Physics-based wireless AI providing scalability and efficiency

Grant Agreement No. : CHIST-ERA-22-WAI-04
Call: CHIST-ERA, Machine Learning-based Communication Systems, towards Wireless AI (WAI)

Deliverable D2.1
Initial report on network architecture, interfaces, and architecture evaluation

Version:	1.0
Due date:	31.08.2024
Delivered date:	27.11.2024
Dissemination level:	PU

Authors

Konstantinos Ntontin (UL, deliverable leader), Thang X. Vu (UL), Nhan Nguyen (UOULU), Markku Juntti (UOULU), Lianet Méndez-Monsanto Suárez (UC3M), Ana García Armada (UC3M), Iñigo Bilbao (EHU), Eneko Iradier (EHU), Marco Di Renzo (CS-UPS), Hanna Bogucka (PUT)

Coordinator

Prof. Ana Garcia Armada

Univ Carlos III de Madrid, Spain

Avda. de la Universidad 30 (28911 Leganés)

Spain

Email: anagar@ing.uc3m.es

Disclaimer

Acknowledgement

This report is funded under the CHIST-ERA grant CHIST-ERA-22-WAI-04.

Version history

Version	Date	Remarks
0.1	10.11.2024	
0.2	15.11.2024	
0.3	20.11.2024	
1.0	27.11.2024	

Executive summary

This deliverable presents the overall architecture of PASSIONATE and the areas of under investigation of the project. Its main chapters are Chapters 1 and 2. In particular, Chapter 1 at first presents that background regarding the use of artificial intelligence in current telecommunications networks and proceeds with motivation behind the physics-based approach followed in the project. Chapter 2 introduces the 3D architecture considered in the project, detailing the nodes that comprise the individual layers, particularly the terrestrial, aerial, and space layers. Finally, the chapter is concluded with an overview of the different areas of interest in the 3D network that will be investigated in the project.

D2.1 reports the initial version of the architecture and gives input to D2.2: Use cases and KPIs and the deliverables of WPs 3-5.

Table of contents

List of abbreviations

AI	Artificial Intelligence
CAPEX	Capital Expenditure
CE-OFDM	Constant Envelope Orthogonal Frequency Division Multiplexing
CN	Core Network
CNN	Convolutional Neural Network
CSI	Channel State Information
DL	Deep Learning
DNN	Deep Neural Network
E2E	End-to-End
FD	Full-Duplex
FM-OFDM	Frequency-Modulated Orthogonal Frequency Division Multiplexing
FPGA	Field Programmable Gate Array
GEO	Geostationary Orbit
GNN	Graph Neural Network
HAPS	High-Altitude Platforms
HBF	Hybrid Analog/Digital Beamformer
HEO	Highly Elliptical Orbit
IoS	Internet of Surfaces
IoT	Internet of Things
LEO	Low-Earth Orbit
LOS	Line-of-Sight
MAC	Medium Access Control
MEO	Medium-Earth Orbit
MIMO	Multiple-Input-Multiple-Output
mMIMO	Massive MIMO
ML	Machine Learning
MmWave	Millimeter Wave
NGSO	Non-Geostationary Orbit
NTN	Non-Terrestrial Network
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	Operational Expenditure
OTFS	Orthogonal Time Frequency Space
PHY	Physical Layer

QoE	Quality of Experience
QoS	Quality of Service
RIS	Reconfigurable Intelligent Surface
RRM	Radio Resource Management
SDR	Software-Defined Radio
SI	Self-Interference
SON	Self-Organizing Network
TN	Terrestrial Network
UAV	Unmanned Aerial Vehicle
UE	User Equipment

List of figures

1 Introduction

1.1 Background

6G is at the gates with the first commercial deployments expect around 2030. 6G is not just considered to be a gradual evolution of 5G, but it will be designed to combine both communication and computing into a hyperconnected world of digital and physical experiences [1]. Several new use cases are foreseen for this new generation of networks, such as: i) enhanced human communication that includes extended reality immersive holographic communications, multimodal communication for teleoperation, and intelligent interaction and sharing of sensation, skills, and thoughts; ii) enhanced machine communication that includes robot network fabric and interacting cobots; iii) enabling services, such as 3D hyper accurate positioning, localization, and tracking, interactive mapping, digital healthcare, automatic detection, recognition, and inspection, smart industry, and trusted composition of services; iv) digital twins; v) autonomous and connected transportation; vi) fixed wireless access; zero-energy Internet of Things (IoT) devices empowered by energy harvesting; and vii) critical communications [2], [3]. Hence, we understand that there is a multitude of new services in 6G networks that need to be facilitated. Those depart from the primary target of achieving high rates that governed the telecommunications generations up until now, with more targets of interest, such as sensing related capabilities, sustainability, and interoperability.

Apart from the multitude of services that would need to be facilitated by the forthcoming generation, the networks 6G and beyond networks are foreseen to heavily increase in size. In particular, while 95% of the Earth's population has cellular coverage, less than 45% of the landmass of the Earth [4] and only 15% of its surface has such coverage [5]. This means that almost 400 million people, a quite large number which is the rest 5%, do not have access to any mobile network. These are people that primarily rural, poor, and sparsely populated areas. Covering those areas poses a great challenge due to the unviability in terms of CAPEX/OPEX and energy consumption, from the operators' point of view, of heavily densifying with base stations an area with a low penetration of users. Instead, what has been brought forward as a radical solution is to integrate the terrestrial networks with non-terrestrial ones and create a 3D network of networks [6].

