IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 20, 2020, accepted September 6, 2020, date of publication September 18, 2020, date of current version October 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025032

6G Architecture to Connect the Worlds

VOLKER ZIEGLER[1], **(Senior Member, IEEE)**, **HARISH VISWANATHAN**[2], **(Fellow, IEEE)**, **HANNU FLINCK**[3], **MARCO HOFFMANN**[4], **VILHO RÄISÄNEN**[3], **AND KIMMO HÄTÖNEN**[3]
[1]Nokia Bell Labs and CTO, 81541 Munich, Germany
[2]Nokia Bell Labs, Murray Hill, NJ 07974, USA
[3]Nokia Bell Labs, 02610 Espoo, Finland
[4]Nokia Bell Labs, 81541 Munich, Germany

Corresponding author: Volker Ziegler (volker.ziegler@nokia-bell-labs.com)

This work was supported by the Nokia.

ABSTRACT The post-pandemic future will offer tremendous opportunity and challenge from transformation of the human experience linking physical, digital and biological worlds: 6G should be based on a new architecture to fully realize the vision to connect the worlds. We explore several novel architecture concepts for the 6G era driven by a decomposition of the architecture into platform, functions, orchestration and specialization aspects. With 6G, we associate an open, scalable, elastic, and platform agnostic het-cloud, with converged applications and services decomposed into micro-services and serverless functions, specialized architecture for extreme attributes, as well as open service orchestration architecture. Key attributes and characteristics of the associated architectural scenarios are described. At the air-interface level, 6G is expected to encompass use of sub-Terahertz spectrum and new spectrum sharing technologies, air-interface design optimized by AI/ML techniques, integration of radio sensing with communication, and meeting extreme requirements on latency, reliability and synchronization. Fully realizing the benefits of these advances in radio technology will also call for innovations in 6G network architecture as described.

INDEX TERMS 6G, architecture, B5G, cellular communication, convergence, orchestration, sub-networks, wireless networks.

I. INTRODUCTION

2030 and beyond will offer a unique set of challenges and opportunities of global relevance and scale: We need an ambitious 6G vision for the communications architecture of the post-pandemic future to simultaneously enable growth, sustainability as well as full digital inclusion. While the COVID-19 related negative impact on consumer purchasing power for handsets and devices likely is temporary, the general and accelerated uptake across industries and sectors of digital collaboration and services as a consequence of the pandemic related shelter-in-place and home office policies will be of long-lasting effect. Going beyond the digital transformation of the 2020s, we envision a future in which human possibilities and capabilities are substantially increased and augmented by 6G technology [1]. Humans will be mesmerized by new experiences through high resolution digital virtual worlds that are indistinguishable and decoupled from their physical location. Humans will be empowered to control their automatons through these virtual worlds which in turn will drive the actuation in the physical world. Humans will be imbued with a sixth sense through numerous networked biological and physical sensors and with the network acting as a sensor and as a source of artificial intelligence. With the emergence of new devices with more intuitive interfaces, new sensing technologies, and the availability of ubiquitous distributed computing, human experience will shift from multimedia to the creation and consumption of new immersive, digital worlds. The 6G network should be architected to achieve an expansion of human experience across physical, biological and digital worlds while at the same time enabling next-generation industrial operations environment beyond Industry 4.0 in dimensions of performance such as positioning, sensing, ultra-reliability, energy efficiency and extreme real-time. Several recent papers discuss related 6G vision and technologies [2]–[6]. Note, however, that assuming the lifetime of a network generation is about ten years, 2030 may well be the beginning year of full-fledged uptake of 6G.

6G networks will provide novel radio and access architecture for both communications and sensing purposes, AI optimized wide area network and data center co-design, as well as dynamic orchestration of personalized services

The associate editor coordinating the review of this manuscript and approving it for publication was Fang Yang.

to revolutionize the long tail of niche consumer interests. While demand for mobile broadband will continue to increase for consumers and enterprise alike, uptake of ultra-reliable and low latency will be largely driven by specialized and local use cases in conjunction with non-public networks, and often with augmented intelligence. This will happen as integral part of automated and secure network transformation. Objects ranging from cars, industrial machines and appliances to watches and apparel, will learn and organize themselves to fulfill our needs by automatically adapting to our behavior, environment and business processes. Energy efficiency is another key design criterion for the design of 6G, since performance of the network will depend on the energy available in the respective architectural domains. One of the most challenging requirements comes from remote control in conjunction with augmented reality and immersive media experience. In addition to extreme ultra-reliable low latency (URLLC) performance requirement, this would demand ultra-high rates of 100Gbit/s or higher allowing uncompressed transmission of high quality 360-degree video. This will necessitate a degree of flexibility and specialization beyond 5G network capabilities. 6G networks must, therefore, be intent and open service driven and, in short, business needs will drive 6G product and service creation. Product and service creation will be an integral part of the automated e2e service workflow which is steered and guided by policy and intent. In other words, use case driven means to meet the diverse needs and preferences of each user or specialized 6G sub-network, whether human, physical machine or digital twin. In summary, the key requirements for 6G architecture include (a) network programmability; (b) deployment flexibility; (c) simplicity and efficiency; (d) security, robustness and reliability; and (e) automation.

The ongoing shift in 5G deployments towards cloud-based networks that is driven by the equipment makers and solution providers taking advantage of IT standards [7], [8] and open source technologies will further accelerate over the decade and become the foundation for the 6G era. In addition to those, the 6G era will see lead momentum from enterprises; such momentum will define sets of key requirements across architectural domains. Open-source and co-create vehicles of shared research and development such as ORAN [9] will act as a catalyst. Next generation 6G networks will not only be agnostic to the type of cloud platform by design but also focus on shared data and requirements for statelessness. While 5G includes edge cloud and virtual radio access networks (vRAN) as powerful architectural domain enabling massive scale access transformation, 6G networks will become fully edge centric and bearer-independent i.e., flow based. Future 6G networks will be programmable platforms in conjunction with the evolution of alternative compute architecture and models. The way forward is likely to include hardware (HW) accelerators to guarantee optimized performance for specific 6G algorithms with an inherent greater degree of parallelism, and higher performance and efficiency. The design of 6G architecture for deep integration into distributed cloud environments is required to meet the challenge of achieving network simplicity, flexibility and programmability.

To explore potential 6G architectural innovation, we find it convenient to decompose the architecture into four building blocks, namely, ''platform'', ''functional'', ''specialization'' and ''orchestration''. In each of these categories we discuss new technology enablers that will likely be prominent in the 6G design. While there is substantial amount of prior work on various 6G air-interface and radio technology enablers, there is relatively few in the area of 6G architecture, which is the topic of this article. In two recent articles [10], [11] on 6G architecture, the authors propose new ways for supporting better QoS and use of artificial intelligence and machine learning for network optimization. The work of [12] deals with optimal placement of 5G distributed unit (DU) in a cloud radio access network (C-RAN) architecture. Our discussion of 6G architecture extends substantially beyond optimization of function placement. In [13], a new security architecture for mobile services with implications on billing is presented. Our paper is focused on the basic network architecture and does not consider new security architecture aspects.

The paper is organized as follows. In the beginning of this article, 5G architectural baseline is defined in Section II, prior to introducing our approach to 6G architectural decomposition in Section III. Sections IV-VII dive deeper into the four main aspects of 6G architectural decomposition – platform, functional, specialization, and orchestration, respectively. We conclude with a summary in Section VIII.

