J. Guerreiro, R. Dinis, P. Carvalho and M. Silva: “On the achievable performance of nonlinear MIMO systems,” IEEE Communications Letters, Vol. 23, No. 10, pp. 1725 - 1729, October 2019.

[C6-63] P. Pedrosa, R. Dinis, D. Castanheira, A. Silva and A. Gameiro: “Joint Channel Equalization and Trackingfor V2X Communications Using SC-FDE Schemes,” IEEE GLOBECOM, 2019

[C6-64] S. Zhidkov and R. Dinis: “Belief propagation receivers for near-optimal detection of nonlinearly distorted OFDM signals,” IEEE VTC’2019 (Spring), 2019.

[C6-65] J. Félix, J. Guerreiro, R. Dinis and P. Carvalho: “Reduced-complexity quasi-optimum detection for MIMO-OFDM signals with strong nonlinear distortion,” IEEE GLOBECOM 2019.

[C6-66] T. L. Marzetta: “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[C6-67] E. G. Larsson, O. Edfors, F. Tufvesson and T. L. Marzetta: “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[C6-68] P. Marsch, S. Brück, A. Garavaglia, M. Schulist, R. Weber, and A. Dekorsy: “Clustering,” in *Coordinated Multi-point in Mobile Communications: From theory to Practice*, Ed. by P. Marsch, G. Fettweis, Cambridge University Press, New York, 2011, pp. 139–159.

[C6-69] R. Fantini, W. Zirwas, L. Thiele, D. Aziz, P. Baracca: “Coordinated multi-point transmission in 5G,” in *5G Mobile and Wireless Communications Technology*, Ed. by A. Osseiran, J. Monserrat, and P.Marsch, Cambridge University Press, Cambridge, 2016, pp. 248–276.

[C6-70] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta: “Cell-free massive MIMO versus small cells,” IEEE Trans. Wireless Commun., vol. 16, no. 3, pp. 1834–1850, 2017.

[C6-71] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai and L. Hanzo: “Cell-Free massive MIMO: A new next-generation paradigm,” IEEE Access, vol. 7, pp. 99878-99888, 2019.

[C6-72] 3GPP TR 38.801, Study on New Radio Access Technology (Release 14), 2017.

[C6-73] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago and F. Wiedmann: “5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications,” IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, Feb. 2014.

[C6-74] Y. Medjahdi, S. Traverso, R. Gerzaguet, H. Shaiek, M. B. Mabrouk, D. L. Ruyet and a. D. R. Y. Louet: “On the road to 5G: Comparative study of physical layer in MTC context,” IEEE Access, vol. 5, pp. 26556-26581, 2017.

[C6-75] Y. Tao, L. Liu, S. Liu and Z. Zhang: “A survey: Several technologies of non-orthogonal transmission for 5G,” China Communications, vol. 12, no. 10, pp. 1-15, Oct. 2015.

[C6-76] R. Gerzaguet, N. Bartzoudis, L. G. Baltar, V. Berg, J.-B. Doré, D. Kténas, O. Font-Bach, X. Mestre, M. Payaró, M. Färber and K. Roth: “The 5G candidate waveform race: A comparison of complexity and performance,” EURASIP Journal on Wireless Communications and Networking, no. 13, 2017.

[C6-77] M. van Eeckhaute, A. Bourdoux, P. de Doncker and F. Horlin: “Performance of emerging multi-carrier waveforms for 5G asynchronous communications,” EURASIP Journal on Wireless Communications and Networking, 2017.

[C6-78] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch and R. Calderbank: “Orthogonal time frequency space modulation,” Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), 2017.

[C6-79] A. Fish, S. Gurevich, R. Hadani, A. M. Sayeed, and O. Schwartz: “Delay-Doppler channel estimation in almost linear complexity,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7632–7644, Nov. 2013.

[C6-80] S. C. Thompson, A. U. Ahmed, J. G. Proakis, J. R. Zeidler and M. J. Geile: “Constant envelope OFDM,” IEEE Trans. Commun., vol. 56, no. 8, pp. 1300-1312, Aug. 2008.

[C6-81] A. U. Ahmed and J. R. Zeidler: “Novel low-complexity receivers for constant envelope OFDM,” IEEE Trans. Signal Process., vol. 63, no. 17, pp. 4572–4582, Sep. 2015.

[C6-82] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo: “A survey of non-orthogonal multiple access for 5G,” IEEE Communications Surveys & Tutorials, 20(3), 2294-2323, third quarter 2018.

[C6-83] L. Zhu, J. Zhang, Z. Xiao, X. Cao and D. O. Wu: “Optimal user pairing for downlink non-orthogonal multiple access (NOMA),” IEEE Wireless Communications Letters, vol. 8, no. 2, pp. 328-331, April 2019.

[C6-84] G. Song and X. Wang: “Comparison of interference cancellation schemes for non-orthogonal multiple access system,” IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, pp. 1-5, 2016.

[C6-85] H. V. Nguyen, V. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz and O. Shin: “Joint power control and user association for NOMA-based full-duplex systems,” IEEE Transactions on Communications, vol. 67, no. 11, pp. 8037-8055, Nov. 2019.

[C6-86] W. Zhang, J. Chen, Y. Kuo and Y. Zhou: “Artificial-noise-aided optimal beamforming in layered physical layer security,” IEEE Communications Letters, vol. 23, no. 1, pp. 72-75, Jan. 2019.

[C6-87] D. Kim, H. Lee and D. Hong, “A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers,” IEEE Communications Surveys & Tutorials, vol. 17, no. 4, 2017-2046, fourth quarter 2015.

[C6-88] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman: “In-band full-duplex wireless: Challenges and opportunities,” IEEE J. Sel. Areas Commun., vol. 32, no. 9, pp. 1637-1652, Sept 2014.

[C6-89] J. Berkmann, C. Carbonelli, F. Dietrich, C. Drewes and W. Xu: “On 3G LTE terminal implementation – Standard, algorithms, complexities and challenges (Invited Paper),” IEEE International Wireless Communications and Mobile Computing Conference, pp. 970-975, Crete, Greece, Aug., 2008.