The non-terrestrial networks encompass flying nodes either aerial one, such as unmanned air vehicles, planes, and high-altitude platforms (HAPS) or space ones, such as low-Earth orbit (LEO), medium-Earth orbit (MEO), and geostationary orbit (GEO) satellites. Due to their much larger footprint on Earth with respect to their terrestrial counterparts, not terrestrial platforms can provide the same coverage that terrestrial networks can provide with much smaller number of nodes. In particular, an indicative study has revealed that in order to provide full 5G coverage in the UK, only 60 HAPS would be requited. In contrast thousands of additional terrestrial small cells would be needed for a full 5G coverage. That difference is substantial and so we understand that in such a cases providing the necessary coverage by non-terrestrial networks (NTNs) is the only viable solution. The same holds with the space segment. In particular, in the next years to come thousands of LEO satellites are expected to fly above Earth, creating mega-constellations. Private ventures have already been realized with aims towards mega-constellations such as the one of Starlink [7].

Based on the above, we understand that 6G and beyond networks are going to be highly complex, both in terms of services that they would need to accommodate, but also due to the inclusion of both the aerial and space layer. Hence, a large network dynamicity is foreseen. More specifically, a list of challenges future networks will face is the following [8]:

- Very high complexity, related to the integration the terrestrial, aerial, and space layers, heterogeneity, agility, and 3D mobility of users and base stations. An indicative

example of how dynamic the network can be is the case of LEO mega-constellations, where thousands of LEO satellites move with velocity 7.8-8 km/s.
- The requirement to support a very large amount of traffic to/from trillions of user equipments (UEs) that include IoT devices.
- The need to enhance users' quality of experience (QoE), enabled by Tbps speeds and reduced latency.
- The requirement to intelligent virtualize and dynamically manage the resources.
- The integration of the capabilities of the user devices in network communication or computation aspects.
- The need to enable computational and caching services at different levels of the network, such as cloud and fog/edge, and user device clusters.

Based on the challenges above, it is clear that the allocation of resources in such a network cannot be performed by manual means. Decisions need to be taken fast that involve a multitude of parameters. Hence, in such an ecosystem network automation through artificial intelligence (AI) seems the only viable approach. In communication networks, automation first appeared for 3GPP Release 8 with the notion of self-organizing networks (SONs) [9]. The autonomy of SONs is achieved by using different algorithms and AI. The following 3 distinct categories define SONs [10]:
- **Self configuration**: It involves automatic recognition and registration of new base stations, adjusting technical parameters to avoid interference, towards the maximization of coverage and capacity.
- **Self optimization**: It focuses on optimizing base station parameters for specific purposes, such as preserving service level agreements during congestion or changing spectrum availability.
- **Self healing**: It enables the network to recover from failures, minimizing service degradation for affected users.

In terms of different optimization areas SONs can operate at, the following are the main areas:
- **Dynamic optimization of spectrum usage**: SONs optimize the use of the radio spectrum, in a dynamic fashion, with the aim of ensuring efficient utilization. The optimization involves adjusting the frequencies used by different network elements to minimize interference and increase coverage and capacity. By allocating spectrum resources in an intelligent wat, SONs ensure that the network can handle varying loads and conditions efficiently.
- **Optimization of control plane resources**: SONs optimize the control plane resources by dynamically adjusting parameters and algorithms with the purpose of improving the network performance. This optimization reduces latency, improves that network performance, and ensures that the control plane can efficiently manage data flows.
- **Avoiding congestion**: Through predictive analytics that analyze real-time network conditions, SONs employ mechanisms to avoid congestion before it occurs. This proactive approach to congestion ensures that the responsiveness and reliability of the network, even under heavy load.
- **Optimal Resource Utilization with Maximum User Experience**: SONs utilize network resources optimally to provide the best possible user experience. This involves balancing the distribution of network resources among users. In this way, they ensure fair access and maximize the quality of service (QoS) for all users. So the network can support many users at once.

- **Self-Healing Mechanisms**: This corresponds to the mechanisms enacted for recovering from failures and degradations in service. In particular, when a network element fails or a service issue appears, other nodes in the network can temporarily operate in the impact area, minimizing service disruption. After the issue is resolved, the network returns to its normal state.

The SONs are further categorized in centralized, distributed, and hybrid. In centralized SONs a central entity, such as a server, gathers all the necessary data from the whole network, and decides optimally for the whole network. The advantage of such an approach is that it leads to globally optimal decisions. However, a typical drawback is the latency to reach this due to the need to gather an immense amount of data to a central entity. For highly dynamic networks, such as the integrated terrestrial-non terrestrial networks, it is very likely that by the time the necessary amount of data is gathered to the central entity, the network and the traffic conditions have substantially changed. This would render any AI-based decision outdated. Furthermore, having single only entity gathering all the information is prone to single-point of failure
A solution to the above issue is given by the distributed SON approach, where the network intelligence and automation are embedded within the individual network elements, such as base stations and routers. The straightforward advantage of this approach is the fast response times since the decisions are taken locally and its robustness to potentially outdated data. Moreover, it avoids the single-point of failure of the centralized approach. However, a drawback of the method is how to effectively perform coordination among many distributed entities and reach a sub-optimal solution that is not far from the globally optimum point. Finally, the hybrid SON approach balances the advantages of the centralized and distributed approaches.