II. 5G ARCHITECTURAL EVOLUTION

In the previous section, we have presented our 6G vision and 6G architecture requirements that stretch performance attributes beyond what 5G architectural evolution will be able to deliver [14]. However, 5G network design has already brought about unprecedented flexibility and transformation compared to previous generations of mobile networks. Hence, it is appropriate to revisit 5G architectural evolution as baseline before an attempt at defining 6G architecture is made. Network elements and network functions are software-controlled and transform from dedicated and specialized hardware (HW) units to mere software (SW) entities, running on standard IT HW as supported through ETSI NFV [7]. Open Networking Foundation [8] has defined the concept of software defined networking, which is supposed to complement the transformation of packet networks in dimensions of virtualization, multi-tenancy and programmability. Distributed cloud computing architecture in conjunction with the capabilities of virtualization, abstracting and sharing physical resources and the automated management and on-demand assignment of virtual resources continue to shape 5G network architectural transformation. A dedicated Network Functions Virtualization Infrastructure (NFVI) Telco Taskforce (Cloud iNfrastructure Telco Task Force, CNTT) has been launched to create an aligned NFVI Framework with common reference model and reference architecture to drive the industry and foster innovation for the evolving 5G era [15]. The advent of cognitive networks in conjunction with Machine Learning (ML) has the potential to enable the network to autonomously adapt its architecture to changing user behavior, user preferences and user experience. 5G new radio (NR) architecture will provide for extreme reliability, low latency and high capacity mobile access networks, and 5G Core will allow flexible network slicing, service-based architecture by design and flow-based optimization.

A. CLOUD NATIVE IMPLEMENTATION

One of the key architectural changes of the 5G era is the transition to a cloud native and microservice architecture; cloud native technologies empower service providers and vendors to build and operate scalable applications in dynamic cloud environments and as fostered and supported by Cloud Native Compute Foundation (CNCF) [16]. A microservice provides a dedicated business function, is independent of other microservices and is an integral part of a service-oriented architecture with published APIs and options for discovery. Modularization in conjunction with virtual machines, containers or combinations thereof will allow to achieve upgrades of distinct software modules with zero service impact and enable independent scaling of such modules. Flexible instantiation and deployment will be facilitated by the container-based approach. Furthermore, cloud-based architecture will enable pooling benefits for multiple clients and tenants of both static and dynamic nature. Another key aspect of microservice based architectural change is the transition of cloud native applications to state-efficient and, in many cases, state-less, extending the shared data layer paradigm to the RAN domain. While data and therefore context independence is a design objective, real-time constraints will need to be taken into account. Open APIs will be a key ingredient of cloud native architecture: open API regime will foster faster time-to-market and service development, access and integration for both internal and external ecosystems and in conjunction with appropriate reward schemes for the service providers. Maybe the most important dimensions of cloud native architecture affect the mode of delivery and orchestration: the transition to "DevOps" [17] paradigm will assure an agile framework for continuous delivery and integration for large scale digital production environments. To further enable ease of operability, management and large-scale deployment, automated and dynamic orchestration will be needed. This will include automatic management of virtual machines and container life cycle as well as aspects of automated monitoring and placement of functions on demand [18]. Utilizing open source tools will enhance interoperability with cloud compute platform. Infrastructure abstraction will allow flexible portability of processes and applications as well as multi-vendor environment.

B. RAN EVOLUTION

The initial deployments of 5G base stations (gNB) have largely followed the traditional model of bare metal distributed units (DU) and radio units (RU) at the cell site and with the centralized unit (CU) either at the cell site or in the central cloud. A deployment with only the CU in the cloud is called vRAN 1.0. The intense interest to expand the ecosystem of RAN vendors by separating hardware from software, the need for flexibility to meet the various requirements of different use cases, and the need for more rapid innovation in the RAN are driving several large operators towards vRAN 2.0 deployments. In vRAN 2.0, the DU is also virtualized and implemented in the edge cloud. The ORAN alliance has been created to facilitate these goals with definition of new open interfaces [9].

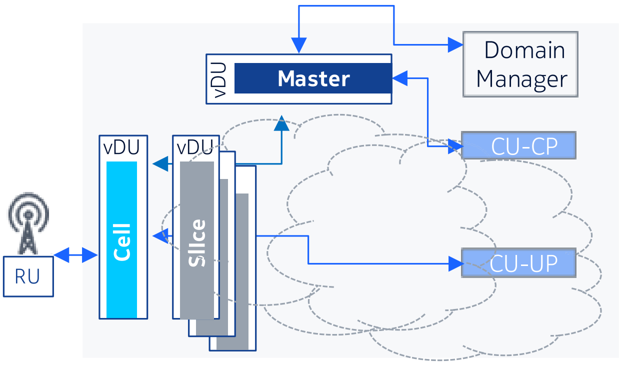

FIGURE 1. Flexible vRAN architecture: functional decomposition of the distributed unit (vDU).

As part of the 5G evolutionary journey and as shown in Figure 1, we expect in the future vRAN 3.0 deployments in which the virtualized DU implementation becomes truly cloud native and decomposed; decomposition is expected into cell, slice and master units with the flexibility to locate the different vDU slices at the cell site, far edge or edge cloud; slices can be sorted by use case categories URLLC, mMTC and eMBB. Therefore, evolved 5G vRAN will provide micro-slicing capability and tailored performance for special area and special purpose networks of great value to many verticals and industry; performance attributes such as capacity can be dynamically scaled on demand.

As described in the previous subsection, 5G architecture is increasingly cloud-native by design: Integrated and cloud native vRAN will be a key enabler of massive scale access transformation.

C. ARCHITECTURE OF THE CORE

The foundations of 5G Core architecture have been specified in 3GPP Rel 15. The 5G Core is fully service based (SBA) with new service-based interfaces (SBIs) and, therefore, decouples service consumer from producer. The 5G Core supports the following new capabilities [19]: improved session management to enable session and service continuity by the "make before break" option, which is essential for URLLC use cases; flow-based QoS framework assuring QoS on an application level; flexible end-to-end and seamless network slicing across RAN, the core and the transport network with UEs being able to simultaneously access more than one slice;

and with access agnosticism to enable seamless mobility by unified registration, authentication, session, mobility and policy management for all access types. In going forward, the 5G Core will continue to evolve through enhancements needed to serve the vertical sectors. In the subsequent 5G releases including Rel 16 and Rel 17, the core network will incorporate full support for Internet of Things (IoT), time sensitive networks (TSN) and non-public networks mainly addressing the requirements of industrial networks [20]. Capabilities for wireline wireless convergence, personal IoT and AI/ML support will be further developed in Rel 18. Incorporation of AI/ML in all subsystems will help achieve service delivery and resource consumption efficiency without compromising end user experience. It is expected that the forthcoming Rel 18 will enable wireless wireline convergence by introducing small indoor base stations with integration of residential LAN with the 5G-LAN. The exact content of Rel 18 will be decided by early 2021, but it is already clear that personal IoT networks leveraging NR-light as per 3GPP Rel 17 study item will be on the agenda; NR light will offer higher data rate, better reliability and lower latency than eMTC and NB-IoT while providing lower cost/complexity and longer battery life than NR eMBB. These anticipated enhancements will impact network exposure functions, policy control and traffic steering functions of the Core network. However, the 5G architectural foundations as laid out in Rel 15 will be maintained.

In the subsequent sections, we explore 6G network architecture beyond the scope of 5G evolution driven by the lessons learnt from the design of 5G architecture and by leveraging the latest advances in technology and research.

III. 6G ARCHITECTURAL FRAMEWORK

6G architecture encompasses building blocks across key architectural domains of a communication network, starting from the physical layer all the way up to the service layer in conjunction with a secure and automated orchestration architecture. We define and formulate architectural 6G building blocks, as illustrated in Figure 2.