[C6-90] A. Elkelesh, M. Ebada, S. Cammerer and S. ten Brink: “Belief propagation list decoding of polar codes,” IEEE Communications Letters, vol. 22, no. 8, pp. 1536-1539, Aug. 2018.

[C6-91] V. Ranasinghe, N. Rajatheva and M. Latva-aho: “Partially permuted multi-trellis belief propagation for polar codes,” IEEE ICC 2020, arXiv preprint arXiv:1911.08868.

[C6-92] A. Balatsoukas-Stimming and C. Studer, “Deep unfolding for communications systems: A survey and some new directions,” arXiv preprint arXiv:1906.05774, 2019.

[C6-93] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein and Y. Be’ery: “Deep learning methods for improved decoding of linear codes,” IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 119-131, Feb. 2018.

[C6-94] F. R. Kschischang and S. Pasupathy: “Optimal nonuniform signaling for Gaussian channels,” IEEE Transactions on Information Theory, 39(3), pp. 913-929, 1993.

[C6-95] N. S. Loghin, J. Zöllner, B. Mouhouche, D. Ansorregui, J. Kim and S.I. Park: “Non-uniform constellations for ATSC 3.0,” IEEE Transactions on Broadcasting, 62(1), 197-203, 2016.

[C6-96] G. Boecherer, F. Steiner and P. Schulte: “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” IEEE Transactions on Communications, 63(12), pp. 4651-4665, 2015.

[C6-97] M. Pikus and W. Xu: “Bit-level probabilistically shaped coded modulation,” IEEE Communications Letters, vol. 21, no. 9, pp. 1929–1932, Sept. 2017.

[C6-98] T. Prinz, P. Yuan, G. Boecherer, F. Steiner, O. Iscan, R. Boehnke and W. Xu: “Polar coded probabilistic amplitude shaping for short packets,” in Proc. IEEE SPAWC, 2017.

[C6-99] P. Schulte and F. Steiner: “Shell mapping for distribution matching,” arXiv:1803.03614, 2018.

[C6-100] W. Xu, M. Huang, C. Zhu and A. Dammann: “Maximum likelihood TOA and OTDOA estimation with first arriving path detection for 3GPP LTE system,” Transactions on Emerging Telecommunications Technologies (ETT), vol. 27, no.3, pp. 339-356, 2016.

[C6-101] H. Wymeersch, G. Seco-Granados, G. Destino, et al: “5G mmWave positioning for vehicular networks,” IEEE Wireless Communications, pp. 80–86, Dec. 2017.

[C6-102] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich and M. Coulon: “Direct localization for massive MIMO,” IEEE Transactions on Signal Processing, vol. 65, no. 10, pp. 2475-2487, May 2017.

[C6-103] M. Kiviranta, I. Moilanen, and J. Roivainen: “5G radar: Scenarios, numerology and simulations,” 2019 IEEE International Conference on Military Communications and Information Systems (ICMCIS), 2019.

[C6-104] H. Kim et al: “5G mmWave cooperative positioning and mapping using multi-model PHD filter and map fusion,” arXiv preprint arXiv:1908.09806, 2019.

[C6-105] R. Koirala et al: “Localization and throughput trade-off in a multi-user multi-carrier mm-wave system,” IEEE Access, vol. 7, pp. 167099-167112, 2019.

[C6-106] C. Aydogdu, G. K. Carvajal, O. Eriksson, H. Hellsten, H. Herbertsson, M. F. Keskin, E. Nilsson, M. Rydström, K. Vanäs and H. Wymeersch: “Radar interference mitigation for automated driving,” submitted to IEEE Signal Processing Magazine, 2019.

[C6-107] J. Khoury, R. Ramanathan, D. McCloskey, R. Smith and T. Campbell: “RadarMAC: Mitigating radar interference in self-driving cars,” 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), London, pp. 1-9, 2016.

[C6-108] A. Dammann, R. Raulefs, S. Zhang: “On prospects of positioning in 5G,” in Proc. IEEE International Conference on Communication Workshop (ICCW), 2015.

[C6-109] W. Xu, A. Dammann and T. Laas: “Where are the things of the internet? Precise time of arrival estimation for IoT positioning,” in ‘The fifth Generation (5G) of Wireless Communication’, A. Kishk, Ed., Intechopen, 2018.

[C6-110] R. di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson and H. Wymeersch: “Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G,” IEEE Signal Processing Magazine, vol. 31, no. 6, pp. 102-112, Nov. 2014.

[C6-111] Y. Polyanskiy: “A perspective on massive random-access,” in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2523-2527, 2017.

[C6-112] H.A. Inan, P. Kairouz and A. Ozgur: “Sparse combinatorial group testing for low-energy massive random access,” arXiv preprint arXiv:1711.05403, 2017.

[C6-113] J. Luo and D. Guo: “Neighbor discovery in wireless ad hoc networks based on group testing,” In 46th Annual Allerton Conference on Communication, Control, and Computing, pp. 791-797, Sep. 2008.

[C6-114] E. Paolini, C. Stefanovic, G. Liva and P. Popovski: “Coded random access: applying codes on graphs to design random access protocols,” IEEE Communications Magazine 53, no. 6, pp. 144-150, 2015.

[C6-115] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli and T. H. Jacobsen: “Uplink grant-free access solutions for URLLC services in 5G New Radio,” 16th International Symposium on Wireless Communication Systems (ISWCS), Oulu, Finland, pp. 607-612, 2019.

[C6-116] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic and P. Popovski: “From 5G to 6G: Has the time for modern random access come?” arXiv: 1903.03063, 2019.

[C6-117] S. Rangan: “Generalized approximate message passing for estimation with random linear mixing,” in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2168-2172, 2011.

[C6-118] Z. Chen, F. Sohrabi and W. Yu: “Sparse activity detection for massive connectivity,” arXiv preprint arXiv:1801.05873, 2018.

[C6-119] S. Haghighatshoar, P. Jung and G. Caire: “Improved scaling law for activity detection in massive MIMO systems,” arXiv preprint arXiv:1803.02288, 2018.

[C6-120] C. Bockelmann et al: “Massive machine-type communications in 5G: Physical and MAC-layer solutions,” IEEE Communications Magazine, vol. 54, no. 9, pp. 59-65, September 2016.

[C6-121] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic and E. de Carvalho: “Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things,” IEEE Signal Processing Magazine, vol. 35, no. 5, pp. 88-99, Sept. 2018.

