Technical Annex to

Strategic Research and Innovation Agenda 2022-27

European Technology Platform Networld Europe

"Smart Networks in the context of NGI"

2024

1 Introduction

The digital society has been built on the backs of the Internet. The large evolutions on associated distributed systems has enabled the digitization of processes and transformation of industries. While the Internet's foundational architecture, developed in the 1970s, relied on a simple client-server connectivity model using the IP protocol suite, it has been adapted to meet modern demands, including the novel cellular network demands. Innovations like content delivery networks (CDNs) and localized domains have built on this foundation to address its limitations, shifting the nature of communication, services, and provisioning. Major Internet players like Google have adapted to evolving service access needs by focusing on improving connectivity within customer access networks, particularly to their Points of Presence (PoPs). Today, most Internet servers are hosted in large-scale data centers, which act as points of presence (PoPs), facilitating service delivery without requiring service providers to manage physical infrastructure. To enhance efficiency, PoPs have been moved closer to end-users by major service providers and CDNs, reducing latency and costs. Consequently, communications and computing has transitioned from server-centric models to service-centric, modular, and virtualized frameworks. These innovations enhance scalability, flexibility, and proximity to end-users, driving efficiency and setting the stage for future advancements.

The traditional monolithic service model has been replaced by modular, collaborative microservices. These independent components cooperate to deliver services, increasing flexibility and scalability. However, this transition introduces challenges such as transaction management and efficient resource allocation. Nevertheless this trend reached even the core functions of the cellular networks, becoming pervasive both in the service provision and on the control aspects of the current critical communication infrastructure. This trend enables modern networks to actively support application-level services by integrating functions like load balancing and security.

Service provisioning has also advanced through virtualization. Lightweight approaches, such as containers and unikernels, allow dynamic microservice deployment. This flexibility underpins modern cloud computing and telco cloud infrastructures. For instance, 5G's service-based architecture (SBA) uses cloud-native microservices to deliver industry-specific solutions. Proper service routing enables mobile operators to localize services within distributed data centers, offering competitive advantages in latency and efficiency.

Network programmability has further enabled dynamic changes to network functionality, allowing real-time adjustments to network elements, functions, and services across all infrastructure segments (access, core, edge, and cloud). Programmability facilitates the creation, deployment, and management of network services, supporting diverse execution environments at the forwarding plane level. On a different level, open and programable systems have reached all aspects of the communication systems, including the link layer components.

The Internet has also expanded far beyond consumer-focused applications like social media or OTT video. Its penetration into industries such as manufacturing, health, and supply chain management, driven by the Internet of Things (IoT), highlights the diversification of service use cases. This raises critical questions about whether the basic Internet networking model, with its "one-size-fits-all" approach, can support such varied services. The realities brought in terms of localized domains by the dominance of the cellular networks have shown that there are different system concepts that can be successful and still rely on Internet concepts to successfully address functional and non-functional requirements. Service invocation models are evolving in these diverse environments. Each framework maps onto specific lower-layer protocols, often leaving IP as the sole common denominator.

The relationships between service providers are becoming more complex as competition and varying trust levels emerge. Delivering trustworthy end-to-end services now requires dynamic, intelligent methods to manage subsystems and microservices of variable reliability, and the ability to deploy reliable and manageable services end-to-end built with a diversity of suppliers with different trust and reliability levels.

In summary, our society is built on collaborative Internet services that have advanced significantly from its Internet inceptions, integrating network programmability, diverse service models, and adaptive trust mechanisms to meet evolving demands across consumer and industrial sectors.

Continued innovation in this domain will need to address current challenges and unlock new opportunities for these distributed computing systems, from the communication to the computing components. This annex to the Strategic Research and Innovation Agenda 2022 is an integral part of the white paper, but focused towards a more technically oriented audience. It discusses concepts and technologies essential for developing such innovative distributed services in a world-wide infrastructure. The diversity of technological domains required for future communication infrastructures highlights the relevance of multiple innovation domains for European Research. In the white paper, a simplified version is presented, but in this annex multiple detailed aspects are covered. We have nine different chapters in this annex, distributed as:

- System Architecture Considerations – analyzing the evolution of systems towards dynamically composed, multi-stakeholder environments, with an increasing softwarization and intelligence of the whole system, and the accompanying challenges.
- Fundamental Enablers for Future 6G Systems – We look at the future of fundamental technologies, like routing, addressing, forwarding, telemetry, and transport technologies.
- Network and Service security – discussing the paths on the increasingly relevant aspects of security in our infrastructure
- Software and AI technologies for telecommunications – addressing the software related challenges of the ongoing network softwarization, the increasing system complexity, the dominant trend of AI, and the enabling of adaptive and customized services.
- Radio Technology and Signal Processing – where the challenges and potential solutions perceived for the future wireless (and mostly cellular) communications are discussed
- Optical networks – a critical component of the backbone (amongst other potentialities) and its perceived evolution is detailed in this chapter.
- Non-terrestrial Networks and Systems – discusses the upcoming closer integration of 3D networks into the overall communication system
- Opportunities for Devices and Components – tackles the unavoidable challenges at the fundamental element level, which will constrain and limit all system developments.
- Future Emerging Technologies – is a final chapter discussing promising technologies that may bring structural changes across all the current communication concepts. Some of these technologies are already being researched but have not yet a clear path (if ever) to the transformational impact it is expected by their wide adoption.

Given the specificities of the different technologies, some slight structure differences exist across chapters, but the structure remains essentially the same across chapters, with the exception of the Future Emerging Technologies. Overall, this document was based in the previous version of the Strategic Research and Innovation Agenda [C1-01] as a baseline, which itself went through a long development process (with almost 200 researchers working on it). The SRIA 2024 discussion started, with a public event in Lisbon, in November 2023, in the Visions For Future Communications Summit. This was followed by a long period of discussion inside the Expert Group of NetworldEurope, where contributions collected from hundreds of experts in Europe, and where different key innovation stakeholders were directly addressed to provide comments. In a final stage, a public consultation was issued, and its comments properly reflected inside the final text. Overall, this SRIA has been the result of extra additional work of a set of 150 volunteers, coming from 85 different entities, improving and expanding the previous work performed for the SRIA 2022. The NetworldEurope community is in debt to all for their selfless efforts, which makes NetworldEurope the European Technology Platforms that provides the shorter updates cycles on the Strategic directions for Europe. The full list of technical editors and contributors is included in the last chapter of the document.

2 System Architecture Considerations

Editor: Artur Hecker

2.1 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 started 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, supported by newer innovations such as content delivery networks (CDN) albeit still relying on separate indirection architectures to the basic IP packet delivery. Some of 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. [C2-1][C2-2][C2-3][C2-4][C2-5][C2-6], among many others)[1].

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 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 [C2-07]. Netflix's estimated 15% share of the Internet traffic is mostly served through localized PoPs [C2-08]. 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.

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

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 [C2-09]. 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 [C2-10] 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 [C2-11] or IETF FORCES [C2-12] (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 [C2-13], 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 an opaque pipe and utilizing service end points 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 [C2-14], 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 system architecture of the last decade has concentrated on dominant commercial requirements, with all other aspects mostly in place to moderate the negative effects of a focus on cost reduction and flexibility, leading to systems with low resilience and survivability. The current pressures in western ecosystems in terms of infrastructure attacks already shows the limitations of our service environment, and the increased multistakeholder characteristics of future environments will only highlight these weaknesses (addressed in more detail in Chapter 4). System resilience will need to become a primary concern on end-to-end service architecture, and service provision.

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.

2.2 System Architecture Vision: Towards Smart Green Systems

With the general move towards collaborative services in the general ICT domain, 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; …), while providing better quality of service (more performance, less latency, adjustable, verifiable trustworthiness levels, etc). Chiefly, if the original Internet was about inter-networking, i.e., best effort bit transport between different networks, future research must address inter-computing, i.e., service execution between different systems, potentially deployed and operated between and by different stakeholders yet accounting for the respective service expectations within the whole chain, including potentially higher value goals like, e.g., trustworthiness or sustainability.

In particular, the same applies to mobile communication systems, which have become a crucial part of the overall Internet ecosystem with the tremendous success of the mobile Internet (cf. smartphone revolution). Indeed, to shorten the paths (and latencies), to reduce general infrastructure involvement (better greenness, risk reduction) and to keep the local operations/data local (better governance), these systems exhibit a unique positioning in terms of standardized omnipresence, best possible locality and realizations already involving both compute and networking resources. However, to achieve this target, their ongoing transformation from single authority domain, mere access networks to dynamically aggregated arbitrary service execution platforms must continue.

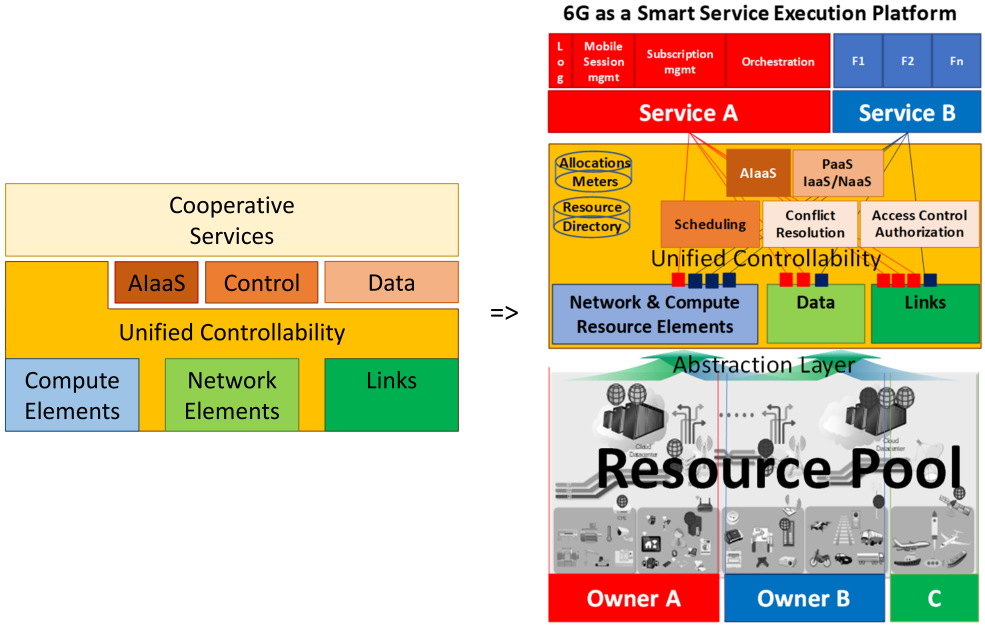
With more and more intelligence and computing power available per resource, in the future, the resources of these systems, configurable and orchestratable dynamically (i.e., also reprogrammable in runtime), do not have to be limited to particular predefined roles and can be used both to deploy/support new services (both network and end-user services) and to better match the requirements of services running over the infrastructure – again, potentially accounting for requirements not necessarily stemming from the service logic per se, like energy consumption reduction, some form of confinement, etc. With this however, *unlike 5G,* 6G will be not only more flexible in both its services and in its realization but will also exhibit much higher dynamics, in service types/loads but also in its own topology. With that *higher dynamics* and the seamless *co-existence of virtual and physical entities*, the currently physically separate islands of 5G and prior systems will often overlap in resources in 6G. This applies both to different domains of one single network (Terminal/RAN/Core), just as it applies to several networks (e.g. run by different MNOs) and to entirely different systems (mobile networks and clouds).

Using the offered large variety of novel challenging ICT services, a massive number of devices will be served by these systems generating, exchanging and treating very large quantities of data. 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. Useful insights can be generated based on the automatic analysis of all that data (e.g., using machine learning methods, ML, and artificial intelligence, AI). Beyond the analysis, AI/ML can also be used to optimize deployment, adaptation, reconfiguration and other decisions or to create better-suited system modularizations and novel entities better suitable for the overall required processing. Hence, *it is paramount to approach AI/ML systemically to correctly assess the relevant trade-offs*: AI/ML instrumentations per se require massive data transfers, are computation-intensive and, ultimately, might consume massive amounts of energy. *Relying on siloed solutions and dedicated implementations limits the usefulness of AI/ML, while it increases both its costs (resources) and the cybersecurity risks (attack surface).*

The postulates above imply that the future network technology will have to support the general Internet economy and the particular needs of the cyber-physical infrastructure, like those encountered in the production industry, 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. 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 operations on 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 a 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 on-board 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 before, during and after the execution, i.e., in particular in runtime, and supports the 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.

Overall, we envision a Smart Green Network as a programmable system based on a common, unified controllability framework spanning all resources to provide each authorized tenant with the required capability to control respectively her resources regardless of their location, type and nature, i.e., 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 unified controllability framework allows a tenant to glue such disparate resource islands to one system of that tenant supporting smart flexible instantiation and adaptive, elastic and correct execution of any service on her resources (Figure 2- 1). For 6G in particular, the resources will stem from all system players, typically from mobile network operators, but also from cloud providers, non-public network providers and might include terminals, where suitable, while tenants could be mobile system operators, particular service providers, vertical industries, enterprise networks, IoT services or private subscriber networks. Interestingly, 6G will have to architecturally embrace the fact that system resources used for service execution might, per se, be provided, e.g., as services, i.e., that the service and its control in general cannot be limited to the strict boundaries of the authority domain of the service operator only, nor to any particular layer. Rather, in this vision of 6G, all system participants are *potentially* both *resource providers* and *service consumers*. In this situation, the properties of the service must be, in general, enforced regardless of (or even in spite of lacking) assurances at the resource layer.



a) b)

Figure 2- 1 - The general Smart Green Networks Concept and its projection to 6G

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 integrates both the existing AI/ML algorithms as well as 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* and of some form of *agreed test*. 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;
- Enable seamless allocation of complex IT objects (ranging from atomic modules to complex services, with the possibility of reusing allocated objects) over some selected as well as over panoplies of such objects, i.e. APIs, interfaces, control hooks, etc.;
- Support seamless fusion of logical and physical worlds, by providing, at each layer, data describing the situation of resources with respect to their involvement into the execution of allocated IT objects (e.g., consumption of energy of a session, cumulated duration of the task execution, memory used by a distributed IT object, volume of transmitted data, number of involved resource elements);

- Integrate autonomics to enable both self-organized, resilient programmability and elastic, correct execution of such IT objects;
- 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:

- pervasive operational control solutions in virtualized environments (section 2.3).
- extensible and flexible data plane solutions (section 2.4),
- efficient yet correct runtime resource allocations and their execution (section 2.5),
- integration of AI/ML at the system level (section 2.6),
- programmable infrastructures composed of versatile devices and subsystems, with an explicit feedback API of the respective resource situation and its current status (section 2.7),

Some added summary considerations on security (section 2.8) will be followed in large detail in Chapter Network and System Security, as well as some summary considerations on sustainability (section 2.9).

It should also be noticed that it is important that the directions captured in this section be **accompanied by the appropriate economic and policy work** in future research to make way for the envisioned new services that go beyond what is currently supported by access networks, 5G, transport networks and data centres. Broader user involvement and incentives for more local resource usage, notably at the (deep) edge, is highly recommended.

2.3 Virtualised Network Control for Increased Flexibility

2.3.1 Programmability is Control

To increase universality and sustainability, 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 running program/task to receive some infrastructure event and to possibly tell to a given resource what to do, both proactively and reactively, including in runtime. Classically, requirements on control plane are intrinsically linked to the data plane requirements through the service logic. Yet, with programmability (e.g., such as that being explored with SDN, P4 or NetApps), *any* service and, hence, any data plane becomes imaginable within limits of the available capabilities and capacities, 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 required security levels 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 that programmability per se creates*. Any programmability solution not able to self-organize or adapt is, therefore, incomplete [C2-10][C2-15]. Network and system control cannot rely on rigid approaches, as any such approach would only be suitable for particular needs and environments (e.g., centralistic control, particular hierarchies, etc). Instead, *novel solutions must be 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 and changing 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 from 3GPP), 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 (typical technologies: OpenFlow SDN, IETF ForCES, ONF P4). Whereas legacy 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 (typical technologies include NetApps, virtual machines, containers, different host virtualization techniques and industrial frameworks such as ETSI NFV).

2.3.2 Separation of control/controllability

The discussion above immediately raises a completely new question of a *service-independent control of resources per se*: as all infrastructure capacities are, in principle, service-independent, in addition to what is done within each service, we need novel means to make sure that the execution of any service-specific element on the infrastructure is correct and sustainable. In other words, while a legacy hardware router routes and a network switch switches for the duration of their respective lifetimes, 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 has had other roles and tasks. Hence, it must be verified that it routes correctly over time of this routing task, and this despite possible task overlap in logic (allocated tasks could result in contradictory operations) or resources (allocated tasks could get an insufficient resource share). 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 the existence and the function of a node used to be the same, so was their control. 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. To distinguish this notion from classical task- or service-specific control, we call this new area *controllability.*

As a side note, controllability is different from the notions of management or administration as well, as management and administration rather refer to a) who is in charge (humans, management platform), b) what is being done (management model) and c) different timings. In contrast, programmability could be between devices; it is typically employed for tasks radically different from the classical management (in particular, not OAMP, not FCAPS, but rather related to some particular function realization) and executed and adapted in runtime, exhibiting time-criticality to the running service. As an example, consider OpenFlow SDN, where SDN controller and SDN switches implement the control plane for traffic switching, whereas the management of the whole domain is done independently of that, e.g., over the so-called north-bound interface and classical

SDN switch management. Previous discussions around these subjects were touched under the concepts of clean slate Internet research, but the challenge here is mostly to address solutions that have a credible path to impact the future 6G networks.

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 in a geographical distribution setup, most network functions rely on distributed realisations, causing the allocation, extension, monitoring or migration of a network function to be much more challenging than the question of copying a software state between two local nodes.

2.3.3 Multi-Tenancy and Ownership

Network virtualization is resource sharing. Therefore, service footprints, projected to the providing (e.g., physical) resources involved into the execution of that service, 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 "listed on that 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 should be aware of energy consumption.

In spite of its expected pervasiveness, resource control solutions need to respect and maintain boundaries of the responsibilities and rights for each stakeholder in the ecosystem, as these are key for a trustworthy 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 of physical devices towards virtual 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.

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.

2.3.4 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 possibly 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 by acting on the resources spanning the control plane. 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**.

2.3.5 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 services, it is exactly this glue that requires standardisation – more so, than any domain-specific management interface, which could be proprietary in principle.
- 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-, IaaS- 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, delayed or delay-tolerant execution, etc).
- Resource control needs to be able to handle questionable data quality, how to proceed when data is not good or biased, developing either fall-backs, 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, partly occupied by loads from other executed services. Such a distributed guaranteed execution under contention and with concurrency is extremely challenging and, currently, can only be solved on very small scales.

Research Theme	**Virtualised Network Control for Increased Flexibility**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Pervasive Resilient Autonomic Resource Control	Mid-term	Highly scalable, distributed, self-organizing routing protocol to provide in-band connectivity between all resources (zero-conf, zero-touch). Should work across a variety of different topologies (sparse, dense, changing), should support mobility and multi-homing of nodes and avoid to create traffic concentrations. Zero-conf and zero-touch are essential so that no configuration errors can break the connectivity, and that failures be auto-corrected. Solutions should notably feature progress as per time axis. Final solutions delivered validated to achieve: - high scalability (number of nodes n>100.000) - high success ratio (delivery of packets to random non-isolated destinations >99.9%) rapid (re)convergence to the expected success ratio under dynamics (<10s for up to 20% random link or node failures)	Flexibility and universality Sustainability (in time) Trustworthiness
Separation of Controllability and Control	Mid-term	Emergence of resource control as a separate domain from service-related control: - Confirm architectural and operational principles for suitable solutions for multi-tenant concurrent control in relevant environments. - Define APIs with authorization from the resource control plane to the service layer, so as to avoid repetitive control and management traffic (e.g. for monitoring, command piggy backing, etc).	Trustworthiness Dependability Better governance Multi-tenancy support

2.3.6 Recommendations for Actions

Research Theme	**Virtualised Network Control for Increased Flexibility**	
Action	Pervasive Resilient Autonomic Resource Control	Separation of Controllability and Control
International Research	Need agreement and standardization	
Cross-domain research	X	X

2.4 Evolving from Data & Forwarding to Function-Rich User Planes

Internet communication has evolved from its original design towards the CDN-assisted provisioning of services in many of today's deployments. However, the role of the network is still largely defined by the original E2E argument, postulated by Salzer et al [C2-16], as communication across network elements between communication endpoints. This argument has driven much of the development of key Internet technologies in recent decades.

However, compute elements are seen as not just crucial elements in future system architecture but as also driving the methods used at the data and forwarding plane. Instead of purely being data-oriented but opaque to the purposes of the communication, following thereby the core of the E2E principle, future exchanges between endpoints must become intrinsically aware of computation and the necessary service and compute-specific information that can improve on the communication resource usage, i.e., the performance, trustworthiness and sustainability of the network.

This intertwining of compute and communication resources is captured in the evolution from today's data (and forwarding) planes to *function-rich user planes* in the form of Compute Inter-Connection (CIC) fabrics, where CIC embodies this strong linkage of both key aspects, namely the computation at the communication ends, the communication itself but also the serialization of individual communication relations to more complex computational tasks, i.e., the chaining of service functions.

Any future design for these evolved user planes MUST enable the interconnection of compute resources across one or many computational tasks that are often jointly evaluated in their overall performance.

While the strength of the Internet is its agreement on key building blocks, most notably its Internet Protocol (IP) with key aspects like addressing and best effort packet delivery, the concept of Limited Domains [C2-17] has long recognized the power of the more service and compute-rich network edge in driving network technology innovations that accommodate the specific stakeholder requirements that those deploying innovative edge solutions bring to the table. As argued in [C2-18] this innovation vehicle through limited domains has driven the larger overall innovation in Internet technologies, while relying on the Internet's reach to ensure the global impact that its development has brought to the wider society.

We assert that many of the developments pertinent for evolving today's data planes into full CIC fabrics will be driven by the same limited domain notion. Through this, we will often see development and deployment of stakeholder-specific (private) solutions, remaining interconnected through the Internet as it evolves. This assertion reconciles the desire for vertical-specific innovation with the need for sustainable evolution of the overall system.

Key to realizing these evolved user planes is the evolution of (current) protocols into a toolbox of solutions that can be deployed into a variety of limited domains. For this, we see these protocols needing to suitably evolve from the three key fundamentals (i.e., principles – key design choices) of today's Internet, i.e., the assurance of *global reachability* through a *robust* packet forwarding mechanism that would provide a *best effort* service to higher layers [C2-19] towards embodying methods that interconnect the distributed pool of compute, storage, and communication resources in a multi-constrained manner beyond the best-effort of the current Internet.

2.4.1 Design Considerations for Evolved User Planes

From the discussion in the previous sub-section, we derive a number of design considerations for evolving data plane into those richer user planes that would ensure a continued support for the services and interactions we have been seeing in the Internet, depicted in Figure 2-2.

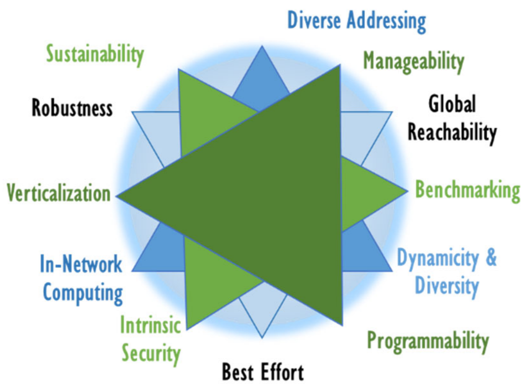


Figure 2- 2 User Plane Evolution - Design Considerations

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 user plane solutions do not need to necessarily address all considerations and we can already see examples for proposed solutions [C2-01][C2-02][C2-03][C2-04][C2-05][C2-06] considering certain aspects described here:

1. 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 enhance this aspect of the current Internet in order to efficiently support the many new services that we see being deployed in a variety of limited domains, 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 [C2-25], information-centric networking [C2-26] and solutions on *semantic addressing* [C2-27] 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 2.1. 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 2.3.
2. Dynamicity & Diversity: As observed in Sections 2.1 and 2.3, 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 user plane will only inadequately support the advances we see in complex SW systems utilizing the Internet, e.g., through container-based micro-services. But this diversity also brings opportunities in utilizing the possible *diversity* of choice, e.g., in endpoint relation, forwarding path, etc. This raises a number of challenges to the off-path (indirection) architecture of the current Internet, utilizing a combination of DNS resolution with following data plane communication between IP locators. If we were to move to a world where communication endpoints (in ever more serendipitous compute elements of a future 6G world) can be rapidly established but also torn down, an explicit resolution step, incurring around 15 to 100ms latency (according to [C2-20]) in modern point-of-presence based deployments, seems to be inadequate for highly dynamic relationships in use cases such as those found in mobile edge computing, distributed AI and others. Instead, approaches to on-path resolution or runtime traffic scheduling, as suggested in [C2-21][C2-22], accommodate the required dynamicity at line-speeds and high frequencies, albeit questioning the placing of traditional application layer functions nearer to the data packet processing. Also, we foresee impacts on transport protocols, moving from the current limited end-to-end view over single path towards rich and diverse many-path approaches, utilizing swarming and, possibly, network coding techniques to improve on robustness and throughput simultaneously.

3. In-Network Computing: Relationships will not only become more dynamic in nature but also more complex in terms of *inter-dependencies* as well as operations on (data) packets in transit. The current model in the Internet treats relationships at the application or session layer, realized through independent connections, managed through protocols like TCP, QUIC and others, with separate resource management schemes, where network operations are (largely) limited to high-performance packet forwarding. Emerging use cases, such as AI, but also approaches to telemetry would greatly benefit from in-network operations going beyond such simple yet efficient forwarding. This concept of *in-network computing* moves computational functions, in alignment with the E2E principle, into the network, transparently exposed to and in orchestration with the application, in order to jointly achieve more complex operations, such as AllReduce, Scatter-Gather and other collective primitives. Here, additional in-packet *metadata* at a lower part of the evolved user plane can build atop existing concepts such as service function chaining (SFC) [C2-09] albeit for parallel not sequential transactions. Such enriched in-network computing capability smudges the edges of what executes an application and network function, albeit preserving the boundaries between application and network through a clear orchestration of the relevant function placement alongside well-established business relations in all phases of the system, i.e., at design stage (when defining the suitable interfaces), at planning stage (when setting the operational boundaries for possible resource executions), and at runtime (when constraining the function placement along those defined boundaries set in the planning stage).
4. Sustainability: 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 *sustainability* MUST be a consideration to the design for an evolved user plane, even to the point where the selection of suitable mechanisms ought to include an *energy consumption 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 [C2-23]. 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 [C2-24]. Furthermore, communication without suitable insights into the usage of compute resources inevitably leads to situations where those compute resources are inefficiently used, e.g., when the often-practiced shortest path routing leads to an otherwise overloaded compute element. Instead, approaches that are compute- and therefore also possibly *energy-aware*, such as proposed in [C2-21][C2-22], are key to building a sustainable user plane that evolves from the pure communication centrism to smarter user plane actions that take the overall task (i.e., its compute, communication, involved data, quality, security and other requirements) into account with the aim to reduce overall energy consumption of the joint compute and communication system.

5. Security plays an important part in user 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.

6. Benchmarking allows for properly reasoning about and evaluating aspect such as sustainability but also security and efficiency of solutions (in particular, when comparing across a number of enabling solutions in the design phase and deployed solutions in the field). For this, well-defined benchmarking suites must exist, where suitable *telemetry* information provides the needed input. Such knowledge obtained through telemetry also provides input into online control mechanisms, thus micro-benchmarking, e.g., transport connections, for an improved congestion behaviour, while macro-scale benchmarking feeds not just into the aforementioned solution comparison at the product level but also into informed policy decisions and the formulation of tendering requirements. An evolved user plane can provide the right technical input into those processes.

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: 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 2.4, 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 plane needs to be assessed.

9. Verticalization: The differing usage of user plane solutions in different Limited Domains places a strain onto the realizing protocols, often leading to many extensions to existing ones, often not being compatible in concurrent use. Hence, it is important in the development of new solutions to fully understand the domain-specific requirements for those extensions and the impact on the existing solutions, without falling into the trap of feature creep. More so, the development of new solutions must consider the possible pruning of capabilities that have proven to be of little or even negative benefit to real deployments, thus reducing the possible overloading of protocols and solutions, thereby also reducing its management in real deployments.

Most crucially, the evolution of the current data plane solutions to the function-rich user plane we outline above requires to continuously evolve existing and possibly develop new enabling solutions in the form of *protocols*. Chapter 3 will discuss this evolution and how the design considerations for future user planes will be reflected in needed research and development on such protocols.

2.4.2 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 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 2.2), 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 2.3, 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 2.4 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 [C2-28] 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 are the key protocol developments to achieve the desirable architectural concepts and functions? Any user plane solution, including existing ones, utilize a set of protocols to realize the desired architecture. Particularly in the context of moving beyond purely cellular architecture in future Smart Networks and Services, it is imperative to identify the needed protocols as building blocks to suitably realize those extended uses and enabling architecture. For this, existing protocols may need extensions or entirely new ones may need development. In collaboration with concrete challenges to realize such protocols, listed in Section 3, an architectural consideration needs to be provided as to where and in which architectural framework those protocols may live.

Research Theme		**Evolving from Data and Forwarding Planes to Function-Rich User Planes**		
	Research Challenges	Timeline	Key outcomes	Contributions/Value
1	**Architectural frameworks** to enable user plane evolution that is economically and technically sustainable	Mid-term	- Architectural blueprints/ frameworks with key roles and interfaces - Economic models to show viability of new communication services Proposed solutions should: -**Confirm architectural concepts** and principles for suitable and novel data plane solutions in relevant environments. - **Define relevant APIs** to enable key exchanges in multi-stakeholder deployment models - Identify key enabling protocols for different architectural concepts that may either be based on existing ones or may require development of new ones	- **Sustainability** through innovation models for evolving user plane and novel concepts for energy-awareness in user plane functions - Improved **innovation** capability through new protocols - **New services** realized via new protocols - Facilitation and identification of **new market entrants** - Faster deployment

2.4.3 Recommendations for Actions

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

1. *Call for internationalized efforts*: given the challenge to evolve the user plane, including through new deployments or test beds, 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 user 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 user plane functionality, strong experimental evidence and very 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 and ongoing SLICES efforts. Open experimentation 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 user 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 explicit technology enablers for an evolved user plane* along the considerations discussed in Section 4

Research Theme	Evolving from Data and Forwarding Planes to Function-Rich User Plane
Action	Challenge 1 architectural frameworks
International Calls	Needed for possible concerted standardization efforts
International Research	Encouraged due to international expertise beyond Europe
Open Data & large trials	Encouraged to broaden insights into markets beyond Europe and replicability of research
Calls to fund technology enablers	Encouraged to provide suitable pool of solutions to realize the desirable user plane(s)

2.5 Sustainability, Efficiency and Resource Management

Sustainability has recently become a major design preoccupation of future ICT. Emerging compute platforms and telecommunications systems will follow the principle of *sustainability by design*. While this is a nice slogan, it is mandatory to clarify, which new technologies or technological advances are required, so as to pre-empt those through ICT research.

In principle, ICT technology has always considered energy efficiency. However, initially, performance was understood as more important than energy efficiency, as indeed the deployment of ICT and its proportion in the overall carbon-dioxide equivalents (CO_2e) emissions was rather low. As the ICT has become more omnipresent, the technology development started focussing on main energy consumers, trying to achieve practicable energy-performance trade-offs, especially where energy was a quality criterion. For instance, with 2G and 3G, the main accent was on extending the battery life of the user equipment. 4G and first 5G generation have achieved remarkable efficiency gains on the air interface. 5GA, has further extended the scope to consider the whole 5G system, from the UE over the RAN to the 5G Core. These developments were driven by the desire to reduce the OPEX of the providing MNOs, whose energy bills represent up to 90% of the overall OPEX. This is reflected, among others, in the Total Cost of Ownership (TCO) KPI in the 5G era.

However, from the larger, societal perspective, sustainability considerations should go beyond efficiency, as the expected increase in the system loads will certainly overcompensate for the achieved gains per bit or per operation. This is due to the so-called *rebound effect*, which has been already observed and confirmed for all prevalent ICT platforms (cf. Wirth's Law, WinTel ecosystem, access networks such as xDSL, transport networks, 5G mobile systems, etc).

Sustainability therefore emerges as an area covering several technological aspects:

- Energy efficiency, still paramount, to be considered by the respective service provider;
- Service user involvement, a relatively fresh notion, to be considered throughout the whole service function chain;
- Carbon-dioxide equivalents (CO_2e) emission reduction, to be considered by all players, i.e., users and all service providers in the service function chain.

This translates to several new problem spaces, currently unaddressed, underestimated or completely overlooked in both the industry and academia. We describe them in more details in the following subsections.

2.5.1 Infrastructure Flexibility vs. Network Capacity Planning

As the trend to an increased network flexibility 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 flexibility measures (marketed as e.g., NaaS or network slicing) turn this principle upside-down: while the infrastructure operator remains neutral to the service, the NaaS instance, slice or service owner is expected to translate the *allocation requirements* onto the infrastructure capacities and capabilities, an exercise that lacks a reliable general methodology. Incapable of correctly translating service to capacity requirements, slice/service 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 allocation 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 flexibility 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 each allocation (NaaS/slice or particular service) can operate, it is the runtime (continuous, real-time, hot) management and control that determines the efficiency and therefore the costs of the provided service. If challenging, beyond best effort services, such as slicing, NaaS, advanced communication services, etc. are to be successfully provided by the telecom infrastructures, the employed technologies must embrace this change and use mechanisms and practices that feed runtime control over a longer timeframe back into the planning and investment cycle for network infrastructure.

Independently of scale, dynamic allocations render 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 of the allocation, 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 drive 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 programmable network infrastructures albeit situated in a many points-of-presence nature of resources, utilised over a possibly huge area of requirements on those.

The change towards network and infrastructure programmability results in an even more radical shift towards the need for runtime optimizations. As programmability advocates a reactive change of the behaviour of the infrastructure in runtime, the efficiency of such changes must be also addressed in runtime: the information on what to optimize is simply not yet available at deployment time, and hence capacity planning is hard to apply.

2.5.2 Programmability 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 allocations and executions of the latter should be addressed at the systemic scale in runtime [C2-29]. Indeed, concurrent resource-competing or semantically contradictory requests at either allocation time or during (elastic) execution must be dealt with to avoid partial operation (e.g., of a path or a slice), generally being useless and, hence, waste of resources, while requiring novel mechanisms for networked garbage collection to free up any erroneous (.e.g, incomplete, orphaned, etc.) resource allocation during such conflict resolution.

While mechanisms exist for handling concurrency at individual component/node level, guaranteed allocations (NaaS, slice, etc.) 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 allocations 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 disjoint 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 [C2-30].

2.5.3 Elasticity: Efficiency Requires Runtime Scheduling

When addressing efficiency, Total Cost of Ownership KPI (TCO) and green ICT become important aspects to consider. For instance, 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 while the mentioned mechanisms address the first part, the second part is still unsolved: the question here is elasticity of the slices per se. 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? [C2-21][C2-22] Note that this cannot be solved within the slice, if the infrastructure owner promises (and sells) extra-functional properties of the allocation; 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 [C2-21] is paramount to address TCO, as solutions to this problem would allow to overprovision slices, without the need to overprovision the underlying infrastructure in the same way. Runtime service scheduling therefore is a potential answer to the questions of elastic and dynamic allocations, currently unsolved. Moreover, if an efficient solution to this problem can be found, logical allocations can – and, for efficiency reasons, should – be implemented as dynamically scheduled entities rather than exclusively reserved (and, therefore, possibly wasted) resource pools for tenants. This has potential to largely increase the resource efficiency by statistical multiplexing (e.g., power of two choices).

2.5.4 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 energy 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 [C2-31]. Even though this effect is due to the increased "popularity" and not to a shortcoming of 5G per se, undeniably, energy efficiency of distributed compute facilities emerges as a required yet insufficient preoccupation. This is not limited to individual data centers, but should consider network and compute at once, e.g., distributed compute, edge computing, mixed virtual/physical networking and interconnected data centers. 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), crude overprovisioning is not a valid approach for radio networks and the computing domain. For instance, modern

DCs reduce the required compute resource and energy for a given load using DC-internal schedulers (e.g., Apache Mesos, Kubernetes K8) and many other means [C2-31][C2-32][C2-33].

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 somehow limited to greening energy power supply (eco-current, not a technological ICT area), potentially using smart grid's demand management, and the "recycling" of waste heat, in particular from the DC cooling systems. Both approaches are already in use in modern datacenters; they 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.

Among suitable techniques, some of which already adopted in the presence of non-virtualized equipment, dynamic adaptation (like adaptive rate and low power idle) and smart sleeping need to be further investigated to extend them to the virtualized environment, where chains of VNFs are implemented by Virtual Machines or Containers. Power states that are immediately identifiable in physical hardware (as defined by the ACPI industrial standard [C2-34]) do not map in a straightforward manner to virtual CPUs, as the interaction with hardware is mediated by software artefacts and hypervisors. Likewise, defining interfaces to convey energy management capabilities of virtual devices, in order to abstract their internals and expose energy-related parameters that can be manipulated by control strategies in the control and management planes, by extending the so-called Green Abstraction Layer (GAL), needs further refinements, although efforts have been already undertaken by ETSI and ITU [C2-35][C2-36] in this direction.

Most of technological measures described above fall into the category of energy efficiency. However, as explained in the introduction to this section, we cannot green the ICT with efficiency measures alone because of the expected higher system load.

Service scheduling by eco-current availability falls into the CO_2e reduction category. It should be noted that very high potentials for ICT's CO_2e reduction are attested to the measures targeting better exploitation of eco-current. Therefore, this general capability ought to be integrated into all modern ICT systems and should become a generalized requirement. This in turn requires transversal interfaces (we need to know, in runtime, at the service layer, which mix of current each involved serving infrastructure element has; we also need to know suitable alternatives), mentioned in the introduction and called "benchmarking interfaces" in Section 3.4. In addition to such interfaces, methods and algorithms should be able to re-route service requests to the most ecologically optimal resources, addressing all potential prerequisites (knowledge of energy mix and loads) and problems that might stem from such re-routing (affinity, context transfers). Ultimately, though, reliance on eco-current is subject to the external dependency from the energy sector. Technologically speaking, nothing can be done from within ICT to get more eco-current or eco-current at more locations, more often, etc.

In this view, **service user involvement** emerges as a novel and promising direction for energy consumption and CO_2e reduction in modern ICT. It is to be considered throughout the whole service function chain (SFC) so as to enable three distinct features:

1. *Service user awareness* of the ecological impact that the used service is responsible for;
2. *Service user incentivization*, i.e., means to push users to rational decisions improving the overall ecological posture of the used services;
3. *Service user actions*, i.e., means for the user to act upon the intended or used service, both prior to and during the service usage.

Note that the user here is not necessarily the end-user, but the respective service user.

Service User Awareness

This feature generally simply tries to account for the native outsourcing nature of ICT: ICT does not work in a compartmentalized fashion, and it is very common to extend the local ICT system through capabilities brought by remote services. As more and more companies and industries migrate towards cloud computing, one could even say that the local ICT system dematerializes completely. This could be rapidly miscounted as reaching net zero locally, however the truth is that the CO_2e emission corresponding to the same service still occurs; it just occurs somewhere else. It remains debatable whether the centralization and specialization in this particular case yields an overall CO_2e gain; on the one hand, the cloud provider certainly has more competence and all interest (TCO) to optimize his own resource footprint to best match the cumulated demand; on the other hand, the remote nature of such service usage makes it hard to account for all CO_2e emissions in the SFC, as, for a fair comparison, one would need to account the use of access and transport networks as well. All this is difficult today, because all ICT domains so far only support domain-wide energy consumption metering. **Service-level energy consumption and CO_2e emission metering is hardly possible anywhere.**

While for the society at large it does not matter, where the CO_2e is accounted for, it does matter for optimizations and reductions, and, hence, it might matter legal entities, regions or states: indeed, if the emissions occur with the respective service provider and the service user is left unaware of the emissions due to his service usage, then optimizing this service usage, the configuration of the service, limiting unnecessary service flows, incentivizing such limitations and other similar measure would be altogether useless, as they all would represent an economic externality for the service user: such optimizations would reduce service provider's OPEX by using less of that provider's carbon credit (e.g., in the EU-ETS regulatory carbon market) without any visible gain for the service user.

To illustrate this, consider as example a user backing up local data to a remote cloud. This user could easily use local data compression (prior to applying any security measures) known to be highly effective to reduce the data volume transmitted through the network. Alas, such compression would increase user's own CO_2e emissions, as doing so requires more effort on the user side (for compression and decompression, respectively). However, if the CO_2e emission of the network is attributed to the network provider, especially under the flat rate pricing, compressing data would actually harm user's own CO_2e footprint. Note how this example shows that a global optimum cannot be reached due to externalities.

That's why we strongly suggest to **raise service user awareness at all stages of the overall SFC**. Hence, each involved service provider should be able to gather energy consumption and, ideally, corresponding CO_2e emissions at the service layer, so as to attribute actually measured quantities to each service instance of interest. This represents a serious departure from the current practice and constitutes a novel type of requirements on the participating service providers. Once this **ecodata** (energy consumption and CO_2e emissions) is gathered, it needs to be exposed to the user. Depending on whether the end-user has knowledge (e.g., through an explicit contract) of a given service function in the SFC or not, ecodata might need to be aggregated or forwarded by various SFC stages, so that it can reach interested stakeholders, for instance end users.

Service User Incentivization

Equipped with the ecodata of used services, the user can make informed decisions, whether local or remote service realization represents a better ecological choice, where and when a service should be used, etc. Future business models could take this into account, so as to provide concrete motivation for rational users to act in an ecology-friendlier way. While the incentivization in the economic sense is not a technological method, ICT needs to support this. A typical example would be **exposure of greener alternative services, service variants or service configurations to the user**, as the user cannot know what levers exist in the service realization of the provider, while provider cannot always use these levers under the existing SLA. In general, only the user can decide to go for a service below the promised SLA; while the incentivization has to find ways, why a service variant would be acceptable or even interesting for the user, the technology has to cater for the existence of such variants and the means for timely exposure of the latter to the users, when it matters (prior, during or post session, etc).

Service user actions

The support for the corresponding actions from the service user is the last piece of the overall service user involvement puzzle. **ICT technology must provide technological means for the users to rapidly (seamlessly) switch to a greener service alternative or variant**.

2.5.5 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 resources or 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 as follows:

- 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 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 allocations has to propagate towards subscriber level and even application level. For instance, an application could use several concurrent allocations during its session in order to best utilise the system as well as to provide superior quality of service compared to any individual allocation. In a view similar to application-driven networking, an application could also explicitly ask for an allocation suitable for 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 allocations 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**. Separation of concerns between Vertical/Network Application Orchestration (VAO/NAO), in terms of organization of micro-services and their interconnection, and Network Functions Virtualization Orchestration (NFVO) will be a facilitator to implement this framework, where slice intent configuration can be conveyed through the mediation of the Operations Support System (OSS).

Research Theme	**Sustainability, Efficiency and Resource Management**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Runtime Service Scheduling (RSS)	Mid-term	Protocols, algorithms, architectures and solutions for dynamic, runtime assignment of resources to tasks, such that the executing system handles each task successfully under that task's specific constraints while explicitly accounting for the resources used by the solution per se and its novel, added constraints. It must achieve overall improvements in: a. either resource usage (energy, capacity, etc) for the same throughput b. or in terms of successful service throughput on an unchanged system Solutions should achieve improvement of: a. Min. 3 times for non-critical services; b. Min. 5 times for services with high guarantees; within systems composed of up to 100k nodes with high task arrival rates from tens of thousands of system nodes.	- Reducing overprovisioning in resources (entities, nodes, links, energy) or, equivalently, increasing system throughput - Improving greenness and sustainability
Conflict Avoidance/ Resolution (CAR)	Mid-term	Protocols, algorithms, architectures and solutions for conflict avoidance or resolution facing concurrent, uncoordinated usage of resources, where the executing system strives to handle tasks successfully under that task's specific constraints, while the throughput of successfully executed critical tasks (i.e., goodput) increases compared to a situation without the said mechanism in place on a fixed resource footprint. Final solutions should demonstrate sublinear scaling of required resources with the increasing load.	- Reducing resource waste - Improving greenness and sustainability - Improving multi-tenancy and governance

		More generally, solutions providing: a. Sublinear increase of the required system resources under the increasing goodput requirement (including the resources required for the proposed solution as such); b. Where this performance improvement is maintained across systems composed of up to 100k nodes.	
Networked Garbage Collection (NGC)	Mid-term	Protocols, algorithms, architectures and solutions for freeing unused, stalled, crashed or incorrect (e.g., partial) allocations on resources. Final solutions capable of: a. Doing this without perceptible negative impact on running task performance, throughput and correctness; b. Freeing, over its runtime, in average more resources than the proposed solution per se uses within the system; c. Supporting systems composed of up to 100k nodes.	- Reducing resource waste - Improving greenness and sustainability
Articulation of needs and provisions from the system to the user/applications (EXP)	Long-term	Protocols, algorithms, architectures and solutions for user-to-system interface, i.e. exposing available resources, their capabilities, requested/authorized status pertaining to the executed application and, in particular, ecodata attributed to the requestor, to the user applications and getting requirements from user applications explicitly or implicitly. Final solutions: a. Supporting legacy applications as much as possible, including solutions for support of applications that use traffic encryption; b. Collecting and attributing relevant ecodata on a service-level, and, if necessary, forwarding eco-data from uplink services, or aggregating eco-data from the remote services used to provide the service instance in question. c. Reusing (extending, integrating, mimicking, maintaining compatibility to) existing methods, where applicable, such as network exposure, ETSI MEC, etc. d. Allow runtime negotiations with compatible, novel applications.	- Improving sustainability and universality - Improving transparency, trustworthiness, governance

2.5.6 Recommendations for Actions

Research Theme	**Sustainability, Efficiency and Resource Management**			
Action	RSS	CAR	NGC	EXP
International Research				X
Cross-domain research	X	X	X	X

2.6 A self-learning, AI-Native, Service Provisioning Infrastructure

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.

2.6.1 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, edge and beyond-edge devices' resources, expected to become integral components of the network infrastructure in, e.g., 6G systems as per Section 3.2.
- AI usage in runtime optimization of service provisioning and coordination mechanisms, or as a replacement to those approaches, providing a dynamic allocation of resources to the services from the overall resource pool as per 3.2, improving the resource multiplexing and hence the OPEX and energy cost of the scheduled services in the system.
- Deploy and provide network services to other operators and/or service providers when requested, or dynamically scaling up and down existing network service based on varying application demands and service requirements, 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 Green Networks – seems 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 (which 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).
- Another challenge is to develop robust AI solutions, which can adapt and adjust to the changes in the environment, with capability of transferring learning from past experiences, also detecting and preventing wrong decision making, when an anomaly occurs in the system. An anomaly in a distributed and complex system, such as a carrier network, could occur due to many reasons, and it is not clear how to detect and react to such anomaly. Should we trust the decision made by the trained policy, or should we fall back to some backup, safer, policy? And how do we detect anomalies in decision making to enable such fallback mechanism? Note that, in some scenarios, we might not be able to detect an anomaly directly but only through its impact on the fitness of decision made by the trained policy. Besides, specific attention should be also paid to the challenges when distributing the learning, e.g., how to enable an efficient learning having access to partial state (environment) information only. This is specifically important if we are deploying deep reinforcement learning techniques.
- In devising control and management strategies (often implying closed-loop control over different time scales), unless applying AI/ML techniques specifically meant to bypass the issue of modelling (in other words, acting as blackboxes, where both system dynamics *and* control are functionally approximated), we need to model a VNF in terms of consumption and performance versus load and configuration [C2-42]. In any case, attempting to model system dynamics does not prevent the application of AI/ML to the synthesis of complex control strategies. A possible approach would be to use models, e.g., at different levels and with different granularity – packet, flow, discrete- (queueing models) or continuous-state (fluid models), where available and feasible, to describe the dynamics of the system, and AI/ML to parametrize the functions expressing optimal control strategies as the solution of Infinite Dimensional Optimization (IDO) – i.e., *functional* – problems (see, e.g., [C2-37]).

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).

2.6.2 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 or parametrized by or even dynamically conceived through AI/ML may also bring significant socio-economic impacts together with improved sustainability models for Network Operators and might constitute a breakthrough in the service development. Indeed, for future platforms and services, AI/ML could be used to dynamically develop suitable network and service functions (NF, SF), e.g., from general stubs or derive best-suitable service-function chains (SFC) both for the operators of the platforms as such and for their users/subscribers. For prosumers, we foresee the proliferation of *personal data platforms* that are tightly connected with network services, and the development of tools allowing any involved principal to control their data and the models established from the latter. This is illustrated in Figure 3-1, where AIaaS is shown as a general block at the new fusion layer between the heterogeneous resources and the cooperative services: in this vision, the service layer is dynamically programmable over a composable infrastructure, with AIaaS kept at the novel abstraction layer. Here, AIaaS could be used to actively change or create service types and instances alike.

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. Consequently, 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 their 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.

2.6.3 Resource-aware AI services

Although numerous AI and ML solutions have been proposed during the last decade for wireless systems, so far, those solutions focus solely on optimizing one or more operational or management aspects and procedures of the networked system. However, as specified above, a necessity can already be identified for a radical evolution for networks in order to introduce a novel paradigm, which will not use AI/ML only for optimizing certain network operations, but will ultimately integrate them as structural enablers of the system itself. To this end, AI will be an integral capability of diverse, heterogeneous network entities, spanning from the core, down to the deep edge; those AI-enabled network entities will be participating with their own heterogeneous types of resources (computation, communication, storage, energy, etc.) in complex operations, both via contributing as well as via consuming resources, towards an AI-as-a-Service (AIaaS) concept [C2-38].

These AI services need to be performed and coordinated in a distributed, or at least decentralized, manner. In a national scale networking context, performing centralized learning is simply too costly, as data would need to be collected and sent all over the network to that centralized entity. Also, such data collection and central training rapidly create bottlenecks and single points of failure in and around the central location. Moreover, if training uses private data, then a centralized learning might not be feasible/authorized.

To this end, an AIaaS operation control framework will be designed and developed, responsible for dynamically switching and selecting among available heterogeneous, and possibly unreliable, resources that will participate in the required AI operations in a distributed manner, targeting to optimize the trade-off that results from the respective resources' selection. Particularly, deep edge devices have usually uncertain and time-varying communication capabilities and are also constrained in computation and storage resources. Besides, distribution of data gathered across these devices could be non-i.i.d (independent and identically distributed). A big challenge therefore is how to perform a reliable learning over such heterogeneous and unreliable set of resources and how to deal with data heterogeneity. [C2-38][C2-39]

Energy is another aspect that should be taken into account. Training a deep neural network model is known to be a very energy-hungry process. Distributing such training on deep edge devices that have limited energy resources or that are running on green energy resources, which are by nature uncertain and time varying, is not very straightforward. How to adjust training complexities to the energy availability at a device? How to benefit from green energy resources? What is the accuracy vs green trade-off?

The widespread adoption of large language models (LLMs) demands significant resources, in terms of communication, computation, memory, and energy, not only for training the models but also for large-scale inference. A large language model can have hundreds of billions or even some trillions of parameters, making the inference (forward-pass) process very resource hungry. Progress toward multi-modal generative AI models, which are usually larger than LLMs and besides text, also accept image, audio and video as their input, further amplify resource requirements. Note that generative AI services are not only used by users for search queries, but being integrated in other fields, such as climate modeling and bioinformatics, slowly but surely making them an essential part of the ICT ecosystem. AI can be extremely relevant at the business levels as well, integrated other concerns (as reliability and safety, e.g.). AI may support an efficient decision-making, e.g., optimise sequencing of activities that run at different IoT/edge nodes, and/or the cloud (referring to critical operations, such as: forecasts/planning in logistics, production, downtimes, resource availability, etc.). An operator can help running these services by sharing the available resources over data plane with AI service providers. However, this is not a straightforward task, given the resource, latency, and accuracy requirements imposed by these services. How to split and distribute such a large model over available resources, including the compute resources at the (deep) edge, without a significant increase in latency? How to adjust model complexity to reduce its inferencing cost (e.g., though parameter quantization or model pruning) without significantly affecting accuracy? What are the accuracy vs. resource footprint vs. latency tradeoffs? How to provide these services in a trustworthy manner, fulfilling also user's privacy requirements? Edge resource constraints bring challenges, but frugal AI methods may provide solutions.

Finally, the most recent cloud computing programming models should be considered. In particular, Function-as-a-Service (FaaS) is very popular because the applications are realized as a composition of short-lived and stateless function calls, which is ideal to deploy following a serverless computing paradigm. Due to the absence of local state and in combination with a flexible container-based virtualization infrastructure, services can be up-/down-scaled in a fast and easy manner and offer fine-grained billing granularity. However, AI/ML applications are very data-intensive, especially during the training phases, which leads to a high communication overhead and unpredictable tail latencies. Can NFV play a role in addressing cold start effects? Where to keep the application's data in a heterogeneous and fast-changing network architecture?

2.6.4 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.
- It is needed an inherently modular and open approach to automation in Software Engineering processes, one that seeks to effectively exploit latest advancements in AI tools, while seamlessly plugging them into existing tools and methodologies. Underlying AI methods will be required to influence the software (e.g., generate code, tests, documentation), but also to edit the model (e.g., assist the user in formalizing requirements, code-to-model, record software behaviour in deployment).
- 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 decision making and reasoning in such distributed environment, providing auditable solutions that may foster future regulations. Complementing this need for supporting the distributed realization of AI/ML (which seems critical for general 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.

- Reliable learning over a pool of either transiently available or generally constrained resources, including specifically deepest edge resources, remains one of the biggest challenges, going beyond the question of full distribution of AI/ML towards full distribution with expendable resources. Besides, AI explainability may be needed in the real-world, requiring additional resources and overcoming problems caused by streaming data.
- Network Functions (NFs) based on AI/ML: In future networks, systems may functionally operate using AI/ML techniques, rather than only traditional methods. As mentioned above, beyond managing a set of running NF instances, AI/ML can be used for automated parameterization of given and even programming of novel NFs. One can imagine code generated in CI/CD pipelines based on observations from NF use, to create more effective versions of NFs, then pushed to deployments to fundamentally improve behavior based on observations over time. In simple cases, configurations could be generated and pushed to production and model updates, where AI/ML engines reside inside the NFs. Rethinking the execution of NFs and code for them, are important areas going forward.
- Cooperative LLM processing: extension of the study related to distributed AI/ML to LLMs and multi-modal generative AI services, with a focus on inferencing tasks with high accuracy and low latency requirements. The idea is to propose protocols and mechanisms that enable the wireless network operators to share their data plane resources with the AI service providers, helping them to run these services in a more cost-efficient manner. (Otherwise, the service providers might need to build their own data centers, which would be very costly, especially for small size companies and start-ups). The challenge is to provide an efficient trade-off between resource footprint vs. latency and accuracy. And all these should be done in a trustworthy environment, preserving user's privacy.
- 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 3.2, 3.3). 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.

- The topic of energy consumption of AI/ML algorithms themselves has started to be investigated. Recent work [C2-40][C2-41] contains some preliminary indications on how to compute the energy consumption during one cycle of inference. Further investigation will be needed, addressing both in-operation and training/adaptation power requirements.
- Novel and better solutions are needed to introduce serverless computing and functional programming aspects in communication and networking architectures, thereby providing the advantages of infinite scalability and flexible orchestration also to AI/ML services. The use of external services, abundant in the cloud but scarce and inefficient in edge/mobile systems, must be redesigned following a frugal approach, for which we need new development patterns and deployment models.

Research Theme	**A self-learning, AI-Native, Service Provisioning Infrastructure**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Distributed, generic AI/ML (platform)	Mid-term	System as distributed joint AI/ML platform for various AI applications, enabling AIaaS, which can create and execute AI services on the fly. Design of APIs for this platform, which would be usable by end-users (e.g., IoT, industry, subscribers), by mobile network operators, and even by the network functions of the system per se, to access, configure, and execute AI services while: a. requiring no exchange of raw data (potentially private, raw data stays at its source) b. reducing resource consumption (communication, computation, memory, energy, etc) c. scaling to systems of up to 100K nodes. Relevant proposals should: a. Design a generic, system-integrated AI/ML platform capable of using up to 100k diverse nodes (different types, capacities, stability profiles) without introducing any single point of failure. b. AIaaS: Design a generic API suitable for both system internal functions (self-optimization) and for different system users (e.g. MNOs, subscribers, verticals) alike. c. Demonstrate that this system can orchestrate and execute AI services over the platform on the fly and in a runtime manner, without requiring human intervention. d. Demonstrate how this system-integrated AI/ML implementation can reduce resource consumption (resources used for learning and inference) on average, i.e. for a set of arbitrary AI/ML tasks, compared to a monolithic centralized AI/ML platform.	-Better locality and governance -Better trustworthiness -Universality of infrastructure -Sustainability of infrastructure
AI/ML computing with and on transient/limited resources	Mid-term	Development of techniques and mechanisms for reliable and robust learning over limited, heterogeneous, and volatile resources that are available at edge and deep edge, which can: a. enable Few-Shot Learning through improving transfer learning capabilities b. scale to networks of 100k nodes and guarantee fairness among different participants (devices, users, etc). Relevant contributions should:	-Further impoving locality and governance -Contribute to digital self-determination, potentially better privacy and data governance -Faster deployment

		a. Show a system achieving stable execution of any AI/ML tasks in spite of potential instability of any used resource, supporting up to 100k nodes and average loads comparable to achievable loads with similar tasks on a system of a comparable stable resource size plus size-independent constant. b. Demonstrate that the system can provide both fairness of participation and fairness of contribution among the participants, guaranteeing that we learn from all data, even those gathered/restored in users with limited resources.	
Cooperative LLM processing	Long-term	Development of protocols and mechanisms for distribution and execution of LLM and generative AI services over shared data plane: a. Enable a reliable and fast execution (inference) of these services over available distributed resources, keeping TTFT (Time To First Token) and TPOT (Time Per Output Token) in the sub-second range, preferably 100-500ms, to ensure that the model is responsive. provide an efficient trade-off between accuracy, latency, and resource footprint (model's complexity), keeping the accuracy in an acceptable range (e.g. 70-90% correctness for general purpose tasks and 95-99% factual correctness for high-stakes tasks such as health-care), while reducing the resource and energy efficiency by a factor of 3-5. b. Scaling to large settings composed of 100K of users. c. Provide a trustworthy environment for executing of these services and improve user's privacy.	-Cost reduction, sustainability -Better trustworthiness -Better privacy -Faster development
Net zero (distributed) AI/ML	Long-term	Development of techniques, protocols and of a new architecture to promote the utilization of renewable energy resources (deployed and distributed all over the network, in the EU often available by energy regulation at the deep edge) for the learning process, so as to: a. Enable a reliable and robust learning, while relying on uncertain and time varying energy resources b. Provide an efficient greenness-accuracy tradeoff c. Scale to large networks d. Preserve privacy Relevant solutions need to: a. Demonstrate a system achieving stable execution of any AI/ML task while relying only on green energy resources that are available and distributed over the network, achieving then the net-zero goal. b. Demonstrate that the system can support large numbers of users and large number of energy domains, and hence is scalable to real world settings.	-Greenness, sustainability -Potentially better privacy
AI/ML driven system updates and evolution (self-updates)	Long-term	Evolutive and automatic adjustment and development of Network and Service Functions. Final solutions should demonstrate a creation of a novel NF type from a stub or e.g. by combining modules from an existing library by AI/ML methods based on the expected features and KPIs.	-Sustainability, universality

2.6.5 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 for **network anomaly/vulnerability pattern identification** and, thus, proven useful in identifying persistent threats/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.

Research Theme	**A self-learning, AI-Native, Service Provisioning Infrastructure**				
Action	Distributed, generic AI/ML	AI/ML on transient / limited resources	Cooperative LLM processing	Net zero (distributed) AI/ML	AI/ML-driven system updates and evolution
International Research			X		X
Open Data		X	X	X	X
Large Trials			X	X	X
Cross-domain research	X	X		X	

2.7 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 3.2 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 2.3.1, 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 2.3 and 2.4 respectively, and resource management considerations, as discussed in Section 2.5, 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 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. The concepts of Joint Sensing and Communication (a.k.a. Integrated Sensing and Communication) are an extreme realization of the potential of exploring a whole world of sensing already embedded in the intrinsics of communications systems, provided by the networking infrastructure itself. 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.** Integrating these trends with the virtualization challenges addressed in Section 2.3 open the area of lightweight virtualization (containers, unikernels,…) as an essential direction that needs to be evolved.

In this regard, novel forms of dynamic edge resource discovery, management and orchestration are required, allowing service provisioning to exploit on-premises deep edge 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. **Future research will need to develop advanced schemes for adaptable and dynamic service provisioning involving deep-edge devices resources, taking into account their dynamic and sometimes hardly predictable availability (moving from deterministic to partly stochastic virtualized platforms).**

2.7.1 Massive Heterogeneous Edge Resources

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 2.2. 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 purpose 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.**

2.7.2 Dynamicity of Edge Resources

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 2.3, **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.

2.7.3 Governance of Edge Resources

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.**

Decentralisation of IoT edge systems has been discussed in several publications, see e.g., [C2-43]. In particular, with the exponential rise in the number of devices, IoT is applied together with edge-centric computing to offer high bandwidth, low latency, and improved connectivity. Moreover, the cloud-centric platforms offer a high amount of data storage, but with deteriorated bandwidth and connectivity that affect the quality of service. The edge-centric Internet of Things-based technologies, such as Multi –Access Computing (MEC) and fog computing, offer distributed and decentralized solutions to resolve the drawbacks of cloud-centric models. However, in order to realize these distributed edge-centric models in the context of 6G, it is needed to realize the concept of decentralized distributed loT Edge Systems, which should at the same time incentivize all the participants in the 6G value chain to share their edge resources.

2.7.4 Edge-Specific Architectural Considerations

The continued growth in video applications including augmented reality (AR) and virtual reality (VR) required by, among others, the emerging applications (cmp. Chapter 2), 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 2.2-2.6).

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 3.3-3.6. 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.

2.7.5 Service Execution on Edge Resources

Processing at the edge in the architecture is essential for ultra-low latency and reliability, while the AI processing is already today often transferred to the mobile 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.

2.7.6 Edge AI

Authors of [C2-44] estimated that 850 ZB of data were generated by people, machines and sensors at the network edge in 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 [C2-45]. 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 concerns 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 [C2-46]. 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 [C2-47], which however is not optimized for working with unbalanced and non-i.i.d. (independent identically distributed) 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 [C2-46], 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 [C2-48], the energy [C2-49]. In model early-exit the inference is performed only with a subset of the network, according to the latency requirements. On the other hand, to reduce 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, 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 solutions 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” [C2-50]. 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 suggested in Section 3.6. The high energy requirements of deep learning solutions suggest that both industry and academia promote the research of more energy efficient AI algorithms [C2-48]. 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 [C2-51] and Green Algorithms [C2-52] should be used to analyse, audit and report the carbon footprint of novel solutions proposed.

2.7.7 Research Challenges

Research challenges in this area include:

- delivery model and APIs, with effective use of ultra-dense and diverse wired and wireless networks (cmp. Sections 6 and 7);

- need of 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. Such management and terminal devices in general are to be integrated with the general architecture vision as per Section 3.2, e.g. as a potential IoT specific "allotment" including both hardware and software objects.
- need of privacy and data management, and the location of processing and data to match legal and moral restrictions on data distribution, access and processing.
- need of policy descriptions, rules and constraints; to be specified in a form that can be enforced by the infrastructure on the services (cmp. Section 2.4).
- IoT architectures applied in 6G, considering the requirements of distributed intelligence at the edge, cognition, artificial intelligence, context awareness, tactile applications, heterogeneous devices, end-to-end capabilities. This is also to be put in the context of the general architecture in Section 2.2.
- Research on distributed intelligence at the edge, cognition, context awareness, tactile applications and integration of heterogeneous devices. Autonomies and distributed intelligence in IoT towards the Internet of Autonomous Things. This a crucial topic for the AI/ML research, and is therefore already mentioned in Section 3.6.
- Need for 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. Dynamic, partly-deterministic and user-centric virtualization, extending service infrastructure to deep-edge devices.
- Need for built-in end-to-end distributed security, trustworthiness and privacy issues in edge computing for IoT, as well as federation and cross-platform service supply for IoT.
- Need for 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.
- Need of deriving 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.
- New 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. See relevant Research Challenges in Section 2.6.
- New AI-IoT integration challenges at the "edge" arise, e.g., new energy- and resource-problems with image recognition at the edge, edge computing implementations (neuromorphic, in-memory, distributed), distributed IoT end-to-end security, swarm intelligence algorithms, etc. See relevant Research Challenges in Section 2.6.
- Need for the design of AI solutions it will be crucial to consider their consumed energy. See Research challenges in Section 2.6.

Research Theme	**Deep Edge, Terminal and IoT Device Integration**		
Research Challenges	Timeline	Key outcomes	Contributions/ Value
IoT architecture	Long-term	A suitable architecture to be executed within the general resource pool as per Section 3.2, customized for the particular needs of IoT, providing: a. Not only individual management of millions of heterogeneous often constrained devices, balancing the needs of the respective organization (efficiency, security, governance) and concerned users/objects (privacy), but also management of collaborative services and tasks executed by the latter. b. Efficient, adaptive, runtime communication environment for particular ultra-dense wireless environments with a capable, multi-modal delivery model. c. Efficient, adaptive, runtime edge computing and swarm intelligence.	
Dynamic, partly-deterministic, user-centric virtualisation	Mid-term	Extension of resource provisioning through dynamic inclusion of deep-edge devices' virtualized resources, to be assessed by: a. extensibility of resources provisioned for a certain CAPEX; b. speed of (re-)configuration of virtualized resources facing change of demand. Final solutions should support: a. Management of dynamic resource provision by deep-edge devices in spite of instability and time limitation of devices' connection b. Working protypes of communication environments, where virtualized resources include deep-edge devices, and planning of their availability takes into account predictions of their future (short-term) availability via stochastic models for resource provisioning c. Large scale systems including management of deep-edge devices are ready for deployment	Flexibility, resource efficiency
Edge intelligence	Mid-term	Particular AI/ML mechanisms suitable for: a. The transient nature of resources in the IoT domain (links, compute resources), cf. Section 3.6. b. The constrained nature of devices (constraints of compute power, of energy, of time) c. Achieving guaranteed convergence of the insight in the swarm environment, i.e. facing the availability of many yet individually weak agents. Design, conceive and demonstrate particular AI/ML mechanisms suitable for: a. The transient nature of resources in the IoT domain (links, compute resources), cf. Section 3.6. b. The constrained nature of devices (constraints of compute power, of energy, of time) c. Achieving guaranteed convergence of the insight in the swarm environment, i.e. facing the availability of many, yet individually weak, agents.	Universality, sustainability, flexibility, potentially better privacy and data governance

2.7.8 Recommendations for Actions

Research Theme	**Deep Edge, Terminal and IoT Device Integration**		
Action	IoT architecture	Dynamic, partly-deterministic, user-centric virtualisation	Edge intelligence
Large Trials	X		
Cross-domain research	X	X	X

2.7.9 Sustainability Considerations

The following taxonomy is used in this section:

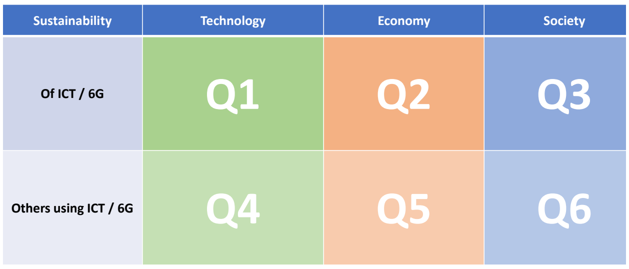


Figure 2- 3 Sustainability impact area

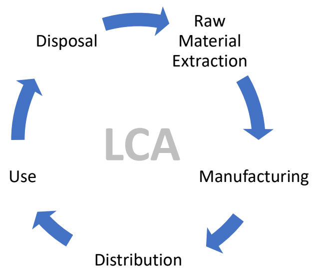


Figure 2- 4 Sustainability lifecycle assessment

The following research topics have clearly articulated relationship to sustainability.

Research Theme /Research Aspect	**Sustainability impact area** Q1-Q6 as per above	**LCA** LCA phase as per above	**Actual impact**
Evolving from Data & Forwarding to Function-Rich User Planes / Architectural frameworks to enable user plane evolution that is economically and technically sustainable	Q1, Q2, Q3	Use phase	A data plane designed with evolution and flexibility in mind promises better sustainability of the deployed infrastructures over time. Helps with sustainability.
Sustainability, Efficiency and Resource Management / Runtime Service Scheduling (RSS)	Q1	Use phase	RSS is dedicated to limiting the resources required for a running task. Dedicated to sustainability – energy or resource efficiency. Can be used to improve CO2e emission reduction.
Sustainability, Efficiency and Resource Management / Conflict Avoidance or Resolution (CAR)	Q1	Use phase	CAR is dedicated to avoid partial allocations and, hence, reduces resource waste due to partial allocations. Helps with sustainability.
Sustainability, Efficiency and Resource Management / Networked Garbage Collection (NGC)	Q1	Use phase	NGC is dedicated to freeing up unused / orphaned allocations. By doing so, it decreases the resource waste. Helps with sustainability.
Sustainability, Efficiency and Resource Management / Articulation of needs and provisions	Q1, Q4	Use phase	EXP explicitly allows and supports, among others, ecodata exchange between the domains. Dedicated to sustainability: required to get user awareness / involvement.

from the system to the user/applications (EXP)			
A self-learning, AI-Native, Service Provisioning Infrastructure/ Net zero (distributed) AI/ML	Q1	Use phase	Net zero AI/ML is dedicated to getting a working AI/ML system without any CO2e emissions. Dedicated to sustainability – reduction of CO2e.

3 Fundamental Enablers for Future 6G Systems

Editor: Luis Miguel Contreras

Key to implementing architectural concepts and visions for the upcoming smart networks and services, realized in the control, service, and user planes of the overall system, is the scoping of the fundamental enablers for such systems and ensuring their evolution through suitable R&I efforts. This chapter outlines the motivation and drivers behind this needed evolution as well as describes those fundamental enablers, without claiming to be exhaustive. Our key messages can be summarized as follows:

- We continue to build upon the system of systems vision that underlies the success of global communication systems so far.
- We, however, recognize the rich and often specialized nature of limited (network) domains, often in the form of technology-rich edge networks that embody the specific requirements of stakeholders developing and deploying them as, for instance, telecom domains, industrial networks, and many others.
- We assert that evolving fundamental enabling protocols and maintaining core ones, e.g., through optimizing existing or even developing novel solutions, is crucial for ensuring sustained innovation towards 6G systems.
- We recognize the importance of the continued evolution of enablers as a means to unlock interoperability barriers for the European community, witnessed through the long-standing European history in relevant innovation through research but also activities in relevant Standard Developing Organizations (SDOs).
- We outline key technology areas in which such that continued evolution will need to happen for the suitable building blocks of future smart networks to emerge, also taking into account deployment and operational considerations during the design phase.
- We derive key research challenges and main actions that will need to be addressed to promote the emergence of innovative solutions embodying those building blocks.

To concretize our key messages, we will first outline the evolution drivers in Section 4.1, embedded into an *interconnected, highly innovation-rich edge system of protocols*. This will be followed in Section 4.2 by a discussion on the enabling technology innovation that we see necessary to realize this vision, followed by the concrete research challenges and recommendations for actions in Sections 4.3 and 4.4, respectively.

3.1 Drivers for the Evolution of Key Enablers

The global communication systems we see today, going far beyond the user-facing Internet, are built on a system of systems vision. We see this vision as a continued foundation moving towards the future digital connected society. Specifically, this vision builds upon but also suitably extends the aspect of interconnecting networks, based on a common and federative protocol suite to foster rich innovation in enabling the various traffic handling solutions across the many access technologies, such as wireless, optical, satellite and others, available, while allowing for the needed multi-domain policy-enabled control of the packet delivery and, ultimately, service provisioning function of the overall networked system. Zittrain [C3-43] and Tarnoff [C3-44] both argue that such system constitutes a generative system with its key characteristic to enable innovation by both being very general purpose and having an extremely low barrier to entry.

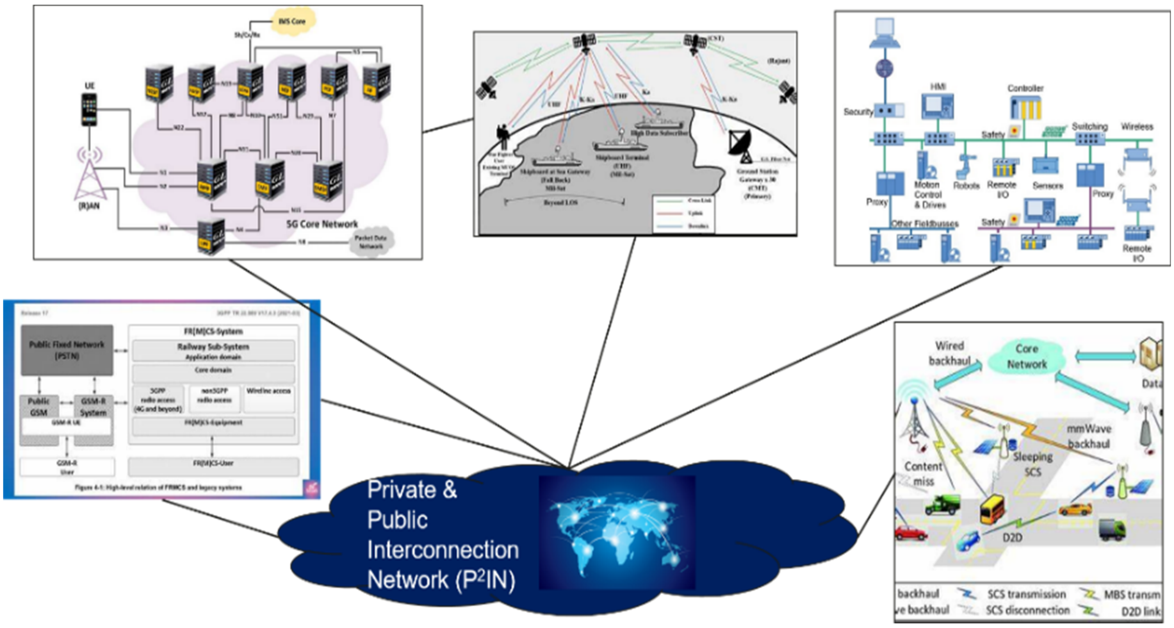


Figure 3-1: System of Systems with Limited Domains

We argue that much innovation stems from the stakeholders creating solutions with global reachability scope but with the very need for interconnection and multi-domain control, defining their own requirements and developing as well as deploying technologies to ultimately communicate end-to-end across the involved systems. Carpenter and Liu [C3-1] define this stakeholder-centric view as a **Limited Domain**, capturing the stakeholder-specific requirements to be embodied in innovation-rich domain-specific networks. Those include (but are not limited to) mobile subsystems, satellite access, industrial, railway, or vehicular networks. They facilitate key aspects for private connectivity facilities and also the public Internet via dedicated interconnection points, as illustrated in Figure 3-1. In doing so, its enabled end-to-end communication suitably supports the service plane realizing the services in the smart networks they embody. In turn, a number of those Limited Domain technologies, such as satellite, may also form the substrate over which to realize the interconnection among Limited Domains themselves.

For this, each Limited Domain controls, realizes, and manages their own methods and protocols as well as their domain-specific traffic handling methods, e.g., for ensuring and managing link diversity, while those methods need interoperation for a proper, secure, and efficient **end-to-end multi-domain communication** to happen, thus enable the generative system [C3-43][C3-44] refer to. Furthermore, emerging networks, such as hybrid quantum networks, may pose new challenges [C3-40], building on or demanding entirely new methods and protocols to be added to the rich mix of existing limited domains.

It is important to note that Limited Domains differ from the concept of an **Autonomous System** (AS) by embodying a suite of technologies, each of which defining their stakeholder-specific requirements, rather than embodying a deployed network domain under the control of a specific owner. As such, the concept of a Limited Domain more appropriately embodies the methods that will make up future smart networks.

Those requirements may be reflected in domain-specific **protocol solutions** (often as extensions to existing solutions) that provide lower latency and/or higher capacity, utilize multi-path capabilities of the (Limited Domain) network environment for higher resilience and others. Key goal is often the reduction of costs in addressing those requirements in the specific network environment in which the solutions are deployed, while supporting new capabilities introduced by advances in, e.g., Network Slicing.

Protocol evolution is important to enable truly smart(er) and more sustainable networks and services for 6G and beyond. Here, we do not see evolution being a limiting concept, only pushing for small, incremental least common denominator solutions, but instead see this evolution happening in (possibly larger) step changes that are driven by the specific needs of stakeholders deploying innovative solutions for their specific needs. This aligns with the views expressed by Trossen, Carpenter, and Crowcroft [C3-2], identifying Limited Domains as a crucial source for innovation in the overall internetworked system, not limited to the public Internet but facilitating many technology solutions we see, for instance, in the mobile subsystem for 5G/6G and industrial networks.

Such innovation is possible assuming ***clear and interoperable interfaces*** being in place between the various Limited Domains and avoiding imposing constraints on external domain for the delivery of an added value services to devices and applications connected to a Limited Domain. This position is exemplified by the many research and development initiatives that ultimately lead to the definition of suitable standards, such as those driving the evolution in mobile systems from 1G to currently deployed 5G Systems (defined by the 3GPP), Layer 2 technologies (IEEE), and (edge) compute systems (ETSI among others), alongside the Internet standardization community in the IETF with significant evolutions of key IP-related and transport layer protocols, even after the latest 'big' step change from IPv4 to IPv6. Furthermore, a sufficiently generative system [C3-43][C3-44] – that is, one which is as general purpose and having as low a barrier to entry as IP – is also a key enabler of continued innovation in the Open-Source community, which the European Commission reports as crucial to European Digital Sovereignty [C3-44].

We can observe evidence of those innovations in standards related to key protocols most Limited Domain networks, including mobile subsystems, utilize. For instance, since the introduction of IPv6 as the basis for addressing in the Internet in the form of RFC 2460 in 1998 (signifying the last big 'step change' in Internet protocols at layer 3 from Version 4 to the now increasingly proliferating Version 6, later updated in RFC 8200), the number of RFCs has increased beyond 9000 in 2023, largely denoting 'protocol track' contributions. Here, many key advances to protocol technologies have been made to areas such as new transport protocols (e.g., QUIC[2]), new routing extensions (e.g., Segment Routing for IPv6[3] but also including new technologies like information-centric networking [C3-33]), many addressing extensions beyond pure endpoint locators (see [C3-3] for a comprehensive overview of those) and more. Many of those developments have been driven by explicit liaisons created with other SDOs, such as the 3GPP, ETSI, and others, thus being embedded in the rich set of the system of systems we outlined in our aforementioned vision.

More so, the many protocols transcend across the various rich limited domains that utilize them. For instance, the core 3GPP specification TS23.501 [C3-25] contains 43 (out of 189) references to IETF specifications, including addressing, routing, transport protocols, HTTP extensions, and security, thus, showing the enormously important role that protocols and their evolution plays for current 5G and future 6G systems to come.

Existing, emerging, and new-to-be-developed protocols are key to realizing Limited Domains but also to binding networked systems together to enable the communication between distributed resources within the system of systems.

---

[2] https://datatracker.ietf.org/doc/rfc9000/

[3] https://datatracker.ietf.org/doc/rfc8754/, among others

We argue that an understanding of enabling methods and protocols is, therefore, imperative to sustainably evolve interconnected networks, not only Internet technologies alone. Key property here is the flexibility by design of developed protocols to serve as building blocks to be realized not simply in one but across several architectures that embody the requirements of the stakeholders deploying them as Limited Domain networks, and beyond. We can see this building block nature well utilized in existing Limited Domains, such as in the usage of IP-based solutions in mobile subsystems, such as documented in the aforementioned TS23.501 [C3-25] for 5G, arranged in a quite different architectural manner than in fixed broadband system that enables Internet access, while yet differently used in an industrial network deployment such as IoT or even emerging satellite networks.

This building block nature is meant to prevent stifling innovation in Limited Domains, thus, guiding them towards a fruitful ecosystem for the system of systems that embodies the digital connected society we aim at building. *For that, the continued evolution of and innovation in protocol technologies, not only those used by the Internet community, is crucial to sustain integration in continuously evolving environments of future use cases to come*.

However, the building block nature also poses a crucial challenge in how well they may integrate if originating from seemingly disparate efforts by different stakeholder sets, particularly in terms of co-existing, e.g., in the joint backbone used for interconnecting between the Limited Domains. Today's communication system already bears witness regarding those challenges, with Section 3.4 of [C3-2] providing many examples for those challenges. *Thus, to ensure the aforementioned desirable evolution of fundamental enabling protocols, we must also ensure their interworking in a complex, interconnected system.*

The liaison model used by SDOs, attempting to actively rely on other SDOs for utilization of solutions in their specific domain, is one approach but there may also exist bigger, architectural opportunities at heart of this issue, which ties into the architectural considerations discussed in Section 3. Feeding those architectural insights back into the protocol engineering community is key to success, e.g., through the formulation of best current practices (an approach long being utilized in the IETF, for instance).

Europe has shown strong leadership in such SDOs effort. For instance, according to data relating to the IETF up until 2020, as published in [C3-10], we can find (industrial and academic) European researchers responsible for around 40% of RFC authorship in 2020, up from about only 20% in 2001, with a consistent share of about 10% of (worldwide) academic authors with 4 European academic institutions in the top 10 academic contributors. Furthermore, IETF leadership, represented as the IETF chair, members of the Internet Architecture Board (IAB), Area Directors or WG Chairs, have repeatedly featured European researchers, with a ratio of up to 40%, aligned with the contribution rate to RFCs. European organisations, both industrial as well as academic, have played a key role in that protocol evolution, with key industry players like Nokia, Alcatel (until merger with Nokia), and Ericsson not just holding key IETF positions but also raking among the top ten authorship organisations for IETF contributions [C3-10].

But at the beginning of any new solution that will enable novel capabilities and/or use cases stands the creation of key protocol innovations that ultimately may make it into relevant SDOs, many of which have come in the past out of European research, such as SIP [C3-14], MPTCP, e.g., [C3-11,12,13], L4S [C3-15], NDP [C3-16], EQDS [C3-9], to name just a few.

This also serves to illustrate the relationship to Europe's Digital Sovereignty – many protocol innovations build upon Open Source; MPTCP's reference implementation notably is within the Linux kernel. The low barrier to both innovating on open standards, as well as implementing such innovations in Open-Source operating systems provides for an exceptionally fast path from research to deployment in production systems.

The key takeaway is that Europe has been leading open research, standards, and solution development for new and evolved protocols, such as through open source, as part of its longstanding history of collaborative, publicly funded research as a concerted effort to ensure European leadership in digital communication technologies through efforts in HorizonEurope and other instruments; this must thus continue in order to ensure successful and well-paced protocol evolution for future Smart Networks.

3.2 Fundamental Enabling Technology Areas

We now discuss the key research areas enabling an innovation-rich system of systems. To start, let us briefly introduce those areas, visualized in Figure 3-2:

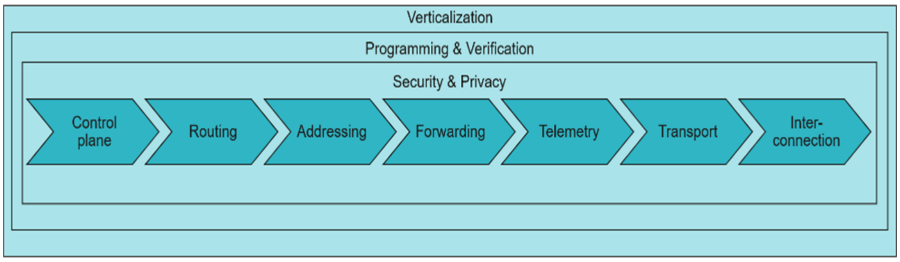


Figure 3 2 Key Research Areas

- *Control plane*: As the basis for any communication to take place, bootstrapping mechanisms (in the form of protocols) are required for building suitable control mechanisms, considering the specific limited domain requirements, while providing flexibility for change that new use cases often demand. Here, the notion of an ephemeral limited domain needs support (e.g., for pervasive computing type of scenarios), bootstrapped at runtime from an ad-hoc formed resource pool for a limited period of time, only to be disbanded after fulfilment of the desired use case. As one key aspect, it is imperative that participating endpoints have suitable interfaces to recognize and manage the relationship with the domain to which they happen to attach. This is particularly crucial when considering endpoints that may 'live' in many Limited Domains, often established for just a limited set of use cases, thus allow for those endpoints to suitably traverse such domains through suitably constraining the endpoint participation in that domain, possibly represented through a Digital Twin at the management layer.
- *Routing*: Once having established a suitable control plane, relations can be established to facilitate the forwarding of packets from one network endpoint to another. The Internet has been built upon the idea of acting upon distributed state for such forwarding, relying on routing protocols to establish that state. However, we have seen novel approaches to, e.g., opportunistic routing, path computation routing, segment routing, and others, that provide alternatives in order to cater to requirements of emerging use cases. Here, the use of concurrent (multi-)path as well as (efficient) multipoint distribution are particularly important, considering the communication patterns of emerging, e.g., AI-centric, use cases, thus enabling transport networks for systems like 5G and beyond that would support use cases such as URLLC, private networks, and others. A key constraint in many Limited Domain is the support for mobility of the connected end systems, which must be suitably supported by the suitable routing and mobility management mechanisms.
- *Addressing*: Key to any communication, based on routing decisions and the suitable forwarding actions, explained next, is the nature of addressing WHAT needs to be exchanged HOW. The

authors in [C3-3] have outlined the problems arising from the addressing ambiguity that IP addresses provide. While rich in possibilities, it has led to a plethora of often competing (and conflicting) extensions to the original endpoint addressing that IP introduced many years ago. [C3-3] posits that it requires an architectural approach to re-think the use of IP addresses, while not hindering the innovation that may be brought into the addressing scheme after its original design.

- *Forwarding*: Once suitable state exists in intermediary network nodes to facilitate an end-to-end exchange of packets, in-network operations can be realized. Those operations have long focused on the most efficient forwarding of packets from an input to an output port. However, with communication semantics being enriched from simple unicast to include group and collective communication, forwarding becomes richer or at least more varied than the widely realized longest prefix match we see in IP networks. Here, among others, information-centric strategies for forwarding but also more efficient bitfield-encoded path-based forwarding strategies have been explored. Programmability will be key to such wider range of forwarding behaviours to enable the deployment of novel solutions without the need for renewed capital expenditure through new hardware, while programmability also opens the route to include other-than-forwarding operations into the mix of operations for intermediary network elements; a trend captured as *in-network computing* in alignment to the generalized resource-centric view presented in the architectural considerations in Chapter 2. Key here, however, is to accommodate the desire of linespeed operation, reflected in the limited cycles-per-packet that one can afford for more complex operations on transferred data.
- *Telemetry*: Knowledge about network operation, including performance and abnormalities, is crucial not just for network management but methods for protocols in general, specifically to routing and forwarding. For instance, optimizations of transport protocols using telemetry information (on transferred bytes along a path) has long been recognized as key for improving network utilization. This, however, requires suitable methods to provide the necessary data needed by those protocols. Key here are methods that may not rely on control plane or out-of-band approaches but use in-band signalling of necessary data upon which fast(er) decision, e.g., on congestion control, can be made.
- *Transport*: Building upon a working control plane infrastructure and the enabling routing and forwarding mechanisms, end-to-end transport may commence. While the Internet has successfully built a (transport) protocol regime that allows the many variants of protocols to co-exist (called *TCP friendliness*), this principle is not applied to environments such as data centres or others. Equally, we can see a move away from a single end-to-end relation towards a set of chained connections, each of which may realize different resource management schemes. Furthermore, the increasing demand for optimizing network utilization through methods of packet spraying, packet trimming or many others, has led to the development of improved solutions for transport, all of which need to co-exist in the shared resource pool of computation and communication that we envision in the future. It is at this level of the protocol stack that we see most impact of the notion of a Limited Domain, embodying the domain-specific resource management regime at an E2E level, while interacting with possibly domain-specific network mechanisms, such as switch support for packet trimming in data centres, when doing so. Ensuring that those mechanisms work

properly across several Limited Domains is a challenge to address, including to supply proper user-network interfaces to allow for configuring participating network elements to ensure the desired end-to-end service.

- *Interconnection*: Interconnection has increasingly moved beyond the current best effort IP delivery, offering not just **delivery guarantees** but also the efficient use of lower layer, such as optical, technologies. The ability to directly use those resources is crucial, possibly leading to interconnection that does not rise to the level of IP at all for efficiency purposes, while other forms of interconnections may reside atop available IP inter-domain links at the transport level. Furthermore, we see a move from a well-established peering backbone with long-running AS-to-AS relations to **dynamically formed interconnection relations**, often temporally limited and/or created catering to specific application requirements. Here, participating domains may even differ in their specific networking semantics, thus requiring methods to dynamically bridge across them to ensure end-to-end communication.

In addition to the above core areas, we also highlight three complementary, system-level challenges that need addressing to ensure suitably evolution. These are briefly outlined next with references to the relevant SRIA chapters in which they are discussed in more detail. We refer the reader to those complementary chapters for more insights:

- *Security & privacy*: The considerations for security are key to any protocol development nowadays, with SDOs like the IETF requiring separate security considerations for each developed solution. But protocols and their usage also pose **privacy challenges**, such as through leaking information to parties outside the direct communication relation, e.g., in the form of protocol numbers, IP addresses and alike, enabling business models (e.g., based on IP geo-locating) but also threat models that need strong consideration in any protocol evolution. Chapter 4 of this document discusses the protocol-related as well as system-level security consideration that need addressing in future developments.
- *Programming & verification*: The realization of developed protocols and formats has tremendously advanced in recent year. Although 'running code' is still key to acceptance of solutions in SDOs like the IETF, the often mission-critical nature of protocols necessitates a more **formal** and **rigorous** approach to verify the current, thus intended operation. This links into the more formal understanding of architectural concepts, as discussed in Section 2.1, but focusses on the aspects of the system for which new solution(s) have been developed. Efforts such as the ELVER[4] project aim to develop such rigorous, mathematically rooted, models and approaches, developing tools and frameworks to build better and more secure computer systems. Section 5.7 discusses those aspects in more detail for the development of secure and safe protocols.
- *Benchmarking*: Comparing protocol performance is an important step in finding the right choice for adopting a solution in a specific system. SDOs have long established efforts on benchmarking, with the IETF having formed a dedicated benchmarking working group[5], while the operations

[4] https://www.cl.cam.ac.uk/~pes20/index-elver.html
[5] https://datatracker.ietf.org/wg/bmwg/about/

community is also active through organizations such as NANOG[6], RIPE[7], and APRICOT[8]. The main challenge in those efforts is that performance characterization must be done and validated only with controlled stimuli in a laboratory environment. The benefits of benchmarking-based measurements are to improve customer quality of experience and service and ensure the competitiveness of networking technologies to optimize customer satisfaction. In addition, it allows to identify network problems and issues or identify the best performing technologies. This enables decision making on where to employ networking technologies and what KPIs to improve. Thus, with efforts having only started recently at SDO level, much is to be done to suitably capture emerging technologies, while also needing to develop methodologies for a benchmark-based research and experimentation approach.

- *Verticalization*: The power of protocol solutions lies in their building block nature to be used often in widely varying contexts. The Internet Protocol (IP) is an excellent example, having penetrated many vertical industries, such mobile subsystems, industrial networks, railways, vehicular, and many others. However, the differing usage in those environments also places a strain by often introducing changes to the original design due to domain-specific requirements, such as through extensions to allow for deterministic bounds to IP packet delivery[9] or through realizing similar capabilities at lower, e.g., Layer2, level through Time-Sensitive Networking[10] capabilities. We need to better understand (a) what those domain-specific requirements for future cross-verticals protocols are and (b) how we can design protocol solutions or extend existing ones to accommodate those future requirements without falling into the trap of feature creep. Possibly even more important from an operational complexity perspective is the pruning of capabilities based on well documented and understood development and deployment insights, thus avoiding overloading protocols and solutions with an overbearing need for backward compatibility that has not materialized in real life deployments.

Let us now discuss the above technology areas in more detail as to their need for and direction towards evolution beyond existing solutions, allowing for formulating a research agenda across those areas.

3.2.1 Control Plane

The operation of Limited Domains often depends on a proper working control plane, e.g., SDN control, Virtualized Infrastructure Management, programming of switches, for which in turn suitable protocols needs to be provided. Due to the complexity of network infrastructures, autonomic management solutions are required that work in a zero-touch manner, i.e., without any manual configuration or administration and self-adapting to changing conditions such as link or node additions and failures [C3-18,C2-19]. *Control plane connectivity that does not depend on any prior configuration* is a key element to not only avoid network failures

[6] https://www.nanog.org/

[7] https://ripe.net/

[8] https://www.apricot.net/

[9] https://datatracker.ietf.org/wg/detnet/about/

[10] https://1.ieee802.org/tsn/

caused by configuration mistakes, but also to be able to maintain or regain control over the networked resources after severe failures.

Another aspect is that ***control across Limited Domains*** may be required, especially for virtual providers that want to combine resources and services from different Limited Domains. A control plane connectivity solution that allows for accessing dedicated subsets of LD resources would easily support combining them for providing a new service.

One may consider a solution to provide scalable **zero-touch control plane** connectivity as foundation for resilient systems: regaining control over the networked resources after experiencing disrupting challenges is key to overall system health. Especially, if one considers today's rather complex infrastructures that use SDN control, practical deployments (e.g., see [C3-34]) typically have their own fallback connectivity solutions for this control plane that avoid further dependencies (but are not necessarily zero-touch). However, having an interoperable single resilient connectivity solution for the control plane would provide huge benefits, especially when virtual resource pools should be controlled that span different limited domains.

More so, nomadic networks are considered in some 6G scenarios that may also form in an ad-hoc manner. These scenarios require a completely **self-organized discovery and control** of resources, because there is no fixed infrastructure that the resources inside the nomadic networks can rely on. Additionally, finding topologically nearby resources or those that are only lightly loaded can be challenging, especially when directly integrated with routing as the resource state may change quite dynamically.

An example for an approach that provides scalable, zero-touch connectivity within Limited Domains or even across different Limited Domains for such, and other, scenarios is KIRA [C3-8]. It provides a basis to support the vision of the **unified controllability** of a resource pool and its programmability (see Chapter 3). Here, every resource needs to run a KIRA instance. Pervasive resilient autonomic resource control requires not only resilient zero-touch connectivity, but also mechanisms for resource (and service) discovery, for which KIRA provides an integrated Distributed Hash Table (DHT) to allow for simple resource and service registration and lookup. Dynamically created distributed controllers can use the DHT to let them rendezvous and coordinate as well as to discover resources and other supporting services. The zero-touch approach of KIRA is based on **self-assigned addresses** and has no dependencies. However, security mechanisms such as certificates or assignment to specific Limited Domains may require some form of prior configuration that is opposed to the zero-touch approach. Alternatively, some information can be configured in a further bootstrapping stage by the help of some dedicated bootstrap services that support automatic configuration. Similarly, it may be advantageous to also support automatic software updates and thereby enabling continuous deployment as well as disseminating (configuration) information to all resources within the whole Limited Domain or just to a resource subset. The latter would require some kind of autonomic **group and membership management**. Here, work similar to or extending that on decentralized identifiers[11] (DIDs) may be used as the basis for verifiable credentials[12] (VCs) that allow for the kind of **decentralized authorization and access control** that would complement the autonomic infrastructure formation. Moreover, as already pointed out in the routing discussion of Section 3.2.2.2, **distributed consensus systems** may be used, not just routing state convergence, but also decentralized identifier verification and assignment.

[11] https://www.w3.org/TR/did-core/

[12] https://www.w3.org/TR/vc-data-model/

3.2.2 Routing

Routing has long evolved from the shortest-path support for unicast communication that we often see as a foundation for end-to-end communication. This evolution has been and will continue to be driven by three major aspects, namely the support for a range of communication semantics beyond unicast, discussed in Section 4.2.1.1, the support for richer constrained routing decisions beyond shortest path, discussed in Section 4.2.1.2, and the tussle between realizing evolved routing capabilities at the network underlay or as a service overlay, discussed in Section 4.2.1.3.

3.2.2.1 Communication Semantics

Communication semantics of the many services we expect from future smart networks (see Section 2) can vary widely beyond a single endpoint-to-endpoint relation.

Specifically, the services discussed in Section 2 and the architectural considerations in Section 3 also impact how routing will need to evolve to accommodate the expected service-rich environments in future networks, enabling native communication between compute and network resources alike. As an example, recent work on anycast [C3-5][C3-41] moves away from static DNS-based assignments of IP anycast to unicast addresses, incorporating instead service information into the addressing (see also Section 3.2.3.1) information, utilized by multi-constrained routing approaches (such as those outlined in [C3-3], suitably extended to anycast relations) to make dynamic network layer decisions on which (service-centric) to choose for an individual packet. Problems that routing approaches need to address here is how to ensure the consistency of routing decisions across service-level relations, which are usually not visible at the network layer (i.e., a packet sent to a chosen anycast endpoint may be followed by many other packets belonging to the same 'service session'). Some of the works on service-centrism have their provenance in information-centric networking (ICN) approaches [C3-23] to routing, where information may not just comprise services but also content or pieces thereof, providing an alternative to existing CDN approaches at the network layer. Although previous ICN research has struggled in its wider vision to replace IP as the foundational network technology, its applicability to limited domains, such as IoT networks, still applies, while key insights, such as dynamic forwarding decisions, constraint-based routing decisions and efficient encodings of addressing information have nonetheless found entry into other non-ICN solutions, see [C3-3] for impact on addressing or [C3-21] for a replacement to IP (multicast) routing based on ICN concepts; a solution that is ultimately listed as one possible realization of the service-based architecture (SBA) in 5G.

Already the original ambition of the Internet outlined the need for supporting at the network layer not just unicast, but also multicast and group communication in general. As discussed in [C3-6], many of the semantics beyond unicast, including the realization of anycast relations, have transitioned in the public Internet towards being realized in parallel commercial architectures provided by large cloud and content delivery network (CDN) players alike, putting additional cost pressure on the underlying unicast-centric network architecture. For instance, OTT video delivery experiences a linear cost curve with growing viewing numbers, a trend that can only be broken by multicast capabilities in the network. As argued in [C3-20] and evidenced in numerous academic and SDO works in recent years, the case for network-level multicast support, particularly within multi-path rich Limited Domains, is being revisited with new emerging solutions, such as the Bit Indexed Explicit Replication (BIER[13]) efforts in the IETF, the use of source routing for multicast (MSR6[14]), or technologies

[13] https://datatracker.ietf.org/wg/bier/about/

[14] https://datatracker.ietf.org/wg/msr6/about/

developed in previous EU-funded efforts, such as [C3-22], showing the ability to significantly reduce future traffic volume by not just utilizing multipoint relations at the network layer but building those relations rapidly.

The latter ability is also key to driving the native support for collective communication into the network layer, replacing current endpoint-centric replication (through middleware solutions like MPI and others) through native and ad-hoc replication in the network. Doing so with standardized solutions will be key to stem the expected explosion of traffic through distributed, federated AI solutions through their scatter-gather semantics.

Although rich communication semantics have also their place in interconnecting Limited Domains (which we discuss separately in Section 3.1), we see those Limited Domains as a first call for developing and deploying novel routing solutions, as can also be seen through the extensions to IPv6 that the comprehensive survey in [C3-3] provides. This scope for extensibility also impacts the methods used for future support for rich addressing (Section 3.2.3) and forwarding (Section 3.2.4) methods.

The key takeaway is that the richer communication semantics of future smart services will impact the protocols and methods used for routing as the key capability exposed by the network over which those service will communicate – unicast as a single method of today's Internet will not suffice.

3.2.2.2 Moving beyond Shortest-Path

Several Limited Domains may be designed to carry traffic belonging to different services with distinct specifications in terms of traffic performance, reliability and robustness, as is the case of networks of satellites. Moreover, there is the need to develop routing strategies able to consider different service semantics described by a combination of fields in the packet header as well as a transported set of instructions. Such routing strategies require a data plane able to support programmable network functions (e.g., forwarding) and services.

One approach to achieving richer path selection is the use of additional semantics in the packet header (see also Section 3.2.3.1 on the richness of such semantics), allowing packets from different services to be marked for different treatment in the network. The packets may then be routed onto different paths according to the capabilities and states of the network links and nodes, in order to meet the performance requirements. For example, one service may need low latency, while another may require ultra-low jitter, and a third may demand very high bandwidth. Possible methods for such semantic-rich path selection may include i) using addresses to identify different device types so that their traffic may be handled differently; ii) expressing how a packet should be handled as it is forwarded through the network; iii) enable Service Function Chaining (SFC)[15]; iv) forwarding packets based on carried data rather than the destination addresses; v) or formatting geographic location information within addresses.

A key focus to those new, possibly semantic-rich, routing approaches is the path selection as a key function of routing, which has seen an evolution from the foundational Dijkstra (shortest path) algorithm. Key here is the basis for making the path selection decision, both in the dimensions of frequency of decision as well as constraint used for the decision.

As for the former, current approaches to IP anycast, for instance, often utilize country information (through the hierarchical DNS) to, for instance, choose a German IP address for a search engine when the (DNS) request to it originated in Germany. The scenarios for service-centric routing at the network layer, however, are much

[15] https://datatracker.ietf.org/wg/sfc/about/

richer in their relationship nature, particularly in terms of dynamicity, where the selection of the 'best' service endpoint may change much faster than a DNS update may ever propagate, thus asking for faster, possibly routing convergence or centralized controller-based approaches that ultimately define the packet forwarding decision. Works like those in [C3-5,24], among others, have shown that fast scheduling decisions can be realized at the network layer, significantly increasing the efficiency of distributed service environments that are driven by the flexibility of virtualization to establish new service endpoints in different network locations.

Also, the basis for those dynamic decisions changes, extending beyond currently used network-centric approaches such as ECMP (equal cost multi-path routing). For instance, the authors in [C3-4] have shown that a partial linear order may capture much richer semantics, e.g., computational capabilities, for routing decisions, including anycast ones. This is showcased in works such as [C3-5] for computational semantics, while [C3-7] utilizes energy constraints to make energy-aware decisions in federated learning scenarios for future 6G systems. SDOs like the IETF have picked up on application- and compute-awareness (e.g., in WGs like CATS[16]). This recognition of considering more than 'just' network parameters into a routing decision, however, poses the challenge of communication often rich service semantics to the network; this is not just a protocol problem but also one at the techno-economic level when such semantics cross system boundaries from the service (or application) provider to the network provider in a vertically disaggregated communication environment; suitable solutions for trust management and minimization of information conveyance, including the avoidance of leakage, are needed here.

Future smart networks will need to support, e.g., IoT smart devices but also highly dynamic computing networks for, e.g., federated learning among end user devices, requiring ultra-reliable, low latency, and high-capacity communication, while also providing the dynamic traffic resulting from the expected mixture of compute, storage, and network resources discussed in Section 3. The expected growth in network traffic and increase in dynamicity of communication relations in those next-generation networks may give rise to new challenges for routing strategy design that will contrast against traditional routing, i.e., convergence over a distributed state. One such direction of approaches may utilize AI-based learning methods, where a federation of AI workers utilizes path and compute utilization to determine the (time-dependent) 'best' path(s) to be used depending on the current network demand and network status, Here, machine learning based routing solutions are utilized to enable learning from past information [C3-28]. Several AI-routing algorithms have already been suggested, such as using LSTM for short term traffic trends prediction [C3-26], while work in [C3-27] realizes routing optimization for wireless sensor network to improve resource and energy efficiency, using reinforcement learning to cluster to nodes based on traffic density, energy and delay. Also, as demonstrated in [C3-29,30,31], machine learning for wireless transport networks may consider the traffic demand and the interferences between active links, leading to a solution choosing the relevant shortest path, while considering the interaction between the active demand.

When it comes to maintaining as well as agreeing on the distributed routing state to forward packets over, the work in [C3-17] suggests using consensus methods provided through a distributed consensus system (DCS), possibly supported by methods developed in groups like the ETSI PDL (permissioned distributed ledger) ISG[17], while the work in [C3-42] pursues an information-centric synchronization approach to attain a converged routing state.

[16] https://datatracker.ietf.org/wg/cats/about/

[17] https://www.etsi.org/committee/pdl

The key takeaway for future routing solutions is that we see the need for much faster, higher frequency decisions for communication relations, while the constraints for choosing the right set of communications will become richer, including service and application-level constraints; this trend will clearly impact the routing protocol space with a need for new solutions to be developed.

3.2.2.3 Underlay or Overlay?

The deployment of network equipment is an expensive capital investment that creates the propensity to push any novel routing solution towards being realized as an overlay, running atop the existing underlay network. Conversely, implementing highspeed packet operations at the overlay impacts the overall solution efficiency, thus favouring underlay solutions.

Concepts such as IPv6 extension headers (as defined in RFC2460 originally, later updated in RFC8200) allow for realizing routing extensions as de-facto overlays albeit close to the packet processing pipeline of IPv6. Specifically, choosing destination (extension) headers lets the packet follow an overlay of enhancing packet processing nodes within the overlay that deploys the new routing solution. Complementing this with techniques like eBPF[18] even allows for pushing packet processing to the Linux kernel or towards an eBPF-enabled NIC in a SW-centric router deployment, thus increasing the possible throughput experienced new solution. Alternatively, programmable data planes, as discussed more in Section 3.2.4.2, may be used for further processing improvements.

Key is that the decision at which 'shim' level of the network layer to implement novel routing solution must consider deployment aspects at the time of designing and ultimately implementing the solution since an overlay vs underlay decision also impacts the possible support (of the new routing solution) through a new market player outside the underlay network operator; we can see those considerations in ongoing SDO work, such as those on CATS and LISP[19] in the IETF, among others.

However, we believe that the flexible support of diverse routing paradigms realized as overlay or underlay solutions can accelerate the integration of novel networking technologies in legacy networks. For instance, the incremental deployment of clean-slate network architectures, such as ICN, is a realistic approach to embody the advantages of next-generation technologies in current networking solutions, thus reshaping the future Internet in a controlled and cost-effective fashion. Assuming the use case of a CDN provider that wants to enrich the set of offered services by incorporating ICN-related functions, the incremental deployment can take place in one of the following indicative scenarios.

In the first scenario, the CDN provider deploys ICN as overlay on top of IP-based protocols, such as IPv6 or SDN, in order to support beyond-unicast transport mechanisms within the CDN. In this scenario, the ICN-enabled CDN is transparent to the end-hosts, thus requiring no modifications at the end-users, yet requiring an interfacing of the IP networking stack (over the MAC layer) to the ICN stack at the boundaries of the CDN networks, namely, the CDN gateway nodes. The interfacing of the two architectures at the gateway nodes is a quite stressful requirement, mainly due to the stateful session-based design of TCP [C3-46]. Therefore, while the interfacing of the TCP flows constitutes a fine-grained solution, it also presents challenging complexity due to the plurality of semantics in TCP connections.

---

[18] https://en.wikipedia.org/wiki/EBPF

[19] https://datatracker.ietf.org/wg/lisp/about/

In the second scenario, the CDN provider deploys ICN as an underlay of an (architecture-agnostic) data dissemination application-level protocol. A data-oriented information model for disseminating data in beyond-unicast paradigms, such as the NGSI-LD API [C3-47], which is available to IP-based end-hosts over HTTP, places the communication semantics at the application-level. In addition, NGSI-LD does not introduce sessions and tightly-coupled end-users, thus significantly minimizing the interfacing complexity and, in turn, rendering the incremental deployment more practical. Thereafter, the gateway nodes convert stateless HTTP requests that convey NGSI-LD requests from loosely-coupled end-hosts to stateless ICN-based messages that convey the NGSI-LD messages to the CDN nodes and vice versa.