6G architectural decomposition into building blocks, as made by Nokia Bell Labs, consists of four major interworking components, which provide an open and distributed reference framework. 6G architectural cloud transformation can be broadly associated with the "het-cloud" component which includes items such as open, scalable and agnostic run-time environment, data flow centricity as well as hardware acceleration, and essentially constitutes the infrastructure platform for the architecture. The "functions" component involves the functional architecture and includes the themes of RAN-CORE convergence, cell free and mesh connectivity as well as information architecture and AI. A big transformational theme of the 6G era is the emergence of specialized networks and associated performance attributes; architectural enablers of flexible off-load, extreme slicing and sub-networks are shown as part of the "specialized" building block. Of key business impact relevance is the "orchestration" component of 6G architectural change which will assure open service enabling and ecosystem play, domain resource monetization as well as cognitive closed loop and automation.

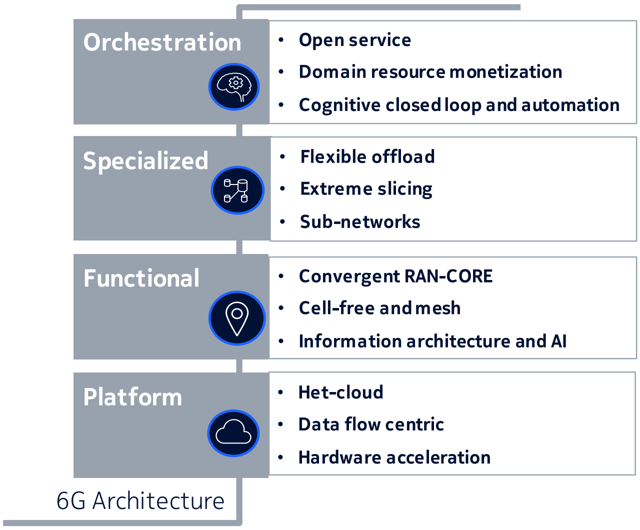

FIGURE 2. 6G architectural framework: building blocks.

IV. HET-CLOUD OF THE 6G ERA

A. OPEN, SCALABLE AND ELASTIC HET-CLOUD

Cloud transformation towards 6G will have many and heterogeneous aspects. Hence, we have coined the term "het-cloud". In the following some of these aspects relevant for 6G architectural transformation will be described. Diversification of cloud-based service delivery platforms into separate private, public, on-premise and edge clouds call for closer co-ordination of distributed computing and communication resources through federated network control and orchestration.

The het-cloud environment is a heterogenous cloud environment with multiple stakeholders to run applications at different sites such as on-prem, far edge, edge and core with a variety of different hardware and software stacks. The clouds can be private, public or hybrid. There are two main benefits of the het-cloud approach. The first one is the ease with which new services can be created, placed, subsequently scaled and moved between the clouds and the efficiency with which they can be executed. The second one is the knowledge of the cloud capabilities to optimize service performance. Such a het-cloud approach is foundational in terms of flexibility and simplicity and well in line with 6G architectural expectations. Also, such a concept will allow highest level of trustworthiness by implementing trusted execution environments (TEE). TEE will guarantee the integrity and confidentiality of both code and states; remote attestation will provide proof of trustworthiness to third-party stakeholders [21].

The unit of execution may vary from a stateless function to a micro-service and to a full service in a container or a virtual machine in the het-cloud environment. The het-cloud will encompass cloud software platform for serverless functions

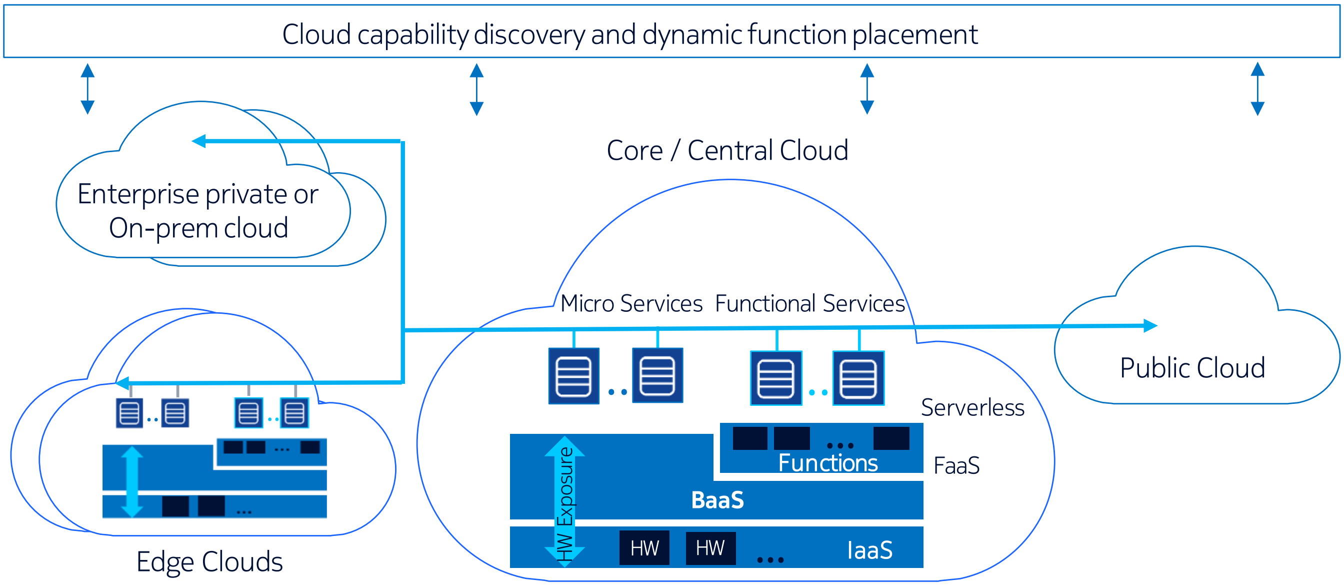

FIGURE 3. 6G het-cloud conceptual view.

in the 'FaaS' or Functions-as-a-Service layer that sits on top of the Backend-as-a-Service (BaaS). The BaaS contains the network related operations and intelligence such as data collection and analysis, logging and monitoring and distributed data storage. Similar to how a UE can offload application execution to the edge clouds, service and network function execution in the het-cloud environment can be dynamically relocated between the connected clouds making up the het-cloud.

The implication of such dynamism is that the network functions and service must be implemented by movable and extensible code. Before relocating the complete service, parts thereof, or a constituent function, the offloading entity needs to discover the capabilities of the connected clouds. For this reason, the connected clouds need to announce their capabilities to their peers. In particular, the announced cloud capability should include exposure of available hardware accelerators such as GPUs and trusted computing platform modules of TEE with remote attestation support and APIs that are integrated into the cloud platform. Other examples of such capabilities include supported computing types (such as IaaS, bare metal, PaaS, SaaS, FaaS), high availability characteristics, processing capacity, latency and type of cloud platform. We thus envision a cloud capability discovery and function placement service as an integral part of the future het-cloud as shown in Figure 3.

With the knowledge about the cloud capabilities of the peering clouds, the offloading party will prepare the suitable execution unit (e.g., stateless function(s), a container, a VM or combinations thereof) for the target environment. From the offloading point of view, the FaaS type target environment is the simplest, but to ensure any stringent performance requirement for a service, would require performance metrics to be exchanged on a per function basis. On the other extreme, when provisioning a service to a bare metal, the HW characteristics needs to be exposed for the capability discovery. Notwithstanding, the unit of execution needs to contain everything else.

A feature of pivotal importance in the het-cloud is the inter-cloud service bus that connects the service components across the het-cloud: The concept of SBA and SBI of 5G must be refined to create a flexible, on-demand inter-cloud service bus. The inter-cloud service bus will seamlessly interconnect micro-services based in containers or in virtual machines with functional services based on serverless and stateless functions. While strict real-time control services are likely to be implemented using traditional micro-services with pre-provisioned compute resource, latency tolerant control and management applications are expected to be implemented through the serverless approach to take advantage of their programming simplicity and automatic scaling capabilities.