[C6-122] E. Bjornson, E. de Carvalho, J. H. Sørensen, E. G. Larsson and P. Popovski, “A Random Access Protocol for Pilot Allocation in Crowded Massive MIMO Systems,” IEEE Transactions on Wireless Communications, vol. 16, no. 4, pp. 2220–2234, April 2017.

[C6-123] G. Fettweis, H. Boche, et al: “The tactile internet,” ITU-T Technology Watch Report, August 2014.

[C6-124] H. Lasi, P. Fettke, HG. Kemper, et al: “Industry 4.0”, Bus Inf Syst Eng, 6: 239, 2014. https://doi.org/10.1007/s12599-014-0334-4.

[C6-125] C. Kalalas and J. Alonso-Zarate: “Massive connectivity in 5G and beyond: Technical enablers for the energy and automotive verticals,” in Proc. of 6G Wireless Summit 2020, Levi, Finland, March 2020.

[C6-126] Ericsson Expert Analytics. [Online]. Available: https://www.ericsson.com/ourportfolio/products/expert-analytics.

[C6-127] M. Agiwal, A. Roy and N. Saxena: “Next generation 5G wireless networks: A comprehensive survey,” IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617-1655, 2016.

[C6-128] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang and L. Zhong: “Argos: Practical many-antenna base stations,” in Proc. of the 18th ACM annual international conference on Mobile computing and networking, pp. 53-64, 2012.

[C6-129] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors and F. Tufvesson: “A exible 100-antenna testbed for massive MIMO,” in Proc. IEEE Globecom Workshops, pp. 287-293, 2014.

[C6-130] S.-W. Jeon, S.-N. Hong, M. Ji, G. Caire and A. F. Molisch: “Wireless multihop device-to-device caching networks,” IEEE Transactions on Information Theory, vol. 63, no. 3, pp. 1662-1676, 2017.

[C6-131] M. Ji, G. Caire and A. F. Molisch: “Fundamental limits of caching in wireless D2D networks,” IEEE Transactions on Information Theory, vol. 62, no. 2, pp. 849-869, 2016.

[C6-132] M. Maddah-Ali and U. Niesen: “Fundamental limits of caching,” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856-2867, May 2014.

[C6-133] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch and G. Caire: “Femtocaching: Wireless content delivery through distributed caching helpers,” IEEE Transactions on Information Theory, vol. 59, no. 12, pp. 8402-8413, 2013.

[C6-134] M. Bayat, R. K. Mungara and G. Caire: “Achieving spatial scalability for coded caching over wireless networks,” arXiv preprint: arXiv:1803.05702, 2018.

[C6-135] K. Wan and G. Caire: “On coded caching with private demands”, submitted to TIT, available online: arXiv:1908.10821 [cs.IT], 2019.

[C6-136] D. Bethanabhotla, G. Caire and M. J. Neely: “Wiflix: Adaptive video streaming in massive MU-MIMO wireless networks,” IEEE Transactions on Wireless Communications, vol. 15, no. 6, pp. 4088–103, 2016.

[C6-137] A. Narayanan, S. Verma, E. Ramadan, P. Babaie, and Z. Zhang: “DeepCache: A deep learning based framework for content caching,” in Proceedings of Workshop on Network Meets AI & ML, pp. 48–53, Aug. 2018.

[C7-1] Photonics21, Vision Paper “Europe’s age of light! How photonics will power growth and innovation”, Nov 2017, online

[C7-2] P.J. Winzer, D.T. Neilson, “From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade,” J. Lightw. Technol., 35(2017)5, Feb. 2017, pp. 1099-1115.

[C7-3] S. Wolf et al., “DAC-Less Amplifier-Less Generation and Transmission of QAM Signals Using Sub-Volt Silicon-Organic Hybrid Modulators,“ J. Lightw. Technol., 33(2015)7, April 2015, pp. 1425-1432.

[C7-4] https://www.etsi.org/committee/f5g ETSI Industry Specification Group (ISG) Fifth Generation Fixed Network (F5G).

[C8-1] https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive

[C8-2] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0477R(02)

[C8-3] https://ec.europa.eu/digital-single-market/en/news/cybersecurity-5g-networks-eu-toolbox-risk-mitigating-measures

[C8-4] https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks

[C8-5] NGMN Alliance: ”Description of Network Slicing Concept”, Version 1.0, January 2016

[C8-6] NGMN Alliance: ”5G Security Recommendations Package #2: Network Slicing” Version 1.0, April 2016

[C8-7] 3GPP: System Architecture for the 5G System. 3GPP-23501, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[C8-8] 3GPP: Procedures for the 5G System. 3GPP-23502, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3145

[C8-9] 3GPP::”Study on the security aspects of the next generation system”. TR 33.899, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3045

[C8-10] 3GPP: ”Security architecture and procedures for 5G System”. TS 33.501, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169

[C8-11] 5G PPP: Phase 1 Security Landscape, https://5g-ppp.eu/white-papers/.

[C8-12] Scott-Hayward, S., Natarajan, S., & Sezer, S.: (2016). A Survey of Security in Software Defined Networks, IEEE Communications Surveys and Tutorials.

[C8-13] Olli Mämmelä, Jouni Hiltunen, Jani Suomalainen, Kimmo Ahola, Petteri Mannersalo, Janne Vehkaperä: ”Towards Micro-Segmentation in 5G Network Security”. EuCNC 2016.

[C8-14] European Commission, 2013, IoT Privacy, Data Protection, Information Security. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=1753.[C8-15]J.L. Hernandez Ramos, A. Skarmeta, (Eds), Security and Privacy in Internet of Things - Challenges and Solutions, within the series: Ambient Intelligence and Smart Environments, (Introduction by E. Del Re), IOS Press, 2020

[C9-0] I. F. Akyildiz and A. Kak, "The Internet of Space Things/CubeSats," in IEEE Network, vol. 33, no. 5, pp. 212-218, Sept.-Oct. 2019, doi: 10.1109/MNET.2019.1800445.