The key takeaway on the question of deploying novel routing solutions is that the gap, in terms of processing capabilities, between an overlay and underlay solution is closing, thus creating commercially viable cases for pushing routing beyond the sunk investment of existing (e.g., IP-based) routers.

3.2.3 Addressing

Communication at the network layer has long been utilizing so-called locator addressing, where endpoint locators are being used in routing solutions to build forwarding tables in intermediary nodes, enabling to transmit a packet from a (network) source to a destination.

The Internet protocol, both in its original v4 and the currently increasingly deployed v6, utilizes the IP address format, such as defined in RFC2460 (with updates in RC8200) for IPv6, with its well-known address field for source and destination.

However, as documented comprehensively in [C3-3], several misalignments have been identified between this original Internet addressing model and the desirable features stemming from different limited domains, resulting in many extensions to that original model. Those extensions have been proposed mostly in Internet edges, i.e., limited domains, since the reliability and scale at the core of the Internet makes the introduction of novel mechanisms more disruptive. On the other hand, extensions can be introduced more easily at Internet edges due to their faster deployment and often small(er) scale deployment.

In the following, we discuss in more detail the many works on extending or introducing new address semantics, together with the desire to do so as an evolution to the emerging wide-spread deployment of IPv6.

3.2.3.1 New Addressing Semantics

As mentioned above, the core semantics for addressing at the (Internet) network layer were introduced, for IPv6 in RFC2460 [C3-36], later updated in RFC8200, inheriting the locator addressing model from its predecessor IPv4.

The authors in [C3-3] have provided a comprehensive analysis on issues with this original (and widely used) model of locator addressing and the many (often standardized) extensions that have been introduced over the many years since its introduction in 1998 (for IPv6) to address some of those issues. We refer the reader to [C3-3] for more details on that analysis and overview of extensions.

In summary, the work in [C3-3] identifies a clear need, evidenced through the many developed extensions, for accommodating new communication semantics beyond addressing a network location. Solutions have been developed, for instance, to address information (items) instead or separate service identification from network locators, with solutions like Information-centric networking (ICN) or Locator/ID Separation Protocol (LISP) specifying many standardized mechanisms to providing such extended addressing semantics. Other issues include improving security or privacy, thus avoiding, for instance, to leak private addressing information to intermediary parties, equally leading to many solutions to address those issues, some of which are deployed today in (parts of) the Internet. Furthermore, the desired improvement of traffic engineering capabilities has led to numerous extensions to addressing semantics, such as through segment routing capabilities, while also seeing solutions to address the efficiency issue arising from the often needed (re-)encapsulation of packets to cater to the various addressing semantics that exist along a network path.

While not explicitly discussed in [C3-3], there exist also number of challenges in appropriately mapping identifiers across semantic boundaries, not just from application to network functions but also across semantic boundaries within, e.g., routing solutions. This creates strain in resolution functions like the DNS or requires the need for introducing entirely new mapping systems, such as done for LISP or similar routing approaches, increasing the complexity of the resulting system further.

The key takeaway is that all those extensions have led to point-wise extensions, some of which do not interoperate when deployed in same network regions, increasing the brittleness of the overall system, represented in increased operational complexity.

Thus, aligned with the main recommendation in [C3-3], a wider (architectural) view is needed on developing those many addressing semantic extensions, including the need for accommodating future, emerging needs. Here, programmable data planes may pave the way for extensibility, where protocol design and deployment may move faster than today.

3.2.3.2 IPv6 Evolution

The work in [C3-3] is limited towards the many works to evolve the original IPv6 addressing semantic. While also including approaches stemming from academic research, finding its way into standardized extensions to the existing basis of the Internet is central in [C3-3].

We believe that this is a key message to the evolution of addressing semantics in general. Thus, while research may (and probably even should) start freed from constraints that stem from existing formats and solutions, most notably IPv6, it should be complemented by developing suitable evolution strategies for the deployed network infrastructure; otherwise, the attempt to boil the ocean through a green field replacement of existing network infrastructure will ultimately fail to substantiate in real-world deployments. Conversely, the suggestion of extensions must avoid feature creep, i.e., the overloading of capabilities. Instead, industrial development and deployment insights should be used to suitably prune capabilities down to those that have found widespread adoption. How to achieve such pruning, however, still remain an open discussion, also ongoing in SDOs like the IETF.

One possible architectural approach to evolving IPv6 (as the key foundation for many deployed network infrastructures) may lie in the utilization of IPv6 extension headers to realize emerging new addressing semantics in a coherent manner. This may, however, require in additional research and development efforts that close the often-cited efficiency gap between processing of IPv6 address fields and the necessary extension-specific parsing of its extension headers, an area where again programmable data planes may provide suitable solutions to achieve just that.

The key takeaway is that evolving IPv6 has long become the approach for successfully extending the original addressing semantics of the Internet, while extensions headers and the increasing knowledge of how to utilize them may provide a suitable path towards deploying the semantic-rich communication solutions that we will need in future networks.

3.2.4 Forwarding

The forwarding of packets in the network has long been focused on processing an addressing information against previously established state for the next hop to send the packet to. With the need for supporting new communication semantics (see Section 3.2.2.1), that processing will be impacted, both in achieving the needed extensibility while providing the needed complexity at high speed.

3.2.4.1 Enabling new forwarding approaches

However, forwarding data in several limited domains may face the challenge of dealing with the intermittent connectivity, as is the case of networks of satellites. One of the potential forwarding strategies aimed to tackle such challenges are based on a Delay Tolerant Networking (DTN) platform. In this context, the DTN bundle protocol allows the chaining of different TCP sessions between custodian nodes, in order to achieve end-to-end connectivity over a set of intermittently connected nodes. From the routing viewpoint, data exchange is rather challenging in such networks since paths between any pair of nodes may never exist or delay may be too long to be accepted by current data transport protocols. There are already a significant number of proposals to opportunistically route data based on time-variant graphs used by DTN, each with a different goal and based on different evaluation criteria.

Another type of forwarding mechanisms that may be suitable to limited domains, mainly the ones with dynamic behavior, are based on Information-Centric Networking (ICN), which addresses data using data identifiers and may encompass different forwarding strategies to forward packets with different data identifiers. This shifts the current host-centric forwarding mechanisms towards a data-centric approach. ICN enables a consumer to request a given data object in the network without any knowledge about the location of the requested data. The paradigm shifting from a host-centric to a data-centric approach brings several benefits to the operation of several limited domains namely the adaptation to intermittent connected networks based on a pull communication model and in-network caching, as well as extra flexibility to handle different types of traffic, based on an extended set of forwarding strategies.

We can also observe a number of new paradigms beyond address-based lookups that are being researched and brought into SDOs. One such example is the field of bitfield-based forwarding, as introduced in [C3-50] for SDN-based systems with the BIER WG[20] effort in the IETF as a standard-based alternative. Key application areas for those technologies are limited domains, as in [C3-50] for an ICN-based IP forwarding system, or shim overlays for wide-area networks, as in the case of BIER. The challenges here are implementability of those new techniques, either requiring new forwarding HW, as in the case of BIER albeit limited to overlay nodes (and tunneled via existing transport networks, such as Ethernet or MPLS), or piggybacking on existing SDN capabilities, as in [C3-50]. The question on implementability and the need for new forwarding plane processing techniques can be extended beyond the forwarding operations, which we do in the following sub-section.

3.2.4.2 Move beyond packet forwarding only

Discussion on doing more than just next hop determination but also compute over the packet content before forwarding, i.e., In-network compute (over packets), including for things larger than just packets.

In the context of an evolution towards a semantic-oriented protocol system, in-network computing may take advantage of forwarding schemes that operate based on other types of semantics than just topological ones. This means that forwarding schemes should take advantage of any semantic information carried in a packet

[20] https://datatracker.ietf.org/wg/bier/about/

header, including references to data objects and service functions capable of invoking computing services by means of their names which are directly mapping to service resources and locations. This obviously brings the important advantage of decoupling in-network computing from the host-centric model, hence making the deployment of a content-based service-oriented protocol system more straightforward. Though still belonging to the research and experimentation domain, these concepts are certainly attractive to achieve a more scalable and dynamic deployment of edge computing services also in limited domains where network changes and mobility of nodes make the application of classical edge computing implementation complicated. In such cases, a named-function forwarding scheme can simplify the service and data discovery operations, as is the case of multi-tier non-terrestrial networks, where data are distributed across multiple space assets and or could be migrated from one NTN node to a neighbor one in order to boost the data access and eventually a task computation.

Moreover, exploitation of semantic routing principles (as outlined in the previous section) would allow for a more agile and dynamic distribution of data and related access independently of the specific underlying network topology that can be highly variable. Though certainly appealing, the power of information centric networks supporting edge computing may come at non-zero cost because of the additional protocol overhead and the change of networking paradigm for which current space systems may not be fully ready for. Deploying a forwarding scheme based on data and service semantics would still require coexisting with traditional IP-based networking models and therefore call for adapted protocol interfaces or mapping between service locations and IP addresses.

In other words, knowledge from research on ICN can also pave the way for a more practical extended forwarding plane. The benefits of extending the operation of in-network forwarding nodes are known, as are their computational costs. In case of transport protocols, in-network forwarding nodes can deliver faster and more accurate congestion events compared to end-to-end approaches. In addition, in-network congestion control is the only credible solution to handle the fairness issues that dominate the Internet due to the proliferation of multipath protocols, the diversity of network types (5/6G, Satellite, Wi-Fi 6, IoT and others) and the various transport solutions (TCP, QUIC, DCCP, MPTCP, MPQUIC, MCQUIC). However, the cost of performing connection-level congestion control at each in-network forwarding node is typically not manageable, thus hindering the application of such schemes.

Previous ICN studies have verified that in-network operation can enhance the performance of data transport, offering more accurate and faster feedback to congestion control algorithms that are either running in the network routers (pure in-network operation) or are collaboratively running at the endpoints and at the (in-network) routers (hybrid operation) [C3-48]. Addressing the potential scalability concerns, numerous studies have introduced designs that reduce the computational cost of in-network operation, commonly distributing specific parts of the pool of transmission flows to different routers, thus aggregating the resources of the in-network devices. For example, to aggregate the storage resources of in-network caching routers, various cooperative distributed in-network caching schemes have been investigated in depth, yielding significant performance advantages [C3-49]. Transplanting the scalability-sensitive ICN-inspired in-network designs to IP-based forwarding nodes can unlock the limits of next-generation networking, namely, introduce enhanced in-network performance in a practical way.

The emergence of large-scale AI training, based on large language models (LLMs), has renewed the possible focus on in-network operations that may improve on training performance, e.g., through latency reduction by pushing aggregation computations into the network. Such NetReduce type of function, as part of a large collective communication capability, requires a controllable placement of in-network computational capability as well as a processing that comes close to the desired line speed of the training system. For this, signaling as well as forwarding processing capabilities will be required that likely extend the in-network functions beyond advanced store-and-forward capabilities previously researched, e.g., in ICN systems.

The key takeaway is the packet processing in the network is evolving beyond purely forwarding packets towards more complex operations, posing challenges not just on preserving desirable line speed rates but also to manage the operational consistency of those enhanced operations within control plane that will allow for a programmable signaling of where which enhanced functionality ought to be used.

3.2.5 Telemetry

Another fundamental enabler for building smart networks and services is formation about the network and its components to derive the right knowledge for decision making, not just at management but also control level. As an example, we have seen efforts to move from congestion signals like packet loss at endpoints, utilized in most transport protocols, to explicit congestion information that is being obtained from the network. Explicit Congestion Notification (ECN) has, for instance, been utilized in the development of the Low Latency, Low Loss, and Scalable Throughput (L4S)[21] protocol, providing such congestion signals through explicit bit marking during the packet's network traversal.

Other works, like the High Precision Congestion Control (HPCC) protocol [C3-66] gathers precise link load information on the return path through an explicit in-network telemetry (INT) approach, adding such telemetry data to the data packet, utilizing programmable switches and network interface cards for doing so. Although originally designed for data centers, HPCC++[22] aims at utilizing commodity hardware, making it amendable to deployments outside data centers.

While telemetry provides many possibly exciting insights into what is happening in the network for a number of applications, its deployment may also quickly overwhelm the system. The work in [C3-67] outlines that not just the traffic overhead but also the required processing of the data at the telemetry endpoints can quickly become an obstacle impossible to overcome. Direct memory access techniques, as described in [C3-67], may lower this burden and facilitate a more direct integration of telemetry data into the memory system of the participating endpoints.

Another obstacle is the need for sufficient support within in-network components but also in network cards at the endpoints. Programmable switches have opened a door here but recent developments in, e.g., P4 deployments, have shown that those programmable capabilities have not widely found deployment from a commercial perspective. FPGAs (field programmable gate arrays) have long provided a route to achieve programmability, as also initially utilized for early SDN (software-defined networking) work, with [C3-68] showcasing how to achieve telemetry with this technology.

Beyond the task of gathering telemetry information in a scalable and efficient manner, the development of suitable data models, e.g., as YANG models[23], is key to elevate the possibly vast pool of information onto the level of knowledge over which decisions can be made or to perform benchmarking within agreed and standardized methodologies and frameworks. Another key challenge is the suitable processing of the telemetry information, in particular for high density networks like data centre networks, where telemetry

[21] https://datatracker.ietf.org/doc/rfc9331/

[22] https://datatracker.ietf.org/doc/draft-miao-tsv-hpcc/

[23] https://datatracker.ietf.org/doc/rfc6020/

information can quickly overburden the processing capacity assigned for it. Work, such as in [C3-73], outlines this challenge in data centre environments, particularly for time series of metrics, utilizing methods for direct memory access into processing nodes to reduce the burden. Pre-processing and hierarchical processing approaches are other key avenues for research to provide suitable telemetry in the light of ever increasing information being gathered.

The key takeaway is that telemetry is fundamental to enable the necessary knowledge to turn networks into properly manageable parts of the overall system, e.g., through developing suitable Digital Twins, for improving the control of data plane protocols, or for the integration into traffic steering policies, among many other usages.

3.2.6 Transport

The introduction of the key transport protocols functions, namely congestion control as well as flow and error control mechanisms, has been facilitating unicast end-to-end communication for many years, not just in the Internet at large but also in transport networks for environments like the cellular subsystem. We argue in this section that the evolution of transport protocols need to evolve beyond TCP (and recent developments like QUIC) to be future proof towards 6G smart networks.

3.2.6.1 Evolving Beyond TCP & QUIC

While those functions, most prominently developed for TCP, have evolved since their initial introduction, research in recent years has seen many developments into new transport protocols, particularly for emerging use cases such as AI, large-scale web search, and datacentre networking in general, many of those new developments initiated and stemming from European activities and researchers (such as multi-path TCP [C3-11,12] and L4S [C3-15] for the Internet, or NDP [C3-16] and EQDS [C3-9] for data centre networks, to name just a few).

The new capabilities in those emerging developments need to cope with properties of new networking environments, specifically those of much higher as well as dynamic demand as well as a more dynamic underlying network topology, leading to significantly increased path density and diversity, while ensuring more controlled latency and utilizing improved capabilities of endpoints and switching nodes alike. Here, it is important to recognize that the original Internet vision of a single end-to-end congestion control mechanism may need strong re-thinking towards a segmented approach that utilizes (Limited Domain) optimized solutions on each segment while packets utilize the end-to-end communication resources.

Conversely, the development of new methods the packet transfer in transport networks may also impact the regulatory framework within which new transport networks will be deployed. For instance, the possible benefit of path diversity, with the suitable transport protocol to utilize such diversity, may impact the spectrum policy for next generation transport networks in that spectrum allocations may allow for more dynamic per link or per area assignments of frequencies in order to provide the desired path diversity.

We argue that Europe must continue to and strengthen its capabilities in new transport protocols, exploring the widened design space that is defined by methods for dividing the end-to-path into explicit proxy-controlled path segments, providing suitable interfaces for user-network interaction, while each segment may employ optimized methods for packet spraying (employing the increased path diversity of future networks), packet pacing (achieving a tighter control over latency in future transport networks), packet trimming (employing increased switch capabilities to signal congestion without packet drop), priority scheduling (employing the increased endpoint and switch capabilities to control the forwarding along defined latency bounds), and packet coding (employing the increased endpoint capabilities for creating redundancies for improved resilience against packet loss), while ensuring an alignment of the **regulatory framework** that exploits the best means of optimal resource utilizations.

But beyond looking at new methods for transport protocols, research must be extended to support the many **new communication semantics** beyond endpoint-centric unicast that is prevalent in the Internet, extending specifically to **collective communication** as a typical AI application pattern, **ad-hoc multicast** to stem the linear cost curve of headend replication for over-the-top chunk delivery applications like video, as well as connect to **information-centric concepts** of applications outside the current HTTP2/3 model (and its increasing intertwining with the concepts of QUIC that allow for a fast albeit limited connection to a centralized data centre for content provisioning).

The outcome of this research must not be limited to a thorough investigation of this widened design space but needs to ultimately lead to specific proposals to the engineering communities in suitable SDOs for new transport protocols to be deployed in future systems.

In the following subsections, we discuss the motivation for those new capabilities, focussing first on the extension from the simple unicast model of TCP in Section 3.2.6.2, while discussing the impact on endpoint components in Section 3.2.6.3.

3.2.6.2 Exploring the Transport Design Space

The starting point for key developments in the transport protocol space has been the utilization of a single, shortest path from the sender to the destination, following a byte stream semantic. Key to the interoperability of any new or evolved transport protocol is the resource management regime of TCP that is applied for flow and congestion control. The term **TCP friendliness** captures the key constraint that any (other) transport protocol must not disrupt the performance of a competing TCP flow along the same or parts of the end-to-end path. Through such strong requirement, not just **backward compatibility** is ensured but also a **fair sharing** of available bandwidth resources is being enforced.

However, we observe that the increasingly **dynamic nature** of the underlying network, including fluctuations of the capacity that are exposed to upper layers with technologies such as mmWave, and, even more so, sub-THz communication, is a terrible match for the end-to-end congestion control loop that is embedded in protocols such as TCP or QUIC. This end-to-end loop assumes a relatively static network bottleneck, and it is ignorant to the true nature of those underlying network dynamics. So far, this issue has been addressed by installing Performance Enhancing Proxies (PEPs) [C3-52, 53, 54] – devices that work as a quick fix by "cheating" TCP. However:

- Since they were never part of the design, PEPs harm the evolution of TCP, e.g. by making assumptions about how the header looks or failing to interpret new options (this has been called "protocol ossification") [C3-55].
- They are limited to functions that can be attained in such a "cheating" manner.

The fact that such devices cause complications has contributed to the decision to let QUIC encrypt everything, including the protocol header. This allows QUIC to implement truly "unharmed" end-to-end congestion control; as a result, it attains much *worse* performance than TCP with PEP support in some scenarios, e.g. over satellite links [C3-56, 57, 58].

QUIC now forces us to **design proxies right**, instead of trying to retro-fit them into a given network ecosystem. With QUIC, such proxies must be known and authenticated; communication with them is explicit, and it allows both ends to agree on the types of services that a proxy will implement. This can enable such proxies to:

- Customize congestion control for a Limited Domain, while correctly interacting with the end-to-end connection,
- Apply techniques such as packet spraying over multiple paths *within* the network, where much greater path diversity is available,
- Dynamically forward traffic *within the network* across multiple paths based on the downstream load, i.e. route around congestion,
- Carry out other functions such as caching *when desired*, i.e. implement suitable functions *depending on the type of traffic* (which traditional TCP PEPs, by design, have to be oblivious to).

Suggestions on how QUIC endpoints could theoretically interact with proxies have already been put forward [C3-58, 59, 60]; this line of work presents an opportunity to design proxies right, not only for QUIC but also other protocols, and it is perhaps the only way forward if we want to enable transport protocols to fully exploit the capacity of a highly dynamic underlying network infrastructure, leading to lower latency and better network utilization altogether.

The consequence of such (explicit) design of a proxy-based system is that a longer end-to-end path is being divided into network segments, such that a separation into Limited Domains is attained for those segments. For one Limited Domain, the ideal choice of the transport protocol and configuration may be very different than for another Limited Domain. More, we observe that along the possibly segmented, end-to-end path, a possibly rich design space exists for new transport protocols, optimally utilizing the resources provided by the Limited Domain. Here, we divide efforts along two orthogonal directions, visualized in Figure 3-2, namely that of control primitives on the one hand and (supported) communication semantics on the other, with the shading showing the focus of efforts in past research and development efforts.

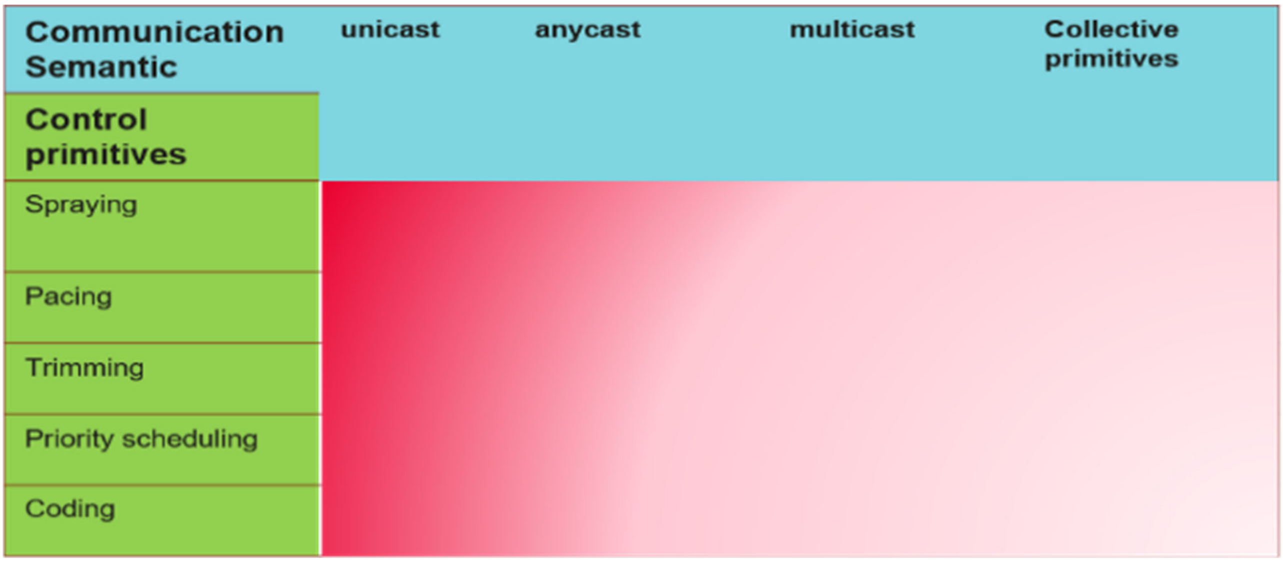


Figure 3-2: Design Space for Transport Protocols in Segments of the E2E Communication

Widening the supported communication semantics is primarily driven by the emerging demands of limited domains. Collective primitives, for instance, are key for DCN as well as 6G-based AI applications, while multicast, particularly ad-hoc multicast relations, are crucial for stemming the exponential growth of video delivery traffic in access networks, whereas anycast plays an increasingly important role when considering compute and network resource interplay, as discussed also in Chapter 2.

Utilizing richer control primitives, compared to the original TCP design for a single path, unicast mechanism, is largely motivated by two factors, namely (i) the better availability of network mechanisms to support those extended primitives and (ii) the improving processing capabilities of endpoints and network elements alike.

The first aspect is particularly crucial for developments of new protocols for limited domains, such as data centres, where the assumption of more homogenous and improved switch capabilities may hold through new(er) rollouts of underlying infrastructure elements. From those control primitives in Figure 3-2, packet spraying has seen most entrance into available solutions, including for the wider Internet, most notably in the form of multi-path TCP [C3-11,12]. The main objective here is the greater utilization of path diversity in networks, where similar 'optimal' paths may exist but also where multiple paths (of differing characteristics) may be used for resilience purposes. The former is particularly true for data centre networks, but also often 5G (and future 6G) networks, where rich and path-diverse underlying transport networks may be utilized more optimally by not limiting E2E traffic to one (initial) path choice. Solutions to spray packets may use existing methods for equal cost multi-path[24] (ECMP) routing or emerging methods for in-network scheduling at runtime, such as those discussed in [C3-5, 24], both of which employ the improved processing capabilities of in-network elements.

The packet pacing primitive, embodied in TCP through its window-driven congestion control mechanism, has seen revisits in recent years through work such as in [C3-9, 16] among others. Such revisit is particularly motivated by the applicability of, for instance, credit-based mechanisms in domains, where a tight control over path length and latency is possible, such as in data centres but also in transport networks for cellular systems. Complemented through methods for network virtualization, those newer resource management regimes may be isolated from legacy TCP streams to achieve the desired TCP friendliness, while optimally running those new mechanisms in a virtualized, network partition.

Packet trimming, proposed in [C3-16], is an example of introducing new network capabilities for the benefit of improving E2E performance. Here, packets are trimmed down to their (packet) headers to allow for congestion handling, while reducing the pressure of the network through the removal of the often much larger payload. Thus, congestion control may continue without the overall packet loss experienced in congestion-sensing schemes like TCP, while retransmission of the trimmed payload will ensure the reliability of transmission, here trading off stability of the network against the possible prolonged but often more constrained competition latency of the transport stream. While originally targeting data centres, due to the requirement for packet trimming support in intermediary switches, solutions may also be employed in other environments, most notably 6G transport networks, where similar switch requirements may be enforced in future infrastructure rollouts. Research is needed, however, in the cost-benefit of such solutions, particularly for supporting high throughput applications.

The same applies to utilizing priority scheduling as another control primitive, which has long been the subject for study to improve utilization and latency of end-to-end streams. With the emergence of Layer2 mechanisms like TSN (time-sensitive networking), the utilization at transport level may see more widespread use in novel solutions.

The use of packet coding was extensively studied in past multicast transport work but has seen a renaissance in more recent work, not just in combination with information-centric networking [C3-39], but also in works like [C3-38] for storage and data centre networks, here combined with the primitives of packet spraying,

[24] https://en.wikipedia.org/wiki/Equal-cost_multi-path_routing

pacing, and trimming for further performance improvements. Key here are the expected improvements of communication endpoints to execute the additional computation needed for packet coding, while reaping the benefits of spraying the coded information for additional information diversity to provide resilience against network-side dropping of packets, thus positively impacting the flow completion time.

A key challenge to move from the uniform byte stream model of TCP to a widened availability of control mechanisms, many of which may be very specific to certain applications, is the problem of exposing application requirements as well as constraints of endpoints and network elements alike to make the right choice for the set of mechanisms to be utilized, and thus ultimately for the right protocol to work optimally. The recent IETF update of the transport layer's API, Transport Services (TAPS), offers the necessary flexibility in the presence of path segments within Limited Domains [C3-61, 62], enabling the dynamic choice of a suitable protocol because it breaks the compile-time binding between applications and i) the underlying transport protocol, ii) the protocol configuration and iii) the network path. For an application running in Limited Domain A, a completely customized protocol for A could be chosen by a Transport Services system, and it could be translated into a Domain B specific protocol or a general-purpose protocol such as TCP or QUIC at the boundary between domains A and B, allowing to map application requirements onto customized protocols that can make ideal use of Limited Domains [C3-63, 64, 65].

The key takeaway here is recognizing and utilizing the richer design space for new transport protocols compared to the current TCP model, where this richer space is driven largely by emerging demands and capabilities of limited domains, most notably DCNs and 6G. More research is needed to examine the set of optimal solutions for categories of key emerging applications, comprehensively investigating and utilizing this richer design space.

3.2.7 Service-& Network-level Inter-Provider Interoperation

The sections above introduce fundamental enabling technology areas and emerging or evolving trends that represents important potential advancements for commercial and operational networks. Thus, these potential advancements will be analyzed and progressed also from the perspectives of inter-provider interoperation, and will point to future opportunities as well as research challenges.

This section complements the above by considering actor domain services and networks, their interconnection, and their interoperation that can support commercial or other contexts of service offering and service use across business or actor (or actor role) relationships. In this context, we address potential service concept evolution that can be enabled by the above technology evolution, and how new service concepts and enablers must be supported by advancements of interoperation along both the horizontal as well as the vertical dimension.

First, we consider the perspectives of the public network and telecommunication service provider (CSP), and highlight broad customer categories such as Consumer, Business, Industry, and Mission Critical customer segments. In today's telecommunication service provider business there are tree very fundamental service domains and ways of interconnection; i) the telecommunication voice, data, and messaging, including roaming (see e.g. GSMA IPX [C3-69]), ii) public Internet best-effort connectivity and traffic exchange, by so-called IP Peering and IP Transit, and iii) service provider provisioned VPN for enterprise customers, largely based on RFC4364 [C3-70]. Their underlying fundamental technical architectures and their business models have shown to be very difficult to evolve.

Building from 5G network slicing, and the support for enterprise or domain specific data networks (see Data Network Name, DNN, in 3GPP [C3-25]) the notion of Logical Network as a Service (LNaaS) serves as an important context for evolving complementary service features and enablers, leveraging from the above technology advancements. Putting extra attention to "beyond connectivity" services and service features and enablers, such security services (considering zero trust), caching, and audio and video processing (supporting e.g. AR, VR, XR) as a service, and enabled by edge computing, will be increasingly important for the telecoms industry. In the more advanced settings, also embedding such functions through capabilities enabled by in-network computing should be considered for the longer term, while for medium term, various scenarios for services adjacent to 5G / 6G should be considered as ways of providing "native" beyond connectivity services.

3.2.7.1 Moving Beyond Best Effort between Internet Endpoints

The resource pooling vision of the Internet has seen a crucial expansion towards an increasingly integrated **compute/communication resource pooling**, where services are fluidly provided in highly dynamic **service clusters** by virtue of virtualization and micro-service technologies (more on this in Chapter 5 on software technologies), utilizing resources both for networking and compute purposes.

Key to the functioning of the joint resource pooling across a clearly delineated service, compute and connectivity pool, realized in a separate stratum each, is an interjoined control plane that dynamically adjusts the relations between suitable resources towards a desired function or use case. Utilising the insights from a dedicated knowledge plane, which provides cross-stratum information through telemetry capabilities, such control functions ensure the workings across service and communication providers, adhering to business boundaries, suitably securing information exchanges, exposing capability information and more.

Figure 3-4 outlines this resource pooling architecture, elaborated in more detail in Chapter 2 on the architectures for future 6G systems.

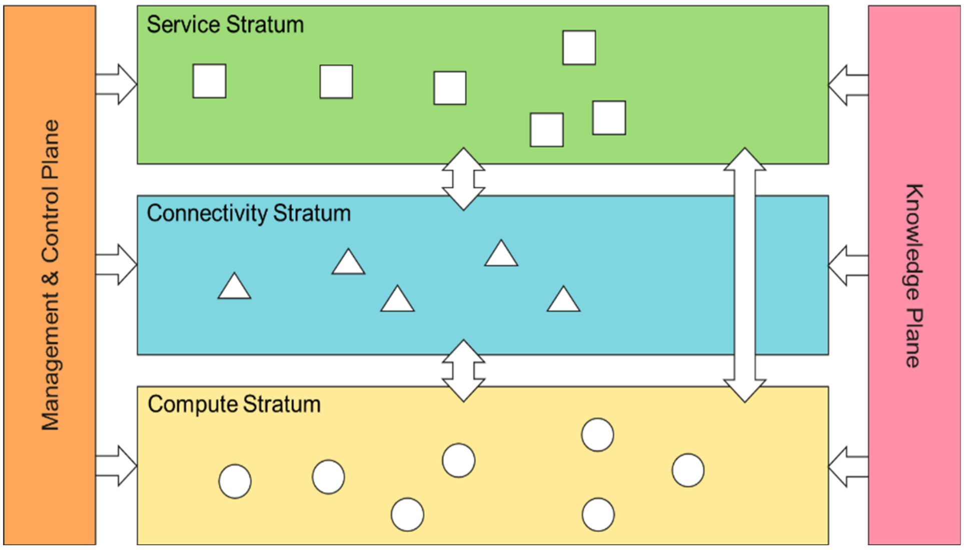


Figure 3-4: Resource Pooling Architecture with Stratum-Spanning Control & Knowledge Plane

Another key trend in service deployment is the establishment of so-called **service clusters**, particular in the form of cloud-native 5G service, realized through the **telecom cloud**. Those service clusters scale both vertically (i.e., adjusting local resource capabilities, e.g., by assigning more resources to a specific service task) and horizontally (i.e., allocating more resources within a cluster in the same or different parts of the network). Also, clusters may interconnect, thus assign workloads and tasks across several clusters for task and load sharing purposes.

This service-level interoperation is mapped upon a system of **networks**, provided by communication service providers (CSPs), either in the form of single or interconnected networks, allowing for tasks to be moved across network locations, e.g., for the purpose of reallocating tasks physically closer to end users. Figure 3-5 illustrates this inter-provider relation that exists at both the service- as well as network level.

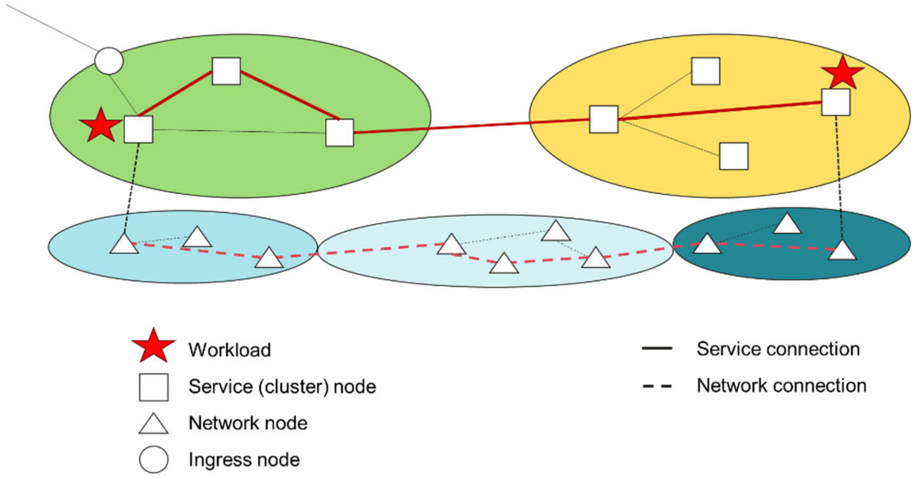


Figure 3-5: Service- and network-level inter-provider interoperation

3.2.7.2 Capabilities & Challenges

To realize this interoperation, a number of key capabilities are required:

1. **Exposure of capabilities**: both, service and network provider need to expose their capabilities both vertically as well as horizontally. For instance, service providers need to expose constraints, such as resource capabilities, to service developers in order to allow for dimensioning service properly. Dynamic service information, such as load information or availability needs exposure to realize service scheduling and load balancing capabilities. Network providers need to expose capabilities such as bandwidth and delay information to allow for guarantees in network connections and thus offer more than just best effort connectivity.
2. **Discovery of capabilities**: As the flipside to item 1, capabilities need discovery, again both horizontally and vertically.
3. **Authentication and Accountability**: With item 2, information may cross business boundaries, e.g., from CSP (communication service provider) to SP (service provider), thus necessitating authentication and accountability interfaces.
4. **Resource chaining**: once suitable resources are discovered, their usage needs to be chained, either as a service execution chaining or network chaining capability.
5. **Information transfer**: along the chain of resource, information flows need to be established, e.g., in the form of chained service invocations or the routing of packets.

For the realization of the above capabilities, a plethora of solutions exist or have emerged, such as [K8S, SRv6, SFC, discovery solution, exposure solutions]. However, the nature of future (6G) systems create additional challenges. Specifically:

a. **Higher dynamicity**: Existing service provisioning platforms like Kubernetes provide flexible vertical and horizontal scaling, although the frequency of change is still limited to the many seconds area. With advances in microservice and virtualization technologies, we can expect a higher fluidity of scaling changes, necessitating significantly faster exposure and discovery functionality. Here, purely increasing the information refresh rate does not suffice due to the increasing signaling load in disseminating information. Aggregation or proactive technique need investigation in order to keep signaling cost in balance, both at service and network level.

b. Permissionless vs permissioned nature: Existing inter-provider relations are assumed to be of permissioned nature, often based on an explicit contractual relation between, e.g., two CSPs for network-level inter-connection. However, 6G scenarios also foresee relations to be built more ad-hoc and at runtime, thus lending itself to a permissionless realization of an authentication and accountability system to reduce the burden of participation on those ad-hoc service and resource providers. This, in turn, may require changes to the capability exposure and discovery in order to allow for a constraint-based integration of the resources into the inter-operation at service and network level.

c. Richer and explicit policies: much of today's interoperation, particularly at the network level, underlies simple policies, most notably a best effort provisioning policy. Emerging 5G but more so 6G scenarios, however, require rich policies for the selection, chaining, and utilization of service- and network-level resources to implement, e.g., guaranteed packet delivery, geographical routing, compute- or energy-aware service load balancing and much more. This will require suitable policy and data models for defining those policies as well as suitable engines to execute them, e.g., in cluster ingress points, path selection and steering elements and other service- and network-level components.

3.3 Main Research Challenges

The research challenges for the Network Control and Policy Stratum are focussed on the main enabling technologies, outlined in the previous Section 3.2 and visualised in Figure 2-2.

All challenges aim at delivering a continued stream of novel methodologies, interfaces, protocols, and reproducible artefacts that can ultimately be transferred into commercial deployments through finding entry into suitable standards. Across all those research challenges, we recommend key actions, discussed in the next sub-section, needed to drive towards those commercial deployment.

Research Theme: Protocol Engineering for Networked Systems				
#	**Research Challenges**	**Time line**	**Key Outcomes**	**Contributions/Values**
1	Evolved future **autonomous control plane** technologies for dynamic, collaborative, protectives, and multi-domain formation of provider networks	Mid-term	Protocols, methodologies, and prototypes as well as suitable standards and guidance with key network technology breakthroughs Proposed solutions should: - Study the necessary **key primitives** for an autonomous multi-domain control plane while allowing for provision for collaborative approaches - Define suitable **protocols** to realize and manage primitives for scales of many (tens of) thousands of resource nodes	- **Sustainability** through reducing operational costs for resource management - Improved **innovation capacity** through accommodating new, particularly highly dynamic use cases - **Trustworthiness and resilence** through autonomous and programmable configuration, reducing possibility for human errors
2	Development of **multi-constraints routing protocols** that offer support for multicast, anycast and collective communication semantics, considering network, service, and compute requirements	Mid/long-term	**Protocols** & **methodologies** as well as suitable **standards** with key network technology breakthroughs Proposed solutions should: - Outline **algorithms** for multi-constraint routing	- **Sustainability** through improved network & compute resource utilization - **Sustainability** through evolving existing routing solutions

			- Define suitable **signalling methods** for providing needed network and service/compute constraints - Define efficient **distributed or centralized routing frameworks** and protocols that may evolve existing deployed solutions - Study suitable architectural and deployment models and technologies to reduce time-to-deployment of novel solutions	- Improved **innovation capacity** for new services and through reduced times to deployment - Better **governance,** e.g., through separating underlay infrastructure from overlay routing provider - **Improved Trustworthiness** through alignment of routing to application requirements
3	Future evolved **addressing framework** to accommodate existing evolved and emerging addressing semantics, while utilizing the existing capabilities of IPv6	Mid-term	**Best practises** and implementation insights and **prototypes**, leading to sustainable evolution of addressing semantics based on emerging needs of Limited Domains Proposed solutions should: - Outline **framework** to utilize IPv6 extension headers (EH) to encode and efficiently process evolved addressing semantics - Provide implementation insights and **prototypes** to improve on EH processing near to current native IPv6 performance	- **Sustainability** through utilizing existing network layer capabilities for future addressing use cases - Improved **innovation capacity** through efficiently implementing novel addressing semantics for future use cases - **Trustworthiness** through moving from solution-specific solution to unified framework, reducing brittleness of overall deployed system
4	Evolved future **data plane technologies** to accommodate emerging communication semantics	Mid-term	**Protocols, methodologies, and prototypes** as well as suitable **standards** with key network technology breakthroughs Proposed solutions should: - Study **requirements** of emerging use cases and communication semantics as to the foundational set of in-network processing primitives - Develop **node architecture** that allows for programmable dataplane operations with in-network capabilities - Develop suitable **deployment approaches** for distributing switchlets in a trusted manner across multiple domains	- **Sustainability** through utilizing existing dataplane capabilities for future use cases, avoiding frequent infrastructure renewal - Improved **innovation capacity** through reducing speed of deployment for new use cases - **Trustworthiness** through industry-wide agreed dataplane execution platform and approach for programmable dataplane execution
5	Future evolved **transport protocols** that offer mechanisms for congestion & flow control, including multi-path and collective comms, that constantly evolve with underlying networks and user workloads	Mid/long-term	**Protocols** & **methodologies** as well as suitable **standards** with key endpoint and proxy-based network technology breakthroughs Proposed solutions should: - **Outline methodologies** for efficient and fair data transport. - Define **suitable user-network interfaces** to allow for service-specific interaction with the network - Develop novel congestion control paradigms	- **Sustainability** through improved network resource utilization - Improved **innovation capacity** for new services - **Trustworthiness** through suitable user-network interface, minimizing or avoiding service requirement mismatch

			- Employ **coding and network telemetry** in combination with hardware offloading, e.g., to smartNICs and/or switches	
6	Evolved Interconnection and interoperation framework & protocols beyond current Internet best effort model	Mid-term	**Architectures**, **primitives** and **protocols** for future multi-domain interconnection Proposed solutions should: - Outline **key primitives**, **roles**, and **interfaces** to enable key exchanges in multi-stakeholder deployment models - Define suitable **protocols** to realize concepts for scales of many hundreds of participating limited domains	- Improved **innovation capacity** through accommodating new, particularly use cases beyond best effort - **Trustworthiness** into economic markets through basis on economic models
7	**Methodologies, infrastructures,** and **toolsets** to develop, train, test, and evaluate network protocols.	Mid-term	**Experimentation frameworks** and **benchmarking methodologies** for novel network protocols, particularly for protocols operating in difficult to access network deployments (e.g., LEO satellite networks) Proposed solutions should - Define suitable methodologies and protocols for extending the **telemetry capabilities** of the networked system, providing the suitable data set for benchmarking as well as improved operations - Define **benchmarking tests** suitable for comparison of performance of key technologies - Provide novel simulation- & emulation-based systems - Enable suitable **experimental capabilities** to benchmark solutions under near-life conditions in a reproducible manner - Define **relevant APIs** for accessing training and benchmarking functionality	- **Facilitation of fair markets** through providing a comparison framework for capital and operational expenditures - **Improved innovation** capability through quantifiable improvements of new technologies and solutions - **Trustworthiness** into product specifications and feature claims
8	Study **applicability of IPv6** in existing and emerging vertical markets	Short-term	Suitable vertical industry insights to distil possible requirements for evolution (of IPv6) or identify roadblocks for adoption Insights should: - Study the **potential roadblocks for adoption** of IPv6 and related technologies in key strategic and emerging vertical markets - Identify the possible **regulatory and market actions** to be taken to remove roadblocks - Identify possible points for **technology evolution** to improve on adoption in identified vertical markets - Recommend suitable **actions** to SDOs and regulators	- **Sustainability** through improving on common technology basis across increasing number of vertical markets - Improved **innovation capacity** through improving adoption of IPv6 and thus drive digitization of new markets & verticals

3.4 Recommendation for Actions

To drive the commercial adoption of developed solutions, stemming from addressing the research challenges in the previous section, we recommend the following actions:

- Connect to key driving verticals: Following our vision expressed in Section 4.1, limited domains are the drivers of innovation for the future NCSP. Thus, it is imperative to bring the relevant domain knowledge from those vertical markets seeking to deploy (their) limited domains into the development of the (horizontal) protocol and system technologies. This will require the establishment of suitable forums and means for exchanging with those, often diverse, communities.
- Reproducible artefacts: Prototyping key technologies has is key to pushing forward not just evaluation but also adoption of technologies. But we must go further by not producing those artecfacts but establishing an ethos of ***reproducibility*** that is aligned with the ACM/IEEE efforts in this space to avoid claims being made that cannot suitably and independently be verified.
- Experimentation: Although strong theoretical foundation is desired for any evolution of future protocols, strong experimental evidence and large-scale open testbeds are crucial to show feasibility but also foster adoption through the operational community. For this, open experimentation facilities in the form of a Digital Infrastructure are required for a large number of third-party experimenters of promising solutions. Key to this is the rooting of experimentation into clear ***benchmarks*** and based on the ***reproducibility*** of results, starting with the aforementioned creation of reproducible artefacts that provide the input into the experimental efforts. This is particularly relevant for the innovations along the multicast/unicast separation, which should be extensively explored in large scale testbeds.
- Internationalized efforts: Given the challenge to continuously evolve the underlying methods and resulting protocols for the NCPS, 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 relevant NCSP technologies as well as through the creation of international expert groups, e.g., in coordination and support actions.
- Link into relevant SDOs: Research has always provided strong input into the relevant SDOs, like IETF, ETSI, …, and that input must continue, or even increase, not being limited to direction contributions, e.g., through solution drafts, but also in initiating new initiatives, e.g., through organizing sidemeetings to build a growing community of interest that will ultimately provide the path into more direct adoption efforts.
- Open Source Funding: as low barrier to entry is a key enabler of rich innovation and a substantial part of Europe's Digital Sovereignty strategy, a) research funding should ideally result in Open Source reference implementations, but also b) specifically target such Open Source efforts that are not necessarily affiliated with commercial entities such as through the European Commission's Next Generation Internet Initiative, exemplifying an exceptionally effective tool for fostering bottom-up innovation.

Research Theme: Protocol Engineering for Networked Systems								
Action	Routing	Addressing	Data Plane	Control Plane	Inter-connect	Transport	Infras & Tools	Suitability of IPv6
Connect to **verticals**	Needed for broad industry and domain insights							
Develop reproducible **prototypes**	Needed for technology development and adoption in standards							
Build & utilize **testbeds**	Encouraged for reproducibility and adoption of solutions as well as benchmarking against well-defined criteria						Compare solutions	

International research	Encouraged due to international nature of industrial and academic research in this space		
Drive SDO **adoption**	Needed for commercial adoption of solutions in telecom and vertical industries alike	Compare solutions	

4 Network and System Security

Editor: Emmanuel Dotaro

4.1 Introduction

The Universal Declaration of Human Rights [C4-1], Art.3 states that "Everyone has the right to life, liberty and the security of person". By many aspects 6G Systems and Services cross security matters. This is actually the case, at least in European Fundamental Rights, for natural persons with respect to their personal data as stated in GDPR [C4-2]. Far beyond, 6G, following 5G enlarged scope, stands as a foundation of Digital Transformations involving natural, legal and up to national security issues. Whereas 6G is expected to be deployed in essential and critical sectors (private or public) of the society, Holistic security must be provided to mitigate inherent risks and ever growing number of Cyber-Attacks [C4-3].

4.2 Vision

As a general principle, Systems & Services evolutions (Architectures, Technologies, Operations, Usages) mandates concomitant evolution of the Cybersecurity. 6G aims at being the enabler of an unprecedented number of use cases encompassing large diversity of architectures (Cellular, Cell-less, Edge, IoT, 3D, mesh, Adhoc, Digital Twins, ...) with massive usage of AI (increasing the Attack Surface) and new technologies. If one consider the diversity of expectations up to Mission Critical but also Human-Centric, the Cybersecurity set of challenges can not be limited to classical hardening of some components or even obsolete perimetric and static approaches.

Last but not least, 6G is crossing multiples digital-related fields such as AI, Data, Micro electronic, Cloud, HPC, Quantum, Sustainability,...from security point of view, it results in a diversity of necessary regulations, knowledge sharing and transverse actions.

Among those actions it is worth to mention existing work achieved in regulation [C4-5][C4-6][C4-7], outcomes from the ENISA developed in a set of reference documents:

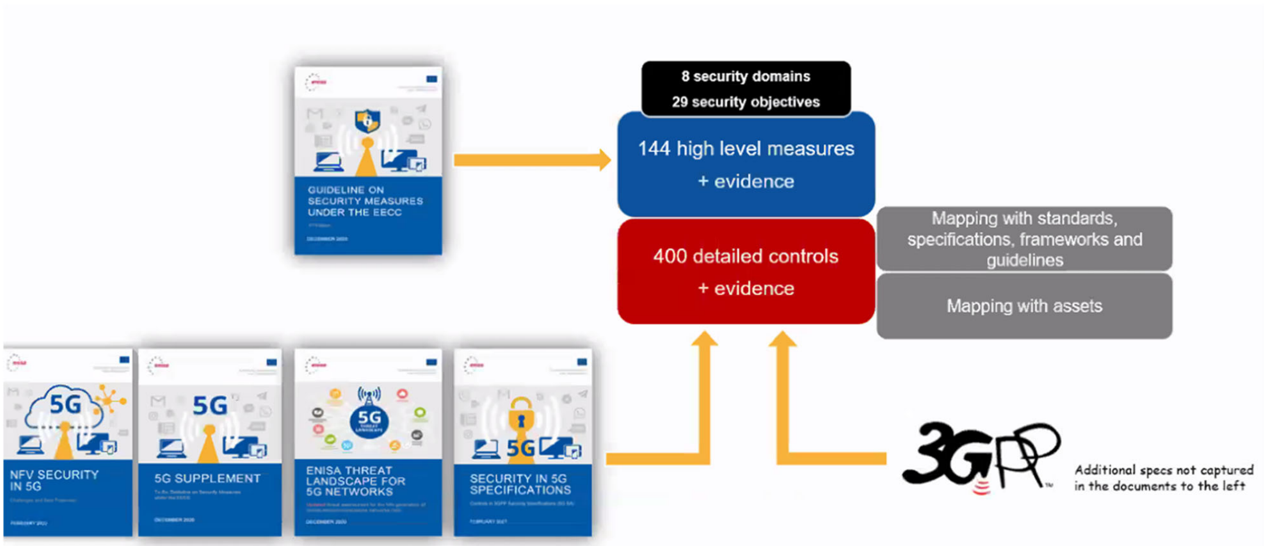


Figure 4- 1 – ENISA overall reference document set

The State of the Art is ever growing on security and specific of security applied to 6G. Useful references should be found (updated on a regular basis) from associations, among others in [C4-8][C4-9][C4-10], as well as a recurrent topic in literature [C4-11][C4-12][C4-13]

- Intrinsic nature of Systems & Services: Holistic & Metamorphic
  - Basically inherited from the Middle-Age, the previous generations of cybersecurity suffer from more and more decoupling with the Systems & services evolutions. While security only makes sense from an End-to-End point of view, the challenge here is to deal with over-complex, highly distributed architectures. Multiplying fragmented multi-lateral, multi-layer, multi-party, (micro-)services "Black Boxes" perimeters with numerous interfaces and exchanges requires to re-think the Holistic Distribution of Security in all its phases (protection, detection, response).
  - Besides spatial distributions challenges, the transient nature of 6G Systems & Services configuration combinations is raising in turn a large set of security challenges. This requires holistic but also adaptive security fitting metamorphic properties of the systems. Among others, yet unsolved, challenges encompass, continuous/predictive assessment of security conditions, incremental certification, (Hybrid)AI-controlled time line of operations,...
  - Moving from Cloud-Native towards AI-Native 6G is opening a specific area for security. On the one hand the whole AI life cycle has to be secured. In particular, a direct consequence is to reinforce data centric security fueling 6G AI, together with xAI and assurance for models and behaviors predictability. On the other hand security will make full use of the AI power enabling enhanced 6G Cyber Threat Intelligence (i.e. OSINT for 6G), smart protection (from physical layer to services), Smart distributed and collaborative Attack Detection, Smart remediation and Response.
  - Security is horizontal and applicable to all Systems & Services parties. But one can wonder what is the global security level resulting from such heterogenous combination, made of more or less opaque (Black Boxes) segments. With well-known benefits (scalability, up-to-date features, consistency,...) of the XaaS paradigm, 6G should integrate Security-as-a-Service (SecaaS) provided by dedicated security vertical pure players.
- Intrinsic needs diversity
  - 6G is aiming at covering such a diversity of usages that it should be obvious that a "one size fit all" security model should either be too (dangerous, de facto trust in providers) basic or too (costly) high grade. It comes to the paradigm of Differentiated Security (DiffSec) Mimicking the multiple QoS-based attributes defining a Digital Service, Security Service Level Attribute (SSLA) should be one of them, based on Quality of Security (QoSec) criteria.
  - Driving research for 6G security foster anticipation of threat/risk analysis and understanding/capture the actual needs consequence of usages/business and the state of the art in terms of attack surface, known, projected vulnerabilities, critical assets… Proper positioning about this balance should motivate priorities.
  - Once 6G will be capable to deliver DiffSec, users (Human or Machines) will be empowered to have a smart usage of Digital Services. Some challenges attached to this simple requirement consist in awareness of Services attributes through adhoc exposure, . request through security policies (ranging from Natural Language, Contextual specific syntax to Intent-based/semantic elicitation) and actual mapping into concrete combined provisioning some vertical applications may require formal proof).

Directions

The multiple 6G security challenges introduced above should be addressed through the following 7 areas :

- Architectures & Strategies: Both Protection and Detection End-to-End Security Distribution (beyond perimetric), encompassing the evolving diversity of 6G architectures (Edge, 3D, Mesh,..) and consequences of "Zero Trust Architecture" developments.. Integration of Security Services including Security as a Service. Differentiated Security architectures. Cooperative Holistic Security across domains, layers, stakeholders. As unavoidable weaknesses will remain, in particular at termination points, Root of Trust distribution and Backboxes Tolerant architectures.
- Data Centric: Data is key for privacy issues but also in control and management as fuel of AI-Native 6G. Beyond (lightweight) Post Quantum Encryption, processing of Data in 6G should be driven by dedicated approaches such as Sticky Policies (Data policy self-support), metadata, binding with impact on hardware processing, Confidential Computing.
- Hardware and Physical layer: 6G is bringing new hardware in the picture such as Intelligent Surfaces. More stringent requirements on Clocks for Time Sensitve Networks. Although 3D or Cell-less architectures changing the attack surface. From Trusted Execution Element on terminals, hardening of newly introduced 6G components (incl. side channel attacks), to jamming and eavesdropping attacks 6G security research landscape must embed systematically security considerations for Components and Physical layer. This should also cover the supply chain and any operations before entering in production.
- Software & Virtualization: Identified in 5G threat landscape as a main source of vulnerabilities Software —from Safe Source Code to the entire life cycle (incl. static/dynamic code analysis, OTA updates, Access Management & privileges) remain a strategic concern. Despite commonalities with the IT domains, 6G being more and more software predominant and disaggregated the question must be addressed with the complexity and authority fragmentation inherent to 6G. Virtualization tools and operations (OS, hyperware, APIs, slice controllers/orchestrators) are key to 6G and should participate to the Holistic approach with secured distributed interactions.
- AI-based Operational Security: the overall goal is the application of the Zero Touch paradigm to security. It results on multiple research direction for smart deployment of Security in such complex 6G architectures. Protection is already massively complex solving policies elicitation and combination. But Detection and response is even more challenging. 6G AI-Native should not only take benefit of atomic AI-based function (xDR) but being able to integrate it in a holistic way with all subsequent interactions between suppliers, providers, users. 6G AI-based security should also encompass Lawful Interception issues as well as Root cause analysis and identification.
- Quantification, Evaluation: There won't be trustworthiness if the Quantification and Evaluation are not there. As introduced above there is a need to define QoSec, provide approaches to evaluate it and maintain this information available from service request to the decommissioning. Thus continuous assessment is a challenging objectives mixing somehow certification complexity with E2E perimeters and dynamicity. One should note that forensic, liabilities and major societal impacts depends on the future capability to evaluate the security quality (requires models, data lakes, potentially Digital Twins, friendly Hacking) and expose it to users,
- Governance: security is based on Standards, Open Source Communities, all of it under multiple regulations. From education (Research Platforms as Cyber Range) to CTI sharing. Research actions should at least integrate State of the Art from (5G) Tool Boxes, ENISA (NIST/NCCoE…) recommendations and further contribute to build a safe and secure ecosystem and make undoubtedly 6G acceptable from societal, industrial, strategic point of views.

Figure 4- 2 – Sources of information relevant to 5G/6G (ENISA)

4.3 6G Security Architectures

Developing security architectures for providing E2E security assurance across the heterogeneity and dynamicity of technologies and architectures envisaged in 6G is a major challenge. The meaning of security from user perspective can't be anything else than End-to-End, hence leading to holistic distribution of security in all its aspects.

The solutions should be able to handle a diversity of 6G scenarios (cell, cell-less, IoT, 3D, private, public environments, integrated Digital Twins) as well Cyber Physical (CPS) scenarios integrating sensing as termination points of the systems.

Beside the complexity of the architectures, the ownership, the control & authority scopes are intrinsically fragmented with increasing interworking requirements and consequences on features distributions & liabilities sharing.

By its multiplicity of architectures patterns, flexible configurations and usage targets 6G is not addressing a single security level. IT makes no sense to provide 6G systems delivering high grade security when not required (for obvious economic reasons) as it makes no sense to underestimate security requirements. Recognizing that one-size-fit-all drives immediately architecture approaches where the resources and features are adequately provided as per the actual needs (time and space). We will talk in the following of Differentiated Security (DiffSec) by analogy with 20th century concept of packet technologies.

Security and trust are close companions, considering evolutions from Cloud-native to AI-native 6G, the question is raised with renewed intensity. Who can we trust an AI-driven system without xAI and security applied in AI deployment? From data collection for statistical AI to Control Loops and distributed smart multi-agent collaboration, securing both AI for 6G and 6G for AI is a pre-requisite of 6G advent. The societal perception of multi-Agents may be negative understanding the intrusive power into the privacy area. Trust in AI would need to be built on Human-Centric usage of the technologies remaining under well-defined boundaries and empowerment of the users.

Open access to digital services should be enabled by awareness of the security conditions associated with those services. Either for Business-to-Customers or Business-to-Business there is a fundamental need to provide integrated architectures qualified with security attributes.

4.3.1 Security Distributions in 6G Architectures

This aspect should address time and spatial distribution of protection, detection and response security capabilities across realistic horizontally and vertically fragmented architectures (multi-layer, multi-provider). This should notably encompass protocols and interfaces for E2E adaptive security delivery (inter-orchestrator, agent-based distributed convergence) ensuring multi-tenancy (e.g., verticals) remediation strategies with regard to business objectives (although vertical specifics).

6G won't exist in isolation; it will converge with existing technologies like cloud computing and the Internet of Things (IoT). Research must focus on developing a holistic security framework that seamlessly integrates security solutions across these the various domains of the architecture (from far edge, through continuum up to applications & services) Architecture patterns considered should be representative of 6G systems and technologies.

As such the diversity of use cases will address Cellular, cell-less, Edge, IoT, 3D (NTN), Public, Hybrid, Private, Mesh, fully distributed D2D/V2V, Adhoc, Vertical specific architectures and Regulations, Distributed Ledger Technologies, Quantum-based architectures

Cooperative holistic approaches involving multi-layers/stakeholders authorities (including compute/ network/ security service providers) will be key together with smart distribution of root of trust AI capabilities to be used in AI-based Operational Security (SecOps)

Security distribution is both spatial and time and the following research challenges cover the two dimensions:

- End-to-End, Multi-lateral (stakeholders), Multi-layer Security functions distribution
- End-to-End Security Policies Decision and Enforcement distribution
- Hardware and Software Root of Trust distribution and adaptation along life cycle
- Digital Twins integration with massive remote management, sensing, monitoring
- Research should focus on a security framework enabling seamless integration across domains (segments from Far Edge to Data Centers, layers up to services/applications, policy, authority) and parties.

4.3.2 Differentiated 6G Security

There are no uniform requirements in security level. Three main axis are potential candidates to constitute the basis of resources and features segregation. The first one is driven by the usages and is based on business objectives. For instance, a business depending on time precision would requires assurance to the availability and precision of time and phase. Another one would like to protect its confidentiality even forbidding activity detection or analysis. A third one may have critical needs to escape from extra-territoriality of foreign countries. As illustrated above, the security differentiation may go far beyond the traditional Confidentiality, Integrity and Availability criteria. Sticking to industrial reality and geopolitics, both sectorial and regional regulation may apply and participate to the adequate 6G systems and Services provisioning.

Once some security differentiation exists and resources being limited, the question of the applicability of priorities and precedence will raise automatically. The limitation will obviously depend on the type of 6G architecture concerned i.e. Drone Swarms (adhoc network) don’t deliver same bandwidth and service availability as fixed cellular network.

- Multi-level Security and architecture profiles, plan-based (user, control, management, service) segregation, resource profile and isolation
- Priority & Precedence policies and mechanisms for security objectives
- Resilient (up to critical) Infrastructures and Services
- Secure-by-design communications for deterministic performances

4.3.3 Secure Artificial Intelligence (statistical, hybrid) for 6G

Before contributing to the security of 6G for AI or AI-based 6G security (in other sections) we discuss here the need of securing massive usage of AI for 6. AI has certainly caught tremendous attention has it has the potential to significantly change the operations of the network. But AI itself is subject to adversarial attacks that require security preserving the integrity (attacks degrading the AI models and models functionality), Availability (attacks interfering with expected operations).

The intrinsic security of the AI process depends on capabilities to prevent adversary to attack the models by input data poisoning, use response to queries in order to steal personal data or learn cyber defence parameters.

Beyond defending AI, societal concerns such as potential biased usage of AI should be purpose of research to guarantee legitimate use of the technology and build trust in 6G and services enabled by 6G.

- 6G-AI models security toolbox including
  - AI environment (training, development, production) evaluation
  - Vulnerability assessment of AI models and their applications
  - Protection measures along life cycle of AI models
  - Specific measures for constrained environment using frugal AI (embedded AI)
- xAI including both Statistical and Symbolic AI building Trust in AI usage in 6G

4.3.4 Human-Centric Multi-Agent & Federative Learning

A specific focus is given to multi-agents and federative learning as promising approaches fitting the distributed nature of 6G. Positions ranging potentially from typical far Edge form factors up to cooperative interdomain applications, the multi-agents & Federative learning should be developed with the intent to limit information spreading taking operations needs but also privacy and Human-Centric protection.

- Private, secure close-to-the-source learning frameworks, decentralized analytics.
- Privacy preservation monitoring and security protocols
- From Edge-AI to the Cloud 6G-enabled secure interworking

4.3.5 Service-Based Architectures

6G Systems and Services users, either Humans or Machines, should be empowered to choose their Digital Services as a function of security information and transparency associated to those services. In turn, Service providers investing in better security should be able to expose their differentiators to users. Through APIs, Catalog of services, market places or any service delivery workflows, security attributes should be exposed. The first research challenge in this domain is the ability to define and expose trustable security attributes. The attributes will then enable smart usage of 6G services and purpose of service agreement, composition End-to-End where security matters.

Either in public or private 6G architectures integrating Security as a Service (Managed or co-managed by Security Service Providers, MSSP) remains an interesting perspective allowing best of bread security maintaining up-to-date knowledge (Cyber Threat Intelligence), benefiting from mutualized Security Operating Centers platforms and tools. Examples of existing Security services are Identity and Access Management as a service, Key management as a service, emerging xDR/monitoring services, etc… Nevertheless, these approaches are often limited to a specific organization, and as such are not capable to provide integrated holistic understanding of security conditions. Vertical integration (from physical layer to Service layer) and Horizontal (multiple domains E2E)

The service-based architecture consist in the two following independent topics:

- Security Attributes exposure & smart usage
- Security as a Service (vertical & Horizontal) Integration
- Secure (as a Service) ZSM

4.3.6 Research Challenges

The following table identifies key research challenges.

Research Theme	*6G Security Architectures*		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Security Distribution in 6G *(4.3.1)* *Holistic Security distribution enabling AI-based Security Operations across multiple layers/stakeholders in 6G architectures (cell/cell-less/3D/…)* ***Mission-based Security Distribution in 6G*** *(4.3.1)* *Covering exhaustive security functions in exhaustive 6G architecture patterns with integration of Sensing and Digital Twins as Termination points and Root of Trust for Mission-based operations (including response)*	*Short-term* *Mid-term (with optional future long-term as per 6G Architectures evolutions)*	Reference architecture & framework for security deployment. Granular interfaces (rom micro-services to organization wide) and protocol exchanges (data models, syntax, semantic) for cooperative E2E security. Beyond obsolete perimetric approaches, integrating evolved 6G architectures with AI at the Edge and Root of Trust for smart remediation. [KPI] **Reference architecture & Framework coverage** <u>2025</u>: main 6G Architectures <u>2028</u>: most 6G architectures with cooperative operations <u>2031</u>: exhaustive applicability with Mission-Oriented applications	Reaching constancy between security and system nature and complexity. Pre-requisite for adoption by most of vertical sectors and final users. Fostering Europe competitiveness in Security aspects of 6G. Actual security transformation mitigating increasing attack surface and system architecture evolution, all with final mission objectives.
Differentiated 6G Security *(4.3.2)* *Definition of actionable Security level criteria & associated control mechanisms*	*Mid-term*	Resource segregation and security profiling in 6G architectures Priority and Precedence mechanisms applications. [KPI] **Multi-Level Security 6G resource profiling and associated mechanism** <u>2025</u>: plan-based segregation (user/control/management/service) <u>2028</u>: flexible profiling of 6G resource as per security level and associated control	Delivery of relevant security level/features through Differentiated services level and user awareness
Secure AI for 6G *(4.3.3)* *Protection of AI usage in 6G*	*Short-Term* *Mid-term*	6G-AI Toolbox including: - 6G-AI life cycle evaluation - Vulnerability assessment & risk analysis - Protection measures - application to constained (frugal) environment - adversarial attack detection - xAI & Trust [KPI] **AI Tool Box Availability & Enforcement**	Pre-requisite, thus enabler of massive AI usage societal acceptance in 6G

		2025: recommendations for most of 6G-AI protection 2028: coverage of constrained use cases 2031: Regulation as per evolution of AI Act, xAI	
Human-Centric Multi-Agents & Federative Learning *(4.3.4)* *focus on distributed approaches enabling Human-Centric protection*	*Mid-Term*	Private & Secure learning frameworks & Federated Learning applicability from Edge to the Data Center	
Service-Based Architecture *(4.3.5)* *Integration of Security either as attributes to digital services in general or as Security services;*	*Mid-Term*	Security Service Level Attributes exposure & usage (awareness, E2E composition) Security Services integration (SecaaS) [KPI] **Security Integration in Services** 2025: available SSLA template for most of service delivery workflows. 2028: Digital Service life cycle integrating provisioning of Security as a Service for most common usages 2031: International mutual recognition of SSLA	Digital decade must come with security but security delivered in similar agile mode as the rest of digital services components. It should also leverage skills and competitiveness of EU security industry

4.3.7 Recommendations for Actions

Research Theme	***6G Security Architectures***		
Action	**DiffSec (4.3.2)**	**AI-based 6G (4.3.3)**	**SBA (4.3.5)**
International Research	**X** *Regional angle considerations for global applicability*		X Global applicability
Cross-domain research		X Sharing with AI community, in particular for xAI	
Regulation/NSA at least ENISA	X May rely on numerous standards and region-based regulation		X Mutual recognition and composition rules

4.4 Strategies and paradigm shift

4.4.1 Beyond perimetric strategies

There is a mismatch between the previous generation security strategies, basically inherited from middle-age (even defence in depth) and the nature of 6G systems. Some promising directions have emerged in then last period for disruptive strategies. Most of them are taking benefits of flexible technologies and are enabled by the technology evolution. 6G research Roadmaps for some of them may be relatively short term (i.e. “Zero Trust”/confidential computing paradigm) others are longer term perspective.

Among known strategies of interest, we can mention:

- Deception aiming at luring the attacks with fake emulated systems,
- Moving Target Defense (MTD) and its potential derivations taking benefits of 6G flexibility may change continuously system morphology to prevent attacks,
- Spatial fragmentation of data, processing, routing making difficult or even impossible re-assembly of the information. This type od approaches being also potentially used for recovery with given N:M replicates of data fragments.

Listed here as one topic, each strategy may be subject to a set of research challenges

- Innovative and disruptive strategies

4.4.2 Black-Boxes and new attack Tolerant Architectures

The postulate here is simple: we will de facto never have homogeneous security levels (or just understanding/trust) across the technologies and architecture of 6G systems. Thus, instead of seeking for security grade elevation, it should be more efficient to study integration of weak or unknown parts with appropriate countermeasures. Most probably those countermeasures will have to be complex, smart, active, detecting, filtering attacks, misbehavior, anomalies in a consistent way as per the overall service objectives. A simplistic example may fit into IoT gateways, filtering data and detecting attacks (DDoS for instance) from low-security objects.

- Countermeasure integration for untrusted sub-systems and Services

4.4.3 Recovery strategies

Similar to Disaster Recovery philosophy, 6G critical systems should anticipate massive and sophisticated attacks able to create a Nation-wide blackout. Obviously, the strategies and mechanisms to be envisaged here must be self-protected. Research challenges may encompass minimal set of resources protection preserving identified critical missions, graceful remediations minimizing service interruptions with scheduled service resilience.

- Mission-critical aware degraded modes and graceful remediations

4.4.4 Per vertical specific security profile

A major application of Security differentiation should be the purpose of vertical sectors specific profile. On a case by case basis 6G should demonstrate its capability to satisfy key requirements in terms of security and security assurance.

- Specific requirements and solutions as per vertical needs (e.g. synchronization, formal proof,;..)

4.4.5 Enhanced authentication, integrity verification

Two main dimensions are here subject to innovative or even disruptive approaches. The first one is about Access Control in the wide sense where various perimeters from data to services require highly dynamic Identity & Access management eventually together with contextual/micro-rights control. The second one is about integrity, just to give an example PQC signatures should ensure integrity of manifest to upgrade IoT firmware; More generally software and service orientations mandate trustable sources through integrity guarantees. This is also quite correlated to some Integrated Sensing Capabilities involved in system command and control.

4.4.6 Continuous Security Assessment: Keeping Pace with Change

The dynamic nature of 6G, with its ever-changing network configurations, software updates (including in a DevSecOps mode), and evolving threats, demands a move from static to continuous security assessments. Research should explore innovative methods for real-time monitoring of security posture, with continuous conformity and vulnerability scanning tools. This will enable proactive identification and mitigation of security risks before they can be exploited and offer potential interworking within the parties involved. This research is to be considered as correlated (source) of smart and dynamic security adjustment, under resource constraints maintaining expected security level expected;

4.4.7 Research Challenges

Research Theme	**Strategies and paradigms shift**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Beyond Perimetric Strategies *(4.4.1)* *What is the future strategies for 6G Cybersecurity.Has the current perimetric already failed?* *Pushing Zero Trust, Confidential Computing, Deception, Moving Target Defense, quantum-based or any disruptive strategies enabling the expected fundamentals of security (CIA)*	*(short to) Long-term*	Innovative and Disruptive strategies applied to or leveraging new technologies and 6G architectures. Examples of such approaches are given in challenges description. The approaches should be promising in terms of capabilities, functional and/or non-functional gains. [KPI] **Security Capabilities** 2025: selection of most promising directions as per preliminary capability gains estimates 2028: some gains from strategies demonstrated and validated and/or first deployments. 2031: Strategies mostly adopted for actual 6G deployments.	Decoupling between systems & technologies and security is a challenge and a risk for the Digital transformation itself. Digital future is dependent on security evolution and security is a valuable differentiator and investment in sustainable democratic way of life. Investment in security research for 6G is paving the way for a safe future.
Black-Boxes and new attack Tolerant Architectures *(4.4.2)* *Integration of blackboxes and new attack surfaces in 6G conditioned by relevant mitigation strategies keeping security level.*	*Short-term*	Despite unavoidable weaknesses and/or black areas in the systems and services, the target here is to develop strategies (policies, solutions) to provide expected security levels. [KPI] **6G Risk Tolerant Architectures** 2025: IoT risks mitigation by means of filtering/monitoring 2028: Most of termination points vulnerabilities mitigations 2031: End-to-End continuum assurance through advanced strategies for most of 6G architectures	
Recovery Strategies *(4.4.3)* *Anticipating hierarchical recovery strategies for Mission-Critical services (6G dependent)*	*Mid-Term*	Scheduled plan for 6G dependent service recovery as per critical needs (up to nation/Region wide) [KPI] Recovery plans maturity 2028: most of essential and critical services performing	Lack of stability worldwide, remind every day the need to anticipate recovery plans as our society is more and more depending on Digital systems and services.

		dependability analysis and recommendations including degraded modes for recovery phase. 2031: recovery plans concomitant deployment with continuous adaptation.	
Vertical Specific Security Profiles *(4.4.5)* *Completion of KPI set per vertical applications. This should encompass security levels and specific attack surface.*	*Short-Term*	Providing matching between specific verticals and 6G security capabilities (potentially any functional and non-functional requirement expected from 6G. 2025: >80% of 6G applications forecast covered. 2028: >90% coverage	Expected value is to enable extended applicability of 6G (and beyond) to specific vertical needs. Thus a clear enabler for economic development but also essential services such as Health or Public safety.

4.4.8 Recommendations for Actions

Research Theme	Strategies and paradigms shift
Action	**Vertical Specific Security Profiles**
Cross-domain research	X Liaison with related programs

4.5 Data Centric Security in 6G

The digital world is data centric, so 6G is as well. Chapter 2 introduced the concept of dynamic resource composition tailored to individual service requirements. This composition includes a virtualized, omnipresent control fabric; a smart CIC fabric within the execution/data plane (also encompassing a virtualized control plane); and an integrated, intelligent management plane with autonomous capabilities. These elements are essential for deploying and enforcing security properties in alignment with security policies. The chapter also explored the smart CIC fabric, which intelligently identifies where resources are needed and available, addressing both compute and data resources. This approach connects locality-aware and resource-aware strategies to create solutions that better meet human users' needs—ranging from efficiency to sustainability—forming the basis of a human-centric networking architecture. Although data-centric security issues are not unique to 6G, the architecture introduces an inherent set of challenges that require both foundational understanding and practical, near-term solutions..

Intra-6G Data protection, often based on advanced cryptographic technologies, is key for privacy and confidentiality. Similar to trackers covering a majority of website, (too) invasive massive data collection and information can be retrieved from behaviours and 6G usages. This constitutes a societal risk for citizens but also an economic risk from Economic Intelligence and its illicit version industrial espionage.

6G Systems and Services will provide much more than "pipes", In-Network or Edge computing has evolved into a standard component. As such 6G will integrate data processing in numerous architecture patterns. Intra-6G Data processing should be secured and should deal with the generalized ciphering. Data may self-described (Sticky Policies) with whom it can be shared, who can simply access, what functions are applicable… In the longer term this raise some Data/Hardware binding Secured optimizations (metadata, neuromorphic designs..).

6G will be AI-native, and Artificial Intelligence capabilities are depending on data quality, integrity and availability. Either for Users' data or System's data, Data security becomes a major concern of AI for 6G. The three classical dimensions of security must be satisfied Confidentiality (multi-party architectures), Integrity (data poisoning for instance in sensing/communications fusion), Availability (Federated, Distributed AI). Massive usage of AI in 6G control and management is directly increasing the Attack Surface and trustable 6G.is in turn a pre-requisite for trustable AI.

Data Centric security in 6G, beyond the integration of AI needs to be assured with respect to the 6G-enabled architecture diversity. We already mentioned In-network and Edge computing but most of the architectures comes with stringent requirements on Data security. This the case for Autonomous Systems (based in turn on Zero Touch 6G), Digital Twins (in particular used for operations of critical systems), Sensor-driven systems, 6G for AI for smart X (X being a city, a building, an industrial system remotely managed,…)

4.5.1 Intra-6G (All type) Data Protection

The main focus is here, the privacy/confidentiality of all type of Data. The data typology in scope is not limited to User's payload data, but also glocalization, mobility, behaviour, usages of 6G. The topics

The related aspects of Data protection in 6G considered is listed as follows:

- Post Quantum Cryptography[25], algorithms, management, Energy Efficient designs
- IoT, embedded form factors, Data security under Size, Weight and Power (SWaP) constraints, secured Data mutualization -multiple Access to sensors)
- Beyond point-to-point paradigm: Distributed Ledger, Tor-like data overlay connectivity
- GDPR conformance and anomaly detection
- Anonymisation, pseudonymisation, counter-measures against inappropriate learning
- Lawful Interception in 6G and antagonism to data confidentiality measures.

4.5.2 Intra-6G Data Processing

Basically, Intra-6G Data processing is addressing 6G In-Network computing capabilities. This research area is following two main directions:

- Recognizing the evolutions towards Data Centric Security, Zero Trust Architectures, or Confidential Computing there is a strong need to provide means to manage what to do with the data, how to expose this information, dedicated software/hardware implementations. Data Security Sticky policies (owner defined): open challenge to attach to the Data the necessary information (Metadata, self-sufficient) usable for its processing in the 6G architecture. Beyond the definition or even standardisation of the syntax and the semantic of this information, the field may provide advanced and sophisticated ontologies for access and usage, identity-based solutions, traceability (for instance by means of watermarking).
- The second ambition is to re-think the Data Plane to make it, by design respectful of Data processing security. This binding between Data and Hardware may benefit from evolution of cryptographic technologies such as Full Homomorphic Encryption (FHE), Multi Party Computation (MPC), neuromorphic designs, metadata processing…

4.5.3 Data powering 6G AI

As AI applications in 6G covers multiple different scope, the Data security will result in various level of requirements. One may for instance consider differently the local data for massive MIMO-implementations

[25] Further discussion on quantum cryptography is done in chapters 7 and 10.

contributing to future modulation and coding schemes compared to data poisoning risk as root of AI for 6G End-to-End control and management. It can address societal concerns such as potential biased usage of AI and includes both the threats directly applicable to user data traffic, and their control and management.

The 6G scope is including a specific interest in federated learning architectures and platforms close to the edge, to enhance data protection, improve inference reliability, and increase autonomy of end clusters.

- Integrity and Availability of Data enabling smart control, management, service
- Data and/or models inter-domain (End-to-End, Multi-layer, Multi-party) distribution secured and trustable approaches. Benefits from promising technologies such as Distributed Ledgers, blockchains or zero Knowledge Proof (ZKP)
- Resilience and recovery responding Data failure
- Abnormal Data detection

4.5.4 Data security (CIA) in relation to exogenous impacts

The data centric security in 6G is a criterion for applicability of System architectures evolution forecasted. According to the usages of Digital Twins, Remote management, robotics, autonomous systems or even indirect contribution to sustainability, the requirements for Data Centric security in 6G will mandate specific security level. One should notice that integrity and availability is becoming more and more important for Data.

- Data security for 6G-enabled from sensing to Digital Twins or remote management including real time availability
- Data security for 6G-enabled autonomous systems
- Data security for 6G-enabled sustainability optimization

4.5.5 Research Challenges

Research Theme	***Data Centric Security in 6G***		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Intra-6G Data Protection *(4.5.1)* *Set of data protection topics applied within 6G systems including cryptographic protection, anonymization, DLT and GDPR conformance check.*	*Short-Term*	Protection of all type of Data within 6G (users' payload, usage data extracted from system, 6G control and management data. [KPI] **All Type of Data protection** 2025: > 90% completeness 2028: 100% completeness	In line with general data protection and its (among other) privacy goals, 6G must demonstrate exemplarity in Human-Centricity
Intra-6G Data Processing *(4.5.2)* *Data processing in 6G driven by sticky policies and confidentiality-preserving technologies*	*Mid-Term*	Solutions providing strong guarantees for data processing in 6G, including user's privacy and policies [KPI] **Secure Data processing in 6G** 2028: availability of solutions for most pof 6G use cases 2031: scalable applications with power constraints	In line with general data protection and its (among other) privacy goals, 6G must demonstrate exemplarity in Human-Centricity
Data powering 6G-AI *(4.5.3)* *Secure Data and its distributions for 6G-AI*	*Short-Term*	Security of Data (AI-feed) life cycle in 6G architectures, including inter-domain and federated learning. [KPI] **6G-AI data security** 2025: integrity assurance solutions for most of 6G applications 2028: full secure life cycle with anomaly detection	As the foundation of digital services, in turn serving numerous verticals, 6G must ensure that its behaviour will not be at risk with biased data and models.

Data Security in relation to exogeneous Impact *(4.5.4)* *Enabling system evolutions such as Digital Twins, Autonomous systems, Sustainable systems by fully secured data transport*	*Long-Term*	Assurance for systems depending on 6G systems and Services enabling smart and innovative applications. [KPI] **Dependability on 6G data transport** 2028: partial mapping of exogenous systems requirements onto 6G 2031: 6G systems & services commitment to most of required guarantees	Most of promising digital services, DT, immersive applications will be strongly dependent on data transported by 6G systems and associated Key performances (delays, availability, integrity,..)

4.5.6 Recommendations for Actions

Research Theme	Data Centric Security in 6G
Action	**Intra-6G Data Processing**
Cross-domain research	X Solutions strongly dependent on cryptographic technologies from cybersecurity domain and microelectronic Hardware capabilities

4.6 Hardware for 6G security & Physical Layer issues

Expected to rely massively on virtualized infrastructures, the directions enabling deployment of security functions in the hardware layers is at least two folds:

The first one, is related of the implementation and distribution of Root of Trust and Secured environment in the overall system complexity. It encompasses Trusted Execution Environment (TEE), (micro)Hardware Security Modules (HSM) in a scalable way from CPU-limited terminals to Data Centers. It may notably contribute to the so-called "Zero Trust" paradigm. Specific security appliances virtualized remain part of any network security architecture and may deserve specific integration issues with the evolution of 6G architectures.

Another direction gaining traction, is to recognize the performances and sustainability issues raised by the load in VMs and containers induced by some infrastructure functions, including filtering and other network security functions. These is illustrated by the Open Programmable Infrastructure/IPDK project aiming at offloading Network, security, learning functions.

Smart usage of security hardware is dependent on two main constraints:

- Although true for non-security specific components the control of the supply chain from design to delivery remain part of the tools required to ensure security.
- Along the life cycle the management of secret elements (keys, attack patterns,…) is a fundamental as weakness on this side may ruin any security level expected from the security hardware.

Still related to 6G & hardware security, there is constant need to take into account the security aspects of any hardware component integrated in the architecture. 6G is coming with new specific components which deserve attention. This is the case in general with sensors involved in 6G being for instance dedicated to sensing/communication fusion or Intelligent Surfaces involved in communication capabilities and such part of the Attack Surface.

Clustered here with the Hardware issues, physical layer and particularly bearer protection is the source of some challenges and a threat landscape on its own. The jamming issues have to be re-considered considering the diversity of architectures (including 3D and cell-less) and critical needs that can occur for vertical applications.

Waveform sophistication and smart but complex access protocols may be source of exploits and malicious usage. Security considerations have to be taken here with as most as possible “by design” protections.

Intention for Eavesdropping/IMSI catcher type of attacks are not going to disappear with 6G. Topics not covered by data protection should be also in 6G scope.

4.6.1 Network Security Hardware

- Root of Trust Distribution and life cycle.
- Security functions Offloading in virtualization infrastructure (from Edge to cloud)
- Trustable Degraded modes, Graceful 6G (private/public) recovery
- Secret Elements distribution management
- Beyond PQC algorithms robustness to side-channel attacks

4.6.2 Securing Network Elements

- Secure LIS/RIS
- Clocks, time/phase distribution
- Flexible Hardware secured deployment
- Advanced Supply chain Assurance, authentication and traceability

4.6.3 Bearer protection

- Anti-jamming and counter measure in 6G
- Protection against intentional source of light (optical domain)
- Authentication, IMSI catchers, …

4.6.4 Research Challenges

Research Theme	***Hardware for 6G security & Physical Layer issues***		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Network Security Hardware *(4.6.1)* *Design and Run of HW-based security in 6G*	*Short-Term to Mid-Term*	Open Framework for security function offloading in virtualization infrastructure. Trustable environments and their usages. Quantum Safe era [KPI] **Availability of trustable infrastructure framework** 2025: System Overarching with security offloading and root of trust. 2028: > 70% Applicability on 6G use cases	Security is a complex combination where secured hardware should contribute to high level of assurance. Added vale from HW is both in interest of control of HW supply chain and overall security.
Securing Network Elements *(4.6.2)* *Security of newly introduced 6G hardware*	*Short-Term*	Secured 6G components and their supply chain	Security considerations can’t be an option but should be mandatory for any 6G component participating to the 6G architectures.
Bearer protection *(4.6.3)* *Denial of Service and eavesdropping mitigations*	*Mid-Term*	Affordable Anti-jamming solutions and Attack Detection	The risk of attacks against bearers is quite Asymmetric…jamming being quite easy and cheap. (critical) services availability and privacy are dependent on bearer protection, raising

			value to mutualized infrastructure.

4.7 Softwarization

Research related to softwarization is occurring in multiple aspects in the architectures, AI-based control and management or confidential computing. The focus addressed here is a subset of software related topics, basically safe code, remaining open research on virtualization enablers and virtualization of security functions.

It starts with the source of any softwarization that is the safe 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 [C4-4] shown the tremendous importance of the software and its life cycle. Next generation and evolution are clearly re-enforcing the trend and the needs.

Many threat Intelligence aspects are already addressed in the literature including a comprehensive survey [CX-X] and numerous softwarization enablers feed on a regular basis the Common Vulnerabilities and Exposure (CVE) list. The architectural 6G trends tend to accentuate the potential issues with more distributions, more interfaces leading to more interfaces/exchanges and thus wider Attack Surface.

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. The issues related to integration of security considerations here are not limited to functional aspects. There are direct impacts on overall performances and sustainability as well as challenge evolution linked to the architecture evolution itself.
- 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 Detection & Response systems (xDR) with potentially sophisticated smart protection and detection inside.

Security is still on a long and complicated path towards fully secure and resilient 6G. Softwarization is dependent of the diversity of the platforms and their specifics (Edge, IoT, space. Softwarization is expected to follow and adapt to the evolving 6G architectures, although matching a wide range of security level expectations.

4.7.1 6G Safe Code life cycle

The agility and operational gains enabled by the softwarization on going since 5G has to be handled with the counterparts in terms of cybersecurity. As any software those taking part of the 6G systems and Services start with code. Drawing a dependence chain, it becomes obvious that software has at least as critical impact as hardware on the overall security.

Safe code is a common concern for the whole ICT domains, integration of advanced research in this area is one of the topics but it should be complemented with specific 6G aspects. A fundamental difference with some IT similar issues may be found in the distribution of the system with multiples authority perimeters, multiple layers, multiple platforms, disaggregated microservices,… Combined with the dynamicity of the systems and the intrinsic Code temporal distribution (design, supply chain, runtime, updates, upgrades, patches, traceability…) it leads to identify 4 distinct research aspects:

- Code analysis, beyond static code analysis. Mapped into 6G architectures it is critical to understand complex interactions and potential effects of vulnerabilities and malicious. Similar and actually not decorrelated from AI a focus on federative approaches should enable high levels of assurance as well as automation of the operations.
- The second critical research aspect is the code timeline with the Updates/upgrades critical phases. Most probably maintaining code security is not scalable without development on incremental analysis or even certification.
- A safe code life cycle should build upon a trusted supply chain. For the benefit of operators, system integrators or verticals users a stable and trusted source of elementary or even integrated (certified) codes would contribute to secure deployment of 6G. It should be understood that multiple verticals deploy systems with very long life cycle/generation (several decades) and would have difficulties to sustain too fast, unchecked code integration.
- Last code specific aspect is traceability, forensic and technologies for identification. On one hand approaches such as watermarking or equivalent may give some guarantee of the sourcing, on the other hand, methodologies based on code morphology (AI-based) should help to identify what is actually at stake in the large protocol stacks and code running in the systems.

4.7.2 Full Security for 6G virtualization

Ensuring fully secured virtualization enablers may appear as a solved problem as large communities use to address those issues during the last period. Unfortunately, literature is listing year after year specific and comprehensive survey [C8-12] on virtualization software generic or implementation-specific components vulnerabilities. Subscribing to Security Alert from Computer Emergency Response Team (CERT) may be sufficient to realize how long and frequent is the Common Vulnerabilities and Exposure (CVE) list. It even concerns on a regular basis well-known Operating Systems, open or not !

Sometimes, a simplistic vision is to believe that problem is solved but limited to a single authority, benefitting from a “god view” on its system. 6G is clearly going in the other direction. The intrinsic nature of networks is to interconnect, whatever the granularity (micro-service, sub-system, high level service, service provider,…) of entity interconnected overall security is only meaningful End-to-End and is not a simple sum of local properties. 6G is often coming with visions of Inter-computing, where computing may result in blackboxes (from system view) running on non-standard features and implementations. Another typical challenge is to actually control and manage security concomitantly to the other digital components of the architecture (communication, compute, storage, AI,..). This may be considered as the more complex, dynamic 6G version of the historical issues of NOC/SOC relationship (different clearance levels, different tools/view scope, different control/management scope).

Security in virtualization is often seen as an isolation/confidentiality issue. By full security here the intention is to cover all the security dimensions: confidentiality, integrity and Availability. The later is a stringent requirement for most of critical verticals while available failover mechanisms or current virtualization rely on lower layers for protection and restoration.

Last but not least, controllers and orchestrators are Policy Decision (PDP) and Enforcement (PEP) points. That is critical asset in the infrastructure to maintain sovereignty in the sense of guaranty on system behavior with respect to expectations from legitimate policy maker.

From the above statements, two research aspects are emerging as contributing to 6G security and use case coverage:

- Resilience, Protection & Restoration in virtualization layers
- Resource scheduling and performances under security (collaborative Inter-Controllers & Inter-Orchestrators policy-based operations)

4.7.3 Virtualized Security Functions

In softwarization domain, a sub-class is made of the necessary virtualized security functions. The availability and integrability of such protection, detection & response functions is just mandatory to be the companion of conventional network functions. As mentioned in previous section, performances, interactions (common syntax, semantic, APIs) across various perimeters and granularities constitute a set of research challenges in order to address evolving cyberthreats (up to Zero-Days vulnerabilities and Advanced Persistent Threats).

- Scalable Cryptographic and filtering functions
- Convergent knowledge sharing, syntax & semantic for Detection & Response. (xDR including NDR)

4.7.4 Research Challenges

Research Theme	**Virtualization**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Safe Code Life Cycle *(4.7.1)* *Code Analysis, updates/upgrades, traceability*	*Short-Term*	Framework and process enabling full life cycle mastering [KPI] **Uncontrolled code ratio** 2025: <10% code in production 2028: Zero unsafe code in critical applications	Reality of softwarization is strongly dependent on the ability to control code/software life cycle. Impact and consequences of safe code is here at the dimension of 6G usages.
Security of 6G virtualization *(4.7.2)* *Protection and restoration mechanisms for virtualization components* *Collaborative, policy-based scheduling*	*Mid-Term*	Fast failover and cooperative protocols for resilient by-design controllers and orchestrators [KPI] **Response time of protection and restoration mechanisms** 2025: <1s (most of) 2028: <50ms (availability)	Virtualization capabilities in terms of security should come with stringent requirements for protection and restoration. State of the Art today in this area is often seconds or even 10's of seconds response time and not compliant with many vertical applications.
Virtualization of Security functions *(4.7.3)* *Frugal cryptography* *xDR*	*Mid-Term*	Development of deployable security functions contributing to CPU, energy consumption reduction [KPI] **Sustainability performance** 2028: <15% supplementary cost (CPU, Energy) in E2E chain	Integration of security in 6G. As any network function should be available as virtualized function keeping its robustness properties and adding sustainability criteria.

4.8 Automation and Intelligence in Operational Security

4.8.1 Security Policy Life Cycle

Assuming that the architecture is providing the relevant Policy Decision Points and Policy Enforcement points distributions, the protocols and APIs. Multiple challenges remain to feed the system with the required intelligence and make it operational.

The first set of challenge, and the starting point of security policies life cycle, is to achieve semantic extraction, provide elicitation of user-driven requirements potentially expressed as security policies. It may also be based on Natural Language Processing (NLP) and put in Intent-based perspective. It should be noticed that talking security policies at the interface of users and providers may come with obligations and consequences on liabilities.

The second set is mapping of the user's requirements into systems policies, configuration, orchestration. As mentioned, many times, 6G is not a monolithic system under a single authority, thus the solutions will rely on various distributions and combinations End-to-End. This type of issues is already difficult with past systems using powerful solvers in bi-lateral relationship. The challenge is therefore difficult to take up and will be even harder considering that some vertical applications may require formal proof to validate the overall policy implementation.

Two priority sets of challenges towards users/providers matching in security policies:

- Intent-based, User requirements semantic extraction
- Solving distribution & combination

4.8.2 Zero touch, autonomic and multi-agent

A promising 6G application is to enable Autonomous Systems for instance Drones Swarms, Robotics, vehicles, etc… Those systems will hardly rely on communications which are not themselves autonomously controlled and managed. Zero touch is the paradigm for 6G and security must follow the same path to avoid blocking 6G while maintaining the expected level of security.

Far beyond self-configuration of security component, hybrid AI will be massively used to provide adaptive protection but also adaptive attack detection and response. The Ai capabilities are intended to be the solution in order to face the extremely large number of events and variations of the problem. The ideal case (long term) being intrinsic reasoning capabilities to deploy response to unknown attacks based on zero days vulnerabilities.

This where Digital Twins may participate to the security operations through modelling of the system, reasoning and validation features allowing smart responses.

Last but not least, coordination of security features should be done in a cooperative holistic approach to cover the scope and complexity of the systems.

The research challenges are listed as follows:

- Adaptive protection, Hybrid AI
- Adaptive detection & response, Hybrid AI, federated multi-modal xDR (Detection Response) mechanisms, eventually LLM-driven or RL-driven.
- Digital Twins for reasoning and validation
- Cooperative holistic security

4.8.3 Root cause and Identification

Attacks may basically be classified into three categories.

- The first one, Criminal is the most known as numerous visible occurrences of those attacks such as Ransomware are subject to communications and impact many users…
- More critical, espionage category may have higher societal impact but is less known that the first one. This is also not the same level of sophistication attacks and take sometimes years to be discovered. This category can already involve states or industrials mandated by states with significant number of skills and efforts. This type of attacks is less frequent but is increasing.

- The last category is nation-wide or even military grade attacks, sophisticated with the potential to create major blackout in the essential interests of a nation.

As baseline for digital transformation and enabler of numerous verticals, 6G will definitively be in the targets of the three categories and will definitely be a vector for the attacks ! In this context and at least for the most dangerous categories, the legitimate authorities must be able to conduct investigations towards identification of attackers. This is particularly complex mixing public/private means, potential fakes made on purpose to accuse someone else,… AI-based security is an indispensable means to overcome these challenge.

- AI-based security for root cause and identification

4.8.4 Research Challenges

Research Theme	**Automation**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Security Policies Life Cycle *(4.8.1)* *Intent-based User to System/services security policies mapping* *Solving E2E policies combinations*	*Mid-Term to Long-Term*	Fluid interface between needs and solutions respectful of security policies [KPI] **Automation of policy life cycle ratio** <u>2028</u>: >50% <u>2031</u>: >75%	Simple, easy to,use and accurate translation of user policies into systems operations and configuration.
Holistic Zero Touch *(4.8.2)* *Adaptive protection* *Adaptive Detection/Response* *Digital Twins for reasoning* *Cooperative Holistic Security*	*Mid-Term to Long-Term*	AI-based security control and Management I 6G Architectures [KPI] **Solution availability** <u>2028</u>: >75% Protection, xDR <u>2031</u>: >50% Reasoning & Cooperative Holistic applicability	Squeezing the most of AI technologies for autonomous operation of security in 6G.
Root Cause and Identification *(4.8.3)* *Multi-Source: forensic, OSINT,…)* *AI-based Identification*	*Long-Term*	Defense against ever growing sophistication of all types of attacks	Strategic capabilities

4.9 Security Quantification and Evaluation

Trustworthiness is not going to be built on declarations (potentially under conflict of interest) basic static measures. A major challenge for 6G is to develop an entire framework allowing definition, evaluation of security levels and awareness/usage of those levels.

4.9.1 Quality of Security (QoSec) and relation to Security Service Level Attributes (SSLA)

The closest analogy to understand the QoSec is the quality of Service in general. Service security is not a new concept and was already mentioned in the ITU-T E.800 in late eighties. Nowadays, criteria definition is the first step and means to evaluate it the purpose of research challenges. The path towards QoSec is potentially inspired by certification history, but needs to integrate 6G complex scopes and also societal aspects such as immunity against foreign laws.

QoSec should be the roots of SSLA and basic enabler of exchanges within Service-based architectures (APIs, Intent-based, Service Catalogues, marketplaces, B2B/B2C interfaces...)

The related research topic is then:

- QoSec definition and evaluation
- Related standards

4.9.2 Continuous assessment of security conformance along life cycle

The security monitoring is facing new challenges as stationarity of the security conditions (resources, features, perimeters, updates variations in time) are not verified in virtualized and flexible 6G architectures. It leads to develop on the fly means to evaluate the security conformance providing continuous assessment. It also the benefit of security to check security much frequently and detect potential anomalies.

Within assessment scope, there is also a challenge to demonstrate E2E provable security with the two dimensions: composition by combining multiple, lateral sub-systems and services, and incremental to deal with change in time of the configuration/composition of the systems/service.

- Continuous assessment of 6G security including the full life cycle supply chain
- E2E provable security including composition and incremental methodologies
- Representative Platforms and data lakes

4.9.3 Research on Economic and Societal impacts, liabilities

Security evaluation is expected to participate to regulation, liabilities, insurance issues and perception of 6G impact in general. At the crossroad of Human Sciences, legal laws and technologies, research topics targeted here are:

- Usage of security quantification and evaluation in 6G societal impacts

4.9.4 Research Challenges

Research Theme	*Security Quantification and Evaluation*		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Quality of Security *(4.9.1)* *Definitions, roles (vs. Risks and Trust), Evaluation*	*Mid-Term*	Integration of Multi-Level Security in Users/service & service/service interfaces. Evaluation framework in relation to Regulation.	Before limited to certification to limited scope, concepts of QoSec should enable knowledge, differentiation, added value and awareness of 6G systems & services. User empowerment.
Continuous Security assessment *(4.9.2)* *On the fly means to capture security conditions variations as per transient system states*	*Long-Term*	Trust by monitoring of security boundaries up to provable conditions	Evaluation capabilities should evolve as systems are evolving. State of the Art is somehow late and poor today compared to the need. Any service should come with means to monitor the service quality, security attribute is one of them.
Economic & Societal impacts *(4.9.3)* *Convergence of 6G security and Societal Impacts*	*Mid-Term*	Using quantification and Evaluation as a Factor of 6G success. Actual figures usable by economic (insurance, industries,…) and society as a whole	Vertical specific regulation or simply cyber-insurance issues, smart usage of secured digital services and means to achieve that have a massive societal impact.

4.9.5 Recommendations for Actions

Research Theme	Security Quantification & Evaluation		
Action	**QoSec**	**Economic & Societal Impacts**	**Research Aspect N**
International Research	**X** *Global consensus & Standards*		
Cross-domain research		X Human Science (Legal Law, Economy, Sociology)	**X**
Regulation Authorities	X Liaison and relationship with Certification/labelling framework		

4.10 Security Governance

There is a potential EU perspective for cooperation among private and public organizations to develop a dedicated 6G Cyber Threat Intelligence (CTI) as well as building cooperative response to incident. This type of logic is already at stake in other ICT domains and fully make sense for 6G.

- Security knowledge sharing for 6G developments
- Compliance to certification frameworks (including evaluation)

4.10.1 Research Challenges

Research Theme	Security Governance		
	Timeline	**Key outcomes**	**Contributions/Value**
6G-CTI *Building common knowledge on 6G security*	***CHOOSE ONE*** *Mid-Term*	CTI platforms and policies across EU or wider)	Strengthening 6G security posture and response, at least starting and reinforcing European cohesion.

4.10.2 Recommendations for Actions

Research Theme	*Security Governance*
Action	**Research Aspect 1**
Cross-domain research	X EU initiative in liaison w/ ECCC, ENISA and NSA

5 Software and AI technologies for telecommunications

Editor: Josef Urban

5.1 Introduction

Software technologies are one of the fundamental enablers of telecommunication networks and they increasingly shape network architectures and capabilities. For example, 5G has adopted a service-based architecture (SBA) for its core functions providing flexibility and scalability. Standardized APIs (e.g. Network Exposure Function – NEF) have been introduced to provide applications with access to network resources and data in a controlled manner. Network slicing is a key concept of 5G, built on NFV, SDN, and the flexible SBA of the 5G core allowing the dynamic creation of multiple virtual end-to-end networks across the same physical infrastructure and offering network services tailored to specific use cases. The cloud has become integral part of the telecommunication infrastructure allowing “as a service” models to expose network capabilities. Software components of network functions are designed and implemented following a cloud native approach and using technologies such as containers and microservices. DevOps approaches are applied to develop, integrate and deploy network services and software updates in an agile way. Open source software has become increasingly important to be witnessed for example in case of open software for the RAN [C5-1] or the use of software solutions provided by the Cloud Native Computing Foundation (CNCF). Open source has been proven to be a successful model allowing competitors to work together towards common platforms and de-facto standards not only in the telecommunication industry, but more general in software-intensive business ecosystems.

This section addresses the ICT continuum, but a notice should be done that computing for telecommunication internal infrastructure and computing provided to users are currently two often separate fields. While internal usage is dominated by questions of efficiency moderated by external factors like privacy and cost, there are larger challenge in the use of computing resource in the network by external actors. Operators face a wide variety of challenges that prevent them to provide easy access to computing resources in locations that are convenient for the external actors; challenges that include operational efficiency, required investment, maintenance, security, privacy and many more. For the external actor, there is also the challenge of interacting with a choice of operators, who on top of that are not active in the entirety of the Union. A strong set of standards are thus needed for exploring and accessing these software opportunities across all (or the majority) of operators.

5.2 Vision

The network softwarization will continue. For example, TM Forum’s ongoing autonomous networks program abstracts “the network as a set of software services and then use intent, automated closed control loops, and machine learning to make networks and operations self-configuring, self-optimizing, self-organizing, self-healing, and self-evolving.” [C5-2] The journey towards autonomous networks will still take some time and requires tackling technological challenges such as the development of effective data frameworks for real-time analysis, the integration of new technologies across network domains, or the security of autonomously operating software units and systems.[C5-3]

Future networks are envisioned to be built over heterogeneous federated clouds, whose resources are homogenously managed by a unified control and orchestration framework. This framework will form a computing continuum in which network functions and services will be created, deployed on demand, subsequently scaled, and seamlessly moved across the federated cloud infrastructure. The computing continuum will be able to optimize autonomously service performance and possibly off-load computation based on in-depth knowledge about the capabilities and resources exposed by the federated clouds. Service meshes and workloads will be based on stateless and serverless functions, microservices running in containers and virtual machines, as well as new advanced concepts that allow to optimize the use of specialized high performance resources including quantum computing resources. The concept of the computing continuum will impact architectures, interfaces, and the disaggregation of networks.

The network softwarization will also include the extensive use of artificial intelligence and machine learning models throughout the network and even at the radio level, and digital twins will allow to simulate, test and manage networks. Software-based capabilities of smart networks will play a significant role for the commercial success of SNS ecosystems by addressing the digital needs for automation, adaptive and customized services, as well as the needs for agility in delivering complex, but reliable and trusted software and services.

The network softwarization requires research to get answers on questions such as how to manage the lifecycle of AI/ML components and to assure access to and the trustworthiness of data required by AI/ML, how to guarantee that a self-adapting AI/ML component will behave within its design parameters, or how to engineer and integrate such a software-intensive system in general so that the growing system complexity can still be managed. It has to be explored how the software needs to be architected so that it is best adapted to the distributed 6G system and benefits most from the capabilities offered by the ICT continuum across devices, edge, and cloud. Even quantum computing resources will become available as part of the ICT continuum raising the question how to integrate and use these special compute resources in 6G. We also need to understand better how the non-functional requirements of sustainability (including energy efficiency), human-centricity, and resilience will impact the software architecture of 6G systems and applications. The software architecture will also be impacted by regulations such as the European data act and the AI act.

The following software research themes have been identified and will be further outlined in the following sections:

- Artificial Intelligence. AI and context- awareness enable capabilities to observe changes in the network and the edge-cloud continuum, to act autonomously on those changes, and to learn for the future. Especially, decentralized AI mechanisms, such as federated learning, split learning, or swarm learning are explored as they meet best the specifics of the distributed and heterogenous network and edge-cloud continuum as well as the needs of a cross-layer orchestration. The enormous success of generative AI raises the question how LLMs can be used in and supported by 6G networks. Further research gaps exist regarding the AI driven prioritization of user and application needs in multi-operator networks, the AI support for intent-based human-network communication, and the AI-based enhancement of user experience while preserving privacy.
- Edge cloud compute continuum. These research area addresses the challenges to design software and applications so that their components can be distributed across the edge-cloud continuum in an optimal way to meet for example performance, security, or usability requirements, including the off-loading of computationally intensive and delay-sensitive tasks to edge nodes. Complementary research is needed on robust frameworks that support applications in making efficient use of distributed network-based computing and sensing resources within a converged computing and communication architectures as well as on intelligent frameworks that enable proactive network customization and optimization. Edge IoT immersive platforms require open and interoperable

solutions that allow seamless, real-time, concurrent collaboration, open APIs, compatible data formats, and protocols.

- Digital twins in the SNS context. Networking and the computing continuum will be important enablers providing the infrastructure and basic services required for the implementation of digital twins. Digital twins will be also used for monitoring and augmenting SNS systems.
- Data frameworks. AI and digital twins rely on having access to accurate and up-to-date data. It is the task of data frameworks to collect, clean, process and store data in a distributed SNS system. Challenges in this context are the alignment of network software and data lifecycle as well as ensuring the compliance with related data regulation.
- Engineering complex, software-intensive, and self-adaptive SNS systems. Managing the software complexity of a system of systems is becoming an increasingly challenging task and requires new operational concepts based on self-adaptation models and relying on AI algorithms as well as new SW engineering approaches for software intensive systems.
- Human centricity and digital trust. SNS-based services should follow a service model that enhance human-centricity, meaning that services need to be trustworthy and ergonomic including easy to use and easy to access. In consequence, the development of software and services need to follow a "human-centricity-by-design" approach.
- Quantum Computing. Quantum computing is an emerging technology in the telecommunication context. The key research questions are: what are the telecom specific complex problems and algorithms suitable for quantum computing and how those algorithms can be implemented and integrated with classical computing?
- Sustainability. SNS systems are expected to support verticals in achieving their sustainability targets. But sustainability should also be a design principle of SNS systems. Research needs to identify the concepts, models, and mechanisms to achieve this twofold objective.
- Software security. The specific focus here is on software or AI as the vulnerable asset that may be attacked, and which may result in harms to itself or other assets it affects.

5.3 Metrics, KPIs and benchmarks

The proposed research themes aim at improvements of various software related aspects such as the performance in service delivery, the human-centricity of network services, or the reliability of network systems. Metrics, KPIs and benchmarks will allow to measure the progress and the improvements that research will achieve with regard to those aspects. There are existing and well-established measurement approaches in software engineering and software deployment that can be used to also measure the achievements of related research. However, in other areas such as human-centricity those metrics might still need to be developed in the context of the research.