The nature of the inter-cloud service bus varies based on the transaction types between the service and their components, e.g., stream data collection vs. real time control plane transaction and, based on their respective need for specific HW resources. In the following, the impact of a dataflow centric approach will be elaborated as well as the importance of HW acceleration. Access to HW acceleration would be needed for computationally intensive services such as AI/ML, media stream processing and security purposes. This emphasizes the need for standardized open APIs for acceleration resources.

B. DATA FLOW CENTRIC NETWORKS: ULTRA FAST DISCOVERY AND FLOW BASED ARCHITECTURE

With an anticipated disaggregation of devices and millions of 6G sub-networks, on premise and edge resources will be highly specialized using dedicated accelerators and access mechanisms with limited scalability implying stronger reliance on the off-loading of traffic. As described in the previous sub-section, on the one hand, network functions and service function chains will be assigned dynamically based on the optimal balance between consumed and available resources. On the other hand, they will be placed based on, connectivity, security, latency and performance (HW acceleration) requirements and energy consumption targets through on-line multi-object optimization algorithms [21]. Therefore, traffic routing between service endpoints needs to be based on data flow characteristics that will disassociate transactions and sessions from application context (use case, location, application), used device, used network functions, storage and transport, enabling in-network caching and replication. We expect the network to become ''cognitive'' with extensive use of AI/ML based optimization for the aforementioned traffic routing as well as for optimal placement of network and service functions. Service discovery i.e., applications, micro-services and functional services locating each other on a network, must operate at the transaction time level to match the changing context and resource allocation situation across the het-cloud. Such approach may lead to introduction of refactoring [22] and distribution of mobile network functions as described in Section V.

C. HARDWARE ACCELERATION

6G experience creation and consumption and the associated dynamic algorithmic requirements, the need for enhanced security paradigm and reduced energy consumption may lead to greatly increased compute complexity, and as a result the definition and leverage of a new computational architecture will be needed. As a first evolutionary step, greatly increased degree of flexibility can be expected from an omnipresent and smart fabric of HW accelerators such as FPGA, GPU, TPU, analog computing chips, and SoCs. Just as 5G vRAN implementations require hardware acceleration for certain layer 1 functions to meet latency and power consumption targets, new processing functions in 6G air-interface will require hardware acceleration. As an example, the use of AI/ML based receiver processing will require hardware acceleration suited for deep learning models. Computing and networking resources must interact seamlessly to ensure availability of the right computational capabilities just in-time. This would mean the creation of a whole new set of computing algorithms, models, tools and platforms, such as enabled by the het-cloud described in Sub-section A, that deal with the full range of distributed resources from full scale data centers to special in-network HW accelerators. While we shall see continued use of accelerators in use today, some new computing alternatives may arise such as analog computing using Ising models [23] to solve specific optimization problems or analog neural processing engines [24]. Accelerator enabled future cloud architecture will be a pre-requisite for 6G experience and performance attributes, enabling new services as well as device and sub-network offloading.

V. 6G FUNCTIONAL ARCHITECTURE

Key design criteria for 6G functions will include objectives of simplification, scalability, flexibility, time-to-market and error reduction. Differentiating control plane (CP) and user plane (UP) will continue to be a key design principle and CP will likely be configured as a chain of services; functional placement can be dynamically optimized and algorithmic placement of micro-services will become the norm. It can be assumed that vRAN and edge cloud will converge i.e., will share the same run-time environment for efficiency reasons. UP paradigm may evolve into pure IP flows across regional and edge data centers. Since we expect many het-cloud scenarios of the 6G era to be serverless [25], novel approaches such as sandboxing to optimize startup delays for function execution and resource footprint will be needed. Session support shall enable persistent functions for continuous data processing while hierarchical storage and messaging layers provide fast, adaptive location-aware storage access and function interactions.

A. RAN-CORE CONVERGENCE

Research into the design options of a new generation of mobile networks offers the opportunity to make the network simpler and more flexible. Network simplification is achieved primarily by harmonizing functions across different entities, eliminating duplicate functionality, reducing system state and processing to the extent possible while still meeting the requirements. In 2G and 3G, the GPRS network included four different nodes in the user plane – GGSN, SGSN, RNC and node B. This was reduced to PGW, SGW, eNode B in the user plane in 4G EPC and simply to the UPF and gNB in the 5G Core. Flexibility is achieved by allowing independent scaling and placement of different functions and ensuring the ability to quickly create new services. Several criteria can be used to optimize functional placement such as latency, security, resilience and energy efficiency. The major trends that have facilitated the flexibility have been the separation of user plane and control plane, virtualization of the Core, and more recently, cloud native implementation of the Core. As discussed in Section II, with the evolution of 5G RAN to vRAN, we see a similar transformation in the RAN with the separation of the base station CU control and user plane functions, cloud native implementation, and centralized placement. Furthermore, the service-based architecture approach of the 5G Core will extend to the RAN in the future. As the Core user plane functions move closer to the edge because of increasing traffic volume and lower latency requirement, there is an opportunity to harmonize the RAN and Core functions to create a simpler network. Figure 4 shows our vision for the converged RAN and Core architecture for 6G. We envision that the network will essentially have a 'Lower Layer Function' (LLF) entity that includes all the latency critical air-interface related RAN functions that are not included in the radio unit, and 'User Plane Micro Services (UPMS)' and 'Control Plane Micro Service (CPMS)' functional entities

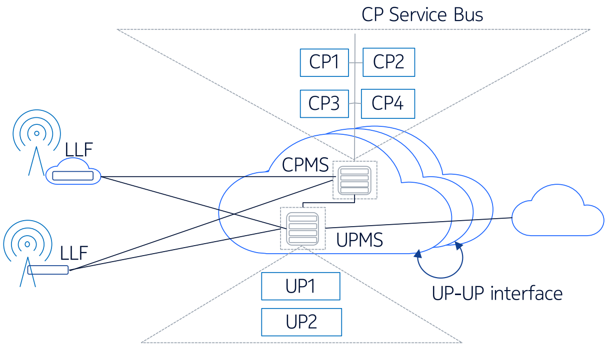

FIGURE 4. **RAN-Core convergence.**

that include all the higher layer RAN and Core capabilities as micro-services. The CPMS includes both RAN and Core services such as radio resource control, radio intelligent control (RIC), mobility management, authentication, radio resource management, etc. The UPMS includes higher layer RAN user plane as well as Core user plane services such as header compression, encryption, QoS policy enforcement and deep packet inspection. The UPMS and CPMS will be based on a framework that exposes APIs for new micro-services to be added to the core set of services that define the two functional entities. The micro services that constitute the UPMS and CPMS can be placed in the het-cloud in a disaggregated fashion. There may be a local and central instance of the UPMS serving different sets of use cases.

RAN-CORE convergence and functional optimization will enable highly specialized RAN (e.g., V2X), and slice specific RAN. Ease of introduction of new services and new types of devices with different radio capabilities and dedicated SW stacks will enhance time to market and network total cost of ownership.

B. CELL FREE AND MESH CONNECTIVITY

The deployment of non-standalone 5G architecture has firmly established the use of dual connectivity in access where each device is connected to LTE and NR cells. Driven by the need for high reliability, NR-NR dual connectivity is also likely to be deployed. In dual connectivity, a device is not associated with only a single cell, but both a master cell and a slave cell. At the same time, integrated access and backhaul (IAB) has been standardized to extend the range of the wireless connection in high bands. IAB nodes are Layer 2 nodes that simply store and forward packets from the donor node and do not maintain any UE control plane or higher layer user plane state themselves. Enabling dual connectivity for IAB nodes and end devices will result in true mesh connectivity, where a device can connect to the network through multiple routes. Mesh connections of even higher density can be achieved by extending dual to multi-connectivity at each hop. Furthermore, with cloud implementation of the 6G CP and UP functions equivalent to CU-CP and CU-UP and higher layer DU functions, devices can become cell-free with state maintained only at the CP and UP anchors in the edge cloud. Packets can be routed through essentially stateless lower layer functions creating the mesh network. Such a cell-free mesh connectivity architecture also facilitates the cell-free concept of [9], where multiple signals for different devices are transmitted simultaneously from multiple radio units to the devices creating a distributed MIMO array thereby achieving very high spectral efficiencies. 6G devices may additionally be locally connected to several proximal peer devices to form 6G sub-networks while maintaining their wide area connectivity.