1.2 Motivation

Based on the above, we understand that via AI SONs offer numerous benefits to communications networks, such as reduction in manual operation of the networks, improved performance, reduction of network downtime, enhanced QoS, cost savings, and proactive network management. Such benefits are expected to be even more pronounced in 6G and beyond networks due to their much higher complexity, dynamicity, and the increased number of services they would need to facilitate. In fact, it is foreseen that 6G communication networks should be the first generation of networks with native AI, so that AI will not merely be an application but an inherent part of the infrastructure, network management, and operations. Yet, a purely data-driven approach has major limitations due to its resource constraints, high complexity, and black-box nature. However, to perform the above in the optimal way requires deep knowledge in network dynamics and machine learning (ML), so to deploy the best approach based on the particular network configuration.
Traditionally, popular ML approaches such as deep neural networks (DNNs) have successfully been used in the areas of computer vison and natural language processing, where accurate statistical models are typically scarce [11]. This is the reason the sue of DNNs by several groups for the resource allocation optimization wireless networks. However, in very complex and dynamic networks, such as the 3D ones, massive data sets would be required to train such DNNs for learning a desirable mapping. In addition, even the case of using DNNs that are pre-trained can result to substantial computational burdens due to their large parameterization [11]. Hence, this can be a bottleneck for low capability devices, such as mobile phones and IoT devices. Finally, such an extreme parameterization and abstractness of DNNs make it totally black regarding how the decisions are taken and any potential of having analytical guarantees. As a result they offer poor reliability and explainability.

In light of this, PASSIONATE will unlock ML for wireless by customizing and accounting-by-design the unique properties (“physics-based”) of the networks they are applied to. Physics-based ML is, in addition, the suitable approach to ensure the scalability, generalization, reliability, and user trust of ML, enabling ML solutions that are technically robust and possibly explainable-by-design. In PASSIONATE, we will develop the understanding and vision of what the application of AI/ML to the wireless network can provide and design use cases that can take advantage of this technology. For these use cases and with the new physics-based AI/ML tools, we will design new PHY, MAC, and RRM techniques and algorithms that achieve the ambitious goals of future mobile networks regarding coverage, data rate, latency, and energy consumption. We will evaluate experimentally by realistic simulations and measurements the achieved gains and contribute to creating data sets that can be used for the community. By advancing the state of the art and stimulating research and technology-based innovation through dissemination, PASSIONATE will create awareness and facilitate the positive impact of advanced wireless communications on society and the economy.

2 Passionate Network Architecture and Areas of Study

2.1 Passionate Network Architecture

The considered 3D network architecture of PASSIONATE is depicted in Figure 1. The network amalgamates terrestrial, aerial, and space layers. They are described in the following:

Terrestrial layer: It involves a multitude of heterogeneous radio, wireless, cellular, ground satellite, and small cells operating at different frequency bands, primarily sub-6 GHz and mmWave. Until 5G, the antennas of these BS nodes have been downwards tilted to serve the ground users. However, PASSIONATE follows the changes considered in the 6G network concept by either equipping them with a large number of antennas to enable 3D beamforming or with uptilted antennas. This would allow, in turn, to better interconnect (integrate) terrestrial BSs or users with other aerial and space users, including flying BSs installed on-board satellite spaceship, aircraft, HAPS or UAVs. The availability of such access networks in the different strata provides a resilient communication network to aircraft, which supports their critical-time operations and safety. It also enables the flying objects (e.g., aircraft, UAVs) to select the air-to-ground links according to their communication performance requirements. Obviously, network scenarios such as traffic or user load balancing or traffic re-routing can be realized to better serve aerial and space users. The TN infrastructure also includes different gateways that connect the aerial and space platforms to the terrestrial core network through feeder links. Finally, we assume that numerous reconfigurable intelligent surfaces (RISs) are deployed in the terrestrial segment, mounted on the facades of buildings for instances. RISs can provide alternative routes for communication through single-hop reflections, especially for communication in mmWave bands that are more susceptible to blockages than their sub-6 GHz counterparts.

