Enabling 6G Smart Factories with O-RAN

André Costa Drummond[1], Osman Tugay Başaran[1], Abhishek Dandekar[1], Max Franke[1], Mihail Balanici[1], Iulisloi Zacarias[1], Ítalo Brasileiro[1], Cao Vien Phung[1], Ehsan Tohidi[1], Naveed Khan[1], Sepideh Kouhini[1], Lorenzo Miretti[1], Zoran Utkovski[1], Emre Durmaz[1], Julius Schulz-Zander[1], Setareh Maghsudi[1], Stefan Schmid[1], Falko Dressler[1], Behnam Shariati[1], Johannes Karl Fischer[1], Ronald Freund[1], and Admela Jukan[1]

[1]Affiliation not available

July 29, 2024

Enabling 6G Smart Factories with O-RAN

André Drummond, Osman Tugay Başaran, Abhishek Dandekar, Max Franke, Mihail Balanici, Iulisloi Zacarias, Ítalo Brasileiro, Cao Vien Phung, Ehsan Tohidi, Naveed Khan, Sepideh Kouhini, Lorenzo Miretti, Zoran Utkovski, Emre Durmaz, Julius Schulz-Zander, Setareh Maghsudi, Stefan Schmid, Falko Dressler, Behnam Shariati, Johannes Karl Fischer, Ronald Freund, Admela Jukan

***Abstract*—One of the most prominent use cases for 6G is the mobile networks for Industry 4.0, which are networks with enhanced characteristics to provide optimized and customized services within a smart factory. This scenario demands high-capacity data communication, which will require the usage of new portions of the electromagnetic spectrum (mmWave/sub-THz). These high frequencies impose new challenges on communication establishment and control, thus requiring tailored solutions from the radio access network (RAN). The Open RAN (O-RAN) specification promotes the RAN functional disaggregation and centralized intelligent control, thus providing the ideal basis for future smart factories' control. Nevertheless, the current O-RAN cannot handle some challenges brought by new technologies in high-frequency regimes. This paper presents these new challenges and proposes extensions to the O-RAN architecture and the underlying transport network infrastructure to support the 6G networks for enabling the future Industry 4.0 smart factories.**

I. Introduction

One of the most prominent use cases for 6G is the mobile networks for Industry 4.0, which are networks with enhanced characteristics to provide optimized and customized services within a smart factory. In this scenario, radio coverage, bandwidth, and dependability/security come with different requirements and opportunities compared to traditional cellular networks [1].

Smart factory automation typically has stringent requirements for determinism, low latency, and reliability. However, some applications, such as surveillance, do not necessarily require particularly high capabilities but may require high data rates. In contrast, support for multiple concurrent connections between machines and sensors is necessary in other applications like manufacturing and industrial IoT. The new smart factory services must embrace all service classes foreseen in the fifth-generation (5G) mobile networks, including enhanced mobile broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra Reliable Low Latency Communication (URLLC) [2] or a combination of those [3].

A. Drummond, I. Zacarias, C. Phung, I. Brasileiro, and A. Jukan are with the Institute for Computer and Network Engineering, Technical University Braunschweig, Germany. A. Drummond is also with the Department of Computer Science, University of Brasilia, Brazil. A. Dandekar, E. Tohidi, J. Schulz-Zander, L. Miretti, and Z. Utkovski are with Fraunhofer HHI, Germany. L. Miretti is also with Technical University Berlin. E. Durmaz was with Fraunhofer HHI at the time of this work. M. Balanici, S. Kouhini, B. Shariati, J. K. Fischer, and R. Freund are with the Photonic Networks and Systems Department, Fraunhofer HHI, Berlin, Germany. O. Başaran, M. Franke, S. Schmid, and F. Dressler are with the School of Electrical Engineering and Computer Science, Technical University Berlin, Germany. S. Maghsudi is with the Faculty of Electrical Engineering and Information Technology at Ruhr University Bochum

The O-RAN Alliance, a consortium of industry and academic institutions, is working on a body of specifications to provide a novel mobile network architecture to telecom operators composed of standardized interfaces that will allow the adoption of multi-vendor solutions to deliver high-performance services on the current 5G mobile networks and beyond. To do that, two main principles are considered: functional disaggregation and centralized intelligent control. In the first, Open Radio Access Network (O-RAN) promotes the Third Generation Partnership Project (3GPP) functional split, where base station (BS) functionalities are virtualized as network functions and spread across multiple network nodes: centralized unit (CU), distributed unit (DU) and Radio Unit (RU). The second principle is the radio access network (RAN) Intelligent Controller (RIC). This new architectural component provides a centralized network abstraction, allowing operators to implement and deploy custom control loops in the RAN [4], [5].

In parallel to the developments/discussions within O-RAN, over the last three years, a broad range of research and technical communities have been defining and refining a general vision for the sixth-generation (6G) mobile networks [6]–[8]. In this context, in June 2022, the O-RAN Alliance founded the Next Generation Research Group (nGRG), a task force created to discuss and mold the evolution of O-RAN specifications to support 6G and beyond. To this date, no recommendations have been proposed. Concomitantly, the researchers and engineers have been discussing novel applications, use cases, and the challenges they brought to fulfill the vision of a future 6G mobile network [3], [5], [9].

With the smart factory scenario, a new spectrum exploitation opportunity arises for high-frequency bands. Currently, the 3GPP already defines the usage of the millimeter wave (mmWave) band for 5G. Moreover, and especially for the smart factory scenarios, the sub-terahertz (sub-THz) bands are considered the next frontier to be explored in 6G networks [10]. Sub-THz bands act as an enabler technology for a variety of smart factory applications, ranging from providing super high bandwidth for eMBB, to allowing small antennas to be installed in IoT devices (mMTC) or enabling high directivity and diversity gains in large arrays of antennas over MIMO techniques, which improves connection robustness for URLLC applications.

A. 6G technologies for enabling smart factories

Fig. 1, on the top, presents a simplified view of the RAN control protocol stack with its distribution among the disaggregated units, as specified in O-RAN. This protocol stack

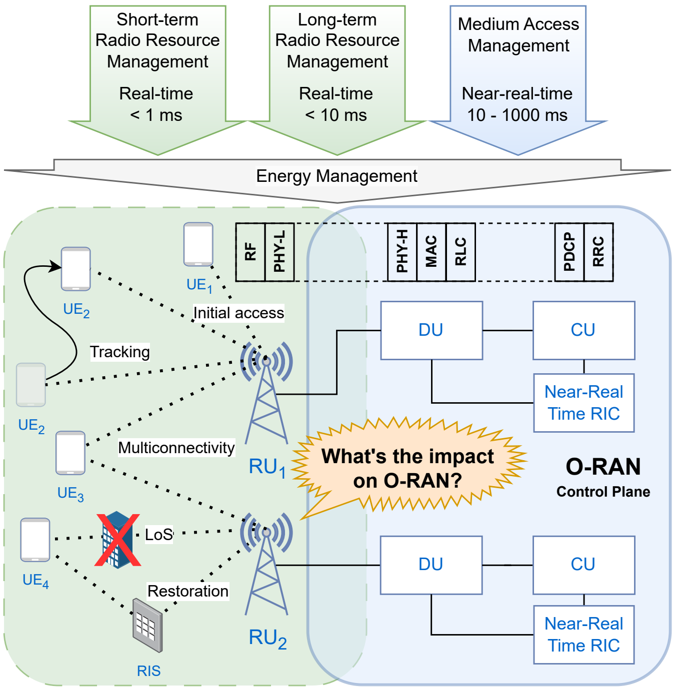

Fig. 1. RAN Management requirements (top) and the main challenges that smart factory networks (left) bring to the O-RAN control plane (right).

includes physical and link layer functions that are used for (i) the short-term radio resource management, which deals with modulation selection, interference mitigation, blockage detection, etc.; (ii) the long-term radio resource management, which controls the radio resource scheduling, beamforming, etc.; and (iii) the medium access management which controls the channel multiplexing, RAN slicing, etc. The lower the function (closer to the RU), the faster its execution needs to be. For higher functions (closer to the CU), there is more time flexibility, thus entering the domain of the near-real-time RIC, as proposed in O-RAN. At this point, it is important to note that the O-RAN Alliance has made a technically inaccurate appropriation of the term real-time to define its RICs. As is well known by the concurrent and distributed systems community, real-time systems must guarantee a response within specified time constraints, often referred to as deadlines [11]. Thus, being real-time does not mean that a system or a function needs to run fast. For the O-RAN specification, a near-real-time action needs to be completed in a time frame from 10 milliseconds up to 1 second. Anything below that is considered real-time, and above that, non-real-time. This document will employ the real-time term defined by O-RAN. Nonetheless, O-RAN does not specify a real-time controller in its current version.

On the left-hand side of Fig. 1, some challenges to User Equipments (UEs) connectivity brought by applying mmWave and sub-THz frequency ranges in smart factories are depicted. When we move from the traditional sub-6GHz frequency range, we lose the radio omnidirectional capability and pass to a directional beam regime to enhance signal strength and establish communication. Thus, a beam-sweeping phase is required to establish the initial access of a UE (e.g., UE_1), a procedure whose complexity rises proportionally to the frequency considered since the higher the frequency, the narrower the beans. Moreover, as a UE moves (e.g., UE_2), it is necessary to maintain the alignment between its beam and the RU beam, thus requiring fast-tracking of the mobility. Furthermore, due to the physical characteristics of high-frequency beams, they can be easily absorbed by obstacles, thus being susceptible to line of sight (LoS) blocking, which will require multi-connectivity solutions or the restoration of the communication by the application of reconfigurable intelligent surface (RIS) (e.g., UE_3 and UE_4), which introduces another device to be considered and controlled by the RAN, requiring a tight synchronization with the RU and UE beam alignment procedure. Finally, given the extremely high power consumption at these frequencies, fine-grained energy management will be required to control the activation of such devices efficiently.

These new high-frequency regimes put pressure on the control plane of the RAN by demanding control loops that need to be executed in real-time, which means at the DU and RU RAN function levels. It is necessary to place computational power on the nodes that will run these functions to meet these requirements, thus allocating it to the edge of the RAN. Moreover, an efficient telemetry data-gathering process is required to feed the RAN functions with the relevant data in a timely and granular fashion.

A big challenge is meeting the strict requirements regarding latency on the control and data plane while achieving interoperability through open architectures and interfaces. These requirements directly affect the so-called network X-haul of mobile networks because it should simultaneously transmit data fast enough to meet the expected QoE, be flexible in the sense that multiple technologies will be deployed in the same network, and reduce deployment costs. Additionally, the transmission to O-RAN presents numerous challenges, particularly in aligning it with optical network technologies. As we delve deeper into the various technologies supporting optical 6G X-haul networks, it is crucial to explore the intricate details of these technologies, their integration challenges, and their potential to meet the demanding requirements of future network architectures.

In the literature, efforts are already made towards implementing real-time control in O-RAN [12]–[15] to support demanding 6G networks use cases, such as smart factories. In [12], an extension to O-RAN is proposed to achieve fast customized control loops by implementing dApps, distributed applications that run at the CUs/DUs. One benefit of pushing the computation to the network's edge is lower overall latency and overhead. Moreover, it allows direct access to DU/CU functionalities, allowing the control of MAC- and PHY-Layer functions. The caveat is that this solution requires enough computational power (hardware acceleration) at the CU and DU nodes, given the high coupling with the dAPPs. Moreover, new interfaces must be standardized to guarantee the exchange of the necessary information between dAPPs and the near-real-time RIC or DUs/CUs in a platform-independent fashion.

Another solution, EdgeRIC, is presented in [13], [14]. EdgeRIC is a real-time RIC co-located with, but decoupled, from the DU that implements μApps, which are applications for real-time control. By co-locating with the RAN, EdgeRIC does not interfere with latency-constrained RAN tasks but can access information and provide control in real-time. The proposed design allows the RAN to continue its normal operation without input from the EdgeRIC. The authors demonstrate that this solution leads to considerable gains in throughput compared to the near-real-time RIC approach [13].

Kaliski et al. [15] introduce the Multi-access Edge Computing (MEC) control loop concept to indirectly allow the UE to communicate with the near-real-time RIC. By doing that, the UE exchange with the RIC can lead to changes in the network configuration, such as deploying novel ML models computed in high-capacity MEC to the UE.