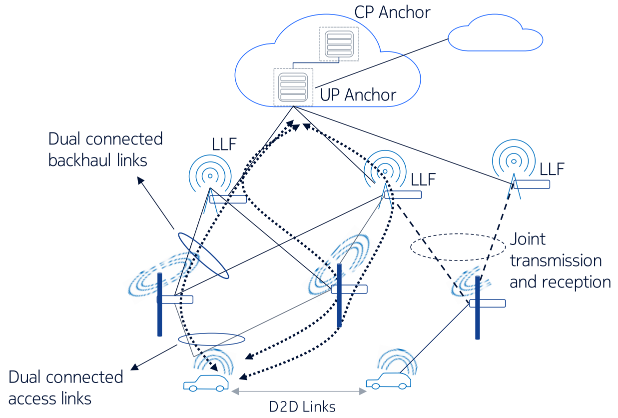

FIGURE 5. **Cell free and mesh connectivity.**

The sub-network connectivity will be required to support and provide extreme reliabilities and latencies as needed for communication and connectivity in the context of Industry 4.0 [26] as further elaborated in section VI A. In other cases, devices may be connected simultaneously to satellite and terrestrial networks, resulting in alternate data paths to the device. 6G architecture should thus be designed to natively support mesh connectivity, which can be achieved through placement of user plane and control plane anchors for the device in the cloud, disassociated with radio cells, to facilitate mobility through such a network.

C. INFORMATION ARCHITECTURE AND AI

A common denominator for key aspects of 6G architecture as described in this article is the need for dynamic reconfigurability i.e., changing placement and resourcing of services and functions while in operation. This new paradigm of flexibility will be governed and orchestrated by autonomous AI/ML based decision-making execution units, which will emerge across all layers of abstraction and cover all parts of the networks. As, at the same time, the data throughput, latency and number of connected devices and subnetworks will grow significantly, the number of reported events within networks will increase accordingly and, thus, enlarge the amount of data produced and transferred within the network. Together these developments set demanding requirements for 6G data and information (D&I) architecture. To support this development, we need to ensure and improve D&I availability within the network. Figure 6 depicts the main components of the D&I architecture for 6G. Even though some of these

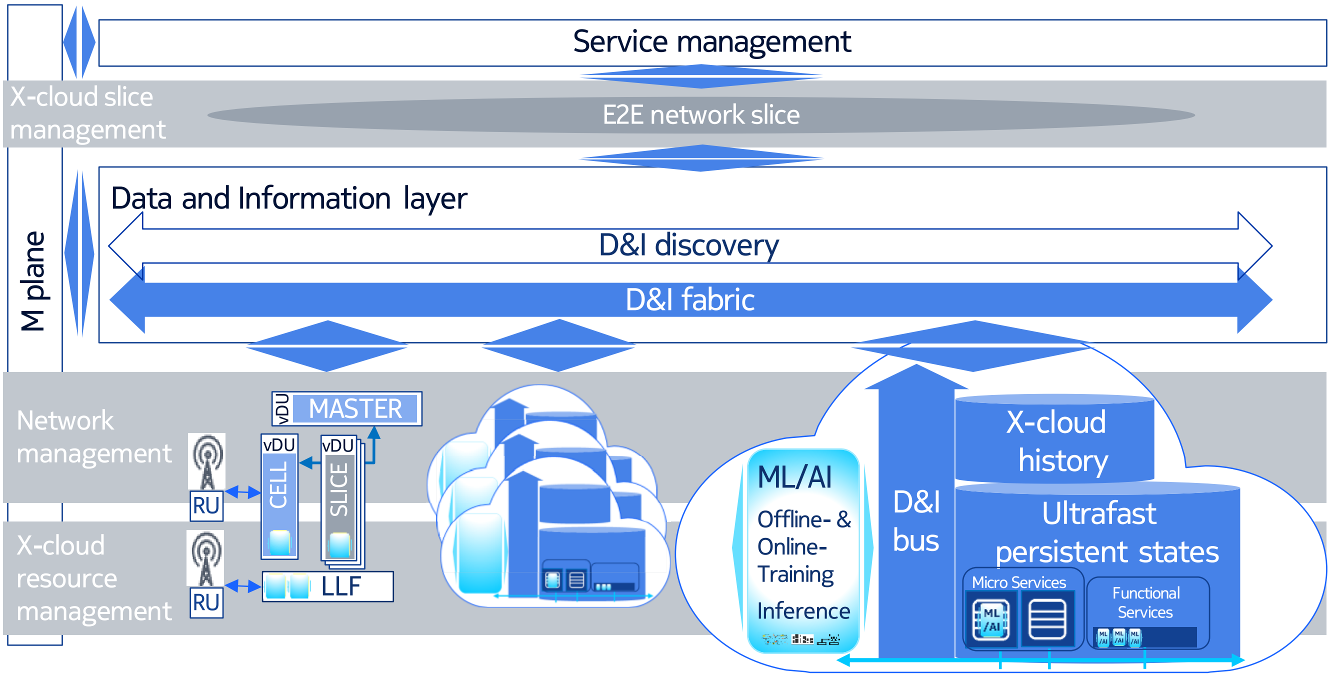

FIGURE 6. Data and information architecture.

components have been introduced already with 5G, all of them need to be enhanced to meet the new challenges.

The D&I services are distributed throughout the network to support the distributed AI/ML based decision making in the het-cloud environment. This approach minimizes the latency between generation of an event or a data point and its use in relevant analysis and inference processes. This set-up supports local analysis and decision making; there is no need to share large amounts of data continuously over the network, when the data can be analyzed or captured to the model locally within the cloud or bare metal unit, where it has been generated. This simplifies, especially, training needed in ML services. Instead of transferring a huge data set, it is possible to forward locally trained models, which can then be utilized elsewhere according to paradigms like federated or transfer learning.

To further facilitate statelessness of network functions and microservices, we suggest using ultrafast persistent storage for maintaining information about system states. This storage will store needed descriptors of active services, functions, resources and other lower level entities used to operate and ensure network performance. Active 6G network function execution units access and change these descriptors as required for function execution and, at the same time, separate management execution units responsible for monitoring and analysis of performance data produce accurate results and information towards D&I consumers.

We need to seamlessly support several data transfer paradigms from, e.g., different types of event streams to continuously pulled or pushed log files or other data bundles. This is facilitated by a D&I bus in the architecture. Through the bus, local execution units can deliver data that is requested by one or more D&I consumers. One such consumer is the ‘X-cloud history storage’, to which is stored only such content that is made available for later need by selected functions or services. At the heart of the D&I architecture is the Data and Information Layer containing the D&I fabric. It connects separated clouds, platforms and bare metal units, and provides consistent capabilities across endpoints and het-clouds. The D&I fabric capabilities include data and information visualization tools, analysis and AI/ML training/inference execution units that can be shared between services and het-cloud platforms. Thus, it simplifies and integrate 6G era D&I management. The fabric is jointly supported by D&I discovery service with catalogue and discovery functions, which enable D&I brokering and retrieval end-to-end over the het-cloud environment.