KPIs that might be applicable in the context of software-related research for future networks include

- KPIs to measure the DevOps performance [C5-4][C5-1]: deployment frequency (how often releases and updates to production are done; lead time for changes (the time it takes to get a commit into production); change failure rate (percentage of deployments causing a failure in production; time to restore (time it takes to recover from a failure in production)
- Measuring the accuracy of machine learning models and the efficiency and failure rate of MLOps
- Sustainability related KPIs that can be defined on the basis of the Energy Efficiency Directive [C5-5]
- Benchmarks and KPIs to assess security by design and secure cloud deployment environments [C5-6]

5.4 Artificial Intelligence

5.4.1 The Role of Decentralized AI

Over the last decade, the softwarization (e.g., Software Defined Networking, SDN) and hyper virtualization of the network infrastructure, coupled with the interoperability of different networking technologies - e.g. short- and long-range wireless, terrestrial and NTN technologies - created the ground to support next generation Internet services. AI-based mechanisms bring the promise of integrating into the networking planes the capability to observe, act, and learn, to improve their performance, and to best serve applications and the user.

Recent trends towards an AI-powered Edge-Cloud continuum (metaOS projects such as NEMO [C5-7], airOS [C5-8]; cognitive computing projects such as e.g., CODECO [C5-9], COGNIT [C5-10], COGNIFOG [C5-11], DECICE [C5-12]; swarm computing projects such as SwarmOS, 1-SWARM [C5-13], Zero-SWARM [C5-14]) address the use and engineering of AI across different vertical domains, with processing being pushed closer to the data sources. Processing at the Edge or across a decentralized, mobile Edge-Cloud requires a new design paradigm for the Edge-Cloud, ensuring that service decentralization, mobility, large-scale dense environments, and multi-tenant support can be provided across 6G environments.

The AI cognitive capabilities can bring a further level of networking automation (basis for self-adaptation) and recommendations (basis for self-awareness and self-healing). However, it is important to highlight that adaptive processes in the context of 6G Edge-Cloud environments supporting heterogeneous mobile devices require distributed behaviour learning and inference techniques that can support decentralization across Edge-Cloud, while preserving the privacy of the raw training data [C5-15]. The approaches have the capability to address specific 6G challenges. For instance, hybrid Federated Learning (FL) addresses the Edge-Cloud continuum as a multi-layer, cluster-based structure [C5-16]. Split Learning (SplitNN) is a more recent distributed and private deep learning technique that can be used across edge-cloud devices while improving in terms of scalability and minimizing the need to share raw data directly [C5-17] [C5-18] [C5-19]. Another relevant learning techniques regards Swarm Learning, a powerful technological concept for industrial applications that involve cyber-physical systems, as it can provide flexibility in learning based on the interaction of IT systems, CPS systems and humans. This decentralized operation of swarm intelligence systems obviates the need for centralizing knowledge, thus offering speed, scalability, and potential for devising optimal solutions. These properties are highly desirable in the case of deployment reconfiguration scenarios. Another relevant dimension is associated with sustainability of the AI processes, which need to be efficiently managed.

Federated learning involves training AI statistical models over distributed and heterogeneous equipment, including data centres, end clusters (e.g., end IoT/IIoT devices, on-premises servers, etc.) or remote devices, while keeping data as much as possible localized. Federated learning brings AI models close to the edge to enhance data protection, improve inference reliability, and increase autonomy of end clusters. The cloud (data centre) plays a federation role for aggregating insights from different IoT edge distributed clusters to generate a federated model shared with each individual cluster. Training in potentially large and heterogeneous networks, such as 6G cellular networks, introduces novel challenges that require a fundamental change compared to standard approaches for large-scale machine learning and distributed optimization, and provide the potential for privacy-reserving and democratic AI training and application. In order to enable this shift to distributed AI models, 6G infrastructure with computing and storage resources along the network, including embedded neural processing resources, will be needed, together with software platforms to support the development and operation of the federated learning pipelines.

The research questions and challenges to be addressed in this context include:

- Up to which point should decentralized AI be considered in the context of 6G use cases requiring an elastic Edge-Cloud support?
- How can AI cope with mobility handling in large-scale 6G sensing scenarios?
- What is the best model for an AI-empowered control and management plane?
- Accessibility to required AI training data and privacy protection of those data.

5.4.2 AI-based Cross-Layer Orchestration

Edge-Cloud architectures are assisting in a service decentralization based on aspects such as virtualization, network softwarization, and AI-based orchestration. Orchestration in this context is focusing on a cognitive, AI-driven orchestration plan, resilience, cooperative behaviour support based on a combination of context-awareness considering metadata from the computational layers, networking layers, data observability layers, together with decentralised AI approaches. Context-awareness in this context refers to the capability of a system to consider knowledge about its environment, to perform specific actions. Via such combination (data-compute-network), orchestration [C5-20] can become more adaptive to challenges such as intermittent connectivity, mobility, and data variability, thus providing better support to complex applications such as digital twins. The capability to use context-awareness towards specific target profiles such as greenness, i.e., a user-based preference in terms of energy consumption, CO2 emissions, etc., provides the grounds to integrate aspects such as energy awareness by design. AI decentralized approaches, such as swarm learning, play a pivotal role in the capability to provide cross-layer cognitive orchestration.

The use of decentralized AI such as SplitNN, GNN and Swarm Learning, mentioned in the former sub-section, is relevant when considering the injection of parameters collected from the network, application requirements, data models and meta-data compliance, as well as user behaviour. In this context it is relevant to address, design, and deploy decentralized AI-based approaches that can take into consideration data sovereignty of a region (e.g., an AS) and that can nonetheless allow for an optimal deployment of 6G services across a mobile, large-scale Edge-Cloud infrastructure. These modules will be enhanced with decentralized decision support algorithms that consider local information only, while being able to contribute to the global optimization through their participation in the swarm network.

By making the network context-aware and cognitive (AI-based), the network will gain a new level of adaptability, being capable of supporting autonomic decisions. The objective of such decisions may aim at improving how the network operates, how it performs in the services delivered to the customers and how the network itself is managed, configured and healed when issues arise. For such reasons, 6G is expected to deliver a network with built-in AI capabilities, where analytics functions are potentially collocated in every network function instance, take part in the business logic and decision-making and provide highly distributed optimizations, potentially within any of the existing and future defined network functions and procedures.

The Service Based Architecture of the fifth-generation mobile network already contemplates the integration of data analytics functions paving the way for the future developments and making the network more proactive. However, in a bigger picture to deliver its promise, 6G networks must be complemented with solutions such as low power wide area networks, transparent end-to-end TSN communication with wired and wireless devices or new interactions with three-dimensional networks, such as the proposed Space-Air-Ground Integrated Networks (SAGIN). It adds a new level of complexity to the expected 6G networks, which, thanks to these built-in AI/ML capabilities, will be able to provide optimization to end-to-end services, even proposing an integrated digital twin of the heterogeneous network itself, supported by these AI/ML capabilities. While cognition and context-awareness need to be handled together, achieving an end-to-end and coordinated proactivity over the heterogeneous 6G ecosystem remains a big challenge to be addressed.

A first challenge is to address this orchestration considering cross-layer approaches that can sustain heterogeneous networking technologies, diversified and hybrid Edge-Cloud environments, while ensuring data privacy and sovereignty.

A few research challenges in this context are:

- What is a “minimum” subset of data that can define an abstracted region, e.g., AS, cluster, RAN? (short)
- Which networking semantic composition models best suit the requirements of 6G use-cases? (mid)
- Which methods can be considered to adequately encode and expose data across a 6G Edge-Cloud infrastructure (mid)
- Is decentralization going to be supported just at a control plane, or also at a data plane? How will cognitive networks impact 6G applications? (long)

5.4.3 Large language models and Generative AI in 6G infrastructure

Large language models and other Generative AI are bringing new challenges and opportunities to 6G.

Generative AI models are extremely resource consuming for both training and usage, and therefore, the current research and application is highly dependent on large data centers in the central cloud. However, potential applications of generative AI, e.g., those involves VR/AR, robotics, self-driving, etc., will demand high-performance connectivity and low-latency models running close to the data sources or data consumers, to generate and deliver content at near real-time. This requires the future 6G networking infrastructure with highly configurable network functions, together with different levels of edge data centres that can efficiently host generative AI models. A 6G network hosting distributed generative AI models would also open the possibility of federated generative AI – the models can be continuously fine-tuned based on the local tasks and contexts, with their knowledge shared and federated via the network.

On the other hand, generative AI, in particularly LLMs, may also change the way 6G networks are managed. Developing from Software-Defined Networks (SDN) to Intent-Based Networking (IBN) shows the trend to bring network configuration and adaptation towards broader range of stakeholder. The capability of LLMs on language and logic will further improve IBN to support intents described in natural languages, allowing even more stakeholders or end-users to be able to control the network services they need.

5.4.4 Application-Centric Services Network

In the current era of wireless communication technologies, the integration of mmWave/cmWave solutions stands as a promising avenue for enhancing network capabilities. Nevertheless, this advancement is not without its challenges. One primary pain point revolves around the increased heterogeneity of networks, resulting in escalated construction costs [C5-21]. Moreover, the deployment of high-frequency networks poses obstacles to seamless service continuity, thereby limiting the flexibility of network selection based on user preferences or application requirements [C5-22]. These challenges underline the necessity for innovative approaches to network construction and management.

Addressing these challenges and charting a course for future advancements entails filling the research gap in developing effective strategies for multi-operator shared network construction within the mmWave/cmWave domain [C5-23]. While the vision of a simplified, federated, and application-centric network architecture holds promise, realizing seamless collaboration among operators and implementing AI-driven prioritization of user and application needs remains a complex endeavour [C5-24]. Future research endeavours could focus on devising efficient protocols and algorithms facilitating seamless coordination among multiple operators while ensuring optimal resource utilization and service quality. Furthermore, comprehensive studies aimed at enhancing application perception within high-frequency networks are imperative, enabling dynamic network selection based on real-time user demands and preferences [C5-25]. Closing these gaps will be instrumental in realizing the full potential of wireless technologies in the next generation of communication networks.

Short-Term Research Challenges:

- Develop efficient protocols and algorithms for AI-driven prioritization of user and application needs within individual administrative domains.
- Address concerns of fairness and stakeholder needs balancing in the prioritization process to ensure equitable service delivery.

Mid-Term Research Challenges:

- Devise protocols and algorithms for seamless coordination among multiple operators, facilitating multi-operator shared network construction within the mmWave/cmWave domain.
- Implement mechanisms for federated network architectures that simplify collaboration among operators while maintaining autonomy and ensuring fair resource allocation.
- Investigate methods for real-time monitoring and adaptation of network resources to dynamic user demands and application requirements.

Long-Term Research Challenges:

- Establish comprehensive frameworks for application-centric network architectures that dynamically adjust network selection based on real-time user preferences and application requirements.
- Develop AI-driven decision-making models that balance the needs of diverse stakeholders, ensuring fairness and optimal resource utilization across administrative domains.
- Conduct in-depth studies on enhancing application perception within high-frequency networks, enabling intelligent network selection and seamless service continuity across different network environments.

5.4.5 Intent communication network

In the contemporary landscape of human-computer interaction, the seamless expression of intent and efficient semantic transfer and execution of communication between humans, digital humans, and intelligent agents remain pivotal challenges. These pain points underscore the necessity for innovative solutions that facilitate effective information exchange while achieving intent transfer, model transfer and maintaining end to end data sovereignty across diverse modalities. The envisioned future entails the realization of multi-modal communication encompassing various modalities such as text, audio, and visual content, as well as object-based communication, all facilitated by artificial intelligence (AI). Furthermore, the integration of digital IDs to manage digital humans and intelligent agents adds another layer of complexity to this communication ecosystem [C5-26].

However, despite the progress made in human-computer interaction, there exists a research gap in developing robust frameworks for achieving seamless intent transfer and model transfer across different communication modalities. Current state-of-the-art approaches often focus on specific modalities or rely on narrow AI models, limiting their scalability and applicability in real-world scenarios. Future research endeavors could explore novel techniques for integrating multiple modalities seamlessly, leveraging advancements in natural language processing, computer vision, and machine learning [C5-27]. Additionally, there is a need for comprehensive studies aimed at addressing the challenges associated with managing digital IDs in dynamic communication environments, ensuring privacy, security, and interoperability. For example, how to ensure data rights ownership and sovereignty is preserved across all environments. Also, potentially heterogeneous identity providers and issues of which ones are trusted by whom. Bridging these gaps will be crucial for realizing the vision of efficient and intelligent communication between humans, digital humans, and intelligent agents in the digital age.

Short-Term Research Challenges:

- Develop robust frameworks for seamless intent transfer and model transfer across various communication modalities, addressing the today's limitations.

Mid-Term Research Challenges:

- Explore novel methods for efficient semantic transfer and execution of communication between humans, digital humans, and intelligent agents, ensuring scalability and applicability in real-world scenarios.
- Conduct comprehensive studies on managing digital IDs in dynamic communication environments, focusing on privacy, security, and interoperability concerns, including issues related to data rights ownership and sovereignty.

Long-Term Research Challenges:

- Establish standardized protocols and frameworks for multi-modal communication encompassing various modalities such as text, audio, visual content, and object-based communication, facilitated by artificial intelligence.
- Address challenges related to the integration of digital IDs for managing digital humans and intelligent agents, including issues of trust, heterogeneous identity providers, and data sovereignty preservation across all environments.

5.4.6 Intelligent Networks Enable Smart Digital Inclusion

In the current landscape of network functions, the lack of AI capabilities presents a significant challenge, hindering the development of high-value solutions for user experience enhancement and operational efficiency. Furthermore, terminal devices often possess limited computing power and energy consumption capacities, posing obstacles to the realization of smart living environments powered by AI technologies. The increasing adoption of AI raises concerns about privacy, as users grapple with the trade-off between enjoying the conveniences of AI-driven lifestyles and protecting their personal information. To address these pain points, the envisioned future involves the evolution of AI-native networks that prioritize enhancing user experience and streamlining operations while preserving privacy. Digital twins emerge as a key enabler, facilitating closed-loop control mechanisms to overcome the limitations of AI and foster inclusive intelligence. Additionally, cloud and network support are integral in providing AI computing capabilities to terminal devices, while the "Smart and Safe" paradigm ensures user privacy is safeguarded without compromising the benefits derived from AI technologies. Moreover, the network plays a vital role in ensuring the security of personal data, thereby alleviating concerns surrounding data leakage [C5-28][C5-29][C5-30][C5-31].

However, despite the strides made in integrating AI capabilities into network functions and addressing privacy concerns, a notable research gap exists in developing comprehensive frameworks for ensuring privacy-preserving AI-native networks. Current solutions often rely on static privacy protection measures or require significant user intervention, which may impede the seamless integration of AI into network operations. Future research endeavours could focus on designing advanced AI algorithms and privacy-preserving techniques that can autonomously identify and mitigate privacy risks in real-time. Moreover, there is a need for innovative approaches to bridge the gap between AI capabilities and user privacy preferences, enabling personalized AI experiences while ensuring user data confidentiality and integrity. Furthermore, there is a need to enhance the accuracy of the results of AI algorithms, including predictions. This is required so that users can have sufficient confidence in applying the AI results in live systems without risking causing major disruptions in the operation of the systems, such as mobile networks. Future research could investigate how to address this issue, e.g. leveraging digital twins to test and verify AI results before applying them on live networks and systems. By addressing these gaps, future AI-native networks can unlock new possibilities for enhancing user experiences and operational efficiency while safeguarding user privacy in an increasingly connected world [C5-32][C5-33][C5-34].

Short-Term Research Challenges:

- Develop advanced AI algorithms and privacy-preserving techniques capable of autonomously identifying and mitigating privacy risks in real-time within AI-native networks.
- Investigate innovative approaches to bridge the gap between AI capabilities and user privacy preferences, ensuring personalized AI experiences while safeguarding user data confidentiality and integrity.

Mid-Term Research Challenges:

- Design comprehensive frameworks for ensuring privacy-preserving AI-native networks, addressing concerns such as data leakage, loss of data control, and unfairness in AI decision-making regarding control decisions.
- Explore methods to enhance the accuracy and transparency of AI algorithms and predictions, particularly in complex network environments, leveraging digital twins for testing and verification.

Long-Term Research Challenges:

- Address the complexities of network heterogeneity and multiple potential identity providers by establishing robust trust relationships and transparent decision-making mechanisms within AI-native networks.
- Investigate strategies for avoiding harm in AI-native networks, considering different levels of harms and balancing proactive prediction with reactive measures to mitigate risks effectively.

5.4.7 Research challenges

Research Theme	*Artificial Intelligence*		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. AI-empowered control and management plane	Mid-term (finished in 5y)	Autonomic and distributed conflict resolution, correctness enforcement and distributed resource scheduling schemes empowered with the advanced resource allocation to support context awareness and cross-layer design for time sensitive 6G applications allowing several user equipment (e.g. AGVs, connected cars, IoT systems) to have specific resources without any conflict Distributed AI approach to avoid single points of failure, to bring learning closer to the event sources and to be able to harness the available yet scattered compute power.	AI-based cognitive and context-aware resource federation and optimization of edge compute continuum

2. *Cross-layer orchestration modelling*	Mid-term	Semantic composition networking models coupled with FaaS to allow for context (data observability, computational resources, network resources) to be used to best produce resource orchestration mechanisms that can support 6G applications depending on context and situational awareness	Semantic data-network-computational composition models that can scale while providing data sovereignty.
3. Infrastructure for large scale application of Generative AI	Mid-term	Method and tools for on-demand composition and configuration of computation, storage and networking resources along the cloud-edge computing continuum, into high-performance, trustworthy and sustainable infrastructure of the resource-demanding application based on Generative AI techniques.	Pave the way for GAI application with significant industrial and societal impacts.
4. Generative AI for next-level intent-based management of the computing continuum	Mid-term	Integration of Generative AI techniques, such as large language models, to interpret stakeholder intents, and achieve conversational management of the resources in the computing continuum; Generation of human-readable report of system operation and management based on observability records, to increase the trustworthiness of automatically managed infrastructure	Human-oriented cognitive computing continuum.
5. Integration and AI/ML optimization of heterogenous networks integrated with 6G SBA	Short-term (finished in 3y)	Federated learning for heterogeneous architectures and different network resources orchestration	AI enablement of 6G
6. Risk assessment & mitigation of AI use	Mid-term	Understand key harms associated with AI including: data poisoning, bias, unfairness in decision making, lack of transparency. Understand how these harms may be caused, which stakeholders / users they affect and how they affect them.	Prevention of harm from AI
7. AI support of end to end data sovereignty for users	Mid-term	Investigate how AI can be used to provide and protect data sovereignty for users, e.g. anomaly / potential unauthorised usage detection; transparent and auditable AI-powered access control decisions and enforcement in the continuum in line with user preferences and wishes.	Greater data owner control

5.4.8 Recommendations

Theme		Artificial Intelligence					
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**	**Subtheme 5**	**Subtheme 6**	**Subtheme 7**
International Collaboration	x Aligning distributed AI approaches	x Aligning distributed AI approaches	x Aligning distributed AI approaches	x Aligning distributed AI approaches	x Aligning distributed AI approaches	x Aligning distributed AI approaches	x Aligning distributed AI approaches
Open Data	x	x	x	x	x	x	x
Open Source							
Large Trials							
Cross-domain research				x Covering human-centric aspects			

5.5 Edge Cloud Computing Continuum

5.5.1 From cloud-native to continuum-native software

Cloud-native architectures [C5-35], based on technologies such as microservices, containers, and serverless architectures, have flexible and easy-to-scale structures to maximize the exploitation of elastic cloud resources.

Today's cloud computing infrastructure will evolve into a computing continuum providing a programmable computing infrastructure across devices, fog, edge, and central clouds and, by that, offering a significantly broader range of choices how to deploy and run microservice-based applications and network services. The choices will be influenced by the profiles of the available hosting compute locations. These profiles include context information for example about hardware architecture, the capacity, available network connectivity, its geolocation, and more.

One of the main challenges will consist of distributing the components of an application across the computing continuum so that requirements of the application in terms of, for example, performance, security, or usability, will be met in an optimal way. This is also true for IoT applications as illustrated in Figure 5-1. The Computing Continuum groups an enormous amount of heterogeneous resources (e.g., CPU, memory), network and services that appear to the user as a unified environment in a seamless and transparent way. The underlying computing space comprises (any fragment of) IoT-edge-cloud hierarchical infrastructure, abstracted in a way that, if reachable through a network connection, enables a distributed application (i.e., composed of containerized workloads) to run on top, and be managed considering its lifecycle. Such federated resources can be anywhere located and distributed, belonging to multiple administrative and technological domains. Today's microservice architectures might need to be reviewed and further developed to support the distribution across the computing continuum in flexible ways. Software design approaches that allow to defer the decision about the software componentisation as long as possible could be one path to achieve the flexibility. Abstraction mechanisms that allow for loose coupling between applications and infrastructure and support dynamic placement optimization could be another way. In any case, the computing continuum will impact the architecture, the interfaces, and the disaggregation of future network functions.

Specific research challenges to be addressed in this context include:

- Research on addressing the end-to-end software capabilities (continuum) of technologies across sensors, connectivity, gateways, edge processing, robotics, platforms, applications, AI, and analytics, including underlying technologies like optical, wireless (cellular and non-cellular) and satellite communications.
- Research on supporting the continuum of intelligence (e.g., ability to acquire and apply knowledge using context awareness) and other edge capabilities, e.g., computing, connectivity, processing, sensing, privacy, security, see Figure 5-1.
- Research on supporting the continuum of edge applications within and across vertical sectors and seamless integration.

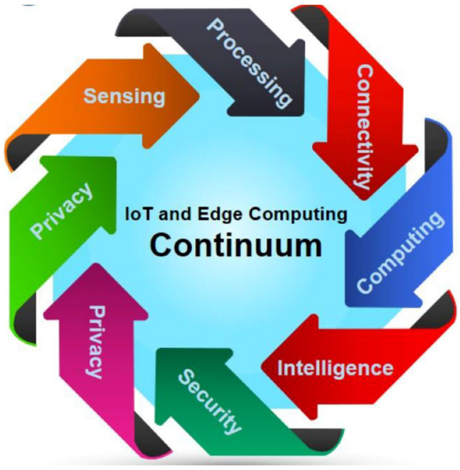


Figure 5-1 Continuum-native software of IoT applications [C5-36]

5.5.2 Task off-loading to the edge

Certain applications benefit from leveraging the vast computational resources of the cloud. These applications involve intensive data processing, complex analytics, or require access to sophisticated Machine Learning (ML) algorithms. By harnessing the power of the cloud, organizations can analyse massive volumes of data, derive meaningful insights, and execute resource-intensive tasks efficiently. On the other hand, some applications demand low latency and real-time responsiveness. These applications operate on the edge, closer to the data sources, to minimise delays and enable near-instantaneous decision-making. Edge computing ensures faster response times, reduced network congestion, and improved overall performance for latency-sensitive applications. For instance, location-based applications or applications for monitoring the environment might comprise tasks with large computational requirements such as pattern recognition. Sensors and field computation units involved in this kind of applications often do not have the compute resources to carry out computationally intensive and delay-sensitive tasks. Edge computing offers the capability to offload those workloads from the end device to an edge node to reduce latency and bandwidth bottlenecks.

These computational units or edge nodes have been specified by 3GPP to be directly integrated in the 5G architecture, in order to improve reliability and round-trip latency. This integration is defined in Technical Specification (TS) 23.501 and is expected to be developed towards the next generation of cellular communications (6G). Examples of industrial applications making use of this architecture are described in [C5-37] and [C5-38] showing that satisfactory results in terms of latency can be achieved.

Each one of the computing layers that enable edge (e.g. IoT) applications, namely IoT Devices, Edge computing, and Cloud computing, exhibit distinct characteristics concerning data handling and sovereignty. IoT devices are responsible for data collection and interaction with the physical environment. They possess limited resources and focus on real-time data processing, often transmitting only essential information to conserve bandwidth. Edge computing, located between IoT devices and the cloud, aggregates, and pre-processes data locally, reducing latency and ensuring faster response times. The cloud, on the other hand, provides extensive computational power and storage capacity, enabling large-scale data analysis and long-term storage. Each layer plays a vital role in the overall architecture, addressing specific requirements and challenges. A growing number of new applications necessitate a dynamic approach that intelligently utilises the resources of all computing layers (IoT, edge, and cloud) to achieve optimal performance while adhering to data sovereignty and privacy restrictions. These applications will leverage the strengths of each layer while optimizing resource utilisation. They will employ intelligent data routing and processing mechanisms to determine where and how to handle data most effectively. By intelligently distributing tasks across IoT devices, edge servers, and the cloud, these applications strike for a balance between real-time responsiveness, efficient data processing, and compliance with data regulations. This dynamic utilization of resources allows organizations to achieve high-performance outcomes, while respecting data sovereignty, privacy, and compliance requirements. In order to ensure QoS using edge computing, the use of offloading computation algorithms must be integrated into Beyond 5G and 6G communications. This integration will provide better support for the use of edge computing in mobile scenarios such as drone or autonomous vehicle-based use cases. In [C5-39], several algorithms are studied to manage the offloading in terms of computation cost and energy savings. Research challenges are about application software architectures that support the offloading of subtasks and the interworking with AI based mechanisms that help to decide about when and where a task should be offloaded and considering the impact that the offloading will have on the overall end-to-end QoS.

5.5.3 Integrated lifecycle management: DevSecOps and CI/CD pipelines

DevOps culture and the use of a continuous integration and delivery pipeline [C5-40] have become common practice in the software lifecycle management, also in the telecommunication industry.

While regular IT CI/CD pipelines and workloads usually operate in a homogeneous and fixed cloud environment managed by a single organisation, telco CI/CD pipelines are split across different organisations – ie. multiple telecom software vendors and communication service providers – and operate in more complex and heterogeneous environments. In telco CI/CD pipelines development, building, and testing are in the vendor part of the pipeline, and deployment and operation in the operator part. All involved parties need to align on the delivery and feedback method across the organizations, covering for example the intervals of delivery by multiple vendors, the timing of integration of the software from multiple vendors, as well as the testing of the integrated software system before going into production. Further challenges include the adaptation of the software to the various hybrid cloud environments of telecom networks and performing software updates without violating stringent SLAs. Effective automation to enable service creation and management on large scale of multi-, hybrid- and edge-cloud systems is needed [C5-41].

In the future, CI/CD pipelines have to deliver software into the more heterogenous and dynamic infrastructure of a computing continuum and have to provide effective support for the entire lifecycle of complex service chains. CI/CD pipelines will need to deliver not only pure software but also AI and ML components, as well as APIs. And all that in a secure way. In the context of service development and deployment over the Edge-Cloud Continuum, it becomes necessary to go one step beyond DevOps, developing the concept of DevSecOps, which means thinking about application and infrastructure security from the beginning and embedding DevOps with security controls providing continuous security assurance. DevSecOps is a natural extension of DevOps to include security-by-design and continuous security testing by automating some security controls in the DevOps workflow. Besides, by adding tests to the DevSecOps methodology to ensure that privacy can be analysed in all phases of the process, the DevPrivSecOps methodology emerges.

As a result, research is needed on the following aspects:

- Adequate abstraction mechanisms that allow to design software in a cloud infrastructure agnostic way.
- AI-based enhancements of Infrastructure as Code techniques for the automatic deployment of software components so that new deployment arrangements can be learned from historical and simulated deployment experiences.

- Automatic generation of test cases and the simulation of test environments making use for example of digital twin technology or chaos engineering.
- Extending the CI/CD tools and techniques to securely allow for bundling software, AI components, APIs, and security into one development, delivery, and deployment process.
- Developing standards to enable secure interoperability in multi-vendor CI/CD scenarios.

5.5.4 Integration of DevOps with business processes

DevOps aligns and automates the collaboration between development and operations teams. As a result, the software development process has become tightly interwoven with the operation of the software and, by that, with the business processes implemented by this software. This accelerates the introduction of new features, the adaption of business processes to changing user and market requirements as well as creates additional opportunities for the network automation to achieve efficiency gains. Network management and orchestration, network slicing, security workflows, or the exposure of network capabilities via APIs are all examples of network functionality and processes that can benefit from a close integration with a CI/CD pipeline.

Feedback loops being integral part of a CI/CD pipeline play an important role in this kind of network automation. Feedback loops provide a constant flow of information that allow to gain insights into the health of the deployed software and related business processes and might trigger automated or semi-automated responses to identified issues. There are two types of feedback loops: development loops along the entire CI/CD pipeline with humans involved, and adaptation loops that monitor system changes and make automated adaptation in real-time. The development loops are running at a slower pace, continuously tunning the adaptation loops. More feedback loops and layers of feedback loops may be added depending on the business processes and applications running on top of the network and using for example network slices.

Research challenges in this context include:

- The investigation of use cases and business processes that can benefit from a close integration with a CI/CD pipeline focusing not only on the technical aspects, but also addressing organizational questions such as how to setup cross-organizational teams and collaboration to make the CI/CD pipeline work.
- The hierarchy and interworking of the feedback loops as well as their integration into network management and orchestration processes.
- The secure exchange of data via the feedback loops which might span across multiple administrative domains and organizations.

5.5.5 Time guarantees on virtualization and containerization

In an edge without time guarantees, virtualization and containerization have no constraints, and share the host resources without limitations. When addressing time sensitive/time engineering applications, such differentiation is needed, and applications should run in virtualized environments capable of having different priorities between services and with pre-emption techniques. Such techniques will enable a deterministic network and edge, capable of processing critical requests within specific time windows (on-time processing). Recent research [C5-42] has been conducted to understand the extent to which time-sensitive payloads can be virtualized. These mostly encompass containerized execution environments resorting to specific Linux kernel fine-tuning or co-kernels to improve determinism. Challenges require an experimental capability that continuously monitors the performance of real-time services using their time utility that potentially degrades with the tardiness of deadline misses[26]. Lastly, the devices deployed in this environment support distinct hardware architectures, in which the orchestrator should be aware. The integration of fine-tuned execution environments with suitable lightweight edge orchestration frameworks, are envisioned as a possible way to tackle the challenge of building a time awareness orchestration framework, suited for constrained edge computing devices. Specific orchestration scheduler is also needed to support the time-sensitive applications.

5.5.6 High-End Experience Network with Converged Communication and Computing

In the realm of mobile computing, the demand for high-end applications such as extended reality (XR) and artificial intelligence (AI) has surged, yet challenges persist in computing performance, power consumption, and battery capacity on mobile terminals. While research efforts on cloud/multi-access edge computing (MEC)-assisted terminals have commenced, there remains a lack of consensus on a distributed native solution to deliver high-end experiences seamlessly. Recent advancements have explored the potential of leveraging high-data rate, low-latency networks to distribute computation across nodes, thereby realizing the vision of ubiquitous computing. Converged computing and communication architectures have emerged as promising approaches to optimize service experience, yet achieving a truly distributed native solution for high-end applications remains an ongoing challenge [C5-43][C5-44][C5-45].

Despite the progress made in exploring cloud/MEC-assisted terminals and converged computing and communication architectures, a notable research gap exists in developing robust frameworks for distributed native applications that can leverage network-based distributed computing effectively. Current state-of-the-art solutions often rely on centralized or partially distributed approaches, limiting scalability and flexibility in dynamically adjusting data paths and computation to ensure uninterrupted service as users move. Future research endeavors could focus on designing efficient algorithms and protocols for distributed native applications that can seamlessly adapt to varying network conditions and user requirements, ultimately guaranteeing a high-end experience across mobile terminals. Additionally, there is a need for comprehensive studies aimed at integrating harmonized communication and sensing functionalities within converged computing and communication architectures, enabling enhanced service experiences in diverse application scenarios [C5-46][C5-47][C5-48].

Short-Term Research Challenges:

- Develop efficient algorithms and protocols for distributed native applications in mobile computing, ensuring seamless adaptation to varying network conditions and user requirements.

Mid-Term Research Challenges:

- Investigate methods to optimize computing performance, power consumption, and battery capacity on mobile terminals, particularly for high-end applications like extended reality (XR) and artificial intelligence (AI).

Long-Term Research Challenges:

- Develop robust frameworks for distributed native applications that can effectively leverage network-based distributed computing, enabling high-end experiences across mobile terminals.
- Integrate harmonized communication and sensing functionalities within converged computing and communication architectures, facilitating enhanced service experiences in diverse application scenarios and optimizing the convergence of computing and communication resources.

[26] Actions on top of libraries such as timeutiles are being explored in projects like aerOS to realize these features.

5.5.7 Network with Device-Edge-Cloud Unified Computing Technology Stack

The current landscape of edge computing faces significant challenges due to the incompatibility between edge and public cloud technology stacks, which impedes the seamless integration of multi-access edge computing (MEC) with each other and with various cloud solutions. This disparity complicates deployment and upgrades, hindering the realization of efficient edge computing environments. Additionally, the absence of an integration-free solution for edge applications necessitates pre-integration efforts, further complicating the development and deployment processes. Moreover, existing protocols lack the inclusion of service programs, limiting transmission to service data only [C5-49][C5-50][C5-51].

The envisioned future of edge computing involves the establishment of a unified technology stack, facilitating a single service technology stack across devices, edges, and clouds. This unified approach entails splitting applications into computing services, which run on distributed nodes, fostering scalability and efficiency. An integral component of this vision is the development of integration-free applications that leverage lightweight, high-performance, and plug-and-play functionalities without requiring pre-integration efforts. Furthermore, there is a pressing need for the development of an application-associated protocol that enables the deployment of applications naturally with data path orchestration, thereby enhancing flexibility and scalability in edge computing environments. Emphasizing information transfer in distributed applications rather than data transformation will be crucial for realizing the full potential of edge computing in future distributed systems [C5-52][C5-53].

Short-Term Research Challenges:

- Investigate methods to streamline deployment and upgrades in edge computing environments by addressing the challenges arising from the incompatibility between edge and public cloud technology stacks, including aspects as a cross-domain continuum ontology to allow a proper, standardized federaton of resources.

Mid-Term Research Challenges:

- Explore strategies for establishing a unified technology stack in edge computing, facilitating a single service technology stack across devices, edges, and clouds to enhance efficiency and scalability.
- Develop integration-free applications for edge computing that leverage lightweight, high-performance, plug-and-play functionalities, reducing the need for pre-integration efforts and simplifying the development and deployment processes.
- A publicly available product composed of strategic commons that will cover: network flexibility, decentralisation identification and management, trust, orchestration and resources incorporation.

Long-Term Research Challenges:

- Design an application-associated protocol for edge computing that enables the deployment of applications naturally with data path orchestration, enhancing flexibility and scalability in edge computing environments.
- Shift focus towards information transfer in distributed applications rather than data transformation, emphasizing the importance of efficient communication protocols for realizing the full potential of edge computing in future distributed systems.

5.5.8 Ultra-distributed Intelligent to Business Network

In the domain of business network infrastructure, the current state-of-the-art faces significant challenges regarding poor network function agility. The customization of network functions often hinders scalability, while the complexity of campus networking presents adaptation challenges. Moreover, the configuration process is cumbersome, leading to slow service provisioning, and maintenance tasks are intricate, making fault locating a time-consuming endeavor. To address these pain points, the envisioned future entails a paradigm shift towards customizable devices that integrate infrastructure only. Intelligent orchestration plays a pivotal role in understanding customer intentions, defining on-site functions, and timely downloading them to the network. The implementation of self-configuration mechanisms facilitates automatic configuration and service provisioning, while self-maintenance capabilities enable quick service recovery and maintenance-free operation. This vision emphasizes the integration of full network functions, including core network functions and customized campus transmission, switching, and security, tailored to the specific requirements of each business network [C5-54][C5-55][C5-56].

However, despite the advancements made in improving network function agility and customization, a notable research gap exists in developing comprehensive frameworks for intelligent orchestration and self-configuration mechanisms. Current solutions often rely on static configurations or rudimentary automation, lacking the adaptability and intelligence needed to dynamically adjust to evolving business requirements and network conditions. Future research endeavors could focus on designing advanced machine learning algorithms and AI-driven approaches that can effectively understand and anticipate customer intentions, enabling proactive network customization and optimization. Additionally, there is a need for innovative techniques for self-maintenance that go beyond reactive fault recovery, ensuring continuous network reliability and performance optimization. Bridging these gaps will be crucial for realizing the vision of agile and customizable business network infrastructures that can seamlessly adapt to the evolving demands of modern enterprises [C5-57][C5-58][C5-59].

Short-Term Research Challenges:

- Develop solutions for improving network function agility by addressing customization hindrances, such as scalability issues, in business network infrastructures.
- Investigate methods to streamline the configuration process and enhance service provisioning speed, reducing the complexity of campus networking and making fault locating more efficient.

Mid-Term Research Challenges:

- Design comprehensive frameworks for intelligent orchestration and self-configuration mechanisms in business network infrastructures, focusing on adaptability and responsiveness to evolving business requirements and network conditions.
- Explore advanced machine learning algorithms and AI-driven approaches to understand and anticipate customer intentions, enabling proactive network customization and optimization.

Long-Term Research Challenges:

- Develop innovative techniques for self-maintenance in business network infrastructures, ensuring continuous network reliability and performance optimization beyond reactive fault recovery.
- Emphasize the integration of full network functions, including core network functions and customized campus transmission, switching, and security, tailored to the specific requirements of each business network.

5.5.9 Edge IoT immersive platforms in the SNS context

One of the future research directions is in the area of converging key technologies like: Immersive technologies, Spatial Computing, artificial intelligence (AI), digital twins (DT), edge IoT and 6G.

Immersive technologies can encompass several systems and services including augmented reality (AR), virtual reality (VR) and mixed reality (MR). These immersive technologies allow end users to interact with digital information and IoT devices more naturally and intuitively.

Spatial computing is encompassing the processes and tools for capturing, processing, and interacting with 3D data; In particular, these technologies enable human-computer interaction that simulates interactions in real-world physical environments rather than being limited to screens and machines.

In this context, 6G, edge IoT digital platforms and marketplaces will be important in providing open, flexible IoT industrial immersive solutions [C5-60]. This requires open, interoperable solutions that allow seamless, real-time, concurrent collaboration, open APIs, compatible data formats, and protocols.

Research and standardisation actions are needed in order to develop the architecture and functionality of a distributed immersive software platform that is able to integrate the 6G edge IoT industrial immersive technologies using a common language that facilitates users' connecting their immersive/digital triplets, avatars, and DTs in the digital, virtual, and cyber worlds; By using this distributed immersive software platform, industries can use interoperable, open systems and standard platforms to connect their immersive/digital triplets, avatars, and DTs with the counterparts of their partners and suppliers, creating larger ecosystems that produce more profound insights and plug-and-play solutions.

Some of the key challenges listed in [C5-60] that apply as well to the development of the distributed immersive software platform, are:

- Interoperability and Standardisation Challenge: Different devices and systems often use proprietary protocols and data formats, APIs, making seamless communication and data exchange difficult. This lack of interoperability can hinder the integration process and limit the potential benefits of a fully connected ecosystem. It is important to focus on the development of open and standardised protocols and data formats, interfaces, and APIs for the distributed immersive software platform.
- Scalability and Flexibility Challenge: As the number of connected devices grows, the system (i.e., distributed immersive software platform) must scale without significant performance losses. Additionally, the system needs the flexibility to integrate new technologies and adapt to changing operational requirements.
- Real-time Data Processing and Decision-Making Challenge: Industrial immersive systems often require real-time or near-real-time data processing and decision-making capabilities. Achieving this level of performance, especially in complex environments with high data volumes, is challenging for the development of the distributed immersive software platform.
- Trust, Data Privacy and Security Challenge: Integrating 6G, IoT, edge computing, and AI increases the amount of data collected, processed, and stored, raising significant data privacy and security concerns. Protecting this data against unauthorised access and ensuring compliance with regulations such as GDPR becomes more complex in a highly interconnected environment.
- AI Model Accuracy and Trustworthiness Challenge: Developing robust and bias-free AI models is difficult, especially when dealing with complex, real-world data. Ensuring these models can operate effectively ethically, and safely (without causing any harms to human) in an industrial context is a significant challenge.
- User Acceptance, Adoption and Training Challenge: The success of integrated systems also depends on user acceptance and the availability of training. Users need to trust the technology, the use of the distributed immersive software platform, and understand how to interact effectively.

All these challenges can be considered as long-term research and standardisation challenges, where RIA type projects are appropriate.

5.5.10 Network compute fabric supporting passive IoT

Today's computing continuums are advancing away from central data centres to the very edge of the network. It enables time-critical and massive-data hungry use cases. As the use cases continue to grow in complexity and criticality, new forms of integrated computing fabric will inevitably be needed to complement today's edge solutions and bring the rest of the computing continuum closer together.

Future deterministic networks that ensure bounded latency, jitter, and packet-loss and out-of-order packet delivery are complex where next-hop forwarding requires service protection along an end-to-end resource reserved explicit routes that can be achieved basically through packet replication, duplicate elimination and/or network coding. Some topics have to be considered:

- Sequencing Information expressed in terms of sequence number or time stamp.
- AI/ML-powered software-defined regenerative payload creation
- Enhancement on Packet Replication Function (PRF), Packet Elimination Function (PEF) and Packet Ordering Function (POF) in addition to Network Coding Function (NCF)

In addition, a large-scale deployment of sensors often requires cost-effective solutions fuelling the introduction of passive IoT devices. These devices are different that they do not possess batteries. A minimal impact that such a device has on the mobile communication system is that the network needs to identify those UEs that are of passive IoT type. Following issues are important to consider:

- New UE type and such information has to be included as part of UE Subscription data
- Geographical location a network can communicate with a given passive IoT device depending on renewable source availability.
- New Location Management Function (LMF) feature
- Making transparent the network compute fabric to users/developers who want to use the heterogeneous edge resource embedded in the network to better support advanced analytics and coordination mechanism.

5.5.11 Research challenges

Research Theme	Edge Cloud Computing Continuum		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Infrastructure-aware microservices model (SNS-native software)	*Short-term (finished in 3y)*	Extended microservices model and container technologies with explicit requirement or preference on the context of the hosting resource	Computing continuum alignment
2. Continuum-native network functions and applications (SNS-native software)	*Short-term (finished in 3y)*	Extending the current container and orchestration platforms to cover network functions and services	Computing continuum alignment
3. Exploring offloading of computationally intensive and delay-sensitive workloads	*Short-term (finished in 3y)*	SW architectures and mechanism for task offloading	Alignment with computing continuum
4. Auto testing on simulated SNS (DevOps and lifecycle management)	*Mid-term (finished in 5y)*	Digital twins of the whole SNS environment and the surrounding physical context to automate the execution of test cases	Computing continuum alignment
5. Abstraction and AI-powered deployment	*Short-term (finished in 3y)*	Cloud agnostic software and automated deployment models for the computing continuum	Computing continuum alignment

models (DevOps and lifecycle management)			
6. Common CI/CD engine for SW, AI, APIs, security, etc	*Mid-term (finished in 5y)*	Concepts and tools for extended CI/CD engines	Computing continuum alignment
7. CI/CD automation use cases incl. organizational issues (Integration with business process)	*Mid-term (finished in 5y)*	CI/CD enabled network automation and organizational optimizations	Computing continuum alignment
8. Hierarchical DevOps loops (Integration with business process)	*Mid-term (finished in 5y)*	Novel DevOps practices where many different processes with different time scales and focuses are coordinated	Computing continuum alignment
9. Time guarantees for containers	*Short-term (finished in 3y)*	Time aware orchestration frameworks for time-sensitive applications	Computing continuum alignment
10. High-End Experience Network with Converged Communication and Computing	*Long-term*	Frameworks for applications to make efficient use of distributed compute and sensing in converged communication and compute infrastructures	Computing continuum alignment
11. Network with Device-Edge-Cloud Unified Computing Technology Stack	*Mid-term*	Unified technology stack that minimizes integration needs of applications	Computing continuum alignment
12. Ultra-distributed Intelligent to Business Network	*Long-term*	Frameworks that enable proactive network customization and optimization	Computing continuum alignment
13. Edge IoT immersive platforms	*Long-term*	SW platforms supporting open and interoperable edge IoT industrial immersive solutions	Computing continuum alignment
14. Abstraction mechanisms for the network compute fabric to support passive IoT	*Mid-term (finished in 5y)*	Programmatic framework for accessing network and compute resources for the use case “passive IoT”	Enabling energy efficiency and sustainability
15. Risk analysis in continuum	*Start mid-term but expected to be a long-term effort due to the challenge’s complexity*	The continuum is highly complex, dynamic and heterogeneous. Threats can propagate from actors, services or data to others. Given the highly complex, composable and dynamic nature of the continuum, assessment of vulnerabilities, threats, risks is needed as well as mechanisms that can be employed at or before deployment time to mitigate them. The risks may be predominantly in the cybersecurity domain, but may cause harms in related domains, such as loss of privacy or damaged reputation.	Protection from cybersecurity vulnerabilities and related harms that result in the highly dynamic and heterogeneous continuum
16. Security balanced with cost and utility		Mitigating risks is highly important, but the measures put in place should not defeat the purpose of the services or data used, otherwise they will be useless. Understanding is needed on how to balance the needs of cybersecurity in the continuum against the utility of the services, components and data within it, with the aim to be able to strike an acceptable balance between protection and usefulness, at reasonable cost.	Provide adequate but not excessive security

5.5.12 Recommendations

Europe must also find a way to overcome some operator's reluctance to offer such services until customer demand is overwhelming: if the service isn't offered before users ask for it, these users will not even try to imagine how to use telecommunication services to fulfill their requirements, but find other ways - thereby preventing commercial demand from ever materializing, and retaining the center of value on over-the-top providers. In this context, international collaboration of common interfaces and standards are essential.

Research Theme	Edge Cloud Computing Continuum						
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**	**Subtheme 5**	**Subtheme 6**	**Subtheme 7**
International Collaboration	X investigate the innovative software paradigm	X investigate the innovative software paradigm	X Investigating common interfaces and standards for offloading	X Digital twins for SW testing as part of the CI/CD pipeline	X Agreement on abstraction models		X Extended CI/CD engines
Open Data / Open Source	X Open Source implementation of microservices and container extensions	X Open Source implementation of microservices and container extensions			X Implementations of the models	X Common CI/CD engine	X Open source support for extended CI/CD engines
Large Trials			X PoCs to test task offloading approaches		**X** Large trials to test and promote the adoption of DevOps in SNS		
Cross-domain research							

Research Theme	Edge Cloud Computing Continuum						
Action	**Subtheme 8**	**Subtheme 9**	**Subtheme 10**	**Subtheme 11**	**Subtheme 12**	**Subtheme 13**	**Subtheme 14**
International Collaboration		X Common standardized solutions	**X** Common standardized solutions	**X** Unified technology stack	X Common customization and optimization framework	**X** Common standardized solutions	**X** Common standardized solutions
Open Data / Open Source		X Support for time-sensitive containers	X Open Source implementation	X Open Source implementation	X Open Source implementation		
Large Trials	X PoC on automation use cases		X PoC of framework	X PoC of framework	X PoC of framework		
Cross-domain research	**X** Joint research with other domains to						

	investigate the integration with business processes						

Research Theme	Edge Cloud Computing Continuum						
Action	**Subtheme 15**	**Subtheme 16**					
International Collaboration	X Security measures	X Security measures					
Open Data / Open Source							
Large Trials							
Cross-domain research							

5.6 Digital twins in the SNS context

5.6.1 Software engineering of telco digital twins

Digital twins are dynamic virtual representations of entities such as assets, persons, and processes. Digital twins are built by developing models for interpreting data at different speeds to be used for creating a one-to-one association with their real-world twins. Digital twins do not exist independently from an enabling software platform. It is thus of paramount importance to investigate software engineering practices and tools for simplifying the development of telco-world twins.

It is also worth to note that, in a broad view there are two main categories of digital twins, offline and real time. On one hand, for the real-time variation, it mimics the behaviour of existing physical products almost simultaneously. From 6G system perspective, ultra-low latency, high availability of connectivity and computational edge resources, high reliability, trustworthiness, and high security are some key requirements to be considered. On the other hand, digital twins that are more computationally intensive and thus cannot keep up in real-time, can also be run after the sensor data has been captured in an offline manner over a public/private cloud. In this case, there might not be extremely hard requirements on 6G systems. What mentioned so far are from the perspective of 6G systems to support digital twin of vertical sectors, e.g. digital twin of an engine or a road. However, the network itself as a physical entity should have its own digital twin too. Defining and realization of a proper and harmonized model with the current available solutions in the vertical sectors for the network gears, e.g. UE, RAN and Core, is an important research direction to follow.

At the same time, the twin's software development life cycle should be restructured for managing requirements, including sustainability requirements (cf., for example, Twin Transition [C5-61]), models, data & metadata, with special attention to the validation and verification procedures and improvement cycles comprising operational feedback. Moreover, new, and more sophisticated software engineering techniques are aimed at taming the complexities of higher-level twins. Considering the very wide range of competencies needed for developing new twins and for making them a more affordable investment for a larger number of companies, more effort on "composable" digital twins' development tools and techniques is needed. Twins of telco networks may begin at the network design stage, facilitate deployment of network nodes, optimise the deployment of edge nodes, and provide integration of information and decision support to enable highly automated and remote operation while ensuring high efficiency, safety, and environmental awareness.

Such digital twins are complex since they must integrate a wide range of information, algorithms, and models for processing real-time operational data as well as a large amount of business-relevant data. They might include simulation and predictive capabilities to support improved operational decision making, possibly in distributed and decentralised contexts, crossing ownership domains. Smart Network Twins could be engineered not only for optimisation and planning purposes but also, for example, for reacting to cyber-attacks, providing an immediate forecast of the consequences, and allowing efficient mitigation of them.

5.6.2 Research challenges

Research Theme	Digital twins		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Managing the life cycle of digital twins	Mid-term (fiished in 5y)	SW engineering approaches for telecom digital twins	Optimisation of network planning and operation
2. Composition and interworking of digital twins	Mid-term (finished in 5y)	Standard interfaces for digital twins	Optimisation of network planning and operation
3. Use of digital twins for security by design and in operation	Mid-term	Investigation into how digital twins can be employed at design-time to support security by design, and in operation (via observation, monitoring, anomaly detection etc) to support security in operation.	Enhance system security via use of digital twins

5.6.3 Recommendations

Research Theme	Digital twins		
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**
International Collaboration		**X** To develop standardized interfaces for digital twins	X Digital twins used for security
Open Data / Open Source	X Open source implementations of digital twin frameworks		
Large Trials		X PoCs for testing and demonstrating the interworking of digital twins	
Cross-domain research	X SW engineering for managing the life cycle of digital twins		

5.7 Data frameworks

5G, and even more 6G society will result in huge amounts of data, and the analysis and application of these massive amounts of data promote the informatization and digitization of society. The focus on data requires, then, important considerations, not only because of the shift toward a data-centric, AI-based O&M that is essential for telco operators per se, but it is quite evident that data will be more and more essential for the whole society.

Data needs to be collected, cleaned, processed and stored in the end-to-end network. Even though it is possible to process data centrally, the integration of different networks in 6G scenarios leads to a more distributed data processing logic. More and more data will be used for algorithm training purposes, and this will lead to the stringent necessity to engineer software AND data operational cycles.

In addition, data frameworks need to be compliant with regulations. For example, the EU data act defines EU-wide and cross-sector harmonized rules on fair access to and use of data. IoT devices and sensors and associated services must allow users to access and use data generated by these devices which means that also a network-based sensing will fall under this regulation.

5.7.1 Research challenges

Research Theme	Data		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Aligned network software and data lifecycle	Short-term (finished in 3y)	Lifecycle management approaches keeping software and data consistent across networks	Increasing dependency on AI and data
2. End-to-end data sovereignty across complex lifecycles	Short term	Data has complex lifecycles – it may be aggregated, split, altered, copied by many different actors. Understanding how to protect provenance, ownership and enforce appropriate rights on data actors of these complex lifecycles is needed	Protecting rights of data owners in complex data lifecycles

5.7.2 Recommendations

Research Theme	Data	
Action	**Subtheme 1**	**Subtheme 2**
International Collaboration	X Compliance with regulation	X Data sovereignty across lifecycle
Open Data / Open Source	X Keeping data aligned with network software	
Large Trials	**X** Keeping data aligned with network software at scale	
Cross-domain research		

5.8 Software Engineering of complex and self-adaptive systems

5.8.1 Managing the software complexity of a system of systems

As described in [C5-62] smart networks will be a highly distributed and decentralised system of systems comprising countless heterogeneous physical and virtual entities and supporting a broad range of services and applications with divergent requirements. All mapping from services to network slices and then to virtual resources will be completely elastic and flexible. There will be no direct relationship between the lifecycle of a service and the lifecycle of the assigned virtual resources. The countless entities need to be managed throughout their specific lifecycle and to have their parameters configured and adapted to a dynamically changing environment. The services have to be of high quality and provided in flexible ways to meet users' demanding expectations, whilst consuming network resources as efficiently as possible to minimize cost. Managing the resulting system complexity will become increasingly challenging and will require new operational concepts based on sophisticated self-adaptation models and relying heavily on AI algorithms.

AI is used to evaluate the current resource status and current service status, and more importantly, to predict any future problems. AI is used to decide how to react to current or predicted certain level of future status changes. AI is also used to optimise the delivered service to avoid unprofitable waste of resources. Although closed loop automation, with limited human intervention brings all the promise of a fully automated business, one of the main ongoing debates is to consider a role for humans in the loop. The issue arises from the concerns about trusting a fully automated system. Humans in the loop can make or break AI success and have to be addressed in a human-centric way.

Developing, debugging and testing such complex cognitive systems will be challenging. The conflicting requirements of extreme flexibility, dynamic adaptation and optimized resource utilization are hard to reconcile in a distributed system of autonomous AI-based subsystems, and the resulting overall system behaviour might be unexpected. To avoid unexpected and even hazardous effects, what is needed is predictable governance for self-adapting AI-based software systems including the traceability of AI.

This implies that AI/ML algorithms have to be made aware of changes in the surrounding system that impact the learned model, and would require re-training. It also involves explainable AI, for transparency of how an AI-based system works and responsibility for the resulting output. Access to high-quality, trustable and sufficient training data needs to be assured, so that no undetected biases find their way into AI systems.

Techniques such as deep neural networks may achieve a high performance in terms of speed and accuracy, but they are generally seen as black-box models due to lack of explanation associated with their outputs. In the context of SLA/Service Level Specification (SLS) management and network orchestration, the features of Explainable AI (XAI) become very desirable in terms of transparency of the model with respect to the intended outcome. Current research contemplates two main approaches aiming at achieving explainable AI models: the first one involves designing models that are inherently interpretable (which means they can easily explain how specific decisions can influence achieving specific objectives); a second approach is to complement the AI black box with the aid of external complementary models. Different levels of model transparency (or explainability) may apply to different categories of AI applications. As a first level, models can explain how conclusions are reached by the AI system in order to improve future decision making, decision understanding and trust from human users and operators. As a second level the external model can allow inspection and traceability of actions undertaken by the AI systems. Traceability will enable humans to get into AI process loops. Depending on the use cases and market/feasibility priorities, pros and cons of various characteristics of human plus AI has to be considered and investigated, i.e. assisted intelligence (improving decisions and actions of people); augmented intelligence (enabling humans to do more than before); autonomous intelligence (adaptation of various situations, acting without human assistance).

5.8.2 Engineering software intensive systems

[C5-62] identifies the need to expand the scope of software engineering to encompass the full range of possible deployments from embedded devices to the cloud, and the full lifecycle of the software including automated operation of self-adapting software intensive systems. This unification of operational and business aspects is not supported by adequate software tools today. Software engineering methods such as UML modelling cannot handle situations where interconnected services are not known in advance, and cannot easily model consequences that may have a legal or ethical dimension. Some aspects previously considered the domain of programmers such as the composition of resources and services will also need to be handled autonomously at run-time. Therefore, we need new engineering approaches which can be applied over the lifecycle of software services and data (including design, implementation and testing); respond to agile changes in self-adapting systems; handle ethical and legal aspects; and support purposeful sharing.

The use of AI presents both challenges and opportunities for software engineers. As previously noted, operating software on a large-scale, distributed, heterogeneous and smart infrastructure requires new approaches such as AI. Can AI also be used to support the evolution of DevOps methods for software design and development? How will those methods enable the design and development of smart components that use AI, and ensure that those components meet ethical, legal, social and economic requirements as they evolve in the presence of new input data? The approaches developed must be able to handle requirements and constraints not only in the use of AI, but also of other novel technologies such as specialised (including quantum) processing devices, and novel modes of human-computer interaction.

5.8.3 Software engineering for the integration of AI into networks

Regarding the introduction of Artificial Intelligence (AI) over wireless networks, software engineering is necessary to provide easy and quick access to AI models along networks by using not only cloud computing but also on-site edge computing. This scope provides a new paradigm to introduce software mechanisms that facilitate the integration of AI in 6G networks. An example is provided in [C5-63] where split-AI is defined to provide access to neural networks from different 6G radio network nodes following a distributed architecture.

The introduction of these mechanisms can boost current scenarios with process offloading requirements, such as mobile robot fleets. In this application, autonomous mobile agent systems must execute different tasks to perform proper navigation, such as object recognition and trajectory calculation. For example, the possibility to execute innovative AI models from close positions such as radio nodes may improve time inference using better computing devices than on-board robot computers.

Furthermore, splitting AI models and offloading the process allows reductions in terms computing cost, what is a key aspect in mobile robot scenarios to avoid obstacles, but also in vehicular use cases. In these applications, AI processing can be offloaded from the vehicle to other vehicles sharing parts of the model, or even to the road infrastructure. The capacity to reduce the latency to apply AI on vehicles can provide more robust obstacle detection and therefore, improved autonomous driving. However, these applications provide also some concerns to be considered, mainly related to the data treatment of the environment (pedestrians, neighbor vehicles, infrastructure…).

5.8.4 Research challenges

Research Theme	Software Engineering of complex and self-adaptive systems		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Testing of self-adaptive systems	*Mid-term (finished in 5y)*	Testing approaches and frameworks for self-adaptive systems	- System resilience
2. Predictive governance for self-adapting AI-based software systems	*Mid-term (finished in 5y)*	Governance framework for monitoring behaviour of AI-based systems	- System resilience
3. New SW engineering approaches	*Mid-term (finished in 5y)*	Overall SW architecture and design approaches for complex systems	- Managing the complexity
4. AI-assisted software design	*Mid-term (finished in 5y)*	AI-based approaches and tools for the design of software intensive systems	- Managing the complexity
5. IP management in AI-assisted software design	*Short term*	Understanding how to manage IP when using AI-assisted software design – i.e. establishing rules governing who owns code when it is created in part via AI that has been trained on a large corpus of existing code	- Rights management
6. Security by design in AI-assisted software engineering	*Short term*	Understand how to avoid introduction of vulnerabilities into software when using AI-assisted development, e.g. inadvertent use of upstream dependencies with vulnerabilities or code that does not employ best practice	- Avoiding vulnerabilities in AI-assisted software engineering

5.8.5 Recommendations

Research Theme	Software Engineering of complex and self-adaptive systems					
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**	**Subtheme 5**	**Subtheme 6**
International Collaboration					X Rights management	X Software security
Open Data / Open Source					X Rights management	X Software security
Large Trials						
Cross-domain research	**X** SW engineering applied to the telco domain	**X** SW engineering applied to the telco domain	**X** SW engineering applied to the telco domain	**X** SW engineering applied to the telco domain		

5.9 Human centricity and digital trust

5.9.1 Data authenticity and trusted digital interactions in dynamically composed service environments

[C5-64] points out that the impact of SNS will be limited if risk-averse users refuse to share data fearing it may be misused or reject advanced applications because they feel manipulated. Without trust, services such as health care or applications such as online elections may not be viable.

Therefore, ways to verify data authenticity and truthfulness will be needed, along with trusted digital interactions, especially in dynamically composed service environments. Trusted identities and authentication services for software and devices as well as humans are essential, along with access control mechanisms that users can understand, to manage their data and protect themselves from manipulation. Telco operators will have the opportunity to play the role of an identity provider and manager and, thus, controlling access to a world of applications.

Smart contracts and distributed ledgers, trusted hardware, and homomorphic encryption may provide links in the chain of trust, but the key is to use measures to support a holistic network of trust between stakeholders. Fact-checking services based on AI may play a role, as may services that govern the AI to ensure that novel technologies and applications remain compatible with societal needs such as the right of individuals to freedom of expression. Authenticity must be demonstrable, not just for data but also for the consequences of using data in AI-enabled decision-making algorithms. Technologies alone will not be enough – they must be deployed in a citizen-centric fashion, giving humans control over their interactions.

To achieve high levels of trust, software engineering methodologies and tools must provide the trust anchors needed by stakeholders: software developers, service operators, business customers and consumers, regulators and certification agents. Certification of products and services will be important, and will play an essential role in regulation to ensure security and safety in sectors such as medical IoT. As certification procedures are expensive, software engineering methods will be needed to implement them for dynamically changing systems in a cost-effective way, e.g. by focusing on critical sub-systems or operational contexts. In some areas, new procedures and standards (similar to ISO 26262) will be needed, e.g. to certify software based on machine learning/AI, for which there are no established methods today.

5.9.2 Human-centric software engineering and codes of ethics for software development

Future networks will enable a rich environment for multi-user interaction and will support "assistive" technologies such as tactile gloves and devices offering gesture recognition and haptic feedback. These devices and the related services will deliver information that is relevant to users' tasks at hand. The interactions will be mediated by algorithms in many cases. The overall system will be a complex constellation of software and content components operated by a multitude of ecosystem participants.

Designing, developing, deploying and maintaining these software-based systems is technically very challenging and many issues remain to be solved [C5-65]. For the requirement engineering, it is essential to understand for example how and who could take control in these spaces, how the information will flow and whether it is necessary to settle protection mechanisms against dominant positions, leading to serious threats to human integrity. The human-centric and ethical issues are often overlooked in the building of such systems.

5.9.3 Research challenges

Research Theme	Human centricity and digital trust		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Collaborative research work required in the context of identity and privacy in open ecosystems	Short-term (finished in 3y) The research activity should progress together with architectural achievements in 6G	Trusted identities and data authentication services offered by SNS	Human-centricity of SNS
2. Privacy and responsible software development	Mid-term (finished in 5y)	6G networks will enable new generation disruptive services. The activity should produce a set of requirements and tools for helping software developers and service providers to embed human-centric aspects.	Human-centricity of SNS
3. End to end data sovereignty for users	Mid-term	Understanding how to ensure that data owners can control access to and usage of their data wherever it is in the continuum, including its use in Generative AI.	Greater data owner control
4. Easy-to-understand & transparent audit of data usage for citizens	Mid-term	Understand technical and legislative measures that can enable citizens to understand easily and quickly what is being done with their data. Vendors hide behind long and complex privacy policies that are seldom read by data subjects before they consent.	Great citizen trust and control

5.9.4 Recommendations

Research Theme	Human centricity and digital trust			
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**
International Collaboration			X Human-centric security	X Human-centric security
Open Data / Open Source				
Large Trials				
Cross-domain research	X Multi- and interdisciplinary	X Multi- and interdisciplinary		

	research to address human-centricity	research to address human-centricity		

5.10 Quantum computing

Quantum Computing is considered as an emerging technology in the context of 6G [C5-66]. However, using quantum computing in the context of 6G requires that the research challenges to integrate quantum computing with classical computing will be solved [C5-67].

Quantum computers operate on qubits which can represent a combination of both zero and one at the same time by exploiting the quantum phenomenon of superposition. Combining qubits into a larger system enables quantum computers to perform multiple calculations with multiple inputs simultaneously. This allows to achieve an enormous - in some cases exponential - speed-up in executing certain algorithms for solving specific problems. For example, multivariable problems could be solved in a significantly more efficient way in quantum computers than in classical computers, in applications such as complex optimization in network planning (e.g. SDN controller placement [C5-68]), large system simulations, and quantum machine learning (e.g. Hybrid quantum deep learning for 6G resource management [C5-69]).

Quantum computers and classical computers will coexist, and hybrid algorithms utilize both. Both quantum computing and classical computing resources will be accessible from the computing continuum and should be usable in an integrated way, similar to the way today's High Performance Computing (HPC) is used. However, the full software lifecycle and software stack for quantum computers are different to those in classical computing. At the lowest machine level, qubits and their interfacing need to be controlled and optimized. At the next level up the software stack, it is about the coding of the algorithm by quantum circuits combing multiple qubits. On top of the stack, the algorithm execution is integrated with other parts of the application software.

Research challenges in this context include:

- The problem categories that are suitable for quantum computers in the telecom domain need to be better understood. Increased interdisciplinary research - including quantum information, computer science, and especially mathematics – is needed to develop domain- and problem-specific quantum algorithms, and to improve the efficiency of existing ones.
- Research into the programming abstractions (e.g. abstract machines, compilers, libraries, programming languages, APIs, etc.) that facilitate the design of hardware-independent quantum programs which nonetheless benefit from the specific features of the underlying quantum hardware.
- More research emphasis on the integration and orchestration of classical and quantum computing. This includes the interplay at the level of algorithms and the lower level of pre- and post-processing of a quantum computing. It also should cover the access, orchestration and integration of quantum computing resources that are available as a service within a computing continuum.
- Benchmarking, testing and debugging of quantum programs is still at a very early stage and needs to be explored at a fundamental level.

5.10.1 Research challenges

Research Theme	Quantum computing		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**

1. Exploring telecom specific problems suitable for quantum computers	Mid-term (finished in 5y)	Quantum algorithms for complex problems in the telecom domain	Alignment with computing continuum
2. Quantum software engineering	Long-term (finished in 7y+)	Implementation and integration of quantum algorithms in the telecom domain	Alignment with computing continuum
3. Quantum application security	Long term	Investigation into how Quantum applications and computing can be attacked, perverted or compromised, plus controls that can be used to mitigate against these events. Aim for Quantum computing / application security by design.	Secure quantum computing

5.10.2 Recommendations

Research Theme	Quantum computing		
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**
International Collaboration			X Quantum software security
Open Data / Open Source		X SDKs for implementing quantum algorithms	
Large Trials		X PoC of implemented quantum algorithms	
Cross-domain research	X Developing Quantum algorithms requires interdisciplinary research		

5.11 Sustainability

5.11.1 Green Softwarized Networks

In-network computation refers to the possibility to offload a set of computational functions from subscribers into networking devices. The core idea of in-network computation derives from the notion of Active Networks [C5-67], where there has been a proposal to replace transmitted packets with small programs (capsules) which would execute when reaching a forwarding device. The introduction of programmable network devices [C5-71], virtualization, as well as the integration of computational power, storage and network are the basis to revisit the OSI stack, and to redesign systems by considering an integrated approach: the combination of networking, distributed systems, storage, and programmable logic.

Such process becomes more robust and faster, by considering ways that integrate "better" computing and networking as is currently under debate in the IRTF COIN working group [C5-72]. One approach is to consider hierarchical approaches to support (control) algorithms in networked devices ("closer" to the source) thus resulting in faster response times. This can reduce information overload by processing and filtering data as "early" as possible in the network. Another approach is to revisit Active networking and explore the possibility of running (parts of) the application in networking devices.

The different technologies and techniques that back up the concept of in-network computing, (e.g. P4 programmable data planes; hybrid SON, see the discussion in 3.2.4.2) provides the means to consider a design that takes into consideration sustainability principles, where a feedback loop can be provided by the network to improve energy efficiency. A few directions are already being taken, towards the recognition of challenges that need to be tackled [C5-73][C5-74][C5-75], such as sustainable network management, sustainable network security, and sustainable network hardware management.

Sustainability, however, goes beyond considering energy efficiency in the design of open solutions (hw and sw), and APIs. It requires the development of mechanisms that allow the network to sense and react to existing resources and surroundings in a sustainable way. This requires relying on an approach that combines monitoring with context-aware/situational-aware offloading.

Furthermore, 6G is also aiming, by digitalisation, to provide significant support in achieving the EU Green Deal and the United Nations Sustainable Development Goals (SDGs). That means that not only 6G itself should be sustainable, but 6G should also enable other industry sectors to achieve the EU Green Deal objectives and the SDGs. In this context 6G is supposed to enable applications for use cases that address aspects such as resource-efficiency, safety, and inclusiveness. Several relevant initiatives in this context such as Manufacturing-X or CATENA-X consider sustainability as a key target in the development of common architectural stacks to be relied respectively by manufacturing or automotive stakeholders. By adopting common standards with focus on energy efficiency and sustainability, these two efforts are providing the basis for a new level of digitisation, considering data-oriented communications to interconnect smart products and even to facilitate cross-sector applications.

The overall 6G architecture needs therefore to evolve taking into consideration these latest efforts, the needs of critical applications along with the needs of time sensitive applications.

A few research challenges are as follows:

- Which energy consumption models can be considered to reach an adequate instrumentation? (short/mid)
- How to perform energy network optimization while taking into consideration application-level energy requirements and energy consumption optimization? (mid)

5.11.2 Research challenges

Research Theme	Sustainability		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Energy consumption models	Short term	Models	Sustainability of SNS
2. Energy network optimization while maintaining service performance	Mid Terms	Add-ons to an extended system management framework	Sustainability of SNS
3. SW based features of private networks to support sustainability in verticals	Mid-term (finished in 5y)	Seamless integration of 6G into vertical domain assets	Enabling energy efficiency and sustainability
4. Optimize energy consumption across end devices	Mid-term (finished in 5y)	Federated intelligence systems for devices	Enabling energy efficiency and sustainability
5. Balancing energy consumption optimisation & security		Understand the challenges, synergies and potential conflicts between optimisation of energy consumption and cybersecurity – i.e. where they support each other and what to do when there is conflict between them, aiming to come to an acceptable balance between the two requirements.	Maintaining both energy optimisation and cybersecurity at the same time

5.11.3 Recommendations

Theme	Sustainability				
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**	**Subtheme 5**
International Collaboration				**X** Common standardized solutions	
Open Data / Open Source	X Access to relevant data	X Access to relevant data			
Large Trials		X PoCs	X PoCs	X PoCs	
Cross-domain research			X Domain knowledge of verticals needed		X Balance between energy consumption and security

5.12 Software security

This section considers software security implications by taking the approach that security is the "the state of being free from danger or threat" (OED), and that threats and dangers lead to harms (specifically "physical injury, especially that which is deliberately inflicted", OED), which refers to injury, degradation, compromise or other detrimental situations to valuable assets including people, data, institutions – anything of value. The specific focus here is on software or AI as the vulnerable asset that may be attacked, and which may result in harms to itself or other assets it affects. The following table describes specific challenges and benefits resulting from research in these areas.

5.12.1 Research Challenges

Research Theme	Software Security		
Research Subtheme	**Timeline**	**Key outcomes**	**Contributions/Value**
1. Relate vulnerabilities in published databases such as CVE (Common Vulnerabilities and Exposures) to systemic risk	Mid-term (finished in 5y)	Research is needed on how to relate CVEs to systemic risk, i.e. to understand how do CVEs, which are easily found from software component SBOMs and databases, map to real-world risks in operational systems on assets that the system's stakeholders care about? There is a more general point about mapping of multiple sources of information, e.g. CVEs, CTI and others to software vulnerability and then onto systemic risk – investigating relative contribution / influence of each component to vulnerability or warning of attack etc.	Improved knowledge on mapping vulnerabilities to real-world risks and harms
2. Mapping anomaly detection to vulnerabilities	Mid-term (finished in 5y)	Many researchers are using machine learning in different situations to detect anomalies, e.g. deviations from the norm, but a key challenge remains regarding whether detected anomalies represent vulnerabilities in a software component under test. Research is needed to map	Mapping systemic anomalies to known vulnerabilities or to characterise potential zero-day vulnerabilities.

		detected system anomalies to known vulnerabilities or to characterise potential zero-day vulnerabilities.	
3. Relate cybersecurity threats and controls to AI	Mid-term (finished in 5y)	Research is needed to understand the interplay between AI vulnerabilities, threats, harms and controls and those from cybersecurity – e.g. how cybersecurity threats can affect AI, e.g. how AI harms can be linked to cybersecurity threats, how cybersecurity controls can manage AI threats and harms	Improved knowledge on how AI harms can be caused or controlled via cybersecurity aspects
4. CVSS base metric modifications	Mid-term (finished in 5y)	Research is needed on how to successfully implement a CVSS v4 approach, where initial base metrics need to be modified by the user to accommodate their specific concerns and environment. The major challenge is how to provide decision support to the practitioners who need to be able to understand how to modify the base metrics based on their concerns and environment – how can they do this reliably and is there a methodology where this modification can be done consistently by different practitioners so it is comparable with others?	Best practice on adapting CVSS to deployment environment

5.12.2 Recommendations

Theme	Software security			
Action	**Subtheme 1**	**Subtheme 2**	**Subtheme 3**	**Subtheme 4**
International Collaboration	**X** Vulnerability – risk mapping	**X** Anomaly – vulnerability mapping	**X** Understanding interplay AI and Security	**X** Best practices
Open Data / Open Source				
Large Trials				
Cross-domain research				

6 Radio Technology and Signal Processing

Editor: Wen Xu

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

1) Spectrum reutilization, interference management, subnetworks, wireless edge caching and RAN evaluation;
2) Optical wireless communications (OWC);
3) Millimeter-wave and terahertz communication;
4) Massive MIMO including ultra- and cell-free massive MIMO (mMIMO), reconfigurable intelleget surfaces (RISs), UAV-assisted communication, near field, and fluid antenna system;
5) Waveform, non-orthogonal multiple access (NOMA) and full-duplex;
6) Enhanced modulation and coding;
7) Integrated sensing and communication (ISAC);
8) Grant-free random access for massive connections;
9) Machine learning empowered physical layer.

6.1 Vision and Requirements

Each generation of wireless communication has provided new services. 5G is no exception, and in addition to the enhanced mobile broadband (eMBB), the new services ultra-reliable low latency communication (URLLC) and massive machine-type communication (mMTC) were introduced.

Over the next decade, digitalisation will continue to evolve as Information and Communication Technologies (ICTs) have the potential to address many of the societal challenges that lie ahead. Thus, the vision for 6G is to include societal needs in addition to the traditional technical requirements for advanced digital services to humans and machines. The usage scenarios include [C6-1]: i) Immersive Communication, ii) Hyper Reliable and Low-Latency Communication, iii) Massive Communication, iv) Ubiquitous Connectivity, v) Artificial Intelligence and Communication, and vi) Integrated Sensing and Communication. The target year for 6G deployment is well aligned with the target year for UNs seventeen sustainable development goals (SDGs), both being 2030. Research should therefore focus from the outset on meeting the requirements of a sustainable 6G. 6G should also properly capture the societal challenges as defined by the SDGs, targeting 6G for Sustainability services [C6-2]. To this end, the 6G system should be assessed using new key value indicators (KVIs) such as digital inclusion, trustworthiness and sustainability, in addition to the traditional service-oriented key performance indicators (KPIs), such as data rates, low latency and high-precision positioning. This gives rise to a number of new challenges, as described below.

Firstly, the traditional KPIs driving the design of 6G should be revisited to capture holistic requirements on CO2 footprint. For this purpose, 6G must become significantly more energy-efficient than previous generations. Unfortunately, advancements in hardware energy efficiency alone are unlikely to offset the increasing power consumption driven by escalating traffic and data rates. With cellular networks already consuming about 0.6% of the world's energy in 2015 [C6-3], and traffic expected to grow by about two orders of magnitude every decade [C6-4], improving the energy efficiency of 6G networks is a critical challenge. To put this into perspective, if the goal is to reduce the absolute energy consumption of cellular networks by a factor of 10 while handling a hundredfold increase in traffic, the energy efficiency per unit of data would need to improve by a factor of 1,000. Achieving this dramatic increase in efficiency cannot be expected through incremental innovations but rather requires a disruptive approach.

Secondly, 6G should go beyond connectivity and become a trusted platform including communication, computation and storage, to provide new capabilities as intelligence at the edge and in the cloud, joint communication and sensing, and highly accurate positioning. This 6G platform will support new applications such as cyber-physical systems (CPS), which take advantage of ubiquitous connectivity, sensing and contextual awareness. It will also support extended reality (XR) and immersive applications utilizing high data rates, low latency, precise positioning and sensing. These applications should be available to customers at a low cost, which requires cost-effective solutions for user devices without compromising the user experience. This includes utilizing edge computing capabilities to minimize latency and computing costs, jointly optimizing processing and communication resources to reduce power consumption, and employing predictive resource allocation to adapt to changing conditions and ensure service continuity. Moreover, solutions at higher layers should be investigated to relax communication latency requirements. For instance, predictive modeling can determine future states based on current and previous states, thus allowing actions and feedback to be communicated in advance.

Technical 6G requirements include Tbps or sub-Tbps data throughput, sub-ms latency, extremely high reliability, everywhere mMTC, extreme energy efficiency, very high security, cm-level accuracy radio positioning, and global coverage though integration of non-terrestrial networks (NTNs) [C6-1][C6-5][C6-6]. These new capabilities and requirements will require continued research and development of radio technologies and signal processing and protocols. A natural way forward to deal with these challenges is to consider electromagnetic spectrum at higher frequencies such as the sub-THz or THz spectrum, as well as infrared and visible light spectrum, since these frequencies offer wider bandwidths for higher data rates and higher resolution positioning and sensing. However, these benefits come at the cost of increased power consumption of radio components. In addition, analyzed use cases show that not all the extreme requirements are needed at the same time. Therefore, various radio options should be provided, optimized according to the extreme requirements and specific use case scenarios [C6-7][C6-8]. Besides, the centimeter and millimeter wave spectrum currently used for 5G and other legacy wireless systems needs to be re-farmed and reused more efficiently. To this end, interference management and coexistence issues should also 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, further research is needed in several other domains, such as modelling, designing and optimizing the use of intelligent reflecting surfaces (IRSs) for communications, positioning and sensing. Further research is needed to integrate high-precision positioning and sensing into mobile communication networks, and to explore the potential for joint sensing and communication using the same hardware and software systems. Additional areas that deserve attention are grant-free random access for massive connections, and wireless edge caching and computing. Moreover, machine learning (ML) and artificial intelligence (AI) have found success in many applications. For the application in communication technologies and radio interface design, further research is still needed on all layers, including semantic communication (see Chapter 10 for more details). Meanwhile, distributed learning and inference over the wireless links will be a common norm in 6G networks, where the radio transmission technologies would merit a fresh look to minimize both spectrum and energy overheads, and to optimize joint communication and AI performance taking into account the specific learning and inference characteristics.

Compared with 5G [C6-9], 6G air interface may need to fulfil more stringent KPIs and requirements, such as

- Energy efficiency (bit/Joule): It is the capability to minimize the energy consumption in relation to the traffic capacity provided.
- Spectral efficiency (bit/s/Hz): This is a metric widely used in Shannon information and coding theory for optimizing a communication system or its building blocks. For example, the 5th percentile user spectral efficiency is the 5% point of the CDF of the normalized user throughput. Note that in the case of very high frequency scenario (e.g. sub-THz), spectral efficiency may not be the most important design metric. User experienced data rate, defined as the 5% point of the CDF of the user throughput, can also be used.
- Peak date rate: Maximum achievable data rate or throughput under ideal conditions per user/devices (in Tbit/s).
- Area traffic capacity (bit/s/km^2): The total traffic throughput served per geographic area. This becomes more and more important for uplink when sensing and distributed AI become common usage scenarios in 6G.
- Coverage: This is usually the 3D global coverage and full connectivity by terrestrial and non-terrestrial coverage.
- Reliability: The success probability (e.g. $1\text{-}10^{-7}$) of transmitting packet of different size within the maximum allowed latency at a certain channel quality.
- Mobility: Maximum speed at which a defined QoS and seamless transfer between radio nodes can be achieved (in km/h), where the nodes may belong to different layers and/or radio access technologies (multi-layer-RAT).
- Air interface latency (user plane): The contribution by the radio network to the time from when the source sends a packet to when the destination receives it (in ms).
- Connection density: Total number of connected and/or accessible devices fulfilling a specific quality of service (QoS) per unit area. Note here in 6G, the QoS could have broader sense than the traditional communication QoS, e.g. it could be sensing resolution or AI accuracy and efficiency.
- Positioning accuracy: It is the difference between the calculated horizontal/vertical position and the actual horizontal/vertical position of a terminal.
- Privacy: Capability of protecting the radio access from eavesdropping.

Besides quantitatively measurable KPIs, such as spectral efficiency in terms of bits per second per Hz, energy efficiency in terms of bits per Joule, reliability in terms of success rate, etc., there are some soft KPIs, such as intelligent level of a network in terms of the network intelligence from the perspectives of action implementation, data collection, analysis, decision and demand mapping according to the degree of manual participation in network operation, coverage which strongly depends on the number of deployed base stations, controllable radio environment in terms of dynamically changing the characteristics of radio propagation environment and creating favorable channel conditions to support higher data rate communication and improving the coverage, privacy, etc.

Figure 6-1 shows the connections between different enabling technologies and different 6G requirements and KPIs.

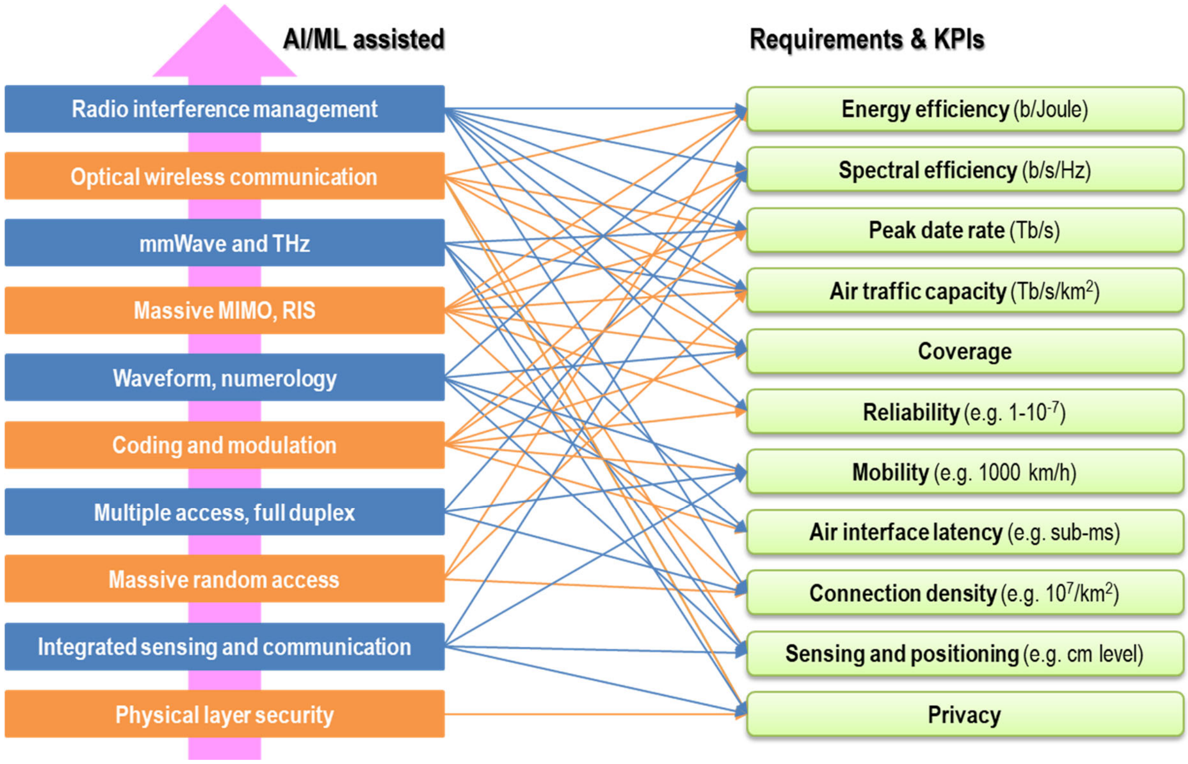


Figure 6-1 Enabling technologies with main contributions to different 6G requirements and KPIs.

6.2 Radio Networking and Interference Management

6.2.1 Spectrum re-farming and sharing

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

A straightforward approach to spectrum reutilization is *spectrum re-farming*, which involves, for example, static allocation of spectrum resources to different RATs. Note that a static nature usually leads to a poor spectrum utilization.

On a more finely resolved time scale, efficient spectrum utilization is achieved by dynamic inter-band resource allocation and scheduling with optimized multi-RAT handover and interference coordination. This has traditionally been based on a centralized radio resource management. However, the associated high signaling overhead motivates the exploration of decentralized strategies. Spectrum sharing 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 does not only increase the performance due to better spectrum utilization, it also makes the network more robust and resilient towards shadowing effects, thus improving reliability and availability. Besides, opportunistic spectrum access opens the door to further improvements by means of multiuser diversity, making it possible for UEs to access spectrum channels, e.g. on a CSMA-like basis.

A key point on the path to 6G networks is the autonomy from human intervention in network configuration, which implies network self-organization and management (SOM) mechanisms that intelligently take into account the characteristics of the environment. To achieve this in the context of joint utilization of licensed and unlicensed spectrum, adaptive and dynamic spectrum sharing strategies are required. In this line, cognitive environment concepts, in which spectrum awareness, e.g. based on a combination of advanced SIGINT (signal intelligence) and AI (artificial intelligence) techniques, can be used to ensure coexistence with existing (e.g. analogue) in-band services. On the other hand, spectrum awareness and reutilization can help to increase security at radio level, e.g. by detecting and counteracting threats such as RF jamming or spoofing. While decentralized SOM approaches can reduce the computational load and signaling overhead compared to centralized SOM at the cost of suboptimal solutions, the trade-off between network load and performance optimality is subject to the characteristics of the multi-band environment.

Such considerations and new concepts for spectrum licensing and reutilization are particularly important in the context of new radio technologies such as the millimeter wave, optical wireless, and terahertz communications discussed below, which have a radically different interference footprint compared to the conventional sub-6 GHz communications. Their highly directional links and susceptibility to blockage reduce interference, which significantly increases the potential gains of spectrum sharing and simplifies its use [C6-10]. Dynamic sharing of the different bands will imply a highly complex scenario in which autonomous AI-based mechanisms are envisioned to be crucial.

6.2.2 Subnetworks and coexistence

While the current generation of mobile communications (5G) has already made a big step towards supporting use cases with strict requirements in terms of latency, reliability, and throughput, 6G is expected to move this one step further, with latency of loop-cycles down to 100 µs, more than six nines reliability and multi-gigabit data rates (not necessarily concurrently). Different scenarios such as factory floor level communications (e.g., within a robot or a production cell, i.e. in-robot and in-production cell communications), sensor/actuator traffic inside a car (in-car communications) and in/on-body communications among smartphones and XR wearables call for these extreme requirements. At the moment, wired solutions are mostly used for some of these deployments, but strong benefits could be provided in terms of flexibility by using wireless.

One approach to achieve those extreme requirements is the introduction of so-called subnetworks [C6-11]: For many scenarios requiring ultra-low latencies, both the origin and the recipient of a given message are close by. The vision is then to have an access point that a) controls and serves the needs of these devices next to each other and b) is, at the same time, a sort of special device connected to an overlay public or private network. This allows avoiding the delays from forwarding the traffic through the core network. Note that subnetworks may also serve non-critical traffic arising within its coverage area targeted to applications outside of the subnetwork. In this case, the access point is required to be attached to the overlay network.

Target scenarios are typically covering rather small areas (few meters up to few tens of meters), therefore many characteristics linked to small cells (pico, femto) apply here as well (e.g., low-power transmissions, antenna configurations, etc).

To be able to meet these extreme requirements, a dedicated air interface (e.g., supporting much shorter symbols), access protocols and diversity mechanisms need to be selected and designed.

In some scenarios it may happen that the subnetwork loses its connection to the overlay network (e.g., a car driving into a tunnel or entering an underground parking lot). As subnetworks may supporting life-critical applications, it is fundamental that mechanisms that allow them running autonomously even when out-of-coverage must be present.

Several subnetworks may be present in close vicinity (e.g., cars on a congested road, production units on a factory floor, people attending a crowded event, etc.). Interference among those subnetworks may then arise and needs to be handled through both respective design choices (e.g., allowing for multiple orthogonal channels) and more sophisticated resource management procedures and interference mitigation techniques (both in a centralized manner exploiting the connection to the overlaying network and decentralized to ensure reliable connections when the subnetwork needs to act autonomously).

A single subnetwork may be required to serve several tens or even hundreds of individual nodes. So, channel access procedures are required. Also, different devices categories are expected: on one side higher-capable devices that can act as subnetwork access points, with potentially higher transmit power to connect to the overlay network and managing connections to the devices within the subnetwork; on the other side lower-capable devices, with limited maximum transmit power, designed to be battery driven and that can be built significantly cheaper.

While there is very limited or close to no mobility between subnetworks (e.g., a sensor within a car does not switch to another car), individual subnetworks may be mobile. That implies the need for mobility techniques to be applied and also makes the interference very dynamic.

As indicated above, there are many open questions to be answered and even more fundamental ones such as: What frequency ranges (FR1: sub-6 GHz, FR2: 24-71 GHz, FR3: 7-24 GHz) are reasonable selections for subnetworks and what are the related implications? How much spectrum do we need? Is licensed or unlicensed access a better choice? Do we use dedicated spectrum, or should we go the ultra-wideband (UWB) underlay route? How can we make the system more pro-active and less re-active? How can we benefit from the fact that subnetworks are attached to an overlay network at least most of the times? How do we integrate the subnetwork into the 5G/6G overlay-network w.r.t. architecture and protocols?

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, utilizing unlicensed bands is important and technologies to bring the quality to the level of licensed spectrum 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 especially for academia but also complementing private simulation systems from industry. 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. Coexistence 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 coexistence needs careful considerations. Scalability and power efficiency are critical for the 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 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. Coexistence 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 [C6-12].

6.2.3 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 [C6-13]. 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 5th Generation (5G) of standards and systems is focused to a certain extend on the traditional view of “increasing peak rates” [C6-14]. In contrast, it is widely recognized that a major driver of the wireless data traffic increase is on-demand access to multimedia content (Wireless Internet). 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.

In light of the above considerations, a novel content-aware approach to wireless network design is needed. Such novel approach should support both “Gigabits per second” and “a few Terabytes per month” for all. 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 (e.g. via broadcasting or multicasting) 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 [C6-15][C6-16][C6-17][C6-18]. This means pre-storing segments of the content files at the base stations, at dedicated "helper" nodes, and also in the user terminals. Further in this direction, the recent introduction of vector coded caching [C6-19] has demonstrated its ability to utilize traditional precoding approaches with a modest amount of cache content at the receivers, in order to offer unprecedented gains in realistic wireless multi-antenna downlink systems for delivery of video on demand (VoD). As the analysis in [C6-20] shows, such systems can boost the rate performance of already optimized (cacheless) MU-MISO and MU-MIMO systems, by a stunning factor of 250%-400%, and can do so for various linear precoders, and under practical considerations such as power allocation and feedback costs. For example, with 32 transmit antennas, a received SNR of 20 dB, and realistic cache-size constraints, vector coded caching offers a multiplicative throughput boost of approximately 310% with Zero Forcing (ZF)/Regularized Zero Forcing (RZF) precoding, and a 430% boost in already optimized Matched Filter (MF)-based systems. Furthermore, vector coded caching accelerates channel hardening, benefiting feedback acquisition, often surpassing 540% gains over traditional hardening methods. These methods call for tighter collaboration between the underlying radio-access network coding strategies and application protocols thus implying non-trivial evolution of 3GPP network and terminal architectures.

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. This behavior is referred to as "full throughput scalability" [C6-21]. 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 [C6-22] 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) and corresponding network and terminal architecture evolution;
- Protocol architectures (e.g., combining edge caching with schemes for video quality adaptation such as Dynamic Adaptive Streaming over HTTP (DASH) [C6-23]);
- AI/ML based content popularity estimation and prediction, to efficiently update the cached content [C6-24].

6.2.4 RAN performance evaluation tools

6G is based on the use of several new types of network elements (e.g., reconfigurable intelligent surfaces/RISs, unmanned aerial vehicles/UAVs assisting communication), frequencies (e.g., THz, FR3), and architectures (e.g., cell-free, non-terrestrial networks/NTNs), and needs to support new services (such as integrated sensing and communication/ISAC), and becomes thus more complex. Mastering the complexity that results from this multiplicity requires the development of new design and system level performance analysis tools.

There are currently two main known methods for these performance evaluation questions: discrete event (system-level) simulation and stochastic geometry (SG). Both need to be researched for future RAN performance evaluation. The advantage of the system level simulation is that it does not require any modeling assumptions. But it is often (though not always) based on proprietary and expensive software tools; it also remains ineffective for optimization and more generally for handling the inherent complexity of 6G RAN. The research effort should focus on the integration of the new features of 6G RAN listed above in major simulation tools (e.g., NS-3) and on the design of new methods for improving the efficiency of discrete event simulation in the context of optimization.

SG is an ancient branch of probability developed by European mining schools in the context of materials science. The use of SG for wireless networks is recent [C6-25][C6-26]. This approach is now used worldwide for the design and dimensioning of cellular networks. It has in particular been used by Orange for the deployment of its last generations of cellular networks. It was also used as a fundamental tool for the design of self-organized networks, WiFi networks, vehicular networks, etc. This approach can be seen as a spatial extension of the approach that consists in representing Internet traffic and queuing networks in routers as functionals of one-dimensional point processes (the IP packet submission processes). In this context, the properties of the communications network are seen as functionals of two- or three-dimensional point processes exploiting the statistical properties of the various network elements (base stations/BSs, users, etc.). In the basic model of the cellular architecture, the BSs are arranged according to a homogeneous Poisson point process of intensity lambda in the Euclidean plane.

Consider for example the downlink where the interference coming from other cells is treated as noise, the user data rate is proportional to log(1 +S/(I +N)) where S is the power of the signal received by the user from the serving BS (the closest), I is the power of the interference from neighbor cells, and N is the power of the thermal noise. In this model, the distribution of the distance to the nearest BS is a Rayleigh law as shown by a classical empty-ball argument. Conditionally on the fact that the serving BS is at distance r, the power of the interference is characterized by its Laplace transform which is the Laplace transform of a Poisson point process of intensity 0 in the ball with center 0 and radius r and with intensity lambda outside this ball. We can then give an explicit form for the distribution of this Shannon rate and therefore quantify the spectral efficiency of such networks in closed form as a function of a few physical parameters [C6-25][C6-26].

The essential point is to see this approach as an instance of statistical physics type analyzes which make it possible to determine the ensemble averages in a system of particles (here the network elements) interacting according to Shannon's law. Indeed, the distribution calculated in this model is also the empirical distribution of rates obtained on the downlink by users located in a large ball of the Euclidean plane. Results of this type can in turn be used to determine which architectures optimize certain metrics, for example economic metrics (see, e.g., [C6-27]). For such a network architecture, we showed that densification becomes counterproductive beyond a certain threshold: for realistic attenuation functions, there exists a threshold beyond which the spectral efficiency decreases. This has important economic consequences.

There is a need for a SG based system-level evaluation of major 6G paradigms, such as ISAC networks, cell-free architectures, RIS-based networks, NTNs, etc. The needs for the last two paradigms are briefly described below.

The SG models allow one to associate a cloud of RIS with each BS. This cloud can for instance be located in the neighborhood of coverage holes. The RISs in this cloud will relay the signal from the BS to the users of the cell. This is particularly useful in case of blockages between BS and users. Explicit integral forms for the spectral efficiency of models of this type generalizing the results of the basic cellular model would be quite useful. Here are some questions that such models would allow to answer: what is the influence on the spectral efficiency of the geometry of RIS clouds? Is it better to have fewer RISs with more controllable reflection elements or the opposite? What is the average gain in spectral efficiency provided by RISs? How does this gain depend on the density of obstacles? In a given configuration, is it better to invest in RIS or to densify BSs?

NTNs are based on constellations of satellites in low or medium orbits. These networks will be integrated into 6G. BSs now form a spatial process with satellites stationed in orbits with a certain inclination. The NTNs already provide universal Internet access. New models based on spherical SG started but should be further developed based on the notion of orbital processes and point processes on these orbits [C6-28]. The notions of coverage and closest connection should be extended and the Shannon rate distribution could be established in principle. Many questions arise at the system level: technical questions about achievable throughput, handovers, scheduling and routing. Economic questions on interactions and interference between 5G terrestrial networks and the NTNs.

Beyond these examples, the essential point is that the developed tools should be used to analyze the interactions between the many RAN paradigms described above. Here is a simple example in line with that last two examples: what about the role of RISs for improving connectivity with NTNs? This should be done in a systematic way for the new 6G RAN paradigms listed above as well as for the interactions with the RAN architecture options: level of centralization or decentralization, MEC, etc. This analysis should be capable of joint deployment optimization and of identifying joint economic opportunities. These tools should also allow for the design of new real-time and non-real-time controllers for all these new paradigms.

6.2.5 Research challenges

Research Theme	**Radio Networking and Interference Management**		
Research challenges	Timeline	Key outcomes	Contributions/Value
Spectrum re-farming and sharing	- Mid- to long-term.	- Advanced methods and protocols to efficiently re-utilize (licensed and unlicensed) spectrum resources.	Improved energy efficiency, spectral efficiency, capacity, throughput, reliability.

	- To be specified in 6G standard.	- Spectrum sharing, e.g. using cognitive radio-based solutions.	
Subnetworks and coexistence	- Mid- to long-term. - May be specified in 6G standard.	- New radio access node setting up a sub-network serving local nodes: 1) vertical use cases (e.g. inside a car or robot); 2) consumer use cases (smart wearables); 3) very low latency, very high reliability, extreme data rates, reduced energy consumption. - Unified and scalable architecture supporting low and ultrahigh data rates.	New use cases for wireless communications. Replace wired connections through wireless systems (e.g. CAN-Bus). Support of the metaverse.
Wireless edge caching	- Mid- to long-term.	- Advanced methods, and protocols, network and terminal architecture evolution. - Coding, e.g., combining edge caching with multiuser physical layer schemes. - Machine learning based content popularity prediction for efficiently updating the cached content.	Improved energy efficiency, spectral efficiency, capacity, throughput, reliability, quality of experience (QoE).
RAN performance evaluation tools	- Mid- to long-term.	- Performance evaluation methodology for RAN and architecture. - Mastering the complexity of new RAN elements deployments and architectural options.	RAN deployment guidelines. New non-real time controllers. RAN economic value guidelines.

6.3 Optical Wireless Communication

The continued exponential growth in mobile traffic [C6-29] means that inevitably the RF part of the electromagnetic spectrum will not be sufficient to be able to drive the cyber-physical continuum which is centered around immersive user experience in an XR environment, digital twins, the convergence of computing, sensing, control and robotics. [C6-30].

It is, therefore, essential to consider the infrared and visible light spectrum, both of which are part of the electromagnetic spectrum for future wireless systems for terrestrial, space and subsea applications. 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 add new capacity – the available unregulated spectrum is 2600 times larger than the entire RF spectrum, and provide enhanced physical layer security features. While most light-based communication systems are based on intensity modulation / direct detection (IM/DD), recently coherent optical wireless systems have been proposed [C6-31].

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) [C6-32]. Moreover, 26 Gbps were demonstrated using a dual wavelength laser device emitting in the visible and infrared spectrum [C6-33]. This work has been extended to a 10x10 WDM system and 105 Gbps were demonstrated at CES 2022 [C6-34]. Furthermore, a record of received data rates of 1.1 Gbps 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 order to meet UN targets. 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 [C6-35].

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) [C6-36]. LiFi enables bi-directional networked communication including multiuser access and handover. The major research efforts in the last 15 years have been focused on enhancing link data rates of IM/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.

Like 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-optimal. 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 leading to data densities of 88 Gbps/m^2 [C6-37].

Furthermore, due to the extremely short wavelength, the active detector sizes are very small, and massive MIMO structures can be implemented at chip-level [C6-38]. 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 has been standardized within IEEE 802.11. The new LiFi standard has received the following reference: IEEE 802.11bb. Similarly, VLC is being standardized in IEEE 802.15.13, while OCC has been standardized 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 [C6-39] 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.

6.3.1 Research challenges

Research Theme	Optical Wireless Communication		
Research challenges	Timeline	Key outcomes	Contributions/Value
Transmitter and detector technology, incl. solar cell data detectors acting as simultaneous energy harvesting devices, avalanche photodiode arrays, and Vertical Cavity Surface emitting lasers (VCSELs), e.g. in >1250 nm spectrum range	- Mid- to long-term.	- Devices that deliver optical-to-electrical (O/E) and electrical-to-optical (E/O) conversion efficiencies, e.g. for electrical bandwidth > 10 GHz at hundreds of mW optical transmit power, and receiver sensitivities less than 40 dBm.	Toward zero carbon footprint of future networks. Tbps wireless multiuser access networks. Extreme MIMO/WDM, > 1000 channels.
Optimized multiuser access and interference management	- Mid-term. - May be specified in 6G standard.	- Achieving extreme networks densification indoors towards 100 Gbit/m^2 at 100X improved energy efficiency in a multiuser scenario.	Extreme network densification, > 100 Gbps/m^2. Sub ms latency.
Optical wireless backhaul	- Short- to long-term. - May be specified in 6G standard.	- Point-to-point backhaul achieving > 1 Tbps at distances up to 10 km in terrestrial environment, and > 10 Gbps in satellite-to-satellite and satellite-to-ground environments.	Achieving pervasive coverage and overcoming digital divide; step-change improvements in network densification.
Bespoke RIS technology for OWC	- Short- to mid-term.	- RIS to support mobility in indoor and outdoor scenarios.	Ultra reliability, mobility.

6.4 Millimeter-Wave and Terahertz Communication

In the last two decades, major device, communication and networking features have led to the development and commercialization of millimeter-wave (mmWave) wireless technology. Today, wireless local area networks operating in the Industrial, Scientific and Medical (ISM) 60 GHz frequency band and orchestrated by the IEEE 802.11ad and the IEEE 802.11ay, are a reality. Similarly, 5G wide area networks operating in the licensed Frequency Range 2 (FR2) between 24 GHz and 71 GHz are already deployed in several countries. Higher data-rates (approaching 20 Giga-bits-per-second or Gbps) and lower latencies (approaching few milliseconds) are some of the promises of mmWave technologies to enable long-awaited applications including immersive augmented and virtual reality, the tactile internet, and autonomous unmanned networks, among others, all within different contexts, from entertainment to education to remote work telepresence. Moreover, besides communications, the mmWave spectrum has also enabled exciting applications in the field of wireless sensing, from precise localization and radar, to the extraction of body features. While the adoption of the mmWave spectrum has not been as prominent as expected in mobile front-haul applications, its use as a fixed wireless access technology is on the rise.

All the aforementioned commercial technologies and the majority of the research solutions explored to date are for systems operating under 100 GHz. However, this has quickly changed in the last decade. Today, there are several major academic and industrial research initiatives worldwide aimed at developing wireless solutions in the sub-terahertz and terahertz bands, or broadly speaking, at frequencies above 100 GHz, broadly defined between 100 GHz and 10 THz [C6-40][C6-41]. The US National Science Foundation Spectrum Innovation Initiative, the Semiconductor Research Corporation (SRC) and DARPA Communication and Sensing at Terahertz frequencies (ComSenTer), the National Natural Science Foundation of China, multiple European projects funded by the Beyond 5G track of the Horizons 2020 program, and several industry-led efforts (e.g., Nokia, Ericsson, Samsung, Huawei) are just a sample set of a growing pool. In addition, the first standard for communications at frequencies above 275 GHz, the IEEE 802.15.3d, was approved in 2017 and amended as the IEEE 802.15.3-2023 last year. Moreover, the International Telecommunications Union (ITU) at the 2023 World Radiocommunication Conference (WRC-23) with resolution COM6/17 defined several frequency ranges in the sub-terahertz band for future 6G network studies.

When moving to these frequencies, not only are there larger contiguous bands for ultrabroadband communication and networking systems, but the smaller wavelength of terahertz signals, which enables the detection of smaller targets, and the higher photon energy of terahertz radiation, which translates interactions at the materials molecular level into unique electromagnetic signatures, give a whole new meaning to what wireless sensing means, opening the door to concepts such as non-damaging imaging, spectroscopy, and atmospheric sensing, all in one [C6-42][C6-43].

Originally, it was thought that the terahertz band would only enable such applications over very short distances, such as in wireless nanosensor networks and the Internet of Nano-Things [C6-44][C6-45], but in addition to those, today, terahertz communications are considered to be at the basis of long-range outdoor applications including ultrabroadband wireless backhaul [C6-46], vehicular networks, and non-terrestrial networks, including satellite communications [C6-47][C6-48].

Of course, such exciting opportunities come with several challenges spanning devices, wave propagation, communication, signal processing and networking. In terms of THz devices, major progress has been achieved in closing the so-called terahertz technology gap. The key hardware building blocks of a THz wireless communication and sensing system include the 1) analog front-ends, 2) the antenna systems, and 3) the digital back-ends. There are three main approaches to the development of analogue *THz front-ends*. First, in an electronic approach, frequency multiplying chains can be utilized to up-convert a microwave signal to the Terahertz band. By moving from Silicon and Silicon-Germanium-based transistors to III-V semiconductor-based transistors and Schottky diodes, on-chip transceivers able to deliver a few hundreds of milliwatts at 300 GHz and a few milliwatts above 1 THz have been demonstrated [C6-49]. Second, in a photonic approach, the different frequency generation based on laser photomixing is at the basis of several THz transceivers operating at a few hundreds of GHz [C6-50]. While their output power is lower than the electronic systems, their phase noise is lower and the potential bandwidth is larger. Third, the direct generation and modulation of the THz signals with plasmonic devices built with graphene and other two-dimensional materials has been proposed [C6-51] [C6-52]. Their high efficiency, combined with their very small footprint (micrometric in size for the entire front-end), is at the basis of future ultra-massive MIMO systems (see Section 6.5). Independently of the approach, the high gain directional *antenna systems* in transmission, reception as well as in reflection are needed to overcome the lack of higher transmission power and high propagation losses. Beyond the fixed high-gain directional antennas and antenna arrays, lenses and metasurfaces can be used to engineer the radiation, propagation and detection of THz signals [C6-53]. Finally, the *high-speed data-converters* and digital signal processors (DSPs), able to sustain multi-Gbps and Tbps are needed. Very high-order parallelization, enabled for example by Radio-Frequency Systems on Chip (RFSoc), is one of the clear paths moving forward [C6-54].

Multibeam antennas are also critical components in future wireless communications networks. The quasi-optical techniques are expected to become part of the design of THz and mmWave antennas. New challenges for research are the design of lenses with metamaterials and transmit arrays based on metavolumes.

Multibeam antenna design also requires the design of low-loss distribution networks, such as groove gap technology, integrated with electronic components. It is necessary to develop very efficient analysis methods for inhomogeneous and anisotropic dielectrics [C6-55].

In parallel to THz technology development, much has been accomplished in terms of understanding THz propagation, by following both physics- [C6-56] and data-driven approaches [C6-57][C6-58]. There are three main phenomena affecting the propagation of THz waves, namely, molecular absorption loss, spreading loss, and blockage. 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 vapor). The absorption does not occur at all frequencies, but only at known absorption peaks and, while it is generally perceived as a problem, it can also be at the basis of enhanced physical layer security [C6-59]. The *spreading loss* accounts for the attenuation due to expansion of the wave as it propagates through the medium and, because of the small effective area of antennas as we move up in frequency, becomes critical at THz frequencies. This is why high-gain directional antennas are needed. High gain antennas exhibit narrow beams, which due to the interaction of Terahertz radiation with materials, make blockage a problem. To overcome this limitation, the use of intelligent reflecting surfaces (IRS) has been motivated, as a way to engineer non-line-of-sight paths [C6-53]. Ultimately, the stochastic multi-path channel models are needed to statistically characterize the channel. In this direction, massive experimental measurement campaigns in different indoor and outdoor scenarios are being performed [C6-60], including in the presence of bad weather [C6-61]. These should guide future THz infrastructure deployments and in the development of tailored real-time channel estimators and equalizers.