The proposed real-time RIC solutions from the literature consider deploying Apps in the DU or CU or co-locating them. Indeed, a modular real-time RIC implementation might be a good choice for reducing the coupling with other functions of the RAN and allowing the centralization of tailored computational resources at the edge. Regardless, a closer look at the 6G smart factory use cases is needed to identify the requirements for low latency control at the RAN [16].

This paper motivates the need for a real-time RIC, considering the challenges of the new high-frequency bands proposed to be deployed in future smart factories. Moreover, it discusses the need for an extended interface to the UE to allow its control in a fine-granular manner and the importance of interfacing with RISs to provide reliability to the communication system. It also discusses the impact of the 6G technologies on the optical transport network (X-haul), which is responsible for interconnecting CUs, DUs and RUs.

Our paper targets academics, practitioners, and industry as well as senior students with an essential background in networking who would like to learn more about RAN and RIC-related aspects in general and real-time specific smart factory use cases in particular.

The document is organized as follows: Section II presents the smart factory scenario and its stringent requirements. Section III presents the evolution of RAN architectures and discusses the challenges of the new 6G technologies in detail. Section IV discusses architecture expansion/modifications to the O-RAN and presents some application scenarios. Section V analyzes the X-haul functional split, interfaces, and architectures, highlighting the key challenges and proposed solutions for seamless integration with O-RAN systems. Finally, Section VI concludes the document.

II. Smart Factories

Historically, mobile network evolution was driven by increasing network bandwidth. Although such motivation is still present in the evolution to 6G, with predictions of global peak data rates of 1 Tbps and peak data rates of 1 Gbps [17] for 95% of UEs, new service classes and Key Performance Indicators (KPIs) originate from new proposed use cases. Industry digitalization is a prominent yet still-to-be-explored niche, aiming to create new services and improve factory performance.

A recent survey from the O-RAN next Generation Research Group (nGRG), where big technology companies were inquired about 6G use cases and potential technology gaps, confirmed a significant interest in specialized vertical industries such as smart factories. Following a similar trend, both the Next G Alliance [18] and the ITU-T Focus Group FG-NET2030 [19] recently published reports on 6G use cases and applications. From the listed reports, it is clear that smart factory – already a trend in fifth generation of cellular networks (5G) – will further evolve, and the network will be required to support even more challenging use cases.

As a foreseen general configuration, non-public (private or campus) networks deployed in smart factories consist of a dense concentration of intelligent mobile robots requiring, at the same time, low latency and high reliability for the control and high data rates for the transmission of high-definition images or maps to be further processed by artificial intelligence (AI) algorithms. Authors see this new service class as a mix of KPIs of URLLC and eMBB services [20]. This new service is expected to deliver user-experienced data rates up to 1 Gbps while meeting latency requirements of 10 ms in the RAN [21]. It is worth noting that the reliability KPI of URLLC services is envisioned to rise from 99.999% to 99.99999% [22].

In the same direction, a mix of characteristics of URLLC and mMTC is required for some factory applications, originating the Massive URLLC (mURLLC) [23] class of service. This class of service is necessary to support, for example, a dense concentration of remote-controlled intelligent mobile robots operating on the factory floor. The remote control of such robots requires critical communication aspects of wireless networks. The production, inspection, and maintenance of critical assets will be automated by unmanned robots and vehicles (such as UAVs). Big-data analytics of the factory processes and cyber-physical systems, such as Digital Twins, will require high data rates and reliable connectivity with the physical systems.

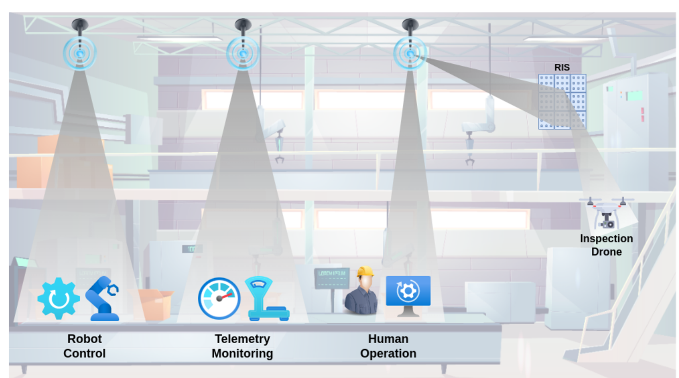

Fig. 2. Smart factory schematic, where multiple radio units (RU) connect to a variety of fixed and mobile devices (UE) either directly or via a reconfigurable intelligent surface (RIS).

Automation of factories will heavily rely on network-enabled robotics and autonomous systems. Applications such as remote control of industrial robots, especially the ones with human-machine interfaces (HMI), pose strict performance requirements for the network. For example, field robots or Unmanned Aerial Vehicles (UAVs) can be employed to inspect smart petrochemical plants, searching for specific situations, such as pipe leaking or abnormal thermal signatures, as depicted in Fig. 2. Many field robots inspect the assets on-site while the human operator monitors the factory and controls the robots from a secure (and probably remote) location. The robots collect information, create a high-precision environmental mapping, and send the data to a Mixed Reality (MR) application that merges the received data with real-time computed data (for example, to recommend preventive maintenance actions). Full context awareness and real-time robot control are required for the reliable and safe operation of the described inspection system, requiring high network throughput, ultra-low latency, and ubiquitous connections simultaneously [18], [24].

The smart factory scenario imposes strict requirements on the RAN deployed on-premises, thus bringing novel challenges to the next-generation mobile infrastructure.

III. 6G RAN Challenges

Following outlining the smart factories' requirements for 6G, we will present the evolution and current state of RAN architectures and highlight the challenges that some newly envisioned technologies will bring. In particular, we will consider sub-THz, reconfigurable intelligent surfaces, ICAS, and RAN telemetry.

A. Evolution of RAN Architectures

Over the years, Cellular networks have evolved from simple systems capable of voice transmissions to networks capable of transmitting data at Gigabits per second at very low latency. Since the complexity of the network increased, the design of mobile networks was split into two main components: the RAN and the core network (CN). From the second generation of cellular networks (2G) onward, we can see in Fig. 3 the evolution of the RAN, keeping backward compatibility wherever possible.

Aiming to reduce the cost of the network and simplify the internal design of RAN components, cloud RAN (C-RAN) was introduced in the fourth generation of cellular networks (4G), splitting the RAN into base band unit (BBU) and remote radio heads (RRHs). Therefore a new interface, called "Fronthaul", was introduced to connect these two blocks. In 5G networks, flexible splitting of RAN is possible, further disaggregating the BBU into CU, DU, and RU, as depicted at the bottom of Fig. 3. Similar to what happened in 4G RAN, a new interface was introduced to connect the CU to the DU, called "Midhaul" while the interface connecting the DU to RUs is still labeled as "Fronthaul". The new design allows the RAN infrastructure to be placed close to the end user, achieving very low latency on the radio interface. As a drawback, the disaggregation of RAN introduces complexity in the management/control plane, motivating the O-RAN Alliance to propose new management/control approaches, such as the introduction of the RICs and its related interfaces as depicted in Fig. 4.

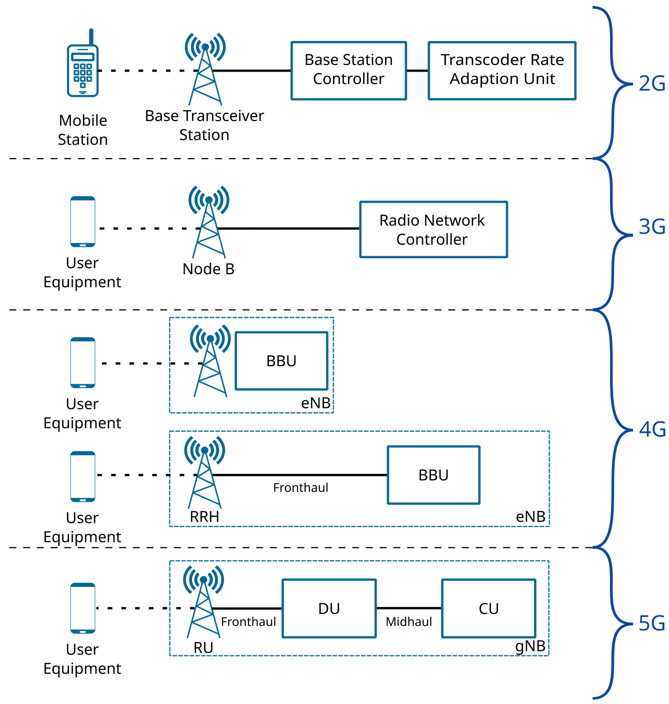

Fig. 3. Evolution of RAN in mobile networks.

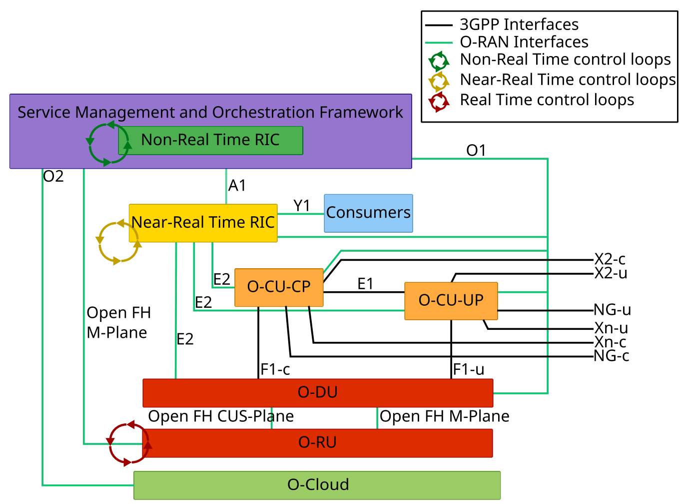

Fig. 4. Based on O-RAN Architecture Description 8.0

B. O-RAN

Conventional black-box radio access network solutions have theoretical and practical limitations. Theoretical limitations are mainly scalability, interoperability, and algorithmic complexity. Besides, practical limitations are high deployment costs, limited optimization capacity, slow adoption of new technologies, and security concerns. Researchers have moved towards more flexible and collaborative radio access network designs to overcome these limitations. Many stakeholders from industry and academia came together to initiate this change, especially the O-RAN Alliance Community, which is taking major responsibility for this innovative re-genesis. The result is a modern architecture composed of open, disaggregated, virtualizable, and cloud-friendly software components.

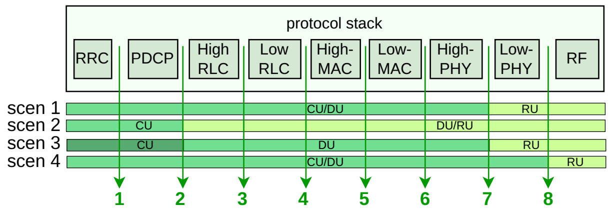

Fig. 5. Functional Splitting and use-case examples.

O-RAN used new functional splits in its architecture parallel with the 3GPP functional disaggregation paradigm [25]. As can be seen in Fig. 3 and Fig. 4, Next Generation Node Bases (gNBs) are subdivided into functional units: Open Central Unit (O-CU), Open Distributed Unit (O-DU), and Open Radio Unit (O-RU). In particular, the data processing load can be handled differently under O-RU and O-DU according to the network's bit rate and latency requirements. 3GPP defines eight main split options. O-RAN also introduced the Split7.2x option (comprising both options 7.2a and 7.2b Splits). Thus, the complexity of the O-RU in the PHY-Layer is reduced, which provides an advantage in terms of latency. With the Split 7.2x option, a fronthaul network that uses the open and simple interface protocol has been designed, paving the way for interoperable next-generation wireless networks.