VI. 6G SPECIALIZED

A. SUB-NETWORKS CLASSIFICATION

Driven by the capability to meet ultra-reliable and low latency requirements, we are beginning to see the use of 5G in vertical industries for industrial automation. The trend will likely further expand resulting in increasing demand for 6G for verticals with application to even smaller range ‘sub-networks’ that can generally operate in a stand-alone fashion but may benefit from connectivity to the wide area network. Examples of sub-networks that will benefit from 6G performance enhancement will range from in-body sub-network, in-robot to in-car and sub-network of swarm of drones.

In Figure 7 we have introduced the scheme of 6G sub-network classification whereby sub-networks are categorized

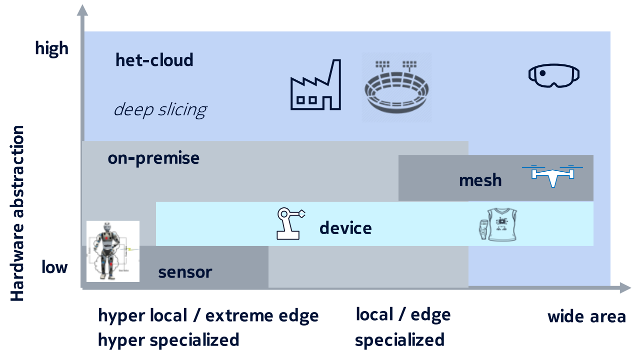

FIGURE 7. 6G sub-networks classification.

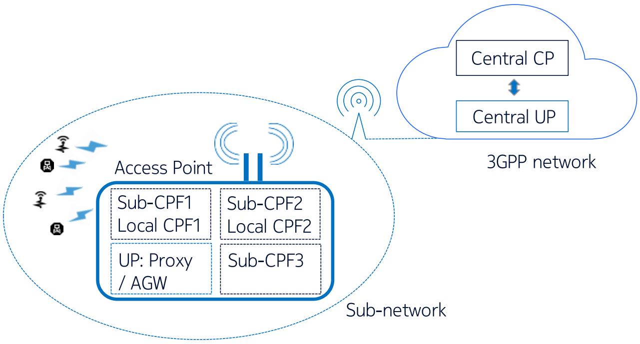

FIGURE 8. Interconnection architecture of sub-network.

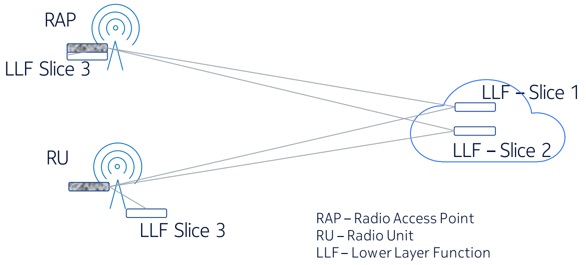

FIGURE 9. Deep slicing and fronthaul mesh.

as a function of both location of network functions within distributed het-cloud on the continuum from hyper-local (such as in-body) to wide area and as a function of the associated degree of hardware abstraction. Whereas in-body networks will require extreme reliability and latency in conjunction with very high density of sensors, swarms of drones will rather require extension of non-terrestrial networks (NTN) related requirements such as link budget optimization and ad-hoc cooperation between flying objects. Energy optimized short-range sub-networks with wireless zero-energy devices (such as passive radio frequency identification tags) will allow for a battery life of up to ten years; at the same time, also specialized wide area use cases such as for environmental monitoring may require sensor devices with very long battery lifetime. Sub-networks will be a key driving factor for 6G architectural change due to sub-networks' local topology in conjunction with the specialized performance attributes required such as extreme latency or reliability. Smart ways of assuring seamless interworking and orchestration of services across special area and wide area networks will deserve dedicated attention.

Sub-networks are defined to work in stand-alone mode as well as connected to wide area connectivity 3GPP network. Such an approach will enable the offloading in both directions as well as enhanced schemes for discovery; at the same time, scope of functionality and protocol stack complexity can be adjusted accordingly. Ultra-fast discovery in conjunction with nested networks may hold special promise for emergency and personal protection purposes: a 6G era smartphone could be imagined to seamlessly take over steering and control function of the autonomously driven car in the case of on-board unit failure.

From an architectural perspective, we expect sub-network CP and central CP functions to facilitate the interconnection of nested networks as shown in Figure 8; local control plane function will be provided by the respective sub-network. While the local CP should work within the sub-network independently of the central CP in the absence of network connectivity, the two CPs will work seamlessly and maintain necessary state transparent to the device whenever network connectivity is available. As the device moves out of the sub-network into the network, as for example, when a drone leaves the swarm, connectivity is uninterrupted. Additionally, the wide area control plane visibility of the devices in the sub-network will enable optimization of services provided to the devices within the sub-network. It should be noted that the lower layers of the wireless interface within the sub-network may be different from that of the 6G network. As appropriate, local protocols can be converted by protocol proxy or access gateway from internal to external protocols.

B. DEEP SLICING

Deep slicing will be a key driving factor for 6G architectural change in conjunction with slice-specific dedicated SW stacks and dedicated HW acceleration. Flexible composition of modular micro-services for slice specific implementations and flexible function placement depending on HW requirements (e.g., HW acceleration for video compression or in general LLF) will enable granular use case instantiation and service level assurance with minimum resource consumption and maximum energy efficiency. Dedicated HW is also motivated by complete separation of traffic flows of a slice for security purposes. Flexible traffic steering (e.g., highly secure areas) and, more broadly, new management and orchestration approaches will provide a powerful option for specialization on demand for the 6G era.

Figure 9 depicts the fronthaul mesh concept for deep slicing purposes, i.e. slicing implemented on separate hardware resources sharing the same radio unit and spectrum. Fronthaul links extend from each radio unit to multiple LLF instances; each LLF instance is connected to multiple radio units. Such a set-up enables slices for a variety of specialized use cases.

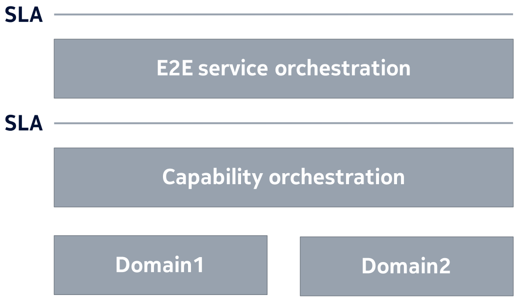

FIGURE 10. Service orchestration architecture: capability orchestration facilitates orchestration of services.

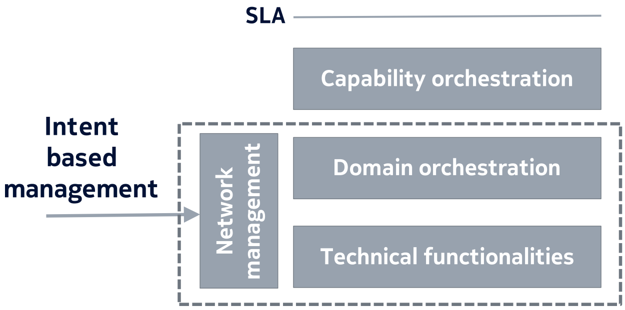

FIGURE 11. Intent-based management: technical domain functionalities to allow capabaility orchestration.

VII. MANAGEMENT AND ORCHESTRATION FOR 6G

Below, we outline a vision for 6G service and domain management supporting cross-stakeholder value networks. We start with an introduction, proceed to describe architecture and domain management, and conclude with an account of service orchestration.

A. MOTIVATION

The purpose of end-to-end orchestration is provision of services using underlying resources. 5G networks introduce network slices as a mechanism for Digital Service Providers (DSPs) to provide unified access to connectivity and computing resources. In 6G, value networks involved in service provisioning are expected to become more complex [27]. Subsequently, service-level integration of wide-area networks with dedicated networks (sub-networks, private, neutral host, etc.) as well as IT domains is expected to be a key feature of 6G. Deep slicing requires dynamic service provisioning supporting use cases and performance attributes tailored to a wide variety of contexts. In addition to configurations associated with orchestration, service management also includes service assurance and service intelligence, which need to be configured automatically in sync with orchestration.