[C9-1] S. C. Burleigh, T. De Cola, S. Morosi, S. Jayousi, E. Cianca, and C. Fuchs, “From Connectivity to Advanced Internet Services: A Comprehensive Review of Small Satellites Communications and Networks,” Wireless Communications and Mobile Computing, vol.2019, pp.1–17, may2019. [Online]. Available:https://www.hindawi.com/journals/wcmc/2019/6243505.

[C9-2] A. Anttonen, P. Ruska, M. Kiviranta. 3GPP non-terrestrial networks: a concise review and look ahead. VTT Research Report nr. VTT-R-00079-19. 01.01.2019. Available at https://cris.vtt.fi/en/publications/3gpp-nonterrestrial-networks-a-concise-review-and-look-ahead.

[C9-3] Y. Wang, J. Yang, X. Guo, and Z. Qu, “Satellite Edge Computing for the Internet of Things in Aerospace,” Sensors, vol. 19, no. 20, p. 4375, oct 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/20/4375

[C9-4] V. Cerf, S. Burleigh, A. Hooke, R. Durst, K. Scott, K. Fall, H. Weiss. Delay Tolerant Networking Architecture. RFC 4838 (Informational). April 2007.

[C9-5] Satellite interference... Have we fixed the problem? http://www.satmagazine.com/story.php?number=1624153681 Retrieved November 2019.

[C9-6] Sreeraj Rajendran, Wannes Meert, Vincent Lenders, Sofie Pollin: Unsupervised Wireless Spectrum Anomaly Detection With Interpretable Features. IEEE Trans. Cogn. Comm. & Networking 5(3): 637-647 (2019).

[C9-7] M. Á. Vázquez, P. Henarejos, J. C. Gil, I. Parparaldo, A. Pérez-Neira, Artificial Intelligence for SATCOM operations, in Proceedings of SpaceOps Operation 22-25 May 2020, Cape Town (South Africa).

[C9-8] M. Á. Vázquez, P. Henarejos, A. Pérez-Neira, E. Grechi, A. Voight, J. C. Gil, I. Pappalardo, On the use of AI for satellite communications, IEEE SSC Newsletter December 2019.

[C9-10] Yoshua Bengio, Andrea Lodi, Antoine Prouvost: Machine Learning for Combinatorial Optimization: a Methodological Tour d'Horizon. CoRR abs/1811.06128 (2018)

[C9-11] Björnson, E., & Giselsson, P. (2020). Two Applications of Deep Learning in the Physical Layer of Communication Systems. arXiv preprint arXiv:2001.03350.

[C9-12] Networld2020 “SatCom Resources for Smart and Sustainable Networks and Services” Nov. 2019, available at https://www.networld2020.eu/wp-content/uploads/2019/11/satcom-ss-whitepaper_v1_final.pdf

[C9-13] A. Vanelli-Coralli, G. E. Corazza, M. Luglio and S. Cioni, "The ISICOM Architecture," 2009 International Workshop on Satellite and Space Communications, Tuscany, 2009, pp. 104-108, doi: 10.1109/IWSSC.2009.5286409.

[C9-14] IEEE Future Networks Initiative, “International Network Generations Roadmap (INGR), An IEEE 5G and Beyond Technology Roadmap - Satellite,” 2019. [Online]. Available: https://futurenetworks.ieee.org/roadmap.

[C9-15] C. A. Hofmann and A. Knopp, "Ultranarrowband Waveform for IoT Direct Random Multiple Access to GEO Satellites," in IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10134-10149, Dec. 2019

[C10-1] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, “Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond,” IEEE Access, vol. 7, pp. 78729–78757, 2019.

[C10-2] J.-B. Dore, Y. Corre, S. Bicais, J. Palicot, E. Faussurier, D. Kt´ enas,´ and F. Bader, “Above-90GHz Spectrum and Single-Carrier Waveform as Enablers for Efficient Tbit/s Wireless Communications,” in 25th International Conference on Telecommunications (ICT’2018), SaintMalo, France, Jun. 2018.

[C10-3] S. Bicaïs and J.-B. Dore, “Phase Noise Model Selection for Sub-THz Communications,” in 2019 IEEE Global Communication Conference (GLOBECOM), December 2019.

[C10-4] S. Bicaïs, J.-B. Dore, G. Gougeon, and Y. Corre, “Optimized Single Carrier Transceiver for Future Sub-TeraHertz Applications”, in International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020.

[C10-5] “IEEE standard for high data rate wireless multi-media networks– amendment 2: 100 Gb/s wireless switched point-to-point physical layer,” IEEE Std 802.15.3d-2017 (Amendment to IEEE Std 802.15.3-2016 as amended by IEEE Std 802.15.3e-2017), pp. 1–55, Oct 2017.

[C10-6] M. Saad, F. Bader, J. Palicot, A. C. A. Ghouwayel, and H. Hijazi, “Single carrier with index modulation for low power terabit systems”, in 2019 IEEE Wireless Communications and Networking Conference (WCNC), April 2019, pp. 1–7.

[C10-7] NEREID Roadmap: https://www.nereid-h2020.eu/roadmap

[C10-8] Simoens F., Meilhan J., Nicolas J.-A., “Terahertz Real-Time Imaging Uncooled Arrays Based on Antenna-Coupled Bolometers or FET Developed at CEA-LETI”, J. Infrared, Millimeter, Terahertz Waves. 2015; 36:961–985. doi: 10.1007/s10762-015-0197-x

[C10-9] A. Visweswaran, K. Vaesen, S. Sinha, I. Ocket, M. Glassee, C. Desset, A. Bourdoux, and P. Wambacq, “A 145GHz FMCW Radar Transceiver in 28nm CMOS”, in 2019 IEEE International Solid- State Circuits Conference - (ISSCC).

[C10-10] PA survey: https://gems.ece.gatech.edu/PA_survey.html

[C10-11] DOTSEVEN European project: http://www.dotseven.eu

[C10-12] P. Chevalier et al., "A 55 nm triple gate oxide 9 metal layers SiGe BiCMOS technology featuring 320 GHz fT / 370 GHz fMAX HBT and high-Q millimeter-wave passives", in 2014 IEEE International Electron Devices Meeting (IEDM).

[C10-13] A. Vais et al., "First demonstration of III-V HBTs on 300 mm Si substrates using nano-ridge engineering », in 2019 IEEE International Electron Devices Meeting (IEDM).