From an architectural standpoint, the aforementioned functional splitting results in different possible placement scenarios of the O-RU, O-DU, and O-CU components, which can be positioned either at different locations, or be partially collocated and centralized. As such, a few main use-cases are currently distinguished [26] and summarized in Fig. 5:

- *Scenario 1*: A C-RAN architecture with distributed RUs and collocated DU and CU in which the fronthaul traffic from the split option 7, with high latency sensitivity and bandwidth requirements, passes through the optical transport network, while the midhaul traffic is local.
- *Scenario 2*: A C-RAN architecture with collocated RU and DU deployed at the cell-site and with CU connected remotely to the DU through the functional split option 2 and interface F1, where the fronthaul traffic is local (normally transported by a cell site router (CSR)), while the midhaul traffic is carried by the underlying optical access network.
- *Scenario 3*: A C-RAN deployment with two RAN splits, where the three components are hosted at separate locations, with the DU placed closer to the cell sites for shorter latency and the front- and midhaul traffic being carried through two separate optical transport networks (O-RAN's choice).
- *Scenario 4*: A C-RAN deployment with the functional split option 8, in which the complexity of RU components decreases even further, and the resource management centralized in the DU/CU is more efficient. On the other hand, due to the system's simplicity, the bit rate on the fronthaul links is constant, very high, and scalable with the number of antennas, which is impractical for scenarios with massive Multiple Input Multiple Output (MIMO) components [27].

Besides this revolutionary disaggregation in RAN, another innovative approach offered by O-RAN is intelligent autonomous network orchestration. It aims to improve network optimization capability with the help of these programmable components called RAN intelligent controllers (RICs). When the O-RAN architecture is examined (Fig. 4), three main components are seen, starting from the top and going down. Begin with the Service Management and Orchestration (SMO) framework and its non-real-time (non-RT) RIC component. The SMO framework connects to RAN nodes of the network through the A1 [28] and O1 [29] open interfaces defined by O-RAN Working Groups (WGs). SMO is responsible for managing, automating, and orchestrating RAN actions through open interfaces, non-RT RIC, and its own framework functions. Non-RT RIC is one of the most essential components of the O-RAN architecture along with near-RT RIC. Closed-loop control is also achieved by implementing AI/ML models via rApps and xApps running on these two programmable controllers [30].

Non-RT RIC focuses on RAN policy management, updates/upgrades, and radio resource management. Thus, it guides the near-RT RIC via the A1 interface. The policies and higher-layer procedures to be applied on the RAN are handled with the help of rApps. The near-RT RIC is located on the network's edge, close to the network elements such as distributed and central units. Thus, it accesses the critical data of the RAN components in a much shorter time, around 10 milliseconds (ms). xApps are software applications and microservices that run on the near-RT RIC. In particular, the near-RT RIC's data-gathering capacity on the RAN enables xApps to be used in various use cases. Intelligent network orchestration can be achieved by deploying AI/ML algorithms on xApps/rApps. They can be used especially in anomaly detection [31], traffic steering, and forecasting scenarios. While rApps execute control loops over 1 s, xApps execute control loops between 10 ms and 1 s.

O-RAN has gone far beyond being a vision. With its innovative architecture, it has provided groundwork for autonomous intelligent control. Realizing autonomous intelligent control requires integrating advanced AI and machine learning techniques and developing sophisticated control algorithms, which is currently a major research effort in 6G.

C. sub-THz communication

The exploitation of the vast portions of available spectrum is the primary motivation of sub-THz communications in smart factories, as it is the preferred mean to deliver the exceptionally high data rates required by the most demanding 6G applications. In fact, while extremely high network capacity can also be achieved at lower frequencies using advanced MIMO technologies (such as massive MIMO and its evolutions), this capacity typically needs to be split among a large number of relatively low-rate UEs due to practical limitations of the propagation channel and of the hardware, which limit the maximum per-UE data rate. In addition, the abundant spectrum offers opportunities for improving latency and sensing capabilities. However, at the same time, sub-THz communications also come with significant technical challenges related to the fundamental characteristics of high-frequency systems. A first major issue is that high-frequency communication is impaired by harsh radio channel conditions due to significant power losses as the signal propagates (path loss), and to high sensitivity to signal blockage. A second major issue is the high energy consumption due to complex hardware and signal processing. Therefore, in order to score the promised gains in practice, sub-THz access networks must counteract the aforementioned harsh radio channel conditions and keep the energy consumption to a tolerable level. To this end, from a physical layer perspective, sub-THz access networks are expected to operate under the following conditions:

- highly directional beamforming along unobstructed propagation paths;
- dense network deployments;
- low complexity hardware and signal processing.

Beamforming is used to counteract the path loss, by focusing the radiated power around specific directions, i.e., beams, instead of spreading it across large region of space as with traditional omni-directional antennas. Only when the beams are properly aligned, the channel is stable enough to be used for communication. Any blockage or misalignment of the beam will lead to severe power drops and hence loss of connectivity. While this beam alignment procedure could be theoretically carried at the physical layer using sophisticated hardware (e.g., fully digital beamforming architectures) and signal processing techniques, in practice, due to the need for low complexity hardware and signal processing, beamforming is typically implemented in 5G mmWave systems by means of so-called "beam management" protocols that select the best beams from a predefined list of available ones instead of calculating a beam ad-hoc. This is list is commonly referred to as codebook. However, to enable communications at reasonable range, i.e. in the order of meters, sub-THz systems are expected to use much larger codebooks of much narrower beams than in current mmWave systems. Therefore, classical beam management protocols based on beam sweeping mechanisms quickly become very inefficient or even impractical.

Furthermore, while reliable data communication is possible after beam alignment, all the control communications required to establish the connection, and hence taking place before beam alignment, are extremely unreliable if not impossible [32]. This also means that an out-of-band control channel is likely needed. This channel could come in the form of the lower frequency bands that are in use today. In fact, the entirety of the control plane can reside in the traditional frequency bands and make use of the spectrum that gets freed up by moving data plane traffic to the mm-Wave and sub-THz bands.

Dense network deployments are likely needed since, in practice, beamforming alone can counteract the path loss problem only up to a certain extent. In addition, dense network deployments are also beneficial in reducing the probability of blockage of the line-of-sight propagation path, which is by far the preferred one. However, to minimize the energy footprint, the network should use some form of energy saving mechanism that, e.g., activates the many RUs spread over the service area only when needed. Overall, the above points directly lead to a significant increase in required coordination between RUs, UEs, and higher load on the fronthaul. This without even mentioning the opportunity of implementing some form of "cell-free" multi-connectivity option to further reduce the impact of signal blockages and to provide seamless connectivity.

There are two additional challenges that RAN architectures could face by the deployment of sub-THz. On one hand, depending on the degree to which one wishes to virtualise physical layer functions, a requirement for a RIC that goes beyond the 10ms offered by the near-RT RIC could arise. On the other hand, the combination of extremely high bandwidth with high instability of sub-THz will lead to issues with existing congestion control mechanisms, resulting in possibly disastrous performance. While this is mostly a transport layer, and thus end-to-end, issue, there should be considerations made in how to alleviate these issues in the RAN. One option could be the extended use of performance enhancing proxies or similar mechanisms as a core part of the RAN. Similarly, it is expected that 6G UEs will make extensive use of multiconnectivity. While this can also be realized on the transport layer, i.e. in the form of multipath TCP or QUIC, there will also be scenarios where packet (de)duplication will need to be handled inside the RAN as to not flood the network core with unnecessary and redundant data.

D. Reconfigurable Intelligent Surfaces

A promising technology to support 6G networks in many smart factory use cases are RIS [33]. RISs enable novel ways of manipulating and extending the wireless channel to establish the desired communication link. Among the potential benefits, one essential use case of RIS is maintaining the benefits of a line of sight (LOS) when the direct LOS between the transmitter and receiver is blocked by providing a "virtual LOS". Although there are similarities to some conventional elements of the network, inherent differences with RIS need to be considered in a network architecture trying to integrate them. The novelty and challenges are related to three features of RISs: (1) RIS is an external element, not a part of the radio unit (RU). (2) RIS can be shared among different RUs, so an entity needs to decide which surface should be allocated to which UE/RU. And (3) RIS lacks the computational capacities of RUs and thus needs to be controlled entirely by the previously mentioned entity. Novel concepts will be required to offload the computations required while keeping the tight latency requirements for mechanisms such as beam steering.

The first set of challenges concerns the information that needs to be collected, such as the user's location and channel conditions and how this information should be considered for configuring the RIS. With RIS providing different channel qualities, there can be a trade-off between delay and channel quality. Furthermore, as in any wireless network, it's critical in an RIS network to mitigate unintentional interference. Unlike with RUs, RISs themselves are not capable of collecting such metrics. Instead, they rely on data gathered by the RU and evaluated by another entity to be configured appropriately. For similar reasons, initial access requires new approaches. Commonly, the RU performs beam sweeping to find new UEs and determine the best beam for each. If a new user instead appears inside the area only covered by an RIS, the RIS side has to perform the beam sweeping. How this is controlled and managed is still open to challenge. When a handover needs to occur, the question arises of how to modify existing procedures to accommodate users served by different RUs via different RISs. Instead of just being performed between RUs, there are now three additional types of handovers (RIS-RIS, RU-RIS, and RIS-RU) that all need separate considerations.

E. Integrated Communication and Sensing (ICAS)

6G networks will enable new applications with extreme bandwidths and high reliability combined with low latency and offer integrated radio sensing services such as localization and detection. This is particularly important for applications such as extended reality, networked robotics, and autonomous systems, common use cases in smart factories. In this respect, it is expected that a tight convergence of communication and sensing will take place in 6G.

Currently, O-RAN focuses on offering AI-based control to the higher layers of the protocol stack, with limited flexibility for the lower layers hosted at DUs/RUs. Indeed, limiting the execution of control applications to the near-RT and non-RT RICs prevents the use of data-driven solutions where control decisions and inference should be made in real-time or within temporal windows shorter than the 10 ms supported by near-RT control loops [12]. Practical examples are user scheduling and beam management for sub-THz communication. Scheduling requires making decisions at sub-millisecond timescales to support ultra-reliable low-latency communications (URLLC) traffic with latency values as low as 1 ms. Similarly, the narrow-beam operation in sub-THz communication comes at the expense of more advanced beam management (time-sensitive) and novel mobility and handover strategies.

These issues may become more prominent with the introduction of novel sensing and localization services in the context of integrating communication and sensing. In addition, introducing new and vertical wireless applications, such as virtual reality industrial applications and mission-critical services, comes with additional reliability and real-time operation requirements. Hence, it is of crucial importance that the evolving O-RAN architectures provide the required flexibility for the support of control decisions at different timescales, including real-time operation with control loops smaller than 10 ms (i.e., beyond what is currently supported by the Non-RT and Near-RT RIC). In this respect, different factors, such as data availability, application requirements, geographical requirements, and network workload, should be accounted for.

At the level of resource management, joint consideration of communication and radio sensing may lead to the definition of new and even more complex multidimensional optimization problems, where not only conflicting key performance indicators (KPIs) of the communication part have to be balanced against each other, but also KPIs of the sensing part. One implication is that the coexistence of sensing and communication implies in practice sharing of the same time-frequency-spatial resources. While the interplay between sensing and communication is expected to introduce some elementary trade-offs in this respect, there are also mutual benefits: sensing and location information can guide communication (e.g., for beam and mobility management), while communication can support localization and sensing by sharing map information between devices. Depending on the level of integration, sharing is not limited to the resources but can also exist at a waveform or hardware level. In general, the convergence of sensing and communication would have multi-fold implications on the system design, including the (i) definition of sensing-related services classes; (ii) development of new medium access control (MAC) and radio resource management (RRM) protocols to allocate the radio resources according to the needs of different sensing and communication services; (iii) extension of the network slicing concept to support sensing-related services.

F. RAN Telemetry

In upcoming 6G networks, different network telemetry solution are needed, fitting their respective domain such as mobile (RAN and Core), transport and Cloud network. First and foremost, in order to enable more control and management in the RAN, the O-RAN Alliance proposed the E2 interface along with the near-RT RIC. This interface is designed between the near-RT RIC and E2 nodes such as CU-C, CU-U, and DU or gNB. The main focus of the E2 interface is to enable self-optimizing networks, anomaly detection, and allocating radio resources at a coarse-grained level. These use cases involve network events and control decisions that occur relatively infrequently, ranging from tens to hundreds per second. This slower pace enables xApps (applications running on the Near-RT RIC) to gather telemetry data, make inferences, and adjust vRAN (virtual Radio Access Network) functions based on predefined control policies. The O-RAN Alliance has defined different ways for the Near-RT RIC to interact with E2 nodes, so-called *E2 interface service models* (E2SMs) [34].