For individual domains, management of resources according to varying business needs of relevant stakeholders needs to be supported. The target is to automate domain management so that business goals are automatically interpreted into technical goals. Cognitive Closed-Loop Automation (CLA) with software agents may be utilized to this end to automatically track the technical goals. Furthermore, management paradigms need to facilitate utilization of domain knowledge as a competitive advantage. Finally, continuous integration / continuous development (CI/CD) allows for keeping functionalities in the network up to date regardless of the development cycles involved.

Within a domain, the goal is to utilize 6G functionalities efficiently and flexibly, while also supporting the previously mentioned service and domain level goals. The interfaces used by service provisioning towards resource domains need to support 6G dynamicity. Examples include dynamic dependencies between IoT sensors, actuators and their effects on the physical and biological worlds. These may include a variety of multimodal interactions of relevance for use case families ranging from smart mobility to human health. Deep slicing (cf. Section VI B) requires microgranular hyper-dynamic orchestration, which may be employed for aggregation of domain resources at an optimal location, e.g., functional services in a serverless cloud domain.

We propose capabilities as abstractions of functionalities within one or more domains under control of a business stakeholder [27]. In the following subsections, we shall discuss architecture, domain management and service management.

B. ARCHITECTURE

The overall architecture is illustrated in Figure 8. End-to-end orchestration creates a service which is provided to customers with a (formal or implied) Service Level Agreement (SLA). Service orchestration uses capabilities provided by domain owner stakeholders, each associated with an SLA. An individual capability is created by orchestrating functionalities within one or more domains managed by the respective stakeholder.

The architecture allows for services to be composed of capabilities from different types of stakeholders such as DSPs, private networks, neutral hosts, IT domains, as well as supporting E2E orchestration by any stakeholder. It allows for controlled access to domain resources as well as autonomy of each domain from the viewpoint of management.

C. DOMAIN MANAGEMENT

The task of domain management in the architecture depicted above is twofold:

- Provide controlled access to domain functionalities via capabilities.
- Facilitate flexible management of each domain's functionalities according to business goals.

Capability orchestration takes care of combining functionalities from one or more domains in control of a stakeholder and exposing them for end-to-end service orchestration. Capability orchestration performs domain-specific actions to achieve this goal, which may involve chaining of microservice instances, for example. In order to enforce SLA, capabilities need to have associated assurance mechanisms. In a multi-domain case, functionalities can be re-allocated dynamically according to loading situation. For example, network-based application can be allocated to execution platforms based on resource utilization.

Capability orchestration may have an intent interface for business goal input; for example, domain owner can express preferences regarding resource utilization level / service quality tradeoff.

Network management is assumed to be controlled by an intent interface, where business goals are transformed into technical representation of the goal state. As described previously, software agents employing CLA (cf. Section VII A) can be employed to track goal state. A domain may have its internal orchestration.

D. SERVICE ORCHESTRATION

In view of anticipated heterogeneity of 6G service value chains, our approach has been designed to support multiple stakeholders from the ground up. An advantage of the proposed approach not yet addressed is support for third party end-to-end service orchestration. An Over-The-Top (OTT) service provider, for example, could leverage capabilities to perform service orchestration itself by selecting optimal combination of capabilities, which can be service instance specific [11]. In some cases, it might make sense to use only single E2E provider like 5G network slice, whereas in another case a more complex combination of capabilities could be optimal.

A parallel novel possibility supported by the proposal is capability marketplace where components for an end-to-end service could be selected. Continuing the previous example, the OTT would not necessarily need contractual context with capability providers but could perform dynamical selection on the marketplace for each service instance.

The capability orchestration approach allows for a flexible way of both monetizing network capacity and for composing services. The operator of a domain could make functionalities simultaneously available in combinations of different scopes, thus potentially attracting a larger customer base and utilizing network resources more effectively.

VIII. CONCLUSION

The journey to the 6G connected worlds is on, 6G architectural research has been successfully initiated. At the same time, commercial and accelerated 5G deployment in most markets worldwide is on-going or will begin soon. The architectural evolution of 5G is far from over as it will likely continue for eight more years or so. Beyond that, the road ahead is characterized by exciting opportunity from real-time communication and synchronization between the physical, digital and biological worlds to create a fundamentally new human experience. From a network point of view, physical can be broadly associated with hardware and system-on-chip architecture, digital will be all about next generation software architecture and digital twin representation and, biological will include the bio-sensors and the new human-machine interfaces, and these will be tightly woven together to achieve the flexibility, simplicity, reliability, security, efficiency and automation required to realize the variety of future applications of 6G to consumer and vertical industries. The het-cloud platform with new cloud computing capabilities serves as the foundation for the 6G network. Simplification can be achieved through convergent RAN-CORE implemented as micro services and facilitates new cell free and mesh architectures. A new data and information architecture will be an essential part of 6G taking into account the important role that data and AI/ML optimization will play in the design and operation of the 6G network. The flexibility of 6G architecture will enable specialization of the network for specific purposes such as sub-networks and optimized slices. Orchestration architecture and intent based automation and networking will be a key enabler across industries and sectors. Several other aspects such as the security architecture, specific protocol choices, charging and policy control and network exposure were not treated in this article, and are topics for future work that are necessary to define the complete 6G architecture.

REFERENCES

[1] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, 2020.

[2] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[3] R. W. Heath, Jr., "Going toward 6G [from the editor]," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 3–4, May 2019.

[4] J. Zhang, K. Kang, Y. Huang, M. Shafi, and A. F. Molisch, "Millimeter and THz wave for 5G and beyond," *China Commun.*, vol. 16, no. 2, pp. 3–4, 2019.

[5] S. Yrjola, P. Ahokangas, M. Matinmikko-Blue, R. Jurva, V. Kant, P. Karppinen, M. Kinnula, H. Koumaras, M. Rantakokko, V. Ziegler, A. Thakur, and H.-J. Zepernick, "White paper on business of 6G," 2020, *arXiv:2005.06400*. [Online]. Available: http://arxiv.org/abs/2005.06400

[6] V. Ziegler and S. Yrjola, "6G indicators of value and performance," in *Proc. 2nd 6G Wireless Summit (6G SUMMIT)*, Levi, Finland, 2020, pp. 1–5, doi: 10.1109/6GSUMMIT49458.2020.9083885.

[7] ETSI. (Oct. 2013). *Network Functions Virtualisation (NFV)*. Accessed: Sep. 14, 2020. [Online]. Available: https://portal.etsi.org/nfv/nfv_white_paper2.pdf

[8] ONF. (Jun. 2014). *SDN Architecture*. Accessed: Sep. 14, 2020. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2013/02/TR_SDN_ARCH_1.0_06062014.pdf

[9] O-RAN. (Feb. 2020). *O-RAN Use Cases and Deployment Scenarios*. Accessed: Sep. 14, 2020. [Online]. Available: https://static1.squarespace.com/static/5ad774cce74940d7115044b0/t/5e95a0a306c6ab2d1cbca4d3/1586864301196/O-RAN+Use+Cases+and+Deployment+Scenarios+Whitepaper+February+2020.pdf

[10] T. Taleb. (2020). *6G Networking, 6G Research Visions 6*. Accessed: Sep. 14, 2020. [Online]. Available: http://jultika.oulu.fi/files/isbn9789526226842.pdf

[11] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, "A survey on green 6G network: Architecture and technologies," *IEEE Access*, vol. 7, pp. 175758–175768, 2019.

[12] Y. Nakayama, R. Yasunaga, D. Hisano, and K. Maruta, "Adaptive network architecture with moving nodes towards beyond 5G era," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–7.