[C10-14] M. Schroter et al., "Physical and Electrical Performance Limits of High-Speed SiGeC HBTs—Part I: Vertical Scaling", IEEE Trans. Electron Devices, Vol. 58, No. 11, Nov. 2011.

[C10-15] U. Peralagu, "CMOS-compatible GaN-based devices on 200mm-Si for RF applications: Integration and Performance", in 2019 IEEE International Electron Devices Meeting (IEDM).

[C10-16] P. Weckx et al., "Novel forksheet device architecture as ultimate logic scaling device towards 2nm", in 2019 IEEE International Electron Devices Meeting (IEDM).

[C10-17] P. Huang, P. Mercier, "A 220µW -85dBm Sensitivity BLE-Compliant Wake-up Receiver Achieving -60dB SIR via Single-Die Multi- Channel FBAR-Based Filtering and a 4-Dimentional Wake-Up Signature," IEEE ISSCC, 2019.

[C10-18] S. Denis, R.Berkvens, M.Weyn, "A Survey on Detection, Tracking and Identification in Radio Frequency-Based Device-Free Localization", Special Issue Surveys of Sensor Networks and Sensor Systems Deployments, December 2019

[C10-19] P. Zand, J. Romme, J. Govers, F. Pasveer, and G. Dolmans, "A High-Accuracy Phase-Based Ranging Solution with Bluetooth Low Energy (BLE)," in WCNC, 2019

[C10-20] Boer, J. Romme, J. Govers, and G. Dolmans, "Performance of High-Accuracy Phase-Based Ranging in Multipath Environments," in 91th IEEE Vehicular Technology Conference, VTC Spring 2020, Antwerp, Belgium, May 25-28, 2020.

[C10-21] D. Vasisht, S.Kumar, D.Katabi, "Decimeter-level localization with a single Wi-Fi access point", NSDI'16: Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation, March 2016, Pages 165–178

[C10-22] E. Bechthum, J. Dijkhuis, M. Ding, Y. He, J. Van den Heuvel, P. Mateman, G-J. van Schaik, K. Shibata, M. Song, E. Tiurin, S. Traferro, Y-H. Liu, C. Bachmann,"A Low-Power BLE Transceiver with Support for Phase-Based Ranging, Featuring 5µs PLL Locking Time and 5.3ms Ranging Time, Enabled by Staircase-Chirp PLL with Sticky-Lock Channel-Switching" IEEE INTERNATIONAL SOLID-STATE CIRCUITS CONFERENCE, ISSCC 2020, San Francisco

[C10-23] R. Al Hadi, H. Sherry, J. Grzyb, Y. Zhao, W. Forster, H. M. Keller, A. Cathelin, A. Kaiser, and U. R. Pfeiffer. "A 1 k-pixel video camera for 0.7-1.1 terahertz imaging applications in 65-nm CMOS". In: IEEE J. Solid-State Circuits 47.12 (Dec. 2012), pp. 2999–3012.

[C10-24] Zheludev, N. I.; Kivshar, Y. S. From metamaterials to metadevices. Nat. Mater. 2012, 11, 917–924.

[C10-25] Silva, A.; Monticone, F.; Castaldi, G.; Galdi, V.; Alù, A. & Engheta, N. Performing mathematical operations with metamaterials. Science 2014, 343(6167), 160-163.

[C10-26] Zhang, L.; Chen, X. Q.; Liu, S.; Zhang, Q.; Zhao, J.; Dai, J. Y.; Galdi, V. Space-time-coding digital metasurfaces. Nature communications 2018, 9(1), 4334.

[C10-27] C.Y. Wu et.al., “Distributed antenna system using sigma-delta intermediate frequency over fibre for frequency bands above 24GHz”, Journal of Lightwave Technology, Feb. 2020.

[C10-28] H. Ramon, et.al., “A 700mW 4-to-1 SiGe BiCMOS 100GS/s analog time-interleaver”, International Solid-State Circuits Conference (ISSCC), San Francisco, Feb. 2020.

[C10-29] S. Ummethala et al., "Capacitively Coupled Silicon-Organic Hybrid Modulator for 200 Gbit/s PAM-4 Signaling," in Proc. of CLEO: Applications and Technology 2019 (Post Deadline Session), San Jose, CA, US, May 2019

[C10-30] F. Eltes et al., "A BaTiO3-Based Electro-Optic Pockels Modulator Monolithically Integrated on an Advanced Silicon Photonics Platform," Journal of Lightwave Technology, vol. 37, no. 5, pp. 1456 - 1462, Mar. 2019

[C10-31] M. He et al., "High-performance hybrid silicon and lithium niobate Mach–Zehnder modulators for 100 Gbit s−1 and beyond," Nature Photonics, vol. 13, pp. 359–364, Mar. 2019

[C10-32] U. Koch et al., “Monolithic High-Speed Transmitter Enabled by Bicmos-Plasmonic Platform” ECOC 2019, PD.1.4, Sep. 2019.

[C10-33] A. H. Atabaki et al., “Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip”, Nature, vol. 556: no. 7701, pp. 349-354, June 2018.

[C10-34] J. Verbist et al., “Real-time 100 Gb/s NRZ and EDB transmission with a GeSi electroabsorption modulator for short-reach optical interconnects,” J. Lightw. Technol., vol. 36, no. 1, pp. 90–96, Jan. 2018.

[C10-35] P. W. Juodalkis, et.al., “Optically sampled analog-to-digital converters”, IEEE Trans. Microwave Theory and Techniques, vol. 49, Oct. 2001.

[C10-36] A. Gatherer et al, Academic Press Library in Mobile and Wireless Communications, 2016

[C10-37] T. J. O’Shea, J. Hoydis An Introduction to Deep Learning for the Physical Layer, IEEE Transactions on Cognitive Communications and Networking, 2017, pp 563-575.

[C10-38] D. Wu, M. Nekovee, Y. Wang, An Adaptive Deep Learning Algorithm Based Autoencoder for Interference Channel, Proc Second IFIP International Conference on Machine Learning for Networks, MLN2019, Paris December 2019.