Telemetry collection mechanisms in O-RAN face several significant challenges [35]–[37]. One notable challenge is about transporting all necessary uplink *IQ samples* from the physical layer (PHY) to the RIC, because RAN generates significant volumes of data at a high frequency. It is an infeasible approach to come up with a single API that collects information for different use cases. Service models are specialized static APIs that are integrated into the virtual Radio Access Network (vRAN) functions and designed based on specific use cases to capture only the essential information. This service model specifies the specific data that can be collected and at what level of granularity. Whenever an xApp developer aims to introduce a new use case, they must create a dedicated service model for it. However, this kind of data collection architecture cannot scale as the number of xApps increases and evolves.

Real-time requirements are making this problem even more severe. Many important vRAN control loops impose strict time constraints that ranges from tens of microseconds to milliseconds. Applications that require real-time control loops are, for instance, radio resource scheduling and power control. Such applications can use a set of pre-defined policies offered by service models, which can run inline inside the RAN functions to eliminate control plane intervention latency. However, current approach requires implementing control policies on a use-case basis and has scalability issues. Considering that any addition of service models has the risk of introducing performance degradations to the time-constrained operation of vRAN functions, RAN vendors are reluctant to add new features and service models.

In other words, the current challenge of the E2 interface is quite similar to that of OpenFlow years ago. OpenFlow was too intended to be device-agnostic. However, it evolved in a way that tightly coupled control apps with the specific flow tables supported by the underlying switches. The current E2 interface architecture does not provide the required flexibility to be extended for different verticals. For instance, if a smart factory use case requires the definition of a new service model with a different set of KPIs, these also need to be supported by the selected RIC and RAN vendors to add support for this service model. Moreover, it would also need to go through a lengthy standardization process, where all O-RAN vendors must be convinced to support it. As observed through the transition from *OpenFlow to P4*, this leads to a proliferation of service models characterized by a significantly slow process and lack of scalability.

G. End-to-End Cross-layer Network Telemetry

The softwarization of future 6G networks can lead to more secure, robust, and innovative networks if modern software engineering methods such as continuous integration (CI) and continuous delivery (CD) are implemented. This, however, requires more fine-grained monitoring, ensuring the proper operation of the entire network and services. The advent of private campus networks enables the adjustment of network solutions and applications to meet customers' needs. Treating the design and development of applications as part of the entire solution is akin to how hyper scalers such as Google and Facebook operate their networks. Following the cross-layer paradigm, network, services, and application monitoring must go hand-in-hand. Today, most network monitoring is mainly achieved by leveraging legacy interfaces such as SNMP or Netflow/SFlow. This, however, only provides very limited insights into the actual application's performance. Frameworks such as OpenTelemetry [38] enable novel Cloud-native interfaces enabling fine-grained, customizable monitoring of services, for instance, running in a containerized or virtualized cluster.

Another challenge is meeting the requirements for end-to-end quality-of-service (QoS) measurements due to the existing segmented network architecture. The current wireless system design focuses on increasing the capacity and the number of supported devices through improvements in the PHY layer. However, increases in QoS by architecture modifications are needed for emerging use cases such as URLLC, industry 4.0, and V2X. The increasing demand for those services emphasizes ensuring end-to-end network quality-of-service (QoS) guarantees with application-level precision. Compliance with *TS 22.261* necessitates implementing a real-time QoS monitoring mechanism for slices [39].

This involves various levels of QoS granularity, such as per set of flows, individual flows, or even individual data packets, particularly for mission-critical applications. It also entails transmitting event notifications to user equipment (UEs) in the event of QoS violations. This mechanism enables timely responses for interconnected automation devices through continuous reports and asynchronous events triggered by predefined loss and latency thresholds.

The current segmented design makes it challenging to cater to the needs of applications that rely on end-to-end QoS assurances. Each network segment or domain is managed by a dedicated orchestrator specializing in that domain, offering detailed configuration options. It is not feasible to expect a single technology or operator to handle all existing and future communication demands. Consequently, the current cellular network architecture lacks a consolidated approach for monitoring the performance of different segments, commonly known as unified Operations, Administration, and Maintenance (OAM).

Accordingly, next-generation O-RAN-based networks need end-to-end cross-layer mechanisms to verify the correct operation of a network, including latency, queuing delay, packet processing delay, etc. This requires mechanisms such as In situ Operations, Administration, and Maintenance (IOAM), a network measurement and monitoring technology that enables real-time traffic sampling at a line rate.

This Section highlighted the main challenges brought by the envisioned technologies in the realm of 6G smart factory networks. The next section will discuss the identified issues further and propose an extension to the O-RAN to tackle them.

IV. O-RAN Evolution towards 6G Smart Factories

With increasing multi-vendor deployments, O-RAN's openness paradigm for fronthaul, midhaul, and management interfaces has found its absolute position in the 5G era. In the evolution of O-RAN towards 6G smart factories, it is necessary to identify new issues on the horizon and some of the issues brought by the expansion of the existing ecosystem. In the following section, we will express solutions for these identified issues with clearer examples and discuss the proposed extensions to the O-RAN architecture (Fig. 6).

As we gear up for the advent of 6G, the evolution of O-RAN emerges as a critical enabler in tackling the challenges posed by cutting-edge communication technologies like sub-THz Communication and Reconfigurable Intelligent Surfaces, required for enabling smart factories. These innovations promise transformative capabilities, including ultra-high data rates, massive connectivity, and unprecedented reliability. However, their integration into existing networks demands novel architectural paradigms and robust standards, which O-RAN's flexibility and openness are poised to provide.

In response to the challenges posed by the emerging communication technologies, we are forging ahead with innovative

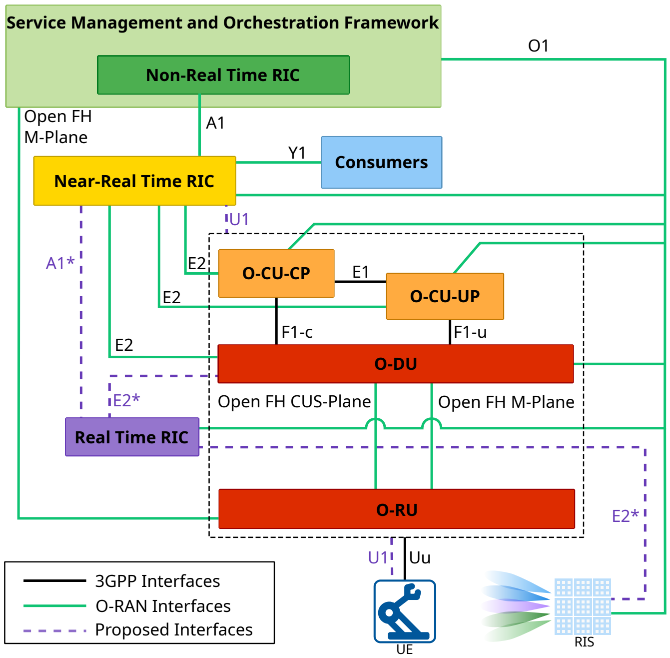

Fig. 6. Proposed extensions to the O-RAN architecture.

solutions, including the design of the new Real-time RIC and novel interfaces within the O-RAN framework. The new real-time RIC architecture (Fig. 6) stands at the forefront of this evolution, offering dynamic orchestration and optimization capabilities essential for efficiently managing complex networks operating at sub-THz frequencies and interfacing with RIS deployments. These RIC enhancements enable real-time adaptation to varying network conditions, ensuring optimal performance and resource utilization in dynamic environments and heterogeneous networking (HetNet). Additionally, introducing new interfaces within the O-RAN architecture facilitates seamless integration of RIS and other advanced technologies, enabling operators to harness their full potential while maintaining interoperability and scalability. Moreover, the new telemetry interfaces further foster the softwarization of upcoming O-RAN-based 6G smart factory networks.

By embracing these advancements, the O-RAN ecosystem is poised to overcome the challenges of next-generation communication technologies, ushering in a new era of connectivity that is agile, efficient, and future-proof.

A. *Necessity for a Real-Time RIC*

The closed-loop control provided by existing near-RT and non-RT RICs is not feasible for control applications below 10 ms. This limitation prevents the realization of Tactile Internet, AR/VR Applications, Real-time AI Inference, and URLLC latency-sensitive scenarios, which are vital use cases of 6G smart factories. As seen in the use case discussed in detail in Section IV-D, there are applications where 1 ms action time is required under the URLLC scenario, such as Factory Automation Applications, AGVs, and Autonomous Mobile Robots (AMRs) use. Similarly, scenarios related to Sub-THz Communication cannot be realized with the near-RT RIC service due to these 10 ms RAN elements' control-loop limitations. Especially when sub-THz communication systems are examined, overcoming beamforming complexity with advanced techniques, dynamic spectrum sharing requires rapid adjustments to spectrum allocation, and access protocols and control loops are required at lower than 10 ms intervals. Considering the technical limitations mentioned, a more time-sensitive RIC is vital to solve the aforementioned limitations.

B. *New Interfaces*

Introducing a real-time RIC requires two new interfaces-one to send enrichment information to real-time RIC (A1*) and the second to control network elements from real-time RIC (E2*). However, introducing new interfaces might lead to increased complexity and implementation costs. To mitigate this, we propose to reuse the same technologies/tech stack as those used to implement E2 and A1 interfaces; however, the service models, procedures, and messages for E2* and A1* might be different from those in E2 and A1. Additionally, interface compatibility is vital in designs where 6G networks are built with O-RAN. The open interfaces offered by O-RAN are required for the inter-operable operation of network components from different vendors to ensure correct standardization. For example, one of the challenges in this regard is the open frounthaul interface, ETSI and O-RAN took responsibility for it to be designed correctly [40].

Similarly, new interfaces to be defined in the combination of 6G and O-RAN should be designed to respond to compatibility concerns. The interface compatibility challenge is overcome by considering the standardization policy, abstraction methodology, and feedback mechanism. Although O-RAN provides open standards as ensured by its vision, we know from the applications in the field that vendors can define proprietary extensions, and they can design completely different use case-specific infrastructures. Considering the various technology stacks, services, and end devices that 6G smart factories will incorporate, the standardization processes of all open interfaces should be carried out with high precision.

We also identified that although the RICs are used to control the RAN, their control over UE is only limited by the functionalities supported by RRC. We have identified a new interface (U1) to enable newer services that allow communication between near RT-RIC and UE.

C. *Proposed Architecture and Solutions*

We now discuss the proposed added-on components and interfaces to O-RAN, which are required to prove it towards 6G in future smart factories.

1) Real Time RIC: We describe the architecture of Real time RIC in Fig. 7. This architecture is intended to be high level and not exhaustive. At its core, Real time RIC has the RT RIC platform which enable the time critical applications (tApps) via the RT RIC APIs. It consists of various subsystems like messaging subsystem for internal messaging, storage subsystem for storing various kinds of data and orchestration and management subsystem for tApps management. The RT RIC Platform also hosts terminations for A1*, O1 and E2* interfaces.

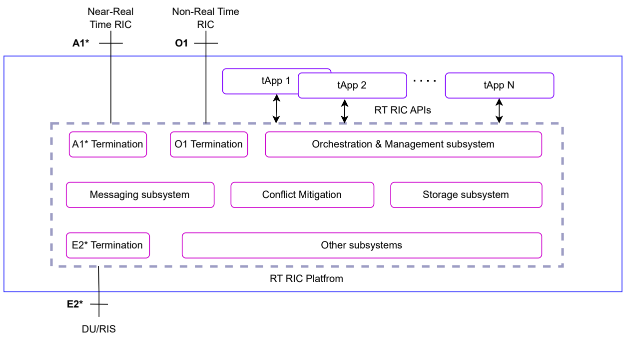

Fig. 7. Real Time RIC architecture.

2) A1 Interface:* This interface is similar to A1 interface. This interface will enable a bidirectional flow of enrichment information between near-RT RIC and RT RIC. This information may include policies, intents, predictions, etc.

3) E2 Interface:* The E2* interface is an extension of the current E2 interface which allows RT RIC to communicate to DU and RIS with very low latency. It supports new service models required to enable real-time data collection and control operations from the DU to RT RIC. The E2* interface enables RT RIC to- a) access data like I/Q sampling data which is critical for low latency applications but is not available over the regular E2 interface, b) to control operations like beam forming at lower latency than that offered by the regular E2 interface. However, the tight latency constraints on the E2* interface mean that its implementation needs to be significantly faster than the regular E2 interface [12], [14].