[13] S. Kiyomoto, A. Basu, M. S. Rahman, and S. Ruj, "On blockchain-based authorization architecture for beyond-5G mobile services," in *Proc. 12th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Cambridge, U.K., Dec. 2017, pp. 136–141.

[14] V. Ziegler, T. Wild, M. Uusitalo, H. Flinck, V. Räisänen, and K. Hätonen, "Stratification of 5G evolution and beyond 5G," in *Proc. IEEE 2nd 5G World Forum (5GWF)*, Dresden, Germany, Sep. 2019, pp. 329–334.

[15] *CNTT*. Accessed: Sep. 14, 2020. [Online]. Available: https://cntt-n.github.io/CNTT/
[16] *CNCF*. Accessed: Sep. 14, 2020. [Online]. Available: https://github.com/cncf/foundation/blob/master/charter.md
[17] L. Bass, "The software architect and DevOps," *IEEE Softw.*, vol. 35, no. 1, pp. 8–10, Jan. 2018.
[18] D.-H. Luong, H.-T. Thieu, A. Outtagarts, and Y. Ghamri-Doudane, "Predictive autoscaling orchestration for cloud-native telecom microservices," in *Proc. IEEE 5G World Forum (5GWF)*, Silicon Valley, CA, USA, Jul. 2018, pp. 153–158, doi: 10.1109/5GWF.2018.8516950.
[19] *System Architecture for the 5G System*, document TS 23.501, 3GPP, Accessed: Sep. 14, 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf
[20] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019, doi: 10.1109/ACCESS.2019.2939938.
[21] R. A. Addad, M. Bagaa, T. Taleb, D. L. C. Dutra, and H. Flinck, "Optimization model for cross-domain network slices in 5G networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 5, pp. 1156–1169, May 2020, doi: 10.1109/TMC.2019.2905599.
[22] M. Pozza, A. Rao, A. Bujari, H. Flinck, C. E. Palazzi, and S. Tarkoma, "A refactoring approach for optimizing mobile networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.
[23] J. Chou, S. Bramhavar, S. Ghosh, and W. Herzog, "Analog coupled oscillator based weighted ising machine," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 14786, doi: 10.1038/s41598-019-49699-5.
[24] A. Shafiee, A. Nag, N. Muralimanohar, R. Balasubramonian, J. P. Strachan, M. Hu, R. S. Williams, and V. Srikumar, "ISAAC: A convolutional neural network accelerator with *in-situ* analog arithmetic in crossbars," *ACM SIGARCH Comput. Archit. News*, vol. 44, no. 3, pp. 14–26, 2016.
[25] P. Aditya, I. E. Akkus, A. Beck, R. Chen, V. Hilt, I. Rimac, K. Satzke, and M. Stein, "Will serverless computing revolutionize NFV?" *Proc. IEEE*, vol. 107, no. 4, pp. 667–678, Apr. 2019.
[26] T. Wild, H. Viswanathan, and M. Uusitalo, *Evolution Beyond 5G for Vertical Industries: Key Drivers and Technology Components*, document, 6G Summit, Levi, Finland, 2019. Accessed: Sep. 14, 2020. [Online]. Available: http://www.6gsummit.com/2019/wp-content/uploads/2019/04/Day3_Session4_Viswanathan_Nokia_Bell_Labs.pdf/
[27] V. Räisänen, M. Elbamby, and D. Petrov, "Cross-stakeholder service orchestration for B5G through capability provisioning," 2020, *arXiv:2008.07162*. [Online]. Available: http://arxiv.org/abs/2008.07162
[28] C. Studer, S. Medjkouh, E. Gonultas, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47682–47698, 2018.
[29] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
[30] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: What it is, and what it is not," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Helsinki, Finland, Aug. 2015, pp. 57–64, doi: 10.1109/Trustcom.2015.357.

VOLKER ZIEGLER (Senior Member, IEEE) received the Dipl.Ing. (M.Sc.) and Dr.Ing. (Ph.D.) degrees from the Department of Electrical Engineering, Universität (TH) Karlsruhe, Germany. In his 6G leadership role with Nokia Bell Labs and CTO, the industrial research and scientific development arm of Nokia, he leads Nokia digital transformation initiatives aimed at exploring and leveraging innovations with customers and partners in the ecosystem for the 6G era. In his previous role as the Head of 5G Leadership and the Chief Architect of Nokia Mobile Networks, he played a key role in defining Nokia e2e 5G offering and positioning Nokia strongly in 5G and associated innovation, technologies, and architecture. Prior to this, he has served as the Head of Strategy roles of the company and as the Head of the North East region, helping gain market share in many key markets, including Scandinavia, Russia, and Turkey. In his more than ten year career with Siemens, he has held business unit leadership, finance, sales and marketing, services, and research and development global roles and senior positions. He has worked as an Information Technology Specialist with the World Bank/IFC in the mid-90s. He has started his career as a Research Scientist with German Aerospace Research/DLR.

HARISH VISWANATHAN (Fellow, IEEE) received the B.Tech. degree from the Department of Electrical Engineering, IIT Madras, Chennai, India, and the M.S. and Ph.D. degrees from the School of Electrical Engineering, Cornell University, Ithaca, NY, USA. Since joining Bell Labs, in October 1997, he has been working extensively on wireless research, ranging from physical layer to network architecture and protocols, including multiple antenna technology for cellular wireless networks, multihop relays, network optimization, network architecture, and the IoT communications. From 2007 to 2015, he was in the Corporation CTO Organization, where he was as a CTO Partner, he advised the Corporate CTO on technology strategy through in-depth analysis of emerging technology and market needs. He is currently the Head of the Radio Systems Research Group, Nokia Bell Labs. He has published extensively with over 100 publications. He is also a Bell Labs Fellow.

HANNU FLINCK received the M.Sc. and Lic.Tech. degrees in computer science and communication systems from the Helsinki University of Technology (Aalto University), in 1986 and 1993, respectively. He has been with the Nokia Research Center and the Technology and Innovation Unit of Nokia Networks in various research positions. He has been involved with a number of international research and university collaboration projects and he is an active participant of the IETF. He is currently the Senior Research Project Manager of Nokia Bell Labs, Espoo, Finland. His current research interests include mobile edge computing, SDN, and content delivery in mobile networks, related 3GPP SA2 topics as well as beyond 5G/6G networks.

MARCO HOFFMANN received the Dipl.Ing. (M.Sc.) and Dr.rer.nat. (Ph.D.) degrees from the Department of Computer Science, Technical University Munich, Germany. In 2004, he joined the Research and Development Department of Siemens. He is currently a Technology Expert for softwarization, cloudification, and 6G architecture, and the Senior Research Project Manager of international projects with Nokia Bell Labs. He has more than 15 years of experience in the telecommunication industry. He was the Consortium Leader and a Board Member of several national and international research projects and a member of company internal and nation-wide Future Internet strategy teams. He is the author of numerous conference papers, journal articles, patents, and patent applications.

VILHO RÄISÄNEN received the M.Sc. degree in theoretical physics from the University of Oulu, in 1992, and the Ph.D. degree in technical physics from the Technical University of Helsinki, in 1998. He is currently the Research Manager of Nokia Bell Labs, Finland, leading a team focusing on cognitive edge architectures. He is the author of two books and a Co-Editor of a third one. His research interests include AI/ML, service and network management, and the Quality of Service. Over a career of more than 20 years at Nokia, he has also experience on customer consulting, system development, and product development.

KIMMO HÄTÖNEN received the M.Sc. and Ph.D. degrees in computer science from the University of Helsinki. He is currently the Principal Scientist of Nokia Bell Labs, Finland. He has more than three decades long continuous career in research including various roles with the University of Helsinki and Nokia research organizations. His main research interest includes data science and its application to communications networks. He has authored or coauthored several tens of publications and been granted numerous patent families.

• • •