Aerial layer: It comprises platforms such as commercial airplanes (aircraft), UAVs, and HAPSs that fly at altitudes of 8-11 km, up to 1 km, and 17-20 km, respectively. In 6G networks, the aerial platforms are envisioned to be boarded with 5G gNBs. These turn them into flying BSs that can serve users within the same network stratum, as well as other users within the strata above and below theirs. The advantage of operating gNBs on-board HAPS stems from their much lower altitude relative to the ground than the satellites installed in the different orbits. Also, the quasi-stationary nature of HAPSs eliminates the need for frequent handovers, like the case with the non-Geostationary Orbits (NGSOs) when they are about to lose visibility with ground terminals. Furthermore, their altitude and size allow them to be equipped with sizeable antennas that can offer high gains. This could enable direct handheld device access even at mmWaves (e.g., Ka band). Moreover, as their service coverage gets wider, relying on their much larger footprint compared to the terrestrial gNBs, it would better prevent the frequent handovers of high-mobility users such as aircraft or UAVs, in contrast with fast-moving ground network mobility, such as trains, where terrestrial cells may fail in providing ubiquitous services. However, what differentiates this stratum more is that the gNBs on-board the platforms deployed within this stratum can act as a relay node enabling the expansion of the visibility of NGSOs to better serve ground users and those in the other network strata. The UAVs have been used for different purposes, including military, cargo, and rescue operations. They can also act as aerial gNBs in the case of natural disasters and high-traffic demand events like sport events. Airplanes (aircraft) require resilient and reliable networks not only to support

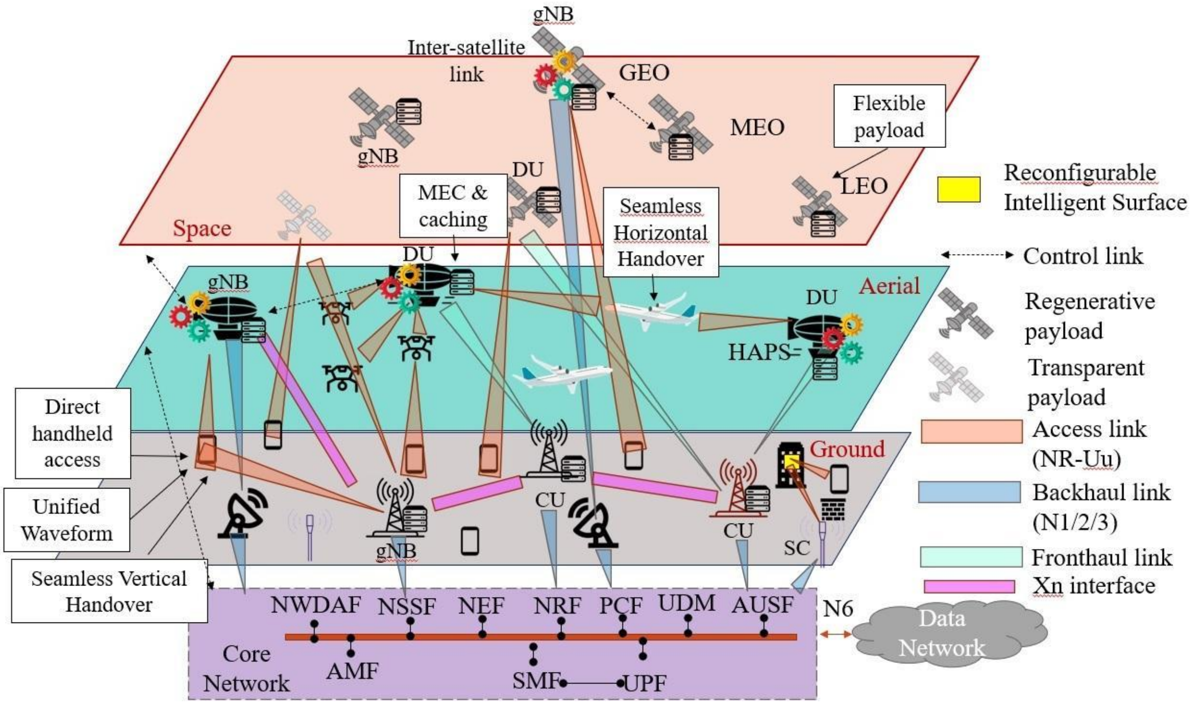

Figure 1: PASSIONATE's architecture.

critical operations or safety, but also to provide continuous broadband connections for passengers on board. This has been challenging and costly with TN networks and can only be expected during the take-off and landing times. However, when the airplanes spend the majority of their flight time at high altitudes, the PASSIONATE integrated 3D network can be deployed to support the required broadband coverage, which can be provided by means of satellites like LEO, NGSO, and HAPS.

Space layer: It is the upper stratum in the 6G PASSIONATE 3D network architecture. It comprises satellites operating at different altitudes in orbit around the Earth. Satellite communications have been around for sixty years; it has become integrated and is increasingly unified as part of the NTN within the 3GPP framework.

Satellite trade-offs

When designing the satellite to provide satellite communications services (satcom), there are, of course, many trade-offs that need to be considered. Factors such as the launch mass, the amount of power that can be generated and the excess heat that can be radiated, and the availability of electronics suitable for the increased levels of radiation above the atmosphere all need to be considered. A related trade-off is for the satellite to be either bent-pipe or regenerative. Bent-pipe satellites have been the predominant GEO satellites for many years as most of the available power budget is used to efficiently receive and retransmit the same radio signal; whereas regenerative satellites demodulate the signal, may process the data in some way and then retransmit it. This does improve the E2E link budget as there is no additive noise from the uplink and downlink transmissions however it also signifies that less power is available for transmission. Two benefits of regenerative satellites supporting NTN are that they facilitate ISLs and allow some aspects of the 5G CN to be instantiated in space thereby improving CP response times. Regenerative payloads are directly connected with the notion of flexible payloads. Flexible Payload is based on the principle of programming the logical resources of hardware boards (e.g., FPGAs) and applies software virtualisation mechanisms on the base OS. Thus, through Flexible Pay¬loads equipping NTN nodes, such as satellites and HAPS, resources such as power, band¬width, and beams, can be adjusted through software on the ground (SDR-based) in accordance with the application demands. The capabilities of each of the corresponding NTN nodes, such as size and generated power, will dictate the FPGA capabilities and the corresponding Flexible Payload functionalities that can be realized.