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, innovative 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 [C6-62]. For longer communication distances, enabled in part by ultra-massive MIMO systems (Section 6.5), new dynamic bandwidth modulations are needed to not only overcome but even leverage the unique distance-dependent bandwidth created by molecular absorption [C6-63]. As of today, both single-carrier (e.g., M-QAM, M-PSK, APSK) and multi-carrier (e.g., OFDM and DFT-spread-OFDM) waveforms for THz systems have been developed, but it is not yet determined which is going to be the waveform(s) for 6G THz systems. On the other hand, non-conventional modulation schemes like zero-crossing modulation (see Section 6.6), have been proposed to mitigate the increased power consumption of the analog frontend at mmWave and THz systems. Independent of the modulation, and like any wired or wireless Tbps communication system, physical-layer synchronization (in time, frequency and phase) becomes a major challenge.

It is relevant to note that the use of large transmit, reflect and receive antenna arrays, lenses and metasurfaces results in Terahertz communication systems to often operate in the near-field (see Section 6.5). For example, a 10 cm antenna array at 120 GHz has a far field of 8 m and, at 300 GHz, it has a far field of 20 m. The far field of a 1 m antenna array or metasurface (e.g., in a base station) at 120 GHz starts at 800 m, or 2 km when operating at 300 GHz. As a result, conventional beam management strategies, which consider far field operation, are no longer valid. Besides trying to "fix" existing techniques to work in the near field (e.g., by compensating for the spherical wavefronts generated by small radiators in the near field), new unique types of wavefronts or beams that are only possible in the near field need to be explored [C6-64]. These include, for example, self-healing Bessel beams that can recover from obstacles [C6-65], or curving airy beams that can bend around corners [C6-66].

Moving up in 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 minimizes the chances of collisions. However, the need for high-gain directional antennas simultaneously at the transmitter and the receiver to establish links over realistic distances makes the beam discovery and tracking a major challenge. In this direction, innovative neighbor discovery strategies that leverage the full antenna radiation diagram as well as new receiver-initiated medium access control protocols have been proposed [C6-67]. At the network layer, multi-hop relaying and routing strategies are needed to support mobile ad-hoc THz networks, especially in outdoor applications. Cross-layer solutions that capture the trade-offs between antenna beamwidth, communication distance, available bandwidth, processing overhead and buffer capacity are needed [C6-68], and the use of different forms of machine learning (ML) can help to operate such networks. 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.

THz phased arrays need to combine agile beamforming and multi-user capabilities. Techniques inspired from lower bands such as hybrid/MIMO beamforming will need to be revisited to take the specificities of the wireless channels and hardware implementations. Specifically, the inter-element spacing becomes so small that the antenna array and interconnect structure must be taken into account in the overall system design. Thermal analysis is also critical, given the small dimensions.

To support the development of the field, new experimental platforms and simulation tools are needed. For the time being, the majority of the experimental platforms developed to date are channel sounders or physical layer testers that rely on non-real-time DSP, and mostly at sub-THz frequencies, but these real-time platforms become a must for testing of anything beyond a point to point link [C6-69].

6.4.1 Research challenges

Research Theme	**Millimeter-Wave and Terahertz Communication**		
Research challenges	Timeline	Key outcomes	Contributions/Value
Development of THz transceivers	Short- to long-term.	Highly efficient THz transceivers with small footprint.	A requirement for any practical THz system.
Programmable, highly directional THz antenna systems	Short- to mid-term.	Smart THz antenna systems (e.g., programmable arrays and intelligent reflecting surfaces) that overcome the high path loss of THz bands.	Improvement of coverage and link robustness.
THz propagation	Short-term.	THz channel models that accurately describe wave propagation in the THz band, across scenarios.	Guide the development of physical layer solutions and network infrastructure deployment.
Radio methods for THz bands	Short- to mid-term. To be specified in 6G standard.	Modulation and coding schemes, beamforming and wavefront engineering techniques, synchronization scheme that take into account the THz band characteristics.	Improvement of spectral efficiency and, thus, bit-rates.
Novel networking protocols for THz bands	Mid- to long-term. May be specified in 6G standard.	New protocols that capture the nature of THz communications and the THz channel.	Reduce latency, improve end-to-end connectivity.

6.5 Massive MIMO

6.5.1 Ultra-massive and extreme 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. As mentioned in Section 6.4, high-gain directional antenna and lensing systems are needed to communicate over distances beyond a few meters.

Similarly, as in lower frequency communication systems, antenna arrays can be utilized 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. Spatial multiplexing can also be used in a point-to-point line-of-sight link, which is called LoS-MIMO. A special case of LoS-MIMO, which is based on concentric uniform circular arrays is called orbital angular momentum (OAM). In the last decade, the concept of massive MIMO (mMIMO) was introduced and heavily studied in the context of 5G systems [C6-70][C6-71][C6-72]. In such schemes, very large antenna arrays with tens to hundreds of elements are utilized 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 [C6-73][C6-74]. In addition, MIMO DPD (digital predistortion) can be used to handle the varying nonlinearities in the power amplifier related to a varying impedance mismatch between the power amplifiers and the antenna array due to mutual coupling [C6-75]. Very large antenna arrays have been proved to be very useful for mmWave communication systems [C6-76][C6-77]. These enlarged arrays can be centralized or distributed, giving birth in the latter case to the cell-free massive MIMO concept.

When moving to the THz-band, antennas become even smaller and more elements can be placed in the same area. Beyond refining mMIMO solutions with more antennas, new solutions that can manipulate radiation in space and frequency in unprecedented ways are needed. This is how ultra-massive MIMO (umMIMO) enters the game. The concept of umMIMO communications, enabled by very dense plasmonic nano-antenna arrays, has been recently introduced in [C6-78] and [C6-79]. Instead of relying on conventional metals, nanomaterials and metamaterials can be utilized 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 a plasmonic wavelength. Such plasmonic nano-antenna arrays can be utilised both at the transmitter and the receiver (1024×1024) as well as in reflection (in the form of RIS) to simultaneously overcome the path 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 [C6-78], different operation modes can be adaptively generated. In *ultra-massive beamforming*, all the nano-antennas are fed with an amplitude- and phase-manipulated version of the same single plasmonic feed. In *ultra-massive spatial multiplexing (um-SM)*, different plasmonic signals are sent through physically or virtually grouped array elements to communicate with different users. Obviously, any combination in between UM Beamforming and UM Spatial Multiplexing is possible. In addition, to maximize the utilization of the mmWave- and THz-channel and enable the targeted Tbps-links, more than one spectral window could be utilized at the same time. In this direction, *multi-band umMIMO* enables the simultaneous utilization of different frequency bands by leveraging the electrically tunable frequency response of graphene-based plasmonic nano-antennas. 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 umMIMO systems to exploit all of the spatial, temporal and frequency diversities, and hence, promise to yield remarkable performance improvements. In general, there are still considerable challenges in terms of cost, implementation complexity and efficiency. Besides the challenges related to the plasmonic nano-antenna array technology, the realization of any kind of umMIMO communication, even at lower frequencies, requires the development of novel accurate channel models able to capture the impact of the very large dimensions of the array, where spatial non-stationarities emerge [C6-80]. Similarly, ways to estimate and equalize an extremely large number of parallel broadband channels are needed. In these cases, the use of ML-driven approaches might be the solution. Recently, the term of holographic MIMO [C6-81] has been used to refer to structures with capabilities very similar to that of umMIMO.

Exploiting very large arrays with ultrawide bandwidth results in extreme MIMO configurations, where conventional assumptions used in beamforming, such as far-field and narrowband, do not hold. This leads to beam squinting, where the beam direction is frequency-dependent, and grating, where multiple main lobes are formed, causing interference. Additionally, with a very large-scale array, it is likely that communication occurs in the near field, and thus conventional far-field approaches might lead to degraded performance. Solutions such as true-time delay (TTD) beamforming and advanced signal processing algorithms should be investigated with the objective of enhancing performance at low cost, considering various architectures and design techniques.

6.5.2 Distributed and cell-free massive MIMO

Network densification is one of the solutions to achieve the high data rates targeted for 5G and beyond [C6-82]. The antennas of such massive MIMO (mMIMO) 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 [C6-83] and the latter as the distributed mMIMO [C6-84]. Distributed mMIMO can be implemented with a cell-based approach where the access points (APs) are divided into disjoint clusters and the 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 [C6-85], but unfortunately it did not provide much practical gains [C6-86]. This can be mainly attributed to the considerable amount of backhaul signaling for Channel State Information (CSI) and data sharing resulting from a network-centric approach to coherent transmission [C6-87], whereby the APs in a cluster cooperate to serve the UEs in their joint coverage region. The practical implementation of JT-CoMP was also hindered by other attributes of LTE, such as frequency division duplex operations and a rigid frame/slot structure, which did not allow for effective channel estimation.

The cell-centric approach can be changed to a user-centric one, where the cluster serving a particular UE is to be determined dynamically by choosing the subset of APs closest to the UE. The basic idea of a cell-free system, denoted as resource pooling for frameless network architecture, was proposed and analyzed already in [C6-88], and a UE-centric JT-CoMP scheme denoted “Cover-Shifts” was proposed in [C6-89]. 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 (CF-mMIMO) 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 CF-mMIMO is considered as a promising technology [C6-90] due to its improvements in terms of spectral and energy efficiency, especially for indoor and hot-spot coverage scenarios [C6-91]. Nevertheless, some crucial questions remain open for CF-mMIMO, such as the relevant initial access, power control, distributed processing considering encoding/decoding, resource allocation, channel modelling, compliance with existing cellular standards and prototype design.

The fronthaul linking the different APs and BS is a challenge in itself. The CF-mMIMO concept in itself generates forward and reverse link throughputs on the fronthaul which are an order of magnitude higher than the aggregated user payload throughput. This results in extreme traffic on the CF-MIMO fronthaul when going to large bandwidths (FR2, FR3 and sub-THz).

Another implementation challenge of CF-mMIMO is the synchronization and calibration of the different APs or BSs. Synchronization refers here to clock frequency and timing and carrier frequency and phase. Synchronization errors can result in EVM (error vector magnitude) degradation and multi-user interference which can dramatically impact the PHY performance.

6.5.3 Reconfigurable intelligent surfaces

A new and revolutionizing technique able to substantially improve the performance of wireless communication networks is smartly changing the propagation characteristics of the wireless channel through the use of reconfigurable intelligent surfaces (RISs), or sometimes referred to as intelligent reflecting surfaces (IRSs) or large intelligent surfaces (LIS), which are made of a large number of low-cost reflecting elements able to independently change the amplitude and/or phase of the incident signal so as to achieve specific propagation effects [C6-92][C6-93][C6-94][C6-95]. Since most of the RISs are nearly-passive, they can be fabricated with light weight, low profile, and even made to be conformal to various objects. As a result, they can be easily deployed in a wide range of scenarios such as walls, ceilings, billboards, lamp-posts, and even on the surface of vehicles to support several applications for smart factories, stadiums, shopping centers, airports, etc. Moreover, RISs can be deployed as energy-efficient auxiliary devices, without the need for modifying the hardware configuration of the end-user devices. This offers high flexibility and compatibility with legacy wireless systems. Overall, RIS can be used to improve the coverage, reduce interference levels, and increase system capacity in a power efficient manner. Additionally, they can be employed to increase physical layer security, positioning accuracy and even support sensing and wireless power transfer.

One approach is to have a large number of 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 [C6-95], as well as accurate positioning [C6-96]. Although the RIS is made of antenna elements with very low complexity, its implementation may still involve considerable challenges possibly due to the large number of antennas and the associated circuitry. Radio stripes [C6-97] are another interesting variant, with the antennas placed over a stripe instead of a surface. Similar to other RIS implementations, radio stripes may provide a low-cost implementation to gain capacity and enable cell-free systems [C6-82].

The communication using RISs and/or radio stripes schemes will require advanced, low complexity techniques for the signal separation, as well as new resource allocation spatial aspects (i.e., which antennas/panels are activated for a given user). To further improve performance, disruptive techniques that take advantage of hardware imperfections such as nonlinear and/or memory effects can be employed [C6-98][C6-99].

The large dimensions of RISs and radio stripes (several tens of meters), together with the relatively short communication ranges (tens of meters or even less), leads to near field communication effects, with its inherent potentials and challenges. The channel estimation can be a considerable challenge due to the large number of parameters to estimate [C6-100]. To overcome these difficulties, parametric channel estimation and tracking techniques [C6-101] can be employed, eventually supported by positioning information. Moreover, the impact of various physical-layer performance limitation parameters has not been investigated yet, including the large-scale fading and spatial correlation. An investigation related to these performance deterioration effects is very challenging but is necessary in order to propose appropriate fading mitigation techniques.

6.5.4 UAV-assisted communication

The adoption of UAV (unmanned aerial vehicle) enabled communications in 6G networks introduces transformative possibilities and unique challenges at the physical layer [C6-102]. Leveraging UAVs as communication nodes offers advantages such as dynamic deployment for enhanced coverage, particularly in areas with limited infrastructure or in emergency situations. UAVs can serve as aerial or flying base stations, improving signal strength, connectivity and localization/sensing capability. However, challenges arise due to the dynamic nature of UAVs, requiring sophisticated mobility models, and due to the existence of swarms of UAVs, requiring to apply coordination strategies and rules to optimize the physical layer performance. For the latter one, UAV corridors have been proposed as virtual aerial pathways designated for the safe and lawful operation of numerous drones [C6-103]. Moreover, the propagation characteristics, including the impact of atmospheric conditions on signal propagation, demand careful consideration for reliable communication links. The physical layer design must address the complexities of UAV mobility, interference management, and efficient resource allocation to harness the full potential of UAV-enabled communication and sensing in 6G networks. Balancing the advantages and challenges at the physical layer is crucial for realizing the seamless integration of UAVs into the evolving communication and sensing landscape of 6G. Therefore, UAV-enabled communication is expected to play an important role in emerging distributed MIMO scenarios, in which massive number of service requests are expected for a short period of time, e.g., in crowded events. In these cases, low complexity and power efficient UAV-association solutions will be proposed, targeting to improve the spectral efficiency, without inducing signal processing overhead [C6-104]. In addition, further research is required for investigating the impact of multi-corridor-assisted UAV networks that are deployed in three-dimensional space at varying heights [C6-105].

6.5.5 Near field communication

Near field communication arises when the distance between the transmitter and receiver antennas decreases and becomes smaller than the Fraunhofer distance[27]. In the context of 6G, the driving forces behind near field communication are the use of higher carrier frequencies, and the advent of distributed mMIMO [C6-106][C6-107][C6-108]. Decreasing wavelength that follows from higher carrier frequencies make antennas of moderate physical size electrically large, and distributed mMIMO makes the antenna physically large.

The boundary between near field and far field communications is often set to the Fraunhofer distance $r_F = 2D^2/l$, where D is the largest physical dimension of the antenna and l is the wavelength. The Fraunhofer distance allows for a phase error of 22.5° which may be too large in some applications. If the distance decreases to the order of a wavelength, we enter the electromagnetic reactive near field where the electric and magnetic field can exist independently of each other.

[27] Near-field communication can also refer to a set of communication protocols for communication between devices at very short distances. This will not be considered here.

In far field communication, we can approximate the spherical wave front as a locally planar wave front. Though mathematically convenient, a planar wave can only be steered in angle. In near field communication, we use spherical waves, which allow to take advantage of the curvature for steering in both angle and distance [C6-109]. Additionally, the MIMO antenna configurations can be very different from large uniform linear arrays to distributed antenna elements or subarrays. The communication distances are also usually short, and the user may be surrounded by the antenna system. The electrically large antennas also allow for line-of-sight MIMO, which is in particular useful for fixed settings such as fronthaul and backhaul.

Near field communication provides many research topics ranging from theoretical to practical [C6-110], including:

- Near field channel models must provide sufficient fidelity while being computationally tractable. New models should bridge electromagnetic modelling of antennas and traditional communication channel models to include electromagnetic mutual coupling between antenna elements, polarization, spherical wave propagation, large antenna systems and realistic propagation environments.
- Measurement validation of models and assessment of to what extent simplified models are useful in 6G deployment scenarios.
- Capacity expressions can be derived for near field communication using electrically large antenna systems. Initial work on electromagnetic information theory has provided upper bound for some cases [C6-111]. Theoretical results would provide insight into how far practical systems are from the limit.
- Practical issues of building and deploying distributed antenna systems, e.g. synchronization [C6-112], RF chains, antennas, and current shaping materials.
- Near field research should/can be applied not only to communication but also to positioning and sensing.

6.5.6 Fluid antenna system

Fluid Antenna Systems (FAS) [C6-113] are an evolution of massive MIMO that includes a new degree of reconfigurability of the propagation channel, that of the antenna itself. The term encompasses several emerging topics such as communications systems based on arrays of liquid antennas and transceivers made of radiating elements (not necessarily liquid) that can be assumed to expand a surface in a seamless way and can be selected to optimize transmission/reception.

Early works are describing initial testbeds with liquid antennas (FAS devices in which the main conductor is liquid at room temperature, either made of metallic alloys or ionised water) [C6-114] and channel models [C6-115]. Arrays of liquid antennas are able to reversibly change their shape to modify their radiation parameters, i.e., resonant frequency, polarization, or radiation pattern. This reconfiguration would allow to save costs while enabling the use of dedicated techniques to take advantage of their features. They may also offer an interesting solution to blockages and challenging scenarios thanks to their flexibility and reconfigurability. Further work is needed in this direction to support the development of signal processing and adaptation algorithms that are consistent with the real behaviours that can be expected. Channel estimation and its feedback [C6-116] are key to maintain a limited overhead while fully exploiting the reconfigurability.

The more general concept on FAS relies on the possibility of activating the optimal port to access the spatial opportunity for favourable channel conditions, when there is a multiplicity of antenna elements, not necessarily liquid. Then, there is also the challenge of optimising jointly the selected ports and beamforming when FAS combines with MIMO systems [C6-117].

6.5.7 Research challenges

Research Theme	Massive MIMO		
Research challenges	Timeline	Key outcomes	Contributions/Value
Ultra-massive and extreme MIMO	Mid- to long-term. May be specified in 6G standard.	Construction of (e.g. plasmonic) nano-antenna arrays esp. for THz-band, considering physical constraints of semiconductor and packaging technologies. Approaches to deal with beam squinting and grating, as well as the near field. Implementable MIMO digital predistortion for wideband massive arrays.	Increase in spectral efficiency and throughput. Low electromagnetic field (EMF) exposure. MIMO digital predistortion provides enhanced radio equipment sustainability.
Distributed and cell-free massive MIMO	Mid- to long-term. May be specified in the 6G standard.	Distributed implementations of cell-free mMIMO encompassing a very large number of antennas. Centralized and distributed algorithms for coordinated transmission / reception involving large numbers of users. Real-time estimation and feedback of a large number of channels. Compliance with existing cellular standards, prototype design, etc.	Improved performance in very crowded scenarios with high user-perceived throughput and low energy consumption. Increased area spectral efficiency, and energy efficiency.
Reconfigurable intelligent surfaces (RISs)	Mid-term.	Implementation of large low-cost RIS arrays, incl. advanced techniques for signal separation. Cost-effective deployment. Advanced algorithms for RIS configuration, channel estimation, etc.	Increased energy efficiency, coverage and capacity. Enhanced sensing and positioning capability. Programmable propagation environments, esp. important for dense networks and in security sensitive scenarios.
UAV-assisted communication	Mid- to long-term.	Low complexity communication techniques that can adapt to dynamic environments. Seamless integration of UAV flying BSs into 6G communication and sensing networks.	Dynamically support the coverage in various scenarios, which results in reduced CAPEX and OPEX. Increase the probability of obtaining line-of-sight links, which results in improved communication, localization and sensing.
Near field communication	Mid- to long-term.	Models, measurement validation and capacity of near field channels. Building and deploying distributed antenna systems considering near field radio propagation, with applications to communication, positioning and sensing.	Increased spectral efficiency, and energy efficiency.
Fluid antenna system (FAS)	Long-term.	Channel reconfigurability at the transmitter and receiver side. Increased number of antennas at the devices with the same form factor.	Improved performance with blockage mitigation. Full reconfigurability providing new degrees of freedom. Potential cost and energy efficiency due to the reduction of RF chains.

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 [C6-118]. CP-OFDM divides the bandwidth into several orthogonal subcarriers. Fine time and frequency synchronization are then required to maintain the subcarrier orthogonality. However, strict synchronization is limiting in certain scenarios. For example, sporadic access in internet of things (IoT) and machine-type communications (MTC) requires relaxed synchronization schemes, in order to limit the length of the signaling overhead [C6-119]. Ideally, the massive number of devices could just transmit their messages asynchronously; being only coarsely synchronized [C6-119]. 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 [C6-120]. Therefore, CP-OFDM is no longer viable in such scenarios. CP-OFDM is also sensitive to phase noise [C6-121], which is larger in state-of-the-art oscillators as we move to higher frequencies. Moreover, the performance of CP-OFDM is challenging in scenarios with very high time-variability, which we find in vehicular applications and high-speed trains.

Several waveforms, e.g. filter bank multi-carrier (FBMC), generalized frequency division multiplexing (GFDM), which is also known as cyclic block filtered multi-tone (CB-FMT) [C6-122], universal filtered multi-carrier (UFMC), and filtered OFDM (f-OFDM) [C6-123] may be more suitable since their subcarriers are better localized in the frequency domain, and therefore limit the inter-carrier interference. A good frequency localization may also be beneficial due to other reasons, e.g. sensitivity to phase noise in mmWave and sub-THz bands, required accuracy of frequency-synchronization, 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 localized in the frequency domain [C6-124]. FBMC is orthogonal, performs 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). There are also proposals to use FBMC with QAM modulation, but in a non-critically sampled system like filter bank orthogonal frequency division multiplexing (FB-OFDM). FB-OFDM can be orthogonal due to the non-critical sampling and does not need a cyclic prefix either [C6-125]. It is reported in [C6-126] that by applying DFT spreading to FBMC, complex orthogonality can be restored. 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. GFDM can be orthogonal or non-orthogonal. Non-orthogonality introduces self-interference even if the transmitters are perfectly synchronized. 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. [C6-127][C6-128].

There are further new waveforms, including orthogonal time frequency space (OTFS) modulation [C6-129] that are proposed to deal with the fast time variability of the channel. OTFS 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 [C6-130]. However, to cope with doubly selective channels, OTFS uses 2D channel estimation and equalization, which creates significant overhead, additional signal processing complexity and high latency if long frames are needed: research is needed to improve on these aspects.

Constant envelope OFDM (CE-OFDM) [C6-131] 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. It 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 [C6-132]. However, CE-OFDM typically achieves significantly lower spectral efficiency than conventional OFDM with M-QAM symbols.

The aforementioned waveforms are linear waveforms that can be synthesized and analyzed within the plain OFDM system by introducing precoding in the time and frequency domains [C6-133], and adjusting various parameters, including CP length and spreading scheme. However, the precoding requires advanced equalization to mitigate the inter-symbol interference [C6-134]. Therefore, a generic linear waveform framework can be used to optimize the precoding matrix for certain scenarios and requirements, such as presenting low PAPR, low OOB, low-complexity equalization, and resilience to synchronization errors. This allows for the exploration of various types of alternative low-complexity linear precoding to replace DFT, such as using Walsh-Hadamard [C6-135].

While linear waveforms can provide high spectral efficiency, other non-linear waveforms aim at improving energy efficiency with lower spectral efficiency. This includes index modulation (see Section 6.7 for further discussions on modulation), which generates sparse signals that allocate few locations in the time, frequency, and spatial resource grid. The information is included in the allocated indexes. Such waveforms can also be processed within the OFDM system. Considering the integration within the OFDM system is essential to maintain backward compatibility and reduce development costs.

Especially when moving to high frequencies and bandwidths, as, e.g., the sub-THz bands further discussed in Section 6.4, hardware power consumption of individual components can form potential bottlenecks. For example, the analog-to-digital converters power consumption is projected to have a quadratic power increase when used for bandwidths larger than approximately 300 MHz [C6-136]. As such, moving from bandwidths of 300 MHz to 3 GHz would result in an 100x increase in ADC power consumption. To address this problem, new modulation schemes are required, such as zero crossing modulation (ZXM) [C6-137], which is specifically designed for receivers with 1-bit ADCs by encoding information in time rather than in amplitude. However, for the realization of communication systems using receivers with temporally oversampled 1-bit quantization many practical aspects like optimal information encoding in the ZXM sequences and receiver synchronization have yet to be studied. On the other hand, such time-encoding based modulation is also a suitable candidate for low rate sensor node communication in the Internet of Things (IoT), requiring the highest possible energy efficiency. For this service class, impulse radio-based transmission schemes like IR-UWB are considered a suitable approach. The interoperability of such unconventional waveforms and also multi-user access are for further study.

Prioritizing energy efficiency in cellular networks requires a focus shift away from peak spectral efficiency and peak data rates as main performance indicators. Energy-efficient networks need to be able to exploit the highly variable data rate demand, analyzed, e.g., in [C6-138], to immensely reduce energy consumption. Dealing with highly variable data rates and a wide range of services with divergent requirements calls for highly adaptive approaches. A potential approach for energy-efficient radio access technologies is the Gearbox-physical layer (PHY) concept [C6-139], which aims to improve energy efficiency by dynamically switching between distinct modulation schemes, each supported by optimized parallel analog front ends. By choosing the most energy-efficient modulation scheme and corresponding front end according to the user’s data rate needs and the available spectrum, the Gearbox-PHY has the potential to significantly enhance energy efficiency in wireless communication systems. The specific design of such a Gearbox-PHY needs to be investigated.

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 [C6-140][C6-141][C6-142]. Challenges for NOMA research include

- *User pairing:* With a careful design, more than two users can be paired to use the same resource [C6-143]. 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:* The design of power control in NOMA can affect other performances such as receiver interference level and throughput. E.g. the work in [C6-144], where the power constraint is jointly allocated in full-duplex NOMA, can be further extended to multi-cell scenario.
- *Physical layer security:* 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. [C6-145]).
- *Code-domain multiplexing:* Different users are allocated different codes and multiplexed over the same time-frequency resources. These schemes include multiuser shared access (MUSA), low-density spreading (LDS), and particularly sparse code multiple access (SCMA), which can be potentially combined with other technologies such as mmWave communications or physical layer security and applied in massive MIMO systems [C6-146]. The main challenge is the design of low complexity SCMA systems, an aspect that still requires research work.

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, joint communication and sensing, and cognitive radio applications [C6-147][C6-148]. 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, and including AI-based algorithms that are capable to perform self-interference cancellation in multiple radio frequency bands. Much work remains to be done, and an inter-disciplinary approach will be essential to meet the numerous challenges ahead [C6-148].

6.6.1 Research challenges

Research Theme	Waveform, Multiple Access and Full-Duplex		
Research challenges	Timeline	Key outcomes	Contributions/Value
Waveforms for cmWave, mmWave, THz, OWC and ISAC applications	Mid- to long-term. To be specified in 6G standard.	Waveforms taking into account the hardware properties and impairments that are important for these applications (phase noise, PAPR, power amplifier non-linearity, ICI in asynchronous communications, …). Approaches to deal with increasing energy consumption by optimizing energy efficiency instead of spectral efficiency.	Improved performances in crowded scenarios, high mobility, also contributing to accurate positioning, with reduced energy consumption. Enabling sustainable networks and support growing data rate demand.
Enhanced NOMA	Mid- to long-term. To be specified in 6G standard.	Code design, resource allocation, and receiver algorithms.	NOMA can provide higher capacity, energy efficiency, device density.
Full-duplex transceivers	Mid- to long-term. May be specified in 6G standard.	Broadband full-duplex RF frontends for massive MIMO. Advanced self-interference cancellation (SIC) schemes.	Full-duplex can increase the throughput and enable spectrum sensing or mono-static radar.

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 [C6-149]. 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 parallelization, 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 6G system, the peak data rate can attain 100~1000Gbps, and it is difficult for the legacy 5G LDPC decoder to support the ultra-high data rate since the legacy 5G NR LDPC design is mainly applicable for block parallel decoder, which has much lower throughput capability compared to row parallel decoder or full parallel decoder. Therefore, a new LDPC design which adapts to row parallel decoder or full parallel decoder should be investigated to support the ultra-high peak data rate of 6G systems. Even with full parallel decoder, the current LDPC design is not likely to achieve the 1000Gbps throughput with an acceptable chip area efficiency. Other coding schemes such as polar codes with SC decoding and new LDPC codes deserve further study. The challenge in designing 6G LDPC codes is how to satisfy the combination requirement of ultra-high throughput, acceptable complexity, comparable performance and flexibility to 5G LDPC codes. 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 [C6-150][C6-151] 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 [C6-152][C6-153]. In addition, if polar codes are considered as channel coding candidates for 6G data channel, a new code extension structure should be investigated to support IR-HARQ functionality. A unified channel coding scheme, e.g. by making polar codes good for long codeword length, or LDPC codes good for short codeword length, is desirable to simplify encoding/decoding hardware at competitive error-correcting performances [C6-154].

The Guessing Random Additive Noise Decoding (GRAND) concept [C6-155][C6-156]tries to recover the user info by guessing the additive noise which is most likely to result in the received codeword. This concept works for any code length, can be adapted to virtually any type of coding scheme and has been shown to be capacity achieving, but has a complexity that is only reasonable at high code rates. Adaptation to soft-inputs, concatenated codes (turbo) and other schemes have already been developed. Further effort is needed to reduce its complexity for medium to low code rates.

Delayed bit-interleaved coded modulation (DBICM) is a generalization of BICM by incooperating a bit delay module [C6-157]. It has the potential to improve the performance of BER performance of BICM by trading-off latency and receiver complexity, e.g. with LDPC code [C6-158], etc.

Even though the 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 [C6-159]. In order to approach capacity, the transmitted symbols need to have a certain probability distribution (e.g., a discrete Gaussian distribution is needed for the transmission over AWGN channels with an average power constraint) and using uniformly distributed symbols results in a performance loss, which can be up to 1.53 dB on AWGN channels. Eliminating this 1.53 dB shaping loss can save 30% of the transmit power.

Several solutions for constellation shaping are proposed to compensate this loss. One option is to optimize the locations of the modulated symbols in the constellation diagram to obtain non-uniform constellations (NUC), as adopted in the ATSC3.0 standard [C6-160]. This scheme is also called geometric shaping and shows improvements compared to uniform signaling. Another approach is the so-called probabilistic shaping [C6-161][C6-162][C6-163][C6-164], 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 [C6-162]. However, a unified and implementable modulation/demodulation algorithm of probabilistic shaping should be studied for different code sizes, different code rates and different modulation orders. In addition, 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.

6.7.1 Research challenges

Research Theme	**Coding and Modulation**		
Research challenges	Timeline	Key outcomes	Contributions/Value
New channel coding	Mid-term.	Channel encoder and decoder for 1) extremely high throughput or/and, 2)	Improved throughput, spectral efficiency, and energy efficiency.

	To be specified in 6G standard.	extremely high reliability or/and, 3) extremely low latency or/and, 4) extremely low power consumption, etc. Unified channel coding scheme for 6G, e.g. by making polar codes good for long codeword length, or LDPC codes good for short codeword length.	Support 6G KPIs, e.g. ultra-high data rate ranging from 100-1000 Gbps, reliability of $1\text{-}10^{-7}$, etc. Simplified encoding / decoding hardware.
New modulation	Mid-term. To be specified in 6G standard.	Modulation and coding scheme trading-off latency and receiver complexity. Advanced modulation and coding scheme with signal shaping loss removed.	Improved spectral efficiency, and energy efficiency.

6.8 Integrated Sensing and Communication

In today's highly interconnected world, especially in the realm of the 'Internet of Things' (IoT), understanding the locations of various 'things' becomes increasingly critical. This is often achieved through positioning and sensing. Upcoming mobile radio systems are expected to play a crucial role in delivering precise positioning for these 'things'. In cellular networks like 4G LTE, positioning involves several base stations (BSs) transmitting reference signals to the user equipments (UEs) in the downlink. Conversely, in the uplink, the UE sends a reference signal to nearby base stations. This method sufficiently meets the Federal Communications Commission's requirements for localizing emergency calls, known as E-911, which demands accuracy of about 50 m [C6-165]. However, the rapid technological evolution has spurred demand for significantly higher positioning accuracy in various applications, such as V2X and smart factories. For example, discovering vulnerable road users (VRUs) in vehicular scenarios requires positioning accuracy as fine as 10 cm [C6-166]. Currently, the 3GPP is exploring positioning accuracies below 20 cm in some cases, utilizing 5G's higher frequencies, larger bandwidths, dense deployments, and device-to-device communication capabilities. Notably, the current 5G NR standard's wireless positioning is limited to locating UEs capable of communication. To monitor physical conditions of environments or objects, radar-like capabilities, referred to as 'sensing' are needed. Integrating communication with sensing is seen as a promising approach in this context. Indeed, traditionally, sensing and communications have been separate functions, often using different entities or frequency bands [C6-167].

The next generation of mobile radio systems is expected to be designed for simultaneous communication, positioning, and sensing. These systems will likely leverage the sensing capabilities of radio frequency (RF) signals in the mmWave and THz bands. The enhanced connectivity and bandwidth provided by 6G technology will enable cooperative devices to use data fusion strategies. These strategies can accurately determine the positions of passive targets in various applications, including traffic and vehicle monitoring, pedestrian detection, and collision avoidance between autonomous guided vehicles. Additional applications include assisted living, accurate positioning, tracking of passive objects, and human-machine interfaces. Sensing can vary from simple object detection to determining position, speed, and specific micro-Doppler signatures, even up to environmental imaging. In addition, sensing will also be able to detect object size, shape, material characteristics, motion state, presence or proximity of adjacent objects, etc. Therefore, sensing is also expected to play an important role in high-precision imaging and environment reconstruction. These capabilities will be used for industrial automation, Internet of Things, V2X, smart homes, public safety, medical care, smart cities, etc. As a result, there is a growing demand for systems that combine sensing and communication capabilities.

Integrated sensing and communication (ISAC) has the potential to transform not only point-to-point communications, like vehicular networks, but also complex mobile/cellular networks. It could revolutionize current communication-only mobile networks. The full integration of communication and sensing functionalities aims to maximize the efficiency of spectrum usage and minimize resources (hardware, energy) in performing both functionalities. Therefore, high-accuracy sensing without weakening wireless communication will be indispensable for future networks, including short-range communication.

ISAC at mmWaves is poised for significant advancements, leveraging multiple antennas for both transmission and reception [C6-168]. Employing THz technologies could yield enormous gains in sensing resolution and accuracy, especially in short-range applications. For UE and passive object positioning within 10cm, 3+ GHz of phase-coherent bandwidth are needed. Carrier aggregation can be used to help make such large equivalent bandwidth available (even without necessarily involving mmWave and THz bands) by stitching together multiple sub-bands that are not necessarily contiguous [C6-169][C6-170]. In order for the aggregated band to be phase-coherent, each sub-band needs to have a fixed phase relation to other sub-band during transmit/receive time.

The use of large-scale antenna arrays, made feasible by shorter wavelengths and larger bandwidths compared to traditional cellular bands, offers substantial benefits for both communication and sensing [C6-171]. Moreover, MIMO technology can deliver high-capacity links to users, such as through spatial multiplexing, while array processing at the sensing receiver can provide accurate direction of arrival (DoA) estimation. Large antenna arrays at the BS result also 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 causes performance degradations and reduces position accuracy. Such methods can additionally leverage the presence of large antenna arrays at the BS [C6-172]. Clearly, having multiple antennas at the UE can improve positioning. 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, which can also assist to develop simultaneous localization and mapping (SLAM) schemes [C6-173]. Note that SLAM may lead to high-complexity for the UE due to iterative nature of the algorithm; hence, environment sensing by network may simplify the process for UE by providing the information about dominant paths which can be used for UE positioning without requiring complicated processing at UE. Such radar-like (sensing) abilities can occur in either bistatic operation (piggybacking on standard positioning reference signals) [C6-174][C6-175], or in monostatic operation (requiring full-duplex processing at the BS) [C6-176]. However, fully harnessing these physical dimensions will require novel signals tailored to fully exploit temporal, spatial, and frequency domain properties [C6-177][C6-178][C6-179].

Cooperation significantly enhances positioning accuracy, especially in scenarios with massive connectivity [C6-180][C6-181]. In cooperative positioning, the User Equipments (UEs) are capable of sending, receiving, and exchanging position-relevant information. A high density of UEs often results in Line-of-Sight (LOS) propagation conditions amongst multiple UEs, markedly improved localization accuracy and coverage. Additionally, cooperative sensing is advantageous as it allows observation of an object from various perspectives, thereby introducing a form of diversity. This diversity not only strengthens robustness but also increases the probability of detection and localization [C6-182][C6-183]. The advent of side-link communication in 5G opens new avenues in cooperative positioning and sensing, encompassing not just signal design but also advancements in protocols and algorithms. These developments are particularly crucial in vehicular contexts and Unmanned Autonomous Vehicles (UAVs), where cooperative link-based relative location information plays a key role in ensuring safety and enhancing global situational awareness.

Accurate positioning and sensing are pivotal in allowing sensing-assisted communications [C6-184]. This enables, e.g., the use of narrow beams aimed at the intended user or the detection of blockages that could disrupt communication links. Furthermore, sensing is instrumental in facilitating sensing-assisted positioning through environmental mapping and identifying non-line-of-sight (NLOS) situations, which are detrimental to positioning accuracy. The significance of these capabilities will intensify as communication systems advance to higher carrier frequencies, notably beyond 0.1 THz. Therefore, exploring new paradigms, such as those integrating communication, positioning, and sensing, is vital. This integration promises to enhance spectral and energy efficiency and reduce latency. Note that in environments with complex propagation characteristics, where traditional modeling is insufficient, data-driven methods, like those based on machine learning, are particularly useful for effective positioning and sensing.

Such more fine-grained sensing tasks often require accurate micro-Doppler extraction through mono-static or multi-static sensing. While the former is well known from the field of radar-sensing, the latter, while more amenable to 5G and 6G deployments, is impaired by the fact that network nodes are asynchronous, i.e., they have different clock sources and oscillators for the Radio Frequency (RF) front-end [C6-185]. This asynchrony causes a time-varying Timing Offset (TO) and Carrier Frequency Offset (CFO). The former appears as a common delay shift for all propagation paths in the Channel Impulse Response (CIR), preventing the correct estimation of actual path delays. The latter is a random frequency shift that destroys phase coherence across subsequent packets, hindering the estimation of the Doppler shift and of the micro-Doppler effect. The micro-Doppler is a frequency modulation of the reflected signal around the main Doppler frequency induced by movements of a target or target components. It is the main signal feature used in fine-grained wireless movement sensing, and has a vast number of applications in target classification, human activity recognition, pervasive healthcare, and person identification, among others [C6-186][C6-187][C6-188]. While existing communication systems use algorithms to compensate for TO and CFO, they treat the sensing parameters (delay and Doppler shift) as part of the undesired offsets and remove them [C6-189]. This makes such existing techniques unfit for sensing purposes, as delay and Doppler are the channel parameters that describe the physical environment. For these reasons, clock asynchrony is the main obstacle to large-scale practical implementations of integrated sensing and communication systems [C6-190], where TO and CFO due to clock asynchrony are to be removed while retaining the delay and the Doppler shift. Potential solutions leverage the correlation in the channel propagation paths and identify static reference paths whose phase can be used to remove frequency offsets from the reflections on sensing targets [C6-191].

Integrated sensing and communication is currently gaining more and more attention by paving the way for a new plethora of services offered by future mobile networks. However, its development poses several challenges, including integrated waveform design, integrated baseband and hardware design, sensing algorithms, multi-band sensing technology cooperation, fusion with other sensing and positioning technologies, computational requirements, etc.

The network infrastructure can be exploited as a large-scale radio sensor, enabling continuous sensing, especially during periods of low communication load. This capability allows the network to gather detailed environmental information. Over time, it can accumulate high SNR data about the static environment, known as clutter, and generate accurate environmental maps. As measurement time increases, the SNR for clutter detection continually improves, leading to precise clutter maps. These maps facilitate high-accuracy target detection by subtracting the clutter from the signal; the resulting data contains information about the target plus noise. This method enhances the network's sensing capabilities, significantly improving the effectiveness of radar sensing-as-a-service.

6.8.1 Research challenges

Research Theme	**Integrated Sensing and Communication**		
Research challenges	Timeline	Key outcomes	Contributions/Value
Integrated waveform design	Mid-term. To be specified in the 6G standard.	New waveform flexible to accommodate communication, positioning, and sensing. Mono-, multistatic sensing supported. Resource allocation and optimal transmission parameters for sensing, positioning, and communications.	Enabler for 6G systems with unprecedented sensing and positioning capabilities. Numerous applications, including safer cities and workplaces.
Multi-band sensing technology	Mid-term. May be specified in the 6G standard.	New technology solutions for sensing at different frequency (e.g. FR3, sub-THz and THz) bands, and/or carrier aggregated bands.	Use of network infrastructure for sensing and positioning. E.g. a new network of cross-sector competences between the ICT industry and the radar / sensing industries.
Distributed and cooperative sensing	Mid- to long-term. May be specified in the 6G standard.	Methods and solutions for distributed sensing with data fusion capabilities. Use of AI for data fusion, object recognition, and environment mapping.	Full exploitation of network resources: pervasive deployment, distributed computation, backhaul and core infrastructure. Enabler for the perceptive network paradigm and key element for mapping the physical world into the digital one.
Sensing aided communication	Mid- to long-term. May be specified in the 6G standard.	Sensing and communication methods with/without full-duplex. Methods to exploit the radar information for communications (e.g., channel estimation).	Improved spectral/energy efficiency, throughput and sensing accuracy.
Sensing aided localization	Mid- to long-term. May be specified in the 6G standard.	Methods to exploit the sensing information for localization (e.g., channel estimation, NLOS detection).	Improved UE positioning accuracy in harsh environments.
Sensing aided environmental reconstruction and clutter detection	Mid- to long-term.	High-precision radio imaging and environmental (incl. clutter) map.	Enhancement of communication, positioning, sensing and radar sensing-as-a-service.

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 [C6-192]. 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. As there is no scheduling of the transmission resource by the base station, the massive random access (MRA) is contention-based. In contention-based MRA, the collisions of multiple packets in the same slot are inevitable. To solve these collisions and support a MRA of high user loading, non-orthogonal multiple-access (NOMA) techniques should be considered. NOMA has been well researched in grant-based schemes. However, in MRA, the transmission is grant-free, the global power control, resource allocation and configuration cannot be used, which poses a challenge to deal with inter-user interference (IUI). The one-dimension discrimination of power domain brought by the near-far effect of MRA is not enough to deal with severe IUI. Therefore, higher-dimension domains like code domain and spatial domain should be introduced. In code domain MRA schemes, the transmitters randomly select their non-orthogonal spread codes [C6-193]. At the receiver side, the codes are detected and used to alleviate IUI. The prior knowledge of the statistic properties of data (e.g., constellation shape), codebook, and CRC result should be fully utilized for advanced blind detection [C6-194].

Spatial domain is an effective way to increase the spectrum efficiency. Although the orthogonality of the spatial domain cannot be guaranteed in MRA transmissions, it is still very efficient as multiple receive antennas increases the degrees of freedom without extra resource consumption. However, using conventional transceiver to acquire spatial degrees of freedom is very challenging in MRA transmission. As there is no coordination of the transmission resources by the base station, active UEs in MRA autonomously select pilot sequences from the predefined pilot sequence set. Inevitably, multiple active UEs may select the same pilot sequence, which is called 'pilot collision'. For a given pilot set, the probability of pilot collision increases rapidly with the increase of the number of active UEs. Pilot collision will lead to miss detection and inaccurate channel estimation of collided UEs, which severely degrades both the suppression of IUI and the compensation of channel distortion experienced by the data symbols. Moreover, considering the extremely simple transmit procedure of MRA, the received signals could experience large time offsets (TO) and frequency offset (FO) : 1) Due to the lack of uplink timing alignment/timing advance (TA) procedure, the transmitted signals of active UEs may arrive at the BS with different time delays, with each UE's delay being determined by its distance to the BS, thus the received signals from the UEs near the edge of the cell would experience large TOs; 2) Due to the lack of tight frequency synchronization, the oscillator misalignment and Doppler effect could cause a large FO. Large TO/FO will further increase the symbol distortion on the basis of distortion induced by wireless multi-path channels, which makes the channel estimation and symbol demodulation more difficult. As a result, it's very challenging for the multiuser detection (MUD) of MRA transmission as it will encounter not only heavy IUI and severe distortion on the received symbols, but also uncontrollable pilot collision. To achieve a better MUD performance for MRA, different transceivers have been proposed. One solution are data-driven methods not relying on pilots via blind receive beamforming and blind equalization [C6-194][C6-195]. Another way is enhancing the pilot design to reduce pilot collisions, for example, multiple independent pilot scheme [C6-196] and extremely sparse pilot scheme [C6-197] can be used.

In MRA, 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 [C6-198] and [C6-199]. 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 [C6-200]. The performance can be further improved using low-rate channel codes in combination with multi-user detection at the physical layer [C6-201]. 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 [C6-202].

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 [C6-203] 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 [C6-204] and [C6-205]. Compressed sensing-based multi-user detection may also be combined with coded random access schemes [C6-206].

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 [C6-207], 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 [C6-208].

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.

It is interesting to notice that 3GPP has already started considering low-latency unsourced random access under the term ``2-step RACH'', in contrast with the conventional 4-step RACH. In 2-step RACH, random access users sent a short preamble (pilot sequence) chosen at random and without coordination with other users in a codebook of possible preambles, followed by data in a subsequent time-frequency resource block called a ``PUSCH opportunity'' (PO). Each preamble points to a given PO in the next slot, such that if the base station detects a preamble, then it knows that the corresponding PO is used by some random access user to send a payload packet. The 2-step RACH scheme is conceptually very similar to schemes studied in the recent communication theoretic and information theoretic literature, which also assume that random access users transmit codewords from a common codebook (preambles) without any coordination, and the goal of the receiver (base station) is to decode the list of ``active codewords'', even not knowing who is transmitting them. Codewords are ``token'' that prepare the base station to the immediate subsequent reception of payload data. In this respect, it is very interesting and relevant to see that the current standardization is being influenced by the recent theoretical research on MRA.

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 [C6-209]. 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 centralized control systems. Similar mechanisms will also be required for evolved Industry 4.0 applications [C6-210]. 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 [C6-211].

6.9.1 Research challenges

Research Theme	**Massive Random Access**		
Research challenges	Timeline	Key outcomes	Contributions/Value
Code design and NOMA for random access	- Mid-term. - To be specified in 6G standard.	- Codes for which a superposition of codewords can be uniquely decoded. - Advanced receiver algorithms to resolve packet collisions for contention-based access.	Improve the area spectral efficiency and reliability by exploiting interference.
Synchronization, channel estimation, and beamforming for random access	- Mid-term. - To be specified in 6G standard.	- Pilot sequence design and autonomous pilot selection schemes. - Channel estimation without tight time/frequency synchronization. - Blind beamforming and equalization algorithms.	Reduce energy consumption and support mobility by relaxing the synchronization requirements, which enables deep sleep modes of inactive devices.
User activity detection	- Mid-term. - May be specified in 6G standard.	- Algorithms to determine the set of active users based on group testing or compressed sensing. - Joint activity and data detection. - AI/ML techniques to exploit correlated activity patterns.	Efficient user activity detection is crucial for grant-free transmission with high device density.

6.10 Machine Learning Empowered Physical Layer

Application of artificial intelligence (AI) and machine learning (ML) in communication systems spans a wide range from optimizing a specific function to end-to-end learning of the entire communication system [C6-212]. AI/ML can also be used to simplify network operations and reduce the need for human participation and supervision. With proper use of AI/ML, the system can automatically address the vast majority of network anomalies and only require human intervention in a small number of extremely unusual cases.

For these reasons, ML techniques are expected to redefine the classical approach in communication system design to achieve global optimization or performance improvement. 6G is expected to be a large-scale and self-organized system that integrates terrestrial and non-terrestrial networks to provide seamless wireless connectivity elsewhere, and ML techniques can play a meaningful role to develop this concept with two different viewpoints. A block-structured strategy focuses on the application of a specific ML algorithm for each block of the communication systems to optimize the individual performance of each block. This methodology inherits the advantages of former models and algorithms. For example, channel estimation can be studied as a regression problem and the selection of modulation and coding scheme (MCS) can be learnt from an exhaustive exploration of the environment [C6-213]. A second perspective addresses end-to-end (E2E) communication and targets at the optimization of the complete communication system, from the transmitter to the receiver. An E2E ML approach may consist in the representation of both blocks as black boxes and the characterization of the transmitter as an autoencoder that can infer MCS through data-driven analysis [C6-214]. ML techniques applied to spectrum/environment awareness may assist in the adoption of both strategies given the scarcity of tools to face 6G challenges such as 3D transmission models, the adoption of new frequency bands (THz), the introduction of new network elements (e.g. RIS), beamforming in ultra-massive MIMO communications, and non-identified interference sources.

The application of ML techniques should be guided by the fact that they do not in general lead to the optimal performance. ML techniques are only advantageous if the considered signal processing problem is nonlinear by nature. Moreover, in case knowledge of the underlying model is available it is always advantageous to use this information and incorporate it into the applied machine learning model.

Regarding IPR and standardization the use of ML techniques provides some challenges and open issues. So far it is unclear if a specific algorithm will be standardized for conformance testing, limiting the application for IPR. Thus, different options regarding the standardization and IPR protection on the application of ML techniques in the physical layer need to be investigated.

In the following, we describe some of the topics where we expect AI/ML research to be important to improve the physical layer design and performance.

- *Hardware:* The dominating hardware architectures for AI/ML algorithms are CPUs and GPUs, but these are not optimal for real-time physical layer algorithms. So-called in-memory computing is a more promising hardware solution for real-time physical-layer ML solutions. Moreover, inspired by the structure and energy efficiency of brains, new neuromorphic hardware architectures have been devised and with them, pulse-based – spiking – neural networks. These and other new architectures and methods to realize learning algorithms should be investigated.
- *Hardware impairment modeling:* In low-cost devices and higher operating frequencies, non-linear effects and other impairments are more pronounced. Complementing the hardware models that exist, AI/ML should be used to compensate for the performance loss that would result if these impairments are not compensated for. Here it is relevant to establish a useful trade-off between algorithmic complexity and hardware simplicity. 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 at a performance loss. Examples include maximum likelihood detection, channel estimation, etc.
- *Overall physical layer:* At the highest level, we can consider the entire physical layer transmitter-receiver chain as an auto-encoder. While promising, it does not make use of the vast knowledge established during a century of communications research. Therefore, a more feasible approach could be to learn only parts of the physical layer, while still training the whole link in an end-to-end manner. Practically, we believe that the greatest benefits can be reaped where there exists a model deficiency or large variations in individual units. The model deficiency can manifest itself in reality being too complex to model with sufficient fidelity at an affordable effort, e.g., some radio propagation channels. Individual variations are expected in e.g., low-cost hardware for IoT or distributed massive MIMO.
- *Channel learning:* The wireless channel can be complex and rapidly changing, in particular as we are progressing to even higher carrier frequencies and into the light spectrum. At higher radio frequencies, communication is often beam-based. To learn the wireless channel and be able to correctly predict time, frequency, and spatial properties allows for performance improvement and overhead reduction. Moreover, massive MIMO and RIS-aided communications require new approaches based on AI/ML as

classical channel estimation is unable to cope with such extremely high complex problems, particularly in multi-RIS systems where the communication path can be supported by more than one RIS.

- *Radio interface design:* AI/ML has the potential to design new signal waveforms or modifying existing signals. The signal constellation is often designed or selected based on the prevailing channel conditions. It has been shown that learning the signal constellation can give improvements both in performance and in control signaling reduction. AI/ML can be used to design schemes beyond QAM, such as fully pilotless waveforms, and find new modulation types for THz carrier frequencies that can be particularly useful for ISAC scenarios. Other functions in the transceiver chain that have received considerable interest include FEC design and decoding. At the receiver, AI/ML can compensate for the loss in performance when operating under non-ideal conditions, e.g., frequency offset, colored noise, interference, cross-talk, etc.

- *Multi-antenna systems:* MIMO, massive MIMO, and distributed MIMO offer rich opportunities for AI/ML research. Current MIMO systems use precoders and beam pair search procedures. The antenna arrays are assumed to be uniform and the RF chains equal. AI/ML can be used to optimize precoders for non-ideal conditions, where the previously mentioned assumptions do not hold. As the number of antenna elements increases and the deployments go from regular arrays to almost random positions, the need for AI/ML algorithms increases even further. Fully digital beamforming requires independent, affordable RF chains. This may lead to increased hardware impairments and variations between individual RF chains. AI/ML solutions would be desirable to compensate for this since manually measuring and calibrating such systems would be prohibitively complex. Moreover, an important aspect in large-scale multi-channel transmission scenarios is the optimization of the energy efficiency. To this aim deep reinforcement learning and/or federated learning approaches are expected to balance the trade-off between the power consumption and achieved throughput [C6-215]

- *Learning over the air:* The wireless medium mixes signals from different sources. This is often a source of nuisance since it causes interference. However, if it can be controlled, it becomes possible to use this to perform “learning over the air”. It should be further investigated how reflective surfaces can be used to create a virtual ML model.

- *Performance vs. resource trade-off:* Many network nodes are energy constrained, e.g. when they are battery-powered or the heat dissipation should be limited. Frequent retraining of large AI/ML models may contribute negatively to the sustainability of future systems. Thus, an important high-level topic is the trade-off between performance and resource/energy use or AI/ML algorithms.

- *Physical layer compression:* AI/ML applications will introduce new data types for air-interface, such as 1) AI model; 2) intermediate feature; 3) inference result, etc. These data are normally of high-dimensionality, containing tens of millions of parameters/coefficients. Transmitting these data gives rise to extensive channel overhead, thereby further affecting latency. To tackle this issue, it is highly desired to develop *physical layer compression* techniques for reducing the amount of transmitted data while warranting AI/ML performance. Further benefits of physical layer compression include: 1) *Privacy protection* - Relevant native data are generated within the RAN, and the nodes that receive and process data are also UEs, BSs, etc., inside the RAN. Neither the original data nor the intermediate results of processing need to go through application layer, which leads to a better data privacy protection. 2) *End-to-end delay reduction* - Compressing and processing native data at physical layer can reduce the delay caused by segmenting and multiplexing data packets at the protocol stack. And the receiver can decompress relevant data in a timely manner to facilitate AI tasks. 3) *Reduced amount of data transmitted across layers* - When the AI data are compressed at upper layer, they need to be transmitted to application layer for compression and decompression before returning to physical layer, giving rise to a lot of overhead to the internal transmission of the protocol stack. Compression at physical layer can avoid this kind of cross-layer data transmission. 4) *Power saving* - Physical layer compression can completely eliminate the power consumption of cross-layer transmission and protocol stack for segmenting and multiplexing of data packets.

- *In-radio network AI computing:* Model training and model interference of AI/ML methods can lead to large amount of computations. Computation efficiency is important to apply AI/ML in radio networks. The computation tasks can be performed by multiple nodes in the network, e.g., jointly by BS and UEs, then the data, parameters of AI/ML models, and outputs of AI/ML models need to be exchanged among network nodes. In this case, communication and computation can be jointly designed. Technologies like over-the-air computing (AirComp) [C6-216] or coded computing [C6-217] can be considered to improve the efficiency or reliability of the computing in radio networks.
- *Physical layer security:* Ubiquitous access and a massive number of low-cost devices will be key features of 6G and, at the same time, will substantially increase the number of potential threats and cyberattacks to the reliability and security of both users and networks. In particular, a major concern is that, due to the propagation medium's broadcast nature, wireless transmissions are exposed to attacks that may require very unsophisticated techniques, e.g. a jamming attack just needs a very powerful transmitter. To guarantee security in wireless networks, two approaches emerge to provide secure wireless communication in addition to network security mechanisms. Physical layer security (PLS) is driven by the exploitation of the physical characteristics of the wireless channels to combat jamming and eavesdropping attacks [C6-218]. While PLS techniques have traditionally leaned on artificial noise generation or diversity, a new wave of ML techniques supported by massive MIMO or full duplex mmWave can face the new security challenges on the radio interface. On the other hand, ML techniques may exploit spectral and signal analysis to detect radio attacks. This paradigm may extend the utilization of signal processing and ML to the detection of jamming, eavesdropping or rogue base stations. Mechanisms for the generation of security keys based on the reciprocity of the uniqueness of the wireless channel between the two communication ends, known as Physical layer Key Generation (PKG), can greatly benefit from the AI/ML approach in complex wireless environments (e.g. multi-hop massive MIMO systems) and high frequency communications where channel models are still not well known.
- *Trustworthiness of AI/ML algorithms:* Topics in this area include explainable AI/ML, uncertainty outside the training distribution, and spoofing. Compared to model-based algorithms, an issue with current-day AI/ML algorithms is that they appear as black-box solutions and their intermediate states cannot (always) be interpreted in a meaningful manner. Explainable AI and how to incorporate existing model-based knowledge in training can bring increased transparency and trust in the models. Mathematical models can be extrapolated to understand their asymptotic behavior. For AI/ML algorithms we cannot do so when the input is far from their training set distribution. Methods should be developed to not only allow AI/ML models to make accurate predictions but also know when they are operating far from their training distribution. Many AI/ML algorithms can achieve super-human performance on e.g., image recognition problems. However, it has been shown that AI/ML algorithms can be tricked into misclassifying images when a noise pattern, imperceptible to humans, is added to the original image. Designing AL/ML algorithms robust to unintentional and intentional spoofing attempts is increasingly important as AI/ML algorithms enter 6G systems.

6.10.1 Research challenges

Research Theme	**Machine Learning Empowered Physical Layer**		
Research challenges	Timeline	Key outcomes	Contributions/Value
AI hardware architecture	Mid-term.	- New architectures and methods to realize real-time learning algorithms.	Improved hardware architectures.
AI-aided hardware impairment modeling	Mid-term.	- Compensate for the performance loss due to non-linear effects in low-cost devices by complementing the hardware models using AI/ML.	Reduce the complexity.
Reduce complexity in neural network modeling of	Mid- to long-term.	- Applying intelligent histogram-based and alternative training data selection mechanisms, together with feature selection and feature extraction techniques can	Reduce neural network training times by two orders of magnitude to enable faster adaptivity, reduce complexity by one order of

multidimensional nonlinear effects		contribute to reduce the dataset dimensionality. The latter can also be applied to prune the neural network and reduce its complexity for multidimensional nonlinear problems.	magnitude and power consumption by a factor between 2 and 5.
Overall physical layer	Mid-term. May be specified in 6G standard.	- Using AI to model scenarios/systems with model deficiencies or mismatches such as communication with non-uniform antenna arrays.	Improved spectral efficiency and throughput.
Radio building blocks	Mid-term. May be specified in 6G standard.	- Using AI to design modulation schemes beyond QAM, pilotless waveforms, beam search procedures in massive MIMO systems, and enhanced channel prediction mechanisms.	Improved spectral efficiency and throughput.
Enhance RAN adaptivity and intelligence without increasing complexity at radio unit (RU) end	Mid- to long-term.	- Transition from RU-centric to cloudified RAN AI-driven PHY building blocks.	Leverage on edge efficient and sustainable power supplies and higher end specialized heterogeneous computing hardware to shorten training times by one additional order of magnitude and simplify the RU architecture.
Physical layer compression	Mid-term. May be specified in 6G standard.	- Coding methods and steps to compress different types of data, such as AI model, intermediate feature, inference result, etc.	Reduces transmission overhead while ensuring AI/ML task accuracy.
In-radio network AI computing	Mid- to long-term. May be specified in 6G standard.	- Apply in-radio network computing techniques to improve the efficiency or reliability of the computing tasks.	Improve the computing efficiency of AI/ML algorithms or reduce the latency of deploying AI/ML models.
Physical layer security enhancement	Mid-term. May be specified in 6G standard.	- Apply ML based techniques to enhance physical layer security schemes to combat the increased number of potential threats and cyberattacks for securing both users and networks.	Increase security risks against cyber attacks.
Trustworthiness of AI/ML algorithms	Mid-term.	- Increase transparency and trust in the models by utilizing ‘explainable AI’. Robust against unintentional and intentional spoofing attempts.	Gain human trust in the network by making the algorithms transparent.
Standardization and IPR of ML algorithms	Mid-term.	- Guidelines regarding the standardization of physical layer components using ML techniques and approaches enabling IPR protection of ML-based solutions.	Assure standard conformity and IPR protection of ML-based physical layer solutions.

7 Optical Networks

Editor: Raul Muñoz

7.1 Introduction

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, residential/business fixed access, and mobile fronthaul/backhaul networks, 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.

From ground-breaking innovations such as new types of optical fibres, new erbium-doped amplifiers for WDM systems, and digital coherent optics for 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.

7.2 Vision

Traffic continues to grow at above 20% rates each year. The transmission of this increased capacity is a well-known challenge with a significant amount of research addressing it. However, all this traffic will have to be switched and routed at some point in the network, and probably at multiple points between customers and data centres, including at future edge-compute locations. The de facto approach to handle this is electronic switching, routing and memory, within IP routers, ethernet switches etc. These devices will continue to consume increasing amounts of energy, occupy larger space and cost more. At the same time, the actual flows entering these switches are also increasing to multi-Gb/s rates: rates that would be more efficiently handled optically, if only there were suitable optical switching alternatives.

Optical communications and networking technologies are essential to provide high-speed, cost-effective, energy-efficient, secure, and reliable connectivity services for 6G, spanning from the fixed access to the transport network, as well as for inter and intra data center communication, as shown in Fig.7-1. Open and disaggregated packet and optical technologies will be further developed to provide a converged packet-optical network with a more granular and large-scale management of flows with dedicated and deterministic QoS in support of B5G/6G mobile networks, IoT/V2X, free-space optics, non-terrestrial networks, and fixed networks (enterprise, residential). The need for extending the cloud towards the network edge (e.g. Street cabinet, Cellsite, RSU) will require the deployment of edge computing integrated with the packet optical networks, providing a wide ecosystem where packet, optical, edge computing and cloud converge.

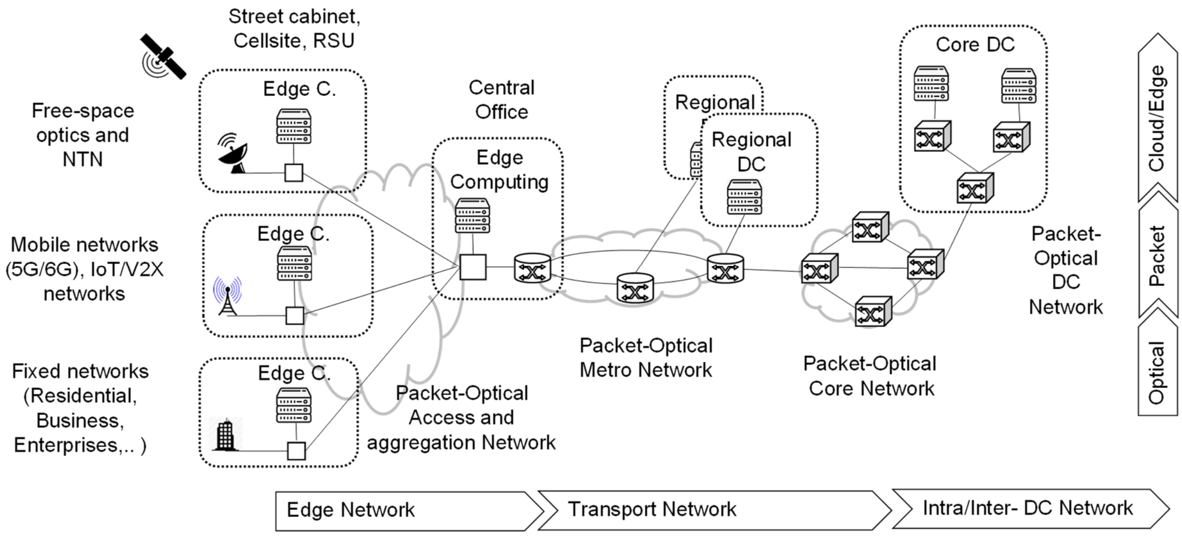


Fig.7-1 End-to-end optical network scenario

The following high-level requirements are identified for the target vision:

- *High-capacity scaling and reliable connectivity* through the adoption of spectrally and spatially multiplexed systems with suitable photonic technologies and devices to support network reconfigurability and dynamicity.
- *Cost-effective and energy-efficient systems* integrating suitable photonic technologies and devices that meet mixed requirements in terms of cost, reach, throughput, power consumption and footprint.
- *Advanced electro-photonic integration* enabling a new generation of optical networking and IT equipment, overcoming the challenges in scaling electronic interconnect speeds. Combining the advantages of optics and electronics is the way forward to deliver unprecedented functionality, compactness and cost- and energy-effectiveness.
- *Access to everything.* Current optical access network solutions will evolve further to also fulfil requirements of future applications demanding ultra-high speed and low latency. New architectures, derived from low-cost Fibre-to-the-Home solutions need to be scalable to support dense 6G deployments at a low-cost point.

- *Deterministic networking in optical/packet networks* to encompass multiple switching capabilities, being able to accommodate traffic flows with a more granular and large-scale management with deterministic QoS.
- *Edge-cloud continuum to* provide a converged packet, optical, edge computing and cloud ecosystem, requiring the deployment of edge computing integrated with the packet optical network based on the need for extending the cloud towards the network edge.
- *Full network programmability*, considering new deployment models for open and disaggregated optical networks, in which open devices and sub-systems result in a programmable optical transport network with open and standard interfaces.
- *Network and service automation* by deploying closed-loop control and advanced telemetry enabling AI mechanisms, using telemetry data for network continuous optimization in a proactive way with machine learning-assisted analytics, that may anticipate to the problems and events, and propose corrective actions.
- *Network multi-tenancy* with intent-based policies and software defined security to deploy smart, secured, and trustworthy end-to-end network and transport slices in open and disaggregated networks.
- *Efficient Integration of optical technologies for radio access networks* to provide optical connectivity to each radio antenna and deliver ultra-high speed and low latency at a cost that is compatible with the revenue generated by smaller and smaller cells.
- *Integration of free space optical technologies* to complement with next-generation technologies, such as 5G and 6G wireless networks to be widely deployed in various indoor (e.g., data centers), terrestrial (e.g., mobile networks), space (e.g., inter-satellite, ground-to-satellite and deep space communication), and underwater systems (e.g., underwater sensing).
- *Secure communications* deploying optical quantum communications and related technologies (such as the adoption of QKD and quantum cryptography in optical networks) in coexistence with the actual deployed network infrastructure.

7.3 Sustainable capacity scaling

Global data traffic in optical networks has been growing a high and steady pace of x2 every 2-3 years over the past 15 years and there is no sign that this pace will be slowing down significantly in the upcoming decade. Hence fiber networks need to urgently adapt. Not all segments will be equally hit. For the sake of efficiency and latency, data will be stored closer to the users of these data, hence metropolitan and edge optical networks will grow considerably faster than long-haul fiber networks. At the same time, cloud providers will continue to massively offload the public internet into their private intranets. Projections of future traffic predict required data rates of 10 Tb/s line interfaces and over 1 Pb/s for optical fibre systems by 2025 [C7-02], while optical interconnect capacity are expected to be aligned with the Ethernet roadmap of line interface speeds (~6.4 Tbit/s in 2030). Networks also need to provide headroom for unexpected traffic increases, as observed in several EU member states during the health crisis of 2020-2022.

7.3.1 Scaling to Petabit/s capacities in core and metro networks

This evolution stumbles upon the most fundamental limits of physics that are: Moore's law on Silicon integration and Shannon's limit on optical fibre capacity, both of which are considerable barriers to growth with their specific challenges. the rate of router interfaces has been surpassing the rate of opto-electronics transceivers per wavelength since the years 2010 and the gap is widening [C7-02]. Since the former generally feed the later, new research efforts are required to radically improve the dense integration of high-speed electronics and optics (separately, or together). However, there is a clear danger that very soon, a two-fold increase in the requested capacity could require doubling the amount of optical/electronic hardware. This would increase costs in a linear fashion, in contrast to revenues from users which have been nearly unchanged over the past decade and depart from future capacity requirement of EU citizens. Obviously, disruptive approaches are now needed.

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, and wavelength selective switches. Advances in fiber technology will facilitate this evolution.In particular, the hollow-core fibre promises a larger bandwidth (up to several hundreds of nm), possibly tuneable during manufacturing to explore wavelength ranges unused today, and lower latency (~30% lower) compared with standard fibres. It is yet unclear when this fibre can be mass-produced but remarkably, its attenuation in laboratories has now gone well below the lowest attenuation with standard fibre. Advantageously, it also comes with a strong reduction of the most detrimental propagation effects (e.g. nonlinear effects), while new operational issues need to be addressed. Moving to this new fibre and to new bands will generate heavy disruption in system and network design and call for significant research efforts to organize the transition.

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 and new optical amplifiers are needed, along with new fibre types.

Finally, 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 never hitting the guaranteed limit of resilience, which has been rendered possible by the availability of a wealth of monitoring data in new optical networks, as discussed in section 7.7.

7.3.2 Next generation terabit/s transceivers

Recent successful innovations will be exploited far beyond the 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 offer enhanced broadband access, to cope with the growth of inter data center communications, to make edge cloud a reality, to allow a new breed of intra-data center networks and support laser non-terrestrial communications. Coherent is the most promising technology to bridge the gap which is caused by the Shannon limit, leveraging "shaped" modulation formats, flexible rates, higher than 200Gbaud symbol rates.

Photonic integration and co-packaged optics will be important for efficient scaling, as full-band WDM comb transceivers could become a prerequisite to support massive space division multiplexing. They will allow for the integration of multiple optical devices into a single platform and the closer integration of the optical devices with their corresponding electronic circuits. Overall, a change of scale in component count per square millimetre will be required to less that 1 pJ/bit/s in the medium-term future.

7.3.3 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Sustainable capacity scaling**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
1 Petabit/s over transcontinental distances Overcome next major milestone of system capacity	Long-term (finished in 7y+)	*Provide mix of technologies to meet the goal, and design rules, including compensation of impairments. *Achieve 2Tbit/s per lambda in 5y	High-capacity scaling and reliable connectivity
Beyond Shannon and Moore in metro Enable massive parallelism in space and wavelength domains	Mid-term (finished in 5y)	*Multi-band WDM amplification (20THz per fiber in 3y, 200THz line amplifier node in 5y) *subsystems for space-division mux (>10 modes or cores or fibers per device)	High-capacity scaling and reliable connectivity Advanced electro-photonic integration
Petabit/s energy-efficient interconnects Cost per bit and power per bit reduction	Mid-term (finished in 5y)	*Make leap in optical integration and co-packaged optics for lower consumption interconnects (<1pJ/bit/s)	High-capacity scaling and reliable connectivity Cost-effective and energy-efficient systems Advanced electro-photonic integration

7.3.4 Recommendations for Actions

Research Theme	**Sustainable capacity scaling**	
Action	**scaling to Petabit/s capacities**	**Next generation terabit/s transceivers**
International Calls	**X**	**X**
International Research	**X** To leverage industry and academic efforts vertically for critical mass.	**X** To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

7.4 New switching paradigms

Technologies such as autonomous driving, augmented/virtual reality, and augmented workspace, are about to become reality in a not so far future. As a consequence, the number of communicating entities, mainly machines, will increase dramatically. This will impose significant challenges for the network. New network architectures with edge clouds close to the end user and centralized clouds with flexible function split are required. The network needs to be automatically managed across a variety of different wireless and wireline transmission and access technologies. In order to enable this flexibility, new switching paradigms are needed to connect real-time programmable optical devices in distributed architectures. Although traditionally, optical switching has been challenging from a material perspective, and optical buffers / memories even more so, the prospects are changing here. Photonic integration with different materials, such as PPLN or others, are beginning to showcase the benefits of introducing optics into the switched part of the overall network. Optical switching within IP router fabrics is one possibility that might not disrupt the network too much as a whole, but entirely new optically-based network architectures should be considered, where the promise of dramatically reduced power consumption is too compelling to ignore. Although optical packet switching, as a concept, has been widely known for 20 years, it is only now that we have both the drivers, in terms of power, space and cost, and the technologies, in terms of new materials and photonic integration, to make substantial progress.

The steep learning curve in photonics integration will also allow optical flow switching approaches, which were previously considered too costly and/or complex. This can pave the way to a new generation of switches with optimized mix between optical and electronic processing functions. They should be operating over multiple wavelength bands and spatial dimensions and have a smart network fabric relying on AI-assisted software programmability and slicing, addressing multiple protocol layers and network domains. This applies also to intra-DC applications, where new switching concepts mixing optical and electronic switching technologies could lead to higher performance and lower power consumption. In addition to ultra-fast switching speeds, the capability of switching on different levels of granularity and a high overall switching throughput, future switching architectures need to take into account the energy-efficiency of the switches itself, but also that of the network they are supporting. Another topic of interest would be disaggregated switching platforms to replace purpose-built solutions.

7.4.1 Ultra-fast Multi-granular Switching Nodes

Flexgrid technology on the optical layer, the utilization of new wavelength bands (beyond C+L+S band) and the advent of multi-core/multi-mode transmission will require new multi-granular switching node architectures. This will allow an even more flexible network slicing in the wavelength as well as in the spatial domain. An operation over multiple wavelengths, wavelength bands and spatial dimensions requires new switch and transponder architectures that 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 (< 1 ms) on the optical layer and taking into account concerns such as amplifier power transients need to be developed.

7.4.2 Switching Architectures guided by Energy-Efficiency

Future networks will face the need to reduce their energy consumption. Therefore, switching architectures need to take energy-efficiency into account at a very early stage and on all network levels. On the hardware level, switching architectures with an intelligent mix between optical and electrical switching functions, will be required. In addition to a power reduction in the switching components itself. On the control & management layer, the switching needs to be optimized to perform switching functions in the domain, electrical or optical, with the lowest power consumption. In that respect the switching operations could benefit from the larger degree of freedom in multigranular switching architectures (wavelength, waveband and space).

7.4.3 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**New switching paradigms**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Multi-granular Switch allowing to switch between space, wavebands and wavelengths	Mid-term (finished in 5y)	Operation in S+C+L-band (1460 – 1625 nm) Operation with multicore fiber Switching granularity from 50 GHz (single wavelength), over 5-8 THz (waveband switching) to 21 THz (complete fiber/core switching)	Deterministic networking High-capacity scaling and reliable connectivity
	Long-term (finished in 7y+)	Operation beyond S+C+L-band	
Switches with fast reconfiguration times	Mid-term (finished in 5y)	< 1ms	Full network programmability
Switching architectures with optimized mix between optics	Long-term (finished in 7y+)		Cost-effective and energy-efficient systems

and electronics for energy-efficient networks			

7.4.4 Recommendations for Actions

Research Theme	**New switching paradigms**		
Action	**Ultra-fast Multi-granular Switching Nodes**	**Switching Architectures guided by Energy-Efficiency**	**Research Aspect N**
International Calls	**X** To leverage industry and academic efforts vertically for critical mass	**X** To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems	
International Research	**X** To leverage industry and academic efforts vertically for critical mass	**X** To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems	

7.5 Deterministic networking

While today's Internet is built on a best-effort traffic paradigm, an increasing number of applications require reliable end-to-end transmission with guaranteed throughput and bounded latency. Examples range from RAN transport over vehicular/robotic/ industrial control to augmented/virtual/extended reality. Stringent network requirements also result from accurate positioning, navigation and timing (PNT) services. Future human-centric use cases such as the "Internet of senses" and holographic communications are expected to increase the demand for deterministic network behaviour even more. Application requirements are diverse and therefore a flexible quality-of-service (QoS) framework is necessary: A control application for instance only consumes low bandwidth but needs high reliability and relies on definite time window for packets arrival. In turn, an extended reality applications may be able to tolerate more timing variations, yet high bandwidth is mandatory for good user experience.

While mechanisms exist to control throughput, latency, jitter and packet loss in packet-optical networks, they often provide statistical QoS only and do not guarantee a deterministic network behaviour when constant latency is more critical than low latency. Available timing signals rarely offer the necessary accuracy and/or reliability to allow precise time synchronization for mission critical applications. End-to-end services are often delivered over heterogenous network infrastructure in which different QoS and traffic-engineering mechanisms are employed and need to interwork which each other. Fundamental architectural questions such as where to control traffic, which data plane support is necessary, and how to facilitate end-to-end service assurance need to be answered in light of a diverse set of application requirements. Novel solutions are needed which trade-off performance improvements against scalability limitations and implementation complexity. Activities should leverage technologies from standards bodies such as ITU-T (e.g. OTN, PON), IEEE (e.g. TSN, 1588, EPON), and IETF (e.g. DetNet, L4S), OIF (e.g. FlexE), address deficiencies and develop novel solutions as extensions.

7.5.1 Resilient solutions for high-precision, network-assisted timing distribution

Precise timing information is not only necessary to operate communication networks, it is also an enabler for the digital transformation of critical infrastructures. Real-time control, positioning and navigation but also accurate event recording and threat mitigation rely on the availability of precise time information. High-precision time distribution becomes an additional service delivered by a new generation of smart networks.

Cost-effective and scalable time distribution solutions are required which can operate over a heterogeneous network infrastructure and are robust against failures and attacks. For high reliability, a resilient timing network combining information from multiple reference sources and offering sufficiently long local hold-over capabilities is necessary. Hardware assistance is required for high-time resolution and low timing error. Pluggable or embedded time synchronisation modules can provide precise timing capabilities to network or user equipment not possessing such capabilities by default. A control, telemetry and analytics framework is required to deliver timing services, assess their quality and take corrective actions where needed.

7.5.2 Reliable data & control plane solutions for deterministic network services

The performance of packet-optical networks is crucially dependent on packet processing and traffic management functions such as shaping and queuing. Deterministic services need to be given preferential treatment without burdening the network with overly complex processing for lower priority services. Flexible data plane and control plane solutions are required which can cope with changing traffic patterns and a variable traffic mix. Architectural trade-offs are needed to avoid network inefficiencies on one hand and insufficient performance on the other hand. Mechanisms to apply back-pressure and to communicate between network and client equipment can help to avoid a QoS deterioration by mitigating network overload conditions.

Some of the most challenging requirements are driven by mobile fronthaul applications and comprise <100µs latency[28], <8ns relative timing error, and several tens of Gb/s throughput. Data plane optimizations are necessary to fulfil the demanding latency and timing requirements, sometimes across multiple networks segments using different transport and switching technologies. Applications in which packets have to arrive in a certain time window need further research, especially if such services have to be delivered over large networks or a heterogenous infrastructure which is only partly timing-aware. If deterministic network services serve mission critical applications, measures need to be taken to protect these services against outages, rogue device behaviour, and attacks by malicious actors.

7.5.3 Tools for service assurance in deterministic networks

The performance of deterministic network services can strongly depend on network size, number of nodes, network traffic as well as on the characteristics of the used network equipment and applied control strategies. Planning tools are required to estimate the attainable service performance and determine an optimized network configuration. Methods of network calculus, heuristics and information about the internal structure of network equipment can be used for this step. Information from network planning but also network telemetry and equipment status (e.g. queue fill levels) can then be leveraged to get a current view of the network and service quality, forecast future evolution, and make adjustments where necessary. Network analytics and machine learning can help to accurately predict the network behaviour. A digital network twin approach may be used to test changes before deploying them in an operational environment.

---

[28] including fibre transmission which adds 5µs/km

7.5.4 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Deterministic networking**		
Research Challenges	**Timeline**	**Key outcomes**	**Contributions/Value**
Resilient network-assisted time distribution	Short-term (finished in 3y)	Accurate time-of-day delivery to network endpoints and user devices Resilience against GPS/GNSS jamming and other attacks, indoor performance w/o GPS/GNSS reception KPI values are application specific	Deterministic networking Secure communicatons
Reliable data & control plane solutions for deterministic network services	Mid-term (finished in 5y)	Reliable transport of deterministic traffic alongside lower priority traffic Guaranteed throughput, bounded latency, high availability KPI values are application specific	Deterministic networking Network multi-tenancy High-capacity scaling and reliable connectivity
Tools for service assurance in deterministic networks	Mid-term (finished in 5y)	Performance prediction of deterministic network services in complex networks, assessment of network optimizations on digital twin Forecast accuracy, network utilization, predicted versus measures service quality	Network and service automation

7.5.5 Recommendations for Actions

Research Theme	**Deterministic networking**		
Action	Resilient solutions for high-precision, network-assisted timing distribution	Reliable data & control plane solutions for deterministic network services	Tools for service assurance in deterministic networks
International Calls		X EU-Japan collaboration	
International Research	X EU collaboration, worldwide standards	X EU collaboration, worldwide standards	X EU collaboration, worldwide standards
Open Data			X Network statistics
Large Trials		X Scalability tests in relevant environment(s)	X Scalability tests in relevant environment(s)
Cross-domain research	X Collaboration with vertical industries (manufacturing, multimedia, …)	X Collaboration with vertical industries (manufacturing, multimedia, …)	X Collaboration with vertical industries (manufacturing, multimedia, …)

7.6 Optical technologies for radio networks and systems

The expected tenfold increase of traffic growth and the tight latency constraints dictated by new 6G services will require a substantial evolution not only of the RAN but also of the architecture and the technology of the underlying mobile transport network. Optics is an enabler for 6G not only as regards the new mobile transport network, but novel optical Interconnect technologies will play a key role in future advanced antenna systems, impacting their architectures. Finally, new advances in photonic integration open new opportunities to apply suitable combinations of optical, radiofrequency, and digital electronics to radio systems. These three application macro-areas lead to a wide set of new challenges for optical technologies in radio access, as illustrated in the next sections.

7.6.1 Optical technologies for radio access networks

Optical technologies are well known for their high bandwidth, transparency to the format of the carried signal and low latency and play a role as important as packet networking in this evolution of the mobile transport networks. However, they need to evolve in parallel to it to achieve a further level of energy efficiency, miniaturization, cost effectiveness and fast reconfigurability. Among optical technologies, wavelength division multiplexing (WDM) is the most promising one as regards supported bandwidth, distance (a prerequisite for centralization) and compatibility with several network topologies (ring, mesh, point-to-multipoint) but it is also the one where the technology evolution needs to be more radical. The current cost figures of essential WDM components (high speed transceivers, wavelength switches, etc.) is at least one order of magnitude higher than the ideal cost figure (<0.1 $/Gbit/s) for an access network. Moreover, in this network segment, the link attenuation is significant due to the presence of optical add-drop nodes and passive splitters. The use of optical amplifiers in outdoor equipment, like a radio unit placed at the top of an antenna pole is problematic for various reasons: additional size and weight (which may be subject to regulatory restrictions or significantly impact on cost), additional power consumption (and related issues in dissipating the generated heat by passive cooling), and impact on equipment reliability (which is critical, considering the operational costs of a radio unit replacement). All this leads to find new technologies to reduce optical amplifiers footprint and power consumption and increasing their lifetime, especially in hard environmental conditions such as operation at high temperatures (which may be higher than 100°C in some conditions). Technologies options and their TRL will be discussed later in this chapter. Alternative actions can be taken at an architectural level, for example using remotely pumped or Raman amplifiers where the most temperature-sensitive components (i.e., the pump lasers) are placed in a protected environment. Even in this case, however, cost implications and compatibility with RAN installation and maintenance practices would need to be investigated. Finally, since fibre's chromatic dispersion becomes very significant when moving at high bit rates, even over short distance, DWDM system in the O band could be developed to boost the system capacity using cost effective intensity-modulated direct-detection optical transceivers. However, this leads to further research challenges related to techniques to mitigate Four-Wave Mixing (FWM) effects as well as new technologies for power and space efficient optical amplification in O-band, both by using doping materials different form Erbium, like Praseodymium or Bismuth, or finding new techniques for linearizing the operation of Semiconductor Optical Amplifiers (SOA) in DWDM systems.

This picture calls for the following research and innovation challenges:

1) The development of high speed coherent optical interfaces (from 100G over 400G to 1.6T) at a much lower cost and power consumption than today. This could exploit new AI-based techniques for joint optimization of signal coding, modulation and pulse shaping at the transmitter and signal equalization at the receiver. Moving into the optical domain functions today performed by the DSP (e.g., digital-

to-analog conversion and polarization recovery) may also be explored, leveraging on recent photonic integration advances.

2) The development of DWDM systems in O-band to increase the system capacity keeping cost and power consumption comparable to those of current short-reach optical interconnection systems. This involves both the development of small size, power efficient optical amplifiers in O-band and the use of new signal coding and modulation techniques that, while making the signal robust with respect to FWM effects, do not require complex digital signa processing.

3) Enabling longer distance and higher link attenuations in a CRAN by means of small footprint optical amplifiers in a pluggable format (which start to be commercialized today) or as a function provided in silicon photonics chips by means of integrated SOAs or doped silicon waveguide, whose TRL is less mature. The capability to work in a harsh environment will be an important enabler for making both the technologies usable in radio systems. Thus, radically reducing cost and getting low power, industrial temperature, zero-touch operation, and high loss budget for moderate capacities are the key features to develop. Dense hardware boards, like the ones used in massive MIMO antenna units operating at high radio frequencies, may require the operation at even higher temperature, close or above 100°C. Remote optical amplification may become appealing in these situations since it places the laser far from system's hotspots.

4) But high capacity and distance are not all: the 6G network will be a dynamic network, ideally capable to provision every type of service on demand and in real time. So, a third challenge is ensuring dynamicity to the underlying transport network to avoid cost and waste of energy for bandwidth overprovisioning. Dynamic bandwidth allocation is something packet networks are very good at doing but energy consumption and latency may unacceptably grow in presence of congested traffic. Optics may be used for traffic offload, but it is basically static. Reconfigurable optical components exist (tuneable lasers, tuneable filters, wavelength selective switches, optical switches) but, despite many research efforts never achieved sufficient cost effectiveness or TRL for mobile transport applications. Innovation Actions to accelerate the time to market of fast configurable (nanoseconds) optical subsystem could help in this sense. This should go in parallel with the research on new system architectures to understand how optical and packet switches can interwork to keep low and deterministic latency, compatibly with the carried service, while preserving the bandwidth advantages of statistical multiplexing.

7.6.2 High speed optical interconnects in radio systems

Traditional fronthaul links, based on digitized samples of the signal on air, were abandoned in 5G in favour of more bandwidth-efficient packet-based solutions, where digital processing functions previously performed at the baseband unit (BBU) were moved back to the antenna. This came at the price of a higher interconnection speed between the digital integrated circuits (ICs) and the radio frequency ICs (RFICs) in the radio unit (RU): in extreme scenarios, the total throughput can be hundreds of Terabit/s. In principle, the same considerations hold in 6G, which is expected to lead to a further increase of capacity. However, recent energy efficient implementations of both optical transceivers, based on co-packaged optics, and digital-to-analog and analog-to-digital converters (DAC/ADC) may revive the transmission of digitized sample (also known as digital radio-over-fiber, RoF). A third possibility is performing at the RU only analog radio frequency functions (an option known as analog RoF or simply RoF). This is a well-known solution but 6G will require a more demanding performance, e.g., in terms of dynamic range, and much lower cost, size and energy consumption compared to current commercial solutions, which are conceived for niche markets and not for mobile systems. What of the three options to use will depend not only on technological considerations but also on their architectural implications, which are significant since al involve the way the radio protocol stack is split.

Regardless of the solution, there will be common constraints:

- Energy efficiency of the order of 1 pJ/bit. Power dissipation is critical in radio systems, mostly relying on passive cooling.
- Low and deterministic latency (a few nanoseconds), including FEC and bits-to-symbols mapping.
- Low Bit Error Rate, <10−15.
- Wide temperature Range. Since an antenna operating outdoor is not actively cooled, the internal operating temperature can be > 100 ∘C. This challenge can be either taken at a technological level, e.g., using lasers or other light sources that can work at high temperature, such as quantum dot lasers or wide arrays of LEDs in massively parallel data transmission, or at n architectural level, studying the impact on external laser sources on system reliability and installation and maintenance practices.

And common technologies:

- Co-packaged optical (CPO) transceivers, i.e., optical transceivers mounted on the same substrate of the IC they are connected to either provide high bandwidth density (hundreds Gbps/mm2) optical interconnects, in digital fronthaul solutions, or highly linear analog transmission, in RoF systems. CPO transceivers are already in production, but not for along transmission. And even for digital transmission, they would need a more energy efficient non-retimed electrical interface, that would affect the way the connected ASICs and their equalizers inside are designed.
- Monolithically integrated optical transceivers, where optical front-end and electronics are integrated on the same chip.
- Passive optical routing solutions (e.g., silicon-on-glass optical interposer or optical Printed Circuit Boards, PCB) to mitigate the high loss, power consumption, signal degradation and electromagnetic interference issues electrical interconnects suffer from at high bit rates. Optical PCBs are the most elegant and compact solution but still have unsolved issues in terms of integration with the electrical layer of the PCB, especially if reflow soldering compatibility is required, and alignment of the connected optical components.

7.6.3 Optically enabled radio functions

Transport networks and high-speed interconnects are two sweet spots for optical technologies but new advances in photonic integration open new opportunities to apply suitable combinations of optical, radiofrequency, and digital electronics to radio systems.

The use of photonic technologies to generate and process radio signals (also known as microwave photonics) is being investigated in the research community for a long time. However, the practical use of microwave photonics technologies was limited so far to niche applications (e.g., military), mainly due to cost. Where the trade-off is between electrical and optical technologies, as regards cost, performance, and energy efficiency, is still a question mark, especially considering the two technologies are investigated by separate research communities. Moreover, most of existing research works often focus on a single aspect (e.g., lowering the phase noise), neglecting the implications of the proposed technique on the whole system. This leads to a need (and an opportunity) to fill these gaps with research and innovation actions to cover the integration of optical and wireless technologies in a single radio system. Possible topics regard the use of photonic components and subsystems (oscillators, mixers, antenna elements) to help wireless systems to scale in bandwidth (e.g., sub-THz) and performance (e.g., low phase noise) in an energy efficient way.

A first related research challenge is the low phase noise (PN) generation of frequency references for clock and radio frequency. Electronic generations schemes are not accurate enough as the frequency rises up to the THz range but also the traditional optical schemes based on the heterodyning of two independent lasers in a photodiode suffer from the frequency and linewidth instability of the laser, making them unsuitable for radio applications. New solutions will require phase-locking of the beating sources, as in Mode-Locked Lasers (MLL). However, state-of-the-art MLL are not compatible with radio applications due to the high cost, power consumption, and large footprint so that developing new integrated photonic components will be necessary. Photonic integrated optoelectronic oscillator is an alternative to MLLs to generate stable low PN radio carriers. They rely on modulating the light's intensity from a laser and feeding back the detected RF modulation to the modulator's input port. A third approach modulates a Continuous Wave (CW) laser with a low frequency carrier signal through a non-linear optical modulator, so generating multiple frequencies of the carriers. The main challenge with this method is achieving sufficient conversion efficiency. Independent of the frequency generation method, the possibility of selecting different laser modes shall be provided for the flexible generation of RF carriers with tunable frequency.

A second research challenge is increasing the density of antenna elements per surface area, as required by massive MIMO systems at very high radio frequencies. This requires integrated photonic solutions for the distribution of optical signals inside the antenna system, as optical waveguides embedded in the Printed Circuit Board (PCB) and glass interposers. With optical waveguides in the PCB, the signal is distributed by means an optical layer below the antenna elements and then routed vertically to PD near the element. A glass interposer could be instead placed on top of the RFICs, so that the optical signal is routed down vertically to the r PD. These solutions his requires the co-packaging of RFIC, TIA, PD and optical waveguides, which is a formidable technology challenge since these components are based on very different technologies.

A third radio-related challenge where optical technologies already demonstrated their potential is beamforming. Current solutions enable squint-free, high pointing accuracy beamforming but suffers from high insertion loss, proportional to the number of AEs and photonic phase shifters or true time delay elements. Integrated optical amplification, either based on III-V integration or rare earths doping of silicon waveguides could mitigate the issue but should not negatively affect the phase noise of the system.

7.6.4 Research Challenges

Research Theme	Optical technologies for radio networks and systems		
Research Challenges	Timeline	Key outcomes	Contributions/Value
High speed WDM optical transmission in RAN (from 100G over 400G to 1.6T) 10 times more cost effective than today. Both coherent and direct detection are in scope. Optical processing and ML-based DSP are among the enabling technologies.	Mid-Term	Energy and cost efficient digital and optical processing schemes for coherent optical interfaces	High-capacity scaling and reliable connectivity. Main application is the 6G mobile transport network, to avoid capacity bottlenecks
Cost-effective optically amplified networks having low power consumption, Industrial temperature operation capability and zero-touch operation as the key features.	Short-Term Mid-Term	Cost efficient, small form factor and plug & play optical amplifiers	Access to everything Enabler for Cloud Ran deployments over high distances, where opex savings and high coordination is ensured by deeply centralized processing functions.
Fast reconfigurable optical RAN based on tuneable lasers, tuneable filters, wavelength selective switches, optical switches) for RAN.	Short-Term	High-TRL fast-reconfigurable integrated optical switches and tuneable lasers	Full network programmability Enabler for optimizing the bandwidth resources in packet RANs where with highly variable traffic load
High bandwidth density (hundreds Gbps/mm2) optical interconnection systems with ~1 pJ/bit energy	Mid-Term	Highly energy efficient, high-performance solutions for co-packaged optical transceivers,	Advanced electro-photonic integration.

efficiency, ~ns latency, <10-15 BER, high temperature operation (> 100 ◦C.).		monolithically integrated optical transceivers and passive optical routing solutions.	High-capacity scaling and reliable connectivity and Cost-effective and energy-efficient systems Application area is 6G RAN equipment (remote units, digital units, x-haul switches) where high capacity must be provided in small space and power dissipation is crucial.
Tuneable **<u>low phase noise (PN) generation of frequency references</u>** for clock and radio frequency. up to the THz range based on integrated photonics schemes.	Mid-Term Long-Term	Integrated photonic based solution for generation of high radio frequency with extremely high accuracy	Efficient Integration of optical technologies for radio access network Enabler for 6G when moving to sub-THz frequencies
High-bandwidth density solutions for the distribution of optical signals inside an antenna system,	Long Term	Solutions for the distribution of optical signals in an antenna system based on the co-packaging of heterogeneous technologies (RFIC, TIA, PD, optical waveguides, etc.) in the same device	Advanced electro-photonic integration Efficient Integration of optical technologies for radio access network Enabler for 6G when moving to sub-THz frequencies, to guarantee high performance (low noise, high output power, high EMF immunity)
MIMO systems based on low noise optical amplification	Long Term	Integrated optical amplification to provide high Tx or Rx in antenna systems based on optical generation and distribution of radio signals	Efficient Integration of optical technologies for radio access network Enabler for 6G when moving to sub-THz frequencies, to guarantee high performance (low noise, high output power,)

7.6.5 Recommendations for Actions

Research Theme	**Optical technologies for radio networks and systems**						
Action	High speed WDM optical transmission in RAN	Cost-effective optically amplified networks	Fast reconfigurable optical RAN	High bandwidth density (hundreds Gbps/mm2) optical interconnection systems	Low phase noise generation of frequency references	Distribution of optical signals inside an antenna system	MIMO systems based on low noise optical amplification
International Calls	X Having a shared program with non-EU based	X Having a shared program with non-EU based		X Involving US market leaders in co-packaged optics		X Need to involve foundries that can integrate	

	research and market leaders in coherent optics and ICs	research and market leaders in optical amplification				PICs and electronic ICs with a mature and reliable process	
International Research	X Having shared program with non-EU based research and market leaders in coherent optics and ICs	X Having a shared program with non-EU based research and market leaders in optical amplification		X Involving US market leaders in co-packaged optics		X Need to involve foundries that can integrate PICs and electronic ICs with a mature and reliable process	
Open Data							
Large Trials	X Trials with leading EU mobile operators in their RAN	X Trials with leading EU mobile operators in their CRAN	X Need to demonstrate production in large scale and high reliability				
Cross-domain research					X It requires skills in both optics and radio systems, a combination tat seldom engineers have	X It requires to involve different expertise: photonic systems, experts, radio designers, PIC and IC technology experts	

7.7 Optical network automation

Optical network automation is key to achieve operators' business goals and in supporting new complex services. Aspects related to automation must be developed in the areas of service deployment, network planning and overall network operation and maintenance. Initial research should be focused in automating repetitive, error-prone tasks or tasks with very well-established workflows, with applications in single domain scenarios, such as service activation (aiming at OpEx reductions due to more efficient workflows and considerably reduced execution times), increased flexibility in offering services, better service level agreement performance, and faster issue resolution. Automation is critical in optical networks supporting increasing data rates given, for example, the complexity of modelling of physical impairments, or the large number of parameters and their interdependencies. Outcomes related to automation in single domains shall form the basis for more ambitious cross-domain automation (across technology layers or network segments). AI/ML solutions in support of network operations should be further developed beyond policy- / expert- /rule- based systems, and control and orchestration architectures should become increasingly modular, leveraging the flexibility of deployment in hybrid clouds while relying on open and standard data models, protocols, interfaces, and frameworks (including, for example, proven and mature open-source projects and initiatives).

7.7.1 Network Telemetry and Optical Network Sensing

This aspect should address activities related to optical monitoring, network streaming telemetry and overall secure and efficient data collection, storage, and subsequent use, with applicability and focus on *large-scale scenarios.* Telemetry systems should allow maximum flexibility, including *at-origin* or *intermediate* filtering and aggregation of data. Research activities range from the definition and subsequent standardization of data models for the telemetry data; the definition of efficient and secure protocols, in terms of latency and encoding to the definition of architectures in support of overall infrastructure monitoring and telemetry. Telemetry should be enabled at a device or system level as well as at the domain or network level. Telemetry systems should unify aspects related to network state synchronization, alarm reporting (incl. threshold crossing alerts), performance monitoring and fault management at metrology grade with well-assessed intervals of confidence. This aspect also encompasses research on *Network Sensing* (the use of network and computing/storage infrastructure to design solutions enabling the detection of events of interest and related applications). This includes, in particular, *optical fibre sensing*, for applications like intrusion detection and prevention, repairing network outages towards self-healing networks, and the systematic use of such sensing techniques – coexisting with actual user traffic -- at scale to predict and pinpoint physical layer issues along with software automation, design, and operational tools to mitigate those issues. The aspect should address applications, services, technologies, and challenges/benefits of network sensing.

- Joint communications and sensing
- Inband sensing and data extraction

7.7.2 Control and Orchestration architectures, protocols and methods for Network Automation

This aspect covers research activities related to the definition of control and orchestration architectures for heterogeneous multi-layer, multi-domain, or multi-technology scenarios, addressing the shortcomings of current (e.g., SDN-based) approaches such as scalability, reliability, or deployment agility, as well as the improved support of emerging cases such as infrastructure sharing. The architectures should exploit and enable multi-tenancy in a more cloudified environment, while empowering users with the capability to manage their own services, while full leveraging network slicing. In a short-term, control and orchestration systems should rely on open data models and interfaces, suitable extended for recent development such as multi-band networking, space division multiplexing (SDM) or improved support for physical impairment modelling in view of beyond 100G systems. Novel or refined architectures should be investigated, including full support of service lifecycle management, integration, and migration of current operators OSS/BSS systems, encompass telemetry and network sensing and apply to heterogeneous environments, such as integrating wired/wireless access networks (i.e., PON/RAN) and transport segments in over-arching control systems. Enable data export and the use of 1st party and 3rd party systems (for resource allocation, path computation, AI/ML systems), including dynamic algorithms and heuristics enabling almost real-time end-to-end quality of experience with solutions to minimize end to end delay/latency and jitter and performing coordinated resource allocation. Further research is needed in the integration of packet and optical networks, ensuring efficient and low-latency service provisioning. In a medium-term, research should address how to enable or adopt the application of IT and DevOps principles for network control and management, further leveraging tools for process automation, data visualization. Support higher abstractions in terms of service definitions, further exploiting and refining intent-based approaches. In the long-term, support truly autonomous networks by integrating monitoring, telemetry and 1st party/ 3rd party AI/ML assisted network operation closed-loops, with dynamic instantiation of customized control and management systems in hybrid clouds in support or consolidated business models of infrastructure sharing and multi-tenancy. Research on the applicability and role of analog computing for network operation with extremely reduced power consumption include:

- Data Lake concept
- RAN / Metro / Core overarching control
- Better integration of packet/optical systems (IPoDWDM)
- Transparent interconnection of domains
- RAN and Optical Transport interworking.
- Joint Resource Allocation (e.g. PON DBA and Metro, RAN)
- Point to Multipoint optical Networks and Optical Multicast
- Energy efficiency at the packet and optical level

7.7.3 AI/ML in support of Network Operation

This aspect encompasses the use of AI/ML-mechanisms in support of network operation, e.g., service and infrastructure management, both in single domains and cooperatively across different domains. This includes: i) the definition of ML models, the use of training data sets; applicability and reusability of existing/previously used models; ii) development of use cases and scenarios involving general resource allocation, function placement (for example, selection of functional splits based on multi-objective problem formulation and dynamic traffic patterns) and iii) Research on distributed self-management control infrastructures based on multi-agent systems able to autonomously coordinate resources near real-time for end-to-end service assurance

- Reuse and Integration of AI/ML platforms
- Applications of Generative AI for network management

7.7.4 Reliability and Security of Control, Orchestration and Management

This research aspect covers activities in support of the reliability and security of the control, orchestration and management functional elements, systems, and devices, considered as critical infrastructure systems. It should be formulated in terms of novel architectures, interfaces, protocols, relevant data security, and overall system integrity. Also related to Research Aspect 4, this includes enablers regarding the usage of the underlying infrastructure (network sensing) for intrusion detection and prevention, to leverage the safety and reliability of network infrastructures. This aspect also should address research on the analysis of attack methodologies against network automation systems (ML-based or not), identification of their vulnerabilities and the definition of mitigation and/or defence countermeasures, or the use of distributed ledger architectures in support of network control and service management, especially in multi-actor scenarios as well as designs addressing security considerations including privacy, avoiding data exfiltration and leakage. This includes:

- Integration of legacy and quantum networking.
- Key management systems and SDN integration
- Secure Interconnection of control plane entities
- Integration of control & secure plane, role of PUFs
- Usage of distributed ledger technologies

7.7.5 Optical Network Digital Twin

This aspect covers the use of digital twins for optical networks. This includes the definition of new use cases and applicability statements of the digital twin concept (including, but not limited, to soft-failure/anomaly detection, localization, identification; dynamic operation and rollback of state changes, root cause analysis, discrete event emulation); mechanisms for state synchronization between the physical entity (e.g., optical device, link, node, network) and the digital twin; techno-economic studies related to savings associated to the used of digital twins.

- Integration of digital twins in control and orchestration architectures.
- Physical Layer Impairment modelling and impact of linear and non-linear effects

7.7.6 Research Challenges

In view of the research aspects, the next table summarizes key research challenges:

Research Theme	**Optical network automation**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Large Scale Telemetry and Efficient and Reliable Network Sensing Leverage operational and instrumental data from devices such as transceivers or ROADMs, while supporting: i) an increasingly high number of devices and ii) low-latency closed-loop systems. Joint communications and sensing Open Data Repositories Enabling controlled data exchange and common repositories from data from multiple sources (e.g., network segments) Related to Research aspects 1, 3, 4	Short-Term & Mid-Term	**Outcomes:** Optical Telemetry systems and platforms, with data models, including scalability analysis. Applications and solutions related to the use of Network Sensing **Properties and KPIs: a**ble to manage large scale systems with aggregated Terabit/s and Petabit/s telemetry data. Support hundreds of thousands of optical devices.	**Requirements:** Full network programmability and network and service automation Contributions/value: Civil engineering Distributed Sensing Health care applications Network operation Failure localization
System Interoperability enabling modular systems and avoiding lock-in; Exploit increasingly complex device/service programmability Addressing technologies such as multi-band, SDM and improved physical impairment modelling. new devices and extended capabilities. Fast Service Creation and Modification with efficient Packet/Optical integration and Networking Leveraging new pluggable/co-packaged interfaces in hybrid scenarios with low latency requirements. Optical bypasses and cost efficiency. Dynamic service provisioning and reconfiguration based on a set of constraints. Leverage programmability of optical devices with operating modes and parameters for increased efficiency. Joint access (PON/RAN) and metro/core control and resource allocation Related to Research aspects 1, 2, 3	Mid-Term	**Outcomes:** Open and Standard Data models (e.g., with YANG) for new devices or extended capabilities for 400G and beyond. Common protocols and frameworks. Promote reuse and adapting industry best practices. Applicability analysis/trade-offs in terms of speed, complexity, efficiency in resource allocation and function placement. Improved hardware designs and system control for reduced optical device reconfiguration latency. **KPIs:** Adoption by SDOs, industry actors and reference implementations. Service Creation O(seconds); Target device reconfiguration operations O(100ms), depending on device complexity and reduce service provisioning time by an order of magnitude compared to manual operation (90% reduction).	**Requirements:** High-Capacity Scaling, Cost-effective and energy-efficient. Full network programmability and network and service automation Contributions/value: Increased interoperability with reduced vendor lock-in. Increased Efficient use of optical spectrum with more advanced algorithms for resource allocation and function placement. Reduced CapEx due to a more competitive market. Reduced OpEx due to automation. Increased user satisfaction and reduced service activation times. Agile operations and reduced OpEx.

Support new services related to **infrastructure sharing enable Multitenancy** in a cloudified environment, while empowering customers with the capability to manage their own services, consolidate network slicing. Related to Research aspects 1, 2, 4.	Mid-Term	Outcomes New architectures in support of optical infrastructure sharing	**Requirements:** Cost-effective. Full network programmability and network and service automation. Network Multitenancy
Agile operation Application of DevOps principles and IT practices for network management. This includes cloudification of OSS/BSS, adoption of open-source projects and frameworks, application of Continuous Delivery/Continuous Integration, Improved workflow automation, network optimization and development of an ecosystem of suitable automation applications. Unified short-term provisioning and long-term network-planning, with closed loops at different timescales Related to Research aspects 1, 2	Mid-Term, Long-Term	Outcomes Novel network automation application ecosystem. Common software frameworks for unified short-term provisioning and long-term planning and dimensioning. Open interfaces for 1st party and 3rd party applications (e.g., resource allocation and function placement)	**Requirements:** Cost-effective. Full network programmability and network and service automation. Network Multitenancy New markets and business opportunities related to specialized services in support of network automation.
Network Domain Automation via i) AI/ML assisted decision-making processes and issuing recommendations and ii) improved resource allocation and function placement algorithms. AI/ML assisted digital twin, including MultiBand/SDM transmission modelling and physical layer impairment information for linear and non-linear effects Interfaces between AI/ML engines and platforms and telemetry/sensing systems Network Domain Automation via AI/ML with Direct Control, truly autonomous networks Network Automation via AI/ML in Cross-domain settings addressing challenges related to trust, security and optimality. Usage of Distributed Ledger Technologies for cross-domaain secure automation. Predict and/or replicate network behaviour based on potential events and actuations Related to Research aspects 1,2, 3,4, 5	Short-Term Mid-Term Long-Term	**Outcomes:** Recommendation based / Direct Control Closed-Loop network automation. Digital Twin Implementations for Optical Networks and Systems with different levels of abstraction, modularity, and reusability. **KPI:** Increased resource efficiency, reduced blocking probability or improved energy efficiency; >25% of OpEx savings compared to manual operation; Reduced time to deploy services in cross-domain scenarios over an order of magnitude shorter than current static practices. Improved prediction of network outages and service impact; Reduced rate of reconfiguration errors; More efficient network planning and capacity upgrades	**Requirements:** Cost-effective. Full network programmability and network and service automation. Network Multitenancy **Contributions/value:** Increased resource efficiency, reduced blocking probability or improved energy efficiency. Efficient network planning and optimization. Seamless and optimum capacity upgrades, including migrating scenarios.
Secure Control Systems Address security requirements incl: privacy, preventing data exfiltration, leakage, functionality bypass, spoofing, or isolation violations Secure Control Plane Joint Legacy and QKD control and interworking Integration of KMS into the SDN control plane	Short-Term & Mid-Term	**Outcomes:** New architectures and functional requirements for software systems. Development of new protocols and interfaces addressing security requirements. Applicability statement and assessment of QKD systems for critical applications **KPI:** Reduced rate of security-related incidents and lower implicationspost	**Requirements:** Secure communications, resilience. **Contributions/value:** Important implications in the overall design of control and orchestration systems. More secure systems and increased robustness against attacks.