4) U1 Interface: 6G enables newer use cases like networked robotics in smart factories. However, unlike in public networks, UEs in smart factories tend to be robots, drones, Autonomous Guided Vehicle (AGV), cars, etc. These UEs are not just communication devices with a human interface but platforms with diverse capabilities. These UEs may have their local controllers to manage these functionalities. They might run their apps (uApps) like xApps on the near RT RIC. Allowing these controllers to exchange enrichment information with near RT RIC is essential for better control decisions, joint optimizations, etc. Currently, UEs have two control interfaces- Non Access Stratum (NAS) for the core network and Radio Resource control (RRC) for the RAN. NAS is used for procedures like registration, authentication, session management, etc., while the RRC interface is used for bearer management, cell selection, power control, etc. Neither of these interfaces allows UE to near RT RIC communication. Therefore, we propose a new interface named U1.

U1 provides a channel for sharing enrichment information between UE controller and near RT RIC. This interface starts at near RT RIC and terminates at the UE. U1 messages are tunneled from UE to the CU and, from there, sent through another tunnel to Near RT RIC (using the same logical connection as E2). U1 is a RESTful interface that uses HTTP2 and JSON to transfer the enrichment information. In the future, this could also be used to share AI models between Near-RT RIC and UE [39].

U1 allows enrichment information to be provided in both directions- from UE to near RT RIC and near RT RIC to UE. We provide examples of both to highlight how U1 can enable new kinds of services in next-generation networks.

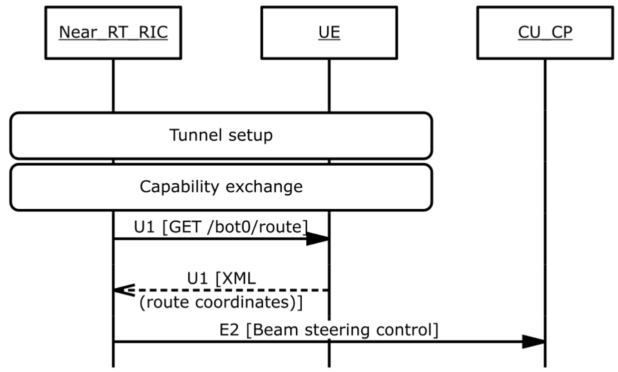

Fig. 8. Route based beam steering using U1.

a) Better coverage using enrichment information from UE: Consider a campus network deployment in a warehouse. The warehouse may have various kinds of connected robots which move around the floor to load and unload packages. These robots typically have onboard controllers which compute the route to be followed for a given task. The robot UEs are connected to the campus network. The onboard robot controllers can use the U1 interface to transfer information on the planned route to the local near RT RIC. This information can then be used as input to beam mobility management xApp [41] to create a beam steering strategy such that the robot will have reliable connectivity along the route. Figure 8 shows a high-level sequence diagram for this procedure.

b) Better throughput using enrichment information from RIC: Consider a radio link predictor xApp on Near-RT RIC. This prediction can be used to prevent radio link failures and degradation. However in scenarios where link degradation is not preventable, the RIC can use U1 to inform the robot UEs about the upcoming degradation in the radio link. The robot UEs can take appropriate actions to limit the effect of link degradation. For example, in the case of TCP traffic, congestion control algorithms can be instructed to be less aggressive in reducing congestion windows on detecting packet losses that are not due to congestion but due to radio link degradation. This would enable UE to maintain throughput levels.

5) In situ Operations, Administration, and Maintenance (IOAM): Latest improvements such as SDN and NFV have disrupted purpose-built hardware architectures and resulted in cloudification of telecommunication networks. This trend is expected to continue for beyond 5G architectures, e.g., by continuing the disaggregation of the radio access and core network as well as increasing the economies of scale and flexibility to allocate resources as needed dynamically expected to continue with.

Accordingly, the softwarization and virtualization of upcoming 6G networks, along with the increasing number of supported use cases on smart factories, requires novel In-band Network Telemetry approaches such as In situ Operations, Administration, and Maintenance (IOAM) or In-band Network

Telemetry (INT) to monitor and verify the correct operation of the network. For instance, INT enables realizing AI/ML-based self-driving mobile networks by collecting fine-grained and real-time telemetry data for use cases beyond 5G [42].

Over the years, the concept of data collection in networking equipment has evolved significantly. Unlike traditional techniques that treat the network equipment as an intermediary black box, advanced monitoring solutions have emerged, offering greater visibility into the networking equipment. Although packet-sampling-based techniques such as sFlow, Netflow, and IPFIX have been proposed in recent decades, they have struggled to keep up with the need for real-time, accurate, and fine-grained measurements. While various methods have been proposed for in-band network measurement, adopting P4-based in-band network telemetry architecture has gained significant momentum due to the advent of programmable data planes. However, at the point of wiring, P4 continues to be a technology for programmable Smart-NICs, due to the high price of programmable network switches. Accordingly, extending IOAM towards the need for 6G is one key pillar for upcoming 6G networks. The requirement of end-to-end QoS measurements and providing visibility in the network, in-band telemetry is a promising concept for O-RAN architecture to adapt 6G.

Due to the aforementioned challenges of upcoming networks, following a decentralized structure, monitoring from various vantage points is necessary. As a result, SmartNICs are quite promising for an O-RAN implementation of INT [43], [44]. In another study, the INT domain is extended to UE using a service app running on the UE [45]. Accordingly, enabling cross-layer end-to-end network telemetry is crucial for O-RAN-based 6G smart factory networks to enable mission-critical applications.

Monitoring real-time latency in current mobile networks is not feasible due to the involvement of multiple service providers such as 5G operators, edge and cloud providers, and transport network operators. Since the existing end-to-end measurement approaches are mainly active and out-of-band, obtaining real-time or fine-grained latency measurements is challenging. As a result, INT is a promising framework for bringing visibility and making fine-grained latency measurements in next-generation cellular networks [45]–[47]. Compared to traditional polling-based approaches, in-band network telemetry can provide fine-grade latency measurements that are better suited for mission-critical applications.

D. Application to Selected Use Cases

Use cases are crucial in the requirements and commercialization of 6G networks and services. They provide a holistic view of the potential applications and benefits of 6G O-RAN networks. Realistic use cases will reveal the requirements more clearly, considering the advantages of the O-RAN fundamental paradigm in the 6G smart factory networks. Guiding technological development, ecosystem collaboration, network and specification design, deployment strategies, and policy decisions to realize the full potential of next-generation wireless communication systems.

1) Beam Mobility Management: 5G+/6G next-generation smart factory networks are designed to operate in mmWave and sub-THz bands. While high data rates are achieved with the high carrier frequencies, challenges arise regarding propagation conditions. The use of narrower signal beams at higher carrier frequencies, higher path loss, and especially the signal blockage due to network area have created a need for innovative solutions. One of these solutions is beamforming applications, but these applications also contain some challenges [48]:

- Beam Management Optimization
- Inefficient Beam Sweeping
- Dense Networks and Rotation Problem
- Lack of Beam Management Response for Downlink (DL)/Uplink (UL)

There are few solutions in the literature to solve these challenges inherent to narrow beams. Recently, Deep Learning (DL)/Reinforcement Learning (RL) techniques have been applied, and valuable results have been obtained [49]. To make DL/RL applications sustainable and manageable in this field, the RAN architecture offered by O-RAN can be used. Beam management can be provided with DL/RL-Assisted Beam Mobility Management xApps to be developed on Near-RT RIC. Although simulation results are obtained in the 5G single-cell network, it would be appropriate to re-evaluate it for 6G dense, time-critical, and heterogeneous networks.

Proceeding by evaluating an application-specific use case will reveal the requirements more clearly. Factory automation environment is a very challenging application area for 5G+/6G networks. There are scenarios with high availability, reliability, and latency requirements. On the other hand, Ultra Reliable and Low Latency Communications (URLLC) traffic constraints are vital in these applications. In such a scenario, DL/RL-Assisted Beam Mobility Management xApp would be designed for synchronous and accurate data traffic between production robots, autonomous mobile robots (AMRs), and automated guided vehicles (AGVs) working in the smart factory environment. Especially considering that the real-time actions that AMRs and AGVs will take according to factory process requirements are between 1ms and 5ms under URLLC conditions, handling this traffic with the current near-RT RIC infrastructure and aforementioned xApp cannot be met in a timely fashion.

In addition to the well-known challenges of beamforming, the challenges posed by the response and processing time between near-RT RIC and Beam Mobility Management xApps may require a new design. Handling the overload on the E2 interface with more innovative RIC Controllers and interface design is possible. Collecting and processing features such as IQ Samples and RSRP/RSRQ from O-DU, O-RU, and UE takes time. Additionally, the inference time of the xApp model running on the RIC adds additional delay. In this case, our proposed RT RIC and tApps would be a solution for time-critical real-time applications. The E2 interface may be insufficient for beam redirection with real-time radio link failure information from AGVs in the smart factory environment. The reason is the lack of telemetry created by the previously mentioned service model dependency. In this case, DL/RL-Assisted Beam Mobility Management tApps/RT-RIC and DU/RIS interaction can manage this decision mechanism much more effectively over the E2* interface. On the other hand, the conflict mitigation sub-module in the near-RT RIC framework is still not fully defined in the O-RAN technical specifications. This sub-module is planned to handle internal or external conflicts arising from xApps.

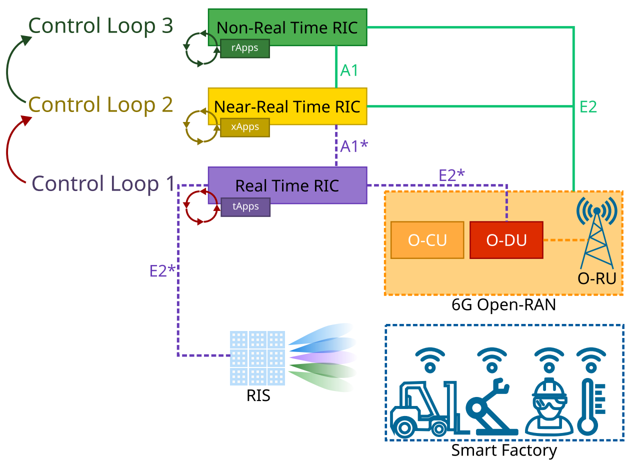

Fig. 9. Multi-Level Control Loops

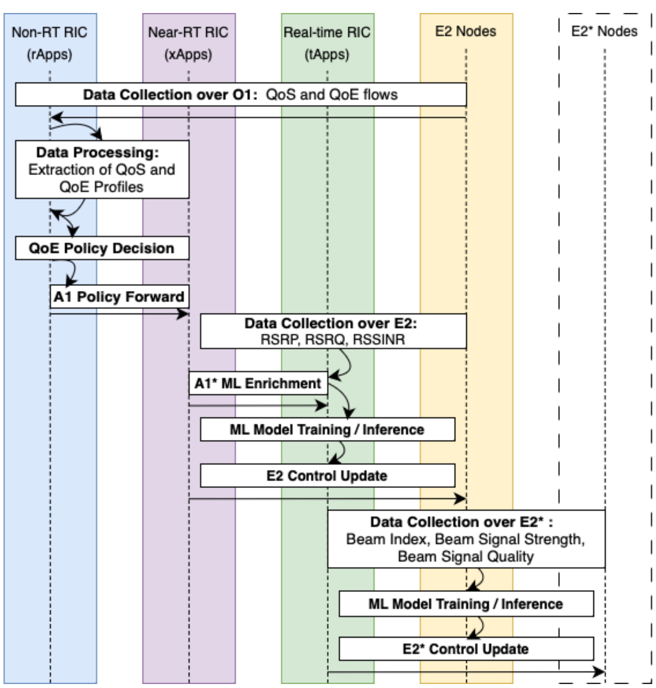

Fig. 10. Information Flow of Multi-Level Control Loops

2) Multi-level Control Loop Interactions: In our proposed use case, AGVs and AMRs operating in a time-critical smart factory environment, the importance of the RT RIC can be expressed through three synchronized multi-level control loops (Fig. 9). The necessity of multi-level control loop interaction has especially come to light in this latency-critical smart factory application. Starting with the first control loop, as shown in Figure 10, UE-specific telemetry will be obtained through DL/RL-Assisted Beam Mobility Management tApps running on RT RIC. Real-time tApps will consider Beam Index, Beam Signal Strength, Beam Signal Quality (Signal-to-Noise Ratio and Signal-to-Interference-plus-Noise Ratio), and Beam direction metrics in Beamforming. In parallel, predictive mobility management will be provided through UE mobility and handover telemetry data such as Speed, Acceleration, and Connection history. The E2* interface will benefit the O-RAN architecture at the physical layer.