Characterisation by orbit

Satellites are usually characterized by orbit type as follows:

- GEO – geostationary orbit where the satellites orbit above the equator so that they appear stationary in the sky from the ground; only a few satellites are needed to provide near global coverage (only the polar regions are excluded), but the propagation delay is too high for some applications and the link budget tends to require higher gain antennas for the user equipment;
- LEO – low earth orbit satellites are much closer to the ground, which improves the link budget performance and reduces the propagation delays at the cost of the satellite moving rapidly across the sky when viewed from the ground and needing many satellites in the constellation to provide continuous coverage;
- MEO – medium earth orbit satellites can be found between LEO and GEO; they provide global coverage with far fewer satellites than LEO with an intermediate propagation delay and link budget performance;
- HEO – highly elliptical earth orbit satellites are a specialist application where a few satellites are launched into the same orbit that drops low (and fast at perigee) before raising to a much higher (and slower at apogee) part of the ellipse. If these orbits are also synchronized to the Earth's rotation, they can provide GEO-like performance in the polar regions when the satellites move relatively slowly across the sky.

PASSIONATE 3D edge computing and storage continuum is another capability that would majorly distinguish the designed 3D network in the sense it would become more responsive to service demands type and volume and match it with the requested infrastructure availability in an efficient manner relying on real-time predictive data analytics. Noting that a key difference between the 3GPP's Release 19 and earlier releases regarding the integration of TNs with NTNs is that Release 19 will include the regenerative payload feature for NTN platforms. This means that the nodes will be able to encode and decode the received information and act as aerial and space gNBs. This, in turn, will enable them to act as edge computing and storage units, which can heavily alleviate the large amount of data that the cloud TNs need to process and store. Hence, the PASSIONATE 3D network is a TN-NTN distributed edge computing and storage network that allows such tasks to be performed in terrestrial, aerial, and space nodes. This can be possible by the existence of inter-space-air-terrestrial links for fast data routing among nodes. The federation of the multi-edge computing deployed in the different 3D network strata will be of paramount importance to support the full single pilot operations. It

will also allow the running of machine learning and artificial intelligence algorithms, training, and testing on the fly to better ensure aircraft operations and navigation.
The above infrastructure contains diverse resources, including radio channels, wireless, cellular and satellite air interface and spectrum, 5G core network functions, as well as added memory, storage, processing, and communication payload to HAPSs and satellites, in addition to the content cache and edge computing distributed in the different network strata. These diverse resources provide opportunities for creating and deploying massive virtualisation assets to support 3D network slices that can meet the emerging requirements of E2E space, aerial and ground applications. It helps space, aerial and ground network operators to immediately open their physical network infrastructure platforms to the concurrent deployment of multiple logical self-contained networks, virtualised and orchestrated according to their specific E2E service requirements. The created network slices are temporarily owned by tenants who have control over multiple layers, i.e., the physical layer, the virtualisation layer, and the service layer in 3D, of a unified 5G infrastructure, while they are also verticals. That is, they integrate the 5G infrastructure vertically on ground, aerial and space networks. The availability of this vertical market multiplies the monetisation opportunities of the network infrastructure as (i) new players, such as the space industry and military, may come into play, and (ii) a higher infrastructure capacity utilization can be achieved by admitting network slice requests and exploiting multiplexing gains. With network slicing, different services, such as space IoT, safety-critical aircraft operations connectivity, UAV connectivity, and mobile broadband, can be provided by different network slice instances. Each of these instances consists of a set of virtual network functions that run on the same infrastructure with a tailored orchestration. In this way, heterogeneous requirements can be provided on the same infrastructure, as different network slice instances can be orchestrated and configured separately according to their specific requirements, e.g., in terms of network QoS. Additionally, this is performed in a cost-efficient manner as the different network slice tenants share the same physical infrastructure.

2.2 Areas of study

Based on PASSIONATE's architecture, we will now present some indicative areas for the use of physics-based AI models.