[C10-39] E. Grimaldi, V. Krizakova, G. Sala, F. Yasin, S. Couet, G. Kar, K. Garello, P. Gambardella, “Single shot dynamics of spin-orbit torque and spin transfer torque switching in 3-terminal magnetic tunnel junctions”, Nature Nanotechnology 15, 111 (2020)

[C10-40] S. Van Beek, B. O’Sullivan, P.J. Roussel, R. Degraeve, E. Bury, J. Swerts, S. Couet, L. Souriau, S. Kundu, S. Rao, W. Kim, F. Yasin, D. Crotti, D. Linten, G. Kar, “Impact of self-heating on reliability predictions in STT-MRAM”, IEDM 2018

[C10-41] K. Garello, F. Yasin, S. Couet, L. Souriau, J. Swerts, S. Rao, S. Van Beek, W. Kim, E. Liu, S. Kundu, D. Tsvetanova, K. Croes, N. Jossart, E. Grimaldi, M. Baumgartner, D. Crotti, A. Furnémont, P. Gambardella, G.S. Kar, “SOT-MRAM 300mm integration for lo power and ultrafast embedded memories”, VLSI 2018

[C10-42] Y. C. Wu, W. Kim, K. Garello, F. Yasin, G. Jayakumar, S. Couet, R. Carpenter, S. Kundu, S. Rao, D. Crotti, J. Van Houdt, G. Groeseneken, G. S. Kar, “Deterministic and field-free voltage-controlled MRAM”, VLSI 2020

[C10-43] R. Delhougne, A. Arreghini, E. Rosseel, A. Hikavyy, E. Vecchio, L. Zhang, M. Pak, L. Nyns, T. Raymaekers, N. Jossart, L. Breuil, S. S. V-Palayam, C.-L. Tan, G. Van den bosch, A. Furnémont, "First demonstration of monocrystalline silicon macaroni channel for 3-D NAND memory devices", Proc. VLSI Technology Symposium, p. 203 (2018)

[C10-44] D. Verreck, A. Arreghini, J.P. Bastos, F. Schanovsky, F. Mitterbauer, C. Kernstock, M. Karner, R. Degraeve, G. Van den bosch and A. Furnémont, "Quantitative 3-D Model to Explain Large Single Trap Charge Variability in Vertical NAND Memory", 2019 International Electron Device Meeting (IEDM) Tech. Dig., p. 755 (2019)

[C10-45] A. H. Du Nguyen, Y. Yu, M. Abu Lebdeh, M. Taouil, S. Hamdioui and F. Catthoor, "Classification of Memory-Centric Computing," ACM Emerging Technologies in Computing, vol. 16, no. 2, pp. 1-26, 2020.

[C10-46] K. Kim, S. Shin and S. Kang, "Stateful logic pipeline architecture," IEEE International Symposium of Circuits and Systems (ISCAS), pp. 2497-2500, 2011.

[C10-47] L. Xie, A. H. Du Nguyen, J. Yu, A. Kaichouhi, M. Taouil, M. Al-Failakawi and S. Hamdiou, "Scouting logic: A novel memristor-based logic design for resistive computing," IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 335-340, 2017.

[C10-48] NetWorld2020 ETP, "5G: Challenges, research priorities, and recommendations", White Paper, September 2014. Available online at: https://www.networld2020.eu/.

[C10-49] A. Paverd, M. Völp, F. Brasser, M. Schunter, A.-R. Sadeghi, Asokan, P. Verissimo, A. Steiniger and T. Holz, "Sustainable Security & Safety: Challenges and Opportunities.," in CERTS 2019; pp. 4:1-4:13, 2019.

[C10-50] Mirai botnet code. Available online at: https://github.com/jgamblin/Mirai-Source-Code.

[C10-51] M. Jian, Y. Zhao, "A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems," IWCMC, 2020

[C10-52] Higginbotham, Stacey. "The internet of trash [Internet of Everything]." IEEE Spectrum 55.6 (2018): 17-17.

[C10-53] https://www.storaenso.com/en/newsroom/regulatory-and-investor-releases/2018/11/stora-enso-introduces-sustainable-rfid-tag-technology-eco-for-intelligent-packaging[C11-1] D. D. Clark, C. Partridge, J. C. Ramming, J. T. Wroclawski: "A knowledge plane for the internet". ACM SIGCOMM 2003 conference.

[C11-2] Marco Ajmone Marsan, Giuseppe Bianchi, Nicola Blefari Melazzi: Living and Fluid Networks: the way ahead? Computer Communications, to be published.

[C11-3] "Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper

[C11-4] Energy and Policy Considerations for Deep Learning in NLP. Emma Strubell, Ananya Ganesh and Andrew McCallum. Annual Meeting of the Association for Computational Linguistics (ACL short). Florence, Italy. July 2019.

[C11-5] I. Akyildiz and J. Jornet: "The Internet of nano-things". IEEE Wireless Communications, vol. 17, no. 6, pp. 58–63, 2010.

[C11-6] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz: "Using any Surface to Realize a New Paradigm for Wireless Communications". Communications of the ACM (to appear), 2018.

[C11-7] G. Piro, G. Boggia, and L. Grieco: “On the design of an energy-harvesting protocol stack for Body Area Nano-NETworks”. Nano Communication Networks, vol. 6, no. 2, pp. 74–84, 2015.

[C11-8] C. Liaskos et al: “Design and Development of Software Defined Metamaterials for Nanonetworks”. EEE Circuits and Systems Magazine, vol. 15, no. 4, pp. 12–25, 2015.

[C11-9] Edgar A. Aguilar, Ravishankar Ramanathan, Johannes Kofler, and Marcin Pawłowski, “Completely device-independent quantum key distribution” Physical Review A 94, 022305, 2016.

[C11-10] Manuel Erhard, Robert Fickler, Mario Krenn, and Anton Zeilinger, “Twisted photons: new quantum perspectives in high dimensions”, Light: Science & Applications, 7, 17146, 2018.

[C11-11] Massimiliano Smania, Ashraf M Elhassan, Armin Tavakoli, and Mohamed Bourennane, “Experimental quantum multiparty communication protocols”, Nature Photonic Journal Quantum Information, 2, 16010, 2016.

[C11-12] Charles H. Bennett, Gilles Brassard, “Quantum cryptography: Public key distribution and coin tossing”, Theoretical Computer Science, 560,7, 2014.