Through tApps running on real-time RIC, incoming AGV or AMR localization data will be used in RIS configuration. In the second control loop, I/Q samples can be fed to the near-RT RIC via the A1* interface, and the second loop can be triggered there. Extended applications such as anomaly detection or traffic steering xApps are designed for different purposes and work on near-RT RIC to detect UEs containing anomalies and take action. However, in the current O-RAN architecture, low-latency data communication cannot be provided for these applications to operate with high reliability and precision. UE-telemetry coming through our proposed A1* interface will allow xApps to work much more efficiently. RSRP, RSRQ, RSSINR, and Physical Resource Block (PRB) data from the UE must be transmitted to anomaly detection or traffic steering xApps with high precision. Collecting this data through RT RIC will fulfill this requirement and increase the learning ability of the AI/ML models running in the second control loop during the offline training phase.

The third and last control loop is Quality of Experience (QoE) rApps running on non-RT RIC via R1 open APIs. Designing QoE rApps for 6G networks on top of the RIC requires a thorough understanding of the lower control loops, network architecture, the specific requirements of the use case, and the expected user experience. Thus, overall network optimization is achieved through three multi-level control loops.

While designing a multi-level control loop flow as shown in Figure 9, the conflict mitigation module is crucial in scenarios where proposed RT RIC, Near-RT RIC, and Non-RT RIC coexist within the network. These RIC instances may have overlapping functionalities and decision-making capabilities, leading to potential conflicts in their actions. The conflict Mitigation Module aims to resolve such conflicts and ensure coordinated and harmonious operations between the RIC instances. Since RT RIC and Near-RT RIC operate on real-time requirements and have varying levels of granularity in decision-making, coordinating their actions is crucial to avoid conflicting decisions between tApps and xApps that could degrade network performance or disrupt services.

Proposed A1*, E2*, and U1 interfaces are necessary for real-time communication and robust and efficient operation of multi-level control loops. Resource allocation management based on QoS demand and RAN Slicing competency will become more optimized through tApps, xApps, and rApps, which will communicate through the mentioned interfaces. While ensuring that offline and online training and classification are sufficient through xApps developed on Near-RT RIC, UE-Centric operations are handled with RT-RIC and tApps.

As the next stage, policy upgrades via the higher layer SMO framework will be useful for cases that are not time-sensitive.

3) RIS: As mentioned throughout this paper, RIS has the potential to overcome the prohibitively expensive deployment costs of the dense antenna placements required for mmWave and Sub-THz. They offer a cheaper and easier-to-integrate alternative that allows for extending the coverage area of RUs. In a potential example use case, as shown in Figure 11, RIS can be used in a campus network setting to cover smart factory floor areas outside the line of sight of a base station. The robots move on pre-determined paths to fulfill certain tasks, e.g., moving a box from a storage rack onto a truck bed. The factory's RIS was placed so that it could cover all areas that the RU did not already cover. When a robot moves out of the LoS of the RU, the RT-RIC is aware of this upcoming change and instructs the RU to illuminate the RIS instead. Unlike RUs, they can not gather telemetry or provide feedback to DU, CU, or RICs. Their operation will likely only consist of receiving codebook-based instructions [50]. However, since they also have to perform beam forming, they have, from a control perspective, [12] similar latency requirements as RUs. While current technologies for RIS only allow reconfiguration in the range of up to 100ms, it is clear that these times will improve in the future. Current times would likely prevent using RIS altogether in cases of high mobility. As such, it is yet to be determined whether or not RISs have to be connected to the NearRT-RIC with the existing E2 interface or to the proposed RT-RIC with the E2* interface. In either case, they will have to feature an O1 interface to allow for management and orchestration tasks, such as provisioning and updating existing codebooks.

Additionally, each RIS can only handle one UE at a time. As such, the decision must be made on which user to serve at every given time. As changes between UEs aren't as time-sensitive as beam-forming operations, the nearRT-RIC can handle this. Depending on the exact architecture, this will either have to be communicated to the RT-RIC to allow it to choose codes accordingly or handled by the nearRT-RIC itself.

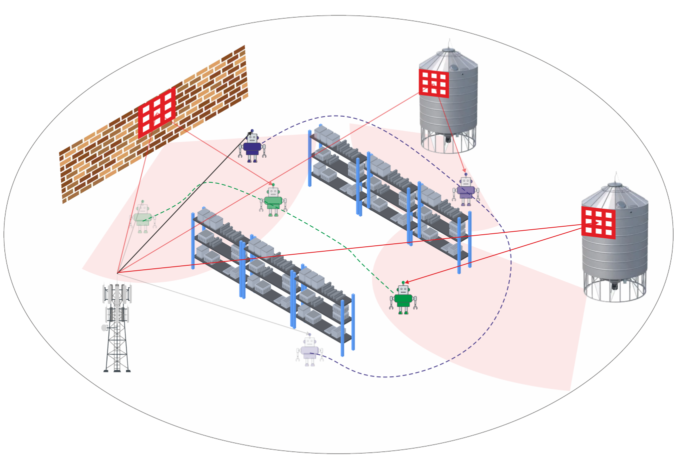

Fig. 11. Example use case for a RIS deployment in a smart factory.

This Section explored the challenges of deploying a 6G smart factory network and proposed solutions based on the O-RAN specification. However, the discussions so far have focused on a functional-level view of the network, disregarding the existing challenges of the actual physical network deployment. The optical data transport infrastructure interconnecting the CUs, DUs, and RUs, known as the X-haul, presents challenges of its own and will be explored in the next Section.

V. X-haul Transport Challenges

The O-RAN paradigm entails interoperability through open architectures and interfaces between network components, enhancing flexibility and QoS for the end users and reducing deployment costs. However, many challenges arise with the O-RAN requirements. Currently no universal optical architecture supports O-RAN, and the best optical network technology depends on the use-case specification. Furthermore, the O-RAN must provide adaptation capabilities to the emerging technologies, such as immersive extended reality (XR), holographic communication, and fully autonomous transportation, which increases the complexity of optimal network planning.

Recent research around O-RAN primarily concentrates on virtualization, network function placement, network slicing, and AI applications within network and application layers. However, many other challenges arise [51], such as the design of network management and control frameworks to support the cost-effective, expandable, and heterogeneous X-haul deployment, ensuring reliability and optimal performance. With these frameworks also comes the necessity of protocols and slicing mechanisms that are not consistently covered in the current state-of-the-art. Not less important is energy efficiency, one of the 6G's KPIs. The integration of X-haul optical architectures needs to be further explored regarding coordinated energy-efficient operation.

Another untackled challenge is the use of AI that comprehends the network architecture and its underlying optical X-haul technologies, especially referring to the complexity of resource allocation on a sub-ms scale. The network endpoints should be able to learn and make autonomous decisions based on behavior and operations [52], and AI algorithms will optimize network performance by optimal resource allocation and prediction.

The following subsections describe in more detail the different technologies that represent strong candidates to support the optical 6G X-haul networks.

A. X-haul Functional Split

The optical X-haul network will play a key role in the future successful deployments of 6G, essential in meeting the high bandwidth requirements in the long run [53]. For data rates up to 100x higher and a transmission latency 1/5x lower than that of 5G, 6G will have much more stringent performance demands towards the optical transport in terms of throughput, delay, synchronization, reliability, and flexibility, thus requiring the support of greater dynamics in network resource allocation to fulfill the multiple RAN deployment scenarios. Up to now, there is no predefined or preferred architecture for the optical transport X-haul.

Considering the multiple scenarios, use cases, and possibilities of functional splitting, a single solution encompassing all requirements is unattainable. Thus, a series of technologies are under consideration to support the X-haul beyond 5G. In this regard, the main challenges related to choosing the appropriate X-haul technology come down to the functional splits of the radio components, discussed in section II. These operator-dependent split use-case scenarios lead to the front, mid, and backhaul protocols to either traverse the optical transport network or to be local, thus defining stricter or more relaxed delay and throughput requirements on the underlying optical network technology.

There are 10 functional splitting options standardized as the next generation fronthaul interfaces (NGFI) [51], in which functional split 8 represents C-RAN architecture, and option 1 is the traditional decentralized RAN architecture. In the context of X-haul systems, the High-Layer Split (HLS) characterizes the functional splitting between CU and DU, in which data is transmitted over midhaul links using F1 interface specifications. Conversely, the Low-Layer Split (LLS) corresponds to the functional split between DU and RU, in which data transmission occurs over fronthaul links using LLS interface specifications. Considering the heterogeneous characteristics of the 6G network components and applications, the design of both HLS and LLS depends on the profile of traffic demands and channel conditions.

As an attempt to understand the impact of different functional splitting options, Bidkar et. al. [54] conduct an X-haul traffic analysis of reference scenarios as defined by 3GPP for system-level radio simulations. In their study, the authors provide a comprehensive comparison of traffic flows across interfaces F1 and options 7.2 and 7.3 of functional splitting. The experiments reveal that split 7.3 presents traffic demand approximately 1.5 times higher in the LLS than in the F1 interface (HLS). It happens due to FEC data, which introduces nearly 50% redundancy. On the other hand, split option 7.2 inserts greater traffic requirements due to the transmission of modulated frequency domain in-phase and quadrature data. This disparity between the data traffic of split 7.3 and split 7.2 becomes more pronounced at lower modulation levels, where fewer bits are encoded per symbol, resulting in traffic demand 9 times higher for split 7.2 compared to split 7.3.

B. X-haul Interfaces

Making the enabling of O-RAN a feasible task, in particular with respect to the fronthaul, requires clear definition of open interfaces for X-haul. These interfaces exhibit varying characteristics depending on the functional split that is chosen. Therefore, a so-called O-X-haul is also required to support O-RAN interfaces. Moreover, simultaneous network planning of both wireless and transport networks is required to support O-RAN architecture, for achieving optimal performance, cost efficiency, and scalability. Therefore, O-X-haul interfaces need to be reconfigurable as well as virtualizable. These abilities enable the system to support various use-cases without upgrading the massive infrastructure [51].

C. X-haul Architecture

Besides the aforementioned architectural aspects commanded by the functional splits, a series of other performance- and cost-related challenges arise. The appropriate X-haul technology for optical transport should provide enough capacity to meet the 6G KPIs, such as peak rates of 100 Gb/s–1 Tb/s, user rates of up to 1Gb/s, and latency of 10-100 μs, among others [20], along with the fixed capital expenditure and energy consumption costs, which tend to increase drastically with the number of interfaces and data rates.

The current fiber structure composing the fronthaul uses predominantly intensity-modulated direct detection (IM/DD) transceivers. Regarding direct detection, some state-of-the-art transceivers can reach data rates up to 100 Gb/s per lane, with a reach of up to 40 km [55]. A higher data rate is achievable after doubling the symbol rate. However, the chromatic dispersion interference reduces the optical signal reach to distances around 2 km, which is way shorter than what is envisioned for 6G fronthaul. For even higher data rates (more than 200 Gb/s), direct detection cannot operate properly, and in this case, the application of coherent detection is the most cost-effective solution [56]. Novel coherent multi-carrier systems are currently being developed and considered for the optical X-haul networks, offering the advantage of high-capacity and low-latency P2P and P2MP connections, simplified aggregation with passive optics, and connections between low- and high-speed transceivers [57]. Although coherent detection enables high bandwidth and longer transmission distances, it has a high implementation cost and adds latency to the network due to the extensive signal processing.

Moreover, the O-RAN Alliance provides recommendations for the X-haul technologies, including the passive, semi-active, and active WDM systems, which, however, based on operator studies, are not conducive to large-scale deployments due to the high cost of tunable optical transceivers and some of the active optical components [58].