Secure ML pipelines Related to Research aspects 1,2,3,4,5			Increased confidence of end users. Large impact on network operation (direct) and end users (indirect)

7.7.7 Recommendations for Actions

Research Theme	Optical network automation				
Action	Research Aspect 1 Telemetry and Sensing	Research Aspect 2 Architectures, Data Models	Research Aspect 3 AI/ML for Netw. Op.	Research Aspect 4 Reliability & Security	Research Aspect 5 Opt. DT
International Calls	X	X		X	X
International Research	X Promote the development of telemetry platforms enabling sharing of data is challenging in a world-wide setting.	X Align research program control and management architectures across SDOs Japan has research groups on optical networking. US several projects and initiatives are US based (TIP, ONF)		X Japan has developed programs targeting reliability post 2011 events.	X US has a solid experience on Digital Twin. The concept was initially developed at NASA.
Open Data	X Promote the sharing of (anonymized) telemetry data from operators and relevant sources		X Promote the sharing of both ML models and ML datasets, enabling reuse and robust/faster training		X Promote the design of Open Digital Twins, in a comparable way to Open-Source Software or Open-Hardware
Large Trials	X Network/Fiber sensing to be demonstrated in operators' networks in environments close to production with coexisting traffic		X Need to test AI/ML systems at a large scale, and to evaluate the required telemetry systems		X Large Scale DT need to be developed and assessed in terms of e.g. computation requirements.
Cross-domain research			X Leverage AI/ML expertise from other domains	X Strong inter-dependencies with Security Experts and Cryptography.	X Leverage the use of Digital Twins in other domains

7.8 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.

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 state-of-the-art software technology to foster efficient and secure implementation of increasingly complex network elements.

7.8.1 Quantum-safe cryptography

A signal on an optical fiber can be tapped once the physical access to the fiber 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 are essential and need 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.

7.8.2 Physical layer security

Physical layer security aims at providing alternatives to algorithmic based encryption, key-exchange and authentication. The topic is already an established research area in wireless communications. Examples are private transmission without keys, deriving secure keys from channel properties or physical unclonable functions (PUF). In many cases physical layer security primitives will be used in addition to the established algorithmic protocols since their security is based on different mechanisms and also possible attacks are very different. Hardware-based authentication mechanisms (e.g., PUF) can enable zero-trust communication and improve the security of modular systems and networks. In addition, such security anchors can be used to prevent product piracy.

7.8.3 Network resilience

Adding redundancy is the conventional, but also expensive way to improve the reliability and resilience of networks. System concepts for low cost and low power implementation of redundancy solutions (e.g., using 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. It should be possible to employ this functionality beyond the borders of a single networking domain. Also, the monitoring of optical distribution networks like passive optical networks should be improved, which is especially difficult due to their passive implementation.

7.8.4 Intrusion detection and mitigation

Based on the data generated by the monitoring solutions, data processing (e.g., by means of ML methods) can help to detect upcoming or hidden problems early and counteract them in advance with the available flexibility. While machine learning often has a competitive edge over conventional algorithms, in most cases it is not clear how a certain ML method gets its results (ML as a ‘black box’). This might lead to unexpected behaviour (e.g., if the input data is not within the standard range), and could be used to fool an ML implementation (c.f., adversarial ML). Secure usage of ML requires exhaustive testing and ideally some degree of explainability, i.e., some insight into the inner working of the algorithm.

7.8.5 Research Challenges

Research Theme	**Security for mission critical services**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Long distance QKD QKD devices for Telco environments with sufficient secure key rates and an attenuation budget considerably beyond current limitations.	Mid-term	KPIs: Attenuation budget, secure key rate	Secure communications. Data privacy and improved security for critical infrastructure
HW-based authentication Concepts for secure authentication and device authenticity	Mid-term	Optical component or macro to include into PIC Improved entropy vs. commercial solutions	Secure communications. Data privacy and improved security for critical infrastructure
AI/ML based intrusion detection Methods to enable secure and robust intrusion/anomaly detection	Mid-term	Truly reliable and secure AI/ML algorithms and concepts to ensure this	Secure communications. Data privacy and improved security for critical infrastructure

7.8.6 Recommendations for Actions

Research Theme	**Security for mission critical services**		
Action	Quantum-safe cryptography	Physical layer security	Intrusion detection
International Calls	X	**X**	X
International Research	X Leverage knowledge in European but non-EU countries	X Cryptography is a truly international endeavour	X Benefit from strong AI/ML footprint in the US
Open Data			X Open training and test data can be essential to progress the community
Large Trials			
Cross-domain research	X Strong interdependencies with classical security	X Strong interdependence with classical security	X Leverage knowledge from strong AI community

7.9 Ultra-high energy efficiency

Communications networks are a pillar of our daily life. It enables to connect between people and machines, providing an enormous range of services.

The ongoing pandemic has strengthened the need for data connectivity relying on higher capacities and forecasts agree that the traffic growth will further intensify, at an annual rate well above 20% that has to be matched by the capacity of the optical transceivers and line systems. At the same time, requirements for reduced latencies in data processing as well as higher energy-efficiencies, become more stringent. Especially the focus on greener networks and processing systems, becomes a top priority for our society. Currently the

ICT sector accounts for 5-9% of electricity use and around 3% of the global greenhouse emissions29. The EC estimates that at this pace, the ICT footprint could increase to 14% of global emissions by 204030,31. Hence, solutions to make sustainable both power consumption and footprint of world-network infrastructures are mandatory. As fibre optics and photonic devices are the key technologies underlying the worldwide telecommunication infrastructure, it will play a highly relevant role in reducing the total power consumption of the ICT. A matter that has been recently highlighted also by the activity of the F5G within the ETSI on green energy. In this context, some key questions arise:

1. How to reduce the overall network energy consumption while increasing capacity?
2. How can photonic integration help to reduce energy consumption? How to replace bulk-optics with PICs for subsystems/functionalities like WSSs/spectral-lanes switching or to replace power consuming electronic processing?
3. How can network controllers be used to reduce energy-critical network functions?
4. Which functions are energy-critical?
5. How to make next-generation networks sustainable and multi-generational so that electronic waste is drastically reduced?
6. Can AI-based network management help to reduce energy consumption?
7. How to optimize service placement in the networks?
8. How to optimize the processing and amount of information, to avoid this being carried out every time and everywhere?
9. What is the impact of distributed compute & storage resources for low latency services on network energy consumption?

The following areas of research might help to answer the above questions

7.9.1 Simplified and fully configurable flexible E2E optical networks

Optical networks consist of a variety of domains and segments, which are traditionally being individually optimized using specific solutions. This might be cost-effective locally, but not in an end-to-end (E2E) fashion across the optical domain. However, as many services, e.g., in access, are terminated after ~20 km, E2E is possible only by assuming some form of bypass.

Next-generation optical networks will cope with even more stringent requirements than today. Existing architectures are sub-optimal and introduce too many unnecessary processing stages and overhead or are relying in electronics processing that introduces scalability limitations due to limited bandwidth. In this sense and from a purely “hardware” perspective, all signal processing functions that can be moved from electronics to photonics will contribute to the saving of energy consumption. Consequently, we must simplify it by developing new network architectures and technologies that enable efficient multi-layer/domain IP routing & optical transport integration. This also requires a high level of intelligence, easiness in managing and deploying. Such a solution would help full configurability of E2E connections, better planning, and lower power consumption.

Lines of research in this field might include: (i) development of technologies that consent to remove unnecessary opto-electronic-opto (OEO) conversions and processing to decrease power consumption and

[29] Although it is worth mentioning that thanks to ICT the number of travelling is reduced.

[30] Shaping Europe’s digital future.

[31] This is gross amount, which does not consider the reduction of power consumption enabled by the ICT.

footprint; (ii) realization of intelligent and configurable components and that can be operated via software, and optimized, e.g., for performance or low power consumption; (iii) Simplification and optimization – by minimizing the amount of packet processing of information, i.e., avoiding when possible IP routing – of the way information is processed. Information is analogue in nature and is important to identify how/where it can be more efficiently processed with electronics, or photonics in the digital (electronics) or analogue (electronics/photonics) domains. This could benefit from concepts relying on electronics/photonics co-integration and by the introduction of configurable programmability in photonics; (iv) optical packet/burst switching in wide network enabled by novel components and new concepts (like e.g. deterministic networking), and by a close research collaboration – aiming also at standards – among operators, vendors, and component manufacturers; (v) realize truly cloudified software-based configurability at the component/subsystems level so that each building block is optimized for specific applications. Programmability and softwarization should be also used at the platform level to optimize the implementation of designs and the performance of the resulting components/subsystems while reducing power consumption and footprint, and finally at the network level with the goal to enhance capacity allocation, reduce OPEX as well as increase network reliability (proactively identify potential failure, optimize planning and operations, etc.)

7.9.2 Energy efficient transceivers

Although the power consumption of individual transmitters and receivers is negligible – in comparison to the total consumed power by the network – their pervasive use and scaling to higher-rates requires careful engineering to achieve the targets of energy savings. There exist several ways to reduce the energy within the transceivers: (i) modern transceivers allow a multitude of transmission modes. These could be exploited to minimize consumed power and spectral occupancy. For example, the FEC overhead can be tuned based on the current specific margins, or, similarly, the spectral occupancy could be adjusted according to the given traffic demands that need to be served. Autonomous and flexible transceivers, also in terms of adaptive spectral occupancy, need to be flexible and follow the dynamically changing traffic variations; (ii) Specific power-hungry functions could be outsourced to photonic devices such as optical FPGAs, which enable dynamic and energy efficient processing via PICs. For instance, the dispersion compensation within the DSP could be carried out by self-adaptable programmable optics, at least for specific link parameters. Optical FPGA could also offload functions/calculations in data centres and high-performance computing infrastructures; (iii) Parallelism in the optical domain is key for scaling/improving performance and can rely on either the spatial (i.e., SDM) or the spectral (i.e. UWB) domain. Transceivers and their building blocks can be designed such that they can be reconfigured and reduce the consumed energy when they are operated under certain conditions; (iv) Co-packaging can significantly reduce the power needs of next-generation optical transceivers, including pluggable ones. Research could deal with the different approaches to interface photonics to the electronics while considering how close the photonics can be to the electronics. So far two solutions have been proposed: (a) low-data rate and (b) high-data rate but limited by SERDES. Research is needed to overcome the limitation caused by co-packaging; (v) Photonic programmable chips to replace ASIC partially or totally.

7.9.3 Energy-aware optical networks and components

Nowadays, optical network architectures are not optimized to minimize the power consumption and achieve energy awareness, although the transition to a greener world is becoming a major sustainability concern for our society. Furthermore, the associated components are also not optimized to avoid waste of energy. As a result a series of new research directions supporting the greener networks is listed hereafter: (i) optimized design of components via parallelism. Often it is not needed to use the best performance, and part of the components composing an optical transmitter might be switched-off, e.g., a DAC does not always need to deliver the best performance, and if intelligently designed, part of it could be switched off - e.g., by turning off un-used capacity and active components, also as function of the instantaneous traffic; (ii) Usage of network telemetry approaches to monitor the actual consumed power and enhancement of control plane operation to enable energy-aware network via software configuration. Guidelines for system CO2 monitoring and requirements; (iii) optimize the application server locations to minimize the power consumption, and latency; (iv) enable network operators to minimize energy and resource requirements through load-adaptive network control; (v) optimized How to optimize services placement in data-centres across the networks? This cannot be done only by networking. Optimize low latency. Optimize where the resources are available.

7.9.4 Zero-electronic waste and scalable optical networks.

Existing optical networks are not scalable. Upgrades to new generations are carried out by generating large electronic waste (e-waste). For example, when moving from 10G to 25G PON, all boxes and transceivers at the end user and within the electrical aggregation stages need to be replaced, regardless it is required or not. Next generation optical networks need to be further developed by relying on improved technologies that enable programmability/reconfigurability, so that the network can be upgraded in a dynamic way only when and where is needed. Some lines of relevant research may include: (i) development of technologies that permit the co-existence of multiple generations of optical devices so that upgrades are local and not network wide. These novel approaches should also enable full interoperability among different domains and vendors; (ii) design of components and network elements e.g., racks, switches, so that the overall power consumption is reduced, and that scalability is optimized; (iii) utilization of new materials and new fibres – e.g., with lower attenuation and nonlinearity – so that the transceivers and network can be simplified; (iv) enable techniques supporting reconfigurability; e.g. to transmit only what it is needed and process only what it is required. E.g., nowadays we transmit 100G / 400G, regardless of the real traffic which is actually transmitted; (v) Consider new cooling techniques by involving collaboration with other fields of research.

Research Theme	**Simplified and efficient high-speed optical networks**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Flexible E2E optical networks Current optical networks are divided into too many domains that require the usage of different OEO conversion stages. Programmable Integrated Photonic Processing hardware Introduce new Optical Node and Transceivers Architectures that are fully dynamic and configurable to support intelligent on demand processing of traffic in an optimized manner.	Long-term (finished within 5-7y)	An ecosystem of devices and components to create E2E flexible optical networks. Highly scalable and flexible, same E2E technology. **KPIs**: (I) reduction of 50% of the transceivers	Sustainability and sovereign digital infrastructure. Adding software-defined intelligence and dynamically optimized performance (spectrum management, capacity allocation and energy saving) to photonic transceivers and switching elements. Incorporation of a future-proof and scalable matched interface between the fibre and the wireless segments of access networks to support 5/6G and future extensions conveying data and sensing information.
Zero-electronic waste and scalable optical networks	Long-term (finished within 5-7y)	A comprehensive strategy to realize fully scalable optical networks.	Sustainability

Current optical networks are not multi-generational. When a new generation is introduced, old and low-speed transceivers cannot talk to newer and higher-speed ones. Need to introduce network architectures and device technologies that are scalable and configurable for that to allow network upgradability. Basic building-blocks of past high-capacity backbone networks/systems may be reutilized in the future in lower performance access/short reach infrastructures (in line with the concept of Cyclic-Economy).		KPIs: (I) 0.15 W/Gbps, a 90% reduction respect to 100G / wavelength platforms in data centers and their interconnect.	

7.9.5 Recommendations for Actions

Research Theme	**Simplified and efficient high-speed optical networks**		
Action	Research Aspect 1	Research Aspect 2	Research Aspect N
International Calls		X Short Comment	
International Research	X Short Comment		
Open Data	X Short Comment		
Large Trials			X Short Comment
Cross-domain research			X

7.10 Optical integration 2.0

The foundation for the development of cost- and energy-efficient systems with high reliability lies in the integration of multiple optical and electrical functionalities, as scaling down the number of high-speed interfaces will reduce the power consumption of the network components. In addition, improving the repeatability of manufacturing will increase the reliability of photonic components and reduce their cost. At the same time, the performance of the components needs to be enhanced to support a wider spectral range for new optical bands and higher speeds for increased data rates per channel. These challenges require the investigation of novel material platforms and, ultimately, the combination of multiple platforms up to the manufacturing level.

To meet all these challenges, the following aspects must be considered in order to enable the photonic layer to support the challenges of the system and networking layers.

7.10.1 Multi-band exploitation

Further expansion of overall system capacity requires exploitation of fiber wavelength windows beyond the C-band. To this end, passive and active optical components need to support ultra-wide-band operation with optical bandwidths exceeding 100nm, posing challenges from the point of view of component design and material properties alike.

7.10.2 High-capacity interfaces for spectrally and spatially multiplexed systems

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.

7.10.3 New materials

Moving to higher channel bandwidths, further performance gains are necessary and might be achieved by (monolithically) adding organic or ferro-electric (e.g., $BaTiO_3$) materials into the Silicon platform, providing potentially very high electro-optical coefficients and reducing the required driving power. This will be beneficial for the integration of optical functionalities on every size scale and allows to meet the demand to drive up analogue bandwidth and data rate per port by a factor of 10 by 2030. Other new materials, e.g., Lead Zirconium Tantalate (PZT) for phase actuation and thin film lithium niobate (TFLN) for modulation, have a potential to significantly contribute to the exploitation of low-power actuation and switching and a wider optical bandwidth range through high performance modulation capability in combination with negligible parasitic propagation losses. The integration with mature photonic platforms such as InP enables on-chip laser integration, leading to compact, energy-efficient, and low-cost transceiver solutions.

7.10.4 Optical chip interconnects

Advances in electronic integration follow Moore's law and 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. Silicon is known to provide good passive optical properties for routing, modulation, and detection of light. 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 micrometre pose a limit on the structure sizes. Further integration of photonic and digital processing functions will require scale adaptation.

7.10.5 Multi-platform manufacturing

For the optical transport use case, typical volumes are in the range of only ~100,000s/year/design, while quality and performance of the transceivers is paramount. These volumes are subscale for typical silicon semiconductor fabs. Scaling III/V wafer processes to 4" and 6", hybridly cointegrating III/V actives with low-loss passive waveguides like stoichiometric Si_3N_4, is essential to scale up optical transport with the ever-increasing long distance internet traffic.

7.10.6 Photonic-electronic integration

As the demand for on-chip functionalities continues to grow, it is expected that electronics technology will keep on focusing on the integration with photonic circuits to keep up with the Moore's Law scaling. The types of electronic-photonic integration are monolithic, heterogeneous, or partially monolithic. The unique advantage of Si photonic-based photonic integrated circuits (PICs) is the CMOS compatibility which enables the seamless integration with electronics and thus, the realization of cost-effective, high yield manufacturable solutions.

7.10.7 Reliability and repeatability

To foster commercial exploitation, PICs require a high degree of repeatability in manufacturing under somewhat varying processing, material and layout conditions. Establishing mature Process Design Kits (PDKs) requires robust passive and active optical components that perform within narrow margins over multiple dies, wafers and wafer-runs. Reducing statistical deviations in PIC performance as opposed to setting up hero-experiments and focusing on on-wafer testability of PICs is likely to enhance industrial uptake.

7.10.8 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Optical Integration 2.0**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 To utilize the full capacity of the fiber, new materials and designs for optical wide-band operation must be developed.	Mid-term (finished in 5y)	Optical components for multi-band operation, based on new materials or robust building blocks **KPI:** Optical operating bandwidth >100 nm	High-capacity scaling and reliable connectivity Increased system capacity while using existing fiber infrastructure
Challenge 2 Parallelized optical transceivers, supporting multiple channels in the spatial and wavelength dimensions	Mid-term (finished in 5y)	Modular transmitter/receiver components to support multiple optical channels Concepts for multi-band operation with unified transceiver components **KPI:** Transceivers for termination of >8 channels with power consumption of < 2 today’s channel equivalents Optical operating bandwidth >100 nm	High-capacity scaling and reliable connectivity; Cost-effective and energy-efficient systems Increased system capacity with less space and less energy consumption
Challenge 3 Integration of optical interconnects between electrical processing modules on a chip	Mid-term (finished in 5y)	Optical interconnections on a chip using multiple material systems including on-chip mode matching / spot size converters Electro-optical interconnection on the same chip allowing monolithic co-integration of RF electronics and photonics **KPI:** Energy consumption < 16 fJ / bit Interconnect loss <0.5 dB/facet	Cost-effective and energy-efficient systems; Advanced electro-photonic integration Higher chip processing power with less energy consumption for connections between processing modules Increased RF-bandwidth systems at reduced footprint, lower cost and lower power consumption.
Challenge 4 Multi-platform manufacturing	Mid-term (finished in 5y)	High-performance, long-distance transceivers using cointegrated best-of-class materials **KPI:** Full O- to L-band coverage Power consumption 2-4x below SotA	High-capacity scaling and reliable connectivity; Cost-effective and energy-efficient systems Increased system capacity with less space and less energy consumption while using existing fiber infrastructure
Challenge 5 Reliable, repeatable and testable PDKs	Short-term (finished in 3y)	PDKs that perform within narrow margins under varying conditions over chips, reticules, wafers, wafer-runs.	Cost-effective and energy-efficient systems; Advanced electro-photonic integration

		Electro-optical on-wafer test vehicles enabling simultaneous testing of optical, DC and RF parameters	Increased reliability of PICs, higher value of foundry services, reduced barriers to enter PIC market

7.10.9 Recommendations for Actions

Research Theme	Optical Integration 2.0				
Action	Multiband exploitation using new materials	Multi-channel transceivers	Integration of chip interconnects	Multi-platform manufacturing	Reliability and repeatability
International Calls	X	X	X	X Establishment of European multi-platform fabs for co-integration of multiple photonic IC technologies	X Improvement of fab performance
International Research	X European cooperation required to achieve goals	X European cooperation required to achieve goals	X European cooperation required to achieve goals		
Open Data					
Large Trials					
Cross-domain research	X Cooperation with Photonics21 PPP	X Cooperation with Photonics21 PPP	X Cooperation with Photonics21 PPP	X Cooperation with Photonics21 PPP	X Cooperation with Photonics21 PPP

7.11 Optical access for FTTH and beyond

The evolution of optical access technologies are still driven primarily by Fiber-to-the-Home (FTTH) network architectures and services. The economies of these networks demand ultra-low cost, still highly performant optoelectronic and digital processing capabilities and since the introduction of 50G-PON also digital signal processing means. The optical system and network architectures must be optimized for efficiently aggregating dynamic traffic to and from many nodes in each area while respecting differentiated service requirements, and at the same time meeting challenging physical layer specifications.

After two decades of increasing system capacities (bit rates) and optical power budgets on the one hand, and of refining the TC Layer (Transmission Convergence, i.e. a system-wide MAC layer) for supporting differentiated service requirements on the other hand, PON-based access solutions have now reached a level of technical maturity and cost efficiency that makes them attractive also for other many-nodes/short-reach networks beyond traditional FTTH. First new market segments can be addressed already today using existing PON systems without the need for modifications: commercially available Passive Optical LANs to be used in place of conventional office and enterprise LANs, as well as for backhaul links - and fronthaul in near future - in small cells mobile networks.

Building on this solid foundation, the future evolution of optical access technologies and architectures will bring about further increased system capacities, highly flexible system and network reconfiguration, meshed and resilient network topologies, coexistence of best effort and deterministic traffic, secured transmission over complex architectures, and much more as will be pointed out in the subsections below. The objectives shall be to make these solutions suited for diverse applications and network scenarios in public and private area environments, with single- and multi-tenant business models, for vertical markets and industrial applications, for 5G and emerging 6G mobile networks, for small intra-datacenter networks, for in-home networking with fiber-to-the-room, for energy grid control with fiber-to-the-grid and more to come.

New opportunities also arise from the availability of a mature passive optical distribution network. On one hand other systems such as quantum-secure virtual point-to-point links can be considered to coexist with legacy and future PON systems on the same fiber plant. On the other hand, the fiber plant itself is being eyed for the use case of distributed fiber optical sensing allowing environmental insights from the fiber plant and the attached systems to enhance reliability and resilience as well as to open new and adjacent markets.

7.11.1 Increased capacities and flexible configuration of access transmission systems

Optical systems for access networks today are worldwide predominantly based on TDM-PON technologies (GPON family (ITU-T) and EPON family (IEEE)). Single channel line rates of commercial systems will soon reach 50 Gb/s, employing NRZ IM/DD modulation and reception schemes, supported by FEC and DSP for mitigating noise and dispersion induced transmission errors. A very high-speed passive optical network (VHSP) supplement project has been started in ITU-T SG15/Q2 with the goal to evaluate system technologies for PON capable of line rates beyond 50 Gb/s. Candidate technologies include IM-DD PON for 100 Gbit/s and higher, as well as coherent PON for 200 Gbit/s and beyond. IM-DD technology has so far been applied in all existing PON system generations including GPON, XGS-PON, 25GS-PON and 50G-PON, because of its simplicity. It is expected that continuing this progression for 100G-class systems will be desirable to satisfy the expectations of network operators. However, longer fiber reach, coexistence challenges and achievement of highest optical power classes demand for further investigations of multilevel signaling (real or complex valued) and field modulation/coherent reception. In an alternative approach, multiple wavelength channels can be combined for achieving higher system capacities beyond 200 Gb/s. For addressing end nodes (users) in complex scenarios with diverse loss budgets and link distances, a third deployment option will comprise a combination of different line rates, modulation formats and FEC / DSP levels on single channels or across multiple parallel wavelength channels. This approach is similar to the Modulation Coding Scheme levels in 5G radio transmission but has not yet been applied in optical access systems. Anticipating complex architectures in future access and similar short reach optical networks, this approach will enable graceful and cost-efficient upgrade options in deployed fiber networks.

The focus of this work shall be on design studies and proof of concept demonstrations of suitable system and network configurations, as well as of enhanced system protocols for managing and operating such multi-dimensional ultra-high-capacity transmission schemes in most flexible and efficient ways, making them useful for many application scenarios as indicated in the introduction of this section. Along with TDM-PON based wavelength channels (TWDM) also non-TDM wavelength channels shall be considered for supporting sustained unshared high-capacity or secure links to individual nodes in the network, preferably as a combined flexibly configurable TWDM/WDM-PON solution. Beside such high-capacity approaches for PON also new use cases for in-home deployments (FTTR) have to be evaluated leveraging GPON and XGS-PON line rates but demanding for another ONU cost erosion taking into account reduced budget classes. For such cases also the capability of centralized management from the operators DC or CO are a key desire.

7.11.2 Flexible realtime and non-realtime resource assignment

In TDM-PON for residential users, the transmission bandwidth per end node (user) is assigned on a non-realtime basis, taking into account predefined minimum (guaranteed), assured and maximum bandwidth values per service, as well as the observed utilization of previously assigned bandwidth and on status reports from ONUs. This established process shall be further improved for optimizing the time varying bandwidth assignments towards more precisely meeting the actual service requirements and thus increasing the overall bandwidth efficiency of TDM-PON systems. For instance, more precise prediction of near future bandwidth requirements per service are expected to be achievable by adding appropriate AI/ML algorithms that can account for observed traffic patterns.

Network slicing in TDM-PON for enabling multi-tenancy business models is another topic that needs precise modeling, comparison and implementation of different bandwidth assignment strategies to enable meeting different prioritization and fairness KPIs, as agreed upon in the operators' SLAs.

The TDM-PON bandwidth assignment algorithms mentioned above are set to accommodate all services (on average over multiple PON frames (125 µs)) in the best possible way before allocating time slots for transmission. However, for challenging service requirements such as low latency, ultra-low jitter, low packet loss rate etc., more predictable, i.e. deterministic, assignment strategies are needed. Direct time slot allocation per service and per frame is a promising approach. For client traffic with precisely predictable bandwidth needs over time (e.g. constant or strictly periodic), the slot allocation on TDM-PON can be preconfigured. For dynamically changing bandwidth needs, a low-level logical interface can be used for mutual realtime communication between client and TDM-PON to dynamically predict varying near future client needs and PON capabilities (e.g. CTI (Cooperative Transport Interface) for 5G fronthaul links over PON). The above strategies are rather simple to design for a single node or only a few nodes on the network. For many nodes, however, and especially when designing for an efficient utilization of the available transport bandwidth, appropriate strategies and algorithms are needed.

Meeting low latency and low jitter KPIs needs realtime and even isochronous realtime synchronization between network nodes, depending on the service precision requirements. The implementation of suitable processes for frequency syntonization and ToD (Time-of-Day) synchronization is addressed in the Deterministic Networking section of this document. However, the implementation of highly precise ToD synchronization processes in PON supporting many end nodes needs special consideration, in particular for sub-nanosecond precision that is required e.g. for precise positioning use cases. The support of time sensitive networking profiles with PON systems become important for fronthaul or industrial applications.

In complex access networks with different segments and technologies (application, mobile, optical, computing), the non-realtime orchestration of coordinated individual realtime resource assignments is crucial for meeting stringent service requirements. Flexible reconfiguration of logical channels, involving time and spectral domains in different segments will add another level of complexity that needs to be addressed in this challenge.

7.11.3 Redundant, meshed and flexible optical layer network architectures

Optical access networks today are deployed on point-to-point (ptp) or point-to-multipoint (ptmp) fiber optical distribution networks (ODN), employing passive optical power splitters (for TDM-PON and WDM-PON) or passive wavelength routers (for TWDM-PON and WDM-PON) for distributing optical signals to the end nodes. The ODNs implement a logical (via time slots in TDM-PON) or physical (via wavelength channels in WDM-PON) star topology from the OLT to all ONUs, in the majority of cases without redundant attachment.

Critical services will require redundant system layouts for improved resilience and ultra-high network availability and service reliability. Redundant ODN layouts have been described for simple tree architectures already in early GPON documents. More sophisticated and resilient architectures will, however, be needed for certain use cases e.g. in the industrial space or in small cells x-haul networks. These architectures shall provide resilient connections in the distribution and drop section of the ODN, and at the same time allow for redundant attachment to the metro or equivalent aggregation networks.

Some latency sensitive use cases will (in addition) benefit from short local interconnects between ONUs in the same ODN or in neighboring ODNs, without going through the central node (OLT). Employing selective optical loop-backs at the passive remote nodes will establish a local meshed topology among the ONUs. Depending on the required interconnection patterns and the group size of interconnected ONUs, as well as on the required local link capacities and allowed latencies, different solutions shall be devised for optimized bandwidth efficiency, cost efficiency, energy efficiency and other KPIs. Additional optical ports on the ONUs may be considered for this overlay network. Solutions are sought for ptmp ODNs (PON), but as well for ptp ODNs.

Deployment related, operational, network migration related considerations are calling for reconfigurable (nominally passive) remote nodes in the ODN. This reconfigurability shall allow for flexible reconfiguration of the interconnection topology on an ODN on time scales well above milliseconds, with low energy consumption as much as possible. Suitable remote node solutions along with the supporting network architectures, including the associated management tools for configuring the remote nodes shall be elaborated and practically demonstrated.

7.11.4 Optical layer multi-tenancy in access networks

The cost of rolling out new fiber connections in access has frequently been a major blocking point in providing early FTTH services and is again considered a blocking point for optical x-haul in future small cells 5G and 6G mobile networks. Sharing the fiber infrastructure among multiple players (either competing or complementing each other in their service portfolio) may become a beneficial business model for such scenarios.

Neutral host models have been implemented in some optical access networks in Europe and in Asia. Those are based on one player leasing a dark fiber infrastructure to one system and service provider, but sharing a dark fiber infrastructure on the optical layer among multiple system and service providers has not been implemented yet. For making this a viable business proposal, appropriate technical additions to current ways of deploying and managing fiber networks as well as operating data services on such networks must be developed. For instance, a neutral host providing the dark fiber must be able to monitor and manage the ODN, however, without having access to telemetry, OAM and management data provided by the service networks using the infrastructure. On the other hand, sharing a fiber link in the wavelength domain e.g. needs definite upper bounds on optical power and other optical transmission system parameters in order to avoid mutual linear and non-linear signal distortions. Related implementation and operation details must be elaborated, both for the neutral host and for the system and service providers, and suitable business models must be evaluated. As a side remark: one such service can be provided by an additional player using the fiber for various kinds of optical fiber sensing projects.

7.11.5 Energy efficiency in PON

Means of energy efficiency in PONs are a critical area aimed at reducing the power consumption of these systems, which are essential for operators in their mission to become CO2 neutral in the future. The research challenges encompass a variety of techniques to enhance energy savings with the goal to avoid compromising performance. Key strategies include the implementation of low-power modes for ONUs during periods of low data demand, optimizing dynamic bandwidth allocation to minimize idle power consumption, and developing energy-efficient protocols that adapt to traffic variations. Additionally, advancements in the design of optical components, such as energy-efficient transceivers, are important. These efforts collectively contribute to making PONs more sustainable, reducing operational costs, and mitigating the environmental impact of expanding data networks.

7.11.6 Distributed fibre optic sensing

Distributed fiber optic sensing leverages the PON infrastructure and PON transceiver technology for advanced monitoring and diagnostic capabilities across deployment areas. This technique involves integrating sensing technologies with fiber optic cables to detect changes in temperature, strain, vibration, and other physical parameters along the length of the fiber. By using methods such as Rayleigh, Raman, and Brillouin scattering, distributed fiber optic sensing enables real-time, high-resolution data collection over large distances, making it invaluable for applications in structural health monitoring, environmental sensing, and security. Also PON transceivers can be utilized for investigating changes in state of polarization or phase. The integration with PONs offers a cost-effective and efficient means to deploy these sensors, utilizing existing telecommunication networks to provide both data transmission and sensing functionalities, thus enhancing the overall utility and value of the fiber optic infrastructureincluding fault management, ODN visualization, and environmental change perception. Key challenges that need to be overcome are the potential cost impact to PON transceivers and the ambiguity of sensing events that can arise from the point-to-multi-point infrastructure.

7.11.7 Research Challenges

Research Theme	**Optical access for FTTH and beyond**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 : Optical systems capacity increase beyond 100 Gb/s by combining diverse modulation coding schemes on multiple TDM and WDM channels (high-speed or quantum secure)	Short-term (finished in 3y)	Scalable architecture solutions and TDM/WDM PoCs for well beyond 100 Gb/s system capacity, applicable to diverse use cases as addressed in the introduction of the section SotA : single TDM-PON channel per system with up to 50 Gb/s serial bit rate (ITU-T), or 2 parallel 25 Gb/s channels aggregated by WDM (IEEE)	Addresses the High Capacity Scaling and the Cost-effective and Energy Efficient Systems requirements of the target vision
Challenge 2 : Dynamic resource assignment, synchronization and orchestration	Short- to Mid-term (finished in 5y)	Non-realtime assignment (2y) SotA : no AI/ML involved Realtime assignment (3y) SotA : first research proposals Precise synchronization (5y) SotA : N/A Orchestration of assignment and synchronization (5y) SotA : N/A	Addresses the Deterministic Networking, the Efficient Integration of optical technologies for radio access networks, and the Network Multi-tenancy requirements of the target vision
Challenge 3 : Redundant, meshed and flexible optical layer network architectures	Short- to Mid-term (finished in 5y)	Quantitative elaboration of redundant network architectures for critical services (2y) Full network architecture studies and PoC for meshed ODNs (3y) Full network architectures and PoC supporting flexible reconfiguration	Addresses the Reliable Connectivity, the Cost Effective and Energy Efficient Systems, and the Deterministic Networking requirements of the target vision

Challenge 4: Optical layer multi-tenancy in access networks	Long-term (finished in 7y)	Elaboration of technical add-ons and design guidelines for enabling optical layer multi-tenancy in fiber access networks, and related business models SotA : N/A	Addresses the Network Multi-Tenancy, the Cost Effective and Energy Efficient Systems, and the High-Capacity Scaling requirements of the target vision
Challenge 5: Distributed Fiber Optical Sensing in PON	Mid- to long-term (finished 7y)	Understanding technical requirements and transceiver design guidelines for DFOS and identification of individual branches in a ptmp network scenario. SotA: N/A	Addresses new use cases of the underlying fiber infrastructure of a PON ODN for sensing requirements of the target vision

7.11.8 Recommendations for Actions

Research Theme	**Optical access for FTTH and beyond**		
Action	Research Aspect 1	Research Aspect 2	Research Aspect 3
International Calls			
International Research	**X** optical components related		
Open Data			
Large Trials			
Cross-domain research	**X** optical components related	**X** AI/ML related	

7.12 Optical wireless transport and access

Future networks will be characterized by ubiquitous and pervasive high speed access of users, often in mobility conditions and in high-user densities. Among the technologies that will be needed to support this trend, Optical Wireless Communication (OWC) is emerging as a promising candidate for massive deployment, covering a wide range of applications. As a broad-spectrum technology, OWC can be assimilated to RF wireless, i.e. a family of many different types of systems, with completely different specs and implementations. As opposed to RF, OWC has potential to provide much higher speed (>>Gb/s) and support high density of users. It can also allow for enhanced security of the information, accurate localization/navigation in indoor and support wireless communication in extreme environments (e.g. underwater).

It is then of paramount importance to view what are the specific areas that can allow for the highest gain on a long-term perspective, especially considering the seamless evolution of 6G networks towards new application areas and ubiquitous connectivity

7.12.1 High-speed Outdoor OWC

Outdoor OWC, often indicated also as FSO (Free Space Optic), can provide high-speed wireless connectivity (>100 Gb/s) to/from fixed and/or moving nodes. FSO can be used to provide fixed point-to-point front-haul connectivity to base-stations. Highly challenging are all applications involving links from ground to drones or high-altitude platforms (HAP), such as the large drones used to provide wireless connectivity to remote areas (see Aquila Project by Facebook). The ultimate frontier would be to exploit FSO in satellite networks, with a complete merging of terrestrial and non-terrestrial networks (NTN), taking advantage of the unmatched speed that FSO can provide (>100 Gb/s), compared to other consolidated technologies.

In all these areas, few remarkable examples prove its viability (e.g., Starlink optical inter-satellite links, NASA laser-link communication to Moon, NASA cubesat-to-ground 200 Gb/s experiments), however the final use of FSO is far from being established and large R&D areas need to be effectively covered.

Developments of outdoor OWC is at its early stages: very limited examples exist of outdoor OWC systems, typically limited over <1 km reach and today OWC systems for NTN inter-satellite links are being proposed by private companies. The key benefit of NTN-OWC is that it can leverage upon existing devices and subsystems developed for fibre communications, However, a large RD effort is required along many research paths. First of all, in terms of new components (telescopes, detectors, tracking algorithms): there is no de-facto architecture for the hardware, which should be sizeable, the solutions for pointing alignment and tracking (PAT) must be properly validated, and the use of adaptive optics (wavefront sensors, deformable mirrors, control software) has to be exactly validated. Moreover, channel modelling of FSO links passing through atmosphere is challenging and not definitely consolidated): it is left for further investigation how slow effects (thin cloud) can determine rerouting of channels and how fast effects (due to turbulence, pointing errors etc.) can be coped by proper system design.

As transversal feature, future OWC can require development of new component technologies. Current systems use existing sources and detectors, adapted, not yet optimized, for OWC: these can be replaced by *ad hoc* developments. Furthermore, key elements, improving performance, footprint, and reliability, could exploit reconfigurable metasurfaces, which can deflect/optimize the wavefront in the optical domain (as currently expected also in RF-domain).

Finally, the typical effects of random unavailability / fadings are expected to have an impact in the network protocols and configuration tools: it is widely expected that channel features should be considered in the choice of services to be deployed over these links whilst the network architecture should be properly adapted and optimized.

7.12.2 Indoor OWC

Different types of OWC systems can also allow for wireless communication in indoor environments. These can support wireless connectivity with a range of different services.

1) indoor navigation for mobile users: can allow users to navigate in (unknown) indoor environments, such as hospitals or museums, thanks to Optical Camera Communications (OCC): a common smartphone, with suitable app, can detect properly modulated ID’s from ceiling LEDs and convert it to meaningful information to the user; a similar technique can be used to guide robots in industrial sites (possibly not using a camera); the reverse approach allows the control centre of the infrastructure (mall, ship etc.) to quickly locate human users in case of emergency
2) high speed access of mobile users (>1 Gb/s) by OWC (sometimes indicated as LiFi), can complement the existing WiFi solutions, providing higher (unshared) bandwidth and much higher density of users than current WiFi; the use of ligthwaves also supports intrinsic security at physical layer, since light can be simply shielded from external sniffers
3) very high speed OWC can be exploited in datacenters, to realize point-to-point links, within specific spatial architectures of Top-of-the-Rack switches;
4) High security wireless data transfer applications with indoor OWC. As light does not penetrate walls, the efficient protection of wireless data is enabled by the unique properties of light. The physical contact to the wireless network can be prevented by using simple opaque layers (e.g., cardboard, plastic, etc.) and sensitive data in medical, industrial or financial sector stay safe. Next to the new dimension of hacking-robustness, also the jamming-robustness of the optical signals is significantly increased.

Large efforts, also funded by EU and national governments, were devoted to LiFi developments in the past few years. There is currently no commercial solution, due to several open issues, such as compatibility with existing protocols in the other parts of the network. Unsolved questions are about the key issues that, till now, prevented wide deployment of the technology.

On the other side, OCC can be suitable for faster applicability, but till now it has received minor attention. It still needs to be investigated, especially in view of social impacts, for emergency and impaired people, as well as for all possible industrial applications. Key issues to face are about energy-efficient solutions that suitably combine LED modulation and signal processing, achieving both low battery-consumption and high location precision.

7.12.3 Underwater OWC

Today, very limited is the knowledge that mankind gained about the underwater areas, which are of crucial relevance for commercial, historical, scientific and strategic issues. Exploration, environmental monitoring and exploitation of the underwater world is now open to activities carried out by autonomous robots, which have presently very limited wireless communication capabilities, whereas cabled communications have obvious limitations. Yet, Underwater OWC (UOWC) is the only feasible means to transmit high-speed data by wireless signals underwater, because other electromagnetic waves are hugely attenuated by water, whilst the only wireless signals that can be used (acoustic waves) present great limitations in terms of bit rates (kbit/s), and also can have significant issues in terms of security (sound waves travel over long distances).

Although in Europe we demonstrated the first 10-Mbit/s real-time transmission in real sea-waters, obviously this requires further optimizations and developments for further applications. We should aim at extending the reach, realibility and speed of the UOWC system, but also test deeply in more complex configurations, such as where robotic swarms can cooperate effectively, thanks to effective communications.

7.12.4 Research Challenges

Research Theme	Optical Wireless Communications		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Stable Outdoor OWC links (refer to Research aspects as per 7.3.x)	Mid-term (finished in 5y)	Optimized OWC for reliable transmission over >50 km reach, with outage probability <1e-6, bit rate >100 Gb/s Optimized OWC for reliable feeder links (ground to satellite) transmission over 500 km reach, with outage probability <1e-3, bit rate 100 Gb/s	Specific channel model, validated by field tests Optimized PAT Optimized telescopes, and suitable optics Optimized protocols/network configuration tools
LiFi networking and portable hardware	Short-term (finished in 3y)	Demonstration of >1 Gb/s LiFi coverage with: -robustness to ambient light -low power consumption -suitable handover -compact transceivers	
Indoor navigation by OCC	Short-term	Demonstration of cm-accuracy in various environments: -robustness to ambient light -low power consumption -exploiting consumer-electronic hardware	Navigation of impaired people (elderly/ blind people in unknown buildings), effective rescue operations in emergency conditions
Underwater OWC	Medium term	develop transceivers for effective underwater communication 10-100 Mb/s, 10-100 m, with low footprint	monitoring and exploring sea (also deep sea)

7.12.5 Recommendations for Actions

Research Theme	Optical Wireless Communications		
Action	Component/subsystem development	Channel modelling for outdoor	UOWC
International Calls	X Having a shared program with non-EU based research and market leaders in devices, subsystems and telescopes (eg movable optical ground stations, adaptive optics)	X	
International Research	X Having shared program with non-EU based research and market leaders in FSO for HAPs and satellites	X Have a shared program with non-EU based companies and research bodies, for deep testing and validation	
Cross-domain research			X It requires to involve different expertise: photonic systems, experts, as well as underwater robotic experts

8 Non-Terrestrial Networks and Systems

Editor: Tomaso de Cola

8.1 Introduction

The evolutionary path from 5G to 6G through the current phase of 5G-Advanced has clearly highlighted the role that NTN will play in the near future, in order to complement terrestrial infrastructures to achieve extended connectivity in very diverse environments, eventually realizing the concept of "connecting the unconnected". In this respect, the vision developed by ITU IMT-20230 recognizing the importance of "ubiquitous connectivity" as one of the key usage scenarios further stresses the importance of an NTN native component within the multi-faceted 6G ecosystem, in terms of sustainable, secure, and flexible data communication architectures whose operations will also build on new distributed intelligence concepts. As a matter of fact, however, the vision of NTN is not to corroborate only the "ubiquitous connectivity" use cases, but to importantly contribute to massive communication and also to further explore potentials that NTN nodes may offer in the future with respect to integrated sensing and communications (ISAC), by possibly combining the technology advanced separately reached by NTN with respect to EO (Earth Observation) and communication missions.

Along these lines, this chapter develops a vision for NTN evolution to 6G and beyond by providing a overall view first at system level and then through its main components in the dedicated subsections.

8.2 The 6G NTN Vision

8.2.1 6G as an umbrella for NTN

5G has now rolled out in most developed countries in the World, and we continue to see improvements and developments in the standards bodies. From 2021 to 2025, NTN including satellites is working towards 5G integration with TNs and full commercial operation. Now is the time to consider what techniques and technologies might feed into standardisation for 6G. The roadmap is shown in Figure 8-1.

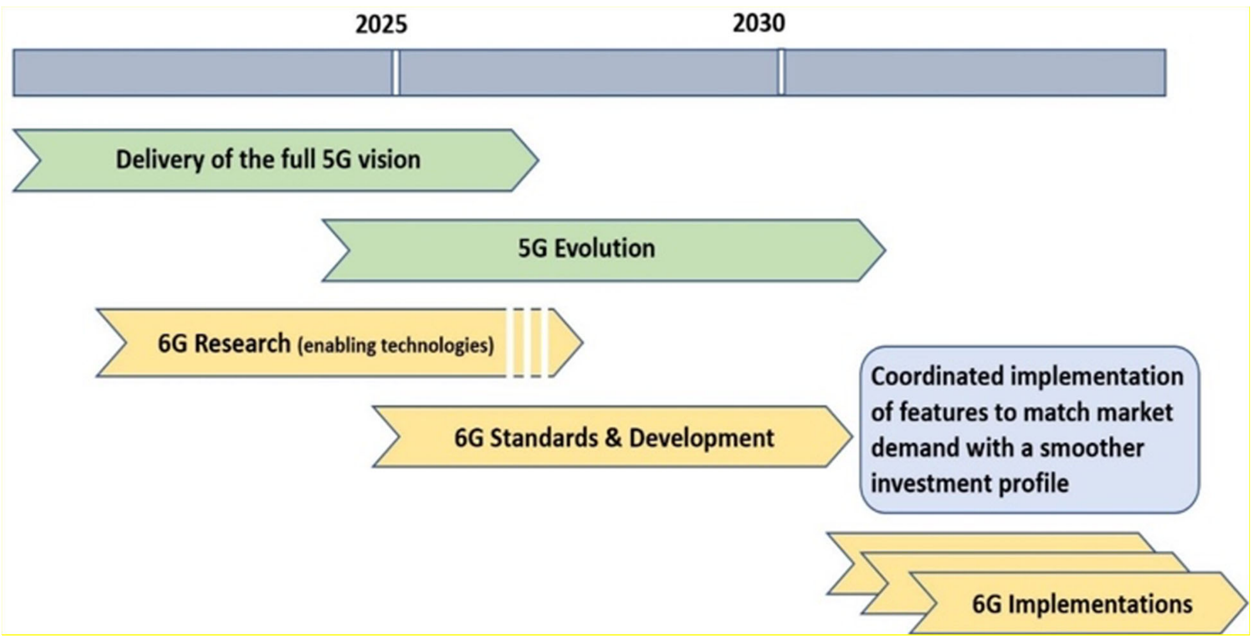


Figure 8-1 5G to 6G Roadmap timelines

6G will be a unified system and must be different from previous generations; it must be designed to meet Global Challenges and, at the same time, include the technology to support cost-effective coverage and radically new innovative services. The vision is that it should consist of converged digital and communications infrastructure in an ecosystem that serves humanity's needs but is also affordable to the users and economical to the providers. In addition, contributing to environmental sustainability and helping towards a net-zero agenda.

As distinct from the start of research for previous generations, in 6G, we have a radically different world order to address. In particular, the following are now crucial challenges:

- Changed working practices in a post Pandemic era
- Climate change and the reduced reliance on fossil fuels
- The need to radically reduce energy usage and improve security

Innovation in technology marches on at an increased pace, with major innovations blurring the boundaries between the physical and virtual worlds and enabling natural interactivities between them. Increased and massive softwarisation, Artificial Intelligence (AI) and Machine Learning (ML), and sensing of the environment are all key innovations. The task for 6G is to embody the new technology advances to address the global challenges in a system that will provide affordable and usable services by the population at costs that are sustainable by the operators.

Scanning the many 6G vision documents from around the world, we abstract the following common themes (note that these are often associated to different types of use cases, transporting different types of traffic, with potentially different requirements):

- New Human-centric services—AR/VR/MR—Teleportation
- KPIs that exceed those possible in 5G in latency and reliability
- Sensing at the user terminal merged with communications
- AI-based network and massive virtualization.
- A 3D space network including UAV's-HAPS-Satellites (NTN)
- New Frequency bands
- Increased security across integrated networks
- Increased network resiliency
- Massive networks of sensors- IoT
- Improved timing and positioning accuracy
- Achieving sustainability and energy reduction targets

The vision of 6G incorporates a range of human multi-sensory experiences enabled by digital solutions and hyperfine geolocation with context awareness provided by massive localized sensors. In addition to human and local information sensing, system-level sensing will be essential for efficient and intelligent system operation. This implies fine time and frequency synchronization to microseconds and guaranteed ultra-low latency (ULL) not provided by 5G. This will enable the provision of a tranche of new services for verticals across telecom networks. The vision is of a hybrid network of networks from short-range and ultra-high capacity to the widest coverage via a new space network dimension (Figure 8-3). NTN will become one of the key enablers in providing coverage, security, and resilience in this 6G vision.

Within NTN, satellites remain key, although High Altitude Platforms (HAPs) and Aerial devices (UAVs) in the troposphere will emerge as new components. Today's satellite communication systems have expanded from

GEOs and MEOs for regional coverage, to include massive constellations of LEOs for global coverage all offering very high throughput and in the LEO case low latency matching the demands of 5G and potentially 6G services. Whilst current satellite generations have considerable digitization on board they remain bent pipe, but the new generations from 2025 onwards will include advanced on-board processing and possibly some on board network functions, and the LEO's will also include optical inter-satellite links(oISL's).

The traditional business to business and broadband satellite delivery market models are now joined by a new market of Direct to Device(D2D), where the Device can be the handheld (HH), IoT sensor terminal or mobile vsat. D2HH has already been demonstrated direct to commercial smartphones for short messaging and 4G by AST Mobile, Lynk and Starlink. These operators are in the process of launching new constellations of satellites and operations will begin in 2024/5. They will either use terrestrial mobile spectrum or L/S band satellite spectrum. Inevitably these will be followed by native 5G NR and then NR-NTN operation as the ecosystem develops to use the new 3GPP NTN standards. For IoT services, constellations of LEO satellites are being launched by Sateliot and OQ for example to exploit the other NTN-NB-IoT standard. Meanwhile Eutelsat OneWeb are about to launch a second-generation constellation for B2B services and Starlink have launched satellites for their D2HH service. The Amazon Kuiper constellation for broadband and D2D is due to start launch in 2025/6 and Telesat's B2B service in the same timescale. Spectrum will continue to be a major consideration with NTN standards moving to Ku band; pressure on S/C bands and moves above Ka band to Q/V/W etc for fixed broadband. Not least will be the issues of frequency sharing between NTN and TN.

The role of satellites has traditionally been to provide coverage into regions not economical for terrestrial infrastructure and to provide resilient backup to terrestrial services. We see these features remaining as key drivers in 6G. Nonetheless, NTN also provides **flexibility, efficiency, service continuity, and fast and low-cost global coverage** (e.g., for IoT applications). As terrestrial networks pursue lower latency service offerings, satellite constellations at very low altitudes (vLEO) with Inter Satellite Links (ISLs) offer comparable and even lower latency for longer links. Thus, these systems are of interest for 5G and will be included in 6G. Due to restricted spectrum and satellite power, capacities have in the past been limited and hence more expensive than terrestrial. However, today using frequency reuse, dynamic resource allocation, and onboard processing, both GEO and LEO satellites have increased to circa 1 Terabit/s, and the costs of the space system have drastically reduced.

NTN will be an integral part of 6G but standards this time will be unified from the start in 6G rather than being bolted on, as with earlier generations. For satellites to play an integrated role in 5 and 6G, some commonality of standards is required. Until recently, satellites remained outside mainstream standards and had developed their own air interface standards — DVB-S2X (and its predecessors), which was initially based on video broadcast. More recently, and seeing the advantages of integration, satellites have joined the 3GPP standards groups responsible for 5G and now 6G standards. The 3GPP Rel. 16 (2017), on which the current rollout of 5G terrestrial networks is based, does not include satellites. However, starting from rel. 17, Non-Terrestrial Networks (NTN) are part of the 5G ecosystems and their development continues into Rel. 18 &19 towards the goal of integrated standards. There is thus a pathway to full unification as shown in Figure 8-2 with the period to 2025 used to getting 5G and satellite integrated and the period up until 2030 having satellite established as a unified part of 6G.

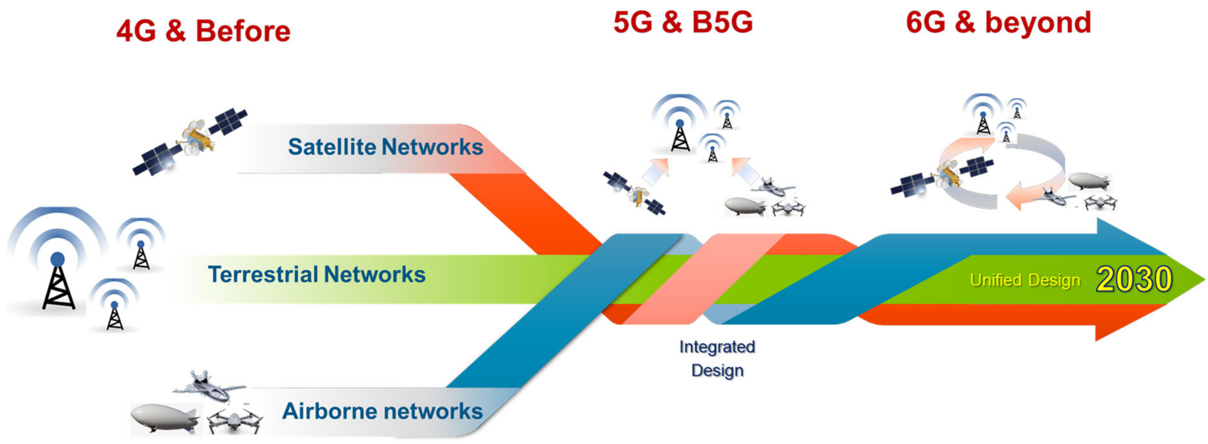


Figure 8-2 Pathway to full integration (source 6G-TakeOff Project – BMBF funded)

8.2.2 Satellites as key components in 6G

The key role of satellites in 6G will be in coverage and resilience, however, they are essential in enabling other critical services such as earth resources, positioning-navigation, and timing (PNT) and for continuous control and connectivity to aerial devices (UAVs) and maritime vehicles. The ultrahigh capacities needed for some 6G services will only be available on short-range terrestrial connections using Terahertz links in urban areas. A range of 6G services will be required by users traveling out of these areas and thus the pathway to 100% coverage can only be economically provided by satellite. The satellite will also be used to backhaul mobile cells in rural and remote areas or/and to provide backup resilience. Of course, connection to ships and aircraft will necessitate satellite backhaul.

In addition to the above it will be seen that the addition of the extra dimension of space to create a 3D network is implicit in the 6G vision and this is where satellites fit into a broader picture. As shown in Figure 8-3 this leads to the concept of a multi-layer network which adds satellites in GEO, MEO and LEO to lower altitude HAPS and even lower aerial devices such as drones; orbits such as HEO (highly-elliptical orbit) and VLEO (Very-Low Orbits), also mentioned above as vLEO) might also be considered. The network architecture connecting these components will be service dependent as some architectures will better suit the requirements of specific services. The network functions can also be distributed amongst the entities to optimise performance. In all cases we will have a highly integrated E2E cross-network system.

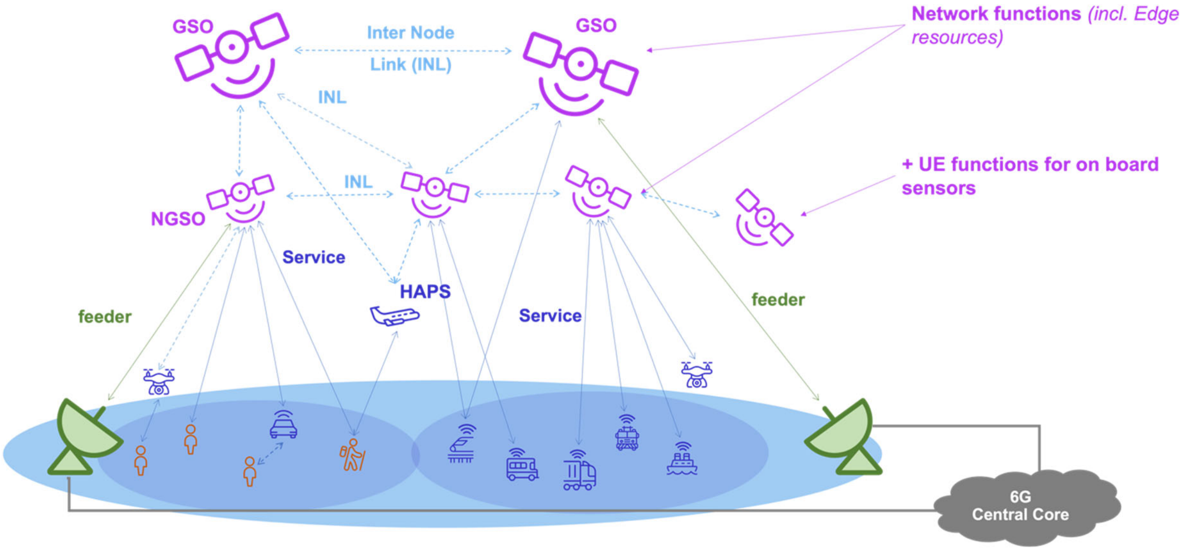


Figure 8-3: 3D space network (source: 6G-NTN project (EU HE SNS))

The 6G standardisation process commences in 2025 and already some emerging themes exist These include the increased use of open systems and the integration across various networks as shown in the 3D layered model. We see 6G as a fabric and an ecosystem bringing together a number of new technologies providing ultra-high resolution, communications at the edge, supported by massive intelligence in the core.

Meanwhile the 3GPP standards group and NTN are continuing their work on 5G+ looking at applications into vehicle to everything (V2X) transport and 5G based IoT. Also considerations regarding energy and spectrum efficiency, location sensing, carrier aggregation and advanced antenna arrays are under way. In other areas advances in software defined radio (SDR) and digital processing will contribute to more flexibility in radio access and in core networks. Security across the whole system will be critical and will be embedded in the design. This will require the use of intelligent firewalls, context-aware domain level protection, and advanced cryptography supported by cloud quantum computing. These and other innovations will feed into the base definitions for 6G.

For 6G we will need new and advanced techniques that enable deeper integration between satellite and terrestrial networks which are seamless from a user perspective, moving from satellite to terrestrial coverage. With the introduction of large LEO satellite constellations with high mobility this introduces new challenges ranging from intelligent and dynamic spectrum sharing to seamless handover and maintenance of QoS.

Key Challenges for satellites in 6G:

- Unified T/NTN architecture based on open networking
- Full network integration of all layers in the 3D SDN Network,
- Direct connectivity to smart phones, outdoor and light indoor and in vehicle,
- Ultra Low Latency support for vertical sectors
- Merging networking and computing -not just at the edge
- Integrated and flexible Air Interface for multi services.
- Ultra-accuracy of positioning and timing
- Integration of sensing and communications (ISAC)
- Embedding AI in network and RAN
- Providing Security across the network- elements
- A modified IP for space networks
- New spectrum and sharing across the network components

- Supporting massive IoT
- Supporting mobile networks (e.g. trains, aircraft, vessels)
- Solving the problem of massive antennas in space
- Contributing to overall sustainability and energy reduction targets.

8.3 Architecture and System-Level Aspects

8.3.1 Multilayer Architecture

It is globally recognized that a pivotal characteristic of 6G systems will be the definition of a 3D multi-layered architecture, with the third dimension expanding legacy flat terrestrial networks (TNs) in the sky thanks to a Non-Terrestrial Network (NTN) component. The multi-layered characteristic of this network stems from the integration of multiple layers in the NTN component, ranging from relatively low altitude air-borne nodes (*e.g.*, Unmanned Aerial Vehicles and High Altitude Platforms) to multi-orbit space-borne elements (from Very Low Earth Orbit to Geostationary and Highly Elliptical Orbit satellites). Such novel and challenging architecture concept will be a key enabler for new use cases and applications, as well as innovative concepts and technologies to achieve the 6G system requirements. Notably, Geosynchronous Orbit (GSO) solutions provide larger coverage for regional/global services with limited jitter, while Non-GSO (NGSO) platforms can provide a better link quality and reduced latency. The challenging objective of a multi-layered network calls for a paradigm shift in network design. In fact, with 5G and 5G-Advanced, a NT component has been integrated in a system optimized for TNs; with 6G, the TN and NTN components must be jointly optimized to fully exploit the potential of 3D networks, [C8-1]. Inter-node links (INLs), optical or at Radio Frequency, will be established both intra-orbit (or horizontal) and inter-orbit (or vertical) and will play a key role. In fact, their characteristics (*e.g.*, availability, latency, quality, as well as the capabilities of the involved nodes) will define a highly dynamic network topology, with potentially different Radio Access Technologies (RAT) and backhaul links capabilities. The realization of 3D multi-layered networks shall be supported by further technology advancements, including highly spectrum efficient and flexible waveforms for both the TN/NTN components, software defined payloads embarking RAN and maybe part of the core network functions, the enhancements of access radio protocols to optimize network mobility, Artificial Intelligence for network optimization and management, multi-antenna solutions, and advanced Radio Resource Management (RRM).

8.3.2 Mobility Management

Mobility management ensures service continuity by seamlessly switching the connection between satellites and across the satellite, aerial, and terrestrial segments. Handover procedures for multi-layer networks need to be defined in a way that ensures transferring the data and voice sessions without interrupting the service during any kind of handover. An enhanced version of the Core Access and Mobility Management Function (AMF) used in 5G should be developed to cover requirements from the 3D architecture. The handover procedures that need to be covered include the following: 1) A handover between satellite beams, called a spotbeam handover. The connection is switched between beams of the same satellite. 2) A handover between satellites, including the consecutive LEO satellites of the same constellation as well as between satellites in different orbital heights. 3) A handover between satellite and terrestrial cells especially when moving outside the terrestrial coverage to the satellite coverage and vice versa. 4) A handover to and from the aerial segment. 5) An inter-satellite link handover where the connection needs to be routed over different ISLs at different time instants. The optimization of mobility management procedures must include advanced beam-pointing approaches such as quasi-Earth fixed cells that can be used to minimize number of handovers between satellites – and most probably AI approaches will have role in this development.

8.3.3 Unified TN-/NTN-Positioning

As it is the goal for 6G to come up with seamless and native integration of TN and NTN in the form of a unified 3D-network, the same applies to the positioning functionality of mobile networks: the positioning solution should exploit both the potential of terrestrial deployments as well as the (new) non-terrestrial network components and the use of both should be enabled in a unified and natively integrated fashion ("6G Unified Positioning Navigation Timing").

The definition of 5G positioning (see 3GPP TS 38.305, "Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN," [C8-9]) relied on terrestrial cellular measurements, *e.g.*, timing- or angle-based positioning measurements between base station locations and UEs. In addition, there was the support of external sensors, for example the use of GNSS or inertial measurement units (IMUs) or barometers. The different sensor information could be acquired via a standardized location services (LCS) framework that offered all the needed procedures, protocols, and functional units like the Location Management Function (LMF), where heterogeneous sensor readings may be combined towards a final positioning result.

For 6G, the task is now to natively enable positioning measurements via NTN nodes. The goal is to support the full range of applicable parameter sets like orbits (*e.g.*, from GSO, HEO to LEO and HAPS), frequency ranges, signal bandwidths and all relevant architecture options, *e.g.*, transparent payloads as well as regenerative payloads including associated gateways or interconnections (*e.g.*, inter-satellite links).

In the community, there is high motivation for a unified TN-/NTN-positioning solution for 6G. Coverage for NTN-positioning can go beyond terrestrial coverage as it is also true for data communications. NTN spectrum can be employed for positioning like for communications in an integrated fashion and support excellent results. For example, wider bandwidth will, as it is well known, provide higher resolution in the time domain. NTN-positioning can also be seen a new "positioning sensor" with new complementary properties compared to terrestrial cellular measurements, *e.g.*, through higher arriving signal elevation angle in certain scenarios and the resulting different susceptibility to multipath. Also, NTN-positioning as an alternative reduces the dependability of GNSS, which is known for its vulnerabilities (jamming, spoofing, user manipulation) and additionally means cost and energy consumption of an additional non-cellular radio module. Orbits at lower altitude might offer advantages by providing a better link budget compared to GNSS.

The challenges during solution design include the support of RAN components moving fast on orbits with long distances at high speeds, while in TN fixed locations close by can be assumed. For some scenarios, there is also the task to provide a meaningful positioning result in case just one or very few space vehicles are in view. In case no prior approximate UE location information is available during startup (for example applicable to UEs without GNSS module), a suitable initial access strategy and solution is needed, which is a design task to be carried out for enabling basic NTN access as well as providing NTN-positioning.

Unified TN-/NTN-positioning is applicable to all narrowband (IoT) and broadband 6G use cases including worldwide asset tracking with cheap and energy efficient tags as well as supporting advanced autonomous driving which may resort to more complex multi-sensor-UEs. Note that 6G use cases can generally be assumed to ask for throughput/low latency/availability including location context as native output of the system.

A possible investigation may deal with NTN-positioning system engineering, system modelling and simulation, later emulation and proof-of-concept, signals and measurements, or unified architecture and protocol design. It may also investigate for certain scenarios or use cases the anticipated complementary gain of NTN-positioning as new sensor within a hybridization scheme being used together with other sensors (*e.g.*, terrestrial Dl-TDOA, UL-TDOA, Multi-RTT, UL-AOA, DL-AOA) for a wide range of different parameter sets (constellations, frequency ranges, bandwidths, antenna and spot beam characteristics, etc.).

8.3.4 Energy-efficient Functional-Split Options and NTN Architecture

The new LEO satellite mega constellations promise ubiquitous high data-rate availability all over the globe thanks to the massive amount of satellites and their user links. However, all these data need to be transmitted back to Earth via satellite feeder link. As comparative analyses in [C8-10]-[C8-11] have shown, these feeder links to gateways (GWs) can easily be the system bottleneck. Using ISLs can help solving this issue in certain scenarios but it introduces additional requirements as well. This requires satellite on-board processing, which is in line with the 3GPP roadmap of moving from allowing only transparent satellite connection to regenerative on-board processing with some or all gNB functions on-board the satellite.

A transparent link via satellite represents of course the simplest case of implementing NTN, where all gNB functions are located in the GW. However, due to various shortcomings, this may serve rather as a starting point for extensions in the next releases. Using a regenerative satellite payload, i.e., an onboard processor (OBP), decouples feeder-link from user-link processing and allows independent error correction. Furthermore, an OBP is the baseline for running ISLs and gNB functions in the sky, which helps to reduce delays in the 3GPP protocol procedures. Just putting the whole gNB processing on each satellite sounds simple and most straight forward but it is also the most energy consuming and expensive approach.

The question which gNB functions shall be processed in the sky leads to a so-called functional split discussion, splitting the RAN functionalities into a centralized (CU), connected to the core network, and a distributed unit (DU), providing the user link. The goal is to reduce the processing requirements at the DU located close to the user-link antenna. Functional split over NTN is one of the key-enabling technologies in the upcoming releases of 6G networks and the recent open RAN (ORAN) extensions. Different functional split options, their implications and interfaces are discussed in [C8-12] with respect to DU processing on-board the satellite and CU processing in the GW. NTN key performance indicators (KPIs) such as delays, interface data rates, or support of multi-connectivity (MC) techniques are discussed, but no unique best choice was found.

As a remedy, a new architecture proposed in [C8-12] tries to balance all these challenges and requirements by putting gNB-DUs onboard multiple satellites connected to a gNB-CU onboard of another satellite. Here, a swarm of satellites hosting the gNB-DUs and the gNB-CU move together as one entity. This implementation allows providing a ubiquitous coverage and supporting of services and applications where the NTN gateways are not available or temporarily not operational. So only the gNB-CU satellites need to be more powerful in terms of processing power. And consequently, the signalling overhead over the feeder-link is significantly reduced.

Such an architecture has a significant impact on the payload of the satellite. As different functions being on different satellites, there might also be different payload processing concepts as they operate on different layers of the signals. OBP on distributed units will be more dedicated to physical layer processing whereas in the CUs higher layer processing is needed with different requirements. To cope with this, a specialized processing is needed and needs to be defined. In general, and most notably also in space, resource and energy-consumption optimised implementations are needed to reduce consumption and power dissipation.

Having identified this beneficial architecture, significantly more investigations are needed to determine the best functional split option and impact on all 3GPP techniques and protocol mechanisms. Furthermore, new 6G technologies may even require to modify/extend these interfaces. And transferring this approach to other scenarios like HAPS swarms offer further room for investigations.

8.3.5 AI enabling architecture

Traditional resource allocation for non-terrestrial networks has been designed on a static basis, regardless of the usage patterns. This approach leads to inefficient resource utilization and high Operational Expenditures (OPEX). Nevertheless, the complexity of non-terrestrial networks is increasing with multi-orbit structures, including multiple layers in the architecture. There is an increasing need to integrate high-throughput space-born network elements (i.e., LEOs and GEOs), with a larger number of beams and flexible payloads, and/or HAPs with terrestrial communication services to achieve ubiquitous global connectivity. Within this context, traditional monolithic allocation of resources is no longer an acceptable option and requires innovative approaches to network design and management.

To ensure adequate functionality and flexibility of future 6G networks, comprehensive and autonomous network monitoring, and management, together with AI-driven optimization [C8-7]-[C8-8], are necessary. This fact motivates the need for an AI-native architecture, primarily based on the Open RAN (Radio Access Network) initiative [C8-5]-[C8-6]. Open RAN's emphasis on disaggregation and virtualization of network functions aligns perfectly with the needs of NTNs. By separating the hardware and software components of the network, NTNs can achieve greater flexibility and scalability. Virtualization allows network functions to be deployed where they are most needed, whether on the ground or in space, enhancing the network's ability to manage resources efficiently and respond quickly to changes in demand. The Open RAN framework, with its open standard interfaces, is designed to foster interoperability among diverse network elements. For NTNs, which must integrate with various terrestrial and space-based components, this interoperability is crucial. The standardized interfaces provided by O-RAN enable seamless communication and coordination across different network layers and elements, reducing the complexity of integration and ensuring a more cohesive network operation.

The orchestration of NTNs by data-driven AI facilitates smarter decision-making processes. With access to real-time data and advanced analytics, an AI-enhanced architecture, can optimize network operations, predict potential issues before they occur, and automate routine tasks. This data-driven approach ensures that NTNs can operate more efficiently, with higher reliability and lower latency, meeting the stringent requirements of modern communication systems. An AI-enabled architecture based on O-RAN addresses these challenges by:

- Reducing Latency: By deploying near-real-time controllers in space, AI can reduce the round-trip time for data processing and decision-making, thus minimizing latency.
- Optimizing Resource Use: AI algorithms can predict demand patterns and allocate resources dynamically, ensuring efficient use of bandwidth and power.
- Enhancing Scalability: The scalable nature of AI and virtualization allows NTNs to support a wide range of missions and services without compromising performance.

One of the key aspects of Open RAN architecture is that it natively embeds intelligence into the RAN and how Machine Learning algorithms can be deployed in the architecture. The data is gathered from the corresponding entities by the corresponding interfaces (*e.g.*, O1, A1 or E2) and managed resources are optimized via two different entities: the Non-real-time Radio Intelligent Controller (RIC) and the near-real-time RIC (involving operations like beam assignment, resource optimization or power control), both operating in different control loops and operating at different time scales (>1 s for Non-real-time RIC and between 10 ms and 1 s for the real-time RIC entity) . More advanced and distributed techniques can be naturally allocated in evolved AI-native architectures, enabling on-ground or on-board placements depending on the latency requirements.

The adoption of an AI-enabled architecture, based on Open RAN, provides a future-proof architecture that can evolve with the technological advancements. It allows for continuous improvement and dynamic capabilities through AI/ML and the ability to incorporate new functionalities and enhancements over time. However, this ambitious integration presents significant challenges. These include the necessity for advanced data processing and control capabilities to manage larger volumes of monitoring information with minimal latency and high energy efficiency.

8.3.6 Summary

Research Theme	**3D Architecture Management**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Definition of a native NTN component: 4.3.1	Short-term (in line with 3GPP Rel. 21)	6G protocols and procedures for jointly optimised TN-NTN 3D networks	Definition of a fully native NTN component in the 6G standard
Challenge 2 High-capacity Inter-Node Links (intra-/inter-orbit): 4.3.1	Medium-term	Definition of protocols, technologies (*e.g.*, waveform, antenna), and spectrum for high-capacity RF/optical INLs	Support of dynamic functional split solutions and resource allocation in the 4D multi-layered architecture
Challenge 3 Routing in 3D multi-layered architectures: 4.3.1	Medium-term	Definition of energy efficient and fast routing algorithms enabling routing with a rapidly varying network topology	Flexible and efficient routing algorithms in the 3D dynamically varying network topology
Challenge 4 Handover procedures for multi-layer networks: 4.3.4	Short-term	Definition of the protocols and procedures for TN/NTN (aerial and space) handover solutions	Support of a fully native NTN component in the 6G standard
Challenge 5 Distributed architectures for energy-efficient functional split: 4.3.7	Medium-term	Design of energy-efficient and reconfigurable functional split in the sky Directions, options and performance criteria for next generation NTN systems and satellite payloads	Flexible and optimised CU/DU allocations over NTN yielding enhanced and energy-efficient performance in order to maximize probability of commercial success and feasibility
Challenge 6 AI for autonomous network monitoring and management	Medium-term	Design and implementation of AI tools, eventually supported by RIC, to autonomously orchestrate the network functions	Optimised autonomous network monitoring and management

Research Theme	**Enhanced Positioning**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Definition of network-based positioning solutions: 4.3.5/6	Short Term	Algorithms and technologies for location services completely network-oriented	Support of location services for GNSS-free user terminals in 6G
Challenge 2 Multi-information positioning technology: 4.3.6	Medium-term	Design of the algorithms exploiting multiple parameters for positioning (NTN nodes at different orbits, frequency ranges, signal bandwidths, and all relevant architecture options) allowing native and tight integration between coms and PNT, but also TN and NTN	Efficient and accurate NTN positioning schemes showing significant KPIs improvements such as coverage (availability), accuracy, latency, robustness/ reduced vulnerability, reduced dependency on 3rd party systems
Challenge 3	Medium-term	Design of positioning algorithms capable of providing accurate	Efficient and accurate NTN positioning schemes

Positioning technology in rapidly changing RAN		location estimates with rapidly changing RAN topologies (speed and type of NTN nodes, compared to relatively fixed and close TN nodes)	

8.3.7 Recommendations for Actions

Research Theme	Architecture Management – part 1		
Action	NTN native component	High-capacity INLs	Routing in 3D networks
International Calls	X The integration of a native NTN component is the focus of 6G NTN, with some projects already addressing it. Still, more calls are needed so as to design/assess the required technologies, develop demonstrators/PoCs, and identify the expected gains and costs	X To support 3D multi-layered networks and adaptive functional split, the technologies of high-capacity INLs must be addressed in international calls seeking performance improvement, interoperability, and open interfaces. This is present in a limited number of projects and more endeavours are thus required.	X The design of routing algorithms and the related technologies to support them are still in their infancy. Researching them in the context of international projects would be beneficial to gather together research institutions and industrial interests.
International Research	X The design and/or adaptation of TN and NTN technologies to support a native NTN component in 6G is fundamental to define the 3D multi-layer network capabilities and performance.	X International research efforts are beneficial for the design of high-capacity INLs, both in terms of the NTN architecture supporting them and the technologies to enable such links.	X The design of routing algorithms and the related technologies to support them are still in their infancy. As such, international research activities are required to develop them.
Open Data			X Considering the highly dynamic environment in which 3D routing schemes will operate, the availability of large open datasets to implement AI/ML solutions would be beneficial.
Large Trials	X In line with the international calls, large trials assessing the feasibility of the identified technologies for a native NTN component are needed, in particular related to TN-NTN interactions	X Large trials to assess the performance of INLs in deployed NTN components is fundamental to assess their feasibility to support both adaptive and advanced functional split solutions and broadband communications.	X Once the 3D routing algorithms have been designed and developed in international calls and research activities, large trials would allow to assess their actual performance.
Cross-domain research	X A native NTN component in 6G networks will be an enabler for many diverse applications and services. Cross-domain research might		X 3D routing is a key enabler for a native NTN component in 6G. As such, it might be beneficial for various applications and services,

	be beneficial to assess the gains of this part of the infrastructure		calling for cross-domain research with different performance objectives.

Research Theme	**Architecture Management – part 2**			
Action	Handover in multi-layered networks		Distributed architectures for energy-efficient functional split	AI for autonomous network monitoring and management
International Calls	X To support 3D multi-layered networks and adaptive functional split, the handover techniques are the key enablers and therefore must be addressed in international calls seeking for robust and secure schemes, interoperability, and open interfaces.	X To support 3D multi-layered networks, distributed and adaptive functional split options must be addressed in international calls seeking for best energy efficiency, trade-offs against needed effort, interoperability, and open interfaces.	X The development of AI-based network management solutions, in particular considering a need for decentralised and self-organising solutions that can increase resilience in the verge of attacks, or natural hazards.	
International Research	X International research efforts are beneficial for the design and assessment of handover techniques for the multitude of different links to allow for a dynamic NTN architecture.	X International research efforts are beneficial for the design and assessment of distributed and adaptive functional split options with focus on the energy-efficiency, practical viability and supporting protocol enhancements.	X International efforts required towards decentralisation of the network management, and analysis of management (control plane) that can quickly adapt to context, and yet be energy-efficient	
Open Data				X Considering the highly dynamic environment in which the native NTN component will operate in 6G, the availability of large open datasets to implement AI/ML solutions is fundamental.
Large Trials	X Handover is a system critical functionality, which is why all aspects must be proven to work properly by trails.		X Functional split is a system critical functionality for proper gNB operation, which is why all aspects must be proven to work properly by trails.	X Cognitive networking involving TN-NTN management requires tools which today are mostly available in the form of simulators.

Cross-domain research			

Research Theme	Enhanced positioning		
Action	Network-based positioning	Multi-information positioning technology	Positioning technology in rapidly changing RAN
International Calls	X Bringing together relevant, strong and innovative players in their fields (6G-TN, positioning, NTN) and role (industry, institutes)	X Bringing together relevant, strong and innovative players in their fields (6G-TN, positioning, NTN) and role (industry, institutes)	X Bringing together relevant, strong and innovative players in their fields (6G-TN, positioning, NTN) and role (industry, institutes)
International Research			
Open Data			
Large Trials		X Involving NTN infrastructure providers at early stage and gaining practical results early	X Involving NTN infrastructure providers at early stage and gaining practical results early
Cross-domain research	X Bringing together technology providers and domains involved in emerging use-cases to develop business models and to solve ethical key questions (e.g. use of data and privacy) right from the start and in parallel (avoiding a gap which was and still is in 5G)	X Bringing together technology providers and domains involved in emerging use-cases to develop business models and to solve ethical key questions (e.g. use of data and privacy) right from the start and in parallel (avoiding a gap which was and still is in 5G)	X Bringing together technology providers and domains involved in emerging use-cases to develop business models and to solve ethical key questions (e.g. use of data and privacy) right from the start and in parallel (avoiding a gap which was and still is in 5G)

8.3.7.1 Key Performance Indicators (KPI)

This Section presented the expected directions in research and innovation for NTN architectures in the 6G ecosystem. As such, the KPIs listed below shall be considered as design principles to be taken into account in developing the system architecture and not all of them might require to be met depending on the specific mission. The requirements for 5G and 5G-Advanced NTN are based on 3GPP specifications, ITU-R IMT-2020 requirements, and activities performed in the framework of ESA and EC funded projects, [C8-15], [C8-16], [C8-17].

Performance requirement	5G/5G-Advanced	6G expectation
Peak Data Rate (DL/UL) Handheld terminals	1/0.1 Mbps outdoor up to 3 km/h	Tens of Mbps up to 250 km/h outdoor
Peak Data Rate (DL/UL) Vehicle/drone-mounted terminals	50/25 Mbps up to 250 km/h 60 cm aperture	Hundreds of Mbps up to 250 km/h outdoor
Peak Data Rate (DL/UL) Large aeronautic or maritime platforms	360/180 Mbps up to 1000 km/h	Thousands of Mbps 1200 km/h outdoor
Positioning accuracy and latency	1 meter, below 100 s	1 meter, below 1 s
Coverage	Outdoor	Outdoor light indoor/in-car for short messages
Reliability	Up to 99.9%	Up to 99.999%
Over-the-air latency	CP: 40 ms UP: 10 ms (eMBB-s and uRLLC-s)	CP/UP in line with IMT-2030 TNs excluding propagation delays (Immersive communications and Hyper Reliable and Low-Latency Communications)
Connection density	Up to 500 per km^2	>1000 per km^2

8.3.7.2 Key Value Indicators (KVI)

In terms of KVIs, similar considerations as those introduced above for the KPIs hold, i.e., the following indicators shall be considered as design objectives that might be partially or fully required depending on the specific NTN mission.

The KVIs for the NTN architecture can be developed along three main domains: economic, societal, and environmental sustainability.

Domain	KVI	Comments
Economic sustainability	Global and affordable network coverage (SDG 8 and 10)	The intrinsic cooperation between TN and NTN will significantly help in bridging the Digital Divide in terms of coverage, affordability, and achievable data rates
	Resilient network coverage and services (SDG 9 and 11)	6G NTN can guarantee the continuity of economic activities in Digital Divide areas or scenarios in which abrupt TN interruptions arise
	Energy consumption (SDG 7 and 9)	Energy consumption and optimization shall be taken into account in the design of all NTN network components. The integration of TN and NTN will help in the transition to renewable energy and reduce the current contribution of the ICT sector to the global carbon footprint
	Circular economy (SDG 9)	The NTN component can both support the lifetime monitoring of manufactured products and integrate circularity in its own production processes
	Eco-design (SDG 9)	The NTN component can support the real-time monitoring of environmental changes to optimize the production processes at industrial level
Societal sustainability	Inclusion (SDG 1, 2, 3, 4, 5, 11, 16)	The NTN component can support the bridging of the Digital Divide in terms of coverage, affordability, and efficiency, thus tackling challenges such as the reduction of inequalities in terms of wealth, hunger, health, education, territories, peace, access to justice and institutions

	Health (SDG 3)	The NTN component will provide a global, seamless, and affordable network coverage, supporting advanced health services
	Trustworthiness (SDG 8 and 9)	All the UN SDGs require support of adequate levels of security, in terms of data confidentiality, integrity, and privacy. The NTN component shall comply with such requirements
Environmental sustainability	Clean water and sanitation (SDG 6)	6G NTN systems can support environmental sustainability by means of global/regional space-based Earth observation missions
	Climate action (SDG 13)	
	Life below water (SDG 14)	
	Life on land (SDG 15)	

8.4 Air Interface

8.4.1 Waveform Design

The physical layer of 5G NR is based on Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM) and Discrete Fourier Transform-spread OFDM (DFT-s-OFDM) as waveforms. Both the CP-OFDM and DFT-s-OFDM have been adopted also in NTN. Indeed, the OFDM-based transmission brings several advantages, such as resilience to time-invariant frequency-selective channels, high spectral efficiency, and flexibility in terms of slot duration and sub-carrier spacing due to the support for different numerologies. However, as it is well known, the CP-OFDM suffers from High Peak to Average Power Ratio (PAPR), which diminishes the efficacy of the high-power amplifier in the satellite, leading to decreased spectral efficiency. Furthermore, high Out-Of-Band (OOB) emissions may pose challenges for coexistence and spectrum sharing of terrestrial networks with non-terrestrial networks and other systems. Therefore, appropriate filtering and spectrum shaping design is required. Another factor impacting the efficiency of the actual NTN waveforms is the high speed of the satellite, which leads to severe issues in UL and DL synchronization. Moreover, looking at the characteristics of the propagation environment and those of the satellite constellation, it is worth mentioning that satellite communications predominantly occur in LoS conditions and, when dense satellite constellations are adopted, more than one satellite is in the UE's visibility. This motivates choosing a reduced cyclic prefix length and exploiting multi-satellite diversity to increase the system throughput. In this framework the research is exploring the use of modulations in the delay-Doppler (DD) dimension for NTN, such as Orthogonal Time Frequency Space (OTFS) and Orthogonal Delay Doppler Multiplexing (ODDM).

Remarkably, DD modulations can be implemented using the existing OFDM modulator and demodulator. Adopting this model allows for some commonality with 5G NR, as it leverages system parameters inherited from the legacy CP-OFDM waveform. This can facilitate a smoother transition when operating in the time-frequency (TF) domain. However, it's important to note that operations in the DD domain will necessitate new algorithms and procedures. One example is channel estimation, particularly when the channel varies on a time-scale shorter than the symbol duration. This scenario is common in NGEO satellite communication systems, where significant Doppler effects are inherent. Ideally, these effects can be mitigated using Global Navigation Satellite System (GNSS) information and satellite ephemeris data. However, in practice, compensation may be imperfect due to inaccuracies in satellite and terminal positioning or synchronization errors. Consequently, the residual Doppler effects result in a time-varying channel. In this context, operating in the DD domain is advantageous because the channel representation becomes sparser and varies on a much larger time scale than in the TF plane. Although the channel in the DD domain can be characterized with fewer parameters, the pilot sequences and grid positions designed for CP-OFDM cannot be reused, necessitating the development of new patterns. Moreover, there are several open research areas that require further investigation, especially for NTN. These include the design of synchronization algorithms, random access protocols, detection schemes, and reference symbols, to mention a few.

A viable approach is to consider an End-to-End data-based physical layer design, i.e., AI native air-interface. Indeed, a learning interface could dynamically adapt and configure customized waveforms, constellations, and pilot signals to utilize the available spectrum more efficiently, thereby enhancing overall performance. This will allow new schemes to optimally adapt to the peculiarities of NTN to enable new features of the air interface, such as light indoor communications.

8.4.2 Multi-antenna solutions

Satellite networks are experiencing a change of paradigm with regard to space antennas, from fixed multibeam architectures (i.e. single or multiple-feed per beam configurations) to large active arrays. These arrays provide new levels of flexibility allowing to dynamically adapt the coverage to the demand needs. However, how to efficiently harness its full potential remains an open topic. Promising solutions such as the combination of user-centric beamforming and advanced scheduling require a proper estimation of the channel state information (CSI) and in the best cases, the inversion of large channel matrices, which increase the computational complexity. Indeed, these have been two of the main barriers for the adoption of precoding in multibeam satellite networks. Moreover, this issue is aggravated by the clear trend of increasing space antenna aperture sizes to increase antenna gains and angular resolutions. This has become fundamental for direct-to-handheld communications, where the gain requirements and low bandwidth operation yields to required antenna aperture of tenths of square meters. An alternative for deploying large apertures is to leverage on the availability of multiple NTN nodes in the envisioned 3D multi-layered architecture, leading to the possible implementation of distributed multi-antenna technologies to improve the link capacity and reliability. Notably, these solutions can be broadly classified as Coherent Joint Transmission (C-JT), in which the transmission is designed so as to let the multiple signals coherently add at the receiving antenna, and Non-Coherent JT (NC-JT), in which multiple connections are generated thus not requiring coherent summation at the user terminal. C-JT solutions, such as federated user-centric beamforming techniques in which multiple NTN nodes define a flying virtual antenna array, are a promising technology gaining an increasing interest, [C8-2], [C8-3]; the major technology challenge for their actual implementation, in general applicable to C-JT, is the need for close-to-ideal INLs in the cooperating set of NTN nodes. In fact, slight misalignments in the time or frequency domains can have a detrimental impact on the system performance, hindering the benefit of such solutions. This is one of the main motivations for the interest in non-coherent approaches, for which the assumption of ideal INLs (or ideal backhaul, in 3GPP terminology) can be relaxed. Among these, we can mention Multi-Connectivity (MC), in which multiple simultaneous PDU sessions are transmitted to the same UE over multiple RAN nodes, [C8-4]; the gNBs involved in the MC scheme are coordinated via the Xn Air Interface and such cooperation can be defined with heterogeneous RANs, *i.e.*, including a TN and an NTN gNB. Notably, in this case the challenge is mainly related to the significantly different characteristics of the user service link involved in the MC scheme. Another solution belonging to the NC-JT family is multi transmission/reception point (multi-TRP), in which multiple TRPs are exploited to serve the same UE, yielding a power gain. To implement this solution, two aspects shall be taken into account: i) whether or not the power of the second TRP can be exploited to boost the performance of the selected user; ii) whether or not the allocation of this extra transmission power can cause harmful interference to other transmissions; and iii) whether or not the transmitted information by the two TRPs is the same. As a consequence, it might be expected that multi-TRP can be beneficial in low load conditions. In the 3GPP ecosystem, multi-TRP can be implemented with single Downlink Channel Indicator (DCI) or multi-DCI approaches; the difference is that, in the latter, the two transmitted information layers are managed over two different PDSCH with different DCIs.

The benefits of large multi-antenna solutions in terms of increased angular resolution translate to increased challenges when data needs to be transmitted/received in a broadcast fashion, for instance for SSB transmission allowing initial cell acquisition in 5G NR. The higher the angular resolution, the higher the number of resources required to cover the large field of view of the satellites. Therefore, efficient beam widening solutions trading off the EIRP loss and the resources required to cover the whole coverage area need to be studied.

All these multi-antenna solutions need antenna hardware which supports digital beamforming. Considering satellite or other flying platform implementations, this will also have an impact on the antenna radiator, its subsequent active devices and the connection to the processing hardware. Architectures are needed which optimally distribute functionality in the different building blocks with the most possible energy efficiency to support as many connections as possible.

8.4.3 Integrated Communications, Sensing, and Positioning

Integrated Sensing and Communication (ISAC) is a key usage scenario for IMT-2030/6G, as recognized by the ITU-R. ISAC is envisioned to play a vital role in the upcoming wireless generation standards. It transcends conventional communication networks and infrastructure to shared hardware and spectral resources, antenna capabilities and efficient signal processing co-design. ISAC for 6G NTN is set to revolutionize various sectors by providing efficient frequency sharing and radio resource management, enhancing adaptive beamforming and beam management, maximizing payload duty cycle use, and contributing to space sustainability.

Moreover, satellite-based Positioning, Navigation, and Timing (PNT) is indeed a unique feature present in NTN that can provide a new dimension to 6G ISAC. The integration of PNT into 6G NTN can overcome the limitations of the GNSS, providing high-precision, robust navigation capabilities, and enabling high-quality communication services, even in GNSS-denied scenarios [C8-18]. Therefore, the combination of NTN-based PNT and ISAC in a Joint Communications, Sensing and Positioning (JCSAP) system is a significant step towards realizing the vision of 6G as a truly global, ubiquitous, and reliable communication system. It not only enhances the capabilities of 6G NTN networks, but also opens up new opportunities for a wide range of applications, especially when combined with Multi-Functional Satellite Systems (MFSS).

The motivation for JCSAP in MFSS is driven by the rapid advancements in satellite technology. Software-defined satellites, which can be completely reprogrammed while in space, offer a new roadmap for services. They enable operators to reconfigure beams as needed, supporting future mobility applications through extremely high speeds, enhanced capacity flexibility, redundancy, and backwards compatibility. Moreover, flexible hybrid beamformers, which significantly reduce the hardware cost and power consumption by employing a small number of RF chains, are poised to transform the fleet, introducing an array of advanced functionalities and capabilities. These technologies collectively motivate the development of JCSAP in 6G NTN, enabling the fine-tuning and steering of resources per user and per service via software-defined on-board processors, digital beamforming, and hopping antennas.

8.4.4 Next Generation Multiple Access and Resource Management

The physical layer impairments related to the use of NTN networks will impact many 5G radio procedures and algorithms. The delays introduced by NTN channel will first require the definition of new Random-Access procedures and timing advance computation algorithms. As delays and UE transmission power profiles are significantly different for 5G terrestrial ones, future procedures will need to face this high variability to keep UE synchronization with satellite infrastructure. Several approaches have already envisioned the use of GNSS to grant synchronization between UEs and infrastructure. If the use of GPS can be envisioned as solution, other solutions will have to be defined in a "GPS less" context. Besides, 6G NTN is envisioned to support different verticals with conflicting QoS constraints (mMTC and emMBB for instance). The coexistence of these different slices at satellite level is particularly challenging due mainly to channel higher delays, higher degree of mobility and strong energy consumption constraints in satellite.

To serve a multi-tenants and high dynamic system, different multiple access techniques in the different domains (time, frequency, spaces, codes, modulations) shall be unified in a unique Multiple Access (MA) technique [C8-20]. Indeed, a single, unified, and general MA scheme would be easier to implement and optimize compared to a combination of multiple MA schemes, each tailored for specific conditions. This simplification is becoming increasingly crucial in multi-functional 6G and beyond networks, where the variety and complexity of services, use cases, and deployments are rapidly expanding. Rate-Splitting Multiple Access (RSMA) offers a promising example of a unified theory of MA schemes, even if, it does not encompass all MA schemes and therefore does not fully exploit all dimensions, such as time, frequency, power, space. How to achieve this ambitious goal is still under investigation.

Resource management in NTNs introduces unique challenges due to the distinctive nature of these communication systems. One of the primary challenges is dealing with dynamic topology and mobility. Satellites operate in various orbits and are in continuous motion, significantly impacting connectivity patterns. This means that resource allocation strategies must be flexible as ensuring reliable connectivity in such a dynamic environment requires resource allocation strategies that can quickly adapt to these changes. Interference and spectrum sharing present another significant set of challenges. NTNs must coexist with existing terrestrial networks, which means they must manage interference and ensure fair spectrum access. Effective interference management is crucial to prevent disruptions and maintain the quality of service. Energy constraints are also a critical concern. Satellites have limited energy resources, so it’s vital to optimize resource allocation while minimizing energy consumption. Balancing communication needs with the available energy supply requires careful planning and innovative solutions. Latency and delay requirements add further complexity to the integration of NTN with coexisting terrestrial networks. Balancing these latency requirements with resource efficiency is challenging. To overcome these challenges, researchers are exploring several promising directions. Machine learning (ML) and artificial intelligence (AI) offer potential solutions through dynamic resource allocation. For example, deep learning (DL) models can anticipate user demands and traffic patterns, while deep reinforcement learning (DRL) techniques assist in making intelligent decisions for resource distribution. Supervised learning (SL) approaches can map the relationship between power levels and performance metrics, and reinforcement learning (RL) methods can develop optimal power control strategies for dynamic network settings [C8-19]. Cross-layer optimization is another promising area of research, involving the coordination of different network layers to ensure efficient spectrum utilization. For instance, at the physical layer, modulation schemes, coding rates, and beamforming can be adjusted to maximize efficiency.

AI algorithms can predict network conditions and adaptively plan and schedule resources like power assignment, frequency and bandwidth settings over time to the different end-users based on real-time data, while also dealing with heterogeneity on the user traffic requests and requirements and aiding in interference prediction by learning interference patterns and optimizing frequency reuse, thereby improving overall system efficiency. The automation of NTN resource management becomes even more relevant when this is used for detection and recovery from failure(s). Pure model-based approaches have shown to be too computationally complex for reactive reconfigurations. This is because networks are becoming denser, with multiple elements with overlapping coverage areas, competing for the same resources. Monitoring the network status and network environment can bring insightful information to quickly identify or even predict network outage and support the preventive and reactive resource reconfiguration. The data-aided design seems to be the key for efficient network management, where the experience is used as side-information to guide future decisions. How this data is recorded, processed, stored and specifically utilized and communicated to the network controller remains a challenge for future wireless generation networks.

The design of radio resource management strategies usually needs to balance competing objectives, ranging from spectral efficiency, resource utilization, maximizing throughout and/or minimizing power expenses, demand matching, cost and fairness. It becomes challenging to achieve an optimal solution for multi-objective problems, as it is rather challenging to determine the metric that captures the trade-off between those objectives.

The continuity of communications will be also a major challenge to address. Proper Handoff mechanisms need to be defined to keep user session continuity in a high mobility environment (Satellite and UE mobility). Investigations must be carried out to determine if handover must be performed at infrastructure (with or without feedback from UE) or at UE level, and if reselection schemes are not enough in some use cases.

Resilience of communications will be also a major issue, especially in a context where selective jamming can be performed on 5G physical channels (PSS, SSS, PBCH, PDCCH, PUCCH, ...). Detection algorithms have first to be designed to detect these new threats. The relevance of machine learning/IA approaches to this specific problem must be assessed. Mitigation procedures need also to be defined afterwards to grant access to the infrastructure and provide a fallback solution.

8.4.5 TN/NTN spectrum coexistence techniques

NTN is a fundamental complement to terrestrial networks to expand and achieve enhanced coverage. Located at distances of few kilometres from the Earth, NTN offer wider view of Earth and, as a consequence, its coverage contains several terrestrial base-stations, which may operate on the same spectral band if not properly coordinated. While initial 5G standard considers limited carriers for NTN coverage, it is question of time that NTN gains more rights in operating in wider spectrum bands. In such situation, different solutions can be envisaged, ranging from complete coordination between TN and NTN up to minimal information exchange between TN and NTN. Obviously, minimal coordination renders the most challenging case. Ensuring service continuity usually suggests going for orthogonal frequency assignments. This situation may revert if some intelligence is equipped at the transceiver nodes, to monitor and detect spectrum congestion, and point towards a more efficient spectrum assignment. Full coordination between TN and NTN, on the other hand, brings the best situation in terms of performance, but suffers from signalling overhead (and lack of standardized way to share this signalling information). Some database-assisted mechanisms and related signaling protocols have been implemented and demonstrated in practice, for example citizens broadband radio service (CBRS). However, the models cannot support very dynamic approaches needed in future networks.

The spectrum coexistence between TN and NTN seems unavoidable and would therefore increase the probability of interference events. Interference detection and classification gains relevance in this case: mechanisms that can notify the network control center and suggest alternatives for mitigation strategies. Artificial intelligence can have a fundamental role in this task, as they can learn the non-altered expected signal from historical recording, and quickly identify the presence of anomalies. Required messaging protocols and sharing mechanisms allowing quick information exchange and adaptation of spectrum use need to be developed.

8.4.6 Air interface definition for FSO links (feeder and ISL)

Recent advancements, including mega-constellations by private investors and 3GPP's standardization efforts [C8-21]-[C8-23], highlight the critical role of satellite systems. While traditional geostationary and low Earth orbits remain essential, innovative approaches are needed to expand satellite service applications. Ground infrastructure must be scaled to meet user demand and prevent network bottlenecks, especially at the interface between terrestrial and non-terrestrial systems. Future satellite networks will feature enhanced connectivity, with satellites communicating with both ground stations and other satellites. This integration aims for seamless global coverage with minimal latency, heavily relying on interconnection capacity within the 3D network. Optical wireless communications (OWC) will be crucial for providing high speed, resilience, and security within these non-terrestrial networks (NTN). Optical inter-satellite links (OISL) offer significant advancements, including higher data rates, reduced interference, improved energy efficiency, and enhanced security [C8-24][C8-25]. However, maintaining precise Pointing, Acquisition, and Tracking (PAT) capabilities in the dynamic environment of LEO satellite constellations is challenging. These mobile satellites require constant adjustments to keep laser beams aligned and minimize signal loss [C8-26]. Overcoming obstacles such as rapid satellite movements due to orbital maneuvers, atmospheric drag and platform micro-vibrations, as well as synchronization errors is essential for optimizing OISL performance and ensuring seamless communication. In satellite-to-ground links, atmospheric conditions degrade optical link performance. Impact of cloud cover can be reduced by exploitation of site diversity and OISLs. Techniques such as adaptive optics, hybrid RF/FSO schemes, beam shaping, beamforming, advanced coding, time-diversity, integrated communication and sensing capabilities in the optical domain, and multi-connectivity strategies can enhance signal robustness and mitigate errors caused by optical channel impairments such as turbulence, scintillation, fog presence, etc. [C8-27]. AI techniques may be useful to minimize the need for switching between optical and radio systems, such as mmWave and TeraHertz frequencies. Developing regenerative satellite payloads to support both optical and radio communications with minimal switching will enhance both optical inter-satellite and feeder links [C8-28]-[C8-30]. Integrating optical fiber with free-space optical networks at the ground segment will be key to achieving a fully optical network, which is essential for a high-capacity 6G transparent network [C8-31]. Finally, experimental validation is the crucial final stage that all proposed new optical technologies must undergo to support 6G and beyond services over NTN. This validation ensures the reliability and performance of these technologies in real-world conditions.

In summary, the main motivations for developing optical wireless communications in Non-Terrestrial Networks (NTNs) are:

- The first generation of mega-constellations, primarily consisting of LEO satellites, will be deployed in 6G. These LEO satellites require high-capacity and fast links due to their high dynamics, reduced visibility time between links, and the need to transport large amounts of traffic. Optical links present a potential solution for this scenario.
- OWC systems can be easily scaled and adapted to various deployment scenarios, from inter-satellite links (ISLs) to satellite-to-ground links. This flexibility makes them suitable for a wide range of applications and network configurations.
- 6G will demand high-capacity networks. Combining multiple bands, including radio and optical, appears promising for creating energy-efficient systems, enhancing robustness to atmospheric impairments, and providing high-capacity networks.
- Integrating optical fiber with free-space networks aims to achieve fully optical networks that offer high capacity and security.

- OWC systems can be integrated with emerging technologies such as AI, machine learning, and adaptive optics to enhance their performance and reliability. These technologies can help in real-time beam alignment, error correction, and optimal resource allocation.

The challenges of free-space optical networks over satellite include:

- Efficiently combining optical and radio systems is necessary to minimize the outage probability of optical links in the presence of atmospheric impairments (satellite-to-ground) and fast dynamic scenarios with high traffic load (inter-satellite links) (Challenge 1, Short-term).
- Achieving Pointing, Acquisition, and Tracking (PAT) systems with reduced latency, robust tracking, and fast pointing is key, especially when dealing with mobile satellites like those in LEO (Challenge 2, Short-term).
- Developing robust air-interfaces that can handle dispersive impairments (similar to optical fiber networks) on one end and turbulence (from free-space satellite networks) and harsh conditions of space e.g. temperature variations, radiations (in the inter-satellite links scenario) on the other. Vertical integration with low complexity intermediate interfaces and with high spectral efficiency is essential (Challenge 3, Mid-term).
- Minimizing outage probability in free-space optical links involves integrating solutions across the network, MAC, and physical layers, such as multiple transport layers, fast routing, adjusting optical waveform parameters, time-diversity and utilizing satellite multi-connectivity diversity (Challenge 4, Mid-term).
- Integrating communication systems with sensing components will develop comprehensive optical communication systems for non-terrestrial networks. This integration will provide added-value services to operators and enhance communication systems by optimally selecting the satellite or ground station for data transmission (Challenge 5, Long-term).
- Validating the integration of optical links through real demonstrations in satellite-to-ground, satellite-to-drone, and inter-satellite links is essential. This includes testing not only optical technology but also its combination with radio technology. This is a significant challenge that future satellite constellations must address (Challenge 6, Long-term).

The results achieved in this assessment remarks that a satellite network composed of O-ISL is characterized by experiencing temporal link outages derived from the transition time to disconnect and connect topology links. These link outages represent the main cause of a set of problems that needs to be addressed. For instance, a link outage may generate a large amount of data loss considering data streams and the expected link capacity (hundreds of GB in case of 10 Gbps interfaces and 17 seconds of outages). This requires developments associated with (1) large buffering systems capable to manage this data, (2) assessment of end-to-end delay increase due to link outages (and buffering), (3) fast data re-routing procedures to minimize data loss, and (4) minimization of the link transition time.

8.4.7 Research Challenges

Research Theme	**Waveform design**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Exploit multi-satellite diversity	Mid-term (finished in 5y)	Waveform design able to handle the differential time of arrival and Doppler effects between the visible satellites.	Increase the system throughput and the link reliability.
Challenge 2 GNSS-independent operation	Mid-term (finished in 5y)	Waveform design able to establish satellite links in presence of user positioning errors.	Capacity to operate with poor GNSS coverage and in denied scenarios.
Challenge 3 AI native air-interface	Long-term (finished in 7y+)	Learning interface that selects the most suitable features according to	Make an efficient use of the resources

		the environment and the propagation conditions.	

Research Theme	**(Distributed) multiantenna systems**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Low-complexity multi-user beamforming and user scheduling combination	Short-term (finished in 3y)	Techniques exploiting the full potential of large arrays considering realistic impairments (e.g CSI acquisition) and constraints (computational resources)	Increasing total satellite-constellation capacity.
Challenge 2 Beam widening solutions for broadcast transmissions	Short-term (finished in 3y	Techniques and antenna architectures trading off EIRP loss and the number of resources required to cover the required field of view	Efficient use of large arrays for broadcast signaling, initial cell acquisition, etc.
Challenge 3 Time and frequency synchronization of multiple distributed satellite platforms	Mid-term (finished in 5y)	Time and frequency synchronization solutions using intersatellite links	Key enabling technology for the realization of large coherent sparse arrays in space
Challenge 4 Low-complexity distributed MIMO processing	Mid-term (finished in 5y)	Distributed algorithms for user-centric beamforming using spatially separated platforms with limited signalling overhead	Distribution of computing tasks in several platforms instead of using a large central processor. The approach presents advantages in particular in terms of modularity, communication overhead and reliability
Challenge 5 Fully flexible resource allocation	Long-term (finished in 7y+)	Advanced scheduling and resource allocation solutions allowing a fully flexible management of time, frequency and space resources with a large coherent sparse array in space.	More sustainable exploitation of system resources
Challenge 4 Formation flying for large swarms	Long-term (finished in 7y+)	Advanced orbit control solutions for large satellite swarms	Unfolding the potential of coherent joint transmission approaches by allowing the design of swarms with more than a few nodes

Research Theme	**Integrated communications, sensing and positioning**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Integration of satellite-based Positioning, Navigation, and Timing (PNT) in 6G	Mid-term (finished in 5y)	High-precision, robust navigation capabilities, and enabling high-quality communication services	Break the dependency on GNSS
Challenge 2 Realization of Multi-Functional Satellite Systems (MFSS) with *communications, sensing and positioning capabilities*	Long-term (finished in 7y+)	Full flexibility in the operation of communication systems with aided sensing capabilities	Contribute to the internet of sense promised by 6G from the NTN

Research Theme	Radio Resource Management		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Random-Access procedures and timing advance algorithms.	Mid-term (finished in 5y)	Procedures need to deal with delays and NTN impairments and support different verticals.	Keep UE synchronization with satellite infrastructure and deal with conflicting QoS constraints.
Challenge 2 Unified Multiple Access technique	Mid-term (finished in 5y)	Replace the multiple access schemes by a single, unified, and general multiple access scheme.	Effective in multi-functional networks, where the variety and complexity of services, use cases, and deployments are rapidly expanding.
Challenge 3 AI based RRM distribution and planning	Mid-term (finished in 5y)	AI based efficient RRM mechanisms to consider the resources scheduling and planning in a competing objectives framework.	Optimized and automated resource management to enhance QoS, deal with traffic heterogeneity and interference.
Challenge 4 efficient Handoff mechanisms	Mid-term (finished in 5y)	Proper Handoff mechanisms to deal with the NTN high mobility environment and decide the required procedures and feedback between UE and NTN infrastructure.	Maintain efficient and reliable communication
Challenge 5 Measures to deal with selective jamming on physical channels	Mid-term (finished in 5y)	Design of detecting algorithms to the physical channel jamming and mitigation procedures.	Maintain access to the NTN infrastructure.