AI/ML for wireless]: ML is widely used to tackle challenging problems in wireless networks. However, ML architectures applied in wireless are typically inherited from other fields (e.g., computer vision and natural language signal processing) and are blindly applied to wireless. When applied in large-scale networks, these ML models result in poor scalability and poor generalizations, with large performance gaps compared with theoretically optimal optimisation methods. To unlock ML for wireless, it is essential to customise and account-by-design the unique properties ("physics-based") of the networks they are applied to, while ML methods for wireless are currently inherited from other fields and ignore the network structure [12]. In particular, the application of deep learning (DL) to wireless communications problems has attracted significant attention, with one of the considered problems being hybrid analog/digital beamformer (HBF) design. Two typical DL techniques are often applied: purely data-driven DL and hybrid model-based DL. The former relies mainly on the learning capability of DNNs, convolutional neural networks (CNNs), or deep reinforcement learning to generate HBF beamformers. Yet, such a purely data-driven DL approach has major limitations due to its resource constraints, high complexity, and black-box nature [13]. Due to the large-scale deployment of massive MIMO (mMIMO) systems, efficient DL models with lower complexity implementations and stable and fast convergence need to be developed. In turn, a physics-based approach to DL encompasses a family of hybrid methodologies for combining domain knowledge with data to realize efficient inference mappings. A leading hybrid methodology is deep unfolding, which leverages DL techniques to improve model-based iterative optimisers in terms of convergence, robustness, and performance [14]. Also, graph neural networks (GNNs) have achieved promising results in applications including beamforming, and it was shown that a GNN trained on a network with 50 users is able to achieve near-optimal performance in a larger network with 1000 users [15]. Furthermore, thanks to the parallel execution, GNNs are computationally efficient. However, despite these empirical successes, the theoretical underpinnings and design guidelines remain elusive, which hinders the practical implementations of GNNs and deep unfolding methods in wireless networks.

Sensing and context awareness: In modern radio communication networks nodes are required to be aware of their context of operation, utilizing information on ambient networks, links, devices, and applications. This allows for the efficiency of network operation and the quality of provided services. PUT presents in [16] a complete overview of AI/ML methods applied for context-awareness in radio communication systems. Additionally, in [32] Spectrum Sensing (SS) architectures based on Federated Learning (FL) algorithms have been evaluated particularly for their effectiveness in SS performance and security. An important part of context awareness in 3D networks are the communication and computing resources at different layers. In [33] PUT has analyzed communication and computing resources vs. arriving requests (from the end-users) based on model-based classification. Again security aspects have been also considered. The identified knowledge gaps are as follows:

- When AI/ML methods are applied to enrich context awareness, they must converge at a required pace due to the system dynamics, and with accuracy tailored to the application.

- A major challenge is to define and incorporate the required functionalities of sensors/nodes for context-information acquisition, storage, and distribution while maintaining low power consumption.

- Without a comparison of the performance of various models on the same data, it is hard to design an approach that works best and to decide in what aspect it could be improved.

- The context awareness in 3D network architecture would require model- and physics-based approach to correctly manage communication and computing resources vs. requirements (demand vs. offer)

- Apart from the performance efficacy of different physics-based and model-based ML algorithms utilizing context awareness, their complexity and security needs to be considered.

Waveforms and transceivers]: The appropriate waveform design for 6G communications remains uncertain, as various use cases will involve the use of both lower and higher frequencies, as well as the ability to support challenging high-mobility scenarios. This is particularly relevant for PASSIONATE network architecture, in which NTNs are integrated into the terrestrial 5G/6G communication networks. However, the traditional orthogonal frequency division multiplexing (OFDM) may result in poor performance under high Doppler spread effects, and, additionally, its high peak-to-average power ratio (PAPR) can push power amplifiers into their nonlinear operational regions, extremely degrading communications performance. Because of this, OFDM needs to be adapted to overcome these challenges, and new waveform designs are being investigated.

Recently, a constant envelope OFDM (CE-OFDM) method was proposed [17] to enhance amplification efficiency; however, it shows performance limitations in high time-varying

conditions. Conversely, orthogonal time frequency space (OTFS) [18] has been suggested for time-varying channels but faces similar PAPR challenges as OFDM. Thus, identifying a waveform that balances all necessary requirements is still an open research area.
UC3M has recently introduced frequency-modulated OFDM (FM-OFDM) [19] featuring a constant envelope and improved resistance to Doppler spread, though its theoretical advantages are yet to be verified in practical settings, where UC3M has started analyzing its practical performance for NTNs.
However, independently of the chosen waveform, current communication transceivers rely heavily on the precision of channel state information (CSI). Acquiring this data usually requires pilot signals, and the necessary overhead rises with both the number of antennas and time-variability of the channel. To bypass the need for CSI acquisition and sharing in massive MIMO, non-coherent methods have been proposed [20], allowing operation in single-user contexts with very high mobility. Extending these approaches to multi-user scenarios remains complex, though early work incorporating AI solutions has been reported by UC3M [21].
Another promising technological advancement for enhancing spectral efficiency are full-duplex (FD) communications. However, the cancellation of the loopback signal or self-interference (SI)—caused by signals leaking from the transmission side into the reception chain—remains a key challenge. Some recent studies indicate that AI-driven signal cancellation may outperform traditional signal processing approaches in this area [22].