A far more scalable and cost-efficient solution that proved its extensive adoption over decades is the passive optical network (PON), offering the advantage of high fiber utilization efficiency by means of P2MP connectivity. Coming in different flavors, such as TDM-, TWDM-, and WDM-PON, these represent potential candidates for the front-, mid-, and backhaul topologies [26], however, with their own limitations. Even with the currently under development and testing Higher Speed HS-PONs (e.g., 50G-PON) and first prototypes of 100G-PON [59], these solutions offer high capacities, but also unacceptably high latencies (in the range of ms) due to the underlying TDMA-based dynamic bandwidth allocation (DBA) [60]. The WDM-PONs solve this issue by dedicating one channel per RU/DU, however, as mentioned earlier, demanding costly tunable transceivers, thus leading to expensive scalability costs.

A viable and highly promising solution currently under development is the spatial division multiplexing-based SDM-PON. This cost-effective network architecture can address energy consumption tied to data traffic. The SDM-PON enhances optical network capacity for long-haul setups without adding complex multiple-input multiple-output (MIMO) and

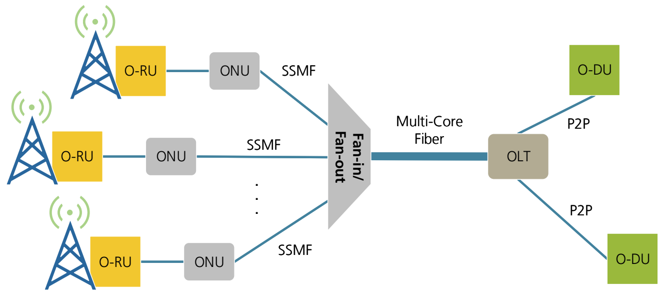

Fig. 12. Deployment example of an SDM-PON system.

digital signal processing (DSP) in access networks. A weakly coupled multicore fiber (MCF) simplifies DSP, enabling use in access networks for high data rates, like in 5G/6G X-haul setups, as shown in Figure 12. With a weakly coupled MCF, each SDM-PON core acts as a parallel P2P Ethernet link, removing the need for extensive DSP. The proposed SDM-PON architecture focuses on a cost-effective point-to-multipoint (P2MP) implementation, achieving variable data rates based on commercial Ethernet TRx availability without intricate DSP or coherent optics [61].

Integrating any of these potential optical X-haul candidates with the O-RAN architecture may require adapting and extending its interfaces and protocols to seamlessly incorporate and harness the benefits of optical transport while maintaining a synergistic overall performance and functionality of O-RAN. As such, on the fronthaul 7.2x control and user plane interfaces between O-RUs and their serving O-DU, the O-RAN Alliance defines that Ethernet encapsulation is a mandatory requirement, with IP encapsulation being optional, and applying only if the transmitting and receiving nodes support IP capabilities. In both cases, the payload is specified to be represented by one or more eCPRI transport headers with respective application data [26]. Moreover, the physical close proximity of the two O-RAN components will be required, in order to meet the ultra-low latency requirement associated with fronthaul. In the synchronization plane, the C-RAN architecture requires accurate synchronization between the O-DU and O-RUs to support Time Division Duplex (TDD) or Carrier Aggregation (CA), which in an optical-based O-RAN fronthaul will demand protocols such as PTP or Synchronous Ethernet (SyncE). When it comes to the management plane, the O-RAN 7.2x requires IP/NETCONF for network management tasks, which means that the IPv4 (optionally IPv6) has to be supported as a mandatory transport protocol [26]. Last but not least, with optical transport regularly employing redundancy for reliability and uninterrupted service, the O-RAN architecture will have to consider back-up links between O-RUs and redundant (beside the serving) O-DUs.

D. X-haul for Cell-free Architectures

One technology particularly relevant for the future 6G is expected to be the cell-free (CF), or cell-less massive MIMO (mMIMO) network architecture, offering a high spectral efficiency allowing for dense network deployments with high flexibility in resource allocation and large energy efficiency – core features native to 6G [62], [63]. Such a CF mMIMO network comprises multiple access points (APs) cooperatively serving the User Equipment (UE) units by means of coherent joint transmission and reception over the same time and frequency resources [64], [65]. The APs are connected through fronthaul links to the central processing units (CPUs) responsible for coordination of resource allocation, whilst the CPUs are interconnected using backhaul links. As investigated and proposed in [63], in O-RAN terms, the AP will be represented by the O-RU and the CPU will be equivalent to the O-DU (Fig. 13). The processing of data signals can thus be performed in a distributed fashion, i.e., locally at each AP, centrally at the CPU, or divided among APs and CPUs [63].

As a result of these deployment use-cases, the requirements towards the fronthaul and backhaul come down to which network unit (AP or CPU) performs the multi-antenna/multi-UE signal processing and channel estimation, and what level of inter-CPU coordination is required. The inter-CPU coordination is essential in the case when one or a few UEs are served by multiple APs connected through different fronthaul links to different CPUs. As a result, in the centralized CPU-based processing scenario, the APs do not require a lot of processing resources, since most of the processing, such as channel estimation, precoding/combining, and data encoding/decoding, is done in the CPU.

By contrast, in the decentralized scenario, the APs estimate locally the channels of its associated UEs, and use this information to process the data signals, whereas only the encoding and decoding of data signals is performed at the CPUs [65]. Contrary to common belief that the more the CPU is involved in the user-data processing, the higher the fronthaul requirements become [66], and hence the decentralized scenario is more optimal and preferable, it has been shown that the opposite is true [67]. With all the processing performed at the CPU level, less signaling and user data have to be carried from the AP through the fronthaul. This is because an AP, typically comprising one single antenna, will often serve two or multiple users. With the processing done at the AP, the received uplink signal, comprising a unidimensional aggregate of each UE's signals, will be split into multidimensional data (one for each UE) for pre-processing. Consequently, the amount of information to be transmitted to the CPU for further processing becomes much larger (proportional to the number of served UEs), leading to higher requirements for the fronthaul capacity. As a result, in terms of signaling, performance and fronthaul requirements, the centralized CPU-based processing is the most optimal cell-free mMIMO paradigm.

Overall, the required fronthaul capacity of the APs is directly proportional to the total number of simultaneous data streams that the AP is capable to support at maximum network load. Similarly, the required backhaul capacity of a CPU is equivalent to the sum rate of data flows supported by its APs at maximum traffic load. In order to set boundaries for these capacity requirements, the maximum number of UEs that can be served by one AP and CPU, respectively, has to be well-defined and limited [65].

A conventional distributed cell-free mMIMO deployment

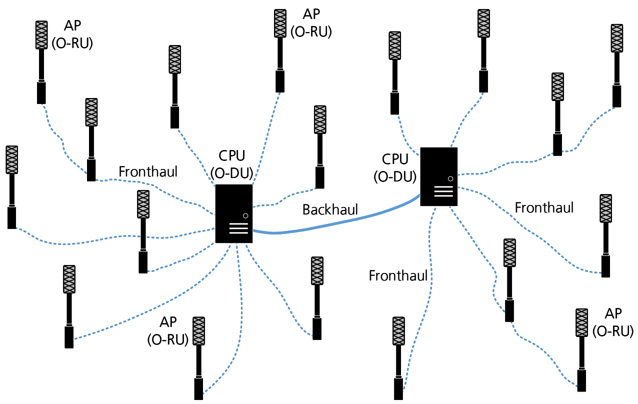

Fig. 13. Conventional cell-free mMIMO architecture with its front- and backhaul links.

requires a star topology, i.e., dedicated optical P2P fiber links between each AP and the serving CPU, which is economically unfeasible given the high number of APs, unless the deployment is done in a fiber-rich area. As a result, some of the earlier discussed P2MP optical access technologies, such as SDM-PONs, will most likely be indispensable to meet the requirements of the large deployment scales.

E. X-haul Main Challenges

To summarize, the main X-haul related-challenges can be divided into those posed by the current optical standards and interfaces to the O-RAN architecture, and vice-versa, which have to be solved in order to make the two compatible and complementary. In particular, open X-haul (O-X-haul) interfaces have to be clearly defined and established, depending on the chosen O-RAN functional split, which introduces extensively varying characteristics for these interfaces. As such, the functional split use-case scenario will determine whether the front-, mid- or backhaul protocols and traffic traverse the optical transport network or remain local, which in turn will define the latency and throughput requirements towards the underlying optical X-haul. Apart from that, integrating any optical transport X-haul technology with the O-RAN, may also require an adaptation and extension of O-RAN's native interfaces and protocols in order to seamlessly integrate the optical transport, while preserving a cohesive overall performance and functionality of the O-RAN system. For instance, the fronthaul control and user plane interfaces between O-RU and O-DU will require Ethernet encapsulation as a prerequisite, with IP encapsulation being optional, while the payload being defined by one or multiple eCPRI transport headers with the corresponding application data [26]. With SDN becoming an integral part of the control and optimization of current and future optical transport networks, APIs and protocols such as RESTful APIs or OpenFlow might be needed and adapted to facilitate communication between SDN controllers and O-RAN components. Additionally, the O-RAN 7.2x dependency on IP/NETCONF for network management tasks, as well as the lack of fully open YANG models covering the full spectrum of initialization, configuration and management functions of the O-RU and O-DU, as well as of the optical components interconnecting these, make the management interface still heavily dependent on proprietary vendor solutions.

Finally, to support the convergence of O-RAN and optical X-haul, a simultaneous network planning of both wireless and transport networks becomes essential for achieving optimal performance, scalability, as well as cost and energy efficiency. An example of such a related architectural challenge is that the employed redundancy specific to optical networks to guarantee service continuity, has to also be considered in the form of back-up connectivity between O-RUs and redundant O-DUs, when planning the wireless network architecture.

VI. Summary and Final Remarks

This paper has presented the challenges of implementing next-generation mobile networks in smart factory scenarios and discusses an extended O-RAN-based solution.

Smart factories will require the network to support all the 5G foreseen service classes (eMBB, mMTC and URLLC) and more, which will require the application of novel technologies such as mmWave and sub-THz transmission, RISs, ICAS, among others. Upon these challenges, using a flexible and extendable architecture is key, thus raising O-RAN as the de facto candidate solution to be explored. However, the current O-RAN specification lacks additional components and definitions to sustain the envisioned future smart factory requirements fully.

The proposed extensions include the addition of an RT RIC module and its companion interfaces, a novel interface between the Near RT RIC and the UE, the support for RISs, and a tighter integration between the O-RAN control and the underlying optical transport X-haul control.

The road to proposing, standardizing, and implementing an architecture supporting the future smart factory mobile networks is long, has many crossroads, and is still incipient. Nevertheless, it is a stepping stone to achieving the future 6G networks.

Acknowledgment

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification numbers: 16KISK020K, 16KISK030, and 16KISK031.

References

[1] Y. Ghildiyal, R. Singh, A. Alkhayyat, A. Gehlot, P. Malik, R. Sharma, S. V. Akram, and L. M. Alkwai, "An imperative role of 6g communication with perspective of industry 4.0: Challenges and research directions," *Sustainable Energy Technologies and Assessments*, vol. 56, p. 103047, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138823000395

[2] Y. Wu, H.-N. Dai, H. Wang, Z. Xiong, and S. Guo, "A survey of intelligent network slicing management for industrial iot: Integrated approaches for smart transportation, smart energy, and smart factory," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1175–1211, 2022.

[3] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[4] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021.

[5] M. Polese, M. Dohler, F. Dressler, M. Erol-Kantarci, R. Jana, R. Knopp, and T. Melodia, "Guest Editorial Open RAN: A New Paradigm for Open, Virtualized, Programmable, and Intelligent Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 241–244, 2 2024.

[6] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6g," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 1–30, 2022.

[7] M. Maier, A. Ebrahimzadeh, A. Beniiche, and S. Rostami, "The art of 6g (TAO 6G): how to wire society 5.0 $invited$," *J. Opt. Commun. Netw.*, vol. 14, no. 2, pp. A101–A112, Feb 2022. [Online]. Available: https://opg.optica.org/jocn/abstract.cfm?URI=jocn-14-2-A101

[8] M. Polese, M. Dohler, F. Dressler, M. Erol-Kantarci, R. Jana, R. Knopp, and T. Melodia, "Empowering the 6G Cellular Architecture with Open RAN," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 245–262, 2 2024.