Research Theme	TN/NTN spectrum coexistence		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Distributed spectrum monitoring, congestion detection	Mid-term (finished in 5y)	Efficient techniques that permit assess the spectrum usage and take coordination decisions	Enabler for TN/NTN coexistence
Challenge 2 Interference detection and classification	Mid-term (finished in 5y)	AI based interference detection and classification algorithms	Enabler for TN/NTN coexistence
Challenge 3 Architectures for efficient full TN/NTN coordination	Mid-term (finished in 5y)	Architecture putting together distributed spectrum monitoring, interference detection and centralised/distributed decisions for efficient TN/NTN spectrum coexistence	More spectrum available for both TN and NTN segments

Research Theme	FSO		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Integration of optical and radio bands	Short-term (finished in 3y)	Integrating optical and radio bands will enhance robustness against channel impairments and reduce satellite link acquisition times. This approach is effective for both inter-satellite and satellite-to-ground links. Key metrics for this challenge include the throughput of the combined system compared to using a single band, and the outage probability of providing service (combined vs. single band).	The contributions will be on physical and MAC layers of 6G networks with algorithms/schemes that support traffic in both optical and radio bands, along with receivers capable of managing information from multiple bands, will enhance link detection capabilities. The resulting system will offer a network with higher capacity, flexibility, and robustness to channel impairments.

Challenge 2 PAT systems able to work in high dynamic scenarios	Short-term (finished in 3y)	The development of mega-satellite constellations will require the use of inter-satellite and ground-to-earth links that support high dynamic scenarios. So, PAT systems with fast acquisition time, robust to channel impairments, and with high tracking capability will be required. The metrics for the PAT system will be : outage probability of losing the link, acquisition time, and tracking speed capability.	The contribution will be to develop synchronization, acquisition, tacking systems that permit to obtain PAT systems for inter-satellite and satellite-to-ground with fast acquisition times, robust to channel impairments, and high tracking capability. Thus, it will be obtained NTN of high capacity, high quality of service and robust to channel impairments.
Challenge 3 Integration of fiber with free-space optical networks	Mid-term (finished in 5y)	Use integrated waveforms that permit to support dispersive turbulence effects, temperature variations, and solar radiation, in a spectral efficient way. Key performance indicators would be quality of services, capacity of the joint system, complexity in the interface between the two networks.	Full optical networks of high capacity enable the perfect integration of terrestrial and non-terrestrial optical networks, ensuring no loss of capacity at their interconnections. This seamless integration supports a wide range of new use cases and services for both non-terrestrial and terrestrial networks.
Challenge 4 Integration of information from multiple layers	Mid-term (finished in 5y)	To minimize the link outage of the optical satellite links, fusion of information multiple layers (network, MAC and physical layer) will be required. Thus, key metrics from different layers will be required to use to achieve designs that optimize the performance in a global view. Time of acquisition of the links, Number of hops, BER, and spectral efficiency will be combined.	New algorithms of routing, combination of multiple transport layers, satellite diversity techniques, efficient physical layer designs such as coding, modulation will outcome. By doing so, network resilience without penalizing in excess the spectral efficiency will be targeted
Challenge 5 Integration of optical communication and sensing	Long-term (finished in 7y+)	The use of AI technologies will permit to develop new optical transceivers that integrate the information from multiple sources in highly dynamic environments.	Development of communication and sensing technologies will help to provide fast PAT systems, reduced complexity optical systems, and additional revenues for operators
Challenge 6 Validation of Optical Links in Non-Terrestrial Networks	Long-term (finished in 7y+)	The development of satellite constellations will require the use of optical links to communicate from satellite-to-satellite, satellite-to-UAV, satellite to ground. Experimental validations of the proposed solutions will be required to provide successful services over NTN.	Development of solutions that integrate optical and radio for NTN in all satellite links, new mitigation techniques of the channel impairments, integrated solutions of multiple transport networks (optical fiber, optical free-space, radio)

8.4.8 Recommendations for Actions

Research Theme	Waveform design		
Action	Exploit multi-satellite diversity	GNSS-independent operation	AI-native air interface
International Calls	X	X	X

	It is addressed in NTN SNS projects. Still, new calls are required to increase the TRL.	NTN SNS projects are investigating alternative positioning methods. Future international calls shall be built upon the solutions developed in previous calls.	Future international calls should target the adoption of a native AI framework for the RAN.
International Research	X International research cooperation is required to further improve the physical layer and the coordination mechanism between satellite access nodes.	X The connection establishment procedure in the presence of positioning errors relies on a more robust air interface. International research will be beneficial to develop the required enhancements.	X The AI-native air interface design is still in its infancy. International research is fundamental to further develop this technology.
Open Data			X Open data sets with close to real operational data is essential to implement new solutions
In-orbit demonstrator	In-orbit validation represents a key milestone for the technology development	In-orbit validation represents a key milestone for the technology development	In-orbit validation represents a key milestone for the technology development
Cross-domain research			The combination of signal processing and data science is beneficial to extract features and build new models.

Research Theme	**(Distributed) Multiantenna systems**			
Action	Colocated multiantenna systems: Low-complexity multi-user beamforming and scheduling for data and broadcast transmissions	Distributed satellite platforms: Time/frequency synchronization and low complexity processing	Fully flexible resource allocation	Formation flying for large swarms
International Calls	X New international calls required following those already running such as NTN SNS projects, with the major aim of increasing TRL.	X International space agencies such as ESA or NASA are funding research activities on distributed systems. Future international calls shall ensure to build upon the technological trends identified during these studies.	X New international calls required following those already running such as NTN SNS projects, with the major aim of increasing TRL.	X International space agencies such as ESA or NASA are funding research activities on distributed systems. Future international calls shall ensure to build upon the technological trends identified during these studies.
International Research	X International research and strong consortiums required to increase TRL targeting in-orbit validations at the end.	X International research must aim at bringing together the experts required to realize an in-orbit demonstrator of the most promising	X International research and strong consortiums required to increase TRL targeting in-orbit validations at the end.	X International research must aim at bringing together the experts required to realize an in-orbit demonstrator of the

		multi-antenna technologies.		most promising multi-antenna technologies.
Open Data	X Open data on user demands required due to the tight link between beamforming and scheduling design	X Open data on user demands required due to the tight ink between beamforming and scheduling design	X Open data crucial to design RRM solutions with high TRL	
In-orbit demonstrators	X In-orbit validation represents a key milestone for the technology development	X In-orbit validation represents a key milestone for the technology development	X In-orbit validation represents a key milestone for the technology development	X In-orbit validation represents a key milestone for the technology development
Cross-domain research	X Requires joint efforts from array design, signal processing and system engineering	X There is a strong need for multidisciplinary research efforts especially in the domain of array processing, formation flying and system engineering		X There is a strong need for multidisciplinary research efforts especially in the domain of array processing, formation flying and system engineering

Research Theme	Integrated sensing and positioning	
Action	Integration of satellite-based Positioning, Navigation, and Timing (PNT) in 6G	Multi-Functional Satellite Systems (MFSS) with communications, sensing and positioning capabilities
International Calls	X This is a hot topic that has been already included in some recent calls, but it deserves a dedicated call aiming TRL increase beyond 4.	X Dedicated calls aiming for satellite -based network sensing applications, architectures and performance are required.
International Research	X International projects with strong consortiums will permit the desired TRL increase	X An international approach is required to assess the potential of NTN technologies in network sensing applications
Open Data		
Large Trials		
Cross-domain research	X PNT and communications traditionally followed independent parallel paths that must meet for successful 6G implementation	X A join effort of radar, Earth observation and communications community is needed to assess the potential of satellite based ICAS.

Research Theme	Radio Resource management - I		
Action	Random-Access procedures and timing advance algorithms	Unified Multiple Access technique	AI based RRM distribution and planning
International Calls	X Some current HE and SNS NTN projects are considering this aspect. More calls are needed to build upon the results and provide more enhancement.	X This item is still in its infancy and requires more international calls.	X Several SNS NTN projects are considering AI for RRM. However, this is a very wide field with high importance to the sector and more calls are needed to enhance the performance.
International Research	X Strong and varied consortiums will help in the research and the collaborative effort can help in to ensure compatibility among different regions and technologies and improve the impact to the standardization efforts.	X International research is fundamental to further develop this technology and to support a variety of communication needs.	X International research must aim at bringing together the experts in AI and telecommunication to improve network performance by dynamically allocating resources based on real-time data and predictive analytics.
Open Data			X Open data crucial to design efficient RRM solutions, improve the accuracy and the robustness of the AI algorithms and increase the TRL
Large Trials	Large scale testing is an essential milestone for this technology.	Large scale testing is an essential milestone for this technology.	Large-scale trials are necessary to evaluate the performance of AI-based RRM systems in diverse network environments.
Cross-domain research			Cross-domain research is crucial in harnessing the potential of AI for RRM. Combining expertise from data science, optimization and telecommunications can lead to innovative solutions that significantly improve network efficiency and user experience.

Research Theme	Radio Resource management - II	
Action	Efficient Handoff mechanisms	Measures to deal with selective jamming on physical channels
International Calls	X More international calls are needed to expand research and development efforts in this area, ensuring better support for high-mobility scenarios and integrated networks.	X More international calls are needed to enhance network security and to further investigate and mitigate the impact of such jamming attacks

International Research	X Strong and varied consortiums will help in developing more effective Handoff mechanisms and ensure interoperability across different network technologies and vendors.	X Collaborative works is required to develop robust detection and mitigation techniques. Short comment
Open Data	Open data can highly enhance the decision of the handoff trigger event and performance metrics and improve the understanding of the handoff dynamics. This will lead to efficient and adaptive handoff strategies.	Open data sets of jamming incidents and countermeasures can provide valuable insights for developing effective mitigation strategies
Large Trials	Conducting large-scale trials is essential to test the performance and reliability of new handoff mechanisms under various conditions	Large-scale trials are necessary to evaluate the effectiveness of anti-jamming techniques.
Cross-domain research	Integrating aspects from signal processing, network theory, and artificial intelligence can lead to innovative solutions that enhance user experience and network performance.	Signal processing, security and network experts are required

Research Theme	**TN/NTN spectrum coexistence**		
Action	Distributed spectrum monitoring, congestion detection	Interference detection and classification	Architectures for efficient full TN/NTN coordination
International Calls	X	X	X
	Rather old topic covered in multiple past calls, but requires a renewed dedicated attention to potentially impact on 6G deployments		
International Research	X	X	X
	International research including main NTN and TN actors required for a successful technology development		
Open Data	-	X AI-based interference detection models require large data sets for their development and training.	-
Large Trials	-	-	-
Cross-domain research	-	-	-

Research Theme	**FSO - I**		
Action	Integration of optical and radio bands	PAT systems able to work in high dynamic scenarios	Integration of fiber with free-space optical networks
International Calls	X In the research calls generally the FSO and radio systems have been studied separately, However, both systems can benefit from each other to achieve integrated systems of high capacity, resilience and security.	X The study of PAT systems in FSO satellite systems is a need, FSO systems have a large bandwidth, but the PAT systems are not enough fast to support fast handovers, multi-connectivity. etc for the use cases of 6G. Calls in these areas are needed to have seamless and transparent 6G satellite communications.	X Initial projects have started to study the integration of optical fiber with free-space optical systems. Toward this regard, the extension to satellite networks is a demand to achieve full optical networks.
International Research	X Multidisciplinary research on the area of integrated FSO and Radio is a need to obtain efficient, low cost and portable	X It is also a need that multiple expertise to develop PAT systems is a need. Expertise from Signal Processing, Optical Systems,	X Teams of FSO and optical fiber networks have to contribute by connecting their different research and testbeds to

	integrated optical and radio systems	Hardware Components have to work together to develop new generation of PAT systems	obtain an experimental pan-European network of FSO and optical fiber network with satellite connectivity.
Open Data	X Provide open data of the optical and radio systems from the satellite-to-ground, ground-to-satellite and satellite to satellite to foster AI research on this area.	X Provide open data of the FSO systems, especially for constellations, to improve the PAT systems and hand-over, multi-connectivity research area.	X Provide open data of the optical fiber and free-space optical networks to develop efficient integration systems of both transport networks.
Large Trials	X Large trials, pilots would be of interest to have the required data and collaborative research for fostering research on this area. Under this concept is included In-Orbit demonstrations.	X Large trials of collaborative research at multi-satellite level would be of interest for testing different PAT systems and reducing the costs of research. Under this concept is included In-Orbit demonstrations.	X Integration of multiple testbeds by resorting to large scale pilots is also of interest for having a fully optical network. Under this concept is included In-Orbit demonstrations.
Cross-domain research	X Yes, this is an area of cross-domain research since multiple research profiles have to be involved (optical, radio, signal processing, component, channel modeling, etc)	X Yes, this is an area of cross-domain research since multiple research profiles have to be involved (signal processing, component level, satellite segment, constellation designer, etc)	X Yes, this is an area of cross-domain research since multiple research profiles are involved (signal processing, optical fiber expertise, free-space optical propagation knowledge, etc)

Research Theme	**FSO - II**		
Action	Integration of information from multiple layers	Integration of optical communication and sensing	Validation of Optical Links in Non-Terrestrial Networks
International Calls	X In the integration of terrestrial-satellite, optical-radio, communications and sensing different layers interact. A general view and methodology for integrating multipole layers should be investigated. So, international research calls on this topic would be wellcome.	X The ISAC in the optical domain is a recent area of research that may help to improve the efficiency of the communication systems and the revenues of the operators by providing added value systems by introducing the sensing capability. So, international calls on this topic would be of interest.	X The In-Orbit demonstration of Optical Links, which be integrated with the Radio, optical fiber, terrestrial networks should be strongly considered in the future.
International Research	X The integration of multiple sources of information calls to the international cooperation to foster a clear process for 3GPP standardization	X In this research topic also is of interest an international collaboration partnership to integrate multiple strategies of integrated optical communication and sensing techniques	X The validation in-orbit of optical links is assumed to be from international consortiums since different segments and technologies are involved.
Open Data	X It will provide open data of the optical inks from free space optical	X It will provide open data of the optical inks from free space optical communications to	X It will provide open data of the optical inks from free space optical communications to

	communications to improve the AI systems that run on top	improve the AI systems that run on top.	improve the AI systems that run on top.
Large Trials	X It is demanded to proceed with large trials to improve the integration between multiple technologies	X It is demanded to proceed with large trials to improve the integration between optical communication and sensing strategies, either at the satellite and the ground segments	X It is demanded to proceed with large trials to improve the integration between optical communication since optical satellite channels may have a different behaviour at different longitudes and latitudes. Missions with multiple satellites would be of interest.
Cross-domain research	X Yes, this is an area where multiple areas of knowledge are combined (e.g., signal processing, routing, physical layer, network layer, etc)	X Yes, this is an area where multiple areas of knowledge are combined (e.g., signal processing, component level, communications and sensing expertise to name a few of them)	X Yes, this is an area where multiple areas of knowledge are combined (e.g., signal processing, FSO satellite segment, FSO ground segment, etc.)

8.4.9 Expected Impact

8.4.9.1 Key Performance Indicators (KPI)

The air interface must contribute to fulfill the KPIs listed in Sections 8.3.1 and 8.3.7.1. In particular, the waveform design should not only contribute the increase in data rates, but also support a growing number of connections for higher network densification and a most robust and reliable operation not depending on GNSS solutions. Multiantenna solutions and radio resource management policies permit adapting beam coverage frequency and power allocation to the user demand distribution, resulting in a higher overall system capacity and so increased user experienced data rates and supported network densification. In the case of distributed multiantenna solutions with coherent joint transmission, a specific KPI is the spatial resolution of the obtained virtual aperture. This KPI can be evaluated via the achievable throughput area density (in bit/km^2). An alternative to increase data rates and supported network densification is to increase the system bandwidth, but since spectrum is a scarce resource, TN/NTN dynamic spectrum sharing will contribute to the most efficient spectrum usage. FSO becomes also vital since the improved air interface between satellites and users requires high-capacity feeder links and /or ISL, to avoid them becoming the system bottleneck. Finally, 6G satellite-based PNT permits breaking the dependency on GNSS and achieving the positioning KPIs. Satellite-based sensing applications and KPIs are still and open research area that requires an in-depth assessment.

8.4.9.2 Key Value Indicators (KVI)

The air interface design must indirectly help to fulfill sustainable development goals listed in Section 8.3.7.1 and discussed in Section 8.3.7.2. Indeed, the air interface contributes the achievement of all SDG in which ICT and specially NTN may have a significant role. More directly, the efficient use of all available resources through TN/NTN spectrum sharing, RRM and multiantenna solutions enable a higher modularity and flexibility of space systems such that resilient services can be guaranteed whereas the energy consumption required to fulfill the KPIs is minimized. A direct impact of the sensing capabilities of satellite-based ICAS on some SDG can be also envisaged but requires and in-depth assessment.

8.5 Time Varying Networks

8.5.1 Network architecture evolution perspective

The radical shift taken from NTN to implement smart processor in space and the extension of the space segment to form multi-orbit multi-service 3D networks has important implications on the evolution overall of the corresponding network architecture. More importantly, the service model to be considered is expected to significantly change in comparison to the currently available one, where essentially services are distributed either via satellite or terrestrial infrastructures, by without any important overlap (i.e. no actual convergence). On the contrary. the vision for 6G ecosystem is to natively support NTN elements, whereby the integration between TN and NTN envisioned at the later stage of 5G standardization is expected to take the form of an actual unification so that common standards and practices can be shared, so as to develop a network of networks, where being data transported over which network should be undistinguishable from final users. In other words, the evolution of the network architecture is such to be act in a polymorphic manner, i.e. able to naturally evolve and adapt to the service demands and traffic characteristics, hence fulfilling principle of autonomic and self-configuration networking. Achieving such an architectural concept appear then even more compelling now with the native inclusion of NTN segments, which considerably differ in characteristics from the TN counterpart especially from a mobility standpoint that important affect the service model. Further to this, the continuous explosion of edge service oriented architecture plays an important role also at network level because of the necessity of deploying routing/forwarding mechanism able to support such a service model even in the case of multi-tier space networks. Then on the other hand, availability of resources distributed across a very space network architecture necessitates a smarter form of network orchestration in order to more effectively monitor the fluctuation of the network conditions and accordingly taken actions on the path to be established to transport data and accordingly establish/migrate microservices and the related service function chaining necessary to achieve modular communication in a sustainable manner. This evolved communication paradigm must however take into account the constraints imposed by the current 3GPP 5G protocol architecture with respect to the interconnection between gNB and 5GCNs, which completely overlooks the nature of future NTNs, where more extreme functional split options might be deployed and the interconnection between several gNB deployed in space will be witnessed hence necessitating a more agile and effective way to establish data communication from a control layer perspective.

8.5.2 Network Orchestration/Management

Orchestration is needed to handle the complexity of application services that are designed and created as chains of micro-services at the application level and as chains of Virtual Network Functions (VNFs) at the network level (where a relevant framework is provided, in particular, by ETSI MANO—Management and Orchestration [C8-32]). An important aspect, however, not always properly evidenced, is that of the separation of concerns between applications’ and network functions’ orchestrators, which has been stressed specifically by some H2020 5G PPP European Projects (among others, 5G-INDUCE (https://www.5g-induce.eu/) and its precursor MATILDA). More specifically, the domain of vertical cloud-native applications, empowered with the service mesh concept [C8-33] and with suitable sidecar proxies that allow the application developer to extend the microservices’ capabilities with the specification of their communications needs, should be the concern of a Network Application Orchestrator (NAO). The NAO should allow application providers and application developers to operate with the mechanisms of the cloud environment they are used to; however, at the same time, it should enable them to fully exploit the advanced communication capabilities offered by 6G, by abstracting the physical network with the slice concept, transparently with respect to the heterogeneous underlying physical infrastructure (including the NTN segment) and providing the means to convey their communication needs and constraints to the Telecommunication Service Provider (TSP), and to constantly maintain this interaction during the lifecycle of their applications. Through the mediation of the Operations Support System (OSS) the TSP can receive the specifications that characterize a particular vertical application via a slice intent, and has the task to configure, deploy and manage the needed resources for the creation of the slice (by means of the Virtual Network Functions Orchestrator—NFVO—provided in the MANO framework), which is then exposed to the NAO through a well-defined Northbound interface, and can be monitored and reconfigured, if necessary, to maintain QoS requirements. In the NTN environment a similar separation concept was introduced a long time ago by the ETSI Broadband Satellite Multimedia (BSM) architecture [C8-34], which created a clear separation between Satellite Independent (SI) and Satellite Dependent (or, more generally, Technology Independent – TI – and Technology Dependent – TD) layers, with the interaction granted through a Satellite(Technology) Independent-Service Access Point S(T)I-SAP. As such, this architectural perspective can play a similar role within the network domain, by separating what is closer to the physical infrastructure, where functionalities may be implemented by means of a mix of VNFs and Physical Network Functions (PNFs), orchestrated by a NTN-Dependent Orchestrator (NTN-DO), from what pertains to the VNFs in the Technology-Independent layer. This separation concept could greatly foster the integration of the NTN segment toward 6G, along the lines that were sketched in [C8-35]. It is worth noting that with the growing importance of LEO small- and nano-scale satellite constellations and, in general, of the hierarchical multilayer structure of LAP, HAP, LEO, MEO and GEO, the complexity of the NTN segment will challenge the traditional orchestration framework. The NTN is characterized by the presence of multiple constellations, intermittent connectivity, intersatellite links and, in general, more distributed data and control plane functionalities. In this framework, the SI-SD separation acquires even more momentum. In this architectural layout, it seems reasonable to conceive a specific NTN-Dependent Orchestrator (NTN-DO). Here, an NTN Control Center (NTN-CC) can play a similar role of mediation point as that of the OSS for the NAO/NFVO interface with respect to the NFVO/NTN-DO interface. The creation of a slice in this case would start from the request of the NAO and would be passed along to the NFVO through the slice intent; in the case that NTN resources were needed, the NFVO would request the appropriate configuration to the NTN-DO via the NTN-CC. It is also worth noting that a similar vision about orchestration of integrated TN/NTN networks is shared by the IEEE INRG Satellite Working Group [C8-36].

8.5.3 IP-Forwarding Payload

Internet access from spaceborne nodes has been provided long time ago with dedicated Satellite Communications (SatCom) systems. This service provision is mainly focused on tunneling or encapsulating Internet Protocol (IP) packets over air standardized protocols, like DVB-S2X [C8-37]. This Internet access service experienced a relevant growth with the apparition of mega-constellations initiatives [C8-38]. Specifically, Starlink is currently providing service to a large number of users around the globe to have Internet access from anywhere. Although this service is currently being provided, the proposed solution remains on the Very-Small-Aperture Terminal (VSAT) architecture [C8-39], in which a custom and privative terminal provides access to the user towards the entire network. Additionally, gateway nodes are located between satellite systems and terrestrial networks to manage interoperability. The associated used protocol follows the same premise of encapsulation or tunnelling, and in some cases is obscure.

On sparse constellations cases, network connectivity remains intermittent and temporal. The resulting network is conceived as a Delay-/Disruptive-Tolerant Network (DTN) [C8-40]. Unlike highly connected networks, DTNs cannot determine a route between two nodes at specific time. Instead, routes over the time can be defined. This novel route definition founds on the capability of the intermediate nodes to store, carry, and forward (SC&F) packets until the next link is established. To achieve this procedure, the Bundle Protocol

(BP) and other DTN-friendly protocol family were standardized [C8-41]. However, as in the previous case, the BP-based architecture is founded on the deployment of Bundle Agents that corresponds to entities capable to execute the BP, and thus the SC&F procedure. This, at the end, corresponds to a proxy or gateway element that differentiate between DTNs and non-DTNs.

In addition, since the creation of the Internet, the volume of exchanged traffic has grown considerably, from less than 100 GB per month in the late 1980s to an expected volume of nearly 400 billion GB in 2023. Due to the fact that the Internet is still growing, in terms of traffic and coverage, with the envisioned integration of NTN and terrestrial networks, there is the need to investigate suitable architectural alternatives to the existing one. One of the most prominent future Internet architectures is Information-Centric Networking (ICN), which addresses data using data identifiers and forwards packets based upon such identifiers instead of host identifier. This shifts the current host-centric Internet paradigms towards a new data-centric approach. ICN enables a consumer to request a given data object in the network without any knowledge about the location of the requested data. The paradigm shifting from a host-centric to a data-centric approach brings several benefits to the operation of large-scale satellite networks, namely the adaptation to intermittent connected networks based on a pull communication model and in-network caching, as well as extra flexibility to handle different types of traffic, based on an extended set of forwarding strategies. While some analysis about the development of ICN based satellite systems have been made, some new ICN based architectures have been proposed to support a universal networking system able to encompass also space borne and airborne platforms

After reviewing these developments, it can be concluded that currently satellite networks cannot be considered as IP-based networks. Although they are capable to tunnel IP traffic, satellite systems do not process at this layer. This results in a set of difficulties like (1) increase of overhead due to accumulative encapsulation [C8-42], (2) required intermediate elements (i.e. gateways and proxies) to achieve end-to-end interoperability, or (3) satellites cannot be directly interacted with IP-based functions (e.g. Core Network functions). It is thus essential to develop and conceive an IP capable to work in satellite network, and integrate them to facilitate network interoperability.

This potential IP extension shall to satisfy satellite systems requirements. In this way, header compression mechanisms, integration of SC&F like in [C8-43], end-to-end or point-to-point congestion control in DTN like in [C8-44], and support for native content-aware networking (i.e., ICN paradigms) among other challenges shall to be addressed to achieve this necessary NTN-enabling technology.

8.5.4 Routing in space

Global- and low-latency connectivity in space requires a routing algorithm for deciding, either at the terrestrial or space source, or in every intermediate node, the directions to be used to reach the terrestrial or space destination. Due to the predictability of the network topology, the conventional approach to routing in NGSO constellations is to centrally compute all the paths in a terrestrial location register, or a GSO satellite if it is a hybrid NGSO/GSO network, and then broadcast the information to all the satellites. In any case, this approach does not scale well, especially in mega constellations, and it creates a dependency on the limited contact times with the terrestrial or the high latencies of the GSO segment. Other challenges of space routing are related to the dynamic imbalanced load and the stringent energy and computing constraints. Large scale satellite networks aim to transfer large volumes of data between tens of thousands of satellites that move continuously at high speeds in different orbits. To support this aim, there is the need to develop new protocols for routing traffic from different services. Such protocols can leverage existing solutions to route data over a large set of mobile devices, based on specific algorithms such as Contact Graph Routing (CGR) as well as new paradigms such as information-centric networking. Future satellite networks should be designed to carry traffic belonging to different services with distinct specifications in terms of traffic performance, reliability and robustness. However, the continuous motion of satellites poses significant difficulties to traditional routing protocols. For instance, as satellite constellations become very large, the routing may never fully converge, resulting in a sub-optimal network. Moreover, there is the need to develop routing strategies able to consider different service semantics described by a combination of fields in the packet header as well as a transported set of instructions. Such routing strategies require a data plan able to support programmable network functions (e.g., forwarding) and services.

DTN Link State Routing protocols, such as CGR, face a new set of challenges in their application within LEO satellite constellations. In particular, they impose significant hardware requirements on satellites. Each satellite is required to store extensive Contact Plans, potentially containing millions of contacts in large networks, such as Starlink. These requirements impact the manufacturing of large quantities of satellites, making constellations less cost-effective. Furthermore, in very large time varying networks, graph-based representations using different kinds of time-varying graphs tend to scale poorly. Network scalability is however key to the operation of large LEO satellite constellations. This is further supported by the predicted levels of network traffic that they are required to handle, given that they are expected to serve densely populated areas with potentially millions of heterogeneous user devices.

All of this leads us to conclude that existing routing protocols for routing in space are not optimally adapted to LEO constellations. IP-based solutions for addressing node mobility are not suitable for LEO constellations. DTN, although designed for routing in space, may be too complex and not optimally aligned with the LEO constellation use case. For this reason, alternative routing paradigms may be exploited, by leveraging the dynamic yet partially structured nature of LEO constellations. Semantic routing is the process of achieving enhanced decisions based on semantics added to IP headers aiming to provide differentiated paths for different services. The additional information or "semantics" may be placed in existing header fields (e.g., the IPv6 Traffic Class field), may be added to new header fields, or it may be encoded in the payload or on additional headers, such as the IPv6 Extension Header. The application of semantic routing allows packets from different services to be marked for different treatment in the network. The packets may then be routed onto different paths according to the capabilities and states of the network links and nodes, in order to meet the performance requirements. For example, one service may need low latency, while another may require ultra-low jitter, and a third may demand very high bandwidth. Examples of existing semantic routing usage in IP-based networks include: i) using addresses to identify different device types so that their traffic may be handled differently; ii) expressing how a packet should be handled as it is forwarded through the network; iii) enable Service Function Chaining (SFC); iv) forwarding packets based on carried data rather than the destination addresses; v) or formatting geographic location information within addresses.

Geographic information may be an interesting system semantics to exploit to route traffic on a LEO constellation, instead of interface identifiers of satellites, which may change due to the intermittent nature of nodes and links. Proposed greedy geographical routing protocols, such as GPSR, are inherently scalable and efficient in highly dynamic networks. However, the local maximum problem and the consequent need for planar graphs to route around obstacles add complexity to the networks. For geographic routing protocols in general, it can be observed that the scalability problems seem to stem from either the need to maintain localized state information, or the need to planarize the network topology graph.

A combination of geographic identifiers with other routing semantics, may be achieved by leveraging the concept of Segment Routing. Segment Routing is already deployed in terrestrial networks and is known for its high performance and ability to enable complex traffic engineering. Hence a potential solution may pass by extending the SR architecture to include geographic segments. This allows the use of greedy geographic routing on specific network segments, i.e., a satellite segment, while seamlessly integrating with the terrestrial infrastructure. Further discussions on these problems are addressed in Chapters 2 and 3, and is still a area of active research (including novel concepts as semantic routing).

8.5.5 Deterministic Networking

The newest evolution of 5G aims to add support for new applications and use cases. Based on this evolution 5G is expected to bring significant enhancements around smarter network management by incorporating AI/ML techniques for beam management and load balancing for instance. 5G aims to support lolatency audio and video streaming services aimed for Extended reality (XR), along with a more energy-efficient use of network resources, and Deterministic Networking (DetNet) capabilities to ensure deterministic data paths for real-time applications with extremely low data loss rates and packet delay variation.

Whilst 5G is about adapting the already established generation for new incremental use cases, 6G is natively designed for the human digital needs of the next decade, including always best-connected service via the integration of NTN.

Satellites and 6G users have several common aspects. Both are moving nodes in adaptive, time-variant networks. Although users move by following complex adaptive patterns, satellites describe more deterministic paths in orbit, and by choosing their orbiting geometry carefully, connected constellations can be deployed to achieve global coverage with low latency and smaller propagation losses.

On the other hand, typical 6G applications, such as cyber-physical continuum has straight quality requirements, since it may comprise operations such as control and automation, wearable robotics and exoskeletons, as well as Extended Reality (XR). All these applications comprise of intimately coupled communication as well as computer loads, while requiring deterministic (i.e., guaranteed) performances.

Considering a 6G NTN to TN scenario that supports critical applications (e.g., industrial applications), then in such scenario there is the need to ensure convergence across NTN to TN considering diverse wireless (and fixed) TSN-capable solutions, such as Wi-Fi 6/7 TSN, and a number of cyber-physical components. Hence, the critical requirements to observe (e.g., bounded latency, low jitter, zero packet loss) have to be addressed from an end-to-end perspective, where the extremes correspond to cyber-physical systems such as IoT sensors, or robots (mobile or not).

Three main challenges arise in this context. The first relates with convergence at a technological perspective, and between the TSN domains/DetNet core. Such convergence needs to take into consideration the following aspects: i) time awareness (and tight synchronization); high reliability and availability; iii) integrated management; iv) standardised interfacing. The second challenge relates with interoperability and TSN integration support in 6G. While 5G is compatible with Ethernet/TSN, there are still several challenges to allow an adequate TSN/Wi-Fi6/7 integration with the 5G core, from a management perspective. The third challenge concerns resilience and self-healing of the overall system. Tight time synchronization is crucial to achieve the requirements of critical applications. From an end-to-end perspective, it is important to address a deployment that can cope with failures and recover fast. Resource management/orchestration requires a more complex approach, eventually integrating estimations of potential hazards into systems that have been traditionally closed, and operating with a tight human control. For further global architecture discussions, Chapter 2 covers these aspects at length, while Chapter 3 discusses protocols, although deterministic networking is particularly highlighted as a challenging topic in Section 7.5.

8.5.6 Mobility Management

Large LEO satellite constellations are proposed to provide global low-latency high-bandwidth Internet connectivity. To achieve low-latency connectivity, LEO satellites are launched at low altitudes (160-2000 km), hence moving fast relative to the ground. Therefore, continuous connectivity requires a large number of satellites. In addition, ground stations and users have frequent handovers/disconnections in communications with satellites. For example, a LEO satellite at 500 km altitude travels at 7.6 km/s and it takes about 95 minutes to orbit the Earth, resulting in a handover every 5 minutes approximately.

We hypothesize that today's LEO constellations can be equipped with a certain amount of computing and storage resources (see section about edge computing). Therefore, handling mobility of users based on a framework able to support data caching and local computing, such as information-centric networking may bring benefits for LEO constellations, such as adaptive forwarding, in-network caching, off-the-grid communication, data mule service, in-network/edge computing, mobility support, and data-centric security.

Specifically, about mobility management, building an IP-based LEO satellite network faces significant challenges, including location management and handover management. In contrast, the information-centric networking paradigm provides a data-centric architecture with a pull communication model, which can better assist users (consumers) to retrieve data in mobile scenarios, i.e., users simply retransmit requests (Interests) after a satellite handover, reducing the complexity of location and handover management.

However, this basic mobility support provided by information-centric networking may incur extra delays, because timeouts are used as signals for network congestion. In this context, Interest retransmission inside the satellite network may improve the overall system performance, since a consumer's forwarder may be able to detect the link change and retransmit stored state. On the other hand, in the case of producer mobility, more research is needed in order to overcome some challenges such as Interest packet loss, Interest retransmission, long handover latency, high costs, and a non-optimal routing path.

8.5.7 Service discovery

In 6G networks, an integrated terrestrial - NTN network can address gaps that currently exist in terrestrial networks. Services carried over such an integrated network are expected to cover broadband services, mobile backhaul, Internet-of-Things and Vehicle-to-Everything. To support the envisioned set of future services, besides being able to route traffic within the space network, there is the need to devise intelligent and flexible networking solutions able to support not only packet switching, but also data storage and processing, while being able to react in real time to the requirements of new operational intents. Such future satellite network aims to reduce data rate requirements, increase energy efficiency, and guarantee end-to-end network connectivity. Therefore, a new network architecture for large-scale multi-orbit satellite systems should be able to abstract a set of networking, storage and computing resources in the form of end-to-end services, deployed to fulfill a set of operational intents that may change over time. To sustain a large set of services, the envisioned network architecture should rely on a programmable data plane, flexible enough to support different services based on a chain of virtual network functions, such as distinct forwarding mechanisms.

From an end-to-end perspective there is the need to develop routing strategies able to consider different service semantics (e.g., load balancing), as well as to exploit a mesh of free space optic links. This goal may require converging optical transport with routing functionality, increasing power efficiency and scalability. Such capability to route other an NTN based on the properties of the services available at the edges of the satellite network may be a novel approach that can be applied to develop a framework in which services instead of communication hosts determine the addressing semantics, in a way that is more aligned with the intermittent nature of satellite networks. Such approach should not aim to replace existing service routing capabilities, most notably Domain Name Service (DNS) as the main form of resolving a service name into a routing locator, since DNS is working well for many general Internet services. However, it is clear that in some specific challenging service scenarios, such as LEO constellations, may benefit from embedding service routing capabilities in the networking stack, without relying on application layer translations or resolution services.

In such service centric approach, the edges of a satellite network may need to have the ability to find where services are stored at the edges of the satellite system and fetch relevant data from other service edges. The edges of a satellite network may also need to keep paths towards other edges, their cost, and availability information which is provided by the IP router. In such context, satellite edges may be able to support service discovery, a service announcement and a service request functionality for service consumers and provider.

Such a service-based edge-to-edge architecture may boost the role of satellite operators in a future integrated terrestrial/NTN network.

8.5.8 Transport protocols over time varying networks

Satellite networks continue to gain importance as a viable option for communication over large distances and to connect places that are not reachable by cellular and wired networks. However, these networks pose unique challenges, including, but not limited to, low bandwidth transmission, challenging physics of orbital movement, and power restriction for on-satellite computations. With Starlink, a satellite network solution is accessible by the general public, leading to increasing adoption worldwide. One way to combat these problems is to optimize transmission protocols for this purpose. In the area of the Transmission Control Protocol (TCP), one latency-improving approach is to apply Performance-Enhanching Proxies (PEPs) to satellite networks.

Another approach is to forgo TCP altogether. An alternative to TCP, QUIC, promises improvements in areas in which satellite networks are challenged, notably latency and low bandwidth. The protocol is based on UDP and compared to other commonly used protocols relatively young, with its first Internet Standard released in 2021 . QUIC is already used for terrestrial networks and deployed by Internet giants such as Facebook and Google . According to Cloudflare, HTTP/3, the HTTP revision based on QUIC, holds a share of around 28% of HTTP requests . The advantages of QUIC for satellite networks usage are actively being researched. Some research is done in the area of performance of QUIC over proxies, and other research looks into the benefits of QUIC over other transport protocols in satellite network application .

While QUIC on satellite networks was already investigated, adaptions to satellite networks that do not sacrifice some desirable properties of QUIC (like security) are scarce. Hence there is the need to further investigate methods to improve QUIC regarding its usage in satellite networks while keeping these properties and analyze the performance of these adaptions regarding common network metrics. Such improvements on QUIC performance may take advantage of service discovery mechanisms, such as in-band service resolution schemes, that may be implemented at the edges of the satellite network to streamline the connection process.

An alternative to TCP and QUIC may pass by developing an edge-to-edge transport layer adapted to the properties of the satellite system, namely being able to keep data exchange in the presence of a dynamic network topology, while achieving transmission reliability and avoiding backlog at intermediary devices. Such transport framework may be devised by leveraging the concept of information-centric networking. The receiver-driven retransmission of the information-centric networking paradigm provides end-to-end reliability, while the retransmission on each satellite minimizes the delay and bandwidth consumption of data recovery. Additional Information-centric networking features that may be leveraged to improve the performance of satellite networks may be mobility support, in-network caching, content-aware traffic management and congestion control.

8.5.9 Networking operations over FSO links

As highlighted in section 8.4.6, operating FSO links is particularly prone to link disruptions especially in the presence of clouds, eventually leading to signal blockage. This shortcoming has an important impact on the overall design of future 6G systems targeting a high level of resiliency and robustness against network faults. In the specific case of satellite constellations, the sudden unavailability of FSO-based links connecting space assets with ground counterparts has important implications on both layers of the network in that 1) a reference satellite gateway may not be able to reach a given satellite and therefore re-routing operations on ground should be carried out to reach an alternate satellite gateway, and 2) a given satellite may not be able to reach a reference gateways and therefore re-routing operations in space should be carried out over the available inter-satellite links. Either way, the support of efficient routing operations on ground and space and an overarching network management framework to monitor link availability and steer consequently the selection of path is of primary importance. This essentially give rise to unprecedented need for integrating time/space diversity concepts for networking operations, which can be partly offered by the existing DTN and ICN network architecture, in that more automatized data forwarding and network control operations must be put in place. Further to this, the exploitation of multi-path data distribution concepts (i.e. based on MPTCP or MPQUIC) is considered attractive to allow for a better exploitation of the available network paths, especially if offered with a reasonable degree of space and time diversity. All in all, the resulting picture is such to expose important challenges with respect to service continuity and network resilience requirements that the existing networking frameworks are not fully able to fulfil. This technology gap is therefore the driver for additional initiatives and research investigation towards a more effective inclusion of exploitation of FSO links in future 6G networks.

8.5.10 Research Challenges

Research Theme	**Time varying networks**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Network orchestration and management	Mid-term	Deployment of multi-tier hierarchic network orchestration frameworks able to coordinate operations in a cross-domain manner, when applied to architecture where both terrestrial and NTN operators are interacting.	Specific development of business model for the interaction and coexistence of MNO and SNO and related mapping of network operations to the orchestration level. Development of consequent cross-domain multi-level network orchestrators.
IP-forwarding payload and routing in space	Short-term	Definition of a NTN payload architecture supporting both 5G/6G operations as well routing in space by means of dedicated UPF and exploitation (where appropriate) of Xn interfaces. Exploitation of novel networking concepts based on ICN/DTN principles to achieved semantic routing in space.	Specific development of advanced payload architecture aligned to 5G/6G architecture functional splitting concepts. Implementation of DTN/ICN capable NTN nodes for enabling semantic data forwarding.
Support for deterministic communications	Mid-term	Management plane, self-healing, capable of addressing the needs of e2edetnet communications, where Ethernet/TSN and Wi-Fi/TSN regions	Contributions towards an interoperable definition of wireless APIs (L2/L3) interconnecting DetNet and Wi-Fi/TSN regions

		are involved and managed by a 6G core	Contributions towards the definition of a self-healing control plane capable of addressing the strict needs of deterministic communications
Mobility management	Short-term	Definition of effective mobility detection and implementation of corresponding network switching operations, followed by user or gateway handover.	Contribution towards the development of mobility management in IP-based networks, by taking advantage of mobility mitigation and detection concepts developed as dedicated micro-services at orchestration level.
Transport protocols	Short-term	Definition of best-practices with respect to the deployment of novel congestion control algorithms and related applicability in new Internet protocols, such as QUIC. More effective support for multi/dual connectivity, by means of multi-path protocols.	Contribution towards the definition and implementation of QUIC-based architecture allowing for more flexible data operations in space.

8.5.11 Recommendations for Actions

Research Theme	**Time Varying networks**		
Action	Routing in space	Service-based networking	Network orchestration
International Calls	X Shared experience to achieve best practice for routing in space, under different system configuration (single vs. multi-tier, LEO/MEO, etc.)	X Extension of the current data networking models to account for the Internet evolution and allow a neater separation between data forwarding and service instantiation across terrestrial and NTN assets	X Achieving cooperation at international level to gather the main lessons learnt about single-domain orchestrators and extend them to the case of 3D networks.
International Research	X Synergy across Europe to reach a common understanding of the best solutions and promote the most promising solutions at standardization level	X Additional investigations and studies aimed at developing new network architectures for seamless integration between TN and NTN at service level.	X Additional investigation towards the design and development of cross-domain orchestrators in the presence of dense 3D networks.
Open Data	-	X Dataset available from service generation	X Datasets available from operators and service providers with respect to traffic records, services/applications statistics…
Large Trials	-	X Through existing satellite platforms, aimed mostly at achieving convergence between NTN ground segments and terrestrial infrastructure. Such demonstrations should be then extended in the long term to real satellite constellations to complete the picture.	X Demonstration of the effectiveness of cross-domain orchestrators in real setups composed of TN and NTN segments.
Cross-domain research	X AI application	X AI application	X AI application

8.5.12 Expected Impact

8.5.12.1 Key Performance Indicators (KPI)

Key Performance Indicators (KPIs) for Edge Computing for data networking in Non-Terrestrial Networks (NTNs) are crucial and pretty much aligned with those outlined for the general architecture, given in Section 8.3.7.1. In particular, the following KPIs are worth being considered:

- Latency: Overall service delay, from the time a service instance is created until the time the service data are distributed to meet the users' demands.
- Energy Efficiency: Energy consumption is a critical factor in NTNs, especially for systems involving satellites. As such data networking must be sustainable also in relation to the possible limitation imposed inherently by NTN nodes.
- Bandwidth usage: overall throughput and capacity offered by the network nodes.
- Reliability: In NTNs, the reliability is a key factor to ensure efficiency of the services in terms of data distribution.

8.5.12.2 Key Value Indicators (KVI)

Key Value Indicators (KVIs) for data networking in Non-Terrestrial Networks (NTNs) are pretty much based on those already defined for the general architecture in section 8.3.7.2 and hence not further repeated here.

8.6 Edge Computing

8.6.1 Scenario

The next generation of mobile networks is expected to be developed under the umbrella term 6G to enable a fully connected, digitized, and intelligent society. The 6G vision is set to serve the advanced version of current 5G applications and some novel scenarios such as holographic communication, Immersive XR, etc. Among others, 6G technology is also expected to revolutionize the traditional services through the support of edge computing.

Computer and communication technologies are the main examples of emerging technologies that reflect how society evolves and integrates such tools into its social structure, and their influence on institutional development and social progress. We recognize that some realities do not match our ideals of life, culture and gender equality, asking how computer platforms affect the equality of opportunities in our society. Consequently, access and computer equality is a very important mission for research.

Cloud computing platforms require enormous investments in their installation, requiring considerable land consumption and energy resources to operate them. Edge platforms, content delivery networks, and other distributed services also depend on the computer infrastructures of the sector's telecommunications service providers and small operators, highlighting the need to already have a market in the region. This is one of the main obstacles to the global deployment of these technologies, which only increases the gap in countries with lower digitalization who cannot overcome digital divides.

The recent European war scenarios have highlighted the long-forgotten realities of the lack of bombed communication infrastructure, relief services and information for the population. One of the answers to these shortcomings was the generously offered Starlink satellite infrastructure to reactivate information channels. However, because these scenarios are well-known in other parts of the world, because of the digital isolation in which they live, they do not attract media attention in the world. These tragic examples contribute to understanding the strategic importance of NTN platforms and how their technology has been limited to improving bandwidth, massive access, life expectancy and equipment miniatures until now. Moreover, it is necessary to take steps and shift paradigms to bring resources into orbit to implement emerging computing services such as artificial intelligence, high-performance computing, storage, content delivery, communication and networking.

In fact, the above-mentioned technologies have been developed mainly for terrestrial and fixed infrastructure, and are unlikely to be re-used in highly dynamic and completely wireless NTN contexts: the stability of the hardware installed on the orbit is very different from that of the Earth, as well as the problems of energy supply and heat dissipation. The terrestrial clouds and NFV systems are designed to ensure 99.99% system reliability by implementing redundant mechanisms to compensate for the possible interruption of services in basic static systems and to perform periodic service and resource migrations in the appropriate way. On the other hand, the system based on orbital computer infrastructure is completely different, much more complex, and dynamic. Only low-orbit satellite visibility time and the need to ensure that all services assigned within the maximum service time provided by the requirements are accessible through inter- and intra-satellite links can be considered, even after visibility time. However, this incredible technological challenge is a concrete response to the need to reduce digital divisions and to provide opportunities for all countries to respond appropriately to challenges we are asked to respond to, such as the recent Covid epidemic. This could undoubtedly lead to digital democracy.

Recently with the advance of various new platforms, non-terrestrial networks (NTN) have acquired a central place in the 6G research. Both aerial and orbital platforms including low altitude platforms (LAPs), high altitude platforms (HAPs), and different satellite constellations can play an important role in creating sustainable and intelligent systems through their added coverage and capacity boosts. NTN platforms can enable edge computing facilities in space through the integration of computation and communication resources onboard. Integrating such edge computing facilities in the traditional EC systems can effectively boost performance. It can tackle several challenges of EC systems including resources, limitations, and security threats mainly due to terrestrial nodes fixed positions, their vulnerability to natural disasters, coverage limitations, etc. Various NTN platforms located at different altitudes in space with differing mobility patterns can serve ground-based vehicular users effectively. With the rapid deployments of 5G systems, initial 5G performances are available. With this, the beyond 5G specifications being a steppingstone into the 6G world are being defined. The 6G vision aims to enable an intelligent society through several new intelligent services and applications.

The EC-integrated NTN platforms, i.e., orbital edge computing (OEC), can enable computing in space, thus facilitating ground/space users with high-quality services in space. OEC-enabled LEO satellite networks can be extremely useful to serve ground users located in remote areas with limited or zero connectivity to terrestrial networks. In recent times, several new LEO constellations such as Starlink, OneWeb, Telesat, etc, tailored to specific missions have populated the space. These constellations can be located at different altitudes, having varying satellite densities, interplane and intra-plane satellite distances, speeds, etc. Thus, ground users can have access to multiple LEO satellites belonging to different constellations for OEC resources. Additionally, satellites can also act as relay nodes to route user requests to other nearby satellites or terrestrial cloud computing facilities. Given the size restrictions and limited coverage, satellites can have reduced OEC resources and thus can serve a limited number of users only. Additionally, with the limited storage resources, each LEO node can be able to store a limited number of services. In a multi-service scenario, users can demand different services based on their specific requirements. This opens a new challenge of proper user-server assignments based upon the users' demands and the availability of the resources at servers in the space with multi-tier edge computing facilities.

By summarizing, the main promising scenarios for implementing edge computing technologies in Non-Terrestrial Networks (NTNs) can be individuated as:

- Beyond 5G (B5G) and 6G Space Infrastructure: The integration of edge computing technologies into the B5G and 6G space infrastructure is one of the most promising scenarios. This involves developing architectures featuring edge computing capabilities aboard satellites to flexibly and dynamically allocate storage and processing to assigned tasks. These systems are set to serve both direct access and transport (backhaul) connectivity, enhancing performance, and user equipment simplification.
- Internet of Things (IoT): The proliferation of IoT devices and the foreseeable bandwidth demand caused by the new 5G networks devices and applications demand alternative solutions and architectures. Edge computing can bring computation and storage resources closer to the end-users and devices, minimizing the need for long-distance communications between edge clients and servers. This results in a reduction in latency and bandwidth usage, a general improvement in connectivity, a more efficient network operation and service delivery, better scalability, faster access to external computation and storage.
- Software Defined Networking (SDN) and Network Function Virtualization (NFV): Low-cost and energy-efficient equipment, along with SDN and NFV approaches, is a promising solution to overcome such an issue, accounting for the joint use of terrestrial, aerial, and spatial communication networks in hierarchical network architectures.

These scenarios are driving the development of a technical and deployment roadmap for edge computing enabled B5G and 6G space infrastructure.

8.6.2 Motivations

Novel services are characterized by an increased amount of processing for being deployed. Despite the growth of processing capacity, also the at user side, it remains relatively low, especially when considering latency-critical, data-intensive applications and AI-based services. The traditional solution has been to resort to cloud computing facilities allowing to reduce the computing burden of new services by enabling users to transfer parts or complete tasks to cloud servers at the cost of considerable computational and communication power. Cloud infrastructures are generally located far from users, on the ground or on the core network, introducing some disadvantages, such as significant transmission costs, traffic congestion, and threats to data security due to long-distance communications. Such problems can be solved by integrating Edge Computing (EC) devices into the network, so that computing resources are closer to end users.

In the past, several initiatives considered the deployment of EC servers on the terrestrial network (TN). Such an approach has achieved great success in enabling new latency-critical services, especially in smart cities and mobility scenarios. However, the limited capacity and coverage of the TN can result in a bottleneck while exploiting the advantages of the EC. EC facilities relying on TNs are becoming increasingly used, as many new users are seeking services with specific requirements. The limited coverage of rural and remote areas, the insufficient service of natural disasters such as tsunamis and earthquakes, new security challenges, poor link budgets and additional interference are among the main challenges to be taken into account when using the TN-based EC platform. The dynamic changes in the resources of the EC server add additional challenges when integrating TN-based EC services due to the presence of different users. With this limitation, the integration of the TN-based EC platform into the TN itself cannot be sufficient to meet new services and applications that will require more stringent requirements in terms of latency and computational resources.

Encouraged by the additional interest shown by new technological developments and several technology giants (such as Facebook and Google), non-terrestrial networks (NTNs) including space and air networks are increasing, mainly to provide global connectivity. New platforms such as satellite constellations, unmanned aerial vehicles (UAVs), small-scale fuel aircraft and balloons are deployed at different heights from ground users to address global connectivity challenges. Improved connectivity, scalability and reliability are some of the advantages of an NTN-based communication platform. The addition of modern communication technologies such as multi-band antennas allows NTN platforms to provide EC-based services for computers on board. Such NTN-based EC platforms can complement TNs to solve various problems, including limited capacity and coverage. However, the higher transmission delay introduced by space network platforms (satellite constellations) is an important challenge considering the NTN platforms used to service crucial latency applications. New low- and high-altitude platforms (LAPs and HAPs) have considerable advantages in terms of reduction in transmission distances, short deployment time and cost, and loss of communications channels compared to space networks.

In summary, the main motivations for developing edge computing technologies in Non-Terrestrial Networks (NTNs) can be summarized in:

- Performance Enhancement: Edge computing in NTNs is highly promising in terms of performance enhancement. It can lower latency, reduce the impact on backhaul, and provide flexibility. This is achieved by moving computation and storage resources closer to the end-users, minimizing the need for long-distance communications between edge clients and servers.
- Infrastructure Optimization: Edge computing can optimize the network and its applications by dynamically allocating storage and processing to assigned tasks. This results in a more efficient network operation and service delivery, better scalability, and faster access to external computation and storage.
- Improved Connectivity: Edge computing satellite systems can serve both direct access and transport (backhaul) connectivity. This can enhance the overall connectivity of the network.
- Security and Privacy: Edge computing can also improve network support for security and privacy. By moving data production closer to data usage, it is easier to achieve higher security, data privacy, and reliability.
- Support for 5G and Beyond: The integration of NTN has emerged as a transformative force in the era of 5G and beyond. 5G technology demands enhanced coverage, lower latency, and increased capacity, all of which can be achieved by leveraging the capabilities of satellite constellations and other non-ground-based technologies.
- Support for IoT Applications: In IoT applications, effective wireless connectivity is not only requisite for the data transmission between the multiple nodes but also represents a key factor to ensure the safety of personnel or citizens in remote locations.

These motivations are driving the development of a technical and deployment roadmap for edge computing enabled Beyond 5G (B5G) and 6G space infrastructure.

8.6.3 Architecture/System

TN and NTN enable EC-based services by integrating edge computing servers and distributed infrastructures. Therefore, multi-layered joint T-NTN consists of different EC platforms, which can be used to provide the user with the requested heterogeneous service. A typical service area is composed of several users who require latency and data intensive services. The network architecture enabled by EC consists of several elements: ground users, the small cells BSs, macro-cell BSs, LAPs, HAPs, LEOs, MEOs and GEOs. Each layer has a specific characteristic in terms of coverage, availability, processing power, communication speed, and provides a heterogeneous system where multiple layers complement each other. Each layer of the EC network architecture can adapt different computing and communication strategies. Several virtualization techniques (VM, containers) can be used to efficiently use EC resources. In addition, SDN-based centralized control methods can be used to manage computing and storage resources on individual EC platforms. Multi-operator-based communication technologies can be adapted to allow communication between EC nodes of the same and different layers.

The reference architecture is composed by a set of multi-layer satellite constellations able to provide edge computing services to the ground-based user set. Each constellation includes a set of satellites distributed in the multiple planes for providing global coverage. A possible multi-tier joint Terrestrial and Non-terrestrial Networking (T/NTN) scenario may be considered with one user layer, multiple satellite constellations, and cloud computing facilities. Several services may be distributed over the edge/cloud facilities based on their storage capabilities. The cloud computing facilities, given their superior nature can provide all services, while the satellites with size limitations are able to provide a subset of services.

The architectural elements for implementing an edge computing based Non-Terrestrial Network (NTN) include:

- Edge Servers: These are the primary computational units in the edge computing architecture. They are responsible for processing and storing data, running applications, and providing services to end-users. Can be located on NTN elements, such as LEO and GEO satellites, as well on terrestrial nodes, depending on the considered application scenario.
- Gateways: Gateways act as the bridge between the edge servers and the end devices. They are responsible for data aggregation, protocol translation, and network connectivity.
- Internet of Things (IoT) Devices: These are the end devices in the edge computing architecture. They generate data that is processed by the edge servers.
- Satellites: In the context of NTNs, satellites play a crucial role. They can host edge computing capabilities to flexibly and dynamically allocate storage and processing to assigned tasks. They serve both direct access and transport (backhaul) connectivity and enhance performance.
- Cloud Infrastructure: This can be a public or private cloud, which can be a repository for the container-based workloads like applications and machine learning models. The cloud can also be a source and destination for any data that is required by the other nodes.

These elements work together to form a distributed and heterogeneous computing environment, enabling edge-computing-capable space infrastructure.

8.6.4 Management

8.6.4.1 Edge Computing resource orchestration and allocation (AIaaS/SaaS/DaaS, etc.)

In 6G networks Edge and Ubiquitous Computing will play a fundamental role due to the transition of the telecommunications architecture towards distributed platforms based on (micro)services that will have to provide computing resources with “zero delay”. This will impose local or proximity processing that cannot be guarantee without encouraging NTN solutions, as well as the service composition models such as SaaS (Software as a Service), FaaS (Function as a Service), AIaaS, etc., and new computing technologies of virtualization such as containers, Unikernels and relative supervisors.

8.6.4.2 NTN management without intervention.

Zero Touch Networks (ZTN) is a term used to indicate those networks that can heal and tune themselves, based on the data signals they collect and analyse across network activity. Zero.

8.6.4.3 Context-sensitive NTN overlays for data sharing.

A key advantage of low-orbit NTN is the limited amount of power requested from ground-based transmitters and receivers. This feature opens up interesting scenarios that allow connectivity to be provided even to small devices in areas that would not otherwise be possible to cover using terrestrial infrastructure. Once the packets arrive at the constellation, the satellites undertake to send the data to Earth. One could imagine the need and opportunity to create data aggregation spanning different satellites, depending on the peculiar characteristics of the satellites, of the users of such data.

A constellation could provide specific mechanisms and solutions to NTN application providers to manage data exchange automatically and dynamically between satellites in an aware context (pertaining to the data itself or also to the metadata).

8.6.5 Application

8.6.5.1 Distributed AI

Recently, Machine Learning (ML) techniques, especially those belonging to the Distributed Learning (DL) class, have gained huge popularity in dynamic wireless scenarios with their added advantages in terms of learning efficiency, reliability, and data security. Various DL methods, such as Federated Learning (FL), Multi-Agent Learning, and Collaborative Learning, are considered in dynamic domains. Additionally, various ML tools and techniques have been considered to form suitable DL methods, such as multi-agent FL, DL with model split, DL with meta-Learning, and DL with swarm learning. In this way, a rich ecosystem of DL methods with specific characteristics, performance, and demand is formed and made available to serve users. From a networking point of view, several new advances have recently been introduced, especially with the innovations of 5G and B5G technologies. Different computing paradigms, such as Edge/Cloud Computing, have been introduced to implement new services and applications with better performance. Technologies, such as network softwarization through Network Function Virtualization (NFV), Software Defined Networking (SDN), and Network Slicing (NS), have revolutionized the networking process and opened the doors to a multitude of applications and services with different demands and additional flexibility. Furthermore, distributed computing and communication technologies, such as the edge-to-cloud continuum, and joint Terrestrial and Non-Terrestrial Networks (T/NTN), have gained huge popularity in terms of capacity, coverage, and reliability for serving end users.

Different networking nodes, such as Terrestrial Base Stations, Low-Altitude Platforms (LAPs), High-Altitude Platforms (HAPs), and Satellite nodes, can be considered as distributed around the service area. Different DL methods can be considered for coping with the requirements of heterogeneous users. A typical EC scenario includes a massive amount of computation, communication, and storage resources distributed over the ground, air, and space networks, these resources can be utilized to create an intelligent network through proper deployment of required DL methods, where each network device is able to host the virtual functions enabling the different DL execution.

FL is a DL paradigm for collaborative model training without sharing the individual element's data. Devices can train local models and update a central server that then aggregates and applies updates to the shared model. However, the drawback of FL lies in the requirement for each client to train the entire resource-intensive ML model. This is particularly true for Deep Neural Networks (DNNs) deployed on end devices with limited resources. Although FL is a privacy-preserving training approach, recently, new concerns have emerged, mainly due to the iterative transmission of local and global model parameters, leading to problems such as poisoning, attacks, and model inversions. Several of these issues can be related to the possible requirement of transmitting a complete model in traditional FL environments. Split Learning (SL) is another DL method that can enable efficient distributed ML model training on resource-constrained devices. With SL, the ML model is split into two parts, where individual parts can be trained on server and client devices. Unlike FL, with SL, only a portion of the model is trained on client nodes, effectively reducing the processing and communication load on resource-limited devices. The communication process is limited to cut-layer activation, ensuring model privacy. With these advantages, SL can be useful in reducing overall training costs and privacy concerns in federated environments and can even enable FL frameworks to train more advanced DNN models. On the other hand, Transfer Learning (TL) solutions from the meta-learning family can increase the efficiency of ML model training by knowledge transfer from previous source tasks to new target tasks. TL accelerates convergence, reduces data needs, and improves ML robustness in diverse vehicular contexts. Therefore, TL can further complement the FL process enabled by SL to further improve the distributed training process. From a networking perspective, Non-Terrestrial Networks (NTNs) have acquired a central position in the 6G vision, mainly to provide global coverage and capacity boosts for traditional terrestrial networks. Different NTN platforms can help Vehicular Users (VUs) to enable flexible intelligent solutions. In particular, High-Altitude Platforms (HAPs), with their reduced transmission distances, higher coverage, and easy and flexible deployments, can complement terrestrial settings to enable efficient DL solutions.

8.6.5.2 TN/NTN Integration

Edge computing technologies are revolutionizing the way we approach network infrastructure, particularly in the context of integrated terrestrial and non-terrestrial networks. This integration is paving the way for a new era of connectivity, characterized by enhanced performance, improved reliability, and increased capacity. In an integrated terrestrial and non-terrestrial network scenario, edge computing technologies are applied to bring computation and storage resources closer to the end-users. This minimizes the need for long-distance communications between edge clients and servers, resulting in reduced latency and bandwidth usage. It also allows for more efficient network operation and service delivery, better scalability, and faster access to external computation and storage. One of the key applications of edge computing in this scenario is in the support for Internet of Things (IoT) devices. The proliferation of IoT devices and the foreseeable bandwidth demand caused by new 5G networks devices and applications demand alternative solutions and architectures. Edge computing can meet these demands by providing low-latency, high-bandwidth connections to these devices, enabling real-time data processing and decision-making capabilities. Furthermore, the integration of non-terrestrial networks, such as satellite constellations, with terrestrial networks can enhance coverage, particularly in remote or underserved areas. Edge computing technologies can be deployed on these satellites, providing computational capabilities in space and reducing the need for data to be transmitted back to Earth for processing. This can significantly reduce latency and improve the performance of applications running on these networks.

However, implementing edge computing in integrated terrestrial and non-terrestrial networks is not without its challenges. These include dealing with intermittent connectivity, latency and coverage constraints, bandwidth limitations, and the need for robust network infrastructure. Despite these challenges, the potential benefits of edge computing in integrated terrestrial and non-terrestrial networks make it a promising solution for the future of connectivity.

In conclusion, edge computing technologies hold great promise for enhancing the capabilities of integrated terrestrial and non-terrestrial networks. By bringing computation and storage closer to the end-users, they can significantly improve network performance, support IoT devices, and enhance connectivity in remote areas. As such, they are set to play a crucial role in the future of network infrastructure.

8.6.5.3 Digital Twins

Digital twins are emerging as a powerful approach for modelling and managing complex systems, including non-terrestrial networks (NTNs). A digital twin is a virtual representation of a physical system or process. It mirrors the real-world entity, capturing its behaviour, interactions, and dynamics. Digital twins can be used for various applications, such as monitoring, simulation, and decision-making. On the other side, NTNs encompass both airborne (e.g., drones, high-altitude platforms) and space-borne (e.g., satellites) elements, playing a crucial role in providing communication services to remote areas, bridging the digital divide, and ensuring global coverage.

NTN complexity arises due to the large number of network entities and users in dynamic and heterogeneous environments. Designing and managing NTNs becomes increasingly costly and challenging. Digital twins offer a solution by providing a detailed virtual representation of the entire NTN ecosystem. Key benefits include real-time monitoring, simulations, and data-driven decision-making. By creating accurate data-driven NTN models, digital twins enable rapid testing and deployment of new technologies and services within NTNs In order to be implemented several technologies empower digital twin development for NTNs: Sensors and connected devices provide real-time data for accurate modelling, AI algorithms enhance predictions and optimize network performance, Space-based Cloud Computing, leveraging on cloud resources from space.

An example involves implementing a data-driven digital twin model within an open radio access network (O-RAN) architecture for NTNs. This facilitates dynamic and service-oriented network slicing, improving efficiency and resource allocation. In summary, digital twins hold great promise for shaping the future of network control and management in the dynamic landscape of non-terrestrial communication systems.

The main arguments for Digital Twins in NTNs can be resumed in

- Real-Time Monitoring and Predictive Maintenance: Digital twins provide a virtual representation of the entire NTN ecosystem. By continuously monitoring network elements (such as satellites, drones, or high-altitude platforms), operators can detect anomalies, predict failures, and perform proactive maintenance. This enhances network reliability and reduces downtime.
- Simulation and Optimization: Digital twins allow network designers and operators to simulate various scenarios. For example, they can model network behavior under different environmental conditions, traffic loads, or hardware configurations. This simulation-driven approach enables better decision-making, efficient resource allocation, and optimized network performance.
- Data-Driven Decision-Making: Digital twins rely on real-time data from sensors, devices, and network components. By analyzing this data, operators gain insights into network behavior, user experience, and potential bottlenecks. These insights inform strategic decisions, such as capacity planning, network expansion, or service deployment.

However, some challenges may arise:

- Complexity and Scale: NTNs involve a large number of interconnected elements (e.g., satellites, ground stations, user terminals). Creating accurate digital twin models for each component can be complex and resource intensive.
- Data Integration and Interoperability: Integrating data from diverse sources (e.g., IoT sensors, satellite telemetry, weather forecasts) into a cohesive digital twin poses challenges. Ensuring interoperability between different data formats and protocols is crucial.
- Latency and Real-Time Requirements: NTN applications often require low latency (e.g., for real-time communication, remote sensing, or navigation). Digital twins must provide timely updates without introducing additional delays.
- Security and Privacy: Protecting digital twin data from cyber threats is essential. Unauthorized access to digital twin models or control systems could have serious consequences for network operations.
- Model Accuracy and Calibration: Digital twin models must accurately reflect the physical behavior of NTN components. Calibration and validation against real-world data are necessary to maintain fidelity.
- Resource Constraints: Space-borne elements (e.g., satellites) have limited computational resources. Designing lightweight digital twin models that balance accuracy with resource constraints is a challenge.

8.6.6 Research Challenges

Research Theme	Edge Computing		
Research Challenges	Timeline	Key outcomes	Contributions/Value
On-board Edge processing Services are requiring an increasing amount of edge processing. When dealing with NTN-based services, such processing should be performed onboard. *(refer to Research aspects as per 4.3.x)*	Short-term (finished in 3y)	Load Balancing Techniques AI processing on board Integrated Networking and service processing	Reduced latency Reduced energy consumption Possible deployment without Cloud Ground infrastructure
On-board VNF Deployment B5G and 6G networks requires an increasing flexibility in terms of network function softwarization and deployment. Proper VNF deployment on board should be considered	Short-term (finished in 3y)	TN/NTN integration Private Satellite based Networks Network Slicing and flexible deployment for heterogeneous services	Increased Resilience Assuring heterogenous services/requirements to users
On-board Intelligence at the edge AI and ML procedures are becoming an essential element for future networks. Despite AI and ML can be considered applications they have proper requirements. They effective deployment at the edge should be considered.	Medium-term (finished in 5y)	AI deployment at the Satellite Edge Distributed Learning deployment AI integration with TN	Implementing Space AI
Orchestration and Management Dealing with Edge computing in spaceborne equipment requires an increased management and orchestration action. ZeroConf is becoming a requirement for AI based orchestration	Medium-term (finished in 5y)	Automated services Autonomous networking AI based management	Smooth network management Increased user resilience

8.6.7 Recommendations for Actions

Research Theme	Edge Computing			
Action	On-board Edge processing	On-board VNF Deployment	On-board Intelligence at the edge	Orchestration and Management
International Calls	X Terrestrial edge processing techniques are a quite mature technology, while its use on satellite is still in an infancy state. International calls may foster their implementation though industrial -research collaborations	X Terrestrial VNF deployment is a technology ready to be deployed, while its use on satellite is still in an infancy state. International calls may foster their implementation though industrial – research collaborations.	X International calls on ML/AI deployment on board may foster their development as well helps in individuating the most promising scenarios.	X Orchestration and Management requires the integration of operators, manufacturers, as well the effort of research institutes. International Calls may strengthen this aspect
International Research			X AI/ML deployment at the edge is a research trend still with low TRL. International	X Next generation Orchestrators employs AI through the ZeroConf approach. International

			collaboration among research institute may help finding new solutions.	Research is envisaged for developing novel solutions.
Open Data			X ML/AI solutions require training data for being deployed. Proper open data should be published as a follow-up of this activity	X ZeroConf approach requires big data for proper training. Open data should be published as a follow-up of this activity
Large Trials	X On a medium-long term large trials are expected by exploiting the presence of mega constellations	X An integrated 5G/B5G/6G network may require the presence of VNF on-board in large deployments		X Orchestrator is strictly connected with VNF deployment and on-board processing management. To this aim large scale deployments should be considered for proper assessment
Cross-domain research	X Edge computing facilities are commodity for several applications. Cross-domain research may be performed where several application domains may gain from on-board processing.		X AI/ML algorithms are commodity for several applications. Cross-domain research may be performed where several application domains may gain from AI/ML on-board.	

8.6.8 Expected Impact

8.6.8.1 Key Performance Indicators (KPI)

Key Performance Indicators (KPIs) for Edge Computing technologies in Non-Terrestrial Networks (NTNs) are crucial for assessing the effectiveness and efficiency of these systems. While specific KPIs can vary depending on the use case and system requirements, here are some general KPIs that are often considered:

- Latency: One of the primary reasons for implementing edge computing in NTNs is to reduce latency. By processing data closer to the source, edge computing can significantly decrease the time it takes for data to travel, resulting in lower latency. In NTNs, this could be measured as the time delay between the data being generated by a device and the processed information being available.
- Energy Efficiency: Energy consumption is a critical factor in NTNs, especially for systems involving satellites. Edge computing can potentially reduce the energy consumption of devices, as less data needs to be transmitted over the network. Efficient energy use can prolong the operational life of these systems.
- Bandwidth usage: By processing data at the edge, less data needs to be sent over the network, which can reduce bandwidth usage.
- Reliability: In NTNs, the reliability of the edge computing infrastructure could be a key Performance Indicator. This could be measured as the uptime of the edge servers or the number of successful data processing tasks.
- Cost: The cost of implementing and maintaining the edge computing infrastructure in NTNs could also be a key indicator.

8.6.8.2 Key Value Indicators (KVI)

Key Value Indicators (KVIs) for Edge Computing technologies in Non-Terrestrial Networks (NTNs) are not explicitly defined in the literature. However, based on the general principles of edge computing and NTNs, we can infer some potential KVIs:

- Flexibility: The ability to flexibly and dynamically allocate storage and processing to assigned tasks is a key feature of edge computing systems in NTNs.
- User Equipment Simplification: Edge computing can help simplify user equipment by offloading some of the processing tasks to the edge nodes.
- Connectivity: In the context of NTNs, providing reliable connectivity in remote areas where terrestrial networks are not available is a key performance indicator.
- Service Level Specifications: These are detailed descriptions of the service level expectations and are often tied to specific use case segments
- Scalability: The ability of the system to handle an increasing amount of work by adding resources to the system.

8.7 Security Considerations

8.7.1 Motivation

The present section deals with security from a holistic point view. That means, by considering security from the eavesdropping the transmitted information to jamming attacks of GNSS signals, key in the satellite communications and their corresponding services (e.g. autonomous car, vehicle to vehicle communications, etc.). Furthermore, it is prospected the strategies of QKD (Section 8.7.2), Federated Block Chain (Section 8.7.3), antijamming techniques (Section 8.7.5), and end-to-end security (Section 8.7.4) over Non-Terrestrial Networks (NTN) to provide a global view of the security. In all sections is introduced the topic, the relevance of the topic to the NTN, and a short description of the challenges that the technologies that provide security has to face at short (<3 years), medium (<5years) and large term (<7 years).

From the QKD over NTN, Satellite-based QKD has demonstrated successful long-distance secure key distribution, overcoming the limitations of terrestrial fiber networks. However, still it has to face challenges on photodetectors efficiency, robust authentication, increasing the key generation rate, and integrating optical fiber with free-space systems over NTN to name a few of them.

Regarding blockchain strategy, it is used to enhance the data privacy and network security by tampering-proof ledger across a large network of devices, which avoids the risks of centralized solutions. By resorting to federated learning blockchain, miner learning rates are increased and network latency is reduced when a pre-selected group of nodes is used for consensus. Challenges in implementing it in NTN include defining a common architecture, addressing connectivity and long delays, and managing high computational costs in resource-limited space environments.

The QKD and federated block chain refers to data security, e.g. unveiling the information to non-legitimate users. However, it is also interesting to protect GNSS signals from interference. GNSS-based positioning, however, can be compromised by interference signals, necessitating their identification, classification, and localization to mitigate them. Toward this regard, Machine learning (ML) approaches have shown promise in interference monitoring, but challenges including diverse jammer types, environmental changes, and the integration of GNSS with 6G signals to enhance positioning accuracy and reliability need to be faced.

From the End-to-end security integrated in TN/NTN networks involves the combination in an efficient way of multiple technologies. By doing so, it is possible to address the broader attack surface of satellite communications. Future research aims to integrate security schemes across multiple layers and satellite constellations, ensuring compliance with privacy rules and achieving high data rates for 6G networks and beyond.

Finally, after providing a summary of the main arguments and challenges of the QKD, federated block chain, GNSS signal robustness and end-to-end security over Non-Terrestrial Networks the following considerations are provided: i) related/threated security assets & potential resulting security needs (Section 0), ii) Proposed security mechanisms (Section 8.7.7), and iii) Security Management mechanisms and related requirements (Section 8.7.8).

8.7.2 QKD on Free-Space NT Networks

Satellite-based QKD has made notable progress, with several successful experiments demonstrating long-distance secure key distribution. The use of satellites, such as the Micius satellite, has enabled QKD over distances of thousands of kilometers, overcoming the limitations of terrestrial fiber networks. These systems typically involve: i) ground stations sending/receiving photons to/from a satellite and ii) photon transmission/reception between satellites through optical inter-satellite links. Thus, satellites act as relays to facilitate key distribution between distant nodes of the network. In the first case, the transmission of the keys suffers from the impairments of the free-space channel, whereas in the second case, the main issues arise from errors in the tracking system. However, these issues are inherent in the optical nature of QKD transmissions. Additionally, the quantum nature of QKD transmissions introduces more impairments: i) the quantum efficiency of the detectors, ii) validity of the protocols, and iii) eavesdropping detection capability. To address the first issue, investigations on single-photon detection with low levels of shot noise and high quantum efficiencies are an active area of research [C8-45], [C8-46]. Regarding the validity of the protocols, several issues still require to be investigated in depth: i) the authentication process, ii) vulnerabilities to detector side-channel attacks, and iii) latencies in the key generation rate. Note that before the transmission procedure for generating the quantum key between Alice and Bob, it is required that both be authenticated [C8-47], [C8-48]. The detector may be another source of vulnerability in QKD implementations. To overcome this vulnerability, the implementations of QKD in terrestrial free-space environments have resorted to the use of the MDI-QKD strategy [C8-49]. To detect eavesdropping in free-space optics, machine learning techniques have been proposed to distinguish channel fades from the presence of eavesdroppers [C8-50]. Finally, both terrestrial and satellite networks are affected by the evolution of quantum computers. Therefore, the integration of QKD in both networks separately and jointly is an active area of investigation. In this regard, the integration of optical fiber and free-space-based systems leverages the strengths of both mediums to create robust QKD networks [C8-51].

The main arguments to support research activities on QKD on Free Space Networks are:

- The evolution of quantum computers will revolutionize security, as classical secure communications will no longer guarantee perfect protection against quantum computer attacks.
- Post-quantum solutions are designed to withstand attacks from quantum computers using classical cryptography. However, while theoretically secure, these algorithms can present vulnerabilities when implemented, which attackers can exploit. Moreover, future advancements in quantum computers may eventually compromise post-quantum solutions.
- Communication security is essential to support the services for which communication networks are designed. If communication security is not maintained, services, especially those involving economic transactions, cannot be implemented, which undermines the future of telecom networks.

As a summary of the main challenges of the QKD on Free Space NT Networks:

- Increasing the quantum efficiency and sensitivity of photodetectors used in quantum communications is a significant challenge due to the very low signal power transmitted by QKD systems, which is comparable to noise levels. (Short-term Challenge 1)
- QKD protocols assume that the transmitter and receiver are legitimate. Ensuring this requires a robust authentication process that is resistant to quantum computer attacks, which is a major challenge for QKD systems. (Short-term Challenge 2)
- QKD protocols need further investigation and implementation for free space communication networks, particularly regarding key distribution using repeaters. (Medium-term Challenge 3)
- The key generation rate of QKD systems is very low, posing a challenge for satellite communications due to high propagation delays. Therefore, strategies to increase the key generation rate are essential in quantum satellite networks. (Medium-term Challenge 4)
- Integrating optical networks involves combining optical fiber-based networks with free space ones. An interface that integrates the quantum systems of both networks needs to be developed to minimize key generation rate loss, enabling an end-to-end full optical network. (Long-term Challenge 5)
- Efficient time and frequency multiplexing of data and key transmissions in free-space optical satellite communications is necessary, considering the different power balances of the two waveforms. (Long-term Challenge 6)

7.1.1.1 Research Challenges

Research Theme	**QKD on Free Space NT Networks**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 Enhancing quantum efficiency and sensitivity in photodetectors	Short-term (finished in 3y)	Key performance metrics include quantum efficiency, sensitivity, signal-to-noise ratio, detection threshold, error rates, and data throughput of photodetectors. These reflect enhanced performance and reliability in quantum communications.	The research explores novel materials, designs, and fabrication techniques for photodetectors to improve quantum efficiency and sensitivity. It also focuses on reducing noise through advanced algorithms, shielding, and using photon-number-resolving detectors and entanglement-based protocols to enhance signal detectability.
Challenge 2 Authentication process of QKD communications	Short-term (finished in 3y)	The key performance metrics in this challenge are: Reduction/Null the Eavesdropping rate of the messages from the legitimate transmitter (Alice). Reduction/Null the probability that Bob accepts information from a non-legitimate transmitter (Eve).	The contribution on this area falls in the integration of QKD with post-quantum-based solutions, developing keys that not only be based on multiple parameters of the physical layer such as the beam.
The protocols of QKD using repeaters to support QKD in satellite networks	Mid-term (finished in 5y)	Key performance metrics include developing and refining QKD protocols for free-space communication, designing efficient repeater architectures, and quantifying key distribution efficiency. The research also focuses on rigorous security analyses and experimentally validating protocols and architectures in real-world scenarios.	Measure the secure key generation rate, maximum distribution distance, error rates (BER and QBER), and security level (including secret key rate and resistance to attacks) in the free-space QKD network to evaluate the efficiency, range, accuracy, reliability, and security of the QKD protocol.
Challenge 4	Mid-term (finished in 5y)	The key performance metrics in this challenge are: Increase the key generation rate of the QKD	Contributions on the area of improving the equalization of the quantum channel, the synchronization

The key generation rate of QKD systems in free space NT networks is very low.		communication systems, reduce the errors in the QKD system by increasing the sensibility of the photodetectors, sensibility to excess noise	procedures and integration with quantum sensing and post-quantum-based solutions may help to speed up the QKD generation rate
Integrating security schemes for free space and optical fiber networks	Long-term (finished in 7y+)	Key performance metrics include designing and implementing interfaces for seamless integration between optical fiber-based and free-space quantum networks. Strategies to minimize loss and optimize performance by mitigating loss, noise, and interference are also crucial.	Integrating different network types requires addressing signal loss from absorption, scattering, and atmospheric effects, minimizing classical and quantum noise, maintaining secure communication channels, reducing latency for real-time applications, and ensuring scalability to meet growing data transmission demands.
Challenge 6 Multiplexing FSO and QKD signals in the same band	Long-term (finished in 7y+)	The key performance metrics in this challenge are: Distance in wavelengths between the channels of FSO and QKD. Power balance between the FSO and QKD systems. Selectivity of the optical filters. Spatial diversity is used to enhance the FSO-QKD transmissions. Error rate in the detection of the FSO and QKD frames.	Strategies for improving the sensibility of the quantum receivers such as the one based on quantum sensing, signal processing for separating both signals. Optical filters with very tight separation, multiplexing techniques with a high selectivity, the use of multiple apertures in the telescopes will be of interest.

8.7.2.1 Recommended Actions:

Research Theme	**QKD on Free Space NT Networks**		
Action	Vulnerabilities of QKD protocols	Enhancing the Key Generation Rate of QKD protocols	Extending QKD to NTN networks
International Calls	X Quantum communications is known to provide perfect security. However, it still demands to relax some constrains as authentication to be a truly secure system	X The first generation of quantum systems, such as CV-QKD and DV-QKD permit to provide perfect secure communications. However, the key generation rate is very low, especially in NTN. So, investigations in this area are still needed.	X QKD systems are devised for developing point-to-point communications. In 6G NTN the satellites may be regenerative. So, the distribution of keys through the NTN should be investigated.
International Research	X Collaborative research on the area of vulnerabilities for quantum system is a must to achieve a consolidated standard on quantum communications over NTN	X Collaborative research on enhancing the key generation rate of QKD protocols is a real need to achieve a consolidated standard on quantum communications over NTN that may be used in 6G or future 3GPP networks	X Collaborative research on QKD systems for satellite networks is a real need since in 6G the first generations of satellite constellations will be deployed for communications.
Open Data	X Open data from the collaborative research to determine the vulnerabilities	X Open data of the quantum and FSO channel is a demand to speed up future ML/AI systems	

	of the QKD protocols and to consolidate and standard on QKD for truly secure non-terrestrial networks	that run on top of the QKD systems	
Large Trials		X Develop large trials to experiment with different strategies for improving the key generation rate, i.e. overcoming the impairments of the FSO/Quantum channel is required. To validate future QKD for NTN in standards	X Develop large trials over NTN networks to experiment with different strategies for transmitting the keys.
Cross-domain research	X Cross-domain strategies for overcoming the vulnerabilities of QKD have to be investigated. Including post-quantum, secure routing, etc multi-band, to name a few of them	X Cross-domain activities are also required in this area, from physics to engineers are welcome to increase the robustness of the quantum signals when they propagate through FSO/Quantum channel	X Cross-domain activities are also of interest to minimize the propagation time and traffic without losing security of the quantum keys through a 6G NTN networks

8.7.3 Federated Network of Blockchain

Decentralized security architectures such as blockchain help increase the privacy of the data and the security at the network level. Specifically, blockchain provides a tamper proof ledger to a large network of devices, unlike the conventional centralized solution which places all trust in a failure-prone central node that is exposed to stored data deletion or tampering attacks due to its single point of failure. The federated learning blockchain may increase the learning rate of the miners and reduce the latency of the network, by a consensus process controlled by a pre-selected group of nodes, rather than being open to the public like in public blockchains.

There are many applications of federated blockchain in space. For example, space-based marketplaces can be used for trading space-based assets and services, such as satellite bandwidth. For data storage, data collected from space missions can be stored on a blockchain network, providing redundancy and security. In future deep space missions, blockchain can manage the inherent time delays in interplanetary communication between Earth and other planets of spacecrafts, ensuring that transactions are recorded and verified accurately even with long communication delays, while providing data integrity and authentication. For traffic management, blockchain can be used to track the positions and trajectories of satellites, helping to prevent collisions by providing a transparent and tamper-proof record of all space objects, or help manage the allocation of orbital slots and frequencies, ensuring fair and efficient use of space resources.

In terrestrial networks, blockchain has been an intensive topic of research that has addressed technological opportunities and challenges related to the applications and architectural options, the consensus mechanisms, or the energy and computational resources demands. In space, some of these challenges apply, and there are new ones specific to the space environment related to the scarcity of resources and connectivity constraints.

The main arguments to support research activities on federated network of blockchain are:

- The wide range of practical problems and services where blockchain can enhance trustworthiness in space, such as creating marketplaces or managing resources, necessitates dedicated research. This research should focus on identifying requirements and defining architectural solutions that can support these diverse applications

- The theoretical potential and performance of blockchain can be severely impacted by the scarcity of computational resources in space and the limited connectivity. This requires a thorough investigation of the specific satellite network problems.
- The integration of this new distributed paradigm in the 5G/6G NTN specification and protocols.

8.7.3.1 Research Challenges

As a summary of the main challenges of the federated network of blockchain:

- Short-term Challenge 1: Develop a unified architecture for federated blockchain where mining is handled by the most capable satellites, while all satellites within the network participate in the blockchain.
- Mid-term challenge 2: Connectivity and long delays
- Long-term Challenge 3: The high computational cost associated with blockchain, and consensus mechanisms is a significant concern for satellites, which often have limited computational resources.
- Long-term challenge 4: extend the architecture for a multi-layer space network that includes deep space communications.

Research Theme	**Federated Network of Blockchain**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Common architecture for federated block chain	Short-term (finished in 3y)	The relationship among all nodes of a federated network of block chain. The latencies, the number of nodes and functionalities will be devised,	The number of nodes for every satellite constellation shell and the functionalities to achieve full security without impact on the latency requirements of the 3GPP.
Connectivity and long delays	Mid-term (finished in 5y)	The federated learning demands exchange of information among the nodes of the network to improve the consensus	Reduce as much as possible the exchanged information by sharing data and increasing the number of miners,
High Computational cost of block-chain	Long-term (finished in 7y+)	To target low delays and a high level of security, many miners should be involved. However, it means that the requirement in terms of complexity for processing all information combined from all miners should be huge.	Parallelization and algorithms that may work in a concurrent way, will be key for reducing the propagation delay of the federated learning on non-terrestrial learning
Multi-layer extension of federate block chain architecture	Long-term (finished in 7y+)	The satellite network will integrate terrestrial, aerial and ground networks. This is a very large architecture, which has fo be federated to support a multi-layer extension to federate block processing.	The integration of federated block chain with inter-satellite link technology to achieve faster exchange of information between the nodes of the NTN

8.7.3.2 Recommended Actions:

Research Theme	**Federated Network of Blockchain**		
Action	Common Architecture for Federated Learning	Increase the connectivity in block-chain	Multi-layer federated Block-chain
International Calls	X Given the large dynamic and wide coverage of the	X Block chain depends on multiple miners that optimize a	X The integration of multi-layers helps to increase the speed of the miners

	3D satellite network, the architecture of the block chain combined with federated learning have to be defined.	cost function. Satellite networks have large propagation delays.	of the block-chain. However, the large distance between of the nodes of non-terrestrial nodes, ask for calls about studying the multi-layer dimension of the federated block chain.
International Research	X Cooperation between multiple research groups is of interest in in international calls due to the different research profile involved in this area	X Cooperation between multiple research groups to increase the throughput of block chain techniques is of interest.	X Collaborative research on federated block chains that involve multiple satellites and at different orbits is of interest to fully develop 3D secure networks.
Open Data		X It is interested to have the channel and real propagation delays for different satellite architectures to foster the research on block chain	X It is interested to have the channel and real propagation delays when it is affected by multiple satellites to foster the research on blockchain
Large Trials		X Provide trials to validate the federated block-chain technique in space	X Develop large trials over NTN networks to experiment with different strategies for implementing the federated learning
Cross-domain research		X Cross-domain research due to foster the development of federated block chain technology multiple profiles are involved (satellite area, learning techniques, etc)	X Cross-domain research due to foster the development of federated block chain technology multiple profiles from different layers are involved (satellite area, physical layer, network layer

8.7.4 End-to-end security for integrated TN/NTN networks

End-to-end security in integrated TN/NTN networks involves implementing robust encryption and authentication mechanisms, ensuring data integrity and confidentiality. Recent efforts focus on developing security frameworks that cater to both terrestrial and non-terrestrial segments, mitigating risks associated with the broader attack surface introduced by satellite communications [C8-52]. Security in communication systems is achieved by introducing cryptographic strategies at the network layer. By doing so, it is possible to achieve a high secure key generation rate, but they cannot provide complete security against quantum computers [C8-53]. To overcome these impairments, it is proposed to resort to multi-layer security. From the network point of view, post-quantum-based solutions are being investigated to increase the robustness of classical cryptography against quantum computer attacks. In this regard, NIST has already provided an initial set of post-quantum algorithms that are robust against quantum computer attacks [C8-54]. From the physical layer point of view, strategies based on QKD and physical layer security are being combined to increase the security of communications. Nevertheless, current QKD solutions, although they can provide full security against quantum computer attacks, have a low-key generation rate [C8-55]. As a result, strategies that combine post-quantum and QKD solutions may help to achieve larger key generation rates and guarantee security against quantum computer attacks. To provide security, initial scenarios have considered QKD links from a single satellite to a ground station. However, in the next year the constellation perspective of the satellite networks will be the main key driver of communications. As a result, the extension of security to consider a global view will be mandatory. Toward this regard, integration of multiple layers and strategies to provide security will be mandatory. Under this umbrella, integration of security schemes from the physical layer, such as QKD, with the network ones, such as quantum and post-quantum cryptography, block-chain based strategies, physical layer security techniques will help to reduce the vulnerabilities of the satellite communications from the constellation perspective, and strategies to prevent tampering attacks to ensure the integrity of the content. Moreover, end-to-end security approaches that provide high security with high data rate will be of high interest for the communication networks of 6G and beyond.

The main arguments to support research activities on end-to-end security on integrated TN/NTN networks are:

- The use of NTN increases the coverage of the communication networks. However, wireless channels are inherently vulnerable to attacks. So, multi-layer approximations that provide security to end-to-end TN/NTN communications is a real need.
- The use of NTN permits to increase the capacity of the TN networks and to guarantee the QoS and the communication service not only in the well-known scenarios of rural and maritime use cases but also in disaster ones, generated naturally or by the humankind.
- TN/NTN networks provide services which without guarantee the security from end-to-end in a fast way is not possible to offer them. So, investigations on how to provide and end-to-end security in a fast way, either in the access and the backhaul links are required.

8.7.4.1 Research Challenges

Challenges of the end-to-end security:

- Short-term Challenge 1 High-security generation rate. Achieving a high-security generation rate in TN-NTN (trusted node to non-trusted node) networks faces challenges such as ensuring robust encryption, authentication, and resistance against various cyber threats.
- Short-term Challenge 2 Key Impairment for multi-layer security: The distribution of the keys suffers from the impairments of the wireless channel. Especially, when multi-hop links are required to implement the end-to-end communications. So, it is a challenge to guarantee the security of the communications, fast transmissions of the keys, and multi-hop links for end-to-end communications.
- Medium-term Challenge 3 Integration of multiple layers that provide security. Integrating multiple layers to ensure security in TN-NTN (trusted node to non-trusted node) networks presents challenges in maintaining compatibility, managing complexity, and addressing potential vulnerabilities across diverse security protocols and implementations.
- Medium-term Challenge 4: Robustness of quantum cryptography over classical. Classical cryptography needs to increase the length of the key to enhance its security against quantum computer attacks. However, this approach reduces the data rate of communications. Therefore, developing strategies that efficiently combine multiple layers is challenging
- Long-term Challenge 5 Time to time increasing the physical layer security in quantum communications. Continuously enhancing physical layer security in quantum communications faces challenges such as adapting to evolving encryption standards, overcoming technological limitations, and mitigating potential vulnerabilities to quantum hacking techniques.
- Long-term Challenge 6: Extending multi-level security to a constellation of satellites. The use of multiple satellites increases transmission coverage but also opens the path to eavesdropping, man-in-the-middle attacks, and data poisoning. Therefore, techniques that enhance security by leveraging information from the physical, MAC, and network layers, as well as combinations of multiple secure

schemes, will be of great interest. Due to the vast array of techniques involved, this is a challenging area of research.

Research Theme	**End-to-end security for integrated TN/NTN network**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Challenge 1 High-security generation rate	Short-term (finished in 3y)	Metrics affecting high-security key generation rate in an integrated terrestrial and non-terrestrial network include signal loss, noise levels, bit and quantum bit error rates, secret key rate, channel efficiency, latency, security vulnerabilities, interference, scalability, and environmental factors.	The metrics collectively improve signal integrity, reduce errors, enhance security, optimize resource utilization, and ensure scalability and reliability, resulting in a highly efficient and secure integrated TN/NTN network.
Challenge 2 Key Impairment for multi-layer security	Short-term (finished in 3y)	The key metrics for this challenge are error rate in the transmission of the key frames when information of multiple layers is used to protect end-to-end link, latency in the generation of the keys when there are channel impairments.	The expected contributions in this challenge are: error correction schemes, routing algorithms, quantum communications for multi-hop links, post-quantum algorithms with information of multiple network layers.
Challenge 3 Integration of multiple layers that provide security	Mid-term (finished in 5y)	Key metrics include encryption strength and authentication efficiency. Also crucial are intrusion detection, resilience against cyber threats, and traffic analysis resistance, ensuring holistic protection and regulatory compliance.	Integrating multiple security layers in NTN and TN networks ensures a resilient, secure, and compliant environment, safeguarding data integrity, confidentiality, and availability across diverse network domains.
Challenge 4 Robustness of keys	Mid-term (finished in 5y)	Key metrics for this area : key length when it is necessary to encode from end-to-end, propagation time of the keys, compared to use information from a single layer respect to the integrated one in the end-to-end approach	The expected contributions in this area fall in: integration of information of multiple layers to reduce the key length, integration of security schemes such as quantum and post-quantum, access and routing schemes with security constraints.
Challenge 5 Time to time increasing the physical layer security in quantum communications	Long-term (finished in 7y+)	Key metrics for improving physical layer security in quantum communications include QKD rate, photon transmission efficiency, quantum SNR, photodetector efficiency, and error rate, ensuring integrity, confidentiality, and reliability of quantum information transmission.	The contributions result in more secure, reliable, and resilient communication systems, paving the way for the widespread adoption and advancement of quantum technologies.
Challenge 6 Extend the multi-level security to a constellation of satellites	Long-term (finished in 7y+)	The key performance metrics of this challenge are: error rate in the transmission of the key, key generation rate, number of links intercepted by the eavesdroppers	The contributions are: new strategies to integrate block chain, quantum and post-quantum techniques, reduced size authentication process, fast pointing, acquisition and tracking algorithms, use multiple transmission reception strategies

7.1.1.2 Recommended Actions:

Research Theme	**End-to-end security for integrated TN/NTN network**		
Action	Integration of Multiple Layers of Different Security Strategies	Integration of TN/NTN systems from end-to-end security perspective	End-to-end security from a NTN satellite view
International Calls	X Integration of strategies for multiple security strategies such as federated block chain, quantum-based security, post-quantum ones, is a real challenge. So, calls in this topic should be promoted	X To achieve a fully transparent TN/NTN that integrates all layers to provide an end-to-end security is a need. To increase the security, efficiency, data rate of the communications without losing security.	X In 6G, there will be deployed the first generation of satellite constellations. Toward, this area the constellation dimension has to be introduced in the end-to-end security since multiple satellites may be involved such as in the multi-connectivity-based configurations
International Research	X Collaborative research in the integration of multiple layers is a need since the expertise of the different strategies is not focused on a single country and domain of investigation (e.g. physics, engineers)	X Collaborative research in the integration of TN/NTN is a need to target a consensual standard of TN/NTN with integrated end-to-end security	X Collaborative research in the end-to-end security in NTN assuming the constellation view is a need since the NTN coverage in 6G will be multinational.
Open Data	X Open data for the different channels and protocols that run over the end-to-end perspective is required to develop future AI/ML systems that run on top of it		
Large Trials			X In the constellation perspective large trials will be necessary to experiment with the end-to-end security.
Cross-domain research	X In this area experts from multiple disciplines of security interact to define the integration between the multiple security schemes to achieve an end-to-end security strategy.	X In this area experts from security areas of TN and NTN will interact to develop an end-to-end transparent TN/NTN security strategy.	X In this area experts from security areas of NTN will interact to develop an end-to-end NTN security strategy.

8.7.5 AI-based Interference Classification for Localization and Positioning Services

The accuracy and reliable localization of vehicles on roads for self-driving cars, toll systems, and digital tachographs are crucial. To achieve this, vehicles typically employ global navigation satellite system (GNSS) receivers to validate their absolute position information. However, GNSS-based positioning can be compromised by interference signals, necessitating the identification, classification, determination of purpose, and localization of the interference to mitigate or eliminate them. Recent approaches based on machine learning (ML) have shown superior performance for interference monitoring. However, their feasibility in real applications and environments is yet to be assessed. To effectively implement (supervised) ML techniques, it is necessary to have training datasets (including sensor data) that incorporate realistic interference signals, including real-world noise and potential multipath effects that may occur between transmitter, receiver, and satellite in the operational area. Additionally, these datasets require reference labels. However, creating such datasets is often challenging due to legal restrictions, as the use of interference from GNSS sources is strictly prohibited. The evaluation of ML-based methods in the literature has been limited to synthetic data or controlled laboratory environments. Consequently, the performance of ML-based methods in practical applications remains unclear. First, a database is recorded, next the interference is detected, classified by employing ML-techniques, and the interference is visualized on a heatmap. One additional source of signal, i.e., 6G data, may increase the positioning accuracy by fusing both input signals. However, in different signal bands, the interferences have different patterns on which the classification accuracy may change.

Challenges. To enhance the robustness of positioning systems by mitigating interferences, it is imperative to initially detect and reliably classify these interferences. However, achieving a dependable classification is impeded by several factors [C8-60][C8-61][C8-62][C8-63][C8-64]: (1) The presence of a multitude of potential (hardware) jammer types, each exhibiting distinct interference [C8-63][C8-64]. (2) Variations in the jammers' frequency, bandwidth, and signal-to-noise ratio, which directly influence the interference patterns. Conducting a feature analysis is crucial for identifying and emphasizing the most significant extracted features [C8-61]. (3) Despite intentional jammers typically being constructed as in-band interferers, numerous unintentional sources of interference contribute to noise in the spectrograms. Employing data augmentation techniques can boost the resilience of ML models. (4) Environmental changes, such as fluctuations in weather conditions (e.g., temperature) or alterations in multi-path scenarios, necessitate the utilization of representation learning, transfer learning, and domain adaptation methods. These changes introduce shift in feature distributions between source and target domain samples. (5) Difficulties arise in transferring models across diverse hardware setups, including low-cost, medium-end, and high-end sensors, smartphone data, and various antennas [C8-64][C8-61]. Continual learning and few-shot learning methods enable adaptation to novel data types. (6) Effectively classifying and promptly adapting to interferences across different locations requires orchestrating ML models by exchanging (sub-)information of new jammers. Federated learning leverages continual learning and few-shot learning to adjust to unfamiliar interference classes and facilitates weight sharing through an aggregation process over a global (central) ML model.

Machine Learning. Both classical and ML methods have demonstrated efficiency in detecting and classifying interference. However, the unpredictable emergence of novel, “undetected” jammer types mandate rapid model adaptation [C8-58][C8-59][C8-60][C8-64]. The objective is to develop ML models that exhibit resilience against diverse jammer types, interference characteristics, variations in antennas, environmental fluctuations, changes in location, and disparate receiver stations. The imperative to adapt to unforeseen jammer types and emerging interference characteristics, opens research avenues for methodologies such as representation learning [C8-60], transfer learning, domain adaptation, continual learning [C8-59], few-shot learning [C8-58], and federated learning.

Continual Learning. Continual learning [C8-59] and few-shot learning [C8-58] denote the capability to sequentially learn consecutive tasks without forgetting how to perform previously trained tasks. In the context of interference classification, such methods are imperative for the adaptation to novel jammer types, including research into optimal strategies for selecting prior task samples and the dynamic configuration of architectures, exemplified by the integration of generalized and specialized blocks.

Federated Learning. By orchestrating ML models across diverse locations and facilitating the exchange of data/information concerning novel jammer types, the models' weights can be updated through an aggregation step. The related research topics encompass the optimization of the aggregation step, the reduction of model parameters to mitigate data transmission volumes, and the implementation of anonymization techniques on model attributes to uphold privacy within federated learning.

6G NTN. Fusing GNSS and 6G signals for localization presents a compelling opportunity to enhance positioning accuracy and reliability, leveraging the complementary strengths of each technology. GNSS provides global coverage with high precision in open environments, while 6G, with its dense network and advanced communication capabilities, excels in urban and indoor settings where GNSS performance often degrades. This fusion not only aims at improving localization but also offers significant advantages in interference classification. While GNSS typically faces multipath and signal blockage, 6G may encounter spectrum congestion and complex urban interference. Since the interference patterns differ between GNSS and 6G signals, integrating data from both sources can lead to more robust and accurate interference detection and classification. The accuracy and reliable localization of vehicles on roads for self-driving cars, toll systems, and digital tachographs are crucial. To achieve this, vehicles typically employ global navigation satellite system (GNSS) receivers to validate their absolute position information. In the context of joint sensing and communication, localization and navigation will be an integral part of 6G technology as well. This topic is particularly important for non-terrestrial navigation systems, where seamless integration of multiple signal sources is essential for maintaining accurate and reliable positioning in diverse and challenging environments.

The main arguments to support research activities on AI-based Interference Classification for Localization and Positioning Services:

- Interference detection algorithms are required in various applications. Hence, well-trained methods that are flexible over different applications, environments, and scenario are important.
- Through the development of federated and continual learning methods for interference classification applications, a knowledge basis can be built up, than can be extended to different applications.

8.7.5.1 Research challenges:

Research Theme	**AI-based Interference Classification for Localization and Positioning Services**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Improved reliability and robustness in detection and classification of interferers (refer to Research aspects as per 4.7.5)	Short-term (finished in 3y) to Mid-term (finished in 5y)	ML model that is trained in a well generalizable manner. The model achieves a high accuracy over different aspects, such as environmental fluctuations and different interference characteristics.	A well-generalized ML model. More reliable interference detection algorithms.
Flexibility and adaptation to environmental changes and robustness to environmental fluctuations	Short-term (finished in 3y)	ML model based on continual or few-shot learning that can adapt to new interference types through environmental changes.	A continual learning method. More reliable interference detection algorithms.
Orchestration of different specialized ML models by exchanging (sub-) information and/ or deployment as federated learning approach	Short-term (finished in 3y)	Federated learning approach with various local models that can adapt to new interference types.	A federated learning method. More reliable interference detection algorithms.
Fusing GNSS and 6G signals for localization	Mid-term (finished in 5y)	A model that can handle both data sources utilizes the positive aspects of both data.	A well-trained ML model that works better in both areas, 5G and 6G.

			More reliable interference detection algorithms.

8.7.5.2 Recommended Actions

Research Theme	AI-based Interference Classification for Localization and Positioning Services		
Action	Flexibility and adaptation to environmental changes and robustness to environmental fluctuations.	Orchestration of different specialized ML models.	Fusing GNSS and 6G signals for localization.
International Calls	X Interesting that new research calls to investigate robustness of the communication systems to atmospheric phenomena be fostered	X ML that be robust to different models be investigated,	X Calls that integrate 6G and GNSS be promoted to improve the communication capability and positioning error
International Research		X So far only a little research in this field for orchestrating ML models for interference classification applications.	X It is a recent area of research in terrestrial communications and in non-terrestrial domain few research has been developed till now.
Open Data	X Promote open data from the GNSS and 6G channels for satellite		
Large Trials	X Measurement campaigns in the real-world.	X Measurements campaigns in the real-world.	
Cross-domain research	X Given the large set of disciplines that participate in this area, cross-domain research will be of interest (signal processing, GNSS, channel knowledge, etc.)	X Given the large set of disciplines that participate in this area, cross-domain research will be of interest (ML, AI, signal processing, GNSS, software, hardware, etc.)	X Given the large set of disciplines that participate in this area, cross-domain research will be of interest (6G, GNSS, signal processing, software, hardware)

8.7.6 Related/treated security assets & potential resulting security needs

Research Theme / Research Aspect	Security asset	Explanation
Security proof (QKD)	Critical data	Developing rigorous mathematical proofs to demonstrate the security of QKD protocols against various types of attacks
Protocol Variants (QKD)	Critical system	Investigating different QKD protocols, such as BB84, E91, and more advanced ones like Measurement-Device-Independent QKD (MDI-QKD).
Device-Independent QKD (QKD)	Critical component	Ensuring security even if the devices used are not trusted or may be partially compromised.
Robustness to interference (Federated Network of Block chain)	Critical data	Federated Systems optimize the systems collectively. However, if the data is corrupted by interference, it may drive to incorrect conclusions at the consensus procedure. So, enhancements of the protocol of federated network of block chain have to be implemented to avoid data

Resource Management System (Federated Learning Block Chain)	Critical system	The Resource and Management system in federated network of block chain has to minimize the resources, the storage and the consumed energy. This is critical in non-terrestrial networks since satellites are energy-limited systems.
Consensus Procedure (Federated Learning Block Chain)	Critical component	Federated Systems need a consensus system to integrate the distributed optimization. This strategy in a non-terrestrial network has to be optimized to minimize the resources used and the delays in the data exchange.
Flexibility and adaptation to environmental changes and robustness to environmental fluctuations. (GNSS)	Critical data	Recorded data as well as reliable availability of data may be critical for certain applications.
Orchestration of different specialized ML models and federated learning. (GNSS)	Critical data	Recorded data as well as reliable availability of data may be critical for certain applications.
Fusing GNSS and 6G signals for localization. (GNSS)	Critical data	Recorded data as well as reliable availability of data may be critical for certain applications.
Robustness to the channel impairments from end-to-end (end-to-end security)	Critical system	The channel impairments may increase the time in the end-to-end communication and integration of multiple keys
Latency vs integration of multiple layers (end-to-end security)	Critical system	The use of information of multiple layers may enhance the level of security. However, it may also reduce the data rate of the communications notably. So, efficient solutions for integrating multiple layers at a higher rate are expected.

8.7.7 Proposed security mechanisms

Research Theme / Research Aspect	Security mechanism type / toolbox	Explanation
Key Management (QKD)	System engineering	QKD provides a continuous supply of fresh, symmetric keys for encryption, reducing the risk associated with key exhaustion and increasing overall security.
Data Encryption (QKD)	System engineering	Quantum keys can be used with classical encryption algorithms like AES (Advanced Encryption Standard) to secure data transmissions, ensuring that even if the data is intercepted, it cannot be decrypted without the key.
Post-Quantum Security (QKD)	System engineering	QKD offers security against future quantum computer attacks, which threaten traditional public-key cryptography methods like RSA and ECC (Elliptic Curve Cryptography).
Data Verification System (Federated Block Chain for NTN)	System engineering	This system evaluates the trustiness of the data by detecting the presence of interferences, if the data has been received from the correct satellite, and if there is any corruption in the data.
Data Encryption System (Federated Block Chain for NTN)	System engineering	The mechanism to encrypt the data. It can be classical cryptography, as quantum, post-quantum and combination of multiple of them
Data Transmission System (Federated Block Chain for NTN)	System engineering	The data transmission of the system through NTN network has to be aware of the presence of eavesdroppers, the links with bad atmospheric conditions, the duration of the links, since the inter-satellite links generally have a reduced visibility period. By doing so,

		it will be possible to select the most secure path against eavesdroppers and atmospheric conditions.
Flexibility and adaptation to environmental changes and robustness to environmental fluctuations. (GNSS protection)	System engineering	Significantly enhanced robustness and reliability of location information and interference detection by using more powerful methods/algorithms (now possible to be implemented) Useful for governmental/ military services, critical industrial applications, avionics and (autonomous) vehicle control (on ground, water, or in air)
Orchestration of different specialized ML models and federated learning. (GNSS protection)	System engineering	Significantly enhanced robustness and reliability of location information and interference detection by using more powerful methods/algorithms (now possible to be implemented) by distributed processing Useful for governmental/ military services, critical industrial applications, avionics and (autonomous) vehicle control (on ground, water, or in air)
Fusing GNSS and 6G signals for localization. (GNSS protection)	System engineering	Significantly enhanced robustness and reliability of location information and interference detection by using more powerful methods/algorithms (now possible to be implemented) and intelligent fusion of data from different sources Useful for governmental/ military services, critical industrial applications, avionics and (autonomous) vehicle control (on ground, water, or in air)
Post-quantum security integration (end-to-end security)	Cryptography	It is assumed that post-quantum cryptography can provide enough robustness to quantum computers. However, it may not be the case with the evolution of quantum computers.
Integrating multiple technologies for security in 6G (end-to-end security)	System engineering	The assumption is that the integration of multiple security schemes permits us to achieve a larger security level. Specifically, to attain security against quantum computers. However, this integration has to be done in a way that quantum computers do not exploit any leakage of information that permits them to break the keys of the end-to-end security scheme.