RIS/Intelligent Surfaces/Holographic Radio: The controllability of the wireless environment is a concept that leverages recent advances in the field of dynamic metasurfaces, enabling RIS [23]. An RIS is a planar surface made of many quasi-passive and low-cost scattering elements, each of which can impose a phase shift/amplitude on the impinging electromagnetic signals in a fully customized way. These surfaces can configure the wireless propagation environment into a transmission medium with more desirable characteristics [24]. An industry specification group was established within ETSI [25].
A promising technology to realize massive arrays in a dynamically controllable and scalable manner at reduced cost and power consumption utilizes large surfaces of radiating metamaterial elements. These are known as dynamic metasurface antennas or holographic MIMO systems [26], which enable the implementation of a new generation of wireless transceivers with thousands of controllable radiating elements and few RF chains. Their benefits include the possibility of realizing many communication modes even in line-of-sight (LOS) channels, which results in a dramatic increase of the spatial capacity density,
Current research on reconfigurable surfaces and dynamic holographic massive MIMO transceivers is insufficient to assess the benefits of these two technologies, especially in the mmWave and higher frequency bands. While these two technologies are both built upon breakthroughs in dynamic metasurfaces, they are usually treated in isolation, without truly leveraging their joint potential for realizing the Internet of Surfaces (IoS) paradigm. Moreover, current communication models for RIS are oversimplified, are not physically consistent, and are often not used correctly. Being a newer concept, the fundamental performance limits of holographic massive MIMO transceivers are not yet understood, especially if these surfaces are large and are deployed in channels dominated by LOS propagation, as channels in high-frequency bands are. The analysis of deploying RIS and holographic MIMO transceivers in wireless networks is limited to simple network topologies, while no system-level assessments have been reported to date except in [27].

Scheduling and RRM: Mobile networks are supported by efficient RRM techniques aiming to allocate the scarce available resources (time, frequency, space) to optimise several KPIs, resulting in a multi-objective optimisation problem. When evolving to a larger number of antennas and users, the number of resources to be allocated increases, making the problem more challenging. Effective RRM requires near real-time processing, and further advances are needed to overcome the exponential complexity of link adaptation, precoding, and scheduling. Several important challenges of multi-segment networking that significantly influence the overall network performance have been recognized in [28], including network control, spectrum management, energy consumption, routing design and handover management, and security issues. Task offloading in IoT systems with the aid of air platforms can be optimized using Q-learning-based solutions [29]. Recently, ML-based solutions have been proposed for load distribution and load balancing in satellite networks [30].

Energy consumption: Normally, the training of widely used ML models rely on a massive amount of data that need to be collected to cloud servers, usually far from where data is generated. Apart from the latency issues that this causes, these data servers consume large amounts of power for their operation. A characteristic example is the one of bitcoin whose total energy annual consumption has been estimated to be the same as the whole country of Norway. Such a problem in telecommunication networks is expected to be exacerbated with the integration of terrestrial networks with their non-terrestrial counterparts. Apart from their very large energy consumption, it is widely argued that ground stations would not be sufficient for the AI-based data processing of the massive amount of data generated by the integrated network [31]. New solutions to reduce the energy consumption in this context are needed.

3 Conclusions

The aim of this deliverable is to introduce PASSIONATE's architecture, that will constitute the basis for which all the upcoming deliverables will rely on. Towards this we have first introduced the limitations of current telecommunication networks with respect to the deployment of AI methods. The methods used primarily rely on back-box approaches that have been successfully used in the past in other areas, such as computer vision, but it's not apparent how they can be customized with respect to telecommunication networks and whether they are appropriate at all to be used. This motivates the use of the so-called "physics"-based AI approaches that leverage the inherent structure of telecommunication networks.

Subsequently, we introduce the considered 3D PASSIONATE architecture that amalgamates the 3 layers, namely the terrestrial, aerial, and space layers. Such a 6G architecture is a characteristic example of the need for customized, physics-based AI solutions for the resource allocation, to handle the large complexity and dynamicity of the network. Finally, areas of interest, based on the 3D architecture, have been introduced that identify the innovations to brought in this project.

References

[1] [Online]. Available: https://www.technologyreview.com/2023/10/26/1082028/ai-powered-6g-networks-will-reshape-digital-interactions/.

[2] "6G Use Case and Analysis," NGMN Alliance, 2022.

[3] [Online]. Available: https://www.qualcomm.com/news/onq/2024/06/path-to-6g-envisioning-next-gen-use-cases-for-2030-and-beyond.

[4] G. Intelligence, "The Mobile Economy 2024," 2024.

[5] W. E. Forum, "What is a satellite-connected smartphone – and how can it help bridge the digital divide?," 2022.

[6] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. M. Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, E. Lagunas and B. Ottersten, "Evolution of Non-Terrestrial Networks From 5G to 6G: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2633-2672, 2022.

[7] [Online]. Available: https://www.starlink.com/.

[8] T. Darwish, G. K. Kurt, H. Yanikomeroglu, G. Senarath and P. Zhu, "A Vision of Self-Evolving Network Management for Future Intelligent Vertical HetNet," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 96-105, August 2021.

[9] COMARCH, "Towards Self-Organizing Networks," 2009.

[10] [Online]. Available: https://nybsys.com/self-organizing-networks-son/#:~:text=In%205G%20networks%2C%20SON%20refers,deliver%20a%20better%20user%20experience..

[11] N. Shlezinger, J. Whang, Y. C. Eldar and A. G. Dimakis, "Model-Based Deep Learning," *Proceedings of the IEEE*, vol. 111, no. 5, pp. 465-499, 2023.