[9] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[10] M. Polese, J. M. Jornet, T. Melodia, and M. Zorzi, "Toward end-to-end, full-stack 6g terahertz networks," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 48–54, 2020.

[11] M. Ben-Ari, *Principles of Concurrent and Distributed Programming*. Addison Wesley; 2nd edition, 2005.

[12] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "dapps: Distributed applications for real-time inference and control in o-ran," *IEEE Communications Magazine*, vol. 60, no. 11, pp. 52–58, 2022.

[13] W.-H. Ko, U. Ghosh, U. Dinesha, R. Wu, S. Shakkottai, and D. Bharadia, "Demo: Edgeric: Delivering realtime ran intelligence," in *Proceedings of the ACM SIGCOMM 2023 Conference*, ser. ACM SIGCOMM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 1162–1164. [Online]. Available: https://doi.org/10.1145/3603269.3610867

[14] ——, "Edgeric: Empowering realtime intelligent optimization and control in nextg networks," 2023.

[15] R. Kaliski, S.-M. Cheng, and C.-F. Hung, "Supporting 6g mission-critical services on o-ran," *IEEE Internet of Things Magazine*, vol. 6, no. 3, pp. 32–37, 2023.

[16] S. Marinova and A. Leon-Garcia, "Intelligent o-ran beyond 5g: Architecture, use cases, challenges, and opportunities," *IEEE Access*, vol. 12, pp. 27 088–27 114, 2024.

[17] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu, X. Huang, J. Hulkkonen, J. M. Jornet, M. Katz, R. Nilsson, E. Panayirci, K. Rabie, N. Rajapaksha, M. Salehi, H. Sarieddeen, T. Svensson, O. Tervo, A. Tolli, Q. Wu, and W. Xu, "White paper on broadband connectivity in 6G," University of Oulu, White Paper 6G Research Visions, No. 10, 2020, https://urn.fi/URN:ISBN:9789526226798.

[18] N. G. Alliance, "6G applications and use cases," Alliance for Telecommunications Industry Solutions, Report, 05 2022.

[19] Focus Group on Technologies for Network 2030, "Representative use cases and key network requirements for network 2030," International Telecommunication Union, Technical Report, 01 2020.

[20] W. Jiang, M. A. Habibi, and H. D. Schotten, "The road towards 6g: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[21] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6g wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[22] A. Pouttu, "6g white paper on validation and trials for verticals towards 2030's," University of Oulu, White Paper 6G Research Visions, No. 4, 2020, http://urn.fi/urn:isbn:9789526226811.

[23] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[24] Z. Wan, Z. Gao, M. Di Renzo, and L. Hanzo, "The road to industry 4.0 and beyond: A communications-, information-, and operation technology collaboration perspective," *IEEE Network*, vol. 36, no. 6, pp. 157–164, 2022.

[25] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, pp. 1376–1411, 2023.

[26] O-RAN Alliance, "Xhaul Packet Switched Architectures and Solutions," O-RAN Alliance (O-RAN), Technical Specification, 03 2022, version 03.00.

[27] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[28] O-RAN, "O-RAN A1 interface: General Aspects and Principles 3.01," O-RAN Alliance (O-RAN), Technical Specification (TS), 03 2023, version 03.01.

[29] O-RAN Alliance, "O-RAN Operations and Maintenance Architecture 4.0," O-RAN Alliance (O-RAN), Technical Specification (TS), 02 2021, version 04.00.

[30] O-RAN, "O-RAN AI/ML Workflow Description and Requirements 1.03," O-RAN Alliance (O-RAN), Technical Specification (TS), 10 2021, version 01.03.

[31] O. T. Başaran, M. Başaran, D. Turan, H. G. Bayrak, and Y. S. Sandal, "Deep Autoencoder Design for RF Anomaly Detection in 5G O-RAN Near-RT RIC via xApps," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2023, pp. 549–555.

[32] X. Song, T. Kühne, and G. Caire, "Fully-/partially-connected hybrid beamforming architectures for mmwave mu-mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1754–1769, 2020.

[33] C. Liaskos, L. Mamatas, A. Pourdamghani, A. Tsioliaridou, S. Ioannidis, A. Pitsillides, S. Schmid, and I. F. Akyildiz, "Software-defined reconfigurable intelligent surfaces: From theory to end-to-end implementation," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1466–1493, 2022.

[34] O. Alliance, "O-ran near-real-time ran intelligent controller, e2 ran control service model (e2sm-rc) 1.02.[link]," 2022.

[35] X. Foukas, B. Radunovic, M. Balkwill, and Z. Lai, "Taking 5G RAN Analytics and Control to a New Level," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. Madrid Spain: ACM, Jul. 2023, pp. 1–16. [Online]. Available: https://dl.acm.org/doi/10.1145/3570361.3592493

[36] X. Foukas, B. Radunovic, M. Balkwill, Z. Lai, and C. Settle, "Programmable RAN Platform for Flexible Real-Time Control and Telemetry," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. Madrid Spain: ACM, Oct. 2023, pp. 1–3. [Online]. Available: https://dl.acm.org/doi/10.1145/3570361.3614065

[37] I. Moerman, D. Zeghlache, A. Shahid, J. F. Santos, L. A. DaSilva, K. David, J. Farscrotu, A. De Ridder, W. Liu, and J. Hoebeke, "Mandate-driven Networking Eco-system: A Paradigm Shift in End-to-End Communications," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. Levi, Finland: IEEE, Mar. 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9083766/

[38] The Linux Foundation, "Opentelemetry." [Online]. Available: https://opentelemetry.io/

[39] 3GPP, "Group Services and System Aspects; Service requirements for the 5G system," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.261, 06 2023, version 19.3.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3107

[40] ETSI, "O-RAN Fronthaul Control, User and Synchronization Plane Specification," European Telecommunications Standards Institute, Technical Specification (TS), 09 2022, version 07.02.

[41] M. Hoffmann, "Beamforming and Open RAN," Aug. 2022. [Online]. Available: https://rimedolabs.com/blog/beamforming-and-open-ran/

[42] A. H. Ahmed and A. Elmokashfi, "Icran: Intelligent control for self-driving ran based on deep reinforcement learning," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2751–2766, 2022.

[43] Y. Feng, S. Panda, S. G. Kulkarni, K. K. Ramakrishnan, and N. Duffield, "A SmartNIC-Accelerated Monitoring Platform for In-band Network Telemetry," in *2020 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN*. Orlando, FL, USA: IEEE, Jul. 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9153279/

[44] D. Giannopoulos, G. Katsikas, K. Trantzas, D. Klonidis, C. Tranoris, S. Denazis, L. Gifre, R. Vilalta, P. Alemany, R. Muñoz, A.-M. Bosneag, A. Mozo, A. Karamchandani, L. De La Cal, D. R. López, A. Pastor, and Á. Burgaleta, "ACROSS: Automated zero-touch cross-layer provisioning framework for 5G and beyond vertical services," in *2023 Joint European Conference on Networks and*

Communications & 6G Summit (EuCNC/6G Summit). Gothenburg, Sweden: IEEE, Jun. 2023, pp. 735–740. [Online]. Available: https://ieeexplore.ieee.org/document/10188293/

[45] D. Scano, F. Paolucci, K. Kondepu, A. Sgambelluri, L. Valcarenghi, and F. Cugini, "Extending P4 in-band telemetry to user equipment for latency- and localization-aware autonomous networking with AI forecasting," *Journal of Optical Communications and Networking*, vol. 13, no. 9, p. D103, Sep. 2021. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=jocn-13-9-D103

[46] Y.-J. Wu, W.-S. Hwang, C.-Y. Shen, and Y.-Y. Chen, "Network Slicing for mMTC and URLLC Using Software-Defined Networking with P4 Switches," *Electronics*, vol. 11, no. 14, p. 2111, Jul. 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/14/2111

[47] I. Pelle, F. Paolucci, B. Sonkoly, and F. Cugini, "Fast edge-to-edge serverless migration in 5g programmable packet-optical networks," in *Optical Fiber Communication Conference*. Optical Society of America, 2021, pp. W1E–1.

[48] Y. Heng, J. G. Andrews, J. Mo, V. Va, A. Ali, B. L. Ng, and J. C. Zhang, "Six key challenges for beam management in 5.5g and 6g systems," *IEEE Communications Magazine*, vol. 59, no. 7, pp. 74–79, 2021.

[49] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for mmwave beam-management: State-of-the-art, opportunities and challenges," *IEEE Wireless Communications*, vol. 30, no. 4, pp. 108–114, 2023.

[50] J. An, C. Xu, Q. Wu, D. W. K. Ng, M. Di Renzo, C. Yuen, and L. Hanzo, "Codebook-based solutions for reconfigurable intelligent surfaces and their open challenges," *IEEE Wireless Communications*, 2022.

[51] C. Ranaweera, C. Lim, Y. Tao, S. Edirisinghe, T. Song, L. Wosinska, and A. Nirmalathas, "Design and deployment of optical x-haul for 5g, 6g, and beyond: progress and challenges," *Journal of Optical Communications and Networking*, vol. 15, no. 9, pp. D56–D66, 2023.

[52] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[53] J. S. Wey and J. Zhang, "Passive optical networks for 5g transport: Technology and standards," *Journal of Lightwave Technology*, vol. 37, no. 12, pp. 2830–2837, 2019.

[54] S. Bidkar, P. Dom, R. Bonk, and T. Pfeiffer, "Mobile xhaul traffic modelling for high-speed tdm-pon," in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4.

[55] S. T. Le, T. Drenski, A. Hills, M. King, K. Kim, Y. Matsui, and T. Sizer, "100gbps dmt asic for hybrid lte-5g mobile fronthaul networks," *Journal of Lightwave Technology*, vol. 39, no. 3, pp. 801–812, 2021.

[56] T. Sizer, D. Samardzija, H. Viswanathan, S. Thai Le, S. Bidkar, P. Dom, E. Harstead, and T. Pfeiffer, "Integrated solutions for deployment of 6g mobile networks," *Journal of Lightwave Technology*, vol. 40, no. 2, pp. 346–357, 2022.

[57] D. Welch, A. Napoli, and et.al., "Digital subcarrier multiplexing: Enabling software-configurable optical networks," *Journal of Lightwave Technology*, vol. 41, no. 4, pp. 1175–1191, 2023.

[58] O-RAN Alliance, "WDM-based Fronthaul Transport," O-RAN Alliance (O-RAN), Technical Specification (TS), 2022, version 02.00.

[59] R. Borkowski, Y. Lefevre, A. Mahadevan, D. van Veen, M. Straub, R. Kaptur, B. Czerwinski, B. Cornaglia, V. Houtsma, W. Coomans, R. Bonk, and J. Maes, "Flcs-pon—an opportunistic 100 gbit/s flexible pon prototype with probabilistic shaping and soft-input fec: operator trial and odn case studies," *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C82–C91, 2022.

[60] S. Bidkar, K. Christodoulopoulos, T. Pfeiffer, and R. Bonk, "Evaluating bandwidth efficiency and latency of scheduling schemes for 5g fronthaul over tdm-pon," *European Conference on Optical Communication*, 2022.

[61] T. Matsui, P. L. Pondillo, and K. Nakajima, "Weakly coupled multicore fiber technology, deployment, and systems," *Proceedings of the IEEE*, vol. 110, no. 11, pp. 1772–1785, 2022.

[62] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6g be?" *Nature Electronics*, vol. 3, p. 20–29, 2019.

[63] V. Ranjbar, A. Girycki, M. A. Rahman, S. Pollin, M. Moonen, and E. Vinogradov, "Cell-free mmimo support in the o-ran architecture: A phy layer perspective for 5g and beyond networks," *IEEE Communications Standards Magazine*, vol. 6, no. 1, p. 28–34, 2022.

[64] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, p. 1834–1850, 2017.

[65] E. Björnson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Transactions on Communications*, vol. 68, no. 7, p. 4247–4261, 2020.

[66] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive mimo communications," *EURASIP Journal on Wireless Communications and Networking*, no. 197, p. 1–13, 2019.

[67] E. Björnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, p. 77–90, 2020.