8.7.8 Security management mechanisms and related requirements

Research Theme / Research Aspect	Security management type	Explanation
Risk Management and Assessment (QKD)	key management, security	Techniques for identifying potential threats to an organization’s assets. Evaluating the potential impact and likelihood of risks.
Cryptographic Techniques (QKD)	signature database update	Advances in symmetric and asymmetric cryptography. Cryptographic methods resistant to quantum computing attacks.
Network Security (QKD)	mechanism reconfiguration	Advanced firewall configurations and intrusion prevention systems (IPS). Advanced firewall configurations and intrusion prevention systems (IPS).
(Federated Block Chain)		
Flexibility and adaptation to environmental changes and robustness to environmental fluctuations. (GNSS protection)	Reliable provision and consistency/verification of data	Various services depend on and rely on correct location data which is often mission critical. Management of different validation/ verification and assessment methods of one or more data sources is essential.

Orchestration of different specialized ML models and federated learning. (GNSS protection)	Reliable provision and consistency/verification of data	Various services depend on and rely on correct location data which is often mission critical. Management of different validation/ verification and assessment methods of one or more data sources is essential.
Fusing GNSS and 6G signals for localization. (GNSS protection)	Reliable provision and consistency/verification of data	Various services depend on and rely on correct location data which is often mission critical. Management of different validation/ verification and assessment methods of one or more data sources is essential.
Efficient Key Management System (end-to-end security)	Key management	The provision of security from an end-to-end may include the integration of multiple keys/ strategies of protecting the signal.

8.7.9 Expected Impact

8.7.9.1 Key Performance Indicators (KPI)

This section summarizes the evolving KPIs for Quantum Key Distribution (QKD), end-to-end security in non-terrestrial networks (NTNs), federated networks of blockchain and GNSS systems. For QKD the main KPIs will be the key generation rate, error rate, quantum stated fidelity and privacy amplification efficiency. Currently, satellite-based QKD systems achieve key generation rates from a few bits per second to several kilobits per second. Future goals aim for key rates of tens of kilobits to a few megabits per second. The focus will be on reducing error rates from 5-15% to 1-3% through improved quantum state preparation, satellite alignment, and advanced error correction. Enhancements are also expected in authentication procedures, quantum state fidelity (from 90-95% to 95-99%), and privacy amplification efficiency (from 50-70% to 70-90%) through better optics, noise reduction, and algorithm optimization. Regarding end-to-end security, the integration of multiple security technologies, such as post-quantum cryptography, QKD, blockchain, and federated learning, will impact KPIs like reliability, throughput, latency, and data rate. These systems must accommodate increased frame lengths and the complexity of exchanging security information, necessitating improvements in throughput, latency, and data rate while ensuring quality of service for 6G networks. In relation to federated network of blockchain, its key KPIs include throughput, latency, energy efficiency, and scalability. Current satellite networks are expected to scale from tens to thousands of nodes with advancements in distributed computing. Resource utilization, currently around 60-70%, is projected to improve to 80-90% with better management and optimization. Energy efficiency will be enhanced by 20-30% through optimized algorithms and energy-efficient hardware. Finally, the GNSS and AI-based interference cancellation are key for improving positioning accuracy, reliability, and data rate in NTNs. These systems must be robust against interference and capable of handling jamming, which can disrupt multiple users. Key KPIs include positioning accuracy, interference removal capacity (SINR at the output), the level of interference mitigated, and the resources used by the ML system for training and operation.

8.7.9.2 Key Value Indicators (KVI)

In Section 8.3.7.2, the Key Value Indicators (KVI) for Non-Terrestrial Network (NTN) architecture are categorized into economic, societal, and environmental sustainability. The next generation of QKD security systems must align with these sustainability pillars. Given the energy limitations of satellite systems, advancements in QKD must focus on energy efficiency. Secure communications are essential for societal applications like remote health services and high-security transactions, while environmental sustainability is vital for monitoring systems such as water quality sensors and earthquake warnings. Security schemes, including federated learning blockchain systems, must support these sustainability criteria by being energy-efficient and resilient to interference. End-to-end security remains a priority to ensure the benefits of sustainable communication architectures. AI-based GNSS interference cancellation schemes also impact economic sustainability by enhancing network coverage and resilience. Addressing these challenges, it will be given an important step forward to achieve the desired quality of service and aligning with the UN Agenda for Sustainability.

8.8 Sustainability Considerations

Research Theme /Research Aspect	Sustainability impact area	LCA	Actual impact
Security	Q1-Q6	Distribution	If no security no distribution of the manufacturing since the loop if LCA may be broken due to: i) inefficiencies of the transport to use, ii) no remote monitoring, iii) no electronic bank payment
3D Architecture Management	Q1-Q6	Distribution	The research aspects related to the 3D architecture management impact sustainability at all levels for the vertical markets and users exploiting the 6G network with a native NTN component.
Routing in space	Q1-Q6	Distribution	Achieving sustainable data communication by exploiting data forwarding in space, by taking into account the operational constraints imposed by NTN nodes, expressed in terms of SWaP figures.
Network orchestration	Q1-Q6	Distribution	Effective monitoring, control, and allocation of resources according to vertical needs in order to optimize the use of the available resources through the entire network over the so-defined network slices.
Edge computing	Q1-Q6	Distribution	More effective distributed power consumption across edge nodes to allow faster and more sustainable access to data
Enhanced positioning	Q1-Q6	Distribution	Enhanced positioning technologies can support more efficient and sustainable monitoring and localization services for many vertical sectors.

9 Opportunities for Devices and Components

Editor: Andre Bourdoux

Progresses in all aspects of the wireless and wireline network are highly dependent on electronic technologies, com ponents 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/wireline 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.

9.1 Vision and Requirements

Wired and wireless networks are in constant evolution with the goal of addressing all relevant societal challenges, support the digitization of the industry, improve communications devices, support new applications (see Chapter 1), support the “more AI” trend, ... To reach these goals, we expect future networks to support very low to very high throughputs, increase area coverage, reduce latencies, improve reliabilities, integrate artificial intelligence, support an ever larger number of verticals.

The requirements on the components (e.g. chips, antennas, ...) and devices (in this chapter, “device” is used with the meaning of “user device” in the broad sense but not “transistor device”) are very broad: they cover all aspects of infrastructure and human and non-human user device hardware and software. The eleven sub-sections of this chapter provide each their set of requirements for different parts of the hardware including radio transceiver hardware, computing and storage, optical hardware, security and IoT.

9.2 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 number of 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 modules 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 a certain extent, 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 far from maturity 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. On top of this the ongoing convergence of communication and sensing is expected to become a key feature of future networks, implying converged front-ends, transceivers and digital signal processing, whilst security and trust considerations may require strict separation of both the sensing and communication concerns such that access to the sensing part can be adequately restricted. This will also drive requirements on the hardware and integration both at the device, circuit and packaging level [C9-57].

9.2.1 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Sub-10GHz RF**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Better and more efficient radio hardware	Long-term (finished in 7y+)	Mix of solutions: architectures, circuits, digital RF, silicon technology node	Cost effective, performant, green radio
Faster and higher resolution converters (DACs and ADCs) at low power	Long-term (finished in 7y+)	Mix of solutions: architectures, circuits, silicon technology node	Useful for all radios

9.2.2 Recommendations for Actions

Research Theme	Sub-10GHz RF	
Action	Better and more efficient radio hardware	Faster and higher resolution converters (DACs and ADCs) at low power
International Calls	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems
International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.3 Millimeter-wave and TeraHertz

9.3.1 THz Communications

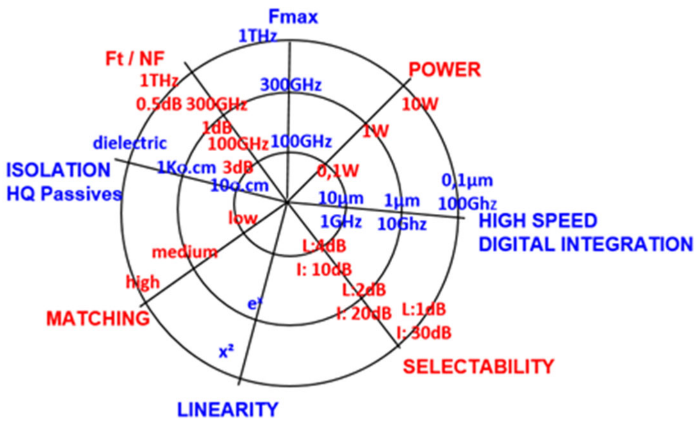
With a massive amount of unused spectrum, the sub-TeraHertz (sub-THz) frequencies between 90 and 300 GHz are candidates to achieve high-data rate wireless and wireline communications and hence to fulfil the requirements of the next-generation of wireless networks and wireline (e.g. data center) networks [C9-1], [C9-2]. THz signals may also be carried over low-cost waveguide structures such as polymer microwave fibre (PMF) as an alternative to high performance copper and optical interconnect use cases over distances from a few cm's upto a few m's.

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 must be compensated by using high-gain antennas, which entails severe constraints on antenna directivity and alignment. Sub-THz has, in similarity with mm-wave frequencies, a vulnerability to precipitation that limits the practical reach of beyond 100GHz links to less than 1-2 km with telecom grade availability (>99.99%). However, combining lower and higher frequency carriers makes it possible to have the lower frequencies backing up the performance during, rare, high precipitation events while still taking advantage of very high peak rates over long hops (>2-5 km). The use of high frequencies also makes it possible to build small and compact radios beyond 100GHz that will work well in urban environments where there is a need for non-intrusive installations. The ability to create high power, wide bandwidth amplifiers on these bands will make it possible to both increase reach and avoid the need for larger antennas. Larger antennas and simplified deployments have driven a need to investigate electronically controlled steerable beams that today 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 [C9-3]. 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 [C9-4].

9.3.2 Solid-state technologies for THz applications:

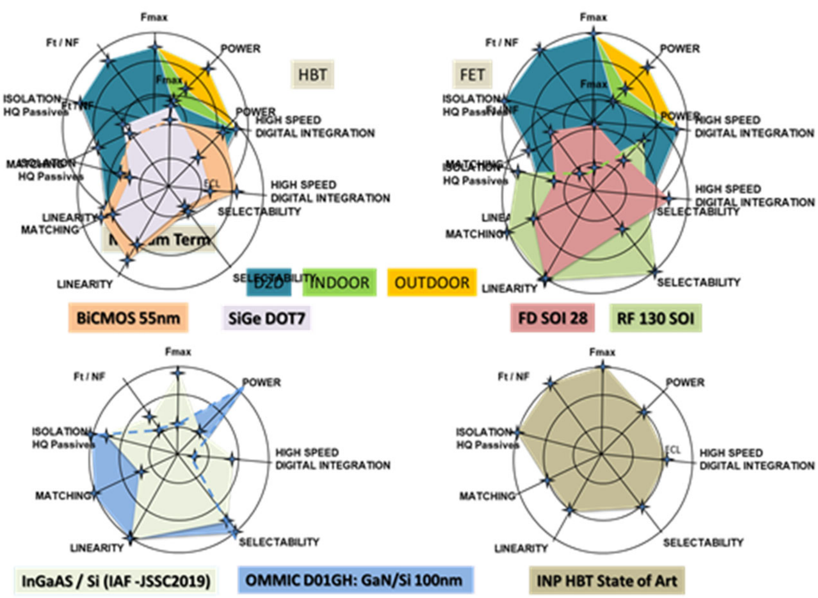
Nowadays, silicon-based technologies offer low-cost solutions for RF and millimetre-wave applications combined with a high complexity in the digital domain. CMOS, however, has its limits in speed and power generation, which become apparent at operation above 100 GHz. This is evidenced in the on-line survey of power amplifiers, maintained by the Georgia Tech University [C9-9]. 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 [C9-7], 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 [C9-7] 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 [C9-10] 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 [C9-13] 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 [C9-14]. 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 [C9-15]. Also here, a wide deployment of GaN devices might benefit from integration on 300 mm wafers and with many Cu metal levels as in [C9-15].

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 [C9-16].

9.3.3 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 [C9-8]. but, using two different circuits, it is an expensive solution.
2. The monolithic CMOS-based THz imagers have recently emerged as low-cost competitors [C9-9, 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.

9.3.4 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. Sparse arrays and distributed architectures, combined with advanced signal processing, are also investigated by many research groups.79GHz radar imaging with wide field-of-view, resolutions of 1 by 1 degree and cm-scale range resolution is experimentally feasible today and radars with wide field-of-view and LiDAR-like resolutions (e.g. 0.1 by 0.1 degree) 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) [C9-9]. 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.

Following the technology trend of 5G networks moving into mm-wave bands, there is an increased interest in developing phased arrays above 100 GHz for 6G, especially around D-band. This frequency band is interesting for its high-bandwidth potential, being able to support joint communication and sensing (JCAS) applications, typically in some kind of hybrid phased array architecture. The development of phased array transceivers operating in D-band compatible with massive MIMO operation necessitate to overcome critical performance bottlenecks of massively parallelized antenna arrays at mm-wave together with novel integration approaches to improve efficiency and related thermal dissipation.

9.3.5 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Millimeter-wave and TeraHertz**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Better and more efficient radio hardware at mm-wave, sub-THz and THz	Long-term (finished in 7y+)	Mix of solutions: architectures, circuits, digital RF, silicon technology node	Cost effective, performant, green radio
Faster (>>1Gsps) and higher resolution converters (DACs and ADCs) at low power	Long-term (finished in 7y+)	Mix of solutions: architectures, circuits, silicon technology node	Useful for all radios with extreme bandwidths
Semiconductor technologies CMOS, BiCMOS, III-V	Long-term (finished in 7y+)	Affordable semiconductor technologies for different market volumes	Enable mm-wave to THz radios

9.3.6 Recommendations for Actions

Research Theme	**Millimeter-wave and TeraHertz**		
Action	Better and more efficient radio hardware at mm-wave, sub-THz and THz	Faster and higher resolution converters (DACs and ADCs) at low power	Semiconductor technologies CMOS, BiCMOS, III-V
International Calls	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.4 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.

9.4.1 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 [C9-17] 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.

Energy harvesting is another interesting approach for devices requiring extremely low energy. Candidate energy harvesting technologies include thermoelectric, photovoltaic, piezoelectric, RF or wireless, wind and vibration energy harvesting.

9.4.2 Spatial Awareness

For spatial radios, we can differentiate between active and device-free localization. In active localization, two (or more) IoT nodes 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 [C9-18].

Currently, channel state information is often based on received signal strength (RSS), mostly because it is easy to implement. However, for 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 [C9-19], a wideband view on the radio channel can be 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 [C9-20], but also Wi-Fi [C9-21].

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 [C9-22].

9.4.3 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 [C9-52]. 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. [C9-53]) but would have to be adopted more widely to the Internet of Everything applications in order to avoid environmental problems.

9.4.4 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Ultra-low Power Wireless**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Battery-free operation and disposable devices	Long-term (finished in 7y+)	Mix of solutions: system concept, protocols, architectures, circuits, energy harvesting, eco-friendly electronics	Green disposable radios

9.4.5 Recommendations for Actions

Research Theme	**Ultra-low Power Wireless**
Action	Battery-free operation and disposable devices
International Calls	X To leverage industry and academic efforts vertically for critical mass
International Research	X To leverage industry and academic efforts vertically for critical mass

9.5 Antenna and Packages

9.5.1 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 [C9-23] 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.

9.5.2 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 of 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) (also known as Reflective Intelligent Surfaces (RIS)) 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) [C9-24], 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. [C9-26] 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 [C9-51
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 [C9-25], 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.

RIS are investigated as a low power and low-cost solution for coverage extension, especially in millimeter-wave frequency bands. Early trials have shown the capability to improve the coverage in areas suffering from poor coverage. However, there are still questions about the maturity, energy consumption, Total Cost of Ownership and deployment of these surfaces. They will need to be assessed and compared with alternatives such as small cells or traditional repeaters.

9.5.3 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Antenna and Packages**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
On-chip antennas, lens-integrated antennas, antenna MIMO arrays	Long-term (finished in 7y+)	Mix of solutions: packaging, interconnects, lenses, on-chip vs off-chip antennas, beamforming/MIMO trade-offs	Small form factor, high efficiency antennas for all frequency ranges
Metamaterials and metasurfaces	Long-term (finished in 7y+)	Breakthrough antennas/antenna arrays and reflective surfaces	Improved coverage, low losses, passive operation

9.5.4 Recommendations for Actions

Research Theme	Antenna and Packages	
Action	On-chip antennas, lens-integrated antennas, antenna MIMO arrays	Metamaterials and metasurfaces
International Calls	X	X

	To leverage industry and academic efforts vertically for critical mass	To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems
International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.6 Optical wireless convergence

9.6.1 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) [C9-27].
- High speed PON to facilitate fixed and wireless convergence

9.6.2 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.

9.6.3 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	Optical wireless convergence		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Radio-over-fibre and other advanced fronthauling technologies	Long-term (finished in 7y+)	Higher efficiency and higher throughput fronthaul	Key technology for O-RAN, distributed/cell-free MIMO
Optically assisted wireless subsystems	Long-term (finished in 7y+)	Better, higher efficiency circuits and building blocks for sub-THz/THz	Improved functionality, higher efficienvy at Sub-THz/THz

9.6.4 Recommendations for Actions

Research Theme	Optical wireless convergence	
Action	Radio-over-fibre and other advanced fronthauling technologies	Optically assisted wireless subsystems
International Calls	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems
International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.7 Baseband Modems

Figure 9-1 shows the range of the processing options that have been explored over the years for base stations baseband modems [C9-36]. 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.

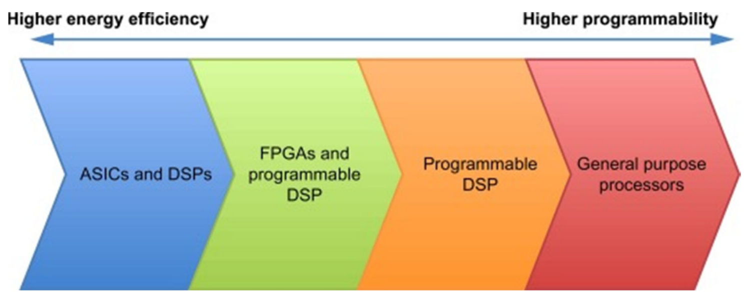


Figure 9-1. Baseband processing options [C9-36].

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. GPUs can also be considered for highly parallelizable tasks that do not require frequent and irregular memory accesses.

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.

For UE modems (e.g. smartphones and IoT devices), ASIC implementations with embedded accelerators, DSP and CPU cores is dominating for power, cost, performance, size perspective. Recently, also dedicated machine learning acceleration is considered within the modem for certain categories of devices. Depending on the advancement regarding machine learning / AI in the PHY, such processing might be more commonplace and demand additional emphasis from an acceleration perspective.

A recent, and potentially disruptive development is the application of deep learning for the physical layer. By interpreting a communication system as an autoencoder, several groups are developing a fundamental new way to think about communication system's baseband design as an end-to-end reconstruction task that seeks to jointly optimize transmitter and receiver components in a single process [C9-37], [C9-38]. 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.

Advances in CMOS scaling (following Moore's law) is crucial to enable progresses in digital implementations. With the current technologies (FinFETs), one approaches several "walls" such as the power wall, performance wall, scaling wall and cost wall. Disruptive approaches are needed to break down these walls to enable evolution towards 1nm (10A) and beyond by the turn of the decade. Promising approaches include gate-all-around (GAA) silicon nanosheet, gate-all-around forksheet, complementary FET (CFET) and exploiting the 3rd dimension. It becomes also essential to have a holistic approach of system-technology co-optimization whereby the application (system), the algorithms, the architecture, the design and the fundamental technologies are considered and optimized jointly.

9.7.1 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Baseband Modems**		
Research Challenges	Timeline	Key outcomes	Contributions/Value

Architecture and processor trade-offs (TPU, GPU, CPU, DSP, ASIC, FPGAs, ASIPs,...)	Long-term (finished in 7y+)	Powerful and Efficient DSP implementation solutions adapted to a broad range of use cases, from simple IoT device to complex base station	High efficiency and computing power
Semiconductor technologies (CMOS scaling towards 1nm and beyond)	Long-term (finished in 7y+)	New technologies such as GAA nanosheet, GAA forksheet and complementary FET to extend Moore's law beyond 1nm	Power, performance, scaling and cost for advanced digital implementations

9.7.2 Recommendations for Actions

Research Theme	**Baseband Modems**	
Action	Architecture and processor tarde-offs (TPU, GPU, CPU, DSP, ASIC, FPGAs, ASIPs,...)	Semiconductor technologies (CMOS scaling towards 1nm and beyond)
International Calls	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems
International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.8 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 must 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.

Spiking neural networks is a promising approach to enable bio-inspired learning with extreme efficiency. This event-driven architecture reduces drastically the amount of computing yet achieves excellent performances. Both digital and analog implementations are possible, with analog implementations being the most energy efficient.

9.8.1 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**AI processors**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Processors for Cloud-AI, Edge-AI and on-device-AI	Long-term (finished in 7y+)	Powerful and Efficient AI processors and memories fitting the energy budget at cloud, edge or device. Innovative architectures such as in-memory compute and spiking neural networks	AI enablement

9.8.2 Recommendations for Actions

Research Theme	**AI processors**
Action	Processors for Cloud-AI, Edge-AI and on-device-AI
International Calls	X To leverage industry and academic efforts vertically for critical mass
International Research	X To leverage industry and academic efforts vertically for critical mass

9.9 Memories

9.9.1 Memory technologies towards 2030

9.9.1.1 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 centre – 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.

9.9.1.2 Clever data mining, and reduced energy consumption

At some point in the flow of data transport, the generated data will need to be analysed 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 down the data flow. 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 centres, for processing, transporting and storing all the data. Without energy consumption optimization, the energy consumption for these operations, data centres worldwide may use almost 8000 terawatt-hours by 2030. (source: https://www.labs.hpe.com/next-next/energy).

9.9.1.3 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-level 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 NAND Flash memory chips for storage. In general, memories located further away from the CPU are cheaper (less expensive per byte), slower, denser and less volatile.

For half a century, Moore's Law has driven the continuous increase of memory densities, and this has translated into cost improvement of memory technologies. However, despite large improvements in memory density, only storage density (NAND Flash devices) has truly kept pace with the data growth rate. With the transition from planar 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.

9.9.1.4 MRAM technologies for embedded cache level applications

Spin transfer torque MRAM (STT-MRAM) technology [C9-39], [C9-40] 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) [C9-41], 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 [C9-42] is another interesting emerging option offering low power, high performance and high-density non-volatile memory solution.

9.9.1.5 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 peripherical 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.

9.9.1.6 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 array, such as phase-change-RAM (PC-RAM), vacancy-modulated conductive oxide (VMCO), conductive bridging RAM (CB-RAM) and oxide RAM (OxRAM).

9.9.1.7 3D NAND... and beyond?

Since its introduction several years ago, 3D NAND [C9-43] has become a mainstream storage technology because of its ability to significantly increase bit density without sacrificing endurance and performance. 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 100 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 [C9-44].

9.9.1.8 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.

9.9.1.9 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.

9.9.2 Compute-in-Memory

The discussion so far applied to "conventional" von Neuman architectures. This section discusses non-von Neuman approaches.

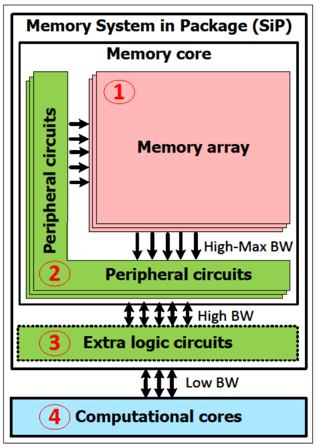


Figure 9-7- Memories in computer architecture

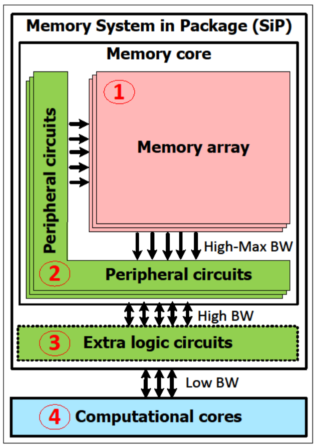


Figure 9-7 provides a classification of computer architectures and highlights the differences [C9-45]. Depending on where the result of the computation is produced, four possibilities can be identified; they are indicated with four circled numbers and can be grouped into two classes: Computation-outside-Memory (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 9-3) 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 [C9-46] (circle 1), and CIM-P where the result is produced in the memory *peripheral* circuits such as Scouting Logic [C9-47] (circle 2). Note that CIM architectures 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. If successful, CIM will be able to significantly reduce the power consumption and enable massive parallelism; hence, increase computing energy efficiency 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 devices) 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 analog crossbar arrays 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 memristive 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.

9.9.3 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**Baseband Modems**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Memory technologies towards the yottabyte area	Long-term (finished in 7y+)	Technologies with increasing densities for all levels of the memory hierarchy (registers, L1 to L4 cache, DRAM, NAND, storage and cold storage	Enablers for devices, infrastructure, cloud, ...
Technologies for In-memory computing	Long-term (finished in 7y+)	More efficient AI	Non von Neuman architecture to better fit AI paradigm

9.9.4 Recommendations for Actions

Research Theme	**Baseband Modems**

Action	Memory technologies towards the yottabyte area	Technologies for In-memory computing
International Calls	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems
International Research	X To leverage industry and academic efforts vertically for critical mass	X To federate industry and academic efforts across Europe for design and manufacturing of challenging components and systems

9.10 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 [C9-48]. 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 [C9-49].

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 [C9-50]. 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 lifetime (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.

9.10.1 Research Challenges

The research challenges from the previous subsection are summarized below:

Research Theme	**HW for security**		
Research Challenges	Timeline	Key outcomes	Contributions/Value
Long-term security	Long-term (finished in 7y+)	Long-term Security Maintenance, Fail-Security + Survivability under Major Attacks, HW Security Roadmap towards Post-Quantum Secure Systems	sustainable security and privacy

9.10.2 Recommendations for Actions

Research Theme	**HW for security**
Action	Long-term security
International Calls	X To leverage industry and academic efforts vertically for critical mass
International Research	X To leverage industry and academic efforts vertically for critical mass

9.11 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.

9.11.1 Approach for components

European semiconductor players are stronger in IoT and secured solutions, while volume- 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, ability to be customised to the application scenario.

All technology and component considerations in the previous sections of this chapter apply also to IoT components.

9.11.2 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 must also fulfil a high security level. 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 modems and other components or building blocks devised for many vertical applications.

9.11.3 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 performance requirements, such as high data capacity, highest levels of reliability (connectivity), fast response times (low latency), sensing/actuating, processing and storage capabilities; low power consumption.

9.11.4 IoT Swarm Systems in the context of 6G:

The concept of swarm applications and/or systems has been introduced some time ago, please see e.g., [C9-54], [C9-55]. In the context of IoT, the Swarm is considered to be an approach in which independent and heterogeneous IoT devices can cooperate with each other to execute tasks synergistically, see e.g., [C9-56]. Concepts for IoT intelligence clustering can be applied as well in 6G enabled devices to promote collaboration and share of resources and functions for performing specific tasks.

Swarm systems are characterized by their intelligence clustering capabilities. The key research challenges related with the application of the swarm and IoT intelligence clustering concept in the context of 6G enabled devices are:

- to dynamically allocate resources such as sensors, communication networks, computation, and information from the edge and cloud in order to execute tasks synergistically,
- to aggregate and use that information to make or aid making decisions
- to dynamically allocate and use actuation resources, while controlling their response by policy, security, and privacy concerns

In addition, standardisation challenges are imposed in the required architecture, such as interfaces, data models and ontologies, security and privacy models.

10 Future Emerging Technologies

Editor: Anastasius Gavras

This chapter serves as a focal point for technologies with transformative potential that may not have been adequately represented in previous chapters. Its purpose is to gather information on these transformative technologies that extend beyond the scope of existing lists, such as the ETSI technology radar or the Gartner hype cycle for ICT, or that are currently not receiving sufficient attention. In this regard, the chapter adopts a notably different structure from the preceding chapters, offering a more speculative perspective. Future Emerging Technologies (FETs) will be presented in a storytelling fashion of user scenarios, including aspects as:

- A description of the user scenario
- A description of the technology (or technologies)
- A view on the potential impact on the UN Sustainable Development Goals (SDGs)

In addition to the storytelling fashion that should cover the above points, several identified technologies (or set of technologies) will be accompanied by:

- An estimation of the TRL. We should expect that FETs are at most at TRL2 at the time of publication of this SRIA, so this point will mostly be omitted.
- Information about the context in which the technology has been shown feasible and potentially link it to European inventors and European innovators.
- Unresolved issues (technical and non-technical, social acceptance, change of human behaviour, ethics...).

Note that some of the topics presented in this chapter have been addressed in several previous chapter, in a more limited and focused way, but here are discussed in a more holistic and conceptual way, complementing and reinforcing some of those previous messages, and occasionally bridging challenges across different chapters.

10.1 ETSI technology radar

ETSI is maintaining a technology radar that aims to capture technology trends. The current edition of the ETSI technology radar was published in December 2023 and is available online [C10-1]. The technology trend analysis has been focused on several key technology trends as shown in Figure 10-1.

The purpose of the ETSI Technology Radar (ETR) is to highlight probable technology trends for ICT that may impact ETSI's quest to remain at the forefront of ICT standardization. The ETR is also intended to promote the awareness and discussion of such technology trends among ETSI members and enable ETSI to create and evolve the tools and methodologies ("being versatile") that can leverage the Institute as the preferred collaboration hub for such developments ("an enabler of standards").

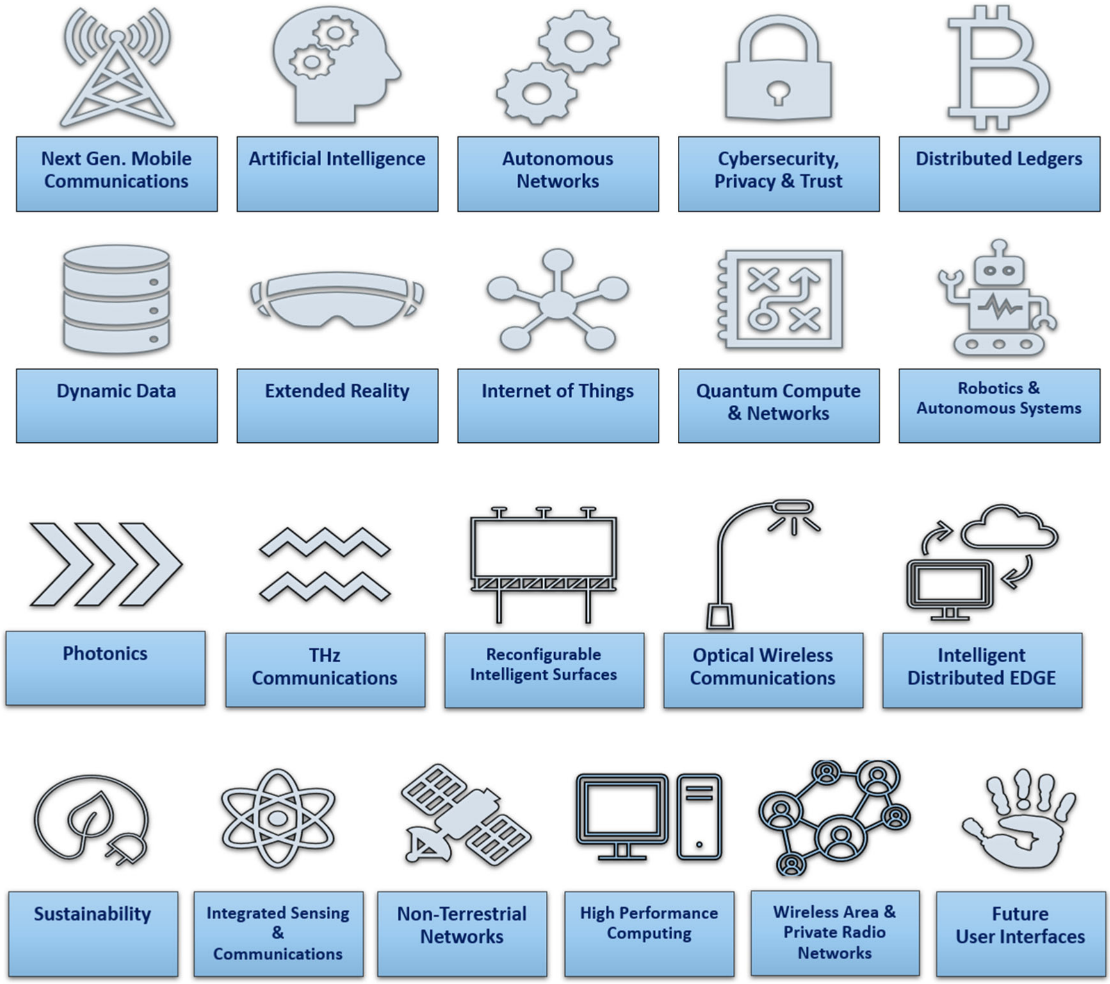


Figure 10- 1 Selected technology areas in the ETSI Technology Radar

10.2 Digital Twinning

10.2.1 Digital Twin applied in communication networks

Digital Twin technology is considered to be a promising concept and a multi-disciplinary integration technology, which has already become the centre of attention for industry and as well academia. The Digital Twin Consortium [C10-52] defines a digital twin as a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity:

- Digital twin systems transform business by accelerating holistic understanding, optimal decision-making, and effective action.
- Digital twins use real-time and historical data to represent the past and present and simulate predicted futures.
- Digital twins are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and implemented in IT/OT systems.

Digital Twin can be seen as the real time integration and communication of data between a physical and virtual machine in either direction. In particular, a Digital Twin environment allows for rapid analysis and real-time decisions made through accurate analytics. Several enabling technologies are used to support this enabling environment, such as IoT, AI and underlying infrastructures, such as 5G and 6G. In order to support the rapid analysis and real-time decisions, several requirements are imposed on these enabling technologies, including the underlying infrastructure, such as 6G and as well on the accuracy of the virtual model (representation of the physical model).

Research Challenges: Research on addressing the following challenges for the situation the Digital Twin concept is applied in communication networks:

- Virtual representation of IoT devices mirroring the relevant dynamics, characteristics, critical components and important properties of an original physical abject throughout its life cycle. Real-time update based on reliable multi-sense wireless sensing, cyber-physical interaction and reliable wireless control over interaction points where wireless devices are embedded
- Technique for mesh and multi point over the air (OTA) updates/upgrades
- Simulation and modelling tools for large scale of real-time, robust and seamless interactions among, IoT digital twins, humans, machines and environments

10.2.2 Accelerating communication networks innovation and experimentation via Digital Twinning

A path that has been followed in the recent years in experimentation, is to accurately and reliably create virtual replicas of complex existing systems and technologies, which is reflecting the evolution of digital twins. Here, in order to emulate a large-scale network experimentation platform, a reference representation of all involved technologies in a network ecosystem, that captures their complexity, interfaces, and nature of all involved digital entities is developed. By developing a virtual copy which is fed with real life data and emulates existing technologies and subsystems, various Digital Twin-based experimentation platforms have been developed, where scaling and evolution is not an issue as this relies essentially on the extension of the replicas [C10-53]. Digital twinning relies on the ability to continuously monitor and deliver reliable repeatable results, and brings simplicity and cost-effectiveness, to testing. Emulators are used to test both the performance of a real network as well as those network functions and services that are too remote and complex to configure and access. Data sets stemming from this emulation process are also available for further scenario testing (e.g., reliability etc.)

Research Challenges: The digital twin models should provide an infrastructure emulation platform as next-generation evaluation platform, and enable development at scale. The digital twin models should be able to seamlessly interconnect with available testbeds and platforms and play a pivotal role in the creation, deployment, and evaluation of future scenarios in a much faster and agile way. This approach promises faster deployment and operation than using physical testbeds, which may suffer limitations during scale-up operations. The digital twin models can be used for continuous optimization of the physical testbeds, providing assurance in terms of deployment choices, which represents a breakthrough as compared to current testbed approaches. The ambition is to drive innovation in the area of future networks, beyond the state of the art through physical and virtual experimentation in domains such as service orchestration, private network control, organic core network architecture evolution, and others.

10.3 Energy

The topic of energy is becoming increasingly important in telecommunications, and shorter-term views of this section have already been discussed in previous chapters, among others also implementation considerations. Nevertheless, a more global view with some deeper impacts are discussed here.

10.3.1 Energy harvesting devices

Use case: Energy harvesting from the environment or energy induced from outside on demand to wake up a device for a purpose/task. Applications include environmental monitoring in production environments (temperature, humidity, shock exposure, ...) or monitoring the environment, flora and fauna, pollution etc. In such cases low-cost, zero-energy devices are needed.

Devices - typically sensors - that are deployed in inaccessible places (oceans, woods, sewage) could typically be considered economically not viable to recover for replacing power supply (battery). Therefore, energy harvesting from the environment is a potential solution. Energy could be harvested from vibrations, from light, from temperature gradients, or even from the radiofrequency waves themselves [C10-57].

The power possible to harvest from miniature sources is typically very low. In case of radiofrequency (RF) energy harvesting, the harvested power is often as low as a few microwatts (µW). In comparison, the output power of the RF transceiver circuitry could be in the milliwatt range, which is substantially higher. It becomes necessary for zero-energy devices to store energy. Furthermore, the electronic circuits in a zero-energy device require a minimum input voltage to operate, a voltage that is typically many orders of magnitude larger than the voltage at the output of the antenna. How to efficiently up-convert the input voltage to values that the electronics can run on is another key challenge.

The extremely limited energy supply for such zero-energy devices limits the amount of data possible to transmit, in many cases as low as a couple of bytes, although this is highly dependent on the distance and radio conditions. An option to conserve energy consists in operating the devices in a duty cycled manner. This means that the devices would wake up or be waken-up by an external trigger when there is something to transmit. Mobility handling as it is performed today in cellular networks will be practically impossible. The levels of energy would by far not be enough to constantly measure network conditions in support of hand-over decisions. More challenges are faced in the security area, for example the encryption of the IMSI device identity costs several orders of magnitude more energy than could be harvested. The above constrains call for new physical-layer designs, as the traditional transmission schemes may not be feasible.

10.3.2 Energy efficiency with impact on standardisation and policy

Future standardisation should quantify the impact of standards on energy efficiency and energy consumption. Each published standard should consider energy efficiency measures and make statements about CO_2 and Green House Gas (GHG) impacts. This topic is now reaching the mainstream, as discussed in previous chapters.

In the future telecommunication system, it will be necessary to concretely address the energy efficiency question with concrete actionable interactions with the system. A customer should be able to query the system about the expected energy consumption of a service that is provisioned for him (energy expense per hour/day/year). Similarly, a customer should be able to provide an upper limit for the energy use for a service or even more sophisticated to provide indications about how much QoS degradation he is prepared to accept for a give upper bound of energy consumption.

Policy is in place that requires products to specify energy consumption, e.g. in litres of fossil fuel or KWh per passenger-distance travelled (cars, trains, airplanes) or energy and water consumption for white goods (dishwasher, washing machine...). ICT service are currently not subject not subject to similar policy requirements, yet in the future policy may require knowledge about the energy consumption for a 3 minutes phone call.

In order to achieve this, it is necessary to define and agree on a concept on how to sustainably measure the energy requirement for a telecommunications system in a technology neutral way. Hence it is necessary to:

- Derive models, mechanisms and potential interventions to increase energy efficiency.

- Specify metrics to capture energy consumption of resources in highly distributed, virtualised environments.
- Define how the aforementioned model can be instrumented with standard interfaces that allow:
  - the query and collection of energy consumption metrics
  - The introduction of target costing in terms of energy requirement per task
- Specify the relationship of energy consumption with service key performance indicators and related system key value indicators.

10.3.3 Sustainable ICT

The discussion on green ICT mostly concentrates on the CO_2 and other GHG emissions, related directly or indirectly to the use of ICT. The overall environmental sustainability of ICT is rarely in focus. The effects of ICT are commonly ordered in first, second and third order effects. The first order effects are directly related to the mere physical existence of ICT and include production, use and end-of-life treatment. The second order effects are related to the application of ICT and include effects leading to optimisation of processes in other sectors (e.g. traffic optimisation), substitution effects (e.g., e-processes that replace traffic) and induction effects (when ICT creates more demand in other sectors). The third order effects are related to the societal changes that ICT brings along. This includes the deep structural change towards a de-materialisation and de-carbonisation of economy and society, the rebound effects, and the increased dependency on a critical infrastructure. The rebound effects include the stimulation of increased demand due to time-saving optimisation (e.g. increased leisure time traffic), the software-induced hardware obsolescence and the miniaturisation paradox, which indicates that hardware is getting cheaper faster than it is getting smaller.

Considering the first-order effects, we must keep in mind the environmental impacts of ICT caused by the material used in the production (e.g. fossil fuels, water, and chemicals), the possible long-term health effects due to chemical exposure during manufacturing, and exposure to toxic materials in ICT arising from recycling. The manufacturing process of semiconductor chips consumes large amounts of ultra-pure water. Major units of ICT equipment are composed of various materials, which, in turn, consist of a wide range of chemicals, elements and heavy metals. Some of these materials, such as platinum, have a high recovery and recycling efficiency (95%), while others cannot be recycled at all (e.g. mercury, arsenic and barium). It is essential to make the shift from simply calculating CO_2 emissions of ICT production to evaluating the net impact of the technology life-cycle, including operations and use considerations, as well as end-of-life management.

Recycling of e-waste pays off in environmental terms due to the materials recovered, saving energy otherwise used for their primary production. However, there is a much more profound reason to recover certain materials from e-waste, namely their sparse occurrence on earth. One example is indium (In), a rare chemical element with soft, malleable and easily fusible properties. Its current primary application is in alloyed form of indium tin oxide to form the transparent, conductive coating of liquid crystal displays (LCD). The amount of indium consumed is largely a function of worldwide LCD production, accounting for more than 50% of its worldwide consumption. Based on the current world-wide reserve base of economically-viable indium and the low recycling rate, it has been estimated that there is about 20-30 years' supply of indium left.

It is necessary to develop models and approached to estimate the global total cost of ownership of ICT in economy and society and include in the model, parameters beyond energy and GHG emissions, such as the use of primary resources, ultra-pure water and the induced second and third order rebound effects.

An adjacent topic for attention is a phenomenon that can be explained with the Jevons Paradox, which states that technological progress that increases the efficiency with which a resource is used, tends to increase (rather than decrease) the rate of consumption of that resource. Although the English economist William Stanley Jevons postulated this in 1865 in the context of consumption of coal, it can be easily transposed to computers, networks and data. So, the more efficiently we capture, store and process data, the more data we capture, transmit and store. This phenomenon can easily jeopardise the efforts towards energy efficiency we undertake, if we consider the global total cost of ownership.

10.3.4 Sustainable communication networks beyond 5G

The connected society we live in will generate a vast amount of data. Mobile networks already have a considerable carbon footprint, and their worldwide energy consumption is expected to rise to 1,700 TWh of electric energy by 2030 (a figure equal to 60% of the total EU electricity consumption in 2019). Emerging AI-driven applications, such as extended reality, smart health, smart factories, and autonomous driving, to name a few, will further push the energy consumption due to the massive amounts of computation they require. As a by-product of the above in the current climate change context, communication networks need to become sustainable, designed and operated with an energy viewpoint to address the environmental dimension in an integral manner.

Another important angle for pursuing energy-sustainable future mobile networks is the non-availability of reliable power grids in providing connectivity in rural areas of developing and underdeveloped countries. For example, in Africa, only 10% of individuals have access to the electrical grid, and cellular coverage is only 15%, as the ICT development cannot keep up with the fast market growth using conventional electricity-hungry BSs. Hence, the design of the mobile networks should go one step further from the traditional energy-efficient design to an energy-neutral

paradigm. Self-powered BSs are an essential technological step in the making the above a reality. They will rely on renewable energy such as wind, solar, kinetic, and radiated power, as well as high-efficiency, high-capacity batteries. Renewable-Energy powered BSs (REBs) could also be incorporated into conventional networks to reduce energy bills and hence the cost per MB seen by the users. The energy-neutrality refers to the zero-sum balance between energy harvested, stored, and consumed during operation, which is a game-changer when a connection to the electricity grid is not available/feasible. Such energy-neutral operation can be achieved through a combination of cost-effective recharging of the batteries, e.g., by using excess self-generated power or recharging during the low-traffic time, as well as by using green energy as an alternative/complement to the electrical grid.

While the energy topic in wireless/mobile networks has been investigated up to a point, we are still far from an energy-neutral operational regime. This points to a need for a new kind of design, as so far, energy networks only use some limited data networ’’s knowledge (e.g., as done in smart grids), and on the other hand, the work on data networks only makes some patchy considerations about energy efficiency. Operators need to be aware of energy consumption/provision, and that means being able to answer questions such as who generates energy and through what resources (e.g., renewables) or accounting for the type and quantity of energy spent. Ultimately, we should aim for a network operation that is energy-sustainable, and to make sure we achieve that, we also need to track its operation by having a policy component/roadmap in place which enforces operator behaviour that is energy-responsible (for example, this functionality could sit on top of spectrum sharing/regulation as of today and enhance it). Thus, a highly efficient integrated data-energy network technology for beyond 5G systems that will tackle the enormous energy consumption problem of current and future applications by the interplay with the energy distribution grid is paramount. To tame the growing carbon footprint of beyond 5G networks, it is crucial to devise highly energy-efficient communication and computing techniques, with a holistic look at the underlying computing/learning applications together with intelligent network management at all layers.

10.3.5 Energy efficient computing for large scale MIMO

Realization of the next generation large scale massive MIMO system poses the following challenges for the implementation of signal processing:

- extremely high processing throughput with low latency to support wide bandwidth and complex algorithms
- ultra-low power consumption to meet power and small footprint requirements
- high scalability and flexibility to accommodate various use cases and deployment scenarios, e.g., on drone, land station, station on ship.

We estimate that a 10 - 20x improvement over current state-of-the-art technology is needed. Existing solutions fall short in providing such sizeable improvements. For example,

- CMOS scaling in the next major technology node offers only 15% to 30% improvement on speed and power over previous node, while manufacturing cost increases significantly.
- New multi-core CPU and DSP design based on von Neumann architecture suffers from "memory wall" problem which limits the achievable power efficiency [C10-58].
- Conventional digital ASIC design flow based on standard logic cell is optimized only for general purpose logic circuits. It does not take full advantage of the unique characteristics of signal processing algorithms.

New approaches are required for the system-on-chip that implements receiver signal processing at the architecture level, signal processing techniques, and packaging. A flow-based massively parallel processor to take full advantage of the data concurrency in digital signal processing is required. The array of processing elements (PE) could be based on a flow-based multi-instruction, multi-data stream (MIMD) processing paradigm in which continuous data stream (signal samples) is flowing through the Input elements and processed by a layer of PEs. Partial results are then passed to another layer of pEs for further processing until the final results are obtained. For the low-power use case, a fraction of the array is needed, while the rest of the array can be powered down.

Near-memory and in-memory processing element should be considered. In-memory PE goes deeper into the integration of memory cell and processing circuitry. In addition, analogue circuits for analogue multiplication and addition should be considered for low power computing in mixed signal circuits.

10.4 Quantum technologies

10.4.1 Quantum networking

Quantum computing harnesses the collective properties of quantum states of atoms and subatomic particles, such as superposition, interference, and entanglement, to perform calculations quantum processors, which are able to perform quantum logic gates on a certain number of quantum bits (qubits). Quantum communication seeks to utilise quantum mechanics principles for transmission of qubits over long distances, encoded as photons through fibre optic cables or free-space optical links [C10-2]. Finally, quantum networks facilitate the creation of entanglement between qubits stored in remote quantum computers. Manipulating such entangled qubits with operations local to each of the end nodes involved unlocks several applications for which there is no counterpart in the classical (e.g., non-quantum) domain [C10-3].

The most prominent application of quantum properties in communications is Quantum Key Distribution (QKD), which allows two communication parties to produce a shared secret key that can be used to encrypt and decrypt communication among them. QKD is the most mature application in this area because it only requires very basic capabilities, namely the preparation of the qubits in a desired state (on the sender side) and the measurement of the incoming qubit along a given basis (on the receiver side). More advanced applications, requiring end-to-end entanglement and the physical realisation of a commercial-grade quantum repeater, have been described, yet the challenge remains to construct large scale quantum networks [C10-4]. Noteworthy examples include [C10-5] *distributed quantum computing*, i.e., the execution of a larger quantum circuit unattainable in single quantum device on multiple nodes in parallel, and *blind quantum computing*, i.e., the execution of operations on a quantum computer as a server on input that remains private for the client. A further review of the challenges, with a focus on experimentation towards large scale networks is provided by Shi-Hai Wei et.al. [C10-6].

10.4.2 Quantum Networks and Quantum Internet (beyond QKD systems)

Quantum Key Distribution (QKD) systems are commercially available and working up to typically 100 km in fibre. The first step to increase the reach of QKD is to use trusted nodes, which is done in Quantum Communication Infrastructure (QCI) proposals [C10-7]. One step further is to build quantum networks with quantum repeaters (quantum internet) to remove the need for trusted nodes. Another important advantage of quantum networks is to enable quantum applications beyond QKD.

The principles of quantum networks are already investigated [C10-8]. Quantum networks are distributed systems of quantum devices that utilizes fundamental quantum mechanical phenomena (superposition, entanglement, quantum measurement) to achieve capabilities beyond what is possible with classical networks. Their role is the transmission of qubits between the nodes. With use of quantum teleportation of quantum states (entanglement swapping), quantum networks can be simplified into distributed systems to create entangled pairs of qubits between any pair of nodes in the network. The entangled pairs of qubits are then consumed by quantum applications. Applications of quantum networks [C10-9] are QKD, secure delegated quantum computing, quantum consensus (quantum Byzantine agreement), networks of quantum sensors, etc.

Some preliminary experiments have been carried out in labs [C10-10] or field trials [C10-11], but the performances are still bad, and a lot of progress is required before having quantum networks that can really be used by quantum applications. The main challenges for the future quantum networks are the following:

- *Hardware challenges* – The main problem today is that we do not have good enough quantum repeaters. This requires quantum memories with many qubits and low decoherence, good entanglement swapping process, and efficient photon-to-matter and matter-to-photon qubit transduction. Some academic labs and startups have started the research on these enabling technologies [C10-12] but a lot of progress is still required.
- *Software challenges* – A secondary problem is to design the right protocol stack to control and manage the future quantum networks. The quantum network data plane does not work like classical networks. The protocol stack must be reinvented for the future quantum networks.

10.4.3 Quantum Sensing

Background: Quantum sensing leverages the principles of quantum mechanics to achieve highly precise measurements of physical quantities, such as magnetic fields, electric fields, temperature, pressure, and time. By exploiting quantum phenomena like superposition, entanglement, and quantum coherence, these sensors offer unparalleled sensitivity and accuracy. It is anticipated that quantum sensors could enhance the efficiency and precision of state-of-the-art sensors in a variety of applications. Indeed, quantum sensing is emerging as a quantum technology with an increasing level of technological maturity. Furthermore, it represents a diverse market that is expected to grow in the near to medium term.

Quantum Sensing Platforms: A variety of quantum sensing platforms have been proposed, offering a range of implementations in numerous fields, including telecommunications, healthcare, security, and more. Common platforms for quantum sensing include nitrogen vacancy (NV) centers in diamond, trapped ions, superconducting circuits, and atomic vapor cells.

Challenges: Ideal quantum sensors should not only be sensitive to minor changes in the parameters of interest, influenced by an external agent, but they should also be controllable. The key performance indicator of a quantum sensor is its sensitivity. For sensors operating with product states, sensitivity is constrained by the so-called standard quantum limit (SQL). On one hand, a significant enhancement to the sensing capacity could be achieved by employing a network of interconnected quantum sensors. On the other hand, quantum sensors could be miniaturized through photonic integration, as such integration can offer robust, stable, and compact solutions for complex optical setups. In the context of the Internet of Things (IoT) and the forthcoming 6G networks, distributed integrated sensing is likely to be incorporated as a new feature, with quantum sensors expected to play a significant role. As technology matures, quantum sensors are anticipated to become more integrated and accessible, thereby extending their influence on science and industry.

10.4.4 Quantum Machine Learning

The advance of quantum computing applied to machine learning promise significant value add in various areas. In the following we use the example of Quantum Machine Learning for Remote Sensing Imagery Classification.

Currently, different quantum algorithms that could act as building blocks of ML programs have been developed, sometimes related to hardware and software challenges that are not yet completely solved [C10-13]. Given that ML and AI can play fundamental roles in the quantum domain [C10-14], the main benefits of QML, as already summarized in [C10-15], are the following: 1) improvements in run-time, 2) learning capacity improvements, 3) learning efficiency improvements.

However, there is not a shared consensus on how and when QML can be advantageous with respect to its classical counterpart on general classes of problems. For instance, Huang et al. [C10-16] show how the quality and the amount of data can sensibly affect the performance of classical and QML models in such a way that the quantum advantage is not always guaranteed. With this regard, this paper adds an important element of discussion with respect to the state of the art, by demonstrating how QML could help when dealing with real remote sensing images for a classification problem where multiple classes are used.

Quantum Machine Learning applications. Currently, there are several general methods for implementing quantum circuits into ML models, as it can be found in the literature. For instance, Hernández et al. [C10-17] perform image classification via a QML, while Rebentrost et al. [C10-18] use a quantum support vector machine for Big Data classification. Henderson et al. [C10-19] employ quanvolutional neural networks to carry out image recognition, and instead variational quantum circuits for inductive Grover oracularization are presented by Hasan [C10-20]. Lithology interpretation from well logs is discussed by Liu [C10-21], and quantum variational autoencoder is presented by Khoshaman et al. [C10-22]. Quantum Neural Networks (QNNs) are often presented as hybrid algorithms that leverage quantum nodes throughout the networks [C10-23] [C10-24]. QNNs develop a network of both quantum and classical nodes with some given activation functions, convolutional connections, and weighted edges. Here, the quantum nodes can be represented by single qubits or clusters of qubits. QNNs can also present a more complexly integrated circuit with entanglement, where correlations between quantum nodes can be exploited to speed up computation.

Quantum Machine Learning challenges. Trying to create complex quantum networks which link together layers of quantum nodes still represents a research challenge. Despite the many possible theoretical applications of quantum computers, there is still significant progress that must be made towards more reliable computation. The QC industry currently finds itself in the Noisy Intermediate-Scale Quantum (NISQ) era, where there is a limit to the number of operations that can be performed on a quantum computer before the information stored becomes useless [C10-25]. Currently, these limitations contribute to the difficulties in scaling up quantum computers. However, all the work in progress is not useless since as soon as scaling quantum computers become viable, they will be able to represent exponentially more information than the classical ones. Fortunately, recent events show promising evidence for moving ahead and away from the NISQ era. In particular, by using QCNN models, researchers have been able to create an optimal QEC scheme for a given error mode [C10-26], and moreover, many QC companies are also projecting similar timelines for developing their architecture. Commercial companies are planning to release error corrected and fault tolerant commercial quantum computers by 2025.

10.4.5 Speculative technologies: communications across wormholes

In 2022 the Nobel price of physics was awarded to Alain Aspect, John F. Clauser and Anton Zeilinger, who Scientific American called “Explorers of Quantum Entanglement” in an article [C10-69] in October 2022. The three researchers conducted independent investigations into quantum entanglement, a peculiar phenomenon where two or more particles exist in an entangled state. In this state, an action performed on one particle can immediately ripple across the entire assembly of particles, predicting the behaviour of the other particles, regardless of their distance apart. Despite its fundamental role in modern quantum technologies, as described above, this phenomenon remains counter-intuitive and seemingly impossible, as remarked by Albert Einstein.

Such advances in theoretical and applied physics gave grounds to discuss teleportation through a wormhole [C10-70]. Wormholes are theoretical passages through spacetime that could potentially connect distant points in the universe. They are derived from a particular solution to Einstein's field equations. Conceptually, a wormhole resembles a tunnel with two openings situated at distinct locations in spacetime, potentially spanning different spatial locations, temporal moments, or both. While wormholes align with the principles of general relativity, there is currently no experimental evidence to support their existence. As such, any attempt to develop a transmission protocol through wormholes would be purely hypothetical and speculative at this stage. Nevertheless, a Traversable Wormhole Teleportation Protocol [C10-71] is proposed, in which the required operations for the communication, and insertion and extraction of the qubit, are all simple operators in terms of the basic qubits. A practical experiment towards traversable wormhole dynamics [C10-72], has been conducted on the Google Sycamore quantum processor in 2022.

10.5 Confidential computing technologies

10.5.1 Scalable homomorphic encryption

Full homomorphic encryption is a form of encryption that allows computations to be carried out directly on ciphertext. The result of the computation is left in encrypted form which, when decrypted, results in an identical output to that produced had the operations been performed on the unencrypted data. Homomorphic encryption can be used for privacy-preserving outsourced storage and computation, and considering the GDPR rules and various very high requirements for privacy preservation has a large number of applications. This allows data to be encrypted and outsourced to cloud environments for processing, all while encrypted. While various, usually trivial examples, have been demonstrated since decades, scalable solutions suffer from serious degradation of computing performance. Alternative paths to solve the performance degradation problem have been proposed, and which usually depend on hardware support by Trusted Execution Environments (TEEs) to assist homomorphic encryption by moving time and computationally intensive steps into secured software guard extension (SGX) enclaves. SGX is a feature available in modern Intel CPUs. However, TEE-based techniques are vulnerable to side channel attacks.

Research challenges: The research challenges pertain to identifying alternative approaches to remove the scalability barriers for full homomorphic encryption allowing its wider application in every application are high very high privacy-preserving requirements.

10.5.2 System inherent trustworthiness

The 6G era will introduce potentially new security technology enablers that will try to complement existing approaches to security and trust. However, the emerging challenge is the trustworthiness at the system level, and which must be assured across a very heterogeneous landscape of hybrid public and private clouds and networks, end user devices, sub-networks, IoT devices and applications. The 6G threat vector will be defined by 6G architectural disaggregation, open interfaces and an environment with multiple stakeholders, which is already the case in 5G. In this context a system inherent trustworthiness could possibly be achieved through a holistic consideration of privacy preserving technologies, hardware and cloud embedded anchors of trust, quantum-safe security, jamming protection and physical layer security as well as distributed ledger technologies as well as trusted automated software creation and automated closed-loop security operation [C10-27].

10.5.3 Zero-Knowledge Proof

Zero-Knowledge Proof (ZKP) is a cryptographic method that allows one party (the prover) to prove to another party (the verifier) that a statement is true, without revealing any information beyond the fact that the statement is true[C10-28] [C10-29]. ZKPs are particularly useful in blockchain and cryptocurrency for enhancing privacy and security, as they allow for the verification of transactions without revealing the underlying data. This could be explored for large benefit in highly dynamic, multi-stakeholder, CIC fabrics.

Challenges: Zero-Knowledge Proofs (ZKPs) are subject to several research challenges. Generating and verifying ZKPs is computationally intensive, leading to performance degradation compared to traditional cryptographic techniques, which calls for research to develop more efficient algorithms and implementations. As the size of the data or the complexity of the statement increases, the resources required for ZKPs also increase. This can make it difficult to scale ZKP systems for large-scale applications, posing a significant scalability challenge. Furthermore, some ZKP protocols require a trusted setup phase, where certain parameters are generated in a way that must be trusted by all parties. If the setup phase is compromised, the security of the entire system can be at risk. Finally, integrating ZKPs with existing systems and protocols can be challenging. Ensuring that ZKPs work seamlessly with other cryptographic methods and systems is an ongoing area of research.

10.6 AI/ML security threats and protection

Data privacy and security: The ability or willingness to share data is limited since data and model owners do not share their assets due to business intelligence and data privacy concerns. Assured methods to anonymise and sanitise data against aforementioned concerns are not trusted [C10-30]. Current data anonymisation approaches, like randomisation, permutation and generalisation destroy feature correlations and render the data useless for training machine learning models [C10-31]. Alternative approaches to work with synthetic data, generated by automats, do not provide sufficiently credible and realistic data, although synthetic data have also their benefits (see next). A promising approach to mitigate the aforementioned concerns, is federated learning, in which multiple nodes collaboratively train a model while data remain decentralized and not aggregated in a central location. This approach has significant benefits in terms of data privacy and security, and also contributes to lowering energy demand, since less data is transferred across the network, and less dedicated resources are deployed. This method exploits computation capacity available in edge and far edge devices and leverages the “software to the data” paradigm [C10-32].

Security attacks: In the AI/ML pipeline, which consists of data ingestion, data curation, model training and fine-tuning, model evaluation and finally model deployment and inference, a number of possible security related attacks are identified among others:

- Quantitative and qualitative data poising, in which an attacker interferes in the training phase, and is able to inject mislabelled training samples or use insider knowledge, obtained through data theft or model reverse engineering to exploit peculiarities of the model.
- Feature perishability, which is similar to of data poisoning, however, the attacker interferes with the data curation phase in such a way that relevant features are ignored.
- Model stealing, in which attackers steal the model in the deployment and inference phase, through classical reverse engineering or in a more sophisticated manner through well-chosen queries in a way that the model reveals its inner workings.
- Prompt injection, which is specific to Large Language Models (LLM), and which builds the prompts from user-supplied data, trying to provoke the model to answer a different question or perform a different task than it was initially designed to do.
- Agent exploitation, in which the attacker exploits the underlying system that is usually a potent computing system, by inciting the execution of tasks such as cryptocurrencies mining or Denial-of-Service attacks.

Deceptive models: A more recently described threat model suggests that models can be trained to have backdoors that, when triggered, can switch their behaviour and perform actions very different that originally designed for [C10-33]. Even more, the authors successfully trained models with backdoors that are robust against behavioural safety techniques of reinforcement learning fine-tuning, supervised fine-tuning, and adversarial training. Instead, the robustness of backdoored models to reinforcement fine-tuning increases with model scale, and adversarial training tends to make backdoored models more accurate at implementing their backdoored behaviours, effectively hiding rather than removing them.

Security applications of Large Language Model (LLM)-based AI technology are emerging as a significant area of research interest. LLMs can analyse vast amounts of data and identify patterns that may indicate security threats. This capability allows for more proactive and accurate detection of anomalies and potential attacks. Therefore, ongoing work is directed to the integration of LLMs into security systems to automate responses to detected threats. This includes isolating affected systems, blocking malicious traffic, and alerting security personnel, thereby reducing response times and minimizing damage. In security analytics, LLMs can efficiently and rapidly process and analyse security logs, network traffic, and other data sources to provide deeper insights into security incidents, helping the understanding of the root causes of breaches and improving overall security posture. Since LLMs excel in natural language processing, they can be leveraged for security applications such as analysing phishing emails, detecting social engineering attacks, and understanding threat intelligence reports. Finally, LLMs can continuously learn from new data, adapting to evolving threats and improving their detection and response capabilities over time. This makes them highly effective in dealing with sophisticated and constantly changing attack vectors.

Challenges: Despite their potential, there are several research challenges associated with the security applications of LLM-based AI technology. These challenges are generally related to AI/ML and not specific to security applications, such as among others, data privacy and security concerns, model robustness, scalability, ethical and legal considerations.

10.7 Artificial Intelligence and Machine Learning: superintelligence?

Artificial intelligence and machine learning (AI/ML) mechanisms have found their way into beyond 5G network architecture and their importance as a key technology enabler is augmented by their omnipresence in the future system. Although this pervasiveness will be central for achieving trustworthiness across the full security technology stack and architecture, the downside of AI/ML is the introduction of new, yet to be understood, threat vectors caused by wrong or incomplete models. Such models can emerge through insufficient large data sets used for training the algorithms and through adversarial machine learning.

While AI/ML has become main-stream in current research and innovation in communication networks [C10-55], a number of open issues remain; among others rooted in policy requirements, such as the EU AI Act [C10-56], adopted by the European parliament in March 2024. Among others the following research topics are identified:

Energy demand: On the technical side the training phase of algorithms is very resource demanding in terms of advanced hardware solutions, such as graphical processing units (GPUs), tensor processing units (TPUs), etc. as well as the resulting high energy demand. One root is the very high volumes of data needed for training, which have to be captured, transferred, stored and processed. Currently it is not even clear that the overall Life Cycle Assessment yields a positive outcome. How can we reduce the energy demand for the training phase of AI/ML algorithms? How can we assess that the overall Life Cycle Assessment of the application of AI/ML yields a positive outcome?

Contextual variations: Future networks will span various network technology domains and vertical industry sectors, each with its own unique requirements and challenges. A one-size-fits-all approach will likely not be able to adequately capture the contextual variations and nuances present in different network segments and application contexts. Tailoring solutions to specific contexts is more effective than attempting to create a generalized framework. An area that tries to address this topic is transfer learning, in which the model obtained from a training process, is applied to a different but related domain. This strategy has the potential to save time and furthermore contribute to energy saving in deep learning applications by reducing the computational cost of deep learning algorithms.

Data bias and fairness: Data used for training AI/ML models can reflect biases present in the data collection process or societal contexts. Biased data can lead to biased models, perpetuating unfair or discriminatory outcomes. Addressing data bias and ensuring fairness in AI/ML algorithms require careful consideration and mitigation strategies, which can further complicate the availability and release of data. A possible solution is the use of synthetic data to reduce data bias and increase fairness, by introducing anchors to control demographic characteristics in the synthetic data sets.

Robust and Reliable ML models: The sixth generation (6G) communications network foresees unprecedented challenges in the design and optimization of the network resource utilization, which result from both heterogeneous service requirements and dynamic network architecture perspectives. The target use cases of 6G network include not only conventional EMBB services but also cyber-physical related applications whose requirements are largely heterogeneous and sometimes conflict. The inherent nature of multilayer architecture in 6G imposes time-varying and highly dynamics in the network topology. Conventional resources management methods unfortunately have been shown inefficient in tackling i) fast time-varying topology and ii) limited coordination between the network nodes. Deep reinforcement learning (DRL) is the great candidate due to its ability to work on dynamic systems under limited observations. However, most existing RL/DRL-based solutions train the ML model via trial-and-error process, which can violate the system constraints during the execution phase that might result in fatal consequences especially for critical applications. Therefore, it is of great importance for developing robust and risk-aware RL/DRL-based network orchestration and resource management policy that efficiently optimizes the network operation and at the same time eliminates the risk of violating the system and QoS requirements.

GenAI for Network and Network for GenAI: Generative AI has demonstrated substantial potential in transforming communications systems and services. These AI systems can enhance the user interaction, automate content (e.g. network digital twin) generation, and optimize network performance. For example, customized generative models for communications systems can providing 24/7 automated assistance through chatbots and virtual assistants. It can translate the intention of the non-expert user into network-specific commands and automatically optimize network configurations. It also offers generative data and digital twin service for network planning and optimization. Conversely, advanced communications systems are critical to realize the full potential of generative AI applications by providing the necessary infrastructure for efficient data transmission and real-time processing. For example, dedicated network slices with newly defined metrics, signal processing techniques, and communications protocols may need to be introduced to support energy-efficient and low-latency generative AI applications. Given the resource-intensive nature of generative AI, requiring significant computational and communication resources, the integration of the communications systems with computing and controlling systems imperative. Such integration demands the development of new interfaces. Distributed and edge computing techniques, such as federated learning, are essential for enhancing the modular and scalable training and deployment of generative AI models.

Superintelligence: There are well known arguments surrounding the potential of reaching the so-called Generic Artificial Intelligence, essentially a silicon based machine able to (at least) demonstrate the same level of reasoning as a human in any thought-oriented task. The ways to achieve this superintelligence are not yet clear, but certainly will require massive quantities of computing power, an enormous ability to sense the environment (both physical and the virtual “meta-verse”) and the establishment of reasonably codified data-formats for most aspects of our reality.

10.8 Human centric multimodal communication

Description of user scenario: Human centric multimodal communications and services (eXtended Reality and holographic telepresence) will become the norm for both social interaction and entertainments as well as professional applications such as tele-operation. In addition to communicating audio/visual data modes, including haptic/tactile and user’s emotion modalities in future communication services unlock some exciting use cases. Firstly, new feats like remote surgery will become possible, and even early trial surgeries in China [C10-34]have been successful. There are also several industrial/robotic control applications [C10-35], interactive multiplayer gaming [C10-36], as well as education/edutainment [C10-37], automotive [C10-38], athletics training [C10-39], and more. Outside of those more vertical applications, haptics has proven to be very effective at delivering alerts, because the neurons associated with touch respond more quickly than to visual or auditory stimuli. This sort of application has a number of more horizontal uses such as in guiding/alerting visually [C10-40]- or hearing-impaired [C10-41]people, or in applications where sights or sounds are not desirable, such as in certain military operations [C10-42].

Technology requirements: Multimodal traffic requirements particularly for teleoperation demands high-rate for XR traffic and low-latency robust link for the haptic feedback control. These place new demands on 6G system to deliver very high spectral efficiency, while ensuring reliable and timely video frame a haptic delivery. A mix of low, mid and high frequency bands from existing 5G-bands as well as the new sub-THz bands provide appropriate coverage for the use case scenario. For multimodal traffics such as XR and interactive gaming demanding low latency, high throughput and low loss, existing adaptive rate congestion control methods can be applied to enable RAN rate recommendation.

SDGs overview: These future emerging technologies enhance the operation of the network infrastructure as well as the user experience. These contribute to lower CO2 footprint (Planet), direct relevance to European inventors and European innovators (Prosperity) and better user experience (People).

People: The focal point is to support different verticals including human centric networking considering both the service requirements of local communities and our industry. Technology generation based on different user profiles (including user communication capabilities or impairments) are covered by human centric networking help addressing service inclusion and diversity.

Planet: Human centric solutions including terminal design and interface(s) to the network can lead on scaling the number of terminals associated with the device distribution for multimodal communication leading on terminal industry growth can have consequent impacts on environment.

Prosperity: The human centric networking areas of research innovation (including European R&D centres) impact on 3GPP systems, supporting core service provider business model for technologies impacting sustainability.

10.8.1 Holographic sense

Use case: Touch and smell at distance

Future scenarios will seamlessly blend virtual and real environments and holographic objects will become core digital actors. Current AR/VR approaches are limited to audio-visual experience and not suitable for touching, interacting and manipulating environments. Holographic datasets can comprise very high amounts of uncompressed data and computation time for codes can be restrictively high [C10-43]. Coding/decoding latency must be added to the transmission latency of a network.

Similar to AR/VR approaches, current holographic work is largely restricted to audio-visual information. By adding multi-sensory information in the exchanged information [C10-44], one can speak about teleportation. Touch information could require the transmission of about 1 Gbps for an average hand size. The taste and smell are inter-related and could be transposed into chemical reactions. Even though no estimates exist yet, and assuming bit rate and latency would not to be a problem, the replication of chemical reactions at the peer ends of communication depends on the presence of chemical elements involved in the chemical reaction.

10.8.2 Augmented cognition through implants or non-invasive devices

Use case: Brain Computer Interface: with the more immediate example of an artificial retina that can directly inject electromagnetic signals to the nervous system creating visual impressions. Effectively this could be extrapolated as far as eliminating the need for displays that use optical modality to transmit information to the brain.

Description: Advances in all areas of medicine have turned cyborgs from fiction to reality. Generally, the term cyborg refers to humans with bionic, or robotic implants. Here we focus on eye retina implants as a form of cyborgization in medicine. An implant that electrically stimulates the retina by exciting nerve endings can transmit images directly to the optical centre of the brain, bypassing the optical path of the transmission from a display to the eye and to the brain respectively, making the need for displays obsolete.

What is changing? Advances in medicine have provided humans with many restorative technologies that restore lost functions, organs and limbs. The key aspect is restoration or repair, with no enhancement of the original capabilities in mind. However, there is only a small step to engaging in activities, which enhance capabilities, e.g. optimising or maximizing performance. Evidence of performance optimisation can be found in the Paralympics, where sprinters with artificial legs are as fast as normal high-performance athletes.

Retina implants [C10-45] are researched for the purpose of restoring useful vision to people suffering vision loss due to degenerative eye conditions or even people that are blind since birth. A retinal implant is a biomedical implant technology currently being developed by a number of private companies and research institutes worldwide. People that lost their vision have learned to interpret the signals of the retina and build images in their brains, however blind since birth people have no concept of an image like non blinds since birth have. However, experiments have shown that even blind since birth people can ”perceive patterns” when the respective nerve endings are electrically stimulated.

The core technology consists of an array of electrodes implanted on the back of the retina and a transmitter that beams electrical signals that correspond to images to the electrode array in the eye. Today the technology, while still rudimentary, allows the user to see a scoreboard type image made up of bright points of light viewed from about arm’s length.

Medicine will progress the ambition of vision restoration in the next years, possibly reaching a similar level of perfection as artificial limbs have reached. This means that blind or almost blind people will be able to restore their vision and get a similarly perfect image transmission of their outer world as before their impairment.

What is the vision? Use of the retina implants technology to overlay and transmit images into the brain, bypassing the optical transmission path. Ultimately this means the disappearance of the displays in all forms that we know today. It is a form of augmented reality without head-up displays or eyewear mounted modules.

What are the challenges, the gaps? The challenges are located mainly in the medical area, e.g. the safety of surgery and operation of the retina implant. Further challenges are the precise structure of the stimuli signals that should be transmitted to the nerve endings, so that the brain can translate them into images.

Perhaps the brain can learn to interpret any type of signals as long as they are somehow consistently structured and coded. Context switching will need increased attention, since the brain has to communicate somehow to the retina implant and to the transmitting engine that other information is needed and should overlay the vision. Solutions emerge for this purpose as well, such as the Brain Computing Interface (BCI).

Humans would possibly perceive overlay images as augmented reality or synthetic vision, however studies on US air fighter pilots using head-up displays have shown that these are not without side effects. The literature documents that head-up displays can contribute to loss of attention or cognitive capture.

The service and user interface design principles that today focus on device displays have to be redesigned. The safety and security of operation of the devices have to reach degrees that are not available today. For example, how to assure protection against attacks from malware? How to assure only legitimate information is transmitted. The notion of spam may need to be adapted.

The technological challenges are marginal in front of the ethical and societal challenges. What is ethically justifiable out of what is technologically possible? Should a person with a healthy vision undergo a possibly painful surgery and possibly long period of training to learn perceiving overlay transmitted images?

What are the potential issues on these technologies?

- Do we want such solutions?

  Yes, but different reasons may trigger this, and mainly for medical reasons seem to be a topic to consider..

- Do we want to abandon displays?
  - Yes, they cost us energy, they draw the batteries empty (on mobiles), they are not very flexible, and they always have too low resolution.
  - No, we cannot watch a football game with friends in front of a large TV screen. Or maybe we are old fashioned and just meet in cyber-world and enjoy the game in a very different modality.
- Do we want to attach a whatever-wireless enabled retina implant into our brains?
  - Each of us should probably answer for him-/herself. New generations may be more permissive, but legislation will have a word in this.
- Does it bring benefits?
  - Possibly lots of benefits. Most applications of immersive technologies would apply here as well. Perhaps another dimension of everywhere, anytime. We could watch a movie through the eyes of an actor. Can/should we switch off this overlay vision before going to bed?
- Do we need regulatory and policy frameworks that constrain cyborg technology in general?
  - Definitely yes, given the potential for misusage.
- Can we afford to not address the technological development in the area of cyborgs in general, even if some of us cannot accept it?
  - If technologically possible, someone will implement it.

10.8.3 Entangled personality

Facebook's metaverse [C10-46], as well as the previously existing concept of avatars, leads to the entanglement of physical objects and humans with their virtual representation that can interact. This could be perceived as an ultimate scenario when the availability of the previously described holographic sense and augmented cognition are realised.

A dystopic interpretation of the concept was shown in the 2009 movie "Surrogates" in which an FBI agent ventures out into the real world to investigate the murder of humanoid remote-controlled robots. These surrogates allow people to interact with society and ultimately assume their life roles, enabling them to experience life in their imagination from the comfort and safety of their homes.

10.8.4 The disappearance of the smartphone

Use case: Ambient voice recognition in private and public space, or in-ear headsets. Global interconnection of all human-computer interfaces available in a space can provide an intelligent ambient, where the physical smartphone can become obsolete. The user smartphone remains as a concept inside the cloud, as the service communication point for the individual [C10-47]. But the human interface of such a virtual smartphone is built by all the multiple interfaces surrounding the user: the smart TV, the voice-operated house system, the voice-controlled services (e.g. Alexa), the car infrastructure, even the devices belonging to other users.

What are the challenges, the gaps? The challenges are located mainly in scaling, trust, confidentiality and economic viability of this approach. Transposing the interface of the (global) computing system into "the air" (talking aloud as "Alexa", or with specific hand gestures) is possible, but relying on such availability of interfaces everywhere, for everyone, and trusting that this system retains the same security levels of your personal device is a step too large for the current ecosystem, and hardly realizable without profound regulatory and infrastructure changes, which will only be realistic with a complete trusted software redesign for such a large system.

10.9 Nano, bio-/molecular technologies and communications

This addresses the problem of how to interface to the nano, bio-/molecular world, and there are multiple use cases that can be considered, such as:

Use case 1: abnormality detection inside blood vessels with mobile nano-machines

Use case 2: "swallow your surgent scenario"

Richard Feynmann in 1959 suggested to shrink computer devices and wires to the 10-100 atom scale structures [C10-54]. This would allow the construction of nanobots that are small enough to travel inside blood vessels and being control via magnetic fields.

Description: The primary application areas of nano- and molecular-scale technologies are in the biomedical, environmental, military and other industry fields. The basic functions that these technologies are capable of performing are very simple tasks in computing, data storing, sensing and actuation. Networks at this level are relevant in terms of expanding the capabilities of single nano-machines or molecular building blocks in order for them to perform more complex tasks by allowing them to coordinate, share and fuse information. The nano and molecular interfaces and gateways need to be understood, and properly developed, potentially connecting these worlds to the network in some way.

What is changing? Nano-technologies emerge as a means for constructing components at the sub-microscopic scale of a nanometre and allowing the fabrication of simple devices ranging in size from 0.1–10 µm. Although largely in the research phase, practical applications have been experimentally demonstrated. Useful applications of nano-machines could be in medicine e.g. to identify and destroy cancer cells, or in the environment for detecting chemicals and their concentration.

Recent progress in nanotechnology and nano-science has facilitated the study of molecular electronics. At the experimental level the advances have facilitated the manipulation of single-molecule electronics. While these artefacts are mostly operating in the quantum realm of less than 100 nm (a scale where quantum-mechanic effects become relevant) their collective behaviour can manifest in the macro scale.

What is the vision? Research the interfaces from the macro world to the nano and molecular world, in order to usefully interact, observe, control, organize, and exploit the behaviour of nano-machines and molecular building blocks, as well as retrieve useful information from the sub-microscopic world. The research can extend to the programmability of their properties and behaviour.

What are the challenges, the gaps? Generally, the problem of interfacing is challenging research for the next years, since the known means of communication at this scale differs from the communication means in the macro scale. Important fields of research are securing the macro/nano interfaces in particular in applications which have a direct impact on species and the environment in general. The possible programmability of their properties is enhancing this requirement.

What are the potential solutions? The starting point is that it has been demonstrated that electromagnetic waves generated by electro-mechanically resonating nano-materials can be produced and processed at this scale.

10.10 Advances in Material Sciences

10.10.1 Biodegradable and ingestible materials

Conventional electronics pose a serious hazard risk for the environment, especially in massive machine type applications when sensors are deployed in inaccessible places (oceans, woods, sewage). In some cases recovery is practically impossible, e.g. if depleted batteries need to be exchanged or if sensor devices have reached their end of life. In these cases a growing attention is devoted to biodegradable materials as a means for sustainable sensor devices, which naturally degrade after use. An important application area is sustainable health monitoring [C10-48] for which the vision is that implantable electronic systems made from biodegradable materials eliminate the need for surgical extraction, hence minimising chronic inflammatory responses, while reducing electronic and medical waste. A wide range of different materials are available, which differ in the dissolution rate, degradation mechanisms and the basic substrate material (organic or inorganic). From a medical point of view the challenges are transient materials for which the degradation rate can be "programmed" at the desired rate. Important ICT challenges, in particular for implantable devices, remain for wireless data and energy transfer. In situ solutions are desired, i.e. solutions that are powered by body movement or heat, or obtain energy via biofluids. Ingestible smart pills equipped with sensors for mobile clinical monitoring, smart drug delivery and endoscopy diagnosis are an area of intensive research according to a patent analysis [C10-49] conducted in 2022. Energy supply for implantable and ingestible biomedical electronic devices has also raised substantial attention in materials research [C10-50], similarly targeted for the field of medicine.

10.10.2 Bioreactors for energy production

A bioreactor is a device used in biotechnology to cultivate biological organisms, such as cells, bacteria, yeast, or other microorganisms, under controlled conditions, optimised to support growth, metabolism and production of desired substances. Bioreactors can also be constructed to produce energy through various process, typically in the realm of bioenergy production. A research field in this area addresses Microbial Fuel Cells (MFCs), that utilise microbial metabolism to generate electricity directly from organic matter. In an MFC, microorganisms oxidize organic compounds, releasing electrons that can be harvested as electrical energy. Bioreactors can be designed to house these microbial communities and facilitate the conversion of organic substrates into electricity. Especially micro-sized MFCs open-up various interesting application opportunities [C10-51] for embedded smart machine type communication devices that can be deployed in areas where their recovery for exchange of batteries is impractical or impossible. Think of an inspection robot that is deployed in the sewage system of a city and "lives and works" there without the need for maintenance or other intervention. The energy for robot movement, inspection and communication is supplied by an MFC and the organic material in the sewage.

10.11 Semantic Communications

Shannon in his mathematical theory of communication addressed how one can efficiently transmit bits or symbols from the transmitter to the receiver, which he called the 'technical problem'. He also described the 'semantic problem', which pertains to the question of how precisely the transmitted symbols convey meaning and the 'effectiveness problem' which deals with how effectively the communication achieves conduct in the desired way. However, he explicitly chose to focus only on the technical problem.

With significant advances in machine learning, and significant amount communication and information processing happening between machines, it is time to consider goal-oriented communications. In these emerging applications, massive amounts of data from many devices with strict latency requirements is required and hence it is necessary to be very efficient with bandwidth and energy. Developing a theory of semantic and effectiveness communications can lead to major increases in communication efficiency. Notions of entropy and mutual information need to be expanded to include the semantics. This topic has been researched before, and is still currently on-going research, but its full realization in all layers of the communication system may have deep transformational changes on the operation of any smart infrastructure. The underlying challenge for sharing semantics between different parties, and essential how to efficiently bootstrap a common knowledge base across all parties remains a hard challenge to be solved, and is not being currently properly covered in all its multidisciplinarity, specially integrating application-level aspects as those arising from XR multi-sensory applications.

10.12 Drone Corridors and 3D Radio Connectivity

Uncrewed aerial vehicles (UAVs), commonly known as drones, are poised to drive significant economic growth and societal change. Due to their low cost and high mobility, UAVs are expected to become crucial for goods delivery, surveillance, search and rescue, and monitoring wildfires, crowds, and assets. As urbanization strains ground transportation, electrical vertical take-off and landing vehicles (eVTOLs) could revolutionize urban mobility by serving as air taxis or ambulances, enhancing the speed, safety, and interconnectedness of transportation systems. Autonomous levitating pods, once considered science fiction, are now being tested and could transform commuting, influencing where we live and work [C10-59].

Use case and technology requirements: As UAVs become more widespread, their flights may be restricted to specific aerial corridors defined by air traffic regulation authorities. Within these corridors, UAVs will need to exchange vast amounts of real-time data with the network, necessitating ultra-reliable wireless connectivity [C10-60]. This connectivity must support safe UAV operations through low-latency control and mission-specific data, encouraging legislators to relax regulations on civilian pilotless flights and paving the way for autonomous UAVs and related vertical markets.

Research challenges: Traditional cellular base stations are optimized for 2D ground connectivity, which limits UAV reach to upper antenna sidelobes and causes signal fluctuations with UAV movement. UAVs flying above buildings may also encounter line-of-sight interference from multiple cells, complicating the decoding of critical command and control messages [C10-61]. Achieving 3D connectivity will require re-engineering ground-based deployments. Traditional trial-and-error cell shaping is not scalable, prompting the need for automated optimization techniques using data-driven models [C10-62]. Alternatively, UAV connectivity could be supported by non-terrestrial networks (NTNs), such as Low Earth Orbit (LEO) satellite constellations, with terrestrial network operators leasing spectrum and infrastructure from NTN providers [C10-63].

10.13 High Frequency Integrated Photonics and Communication Systems

Background: The telecommunications sector is shifting its focus towards next-generation networks, namely B5G/6G. A key technological advancement in this area is the transition to the upper millimetre wave (100-300 GHz) and THz (0.1 -10 THz) spectrum. The larger bandwidth and shorter wavelengths offer the benefits of higher data rates and greater miniaturisation. This also allows for denser antenna arrays on integrated platforms, with the potential for integrated directional beamforming.

Integrated Sensing and Communications (ISAC): Concurrently, the high photon energies of THz radiation can enhance the resolution and accuracy of radar and imaging systems, as well as enable new sensing capabilities.

Notably, specific frequencies in the THz band are significantly affected by molecular absorption. While traditional communication systems have avoided these frequencies, a shift in perspective reveals that the inherent characteristics of the THz range make this spectrum ideal for joint communication and sensing (JCAS) systems.

Challenges: Communications in the THz band face some remarkable challenges. The very high attenuation significantly limits the coverage, implying that new techniques must be developed for coverage extension and energy efficiency enhancement. New waveforms with multiple numerologies must be developed to achieve full integration of the sensing and communication subsystems. At the same time, receiver impairments such as phase noise play an important role in THz frequencies and new approaches to minimize their effect must be investigated. At present, there is a scarcity of integrated systems capable of operating in frequencies up to THz for both communication and sensing applications. High frequency components face technological constraints and integration challenges within photonic circuits. These limitations include high power consumption, complex system design, and difficulties in achieving high efficiency and miniaturisation for THz devices. Among these, the most significant is the effectiveness and efficiency of current THz technologies, which impede their practical application and widespread use. Firstly, the demand for ultrafast broadband connectivity and precise environmental monitoring in various applications (such as safety, industrial control, and infrastructure monitoring) cannot be fulfilled with existing solutions. Secondly, the absence of miniaturised, integrated devices that can operate efficiently at such high frequencies restricts the potential for new applications and technologies that could leverage its capabilities. However, by combining advancements in high frequency components with photonic circuit integration, these limitations could be overcome, leading to the creation of compact, accurate, and cost-effective devices for mm and THz communication and sensing. With complementary technologies operating at these high frequencies, high frequency Integrated Photonics holds the promise of revolutionising ultrafast broadband connectivity and real-time environmental monitoring, thereby unlocking new applications and enhancing existing ones.

10.14 Hidden Communication Systems and Weird Networks

In an era where digital surveillance and censorship are increasingly prevalent, the need for secure and covert communication channels has never been more critical. At the same time, the need to detect hidden communication channels and prevent the illegal and malicious activities coordinated and conducted in hidden networks has emerged. Hidden Communication Systems (HCS) and Weird Networks have grown as vital research areas dedicated to addressing these challenges. HCS focus on creating networks that can operate undetected, ensuring user anonymity and data privacy. Weird networks, on the other hand, explore unconventional communication pathways within existing infrastructures, leveraging their inherent properties to enhance security.

Over the years, several methods have been developed to secure and convert communication channels to protect users from surveillance, espionage, and censorship. One of the most widespread solutions for such technology has been The Onion Router network, i.e., Tor network, whose principle was developed by the U.S. Navy in mid-1990s to protect government communications. The initial launch of Tor network happened in September 2002 [C10-64] [C10-65]. Like the Tor-network, the hidden communication solutions and weird networks are implemented currently as overlay networks on top of the legacy Internet protocols, without affecting underlying physical wired Internet or wireless access networks. However, recently the interest in alternative solutions and methods have been also considered, including for example the quantum and nano communications mentioned in previous chapters.

“Weird Networks” are communication networks that operate outside the original specifications of a network. These networks can be formed spontaneously, not being part of the network’s intended design. Weird Networks are use-case agnostic and can be applied in various use-case scenarios and verticals, from internet freedom efforts to secure critical communications.

10.14.1 Research challenges and directions

The research in Weird Networks and HCS focuses especially on: a) Developing models to understand and detect hidden networks, including the creation of domain-specific languages and formal analysis tools to verify the properties and protocols of weird networks, and methods to reveal and counter steganography. b) Resilient Anonymous Communication, including enhancing communication systems to ensure they remain anonymous and resilient against censorship and surveillance, for example, enabling advanced cryptography for data and fuzzing and hiding communication endpoints, as well as enabling protocol steganography or utilizing other types of steganography to hide data. c) Developing adaptable and secure network architectures, which allow dynamic adjustments to network configurations, improving both performance and privacy. The adaptable network architectures can leverage, for example, software-defined networking solutions and the efficient use of AI-based network management and configuration methods to create and detect dynamic and hidden networks.

Future orientations of the research include developing new mathematical guarantees for the privacy and performance of hidden networks, exploring how new communication pathways can emerge within networks to enhance their robustness and security, as well as supporting Internet freedom efforts by developing technologies that enable secure communication in challenging environments.

- Design and Deployment: Strategies for developing robust hidden networks capable of withstanding detection and attacks, including protocol design for data and control plane protocols.
- Anonymity and Privacy: Advanced cryptographic techniques and network obfuscation methods to ensure user anonymity and data privacy, including post-quantum cryptography solutions for Internet and mobile systems.
- Performance Optimization: Balancing the trade-offs between network performance, scalability, and the level of concealment.
- Testing and Validation: Creating realistic testing environments to validate the effectiveness of hidden communication systems.

Some of the other research and development directions considered for enabling HCS and Weird Networks include:

- Internet of Sound: Utilizing sound waves and, for example, background sounds or noise for covert data transmission.
- Light Communications: Employing visible light communication (VLC) and Li-Fi technologies for secure data transmission or utilizing other frequency ranges such as infrared signals outside the regular communication networks.
- Biological Systems: Leveraging biological organisms and processes for data transmission. Some of the techniques and technologies such as utilizing extracellular vehicles and intercell communication have already impact on medical applications and biosensing but may provide a way to create covert communication channel for stealth molecular communications.
- Primary Particle Communications: Investigating the use of fundamental and exotic particles for communications, for example, muons for deep underground or underwater communication and

hiding communications. Quantum communication is usually included in primary or elementary particle communications.

10.14.2 Emerging use cases and application areas

Weird networks have the potential to revolutionize various verticals by enhancing security, resilience, and adaptability. Some potential applications across different sectors can be identified as below:

Vertical sector use cases and application areas for HCS and Weird Networks	
Traffic Sector	**Smart Traffic Management**: The deployment of covert communication channels for traffic sensors and cameras to ensure data integrity and privacy. This can help in real-time traffic monitoring and management, reducing congestion and improving safety.
	Autonomous Vehicles: Secure and resilient communication networks are crucial for autonomous vehicles. Weird networks can provide hidden and reliable communication pathways, ensuring that vehicles can communicate without interference or detection.
Healthcare	**Patient Data Security**: Protecting sensitive patient information is paramount. Weird networks can create secure communication channels for transmitting patient data, ensuring privacy and compliance with regulations.
	Remote Monitoring: Secure networks can support remote patient monitoring systems, allowing healthcare providers to track patient health data in real-time without compromising security.
Finance	**Secure Transactions**: Financial institutions can use weird network and hidden network solutions to secure transactions and protect against cyber-attacks. This includes not only encrypting of data but also creating hidden communication channels for sensitive financial data.
	Fraud Detection: Enhanced network security can help in detecting and preventing fraudulent activities by ensuring that communication channels are secure and monitored.
Manufacturing	**Industrial IoT**: Weird network solutions can support the deployment of secure and resilient communication pathways for industrial IoT devices, ensuring that manufacturing processes are not disrupted by cyber-attacks.
	Supply Chain Security: Secure communication networks can help in tracking and managing supply chains, ensuring that data is protected and tamper-proof.
Public Sector	**Emergency Services**: Weird networks can provide secure and reliable communication channels for emergency services, ensuring that they can operate effectively during crises.
	Government Communications: Protecting sensitive government communications is crucial. Solutions can create hidden and secure communication pathways for government officials.
Utilities	**Smart Grid Security**: Ensuring the security of smart grids is essential for reliable energy distribution. Weird networks can provide secure communication channels for smart grid components.
	Water Management: Secure hidden networks can support the monitoring and management of water resources, ensuring that data is protected, and systems are resilient.

10.14.3 Long-, medium- and short-term actions

We could define also short-, medium-, and long-term actions for the HCS and Weird Networks research topics. Including in the short term, investigation and analysis of existing Weird Network solutions, traffic hiding and obfuscation methods, building real-world experimentation platforms and tools, including simulation/emulation and AI tools, for analysis and sampling the traffic in overlay networks. The analysis work would concentrate on the properties such as usability of different weird network and HCS methods for use cases, analysis of performance and complexity (including the trade-offs for network utilization and quality, security, and sustainability).

The long term and mid-term action also include study, discovery, and realization of weird new network solutions, and focus also on new physical network layers, not only the solutions as overlay of Internet.

Short/medium term:

The short- and mid-term actions and topics may include:

- Detection and testing: Developing methods for detecting and testing overlay weird networks and traffic obfuscation methods, building tools and experimentation platforms.
- Protocol enhancements: Improving existing internet protocols and integrating also new technologies such as LiFi and the Internet of Sound as HCS methods, study 6G enablers, advanced architecture, and protocol solutions in context of Weird Networks and hidden communication channels.
- Steganography and steganalysis: Enhancing and analysis of steganographic techniques in the overlay networks, utilizing AI tools to analyze and enable hidden communication.
- Artificial Intelligence: Use of AI and machine learning for traffic anomaly and hidden communication channel detection, development of adversarial machine learning to detect vulnerabilities of machine learning models, and enhance the resilience and robustness of weird networks, and enable, for example, adaptive security measures.
- Quantum communication and cryptography: Enabling the secure connectivity with quantum key distribution (QKD) and enhancement of existing trial and proof-of-concept infrastructures, study especially on distributed system architecture, and how the QKD can enable the secure and resilient telecommunications, validation of post quantum cryptography (PQC) solutions for Internet and mobile systems, investigating of quantum-based steganographic methods and utilization of quantum computing for analyzing hidden traffic, as well as design and use of hybrid quantum – classical algorithms and computing.

Long term:

The long-term actions and topics may include:

- Biological – Bio-inspired networks: Exploring the bio-inspired data transmission system in communication networks, using biological system as data storage (including DNA), utilizing inter-cellular communication mechanism (e.g., extracellular vesicles) in data transmission and integration with electrical/optical networks and services,
- Quantum Internet and Quantum Network Infrastrucuture: Development and design of large-scale the quantum network infrastructure based on secure quantum key distribution for critical services, study of Quantum Internet able to maintain, store and forward quantum state, solutions for connecting distributed European quantum computer infrastructure, development of quantum memory and router/repeater solutions, further develop the quantum AI and machine learning for HCS and Weird Networks.
- Muon and primary particle communications: Study on utilizing muon-based communication to conceal traffic, enable underwater and underground communication channels, study of utilizing Muongraphy (Muon tomography) for steganography and covert communication, development of muon detector, receiver solutions for communication, study of alternative fundamental particles for communication and network technology purposes.
- Advanced magnetic communication systems: Further research on how magnetic fields, magnetic induction and gravitation fields can be used as communication channel, especially in situation where traditional RF frequencies can not be used, utilizing very low frequencies underwater and underground networks beyond short range magnetic induction.

10.15 Mastering complexity

It is argued in Chapter 6 that there is a vital need for tractable modelling and optimization tools allowing one to handle and master the joint complexity of the many new radio paradigms to be introduced in 6G.

However, more global network architecture questions, like e.g. their hierarchical organization, their level of centralization/decentralization, the level of virtualisation, or more concretely the balance between core and edge computing, also contribute to the complexity of 6G. These questions and their impact on network performance metrics are most often addressed independently using queuing network or resource sharing theory. However, the joint complexity of these architecture questions and of the new radio paradigms alluded to above is not taken into account yet, and should be addressed for 6G to reach the promised performance targets.

This joint analysis can be conducted using either discrete event simulation or by developing unified stochastic geometry tools that incorporate the new 6G RAN components (as described in Chapter 6). These architectural choices, which have been partially introduced in the literature [C10-66] [C10-67] [C10-68], along with the relevant resource sharing paradigms, can potentially be used to master complexity in this context.

11 References

Introduction

[C1-1] NetworldEurope, Strategic Research and Innovation Agenda, Strategic Research and Innovation Agenda 2022 Technical Annex, https://bscw.sns-ju.eu/pub/bscw.cgi/d95665/SRIA%202022% 20Technical%20Annex% 20Published.pdf

System Architecture Considerations

[C2-1] S. Ren et al., "Routing and Addressing with Length Variable IP Address", Proc. of ACM SIGCOMM NEAT workshop, 2019.

[C2-2] R. Li et al., "A Framework for Qualitative Communications using Big Packet Protocol", Proc. of ACM SIGCOMM NEAT workshop, 2019.

[C2-3] 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.

[C2-4] Pouzin Society, "RINA: Building a Better Network", available online: http://pouzinsociety.org/education/highlights, Last accessed May 20, 2020.

[C2-5] D. Trossen et al., "Service-based Routing at the Edge", available online: https://arxiv.org/abs/1907.01293v1. Last accessed May 20, 2020.

[C2-6] SCION Internet Architecture, online: https://www.scion-architecture.net/. Last accessed May 20, 2020.

[C2-7] "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.

[C2-8] "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.

[C2-9] IETF SFC WG. Available Online: https://datatracker.ietf.org/wg/sfc/about/. Last accessed May 20, 2020.

[C2-10] IETF ANIMA WG. Available Online: https://datatracker.ietf.org/wg/anima/about/. Last accessed May 20, 2020.

[C2-11] IETF COIN RG. Available Online: https://irtf.org/coinrg. Last accessed May 20, 2020.

[C2-12] IETF FORCES WG. Available Online: https://datatracker.ietf.org/wg/forces/about/. Last accessed May 20, 2020.

[C2-13] IETF QUIC WG. Online: https://datatracker.ietf.org/wg/quic/about/. Last accessed May 20, 2020.

[C2-14] 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.

[C2-15] Roland Bless, Zoran Despotovic, Artur Hecker, Martina Zitterbart, "KIRA: Distributed Scalable ID-Based Routing With Fast Forwarding", in proc. IFIP NETWORKING 2022, June 13-16 2022, Catania, Italy

[C2-16] J.H. Saltzer, D.P. Reed and D.D. Clark, "End-to-End Arguments in System Design", ACM Transactions in Computer Systems, pp. 277-288, vol. 2, ACM, 1984.

[C2-17] B. Carpenter and B. Liu. 2020. Limited Domains and Internet Protocols. RFC 8799. IETF. http://tools.ietf.org/rfc/rfc8799.txt.

[C2-18] B. Carpenter, J. Crowcroft, D. Trossen, "Limited domains considered useful", ACM Computer Communication Review, pp. 22-28, volume 51, issue 3, ACM, 2021.

[C2-19] 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.

[C2-20] A. S. Asrese, S. J. Eravuchira, V. Bajpai, P. Sarolahti, J. Ott, "Measuring Web Latency and Rendering Performance: Method, Tools \& Longitudinal Dataset", IEEE Transactions on Network and Service Management, vol. 16, num. 2, pp. 535-549, IEEE, 2019.

[C2-21] M. Bloecher, R, Khalili, L. Wang, P. Eugster, "Letting off STEAM: Distributed Runtime Traffic Scheduling for Service Function Chaining", in proc. IEEE INFOCOM 2020.

[C2-22] K. Khandaker, D. Trossen, R. Khalili, Z. Despotovic, A. Hecker, J. Carle, "CArDS: Dealing a New Hand in Reducing Service Request Completion Times", IFIP NETWORKING 2022.

[C2-23] 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.

[C2-24] “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.

[C2-25] M. Caesar et al., “ROFL: routing on flat labels”, ACM SIGCOMM Computer Communication Review, August 2006, https://doi.org/10.1145/1151659.1159955.

[C2-26] G. Xylomenos et al, “A Survey of Information-Centric Networking Research”, IEEE Communications Surveys & Tutorials (Volume: 16, Issue: 2, Second Quarter 2014).

[C2-27] 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.

[C2-28] 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.

[C2-29] Maja Curić, Georg Carle, Zoran Despotovic, Ramin Khalili, and Artur Hecker, “SDN on ACIDs”, in Cloud-Assisted Networking workshop, in proc. of ACM CONEXT 2017, Incheon, South Korea, December 2017.

[C2-30] M. Curic, Z. Despotovic, A. Hecker, G. Carle, “FitSDN: Flexible Integrated Transactional SDN”, IEEE LCN 2019.

[C2-31] 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.

[C2-32] E. Masanet, A. Shehabi, N. Lei, S. Smith, J. Koomey, “Recalibrating global data center energy-use estimates,” Science, vol. 367, no. 6481, pp. 984-986, Feb. 2020.

[C2-33] E. Masanet, J. Koomey, “Does not compute: Avoiding pitfalls assessing the Internet’s energy and carbon impacts,” Joule, vol. 5, pp. 1-4, July 2021; https://doi.org/10.1016/j.joule.2021.05.007.

[C2-34] Advanced Configuration and Power Interface Specification. Available: https://uefi.org/sites/default/files/resources/ACPI_6_2.pdf.

[C2-35] ETSI ES 203 682 V1.1.0, “Environmental Engineering (EE); Green Abstraction Layer (GAL); Power management capabilities of the future energy telecommunication fixed network nodes; Enhanced Interface for power management in Network Function Virtualisation (NFV) environments,” 2019.

[C2-36] ITU-T L.1362, “Interface for power management in network function virtualization environments – Green abstraction layer version 2,” 2019.

[C2-37] R. Zoppoli, M. Sanguineti, G. Gnecco, T. Parisini, Neural Approximations for Optimal Control and Decision, Springer Nature, Cham, Switzerland, 2019.

[C2-38] Martin Rapp, Ramin Khalili, Kilian Pfeiffer, Jörg Henkel, “DISTREAL: Distributed Resource-Aware Learning in Heterogeneous Systems”, 36th AAAI 2022, Feb 2022, Vancouver, Canada.

[C2-39] Jing Tan, Ramin Khalili, Holger Karl, Artur Hecker, “Multi-Agent Distributed Reinforcement Learning for Making Decentralized Offloading Decisions”, in proc. IEEE INFOCOM 2022.

[C2-40] NGMN’s report “Green Future Networks – Network Energy Efficiency“, v1.0, Nov. 2021, available online https://www.ngmn.org/wp-content/uploads/211009-GFN-Network-Energy-Efficiency-1.0.pdf.

[C2-41] E. Hossein and F. Fredj, “Energy Efficiency of Machine-Learning-Based Designs for Future Wireless Systems and Networks”, Editorial, *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1005-1010, Sept. 2021.

[C2-42] 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.

[C2-43] K Yeow, A Gani, RW Ahmad, J.J.P.C. Rodrigues, Kwangman Ko, "Decentralized consensus for edge-centric internet of things: A review, taxonomy, and research issues", Vol 6, pp. 1513 - 1524, IEEE Access, 2017.

[C2-44] Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper

[C2-45] 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.

[C2-46] H. Kim, J. Park, M. Bennis, and S.-L. Kim, “On-device federated learning via blockchain and its latency analysis,” arXiv:1808.03949, 2018.

[C2-47] 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.

[C2-48] 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.

[C2-49] 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.

[C2-50] R. Zoppoli, M. Sanguineti, G. Gnecco and T. Parisini, “Neural Approximations for Optimal Control and Decision”, Springer, 2019.

[C2-51] A. Lacoste, A. Luccioni, V. Schmidt, Victor and T. Dandres, "Quantifying the Carbon Emissions of Machine Learning", arXiv preprint arXiv:1910.09700, 2019

[C2-52] S&P Globl Market Intelligence, “2024 Trends in Datacenter Services & Infrastructure”, 451 Research 2024 Preview, 2023.

[C2-53] Alex de Vries, “The growing energy footprint of artificial intelligence”, Joule, 7(10):2191–2194, 2023

[C2-54] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. Aga, J. Huang, C. Bai, et al, “Sustainable ai: Environmental implications, challenges and opportunities”, Proceedings of Machine Learning and Systems, 4:795–813, 2022.

[C2-55] AI-RAN Goes Live and Unlocks a New AI Opportunity for Telcos, https://developer.nvidia.com/blog/ai-ran-goes-live-and-unlocks-a-new-ai-opportunity-for-telcos/?ncid=so-link-911123-vt26&linkId=100000308020822

[C2-56] M. Xu, W. Yin, D. Cai, R. Yi, D. Xu, Q. Wang, B. Wu, Y. Zhao, C. Yang, S. Wang, Q. Zhang, Z. Lu, L. Zhang, S. Wang, Y. Li, Y. Liu, X. Jin, X. Liu, “A Survey of Resource-efficient LLM and Multimodal Foundation Models”, arXiv preprint arXiv:2401.08092, 2024

[C2-57] LLM Inference Performance Engineering: Best Practices https://www.databricks.com/blog/llm-inference-performance-engineering-best-practices

Fundamental Enablers for Future 6G Systems

[C3-1] NetworldEurope Strategic Research and Innovation Agenda (SRIA 2022), available at https://www.networldeurope.eu/sria-and-whitepapers/

[C3-2] B. Carpenter and B. Liu. 2020. Limited Domains and Internet Protocols. RFC 8799.

[C3-3] Dirk Trossen B. Carpenter, J. Crowcroft. 2021. *Limited Domains Considered Useful*. ACM Computer Comm. Review, Vol. 51. Issue 3.

[C3-4] L. Iannone (ed.), *IP Addressing Considerations*, available at https://datatracker.ietf.org/doc/draft-iannone-ip-addressing-considerations/, 2023

[C3-5] M. A. Ferreira J. L. Sobrinho. 2020. Routing on Multi Optimality Criteria. In ACM SIGCOMM.

[C3-6] R. Khalili Z. Despotovic A. Hecker J. Carle K. Khandaker, D. Trossen.2022. CArDS: Dealing a New Hand in Reducing Service Request Completion Times. In IFIP Networking.

[C3-7] M. Beck, T. Moore, “How We ruined the Internet”, submitted to Communications of the ACM, 2023, available at http://arxiv.org/abs/2306.01101

[C3-8] Magoula, L., Koursioumpas, N., Thanopoulos, A.-I., Panagea, T., Petropouleas, N., Gutierrez-Estevez, M. A., & Khalili, R. (2023). A Safe Genetic Algorithm Approach for Energy Efficient Federated Learning in Wireless Communication Networks.