[C11-13] Wei Zhang, Dong-Sheng Ding, Yu-Bo Sheng, Lan Zhou, Bao-Sen Shi, and Guang-Can Guo, “Quantum Secure Direct Communication with Quantum Memory”, Physical. Review Letters, 118, 220501, 2017.

[C11-14] Donaldson, R. J. et al. “Experimental demonstration of kilometer-range quantum digital signatures”, Physical Review A, 93, 012329, 2016.

[C11-15] Hiroyuki Shibata, Toshimori Honjo, and Kaoru Shimizu, “Quantum key distribution over a 72 dB channel loss using ultralow dark count superconducting single-photon detectors “, Optics Letters 39, 17, 2014.

[C11-16] Sheng-Kai Liao et al.: “Long-distance free-space quantum key distribution in daylight towards inter-satellite communication” Nature Photonics, 11, 509, 2017.

[C11-17] T. Shinada, et al.: “Deterministic doping to silicon and diamond materials for quantum processing” In: Nanotechnology (IEEE-NANO), 2016 IEEE 16th International Conference on. IEEE, 2016. p. 888-890.

[C11-18] Tiecke, T. G., Thompson, J. D., de Leon, N. P., Liu, L. R., Vuletić, V., & Lukin, M. D.: (2014). “Nanophotonic quantum phase switch with a single atom” Nature, 508(7495), 241.

[C11-19] Azuma, K., Tamaki, K., & Lo, H. K.: (2015). “All-photonic quantum repeaters” Nature communications, 6, 6787.

[C11-20] Marandi, A., Wang, Z., Takata, K., Byer, R. L., & Yamamoto, Y.: (2014). Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. Nature Photonics, 8(12), 937.

[C11-21] A. N. Sloss and S. Gustafson, “Evolutionary Algorithms Review,” 2019.

[C11-22] M.J. Lopez-Morales et al., “Artificial Intelligence-aided Constellation Design for Multi-User MDPSK Non-Coherent Massive SIMO”, IEEE Global Communications Conference (Globecom), Taipei, Dec. 2020.

[C11-23] R. T. Jones, et al., “Deep learning of geometric constellation shaping including fibre nonlinearities,” in 2018 European Conference on Optical Communication (ECOC), Sep. 2018.

[C11-24] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, Dec 2017.

[C11-25] C. Lee, H. B. Yilmaz, C. Chae, N. Farsad and A. Goldsmith, "Machine learning based channel modeling for molecular MIMO communications," 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, 2017.

[C11-26] R. Shrestha e al. "A wire waveguide channel for terabit-per-second links", Applied Physics Letters, vol. 116, no. 13, Mar. 2020.

[C11-27] H. D. Trinh, E. Zeydan, L. Giupponi, P. Dini, Detecting Mobile Traffic Anomalies Through Physical Control Channel Fingerprinting:A Deep Semi-Supervised Approach, IEEE Access, Vol. 7, pp. 152187-152201, October 2019.

[C11-28] "Network 2030: A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond", FG-NET-2030 ITU-T, May 2019

[C11-29] Nick Szabo: "Smart Contracts: Building Blocks for Digital Markets". (1996).

[C11-30] Satoshi Nakamoto: "Bitcoin: A peer-to-peer electronic cash system". (2008), https://bitcoin.org/bitcoin.pdf.

[C11-31] Multichain: Three (non-pointless) permissioned blockchains in production. 2017, Quite taken from https://www.multichain.com/blog/2017/11/three-non-pointless-blockchains-production/.

[C11-32] L. S. Sankar, M. Sindhu, M. Sethumadhavan: "Survey of consensus protocols on blockchain applications." 4th IEEE Int. Conf. on Advanced Computing and Communication Systems (ICACCS), 2017.

[C11-33] Certificate Transparency: General Transparency. https://www.certificate-transparency.org/general-transparency

[C11-34] H. Lasi, P. Fettke, HG. Kemper, et al: "Industry 4.0", Bus Inf Syst Eng, 6: 239, 2014. https://doi.org/10.1007/s12599-014-0334-4

[C11-35] Falko Dressler and Stefan Fischer, "Connecting In-Body Nano Communication with Body Area Networks: Challenges and Opportunities of the Internet of Nano Things," Elsevier Nano Communication Networks, vol. 6, pp. 29–38, June 2015.

List of Contributors

Rui L. Aguiar (editor)	Instituto de Telecomunicações, Univ. de Aveiro
Ahmed Akan	C Tech Bilisim Tech
Ian F. Akyldiz	Georgia Institute of Technology
Marco Andrenacci	MBI
Miguel Ángel Vázquez	CTTC
Antonio Arcidiacono	EBU
Jordi Arjona	Instituto Tecnológico de Informática (ITI), Spain
Jorge Arjona Aroca	ITI
Achim Autenrieth	ADVA
Arturo Azcorra	University Carlos III of Madrid, Imdea Networks
Nikolaos Bartzoudis	CTTC, Spain
Didier Belot	CEA-LETI
Jose Manuel Bernabeu Auban	ITI
Carlos Jesus Bernardos Cano	UC3M
Sébastien Bigo	Nokia Bell Labs
Gregory Blanc	IMT
Nicola Blefari-Melazzi	CNIT
Marja-Matinmikko Blue	Unversity of Oulu
Harald Bock	Infinera
José Bonnet	Altice Labs
André Bourdoux (editor)	IMEC
Raffaele Bruno	National Research Council – Italy
Albert Cabellos	Universitat Politècnica de Catalunya
Giuseppe Caire	TU Berlin
Gino Carrozzo	Nextworks
Roberto Cascella	ECSO
Ramon Casellas	CTTC
Francky Catthoor	IMEC
Fabio Cavaliere	Ericsson
Nicolas Chuberre	Thales Alenia Space
Claudio Cicconetti	National Research Council – Italy
Nicola Ciulli	Nextworks
Daniel Corujo	Instituto de Telecomunicações, Univ. de Aveiro
Emmanuel Dotaro (Editor)	Thales
Falko Dressler	TU Berlin
Jan Craninckx	IMEC
Filippo Cugini	CNIT
Francesco D'andria	ATOS Spain
Franco Davoli	University of Genoa and CNIT S2N National Lab, Italy
Tomaso De Cola	DLR
Hervé Debar	ECSO, IMT
Thomas Delamotte	Bundeswehr University of Munich
Paolo Dini	Centre Tecnologic de Telecomunicacions de Catalunya CTTC, Spain
Rui Dinis	Instituto de Telecomunicações (IT) / Nova University of Lisbon
Jordi Domingo-Pascual	Universitat Politècnica de Catalunya (UPC)
Michael Eiselt	ADVA
Jörg-Peter Elbers (Editor)	ADVA
Omar Elloumi	Nokia
Barry Evans	University of Surrey