[12] A. Zappone, M. D. Renzo, M. Debbah, T. T. Lam and X. Qian, "Model-Aided Wireless Artificial Intelligence: Embedding Expert Knowledge in Deep Neural Networks for Wireless System Optimization," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 60-69, Sept. 2019.

[13] A. Jagannath, J. Jagannath and T. Melodia, "Redefining Wireless Communication for 6G: Signal Processing Meets Deep Learning With Deep Unfolding," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp. 528-536, Dec. 2021.

[14] V. Monga, Y. Li and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18-44, March 2021.

[15] Y. Shen, Y. Shi, J. Zhang and K. B. Letaief, "Graph Neural Networks for Scalable Radio Resource Management: Architecture Design and Theoretical Analysis," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 101-115, Jan. 2021.

[16] M. Wasilewska, A. Kliks, H. Bogucka, K. Cichoń, J. Ruseckas and G. Molis, "Artificial Intelligence for Radio Communication Context-Awareness," *IEEE Access*, vol. 9, pp. 144820-144856, 2021.

[17] S. C. Thompson, A. U. Ahmed, J. G. Proakis, J. R. Zeidler and M. J. Geile, "Constant Envelope OFDM," *IEEE Transactions on Communications*, vol. 56, no. 8, pp. 1300-1312, August 2008.

[18] P. Raviteja, K. T. Phan, Y. Hong and E. Viterbo, "Interference Cancellation and Iterative Detection for Orthogonal Time Frequency Space Modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6501-6515, Oct. 2018.

[19] J. L. Hernando and A. G. Armada, "Frequency-Modulated OFDM: A New Waveform for High-Mobility Wireless Communications," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 540-552, Jan. 2023.

[20] K. Chen-Hu, Y. Liu and A. G. Armada, "NON-COHERENT MASSIVE MIMO-OFDM FOR COMMUNICATIONS IN HIGH MOBILITY," *ITU Journal on Future and Evolving Technologies*, vol. 1, no. 1, Dec. 2020.

[21] M. J. Lopez-Morales, K. Chen-Hu, A. Garcia-Armada and O. A. Dobre, “Constellation Design for Multiuser Non-Coherent Massive SIMO Based on DMPSK Modulation,” *IEEE Transactions on Communications,* vol. 70, no. 12, pp. 8181-8195, Dec. 2022.

[22] I. Bilbao, E. Iradier, J. Montalban, Z. Hong, L. Zhang, W. Li and Y. Wu, “AI-based Inter-Tower Communication Networks: First approach,” in *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB),* Bilbao, 2022.

[23] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. d. Rosny, A. Bounceur, G. Lerosey and M. Fink, “Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come,” *EURASIP Journal on Wireless Communications and Networking,* vol. 129, May 2019.

[24] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang and E. Björnson, “RIS-Aided Wireless Communications: Prototyping, Adaptive Beamforming, and Indoor/Outdoor Field Trials,” *IEEE Transactions on Communications,* vol. 69, no. 12, pp. 8627-8640, Dec. 2021.

[25] [Online]. Available: https://www.etsi.org/committee/1966-ris). .

[26] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di and M. Debbah, “Holographic MIMO Surfaces for 6G Wireless Networks: Opportunities, Challenges, and Trends,” *IEEE Wireless Communications,* vol. 27, no. 5, pp. 118-125, Oct. 2020.

[27] B. Sihlbom, M. I. Poulakis and M. D. Renzo, “Reconfigurable Intelligent Surfaces: Performance Assessment Through a System-Level Simulator,” *IEEE Wireless Communications,* vol. 30, no. 4, pp. 98-106, August 2023.

[28] N. Kato, Z. M. Fadlullah, F. Tang, B. Mao, S. Tani, A. Okamura and J. Liu, “Optimizing Space-Air-Ground Integrated Networks by Artificial Intelligence,” *EEE Wireless Communications,* vol. 26, no. 4, pp. 40-147, August 2019.

[29] S. Gu, Q. Zhang and W. Xiang, “Coded Storage-and-Computation: A New Paradigm to Enhancing Intelligent Services in Space-Air-Ground Integrated Networks,” *IEEE Wireless Communications,* vol. 27, no. 6, pp. 44-51, Dec. 2020.

[30] Y. Zhao, H. Yao, Z. Qin and T. Mai, “Collaborate Q-learning Aided Load Balance in Satellites Communications,” in *2021 International Wireless Communications and Mobile Computing (IWCMC),* Harbin City, 2021.

[31] Y. Qiu, J. Niu, X. Zhu, K. Zhu, Y. Yao, B. Ren and T. Ren, “Mobile Edge Computing in Space-Air-Ground Integrated Networks: Architectures, Key Technologies and Challenges,” *MDPI . Sens. Actuator Netw.,* vol. 11, no. 4, Sept. 2022.

[32] M. Wasilewska, H. Bogucka, H. V. Poor, “Secure Federated Learning for Cognitive Radio Sensing”, *IEEE Communications Magazine*, Vol. 61, no. 3, March 2023, pp. 68 - 73

[33] Y. Salmi, H. Bogucka, “Poisoning Attacks Against Communication and Computing Task Classification And Detection Techniques”, *Sensors*, 2024, No. 2: 338, pp. 1–19,

Annex 1