[C3-9] Bless, R., Zitterbart, M., Despotovic, Z., and A. Hecker, "KIRA: Distributed Scalable ID-based Routing with Fast Forwarding", June 2022, available at https://doi.org/10.23919/IFIPNetworking55013.2022.9829816

[C3-10] V. Olteanu, H. Eran, D. Dumitrescu, A. Popa, C. Baciu, M. Silberstein, G. Nikolaidis, M. Handley, C. Raiciu, An edge-queued datagram service for all datacenter traffic, NSDI 22

[C3-11] S. McQuistin, M. Karan, P. Khare, C. Perkins, G. Tyson, M. Purver, P. Healey, W. Iqbal, J. Qadir, and I. Castro. 2021. Characterising the IETF through the lens of RFC deployment. In Proceedings of the 21st ACM Internet Measurement Conference (IMC '21). Association for Computing Machinery, New York, NY, USA, 137–149.

[C3-12] Paasch, Christoph; Detal, Gregory; Duchene, Fabien; Raiciu, Costin; Bonaventure, Olivier (2012). "Exploring mobile/WiFi handover with multipath TCP". Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: Operations, challenges, and future design - Cell Net '12. ACM SIGCOMM workshop on Cellular Networks (Cellnet'12)

[C3-13] Raiciu; Barre; Pluntke; Greenhalgh; Wischik; Handley (2011). "Improving datacenter performance and robustness with multipath TCP". ACM SIGCOMM Computer Communication Review. 41 (4): 266

[C3-14] Hesmans, Benjamin; Bonaventure, Olivier (2016). "An enhanced socket API for Multipath TCP". Proceedings of the 2016 workshop on Applied Networking Research Workshop - ANRW 16. pp. 1–6.

[C3-15] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg, “SIP: Session Initiation Protocol”, RFC 2543, Internet Society, 1999

[C3-16] B. Briscoe, K. De Schepper, M. Bagnulo, G. White, “Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture”, RFC9330, Internet Society, 2023

[C3-17] M. Handley, C. Raiciu, A. Agache, A. Voinescu, A. W. Moore, G. Antichi, M. Wójcik, “Re-architecting datacenter networks and stacks for low latency and high performance”, ACM SIGCOMM, 2017

[C3-18] M. McBride, D. Trossen, D. Guzman, T. Martin, “BGP Blockchain”, IETF draft, 2024, available at https://datatracker.ietf.org/doc/draft-mcbride-rtgwg-bgp-blockchain/

[C3-19] T. Eckert (Ed.) and M. Behringer, “Using an Autonomic Control Plane for Stable Connectivity of Network Operations, Administration, and Maintenance (OAM),” RFC 8368 (Informational), RFC Editor, Fremont, CA, USA, May 2018. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8368.txt

[C3-20] E. Coronado, R. Behravesh, T. Subramanya, A. Fernàndez-Fernàndez, M. S. Siddiqui, X. Costa-Pérez, and R. Riggio, “Zero Touch Management: A Survey of Network Automation Solutions for 5G and 6G Networks,” IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2535–2578, 2022.

[C3-21] D. Trossen, J. Crowcroft, “If Multicast is the Answer -- What was the Question?”, arxiv, 2022, available at https://arxiv.org/abs/2211.09029

[C3-22] P. Mendes, J. Finkhaeuser, L. Contreras, D. Trossen, “Use Cases and Problem Statement for Routing on Service Addresses”, IETF draft, 2023, available at https://www.ietf.org/archive/id/draft-mendes-rtgwg-rosa-use-cases-01.html

[C3-23] M. J. Reed; J. Riihijärvi; M. Georgiades; N. Fotiou; G. Xylomenos, D. Trossen, “IP over ICN - The better IP?”, In European Conference on Networks and Communications (EuCNC), 2015

[C3-24] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher and B. Ohlman, "A survey of information-centric networking," in IEEE Communications Magazine, vol. 50, no. 7, pp. 26-36, July 2012

[C3-25] M. Blöcher, R. Khalili, L. Wang, and P. Eugster, “Letting off STEAM: Distributed Runtime Traffic Scheduling for Service Function Chaining”, In IEEE INFOCOM, 2020

[C3-26] 3GPP, TS23.501, “System architecture for the 5G System (5GS)”, 2024, available at https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[C3-27] Fiandrino, C., Zhang, C., Patras, P., Banchs, A., & Widmer, J. (2020). A machine-learning-based framework for optimizing the operation of future networks. IEEE Communications Magazine, 58(6), 20-25.

[C3-28] Arya, G., Bagwari, A., & Chauhan, D. S. (2022). Performance analysis of deep learning-based routing protocol for an efficient data transmission in 5G WSN communication. IEEE Access, 10, 9340-9356.

[C3-29] Langpoklakpam, B., & Murry, L. K. (2023). Review on Machine Learning for Intelligent Routing, Key Requirement and Challenges Towards 6G. Computer Networks and Communications, 214-230. Trends in Electronics and Informatics (ICOEI) (pp. 1094-1101). IEEE.

[C3-30] Paul, R., Cohen, K., & Kedar, G. (2023). Multi-Flow Transmission in Wireless Interference Networks: A Convergent Graph Learning Approach. arXiv preprint arXiv:2303.15544.

[C3-31] Gahtan, B., Cohen, R., Bronstein, A. M., & Kedar, G. (2023, October). Using Deep Reinforcement Learning for mmWave Real-Time Scheduling. In 2023 14th International Conference on Network of the Future (NoF) (pp. 71-79). IEEE.

[C3-32] Danilchenko, K., Kedar, G., & Segal, M. (2023, June). Reinforcement Learning Based Routing For Deadline-Driven Wireless Communication. In 2023 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 30-35). IEEE.

[C3-33] McKinsey, “Technology and telecommunications B2B customer buying trends: Bright horizons with some warning signs”, Pulse Survey, Feb 2024, available at https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/technology-and-telecommunications-b2b-customer-buying-trends-bright-horizons-with-some-warning-signs?stcr=BE8CFD05D3B64ECD81A641D224C966FF&cid=other-eml-dre-mip-mck&hlkid=3f3d24e4cead473bbee513ad8a5a403f&hctky=12197129&hdpid=4daeff29-9fd1-498e-9907-0a69e90936bf

[C3-34] G. Xylomenos et al., "A Survey of Information-Centric Networking Research," in IEEE Communications Surveys & Tutorials, vol. 16, no. 2, pp. 1024-1049, Second Quarter 2014

[C3-35] M. Denis et al., “EBB: Reliable and Evolvable Express Backbone Network in Meta,” in Proceedings of the ACM SIGCOMM 2023 Conference, ser. ACM SIGCOMM ’23. New York, NY, USA: Association for Computing Machinery, 2023, p. 346–359. [Online]. Available: https://doi.org/10.1145/3603269.3604860

[C3-36] A. D. Ferguson et al., “Orion: Google’s Software-Defined Networking Control Plane,” in 18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21). USENIX Association, Apr. 2021, pp. 83–98. [Online]. Available: https://www.usenix.org/conference/nsdi21/presentation/ferguson

[C3-37] S. Deering, R. Hinden, “Internet Protocol, Version 6 (IPv6) Specification”, RFC2460

[C3-38] M. Kühlewind, B. Trammell, C. Perkins. 2017. Post sockets: Towards an evolvable network transport interface. In Proceedings of 2017 IFIP Networking Conference (IFIP Networking) and Workshops.

[C3-39] M. Alasmar, G. Parisis, and J. Crowcroft. 2021. SCDP: Systematic Rateless Coding for Efficient Data Transport in Data Centers. IEEE/ACM Trans. Netw. 29, 6 (Dec. 2021), 2723–2736.

[C3-40] M.-J. Montpetit, C. Westphal, and D. Trossen. 2012. Network coding meets information-centric networking: an architectural case for information dispersion through native network coding. In Proceedings of the 1st ACM workshop on Emerging Name-Oriented Mobile Networking Design - Architecture, Algorithms, and Applications (NoM '12). Association for Computing Machinery, New York, NY, USA, 31–36.

[C3-41] L. Fan, Z. Han, "Hybrid Quantum-Classical Computing for Future Network Optimization," in IEEE Network, vol. 36, no. 5, pp. 72-76, September/October 2022

[C3-42] Y. Li, Z. Han, S. Gu, G. Zhuang and F. Li, "Dyncast: Use Dynamic Anycast to Facilitate Service Semantics Embedded in IP address," *2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR)*, Paris, France, 2021, pp. 1-8

[C3-43] L. Wang, V. Lehman, A. K. M. Mahmudul Hoque, B. Zhang, Y. Yu and L. Zhang, "A Secure Link State Routing Protocol for NDN," in *IEEE Access*, vol. 6, pp. 10470-10482, 2018

[C3-44] Zittrain, Jonathan. The Future of the Internet – And How to Stop It. ISBN 978-0300151244

[C3-45] Tarnoff, Ben. Internet for the People: The Fight for Our Digital Future. ISBN 978-1804290392

[C3-46] European Commission report: Recommendations and roadmap for European sovereignty on open source hardware, software and RISC-V Technologies, https://digital-strategy.ec.europa.eu/en/library/recommendations-and-roadmap-european-sovereignty-open-source-hardware-software-and-risc-v

[C3-47] Trossen, Dirk, et al. "IP over ICN-the better IP?." 2015 European conference on networks and communications (EuCNC). IEEE, 2015.

[C3-48] Privat, Gilles, and Alexey Medvedev. "Guidelines for Modelling with NGSI-LD." ETSI White Paper 42 (2021).

[C3-49] Tanaka, Daichi, and Masatoshi Kawarasaki. "Congestion control in named data networking." Proceedings of the IEEE international symposium on local and metropolitan area networks (LANMAN). 2016.

[C3-50] Zhang, Meng, Hongbin Luo, and Hongke Zhang. "A survey of caching mechanisms in information-centric networking." IEEE Communications Surveys & Tutorials 17.3 (2015): 1473-1499.

[C3-51] M. J. Reed, M. Al-Naday, N. Thomos, D. Trossen, G. Petropoulos, S. Spirou, Stateless multicast switching in software defined networks, ICC 2016, Kuala Lumpur, Maylaysia, 2016

[C3-52] M. Nottingham. 2020. The Internet is for End Users. RFC8890

[C3-53] J. Border, M. Kojo, J. Griner, G. Montenegro, Z. Shelby, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations", RFC 3135 (Informational), June 2001.

[C3-54] M. Polese, M. Mezzavilla, M. Zhang, J. Zhu, S. Rangan, S. Panwar, and M. Zorzi. 2017. milliProxy: A TCP proxy architecture for 5G mmWave cellular systems. In Asilomar Conference on Signals, Systems, and Computers, IEEE, pp. 951–957.

[C3-55] D. Hayes, D. Ros, and Ö. Alay. 2019. On the Importance of TCP Splitting Proxies for Future 5G mmWave Communications. In Proceedings of IEEE LCN.

[C3-56] G. Papastergiou, G. Fairhurst, D. Ros, A. Brunstrom, K.-J. Grinnemo, P. Hurtig, N. Khademi, M. Tuexen, M. Welzl, D. Damjanovic and S. Mangiante: "De-ossifying the Internet Transport Layer: A Survey and Future Perspectives", IEEE Communications Surveys and Tutorials 19(1), Firstquarter 2017. DOI 10.1109/COMST.2016.2626780

[C3-57] J. Border. 2020. Google QUIC over Satellite Links. Presentation, IETF PANRG interim, June 3, 2020. https://datatracker.ietf.org/meeting/interim-2020-panrg-01/materials/slides-interim-2020-panrg-01-sessa-google-quic-over-satellite-testing-update

[C3-58] L. Thomas, E. Dubois, N. Kuhn, E. Lochin. Google QUIC performance over a public SATCOM access. International Journal of Satellite Communications and Networking, 2019, 37 (6), pp.601-611.

[C3-59] Kuhn N, Michel F, Thomas L, et al. QUIC: Opportunities and threats in SATCOM. Int J Satell Commun Network. 2022; 40(6): 379-391. doi:10.1002/sat.1432

[C3-60] G. Yuan, D. K. Zhang, M. Sotoudeh, M. Welzl, K. Winstein: "Sidecar: In-Network Performance Enhancements in the Age of Paranoid Transport Protocols", ACM HotNets, November 14-15, 2022. DOI https://doi.org/10.1145/3563766.3564113

[C3-61] G. Yuan, M. Sotoudeh, D. K. Zhang, M. Welzl, D. Mazières, K. Winstein: "Sidekick: In-Network Assistance for Secure End-to-End Transport Protocols", Usenix NSDI '24, Santa Clara, CA, USA, 16-18 April 2024.

[C3-62] M. Welzl, S. Islam, M. Gundersen, A. Fischer: "Transport Services: A Modern API for an Adaptive Internet Transport Layer", IEEE Communications Magazine 59(4), April 2021. DOI 10.1109/MCOM.001.2000870

[C3-63] B. Trammell (Ed.), M. Welzl (Ed.), T. Enghardt, G. Fairhurst, M. Kuehlewind, C. Perkins, P. Tiesel, C. Wood: "An Abstract Application Layer Interface to Transport Services", Internet-draft draft-ietf-taps-interface-26, March 2024. In RFC Editor Queue.

[C3-64] K. Ciko, M. Welzl, M. Marek: "TAPS and RINA: Do they fit together?", 7th International Workshop on the Recursive InterNetwork Architecture (RINA 2020), co-located with IEEE ICIN 2020, Paris, France, 24 February 2020. DOI 10.1109/ICIN48450.2020.9059406

[C3-65] A. Rabitsch, G. Xilouris, T. Anagnostopoulos, K.-J. Grinnemo, T. Sarlas, A. Brunstrom, Ö. Alay, and G. Caso. 2021. Extending network slice management to the end-host. In Proceedings of the 1st Workshop on 5G Measurements, Modeling, and Use Cases (5G-MeMU '21). Association for Computing Machinery, New York, NY, USA, 20–26. https://doi.org/10.1145/3472771.3472775

[C3-66] T. Krüger and D. Hausheer. 2021. Towards an API for the Path-Aware Internet. In Proceedings of the ACM SIGCOMM 2021 Workshop on Network-Application Integration (NAI'21). Association for Computing Machinery, New York, NY, USA, 68–72. https://doi.org/10.1145/3472727.3472808

[C3-67] Y. Li, R. Miao, H. Harry Liu, Y. Zhuang, F. Feng, L. Tang, Z. Cao, M. Zhang, F. Kelly, M. Alizadeh, and M. Yu. 2019. HPCC: high precision congestion control. In Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM '19). Association for Computing Machinery, New York, NY, USA, 44–58. https://doi.org/10.1145/3341302.3342085

[C3-68] J. Langlet, R. B. Basat, G. Oliaro, M. Mitzenmacher, M. Yu, and G. Antichi. 2023. Direct Telemetry Access. In Proceedings of the ACM SIGCOMM 2023 Conference (ACM SIGCOMM '23). Association for Computing Machinery, New York, NY, USA, 832–849. https://doi.org/10.1145/3603269.3604827

[C3-69] W. Feng, J. Gao, X. Chen, G. Antichi, R. B. Basat, M. Mingchao Shao, Y. Zhang, M. Yu. 2024. F3: Fast and Flexible Network Telemetry with an FPGA coprocessor, to appear in ACM CoNext 2024

[C3-70] GSMA Official Document IR.34. “Guidelines for IPX Provider networks (Previously InterService Provider IP Backbone Guidelines)”, Version 17.0, 18 May 2021.

[C3-71] IETF RFC4364. “BGP/MPLS IP Virtual Private Networks (VPNs)”, February 2006.

[C3-72] H. Lønsethagen *et al*., "Toward Smart Public Interconnected Networks and Services - Approaching the Stumbling Blocks," in *IEEE Communications Magazine*, vol. 61, no. 11, pp. 184-190, November 2023, doi: 10.1109/MCOM.006.00210.

Security

[C4-1] https://undocs.org/en/A/RES/217(III)

[C4-2] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[C4-3] https://www.enisa.europa.eu/publications/telecom-security-incidents-2021

[C4-4] https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks

[C4-5] https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive

[C4-6] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0477R(02)

[C4-7] https://ec.europa.eu/digital-single-market/en/news/cybersecurity-5g-networks-eu-toolbox-risk-mitigating-measures

[C4-8] https://www.nextgalliance.org/white_papers/trust-security-and-resilience-for-6g-systems/

[C4-9] https://www.gsma.com/security/resources/gsma-security-landscape-report-2022/6g-ia

[C4-10] https://bscw.5g-ppp.eu/pub/bscw.cgi/d381923/INSPIRE-5Gplus_White_Paper_HLA_FV_GA_Abbrev.pdf

[C4-11] Scott-Hayward, S., Natarajan, S., & Sezer, S.: (2016). A Survey of Security in Software Defined Networks, IEEE Communications Surveys and Tutorials.

[C4-12] Olli Mämmelä, Jouni Hiltunen, Jani Suomalainen, Kimmo Ahola, Petteri Mannersalo, Janne Vehkaperä: ”Towards Micro-Segmentation in 5G Network Security”. EuCNC 2016.

[C4-13] P.Porambage and al. The roadmap to 6G security and Privacy, IEEEE Open Journal of the Communication Society, May 2021

Software and AI technologies for telecommunications

[C5-1] Open Source Software for RAN: https://www.o-ran.org/software

[C5-2] TM Forum, IG1230 Autonomous networks technical architecture, Dec 2022

[C5-3] Capgemini Research, Networks with intelligence, 2024

[C5-4] 2021 Accelerate State of DevOps Report, available at: https://cloud.google.com/devops/state-of-devops?utm_source=google&utm_medium=blog&utm_campaign=FY19-Q3-global-demandgen-website-wd-gcp_gtm_stateofdevops

[C5-5] Proposal for a directive of the European Parliament and of the Council on energy efficiency, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0558

[C5-6] CIS benchmarks, available at: https://www.cisecurity.org/cis-benchmarks/

[C5-7] https://meta-os.eu/

[C5-8] https://aeros-project.eu/

[C5-9] https://he-codeco.eu

[C5-10] https://cognit.sovereignedge.eu

[C5-11] https://cognifog.eu

[C5-12] https://www.decice.eu
[C5-13] https://www.1-swarm.eu/
[C5-14] https://zero-swarm.eu/
[C5-15] Vepakomma, P., Swedish, T., Raskar, R., Gupta, O., & Dubey, A. (2018). No peek: A survey of private distributed deep learning. arXiv preprint arXiv:1812.03288.
[C5-16] Huang, Anbu, et al. "StarFL: Hybrid Federated Learning Architecture for Smart Urban Computing." ACM Transactions on Intelligent Systems and Technology (TIST) 12.4 (2021): 1-23
[C5-17] Kairouz, P., et al. "Advances and open problems in federated learning." Foundations and Trends® in Machine Learning 14.1–2 (2021): 1-210.
[C5-18] Thapa, C. et al. "Advancements of Federated Learning Towards Privacy Preservation: From Federated Learning to Split Learning". Federated Learning Systems. Studies in Computational Intelligence, vol 965. Springer, 2021
[C5-19] Singh, A., et al. "Detailed comparison of communication efficiency of split learning and federated learning." arXiv preprint arXiv:1909.09145 (2019)
[C5-20] Paolo Bellavista, Luca Foschini, and Alessio Mora. 2021. Decentralised Learning in Federated Deployment Environments: A System-Level Survey. ACM Comput. Surv. 54, 1, Article 15 (January 2022), https://doi.org/10.1145/3429252
[C5-21] A. T. Hashmi, S. D. Blostein, and L. Hanzo, "Emerging Terahertz Wireless Communications: Applications, Challenges, and Opportunities," IEEE Transactions on Terahertz Science and Technology, vol. 11, no. 2, pp. 103-133, Mar. 2021
[C5-22] M. Di Renzo, A. Alkhateeb, D. Malliavin, M. Debbah, and A. Zappone, "Smart Probabilistic Combinatorial Optimization for the Deployment of Intelligent Reflecting Surfaces (IRSs)," IEEE Transactions on Communications, vol. 70, no. 4, pp. 1678-1692, Apr. 2022
[C5-23] X. Liu, H. Yu, and V. C. M. Leung, "A Survey on Multi-Operator Terahertz Communications: Technologies, Challenges, and Future Directions," IEEE Wireless Communications, vol. 29, no. 1, pp. 52-58, Feb. 2023
[C5-24] Y. Liu, H. C. Yang, C. Wang, J. G. Andrews, and Z. Ding, "AI-Enabled Resource Allocation in Terahertz Communication Networks: Opportunities and Challenges," IEEE Journal on Selected Areas in Communications, vol. 41, no. 4, pp. 891-905, Apr. 2023
[C5-25] H. C. Huang, H. G. Ryu, and M. L. Sichitiu, "Application-Centric Networking in Terahertz Communication Systems: State-of-the-Art and Future Directions," IEEE Communications Surveys & Tutorials, vol. 26, no. 4, pp. 2881-2906, Fourth quarter 2024
[C5-26] Zhang, Y. Wang, X. Peng, and H. Li, "Smart Agent-Based Multi-Modal Human-Computer Interaction System Using Deep Learning," IEEE Access, vol. 9, pp. 22238-22252, Jan. 2021
[C5-27] S. Lee, H. Kim, S. Lee, and S. J. Lee, "Multi-Modal Human-Robot Interaction System for Autonomous Vehicles Using Deep Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 1123-1133, Mar. 2022
[C5-28] Y. Liu, C. Wang, and Z. Zhang, "AI-Driven Network Functions: Challenges and Opportunities," IEEE Network, vol. 35, no. 5, pp. 262-268, Sep./Oct. 2021
[C5-29] W. Chen, H. Zhang, and S. Li, "Digital Twins for Intelligent Networking: State of the Art and Future Directions," IEEE Communications Magazine, vol. 59, no. 6, pp. 156-163, Jun. 2022
[C5-30] L. Zhang, Y. Xu, and X. Wang, "Privacy-Preserving AI Computing in Edge Networks: Challenges and Solutions," IEEE Transactions on Network Science and Engineering, vol. 9, no. 4, pp. 2079-2090, Oct. 2023
[C5-31] H. Wang, Q. Liu, and Z. Wu, "Smart and Safe: Privacy-Preserving AI for Smart Living," IEEE Internet of Things Journal, vol. 10, no. 1, pp. 433-445, Jan. 2024
[C5-32] J. Li, S. Wang, and X. Yang, "Towards Autonomous Privacy Protection in AI-Driven Networks," IEEE Transactions on Information Forensics and Security, vol. 17, no. 6, pp. 1221-1234, Jun. 2022
[C5-33] A. Zhang, Y. Chen, and H. Liu, "Privacy-Preserving AI Models for User-Centric Networks," IEEE Transactions on Mobile Computing, vol. 21, no. 3, pp. 1091-1104, Mar. 2023
[C5-34] C. Wang, L. Li, and Y. Zheng, "Balancing AI Capabilities and Privacy Protection in Network Operations," IEEE Transactions on Cognitive Communications and Networking, vol. 10, no. 1, pp. 49-62, Mar. 2024
[C5-35] CNCF Cloud Native Definition v1.0, available at: https://github.com/cncf/toc/blob/main/DEFINITION.md
[C5-36] DS. Dustdar, V. Casamajor Pujol, P. Kumar Donta, "On distributed computing continuum systems", IEEE Transactions on Knowledge and Data Engineering, IEEE, pages: 14, 13 January 2022
[C5-37] Coutinho, R.W.L., Boukerche,A., 2022. Design of Edge Computing for 5G-Enabled Tactile Internet-Based Industrial Applications. IEEE Communications Magazine.
[C5-38] Shih-Chun Lin, Kwang-Cheng Chen, Ali Karimoddini, 2021. SDVEC: Software-Defined Vehicular Edge Computing with Ultra-Low Latency. IEEE Communications Magazine.
[C5-39] Shiwei Lai, Rui Zhao, Shunpu Tang, Junjuan Xia, Fasheng Zhou, Liseng Fan, 2021. Intelligent secure mobile edge computing for beyond 5G wireless networks. Elsevier, physical communication, vol. 45.

[C5-40] Bass, L., Weber, I. and Zhu, L., 2015. DevOps: A software architect's perspective. Addison-Wesley Professional.

[C5-41] Gartner: "Future of IT Infrastructure Is Always On, Always Available, Everywhere" https://www.gartner.com/en/newsroom/press-releases/2018-12-03-gartner-says-the-future-of-it-infrastructure-is-always-on-always-available-everywhere

[C5-42] 5G-ACIA White Paper, 2021. Integration of 5G with time-sensitive networking for industrial communications. 5G Alliance for Connected Industries and Automation.

[C5-43] Y. Zhang, C. Wang, and X. Liu, "Distributed Computing-Assisted Communication: Enabling Ubiquitous Computing for High-End Applications," IEEE Transactions on Mobile Computing, vol. 20, no. 8, pp. 2806-2819, Aug. 2021

[C5-44] M. A. K. Ahmed, M. S. Hossain, and A. Ghoneim, "Converged Computing and Communication for Harmonized Communication and Sensing: A Comprehensive Survey," IEEE Access, vol. 9, pp. 54303-54320, Apr. 2021

[C5-45] A. Kumar, S. K. Das, and J. Fang, "Optimizing High-End Experience Applications in Distributed Native Computing Environments," IEEE Transactions on Services Computing, vol. 15, no. 2, pp. 907-919, Mar-Apr. 2022

[C5-46] R. Li, Y. Xiao, and Y. Gao, "Toward Seamless Integration of High-End Applications in Distributed Native Computing: Challenges and Opportunities," IEEE Transactions on Parallel and Distributed Systems, vol. 33, no. 3, pp. 718-731, Mar. 2022

[C5-47] Z. Wang, X. Chen, and W. Zhang, "Dynamic Resource Management for High-End Application Services in Mobile Edge Computing," IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 1755-1768, Jul. 2023

[C5-48] Y. Wu, C. Liu, and L. Feng, "Harmonized Communication and Sensing: Challenges and Opportunities in Converged Computing Environments," IEEE Transactions on Network Science and Engineering, vol. 9, no. 1, pp. 1-14, Jan. 2024

[C5-49] S. Wang, H. Yang, and X. Luo, "A Comprehensive Survey on Edge Computing: Principles, Technologies, and Challenges," IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2348-2398, Fourth quarter 2021

[C5-50] Y. Sun, C. Yao, and J. Zhang, "Edge Computing and Networking: Challenges and Opportunities," IEEE Internet of Things Journal, vol. 9, no. 1, pp. 302-316, Jan. 2022

[C5-51] X. Chen, Y. Wang, and H. Li, "Service-Program Oriented Communication for Edge Computing," IEEE Transactions on Industrial Informatics, vol. 18, no. 2, pp. 903-913, Feb. 2023

[C5-52] G. Xu, W. Zhang, and C. Wang, "Integration-Free Edge Applications: Challenges and Opportunities," IEEE Transactions on Services Computing, vol. 16, no. 4, pp. 1061-1075, Jul-Aug. 2023

[C5-53] L. Liu, Y. Zhang, and X. Liu, "Towards a Protocol for Application-Associated Data Path Orchestration in Edge Computing," IEEE Transactions on Cloud Computing, vol. 12, no. 6, pp. 2276-2289, Jun. 2024

[C5-54] A. Li, X. Chen, and W. Zhang, "Towards Intelligent Orchestration for Agile Business Network Infrastructure," IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 801-814, Jun. 2021

[C5-55] B. Wang, Y. Wang, and H. Liu, "Self-Configuration Mechanisms for Scalable Business Network Infrastructure," IEEE Transactions on Network Science and Engineering, vol. 8, no. 4, pp. 2357-2369, Oct. 2022

[C5-56] C. Zhang, J. Li, and Y. Liu, "Integrating Full Network Functions for Customizable Business Network Infrastructures," IEEE Journal on Selected Areas in Communications, vol. 41, no. 9, pp. 2080-2092, Sep. 2023

[C5-57] Y. Zhao, S. Wu, and L. Wang, "Intelligent Orchestration and Self-Configuration for Agile Business Network Infrastructure: Challenges and Opportunities," IEEE Communications Magazine, vol. 59, no. 3, pp. 98-104, Mar. 2021

[C5-58] W. Zhou, Z. Zhang, and Q. Wu, "Machine Learning Approaches for Self-Maintenance in Business Network Infrastructures," IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 6511-6522, Sep. 2021

[C5-59] H. Wang, L. Li, and X. Liu, "Towards Continuous Self-Maintenance in Business Network Infrastructures: A Framework and Challenges," IEEE Transactions on Network and Service Management, vol. 19, no. 3, pp. 1243-1256, Sep. 2022

[C5-60] AIOTI, "Edge IoT Industrial Immersive Technologies and Spatial Computing Continuum, Release 1", available at https://aioti.eu/wp-content/uploads/AIOTI-Paper-Edge-AI-IoT-Immersive-Technologies-Published.pdf

[C5-61] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL. Twinning the green and digital transitions in the new geopolitical context, available at: com_2022_289_1_en.pdf (europa.eu)

[C5-62] NESSI, "Software and Smart Networks and Services," 2020, available at: https://nessi.eu/wp-content/uploads/2020/09/NESSI-Software-and-SNS-issue1.pdf

[C5-63] S. Schwarzman et al., "Native Support of AI Applications in 6G Mobile Networks via an Intelligent User Plane". IEEE WCNC 2024. Accepted

[C5-64] Orgalim. Environment: Orgalim position on the Sustainable Products Initiative. https://orgalim.eu/position-papers/environment-orgalim-position-sustainable-products-initiative-0

[C5-65] NESSI, "Software and Human Centricity", 2022, available at: https://nessi.eu/wp-content/uploads/2022/04/NESSI-Human-Centricity-issue-1.pdf

[C5-66] Diego Gabriel Soares Pivoto et al., (2023, July). A Detailed Relevance Analysis of Enabling Technologies for 6G Architectures. IEEE Access.

[C5-67] NESSI, “Software and Quantum Computing”, 2022, available at: https://nessi.eu/wp-content/uploads/2022/05/NESSI-Quantum-Computing-issue-1.pdf

[C5-68] Swaraj Shekhar Nande, Osel Lhamo, Marius Paul, Riccardo Bassoli, Frank H P Fitzek. Quantum Machine Learning for Controller Placement in Software Defined Networks. Quantum Machine Learning for Controller Placement in Software Defined Networks (techrxiv.org)

[C5-69] M Ashwin, Abdulrahman Saad Alqahtani, Azath Mubarakali, B Sivakumar, (2023, March). Efficient resource management in 6G communication networks using hybrid quantum deep learning model. Efficient resource management in 6G communication networks using hybrid quantum deep learning model - ScienceDirect

[C5-70] Calvert, Kenneth L., Samrat Bhattacharjee, Ellen Zegura, and James Sterbenz. "Directions in active networks." IEEE Communications Magazine 36, no. 10 (1998): 72-78

[C5-71] Schwartz, B., Jackson, A. W., Strayer, W. T., Zhou, W., Rockwell, R. D., & Partridge, C. (1999, March). Smart packets for active networks. In 1999 IEEE Second Conference on Open Architectures and Network [3] Programming. Proceedings. OPENARCH'99 (Cat. No. 99EX252) (pp. 90-97). IEEE

[C5-72] https://www.irtf.org/coinrg.html

[C5-73] Ahmad, Ijaz, et al. "Sustainability in 6G Networks: Vision and Directions." 2023 IEEE Conference on Standards for Communications and Networking (CSCN). IEEE, 2023

[C5-74] Zhang, S., Lim, W. Y. B., Ng, W. C., Xiong, Z., Niyato, D., Shen, X. S., & Miao, C. (2023). Towards green metaverse networking: Technologies, advancements and future directions. IEEE Network

[C5-75] Galis, A. (2024). Future Sustainable Internet Energy-Defined Networking. Future Internet, 16(1), 23

Radio Technology and Signal Processing

[C6-1] ITU-R M. 2160, “Framework and overall objectives of the future development of IMT for 2030 1and beyond”, 2023, https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2160-0-202311-I!!PDF-E.pdf.

[C6-2] 5G-IA, “European vision for the 6G network ecosystem,” 2021, https://5g-ppp.eu/wp-content/uploads/2021/06/WhitePaper-6G-Europe.pdf.

[C6-3] GSMA, “2019 mobile industry impact report: Sustainable development goals,” 2019.

[C6-4] G. P. Fettweis and H. Boche, “6G: The personal tactile internet—and open questions for information theory,” IEEE BITS Inf. Theory Mag.,171–82, 2021.

[C6-5] N. Rajatheva et al, “White paper on broadband connectivity in 6G,” 2020, arXiv preprint: https://arxiv.org/abs/2004.14247.

[C6-6] Next G Alliance Report, “6G applications and use cases,” https://nextgalliance.org/white_papers/6g-applications-and-use-cases/, June 2022.

[C6-7] Hexa-X-II, “Deliverable D1.2: 6G use cases and requirements”, Dec. 2023.

[C6-8] Hexa-X-II, “Deliverable D4.2: Radio design and spectrum access requirements and key enablers for 6G evolution”, Oct. 2023

[C6-9] Report ITU-R M.2410-0, “Minimum requirements related to technical performance for IMT-2020 radio interface(s),” 11/2017, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2410-2017-PDF-E.pdf.

[C6-10] 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.

[C6-11] G. Berardinelli et al, “Extreme communication in 6G: Vision and challenges for ‘in-X’ subnetworks,” IEEE Open Journal of the Communications Society, vol. 2, pp. 2516-2535, 2021.

[C6-12] 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.

[C6-13] Cisco Annual Internet Report (2018–2023) White Paper, 2020.

[C6-14] 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.

[C6-15] 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.

[C6-16] 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.

[C6-17] M. Maddah-Ali and U. Niesen, “Fundamental limits of caching,” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856-2867, May 2014.

[C6-18] 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.

[C6-19] E. Lampiris and P. Elia, "Adding transmitters dramatically boosts coded-caching gains for finite file sizes," IEEE J. Sel. Areas Commun., vol. 36, no. 6, pp. 1176-1188, June 2018.

[C6-20] H. Zhao, A. Bazco-Nogueras and P. Elia, "Vector coded caching multiplicatively increases the throughput of realistic downlink systems," IEEE Transactions on Wireless Communications, vol. 22, no. 4, pp. 2683-2698, April 2023.

[C6-21] M. Bayat, R. K. Mungara and G. Caire, “Achieving spatial scalability for coded caching over wireless networks,” arXiv preprint: arXiv:1803.05702, 2018.

[C6-22] K. Wan and G. Caire, “On coded caching with private demands”, submitted to TIT, available online: arXiv:1908.10821 [cs.IT], 2019.

[C6-23] D. Bethanabhotla, G. Caire and M. J. Neely, “Wiflix: Adaptive video streaming in massive MU-MIMO wireless networks,” IEEE Trans. Wireless Communications, vol. 15, no. 6, pp. 4088–103, 2016

[C6-24] 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.

[C6-25] F. Baccelli and B. Blaszczyszyn, “Spatial modeling of wireless communications – A stochastic geometry approach”, Foundations and Trends in Networking, NOW Publishers, 2009.

[C6-26] J. G. Andrews, F. Baccelli and R. K. Ganti, “A tractable approach to coverage and rate in cellular networks,” *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122-3134, 2011.

[C6-27] A. Alammouri, J. G. Andrews, and F. Baccelli, “A unified asymptotic analysis of area spectral efficiency in ultradense cellular networks” IEEE Trans. Information Theory, vol. 65, no. 2, 2019.

[C6-28] N. Okati, T. Riihonen, D. Korpi, I. Angervuori, and R. Wichman, “Downlink coverage and rate analysis of low earth orbit satellite constellations using stochastic geometry,” IEEE Trans. Commun., vol 68, no. 8, 2020.

[C6-29] 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.

[C6-30] 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.

[C6-31] Y. Tan and H. Haas, “Coherent LiFi system with spatial multiplexing,” IEEE Transactions on Communications, vol. 69, no. 7, pp. 4632-4643, July 2021.

[C6-32] 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.

[C6-33] C. Lee et al, “26 Gbit/s LiFi system with laser-based white light transmitter,” Journal of Lightwave Technology, vol. 40, no. 5, pp. 1432-1439, March, 2022.

[C6-34] C. Chen *et al*, "100 Gbps indoor access and 4.8 Gbps outdoor point-to-point LiFi transmission systems using laser-based light sources," *Journal of Lightwave Technology*, vol. 42, no. 12, pp. 4146-4157, June, 2024.

[C6-35] 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.

[C6-36] 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.

[C6-37] E. Sarbazi, H. Kazemi, M. Dehghani Soltani, M. Safari and H. Haas, “A Tb/s indoor optical wireless access system using VCSEL arrays,” IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2020.

[C6-38] H. Kazemi et al, “A Tb/s indoor MIMO optical wireless backhaul system using VCSEL arrays,” *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 3995-4012, June 2022.

[C6-39] D. Tsonev, S. Videv and H. Haas, “Unlocking spectral efficiency in intensity modulation and direct detection systems,” IEEE J. Sel. Areas Commun., vol. 33, no. 9, pp. 1758-1770, Sept. 2015

[C6-40] I. F. Akyildiz, J. M. Jornet and C. Han, “TeraNets: Ultra-broadband communication networks in the terahertz band,” IEEE Wireless Communications Magazine, vol. 21, no. 4, pp. 130-135, August 2014.

[C6-41] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, J. M. Jornet, “Terahertz band communication: An old problem revisited and research directions for the next decade,” *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4250-4285, June 2022.

[C6-42] S. Aliaga, A. J. Alqaraghuli and J. M. Jornet, “Joint terahertz communication and atmospheric sensing in low earth orbit satellite networks: Physical layer design,” 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2022.

[C6-43] L. T. Wedage, et al, "Climate change sensing through terahertz communication infrastructure: A disruptive application of 6G networks," *IEEE Network*, vol. 38, no. 3, pp. 261-268, May 2024.

[C6-44] I. F. Akyildiz and J. M. Jornet, “The internet of nano-things,” IEEE Wireless Communication Magazine, vol. 17, no. 6, pp. 58-63, December 2010.

[C6-45] S. Abadal, C. Han, V. Petrov, L. Galluccio, I. F. Akyildiz and J. M. Jornet, "Electromagnetic nanonetworks beyond 6G: From wearable and implantable networks to on-chip and quantum communication," IEEE J. Sel. Areas Commun., vol. 42, no. 8, pp. 2122-2142, Aug. 2024.

[C6-46] P. Sen, J. V. Siles, N. Thawdar, J. M. Jornet, “Multi-kilometre and multi-gigabit-per-second sub-terahertz communications for wireless backhaul applications,” Nature Electronics, Dec., 2022.

[C6-47] S. Nie and Ian F. Akyildiz. “Channel modeling and analysis of inter-small-satellite links in terahertz band space networks," *IEEE Trans. Communications*, vol. 69, no. 12, pp. 8585-8599, Dec. 2021.

[C6-48] A. J. Alqaraghuli, J. V. Siles and J. M. Jornet, “The road to high data rates in space: Terahertz vs. optical wireless communication,” IEEE Aerospace and Electronic Systems Magazine, vol. 38, no. 6, pp. 8-13, June 2023.

[C6-49] J. V. Siles, K. B. Cooper, C. Lee, R. H. Lin, G. Chattopadhyay, and I. Mehdi, “A new generation of room-temperature frequency-multiplied sources with up to 10× higher output power in the 160-GHz–1.6-THz range,” IEEE Transactions on Terahertz Science and Technology, vol. 8, no. 6, 596-604, 2018.

[C6-50] J. Shi, M. C. Lo, L. Zhang, O. Ozolins, ... and L. K. Oxenløwe, “Integrated dual-laser photonic chip for high-purity carrier generation enabling ultrafast terahertz wireless communications,” Nature communications, 13(1), 1388, March 2022.

[C6-51] J. M. Jornet and I. F. Akyildiz, "Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks," IEEE J. Sel. Areas Commun., vol. 31, no. 12, pp. 685-694, Dec. 2013.

[C6-52] J. Crabb, C.-R. Xavier, J. M. Jornet, and G. R. Aizin, “Hydrodynamic theory of the Dyakonov-Shur instability in graphene transistors,” Physical Review B 104, no. 15, 2021.

[C6-53] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, “A new wireless communication paradigm through software-controlled metasurfaces,” IEEE Communications Magazine, vol. 56, no. 9, 162-169, 2018.

[C6-54] H. Abdellatif, V. Ariyarathna, S. Petrushkevich, A. Madanayake, J. M. Jornet, “A real-time ultra-broadband software-defined radio platform for terahertz communications,” in Proc. of the IEEE Conference on Computer Communications Workshops (INFOCOM) - Demo Track, May 2022.

[C6-55] Y. J. Guo, M. Ansari, R. W. Ziolkowski and N. J. G. Fonseca, “Quasi-optical multi-beam antenna technologies for B5G and 6G mmWave and THz networks: A review," IEEE Open Journal of Antennas and Propagation, vol. 2, pp. 807-830, 2021.

[C6-56] 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.

[C6-57] Y. Chen, Y. Li, C. Han, Z. Yu, and G. Wang, “Channel measurement and ray-tracing-statistical hybrid modeling for low-terahertz indoor communications,” IEEE Transactions on Wireless Communications, vol. 20, no. 12, 8163-8176, 2021.

[C6-58] Y. Xing, T.S. Rappaport, and A. Ghosh, “Millimeter wave and sub-THz indoor radio propagation channel measurements, models, and comparisons in an office environment,” IEEE Communications Letters 25, no. 10, 3151-3155, 2021.

[C6-59] C. Han, W. Gao, N. Yang, and J.M. Jornet, “Molecular absorption effect: A double-edged sword of terahertz communications,” IEEE Wireless Communications, vol. 30, no. 4, pp. 140-146, 2022.

[C6-60] C. Han et al, “Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis,” IEEE Communications Surveys & Tutorials 24.3, pp. 1670-1707, 2022.

[C6-61] P. Sen, J. Hall, M. Polese, V. Petrov, D. Bodet, F. Restuccia, T. Melodia and J. M. Jornet, “Terahertz communications can work in rain and snow: Impact of adverse weather conditions on channels at 140 GHz,” in the Proc. of the 6th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems (mmNets), October 2022.

[C6-62] 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.

[C6-63] D. Bodet, P. Sen, Z. Hossain, N. Thawdar, and J.M. Jornet, “Hierarchical bandwidth modulations for ultra-broadband communications in the terahertz band,” IEEE Transactions on Wireless Communications, 22(3), 1931-1947, 2022.

[C6-64] A. Singh *et al*, "Wavefront engineering: Realizing efficient terahertz band communications in 6G and beyond," *IEEE Wireless Communications*, vol. 31, no. 3, pp. 133-139, June 2024.

[C6-65] H. Guerboukha, B. Zhao, Z. Fang, E. Knightly, and D. M. Mittleman. "Curving THz wireless data links around obstacles," Communications Engineering 3, no. 1: 58, 2024.

[C6-66] I. V. Reddy, D. Bodet, A. Singh, V. Petrov, C. Liberale, and J. M. Jornet, “Ultrabroadband terahertz-band communications with self-healing bessel beams." Communications Engineering 2, no. 1: 70, 2023.

[C6-67] Q. Xia, Z. Hossain, M. Medley, and J. M. Jornet, "Synchronization and medium access control protocol for terahertz-band communication networks," IEEE Transactions on Mobile Computing, vol. 20, no. 1., pp. 2-18, January 2021.

[C6-68] Q. Xia and J. M. Jornet, "Multi-hop relaying distribution strategies for terahertz-band communication networks: A cross-layer analysis," IEEE Transactions on Wireless Communications, vol. 21, no. 7, 5075-5089, 2021.

[C6-69] P. Sen, V. Ariyarathna, A. Madanayake, and J. M. Jornet, "A versatile experimental testbed for ultrabroadband communication networks above 100 GHz," Computer Networks 193: 108092, 2021.

[C6-70] 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.

[C6-71] 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.

[C6-72] 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.

[C6-73] 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.

[C6-74] T. Laas, J. A. Nossek, S. Bazzi and W. Xu, "On reciprocity in physically consistent TDD systems with coupled antennas," IEEE Trans. Wireless Communications, vol. 19, no. 10, pp. 6440-6453, 2020.

[C6-75] C. Fager, T. Eriksson, F. Barradas, K. Hausmair, T. Cunha and J. C. Pedro, "Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context," IEEE Microwave Magazine, vol. 20, no. 5, pp. 35-49, May 2019

[C6-76] 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.

[C6-77] 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.

[C6-78] A. Singh, M. Andrello, N. Thawdar and J. M. Jornet, "Design and operation of a graphene-based plasmonic nano-antenna array for communication in the terahertz band," IEEE J. Sel. Areas Commun., vol. 38, no. 9, pp. 2108-2117, Sep. 2020

[C6-79] 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.

[C6-80] E. De Carvalho et al, "Non-Stationarities in extra-large-scale massive MIMO," IEEE Wireless Commun., vol. 27, no.4, pp. 74–80, Aug. 2020.

[C6-81] C. Huang et al, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," IEEE Wireless Commun., vol. 27, no.5, pp. 118–125, Oct. 2020.

[C6-82] G. Interdonato, E. Björnson, H. Ngo, P. Frenger and E. Larsson, "Ubiquitous cell-free massive MIMO communications," EURASIP J Wireless Com Network, Aug. 2019.

[C6-83] 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.

[C6-84] 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.

[C6-85] 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, pp. 139–159, 2011.

[C6-86] 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, pp. 248–276, 2016.

[C6-87] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO Communications," EURASIP J. Wireless Commun. and Networking, 2019.

[C6-88] X. Xu, D. Wang, X. Tao, T. Svensson, "Resource pooling for frameless network architecture with adaptive resource allocation", SCI. CHINA Inf. Sci., vol. 56, 2013.

[C6-89] V. Jungnickel et al, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," IEEE Commun.Mag., vol. 52, no. 5, pp. 44–51, 2014.

[C6-90] 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.

[C6-91] 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.

[C6-92] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5398-5409, Nov. 2019.

[C6-93] C. Huang, A. Zappone, M. Debbah and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in Proc. IEEE ICASSP, 2018.

[C6-94] R. Liu, Q. Wu, M. Di Renzo and Y. Yuan, "A path to smart radio environments: An industrial viewpoint on reconfigurable intelligent surfaces," IEEE Wireless Communications, Jan. 2022.

[C6-95] S. Hu, F. Rusek and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," IEEE Trans. Signal Processing, vol. 66, no. 10, pp. 2746–2758, May 2018.

[C6-96] S. Hu, F. Rusek and O. Edfors, "Beyond massive MIMO: The potential of positioning with large intelligent surfaces," IEEE Trans. Signal Processing, vol. 66, no. 7, pp. 1761–1774, April 2018.

[C6-97] P. Frenger, J. Hederen, M. Hessler and G. Interdonato, "Improved antenna arrangement for distributed massive MIMO," Patent application WO2018103897, 2017.

[C6-98] J. Guerreiro, R. Dinis and P. Carvalho, "On the optimum multicarrier performance with memoryless nonlinearities," IEEE Trans. Commun., Vol. 63, No. 2, pp. 498 - 509, February 2015.

[C6-99] 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.

[C6-100] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," arXiv: 2203.03176, 2022.

[C6-101] 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

[C6-102] Y. Zeng, R. Zhang and T.J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," IEEE Communications Magazine, 54 (5), pp. 36-42, 2016.

[C6-103] C. Xu, X. Liao, J. Tan, H. Ye, and H. Lu, "Recent research progress of unmanned aerial vehicle regulation policies and technologies in urban low altitude," IEEE Access, 8, pp. 74175-74194, 2020.

[C6-104] P. S. Bithas, V. Nikolaidis, A. G. Kanatas and G. K. Karagiannidis, "UAV-to-ground communications: Channel modeling and UAV selection," IEEE Trans. Commun., vol. 68, no. 8, pp. 5135-5144, Aug. 2020.

[C6-105] H. K. Armeniakos, K. Maliatsos, P.S. Bithas, and A.G. Kanatas, "A stochastic geometry-based performance analysis of a UAV corridor-assisted IoT network," Frontiers in Communications and Networks, 5, 2024.

[C6-106] H. Q. Ngo, G. Interdonato, E. G. Larsson, and J. G. Andrews, "Ultradense cell-free massive MIMO for 6G: Technical overview and open questions," Proc. IEEE, May, 2024.

[C6-107] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," Foundations and Trends® in Signal Processing, vol. 14, no. 3-4, pp. 162-472, 2021.

[C6-108] H. Lu, Y. Zeng, C. You, Y. Han, J. Zhang, Z. Wang, Z. Dong, S. Jin, C.-X. Wang, T. Jiang, X. You, and R. Zhang, "A tutorial on nead-field XL-MIMO communications towards 6G," arXiv: 2310.11044v3, Apr., 2024.

[C6-109] P. Ramezani, A. Kosasih, A. Irshad, and E. Björnson, "Exploiting the depth and angular domains for massive near-field spatial multiplexing," IEEE BITS the Information Theory Magazine, vol. 3, no. 1, pp. 18-26, Mar., 2023.

[C6-110] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field communications for 6G: Fundamentals, challenges, potentials, and future directions," arXiv: 2203.16318v1, Mar., 2022.

[C6-111] J. Zhu, Z. Wan, L. Dai, M. Debbah, and H. V. Poor, "Electromagnetic information theory: Fundamentals, modeling, applications, and open problems," arXiv: 2212.02882v2, Jun., 2023.

[C6-112] E. G. Larsson, "Massive synchrony in distributed antenna systems," IEEE Transactions on Signal Processing, vol. 72, s. 855-866, Jan., 2024.

[C6-113] K.-K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Fluid antenna systems," IEEE Trans. Wireless Commun., vol. 20, no. 3, pp. 1950–1962, Mar. 2021.

[C6-114] J. O. Martinez, J. R. Rodriguez, Y. Shen, K.-F. Tong, K.-K. Wong, and A. G. Armada, "Toward liquid reconfigurable antenna arrays for wireless communications," IEEE Communications Magazine, vol. 60, pp. 145–151, Dec. 2022.

[C6-115] J. O. Martinez and A. G. Armada, "Realistic correlation modeling for fluid antenna systems", in Proc. IEEE International Conference on Communications (ICC), pp. 1-, Jun. 2024.

[C6-116] P. Mukherjee, C. Psomas, and I. Krikidis, "On the level crossing rate of fluid antenna systems," in IEEE SPAWC, pp. 1–5, July 2022.

[C6-117] K. Wong, K. Tong, C. Chae, "Fluid antenna system—Part II: Research opportunities", IEEE Communications Letters, Vol. 27, no. 8, pp, 1924 – 1928, Aug. 2023.

[C6-118] 3GPP TR 38.801, Study on New Radio Access Technology (Release 14), 2017.

[C6-119] G. Wunder et al, "5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications," IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, Feb. 2014.

[C6-120] Y. Medjahdi, et al, "On the road to 5G: Comparative study of physical layer in MTC context," IEEE Access, vol. 5, pp. 26556-26581, 2017.

[C6-121] A. G. Armada, "Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM)," IEEE Trans. Broadcast., vol. 47, no. 153–159, Jun. 2001.

[C6-122] M. Girotto and A. M. Tonello, "Orthogonal design of cyclic block filtered multitone modulation," IEEE Trans. Commun., vol. 64, no. 11, pp. 4667–4679, Nov. 2016.

[C6-123] J. Abdoli, M. Jia and J. Ma, "Filtered OFDM: A new waveform for future wireless systems," in *IEEE SPAWC*, pp. 66-70, 2015.

[C6-124] 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.

[C6-125] X. Yu, Y. Guanghui, Y. Xiao, Y. Zhen, X. Jun and G. Bo, "FB-OFDM: A novel multicarrier scheme for 5G," in Proc. Eur. Conf. Netw. Commun. (EuCNC), pp. 271-276, Jun. 2016.

[C6-126] R. Nissel and M. Rupp, "Pruned DFT-Spread FBMC: Low PAPR, low latency, high spectral efficiency," IEEE Trans. Commun., vol. 66, no. 10, pp. 4811-4825, Oct. 2018.

[C6-127] 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.

[C6-128] 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.

[C6-129] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch and R. Calderbank, "Orthogonal time frequency space modulation," in Proc. IEEE WCNC, 2017.

[C6-130] 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.

[C6-131] 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.

[C6-132] 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.

[C6-133] A. Nimr, M. Chafii and G. Fettweis, "Unified low complexity radix-2 architectures for time and frequency-domain GFDM modem," IEEE Circuits and Systems Magazine, vol. 18, no. 4, pp. 18-31, Fourthquarter 2018.

[C6-134] R. Bomfin, M. Chafii, A. Nimr and G. Fettweis, "A robust baseband transceiver design for doubly-dispersive channels," IEEE Trans. Wireless Commun., vol. 20, no. 8, pp. 4781-4796, Aug. 2021.

[C6-135] R. Bomfin, A. Nimr, M. Chafii, and G. Fettweis, "A Robust and Low-Complexity Walsh-Hadamard Modulation for Doubly-Dispersive Channels," IEEE Communications Letters, vol. 25, no. 3, pp. 897-901, March 2021.

[C6-136] B. Murmann, "ADC Performance Survey 1997-2023," [Online]. Available: https://github.com/bmurmann/ADC-survey.

[C6-137] G. Fettweis, M. Dörpinghaus, S. Bender, L. Landau, P. Neuhaus, and M. Schlüter, "Zero crossing modulation for communication with temporally oversampled 1-bit quantization," in Proc. Asilomar Conference on Signals, Systems, and Computers (ASILOMAR), 2019.

[C6-138] F. Gast, M. Dörpinghaus, F. Roth, and G. Fettweis, "A new spatio-temporal model for data rate distributions in mobile networks," Dresden, Germany, Mar. 2024, pp. 103–108.

[C6-139] G. P. Fettweis and H. Boche, "6G: The personal tactile internet—and open questions for information theory," IEEE BITS Inf. Theory Mag.,1, no. 1, pp. 71–82, 2021.

[C6-140] 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), 2298-2323, third quarter 2018.

[C6-141] Y. Yuan, Z. Yuan and L. Tian, "5G non-orthogonal multiple access study in 3GPP," IEEE Communications Magazine, vol. 58, no. 7, pp. 90-96, July 2020.

[C6-142] Y. Yuan et al, "Non-orthogonal transmission technology in LTE evolution," IEEE Communications Magazine, vol. 54, no. 7, pp. 68-74, July 2016.

[C6-143] L. Zhu, J. Zhang, Z. Xiao, X. Cao and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," IEEE Wireless Commun. Letters, vol. 8, no. 2, pp. 328-331, April 2019.

[C6-144] 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.

[C6-145] 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.

[C6-146] M. Rebhi, K. Hassan, K. Raoof and P. Chargé, "Sparse code multiple access: Potentials and challenges," IEEE Open Journal of the Communications Society, vol. 2, pp. 1205-1238, 2021

[C6-147] 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.

[C6-148] Z. Yuan, Y. Ma, Y. Hu, and W. Li, High-efficiency full-duplex V2V communication, in Proc. 2nd 6G Wireless Summit, Levi, Finland, Mar. 2020.

[C6-149] 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.

[C6-150] 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.

[C6-151] 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.

[C6-152] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," arXiv preprint arXiv:1906.05774, 2019.

[C6-153] 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.

[C6-154] S. Miao, C. Kestel, L. Johannsen, M. Geiselhart, L. Schmalen, A. Balatsoukas-Stimming, G. Liva, N. Wehn, and S. ten Brink, "Trends in channel coding for 6G," *Proc. IEEE*, 2024.

[C6-155] K. R. Duffy, J. Li and M. Médard, "Capacity-achieving Guessing random additive noise decoding," *IEEE Transactions on Information Theory*, vol. 654023-4040, July 2019.

[C6-156] K. R. Duffy, M. Médard and W. An, "Guessing random additive noise decoding with symbol reliability information (SRGRAND)," *IEEE Trans. Communications*, vol. 70, no. 1, pp. 3-18, Jan. 2022.

[C6-157] H. Ma, W. Leung, X. Yan, K. Law and M. Fossorier, "Delayed bit interleaved coded modulation," in 2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), pp. 86–90, 2016.

[C6-158] Y. Liao, M. Qiu and J. Yuan, "Design and analysis of delayed bit-interleaved coded modulation with LDPC codes," IEEE Transactions on Communications, vol. 69, pp. 3556–3571, 2021.

[C6-159] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," IEEE Transactions on Information Theory, 39(3), pp. 913-929, 1993.

[C6-160] 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.

[C6-161] 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.

[C6-162] M. Pikus and W. Xu, "Bit-level probabilistically shaped coded modulation," IEEE Communications Letters, vol. 21, no. 9, pp. 1929–1932, Sept. 2017.

[C6-163] 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.

[C6-164] O. İşcan, R. Böhnke and W. Xu, "Shaped polar codes for higher order modulation," *IEEE Communications Letters*, vol. 22, no. 2, pp. 252-255, 2018.

[C6-165] 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.

[C6-166] H. Wymeersch, G. Seco-Granados, G. Destino, et al, "5G mmWave positioning for vehicular networks," IEEE Wireless Communications, pp. 80–86, Dec. 2017.

[C6-167] M. Chiani, A. Giorgetti and E. Paolini, "Sensor radar for object tracking," Proceedings of the IEEE, vol. 106, no. 6, pp. 1022-1041, Jun. 2018.

[C6-168] J. A. Zhang, M. L. Rahman, K. Wu, X. Huang, Y. J. Guo, S. Chen, and J. Yuan, "Enabling joint communication and radar sensing in mobile networks – A survey," IEEE Commun. Surveys Tuts., pp. 1–41, 2021.

[C6-169] Hexa-X, "Hexa-X architecture for B5G/6G networks – final release," 2023, https://hexa-x.eu/wp-content/uploads/2023/07/Hexa-X-D1.8-Final.pdf.

[C6-170] W. Xu, S. Shojaee and K. Manolakis, "Carrier-aggregated timing estimation for radio positioning," in Proc. *IEEE VTC2019-Spring*, 2019.

[C6-171] L. Pucci, E. Paolini, and A. Giorgetti, "System-level analysis of joint sensing and communication based on 5G new radio," IEEE J. Select. Areas in Commun., vol. 40, no. 7, pp. 2043–2055, Jul. 2022.

[C6-172] 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.

[C6-173] J. Yang, C.-K. Wen, and S. Jin, "Hybrid active and passive sensing for SLAM in wireless communication systems," IEEE J. Sel. Areas Commun., vol. 40, no. 7, pp. 2146-2163, 2022.

[C6-174] H. Kim et al, "5G mmWave cooperative positioning and mapping using multi-model PHD filter and map fusion," arXiv preprint arXiv:1908.09806, 2019.

[C6-175] L. Pucci, E. Matricardi, E. Paolini, W. Xu, and A. Giorgetti, "Performance of a 5G NR-based Bistatic Joint Sensing and Communication System," in Proc. IEEE ICC Workshop, 2022.

[C6-176] M. Kiviranta, I. Moilanen, and J. Roivainen, "5G radar: Scenarios, numerology and simulations," in IEEE International Conference on Military Communications and Information Systems (ICMCIS), 2019.

[C6-177] 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

[C6-178] C. Aydogdu et al, "Radar interference mitigation for automated driving: Exploring proactive strategies," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 72-84, July 2020.

[C6-179] 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.

[C6-180] A. Dammann, R. Raulefs, S. Zhang, "On prospects of positioning in 5G," in Proc. IEEE International Conference on Communication Workshop (ICCW), 2015.

[C6-181] 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.

[C6-182] E. Favarelli et al, "Map fusion and heterogeneous objects tracking in joint sensing and communication networks," European Radar Conf. (EuRAD), Berlin, Germany, Sept. 2023.

[C6-183] E. Matricardi, L. Pucci, E. Paolini, W. Xu, and A. Giorgetti, "Performance Analysis of a Multistatic Joint Sensing and Communication System," in IEEE PIMRC, Toronto, Canada, Sept. 2023.

[C6-184] 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.

[C6-185] J. A. Zhang, K. Wu, X. Huang, Y.J. Guo, D. Zhang, and R. W. Heath, "Integration of radar sensing into communications with asynchronous transceivers," IEEE Communications Magazine, 2022

[C6-186] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, "RAPID: Retrofitting IEEE 802.11ay access points for indoor human detection and sensing," IEEE Transactions on Mobile Computing,

[C6-187] X. Li, Y. Cui, J. A. Zhang, F. Liu, D. Zhang, and L. Hanzo, "Integrated human activity sensing and communications," IEEE Communications Magazine, 2022.

[C6-188] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," IEEE Transactions on Mobile Computing, 2022.

[C6-189] Y. Chen, X. Su, Y. Hu, and B. Zeng, "Residual carrier frequency offset estimation and compensation for commodity WiFi," IEEE Transactions on Mobile Computing, vol. 19, no. 12, pp. 2891–2902, 2019.

[C6-190] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," IEEE Journal of Selected Topics in Signal Processing, 2021.

[C6-191] J. Pegoraro, J. O. Lacruz, T. Azzino, M. Mezzavilla, M. Rossi, J. Widmer, and S. Rangan. "JUMP: Joint communication and sensing with unsynchronized transceivers made practical," IEEE Transactions on Wireless Communications, Feb. 2024.

[C6-192] Y. Polyanskiy, "A perspective on massive random-access," in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2523-2527, 2017.

[C6-193] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang and J. Xu, "Multi-user shared access for internet of things," in Proc. IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, 2016.

[C6-194] Z. Yuan, Y. Hu, W. Li and J. Dai, "Blind multi-user detection for autonomous grant-free high-overloading multiple-access without reference signal," IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, 2018.

[C6-195] Z. Yuan, W. Li, Y. Hu, H. Tang, J. Dai and Y. Ma, "Blind multi-user detection based on receive beamforming for autonomous grant-free high-overloading multiple access," IEEE 2nd 5G World Forum (5GWF), Dresden, Germany, pp. 520-523, 2019.

[C6-196] Z. Yuan, W. Li, Z. Li, Y. Ma and Y. Hu, "Contention-based grant-free transmission with independent multi-pilot scheme," in Proc. IEEE VTC2020-Fall, Victoria, BC, Canada, 2020.

[C6-197] Z. Yuan, Z. Li, W. Li, Y. Ma and C. Liang, "Contention-based grant-free transmission with extremely sparse orthogonal pilot scheme," in IEEE VTC-Fall, 2021.

[C6-198] H.A. Inan, P. Kairouz and A. Ozgur, "Sparse combinatorial group testing for low-energy massive random access," arXiv preprint arXiv:1711.05403, 2017.

[C6-199] 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.

[C6-200] 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. 148-150, 2015.
[C6-201] 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.
[C6-202] 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.
[C6-203] 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.
[C6-204] Z. Chen, F. Sohrabi and W. Yu, “Sparse activity detection for massive connectivity,” arXiv preprint arXiv:1801.05873, 2018.
[C6-205] S. Haghighatshoar, P. Jung and G. Caire, “Improved scaling law for activity detection in massive MIMO systems,” arXiv preprint arXiv:1803.02288, 2018.
[C6-206] 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.
[C6-207] 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.
[C6-208] 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.
[C6-209] G. Fettweis et al, “The tactile internet,” ITU-T Technology Watch Report, August 2014.
[C6-210] H. Lasi, P. Fettke, H.-G. Kemper, et al, “Industry 4.0”, Bus Inf Syst Eng 6, pp. 239–242, 2014.
[C6-211] 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.
[C6-212] S. Ali, W. Saad & D. Steinbach (Eds.), “White paper on machine learning in 6G wireless communication networks,” University of Oulu, http://urn.fi/urn:isbn:9789526226736, 2020.
[C6-213] H. Ye, G. Y. Li, and B. Juang, “Power of deep learning for channel estimation and signal detection in OFDM systems,” IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 114–17, 2018.
[C6-214] T. OShea and J. Hoydis, “An introduction to deep learning for the physical layer,” IEEE Transactions on Cognitive Communication Networks, vol. 3, no. 4, pp. 563–75, 2017.
[C6-215] A. Giannopoulos, S. Spantideas, N. Kapsalis, P. Karkazis and P. Trakadas, “Deep reinforcement learning for energy-efficient multi-channel transmissions in 5G cognitive hetnets: Centralized, decentralized and transfer learning based solutions,” IEEE Access, vol. 9, pp. 129358-129374, 2021.
[C6-216] K. Yang, T. Jiang, Y. Shi and Z. Ding, “Federated learning via over-the-air computation,” IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2022-2035, March 2020.
[C6-217] S. Li and S. Avestimehr, “Coded computing: Mitigating fundamental bottlenecks in large-scale distributed computing and machine learning,” Foundations and Trends in Communications and Information Theory, vol. 17, no. 1, pp 1-148, 2020.
[C6-218] J. D.V. Sánchez, L. Urquiza-Aguiar, M.C.P. Paredes and D.P.M. Osorio, “Survey on physical layer security for 5G wireless networks,” Annals of Telecommunications, 76 (3), 155-174, 2021.

Optical Networks

[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 and D. T. Neilson, “From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade,” Journal of Lightwave Technol., vol. 35, 5, pp. 1099 – 1115, 2017.
[C7-3] Z. Ghassemlooy, et al., "Emerging Optical Wireless Communications-Advances and Challenges," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 9, pp. 1738-1749, Sept. 2015
[C7-4] V. Jamali, et al., "Intelligent Reflecting Surface Assisted Free-Space Optical Communications," in IEEE Communications Magazine, vol. 59, no. 10, pp. 57-63, October 2021,
[C7-5] H. Elgala, et al "Indoor optical wireless communication: potential and state-of-the-art," in IEEE Communications Magazine, vol. 49, no. 9, pp. 56-62, September 2011,
[C7-6] G. Cossu, et al. "Sea-Trial of Optical Ethernet Modems for Underwater Wireless Communications," J. Lightwave Technol. 36, 5371-5380 (2018)

Non-Terrestrial Networks and Systems

[C8-1]. A. Guidotti et al., "Role and Evolution of Non-Terrestrial Networks Toward 6G Systems," in IEEE Access, vol. 12, pp. 55945-55963, 2024, doi: 10.1109/ACCESS.2024.3389459.

[C8-2]. A. Guidotti, A. Vanelli-Coralli and C. Amatetti, "Federated Cell-Free MIMO in Non-Terrestrial Networks: Architectures and Performance," in IEEE Transactions on Aerospace and Electronic Systems, doi: 10.1109/TAES.2024.3362769.

[C8-3]. D. Tuzi, T. Delamotte and A. Knopp, "Satellite Swarm-Based Antenna Arrays for 6G Direct-to-Cell Connectivity," in IEEE Access, vol. 11, pp. 36907-36928, 2023, doi: 10.1109/ACCESS.2023.3257102.

[C8-4]. M. Majamaa, H. Martikainen, J. Puttonen and T. Hämäläinen, "Satellite-Assisted Multi-Connectivity in Beyond 5G," 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Boston, MA, USA, 2023, pp. 413-418, doi: 10.1109/WoWMoM57956.2023.00073.

[C8-5]. M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376-1411.

[C8-6]. R. Campana, C. Amatetti and A. Vanelli-Coralli, "O-RAN based Non-Terrestrial Networks: Trends and Challenges," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 268-269

[C8-7]. Fontanesi, G. & Ortiz, Flor & Lagunas, Eva & Monzon Baeza, Victor & Vázquez, M. & Vasquez, Juan & Minardi, Mario & Vu, H. & Jubba Honnaiah, Puneeth & Lacoste, C. & Drif, Youssouf & Abdu, Tedros & Eappen, Geoffrey & Ur Rehman, Junaid & Garcés-Socarrás, Luis & Martins, Wallace & Henarejos, Pol & Al-Hraishawi, Hayder & Merlano Duncan, J.C. & Chatzinotas, Symeon. (2023). "Artificial Intelligence for Satellite Communication and Non-Terrestrial Networks: A Survey", arXiv preprint (2023).

[C8-8]. C. J. Vaca-Rubio, L. Blanco, R. Pereira, and M. Caus, "Kolmogorov-Arnold Networks (KANs) for Time Series Analysis", arXiv preprint (2024).

[C8-9]. 3GPP TS 38.305, "NG Radio Access Network (NG-RAN); Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN," April 2024.

[C8-10]. I. del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," Acta Astronautica, 2019.

[C8-11]. N. Pachler, I. del Portillo, E. F. Crawley and B. G. Cameron, "An Updated Comparison of Four Low Earth Orbit Satellite Constellation Systems to Provide Global Broadband," 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 2021, pp. 1-7, doi: 10.1109/ICCWorkshops50388.2021.9473799.

[C8-12]. Rashid Khouli, Leo Frank, Alexander Hofmann, "Functional Split Evaluation in NTN for LEO Satellites," Proceedings of the International Communications Satellite Systems Conference (ICSSC) 2023, Bradford, UK, Oct. 2023.

[C8-13]. S. Choi and F. Baccelli, `Cox Point Processes for Multi-Altitude LEO Satellite Networks'', to appear in IEEE Trans. Veh. Technol. 2024.

[C8-14]. J. Andrews, F. Baccelli and R. Ganti, ``A Tractable Approach to Cellular Network Modeling'', IEEE Transactions on Communications, Vol. 59, No. 11, 2011.

[C8-15]. HORIZON-JU-SNS-2022 Project 6G-NTN (6G Non-Terrestrial Networks), Deliverable 2.3 "Report on System Requirements," January 2024. Available: https://www.6g-ntn.eu/public-deliverables/

[C8-16]. ESA Project EAGER (Innovative Technologies and Techniques for SatCom Beyond 5G), White Paper "Architectures, services, and technologies towards 6G Non-Terrestrial Networks," February 2023. Available: https://www.eagerproject.eu/wp-content/uploads/sites/82/2023/02/EAGER_White_Paper_long_version_v3.pdf

[C8-17]. Nicolas Chuberre, "Integration of NTN in 6G: Requirements, Enablers and Technology Building Blocks," ETSI Conference on "Non-Terrestrial Networks, A Native Component Of 6G," April 2024.

[C8-18]. W. Stock, R. T. Schwarz, C. A. Hofmann and A. Knopp, "Survey On Opportunistic PNT With Signals From LEO Communication Satellites," in IEEE Communications Surveys & Tutorials

[C8-19]. A. Celik and A. M. Eltawil, "At the Dawn of Generative AI Era: A Tutorial-cum-Survey on New Frontiers in 6G Wireless Intelligence," in IEEE Open Journal of the Communications Society, vol. 5, pp. 2433-2489, 2024

[C8-20]. Clerkx et all, Multiple Access Techniques for Intelligent and Multi-Functional 6G: Tutorial, Survey, and Outlook, arXiv:2401.01433

[C8-21]. TR 38.811, "Study on New Radio (NR) to support non-terrestrial networks", REL-15

[C8-22]. I. Del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," Acta Astronautica, vol. 159, pp. 123–135, 2019.

[C8-23]. N. Pachler, I. del Portillo, E. F. Crawley, and B. G. Cameron, “An updated comparison of four low earth orbit satellite constellation systems to provide global broadband,” in 2021 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2021, pp. 1–7

[C8-24]. D. Bhattacharjee, A. U. Chaudhry, H. Yanikomeroglu, P. Hu and G. Lamontagne, "Laser Inter-Satellite Link Setup Delay: Quantification, Impact, and Tolerable Value," *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023, pp. 1-6, doi: 10.1109/WCNC55385.2023.10119016.

[C8-25]. R. M. Calvo *et al.*, "Optical Feeder Links for Future Very High-Throughput Satellite Systems in B5G Networks," *2020 European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020, pp. 1-4, doi: 10.1109/ECOC48923.2020.9333405.

[C8-26]. M. Neinavaie, J. Khalife and Z. M. Kassas, "Acquisition, Doppler Tracking, and Positioning With Starlink LEO Satellites: First Results," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 3, pp. 2606-2610, June 2022, doi: 10.1109/TAES.2021.3127488.

[C8-27]. M. M. Amay, J. Bas, M. Á. Vázquez and A. Pérez-Neira, "Modified Hierarchical Modulation for Hybrid RF-FSO Satellite Communication," *2024 27th International Workshop on Smart Antennas (WSA)*, Dresden, Germany, 2024, pp. 1-7, doi: 10.1109/WSA61681.2024.10512057.

[C8-28]. Baltaci A, Shortt K. Investigation of the Influence of LEO Constellation Dynamics on Optical Inter-satellite Links. In 2023 IEEE International Conference on Space Optical Systems and Applications (ICSOS) 2023 Oct 11 (pp. 121-127). IEEE.

[C8-29]. Consultative Committee for Space Data Systems, “Optical high data rate (HDR) communication – 1064 nm experimental specification”, CCSDS, Tech. Rep. CCSDS 141.11-O-1, 2019.

[C8-30]. E. Fernandez, J.A. Fraire, S. Figuerola, A. Camps, and J.A. Ruiz-de-Azua. “A Case Study of an Hybrid RF and Optical Inter-Satellite Link Terminal to Enhance Optical Pointing”. In European Conference on Networks and Communications (EUCNC). 2024

[C8-31]. Y. Liu, X. Li, D. Li, L. Zhang, and S. Huang, "Free-space optical (FSO) feeder link planning in space-ground integrated optical networks (SGIONs): trade off throughput and dynamics," Opt. Express 31, 35396-35418 (2023).

[C8-32]. ETSI GS NFV-IFA 031 V3.4.1: Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Requirements and interfaces specification for management of NFV-MANO, 2020.

[C8-33]. W. Li, Y. Lemieux, G. Gao, Z. Zhao, Y. Han, “Service mesh: challenges, state of the art, and future research opportunities,” Proc. 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE), San Francisco East Bay, CA, USA.

[C8-34]. ETSI: Satellite Earth Stations and Systems (SES). Broadband Satellite Multimedia, IP over Satellite. ETSI Technical Report, TR 101 985, V.1.1.2, 2002.

[C8-35]. L. Boero, R. Bruschi, F. Davoli, M. Marchese, F. Patrone, "Satellite networking integration in the 5G ecosystem: Research trends and open challenges", IEEE Network, vol. 32, no. 5, pp. 12-18, Sept./Oct. 2018.

[C8-36]. IEEE INRG International Network Generations Roadmap, 2023 edition. [Online]. Available: https://futurenetworks.ieee.org/home/sign-in?notauth=1.

[C8-37]. Dmitriev DD, Gladyshev AB, Ratuschnyak VN, Grithan OB. Software and hardware complex for the development and research of methods for broadband access to multimedia resources and the Internet. InJournal of Physics: Conference Series 2020 Apr 1 (Vol. 1515, No. 3, p. 032041). IOP Publishing.

[C8-38]. Del Portillo I, Cameron BG, Crawley EF. A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. Acta astronautica. 2019 Jun 1;159:123-35.

[C8-39]. Kafle VP, Sekiguchi M, Asaeda H, Harai H. Integrated Network Control Architecture for Terrestrial and Non-Terrestrial Network Convergence. IEEE Communications Standards Magazine. 2024 Mar 12;8(1):12-9.

[C8-40]. Vasilakos A, Zhang Y, Spyropoulos T. Delay tolerant networks. Boca Raton, FL, USA: CRC press; 2016 Apr.

[C8-41]. Pitts RL. CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS) WHITE PAPER ON THE USE OF BUNDLE PROTOCOL (BP) VERSION 7. InCCSDS Biannual Conference 2020 Oct 8.

[C8-42]. Fuste O, Marin-de Yzaguirre M, J.A. Ruiz-de-Azua. “Implementation of a protocol stack with DTN protocols for IoT services deployed from Non-Terrestrial Networks”. In74th International Astronautical Congress (IAC) 2023.

[C8-43]. Ward T. Delay tolerance for constrained IPv6 networks (Doctoral dissertation, University of Southampton).

[C8-44]. M. Blanchet, C. Huitema, and D. Bogdanović. “Revisiting the Use of the IP Protocol Stack in Deep Space: Assessment and Possible Solutions“. Draft-many-deepspace-ip-assessment-01. Online: https://datatracker.ietf.org/doc/draft-many-deepspace-ip-assessment/01/ (Last access 22nd May 2024)

[C8-45]. Kocak, Fatma, Kılıç, Adnan & Tapan, İlhan. (2016). Investigation of Shot Noise in Avalanche Photodiodes. Acta Physica Polonica A. 129. 721-723. 10.12693/APhysPolA.129.721.

[C8-46]. Jompol, Y., Roulleau, P., Jullien, T. et al. Detecting noise with shot noise using on-chip photon detector. Nat Commun 6, 6130 (2015). https://doi.org/10.1038/ncomms7130

[C8-47]. Y. Luo, H. -K. Mao, Q. Li and N. Chen, "An Information-Theoretic Secure Group Authentication Scheme for Quantum Key Distribution Networks," in IEEE Transactions on Communications, vol. 71, no. 9, pp. 5420-5431, Sept. 2023, doi: 10.1109/TCOMM.2023.3280561.

[C8-48]. F. Hou, Q. Zhou, Z. Jin and Y. Li, "Authenticated QKD Protocol Based on Single-Photon Interference," in IEEE Access, vol. 8, pp. 135357-135370, 2020, doi: 10.1109/ACCESS.2020.3009090

[C8-49]. Wang, W., Tamaki, K. & Curty, M. Measurement-device-independent quantum key distribution with leaky sources. Sci Rep 11, 1678 (2021). https://doi.org/10.1038/s41598-021-81003-2

[C8-50]. Huang, Wenti etal, Quantum hacking of free-space continuous-variable quantum key distribution by using a machine-learning technique, Physical Review A, 100,1 012316 (2019). https://link.aps.org/doi/10.1103/PhysRevA.100.012316

[C8-51]. M Stanley et al, Recent Progress in Quantum Key Distribution Network Deployments and Standards 2022 J. Phys.: Conf. Ser. 2416 012001

[C8-52]. Li, W., Zhang, L., Tan, H. et al. High-rate quantum key distribution exceeding 110 Mb s–1. Nat. Photon. 17, 416–421 (2023). https://doi.org/10.1038/s41566-023-01166-4

[C8-53]. Kamil Brádler *et al* 2016 *New J. Phys.* 18 073030DOI 10.1088/1367-2630/18/7/073030

[C8-54]. Diamanti, E., Lo, HK., Qi, B. et al. Practical challenges in quantum key distribution. npj Quantum Inf 2, 16025 (2016). https://doi.org/10.1038/npjqi.2016.25

[C8-55]. Weedbrook, Christian and Pirandola, Stefano and Lloyd, Seth and Ralph, Timothy C., Quantum Cryptography Approaching the Classical Limit, PhysRevLett.105.110501, 2010.

[C8-56]. Muhammad Zulfiqar Ali, Abdoalbaset Abohmra, Muhammad Usman, Adnan Zahid, Hadi Heidari, Muhammad Ali Imran, Qammer H. Abbasi, IET Quantum Communication, 2023, 4, 112-124. https://doi.org/10.1049/qtc2.12060

[C8-57]. Adarsh Jain, Abhishek Khanna, Jay Bhatt, Parthkumar V. Sakhiya, Shashank Kumar, Rohan S. Urdhwareshe, Nilesh M. Desai, Development of NavIC synchronized fully automated inter-building QKD framework and demonstration of quantum secured video calling, Optik, Volume 252, 2022,168438, ISSN 0030-4026, https://doi.org/10.1016/j.ijleo.2021.168438.

[C8-58]. Felix Ott, Lucas Heublein, Nisha Lakshmana Raichur, Tobias Feigl, Jonathan Hansen, Alexander Rügamer, Christopher Mutschler. Few-Shot Learning with Uncertainty-based Quadruplet Selection for Interference Classification in GNSS Data. In ICL-GNSS, June 2024.

[C8-59]. Nisha L. Raichur, Lucas Heublein, Tobias Feigl, Alexander Rügamer, Christopher Mutschler, Felix Ott. Bayesian Learning-driven Prototypical Contrastive Loss for Class-Incremental Learning. In arXiv preprint arXiv: 2405.11067, May 2024.

[C8-60]. Dorsaf Jdidi, Tobias Brieger, Tobias Feigl, David C. Franco, J. Rossouw van der Merwe, Alexander Rügamer, Jochen Seitz, and Wolfgang Felber. Machine Learning Compression for GNSS Interference Analysis. In Proc. of the Intl. Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+), pages 1176-1208, Denver, CO, September 2022. doi:10.33012/2022.18493.

[C8-61]. Nisha L. Raichur, Tobias Brieger, Dorsaf Jdidi, Carlo Schmitt, Birendra Ghimire, Felix Ott, Tobias Feigl, J. Rossouw van der Merwe, Alexander Rügamer, and Wolfgang Felber. Machine Learning-assisted GNSS Interference Monitoring Through Crowdsourcing. In Proc. of the Intl. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+), pages 1151-1175, Denver, CO, September 2022. doi:10.33012/2022.18492.

[C8-62]. Tobias Brieger, Nisha L. Raichur, Dorsaf Jdidi, Felix Ott, Tobias Feigl, J. Rossouw van der Merwe, Alexander Rügamer, and Wolfgang Felber. Multimodal Learning for Reliable Interference Classification in GNSS Signals. In Proc. of the Intl. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+), pages 3210-3234, Denver, Colorado, September 2022. doi:10.33012/2022.18586.

[C8-63]. J. Rossouw van der Merwe, David C. Franco, Dorsaf Jdidi, Tobias Feigl, Alexander Rügamer, and Wolfgang Felber. Low-cost COTS GNSS Interference Detection and Classification Platform: Initial Results. In Intl. Conf. on Localization and GNSS (ICL-GNSS), pages 1-8, Tampere, Finland, June 2022. doi:10.1109/ICL-GNSS54081.2022.9797025.

[C8-64]. J. Rossouw van der Merwe, David Contreras Franco, Jonathan Hansen, Tobias Brieger, Tobias Feigl, Felix Ott, Dorsaf Jdidi, Alexander Rügamer, and Wolfgang Felber. Low-Cost COTS GNSS Interference Monitoring, Detection, and Classification System. In MDPI Sensors, volume 23(7), March 2023. doi:10.3390/s23073452.

Opportunities for Devices and Components

[C9-1] 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.

[C9-2] 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.

[C9-3] S. Bicaïs and J.-B. Dore, "Phase Noise Model Selection for Sub-THz Communications," in 2019 IEEE Global Communication Conference (GLOBECOM), December 2019.

[C9-4] 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.

[C9-5] "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.

[C9-6] 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.

[C9-7] NEREID Roadmap: https://www.nereid-h2020.eu/roadmap

[C9-8] 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

[C9-9] 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).

[C9-10] PA survey: https://gems.ece.gatech.edu/PA_survey.html

[C9-11] DOTSEVEN European project: http://www.dotseven.eu

[C9-12] 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).

[C9-13] 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).

[C9-14] 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.

[C9-15] U. Peralagu, "CMOS-compatible GaN-based devices on 200mm-Si for RF applications: Integration and Performance", in 2019 IEEE International Electron Devices Meeting (IEDM).

[C9-16] P. Weckx et al., "Novel forksheet device architecture as ultimate logic scaling device towards 2nm", in 2019 IEEE International Electron Devices Meeting (IEDM).

[C9-17] 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.

[C9-18] 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

[C9-19] 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

[C9-20] 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.

[C9-21] 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

[C9-22] 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

[C9-23] 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.

[C9-24] Zheludev, N. I.; Kivshar, Y. S. From metamaterials to metadevices. Nat. Mater. 2012, 11, 917–924.

[C9-25] Silva, A.; Monticone, F.; Castaldi, G.; Galdi, V.; Alù, A. & Engheta, N. Performing mathematical operations with metamaterials. Science 2014, 343(6167), 160-163.

[C9-26] 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.

[C9-27] 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.

[C9-28] 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.

[C9-29] 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

[C9-30] 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

[C9-31] M. He et al., "High-performance hybrid silicon and lithium niobate Mach–Zehnder modulators for 100 Gbits−1 and beyond," Nature Photonics, vol. 13, pp. 359–364, Mar. 2019

[C9-32] U. Koch et al., “Monolithic High-Speed Transmitter Enabled by Bicmos-Plasmonic Platform” ECOC 2019, PD.1.4, Sep. 2019.

[C9-33] 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.

[C9-34] 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.

[C9-35] P. W. Juodalkis, et.al., “Optically sampled analog-to-digital converters”, IEEE Trans. Microwave Theory and Techniques, vol. 49, Oct. 2001.

[C9-36] A. Gatherer et al, Academic Press Library in Mobile and Wireless Communications, 2016

[C9-37] 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.

[C9-38] 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.

[C9-39] 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)

[C9-40] 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

[C9-41] 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

[C9-42] 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

[C9-43] 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)

[C9-44] 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)

[C9-45] 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.

[C9-46] K. Kim, S. Shin and S. Kang, “Stateful logic pipeline architecture,” IEEE International Symposium of Circuits and Systems (ISCAS), pp. 2497-2500, 2011.

[C9-47] 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.

[C9-48] NetWorld2020 ETP, “5G: Challenges, research priorities, and recommendations”, White Paper, September 2014. Available online at: https://www.networld2020.eu/.

[C9-49] 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.

[C9-50] Mirai botnet code. Available online at: https://github.com/jgamblin/Mirai-Source-Code.

[C9-51] M. Jian, Y. Zhao, “A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems,” IWCMC, 2020

[C9-52] Higginbotham, Stacey. "The internet of trash [Internet of Everything]." IEEE Spectrum 55.6 (2018): 17-17.

[C9-53] 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.

[C9-54] J. M. Rabaey "The swarm at the edge of the cloud - A new perspective on wireless", Symposium on VLSI Circuits - Digest of Technical Papers, IEEE, pp. 6-8, 2011.

[C9-55] E. A. Lee, J. Rabaey, B. Hartmann, J. Kubiatowicz, K. Pister, A. Sangiovanni-Vincentelli, S. A. Seshia, J. Wawrzynek, D. Wessel, “The swarm at the edge of the cloud” IEEE Des. Test., vol. 31, no. 3, pp. 8–20, Jun. 2014.

[C9-56] G. Fedrecheski, L. Caroline, C. De Biase, P. C. Calcina-Ccori, M. Knorich Zuffo, "Attribute-Based Access Control for the Swarm With Distributed Policy Management", Volume: 65, Issue: 1, IEEE Transactions on Consumer Electronics, IEEE, 2019.

[C9-57] Position paper, "Research Priorities on Microelectronics for 6G Networks R&I Activities", Editors : Bernard Barani, Alex Kaloxylos, available at https://6g-ia.eu

[C9-58] Orange WhiteOrange White Paper, "Mobile Network Technology Evolutions Beyond 2030", April 2024, available at https://hellofuture.orange.com/en/

Future Emerging Technologies

[C10-1] ETSI Technology Radar online https://www.etsi.org/technologies/technology-radar

[C10-2] Juan Yin et al., “Satellite-based entanglement distribution over 1200 kilometers”. Science356,11401144(2017). DOI:10.1126/science.aan3211

[C10-3] Jessica Illiano, Marcello Caleffi, Antonio Manzalini, and Angela Sara Cacciapuoti. “Quantum Internet protocol stack: A comprehensive survey”. Comput. Netw. 213, C (Aug 2022). https://doi.org/10.1016/j.comnet.2022.109092

[C10-4] Kozlowski, Wojciech and Stephanie Wehner. “Towards Large-Scale Quantum Networks.” Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication (2019).

[C10-5] C. Wang, A. Rahman, R. Li, M. Aelmans, K. Chakraborty, “Application Scenarios for the Quantum Internet”, IRTF RFC 9583, June 2024.

[C10-6] Wei, Shi-Hai *et al.* Towards Real-World Quantum Networks: A Review. Laser & Photonics Reviews 16 (2022).

[C10-7] See EuroQCI: https://digital-strategy.ec.europa.eu/en/policies/european-quantum-communication-infrastructure-euroqci.

[C10-8] W. Kozlowski, S. Wehner, R. Van Meter, B. Rijsman, A. S. Cacciapuoti, M. Caleffi, S. Nagayama. Architectural Principles for a Quantum Internet. RFC 9340, QIRG @IRTF, https://datatracker.ietf.org/doc/rfc9340/.

[C10-9] C. Wang, A. Rahman, R. Li, M. Aelmans, K. Chakraborty. Application Scenarios for the Quantum Internet. RFC 9583, QIRG @IRTF, https://datatracker.ietf.org/doc/rfc9583/.

[C10-10] S. L. N. Hermans, M. Pompili, H. K. C. Beukers, S. Baier, J. Borregaard and R. Hanson. Qubit teleportation between non-neighbouring nodes in a quantum network. Nature volume 605, pp. 663–668 (2022),

[C10-11] Yong Yu, Fei Ma, Xi-Yu Luo, Bo Jing, Peng-Fei Sun, Ren-Zhou, Fang, Chao-Wei Yang, Hui Liu, Ming-Yang Zheng, Xiu-Ping Xie, Wei-Jun, Zhang, Li-Xing You, Zhen Wang, Teng-Yun Chen, Qiang Zhang, Xiao-Hui Bao, and Jian-Wei Pan. Entanglement of two quantum memories via fibres over dozens of kilometres. Nature volume 578, pp. 240–245 (2020).

[C10-12] O. Ezratty. Understanding Quantum Technologies. https://www.oezratty.net/wordpress/2023/understanding-quantum-technologies-2023/.

[C10-13] Biamonte, J., Wittek, P., Pancotti, N. *et al.* Quantum machine learning. *Nature* 549, 195–202 (2017). https://doi.org/10.1038/nature23474

[C10-14] V. Dunjko and H. J. Briegel. Machine learning & Artificial Intelligence in the quantum domain: a review of recent progress. Reports on Progress in Physics, vol. 81, no. 7, p. 074001, jun 2018, doi: 10.1088/1361-6633/AAB406

[C10-15] F. Phillipson. Quantum Machine Learning: Benefits and Practical Examples. QANSWER, 2020, pp. 51–56.

[C10-16] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean. Power of data in quantum machine learning. Nature Communications 2021 12:1, vol. 12, no. 1, pp. 1–9, may 2021, doi: 10.1038/s41467-021-22539-9

[C10-17] H. I. G. Hernández, R. T. Ruiz, and G.-H. Sun. Image Classification via Quantum Machine Learning. arXiv:2011.02831, 2020

[C10-18] P. Rebentrost, M. Mohseni, and S. Lloyd. Quantum Support Vector Machine for Big Data Classification. Physical Review Letters, vol. 113, no. 13, Sep 2014, doi: 10.1103/physrevlett.113.130503

[C10-19] M. Henderson, S. Shakya, S. Pradhan, and T. Cook. Quanvolutional Neural Networks: Powering Image Recognition with Quantum Circuits. arXiv:1904.04767, 2019

[C10-20] A. I. Hasan. IGO-QNN: Quantum Neural Network Architecture for Inductive Grover Oracularization. arXiv:2105.11603, 2021

[C10-21] N. Liu, T. Huang, J. Gao, Z. Xu, D. Wang, and F. Li. Quantum-Enhanced Deep Learning-Based Lithology Interpretation from Well Logs. IEEE Transactions on Geoscience and Remote Sensing, 2021

[C10-22] A. Khoshaman, W. Vinci, B. Denis, E. Andriyash, H. Sadeghi, and M. H. Amin, “Quantum variational autoencoder,” Quantum Science and Technology, vol. 4, no. 1, p. 014001, sep 2018, doi: 10.1088/2058-9565/AADA1F

[C10-23] J. Liu, K. H. Lim, K. L. Wood, W. Huang, C. Guo, and H.-L. Huang. Hybrid quantum-classical convolutional neural networks. Science China Physics, Mechanics and Astronomy 2021 64:9, vol. 64, no. 9, pp. 1–8, aug 2021, doi: 10.1007/S11433-021-1734-3

[C10-24] S. Oh, J. Choi, and J. Kim. A Tutorial on Quantum Convolutional Neural Networks (QCNN). arXiv:2009.09423, vol. 2020-October, pp. 236–239, sep 2020

[C10-25] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik. Noisy intermediate-scale quantum (NISQ) algorithms. arXiv:2101.08448, 2021

[C10-26] I. Cong, S. Choi et al. Quantum convolutional neural networks. Nature Physics, vol. 15, no. 12, p. 1273–1278, 2019, doi: 10.1038/s41567-019-0648-8.

[C10-27] V. Ziegler, P. Schneider, H. Viswanathan, M. Montag, S. Kanugovi and A. Rezaki. Security and Trust in the 6G Era. IEEE Access, vol. 9, pp. 142314-142327, 2021, doi: 10.1109/ACCESS.2021.3120143

[C10-28] Zero-knowledge proof - Wikipedia. https://en.wikipedia.org/wiki/Zero-knowledge_proof

[C10-29] Zero-Knowledge Proof (ZKP) — Explained | Chainlink. https://chain.link/education/zero-knowledge-proof-zkp

[C10-30] “Anonymous” Data Won’t Protect Your Identity, 2019, online https://www.scientificamerican.com/article/anonymous-data-wont-protect-your-identity/

[C10-31] https://mostly.ai/what-is-data-anonymization

[C10-32] C. Thuemmler, J. Mueller, S. Covaci, T. Magedanz, S. de Panfilis, T. Jell, A. Schnieder, A. Gavras. Applying the Software-to-Data Paradigm in Next Generation E-Health Hybrid Clouds. 2013 10th International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 2013, pp. 459-463, doi: 10.1109/ITNG.2013.77.

[C10-33] E. Hubinger *et al.* Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training. arXiv e-prints. Jan 2024. arXiv:2401.05566. doi: 10.48550/arXiv.2401.05566

[C10-34] China Performs First 5G Remote Surgery, 2019, https://www.pcmag.com/news/china-performs-first-5g-remote-surgery

[C10-35] Glasgow University demonstrates 5G robotic arm teleoperation, 2021, https://www.landmobile.co.uk/news/glasgow-university-demonstrates-5g-robotic-arm-teleoperation/

[C10-36] ‘Call Of Duty: Black Ops Cold War’ lets PC players enable trigger haptics, https://www.nme.com/news/gaming-news/call-of-duty-black-ops-cold-war-lets-pc-players-enable-triggerhaptics-2995925

[C10-37] Pan-tilt Ultrasound Mid-Air Haptics, 2020, https://inria.hal.science/hal-02479022/file/Howard-PUMAH-demo.pdf and https://hal.inria.fr/hal-02479022/file/Howard-PUMAH-demo.mp4

[C10-38] Surface Haptics: A Safer Way for Drivers to Operate Smooth-Surface Controls, 2020, https://www.electronicdesign.com/markets/automotive/article/21145025/surface-haptics-a-safer-way-fordrivers-to-operate-smoothsurface-controls

[C10-39] 'Haptic feedback' virtual reality Teslasuit can simulate everything from a bullet to a hug, 2021, https://www.abc.net.au/news/science/2021-04-01/vr-teslasuit-simulates-virtual-reality-touch-hapticfeedback/100030320

[C10-40] A new kind of haptic wearable: GPS for the blind, 2021, Flavio Meneses, Daniel Corujo, Rui Aguiar. Virtualization of Customer Equipment: Challenges and Opportunities. ITL internet technology letters Vol. 3, Nº 6, May, 2020f https://www.freethink.com/series/ramen-profitable/gps-for-the-blind

[C10-41] This Haptic suit lets you ‘hear’ music through your skin, 2020, https://www.auganix.org/this-haptic-suit-lets-you-hear-music-through-your-skin/

[C10-42] Researchers develop hands-free, eyes-free navigation for Soldiers, 2014, https://www.army.mil/article/125405/researchers_develop_hands_free_eyes_free_navigation_for_soldiers

[C10-43] Jon Karafin, Light Field Lab Inc. online https://mpeg.chiariglione.org/sites/default/files/events/08_KARAFIN_LightFieldLab_MPEGWorkshopLB_v01.pdf

[C10-44] Rahim Tafazolli, online https://www.itu.int/en/ITU-T/Workshops-and-Seminars/20190218/Documents/Rahim_Tafazolli_Presentation.pdf

[C10-45] https://en.wikipedia.org/wiki/Retinal_implant

[C10-46] The term originates in the 1992 science fiction novel “Snow Crash” by Neal Stephenson

[C10-47] Flavio Meneses, Daniel Corujo, Rui Aguiar. Virtualization of Customer Equipment: Challenges and Opportunities. ITL internet technology letters Vol. 3, Nº 6, May, 2020f

[C10-48] Hosseini, Ensieh S., Dervin, Saoirse, Ganguly, Priyanka, Dahiya, Ravinder. Biodegradable Materials for Sustainable Health Monitoring Devices. ACS Applied Bio Materials 2021 4 (1), 163-194. doi: 10.1021/acsabm.0c01139

[C10-49] Litvinova O, Klager E, Tzvetkov NT, Kimberger O, Kletecka-Pulker M, Willschke H, Atanasov AG. Digital Pills with Ingestible Sensors: Patent Landscape Analysis. Pharmaceuticals (Basel). 2022 Aug 19;15(8):1025. doi: 10.3390/ph15081025. PMID: 36015173; PMCID: PMC9415622.

[C10-50] Yang SY, Sencadas V, You SS, Jia NZ, Srinivasan SS, Huang HW, Ahmed AE, Liang JY, Traverso G. Powering Implantable and Ingestible Electronics. Adv Funct Mater. 2021 Oct 26;31(44):2009289. doi: 10.1002/adfm.202009289. Epub 2021 Feb 4. PMID: 34720792; PMCID: PMC8553224.

[C10-51] Hsiang-Yu Wang, Angela Bernarda, Chih-Yung Huang, Duu-Jong Lee, Jo-Shu Chang. Micro-sized microbial fuel cell: A mini-review. Bioresource Technology, Volume 102, Issue 1, 2011, Pages 235-243, ISSN 0960-8524, https://doi.org/10.1016/j.biortech.2010.07.007.

[C10-52] https://www.digitaltwinconsortium.org/

[C10-53] https://5g-acia.org/whitepapers/using-digital-twins-to-integrate-5g-into-productionnetworks/

[C10-54] Richard Feynmann, “There is plenty of room at the bottom”, Pasadena, Dec. 1959, https://web.archive.org/web/20170105015142/http://www.its.caltech.edu/~feynman/plenty.html

[C10-55] İbrahim Yazici, Ibraheem Shayea, Jafri Din. A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. Engineering Science and Technology, an International Journal, Volume 44, 2023, 101455, ISSN 2215-0986, https://doi.org/10.1016/j.jestch.2023.101455.

[C10-56] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[C10-57] Ericsson blog - Zero-energy devices – a new opportunity in 6G, retrieved June 2022, https://www.ericsson.com/en/blog/2021/9/zero-energy-devices-opportunity-6g

[C10-58] Brian Rogers, et. Al. “Scaling the Bandwidth Wall: Challenges in and Avenues for Chip Multi-Processor Scaling,” ISCA’09, June 20–24, 2009

[C10-59] G. Geraci, et al., “What will the future of UAV cellular communications be? A flight from 5G to 6G,” IEEE Communications Surveys and Tutorials, 2022.

[C10-60] S. Karimi-Bidhendi et al., “Optimizing cellular networks for UAV corridors via quantization theory,” IEEE Transactions on Wireless Communications, 2024.

[C10-61] Y. Zeng, et al. “UAV communications for 5G and beyond”, Wiley – IEEE Press, Dec. 2020.

[C10-62] M. Benzaghta et al., “Designing cellular networks for UAV corridors via Bayesian optimization,” in Proc. IEEE Globecom, 2023.

[C10-63] G. Geraci, et al., “Integrating terrestrial and non-terrestrial networks: 3D opportunities and challenges,” IEEE Communications Magazine, 2023.

[C10-64] The Tor -project, https://www.torproject.org/about/history/ [online: accessed on 10.10.2024]

[C10-65] Dingledine, Roger (20 September 2002), pre-alpha: run an onion proxy now! (seul.org), freehaven.net (mailing list) announce [online: accessed 10.10.2024].

[C10-66] Stochastic geometry and architecture of communication network, F. Baccelli, M. Klein, M. Lebourges and S. Zuyev, Telecommunications Systems, 7, pp.~209--227, 1997

[C10-67] Modeling and Analysis of Stochastic Mobile-Edge Computing Wireless Networks, Yixiao Gu; Yao Yao; Cheng Li; Bin Xia; Dingjie Xu; Chaoxian Zhang, 2021 https://ieeexplore.ieee.org/abstract/document/9385418

[C10-68] Wireless Networks for Mobile Edge Computing: Spatial Modeling and Latency Analysis (Extended version) Seung-Woo Ko, Kaifeng Han, Kaibin Huang, https://ieeexplore.ieee.org/document/8370831

[C10-69] https://www.scientificamerican.com/article/explorers-of-quantum-entanglement-win-2022-nobel-prize-in-physics1/

[C10-70] Leonard Susskind,, Ying Zhao. Teleportation through the wormhole. Phys.Rev.D 98 (2018) 4, 046016. doi: 10.1103/PhysRevD.98.046016, arXiv:1707.04354

[C10-71] Ping Gao, Daniel Louis Jafferis. A traversable wormhole teleportation protocol in the SYK model. Journal of High Energy Physics 07 (2021) 097. DOI: 10.1007/JHEP07(2021)097. arXiv:1911.07416

[C10-72] Jafferis, D., Zlokapa, A., Lykken, J.D. et al. Traversable wormhole dynamics on a quantum processor. Nature 612, 51–55 (2022). https://doi.org/10.1038/s41586-022-05424-3

Editors

Rui Aguiar, Instituto de Telecomunicações
Ari Pouttu, University of Oulo
Jyrki Huusko, VTT

Artur Hecker, Huawei (chap.2)
Luis Miguel Contreras Murillo, Telefonica (chap. 3)
Emmanuel Dotaro, Thales (chap. 4)
Josef Urban, Nokia, (chap. 5)
Wen Xu, Huawei (chap. 6)
Raul Muñoz, CTTC (chap. 7)
Tomaso de Cola, DLR (chap. 8)
Andre Bourdoux, imec (chap. 9)
Anastasius Gavras, Eurescom.eu (chap. 10)

List of contributors

Alessandro Bassi, Eurescom
Alessandro Guidotti, University of Bologna
Alessandro Vanelli-Coralli, University of Bologna
Ana Aguiar, University of Porto
Ana G. Armada, Universidad Carlos III de Madrid
Andre Bourdoux, IMEC
Andrea Giorgetti, University of Bologna
Andreas Knopp, Universität Bundeswehr Munich
Andrés Meseguer, ITI
Andrew Lord, BT
Antonio Napoli, Infinera
Antonio Skarmeta, Univesidad de Murcia
Arman Shojaeifard, InterDigital, UK
Augusto Casaca, INOV
Barry Evans, University of Surrey
Beatriz Soret, Uni Malaga/Aalborg
Bengt Holter, SINTEF
Bengt Sahlin, Ericsson
Benjamin Wohlfeil, Adtran,
Carla Amatetti, University of Bologna
Carsten Griwodz, University of Oslo
Chiara Lombardo, University of Genua
Christian Rohde, Fraunhofer IIS
Claudio Cicconetti, IIT-CNR
Colja Schubert, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute.
Constantinos Papadias, American College of Greece
Cristian Vaca, CTTC
Damir Filipovic, AioTI
Daniel Corujo, IT Aveiro
Daniel Kofman, IMT Telecom Paris
Daniele Tarchi, University of Florence
Didier Arnaud Bourse, Nokia
Diego Lopez, Telefonica
Dirk Trossen, Huawei
Enrico Del RE, Firenze University
Elisa Rojas, Universidad de Alcalá
Ernesto Ciaramella, Scuola Superiore Sant'Anna
Esther Payne, the Librecast project
Eva Fogelstrom, Ericsson
Eva Lagunas, Uni Luxembourg/SnT
Fabio Cavaliere, Ericsson
Fabio Martinelli, CNR
Felix Ott, Fraunhofer IIS
Francisco Muro, Uni Malaga
Franco Davoli, CNIT
Francois Baccelli, INRIA, ENS
Frank Schaich, Nokia
George Carle, TUM
George Parisis, Sussex University
George Polyzos, AUEB
George Xylomenos, AUEB
Georgios Karagiannis, Huawei
Gerhard Fettweis, TU Dresden
Gil Kedar, Ceragon
Giovanni Geraci, Telefonica
Giuseppe Caire, TU Berlin
Harald Haas, University of Cambridge
Helmut Griesser, Adva Network Security
Henk Wymeersch, Chalmers University of Technology
Hervé Debar, IMT Telecom Sud Paris
Hugo Tullberg, Ericsson
Hui Song, Sintef
Husnain Shahid, CTTC
Ian F. Akyldiz, Georgia Institute of Technology
Izad Ahmad, VTT
Ioannis Tomkos, University of Patras
J. Joaquín Escudero Garzás, Gradiant
Jens Finkhaeuser, Interpeer
Joan A. Ruiz-de-Azua, i2CAT
Joan Bas, CTTC
Joerg Ott, TUM
Joerg Widmer, IMDEA Networks
Johannes Fisher, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute.
Jorge Querol, Uni Luxembourg/SnT
Jörg-Peter Elbers, ADTRAN
Jose F. Monserrat, UPV
Josef Urban, Nokia Bell-Labs
Josep M. Jornet, Northeastern University
Juraj Poliak, DLR
Lorenzo Favalli, University of Pavia
Ludovic Noirie, Nokia Bell-Labs
Luis Blanco, CTTC

Luis Miguel Contreras Murillo, Telefonica
Luis Perez-Freire, Gradiant, AIOTI
Luis Velasco, UPC
Madhusanka Liyanage, University College Dublin
Marco Ruffini, Trinity College Dublin (TCD)
Marcus Brunner, Huawei
Marius Caus, CTTC
Markus Höyhtyä, VTT
Martin Reed, Essex University
Martin Schell, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute
Michael Montag, Nokia
Michael Welzl, University of Oslo
Miguel Ángel Vázquez, CTTC
Miguel Rio, UCL
Mika Rantakokko, VTT
Muhammad Shuaib Siddiqu, I2CAT
Musbah Shaat, CTTC
Nandana Rajatheva, University of Oulu
Nikolaos Bartzoudis, CTTC
Paolo Di Francesco, Level42
Paulo Mendes, Airbus Defence & Space
Petros S. Bithas, National and Kapodistrian University of Athens
Philippe Chanclou, Orange
Philippe Martins, INRIA
Pietro Savazzi, University of Pavia
Pouria Khodashenas, Huawei
Qi Liao, Nokia Bell-Labs
Raffaele Bolla, University of Genua
Rainer Wansch, Fraunhofer IIS
Ramin Khalili, Huawei
Ramon Casellas, CTTC
Raymond Knopp, EURECOM
Rene Bonk, Nokia Bell Labs Germany
Roberto Bruschi, University of Genua
Roberto Cascella, ECSO
Roland Bless, KIT
Rui Dinis, Instituto de Telecomunicações (IT) / Nova University of Lisbon
Rute Sofia, Fortiss
Samuel Marchal, VTT
Satyendra Kumar Mishra, CTTC
Sébastien Bigo, Nokia Bell Labs
Sergio Cozzolino, Telecom Italia
Sergio Fortes, Uni Malaga
Silvia Liberata Ullo, Invesity of Sannio
Steve Taylor, IT Innovation Centre
Thang Xuan Vu, University of Luxemburg
Thomas Delamotte, Universität Bundeswehr Munich
Tim Hentschel, Barkhausen Institut
Tommy Svensson, Chalmers University of Technology
Vittorio Curri, Politecnico di Torino (Polito)
Xavier Artiga, CTTC
Yvan Pointurier, Huawei
Yvette Koza, ZTE

NETWORLD EUROPE

https://www.networldeurope.eu/

Cite as: Networld Europe, "Technical Annex to Strategic Research and Innovation Agenda 2022-27", Eds. Rui Aguiar, Ari Pouttu, Jyrki Huusko, Artur Hecker, Luis Miguel Contreras Murillo, Emmanuel Dotaro, Josef Urban, Wen Xu, Raul Muñoz, Tomaso de Cola, Andre Bourdoux, Anastasius Gavras, 2024, DOI: https://doi.org/10.5281/zenodo.15165400