Gerhard Fettweis	TU Dresden
Sérgio Figueiredo	Altran
Valerio Frascolla	Intel
Ronald Freund	Fraunhofer HHI
Alex Galis	University College London
Ana Garcia	University Carlos III of Madrid
Lorenza Giupponi	CTTC
Stein Gjessing	University of Oslo Norway
Helmut Grießer	ADVA
Harald Haas	University of Edinburgh
Said Hamdioui	TU Delft
Artur Hecker (Editor)	Huawei
Tim Hentschel	Barkhausen Institut, Germany
David Hillerkuss	Huawei
Marco Hoffmann	Nokia
Bengt Holter	SINTEF (Stiftelsen for industriell og teknisk forskning)
Marko Höyhtyä	VTT Technical Research Centre of Finland
Hans-Peter Huth	Siemens
Jyrki Huusko (Editor)	VTT Technical Research Centre of Finland
Riku Jäntti	Aalto University
Jenny Johnson	mVCE (The Centre of Excellence in Mobile and Personal communications Ltd.)
Josep M. Jornet	State University of New York at Buffalo
Liu Ju	Huawei
Ana Juan	ATOS Spain
Charalampos Kalalas	CTTC
Alexandros Kaloxylos	5G-PPP
Gour Kar	IMEC
Georgios Karagiannis	Huawei
Holger Karl	Paderborn University
Andreas Knopp	Bundeswehr University Munich
Raymond Knopp	EURECOM
Florent Krichner	SPARTA, CEA
Evgeny Kucheryavy	Tampere University of Technology
Eva Lagunas	University of Luxembourg
Yao-Hong Liu	IMEC-NL
Michele Luglio	University of Rome Tor Vergata
Antonio Manzalini	TIM (Telecom Italia) – Italy
Mario Marchese	University of Genova
Josep Martrat (editor)	ATOS Spain
Cyril Michel	Thales Alenia Space
Enzo Mingozzi	University of Pisa
Marco Miozzo	Centre Tecnologic de Telecomunicacions de Catalunya
Werner Mohr (editor)	Nokia
Alain Morad	InterDigital
Raul Muñoz	CTTC
Maziar Nekovee	University of Sussex
Peter Ossieur	IMEC
Daniel Pakkala	VTT Technical Research Centre of Finland
José António Palma	Altice Portugal
Aarno Pärssinen	University of Oulu
Andrea Passarella	National Research Council – Italy
Ana Pérez	CTTC

Luis Perez-Freire	Gradiant, AIOTI
Ullrich Pfeiffer	University of Wupperthal
Yvan Pointurier	Huawei
Ari Pouttu (editor)	University of Oulu
Ayman Radwan	Instituto de Telecomunicações
Parm Raewal	Vodafone
Nandana Rajatheva	University of Oulu, Finland
Aurora Ramos	ATOS Spain
Enrico del Re	University of Florence/CNIT
Johannes Riedl	Siemens
Elisa Rojas Sánchez	Universidad de Alcalá
Marco Ruffini	Trinity College Dublin
Daniel Sáez-Domingo	ITI
Natalie Samovich	Enercoutim
Jose Manuel Sanchez	Integrasys
Pouria Sayyad Khodashenas	i2CAT Foundation
Sachin Sharma	National College of Ireland
Colja Schubert	Fraunhofer HHI
Egon Schulz	Huawei Germany
Kari Seppänen (editor)	VTT Technical Research Centre of Finland
Pablo Serrano	University Carlos III of Madrid
Alena Siarheyeva	Yncréa– Méditerranée
Dimitra Simeonidou	University of Bristol
Muhammad Zeeshan Shakir	University of the West of Scotland
Apoorv Shukla	TU Berlin
Rute Sofia	Fortiss
Hui Song	SINTEF, Norway
Alexandros Stavdas	University of Peloponnese
Juergen Sturm	Texas Instruments
Panagiotis Trakadas	National and Kapodistrian University of Athens
Dirk Trossen	Huawei ERI
Attilio Vaccaro	MBI
Alexandre Valentian	CEA-LETI
Alessandro Vanelli-Coralli (editor)	University of Bologna
Angeles Vazquez-Castro	University Autonoma Barcelona
Luís Velasco	UPC
Ovidiu Vermesan	Sintef Digital
Panayiotis Verrios	ACTA Lda
Piet Wambacq	IMEC
Rainer Wansch	Fraunhofer Institute
Michael Welzl	University of Oslo Norway
Joerg Widmer	IMDEA Networks, Spain
Henk Wymeersch	Chalmers University of Technology, Sweden
George Xilouris	NCSR Demokritos
Dehao Wu	University of Sussex
Wen Xu (Editor)	Huawei Germany
Seppo Yrjölä	Nokia
Florian Zeiger	Siemens
Martina Zitterbart	KIT
Zhi Zhou	ZTE Germany

About NetWorld2020

NetWorld2020 is the European Technology Platform for communications networks and services. Communications networks enable interaction between users of various types of equipment, either mobile or fixed. They are the foundation of the Internet. The NetWorld2020 European Technology Platform gathers 1000 players of the communications networks sector: industry leaders, innovative SMEs, and leading academic institutions. NetWorld2020 has as mission developing position papers on technological, research-oriented and societal issues, in order to strengthen Europe’s leadership in networking technology and services so that it best serves Europe’s citizens and the European economy.

About the SRIA

The Strategic Research and Innovation Agenda is organized by the Networld2020 European Technology Platform to provide a long-term view about the future of telecommunications in Europe in cooperation with the 5G-Infrastructure Association (5G-IA). It collects the views from industry and academia, and received inputs from Alliance for Internet of Things Innovation (AIoTI) and from the Networked European Software and Services Initiative (NESSI).