Mandate-driven Networking Eco-system: A Paradigm Shift in End-to-End Communications

Ingrid Moerman*, Djamal Zeghlache†, Adnan Shahid*, Joao F. Santos‡, Luiz A. DaSilva‡, Klaus David§,
John Farserotu¶, Ad de Ridder‖, Wei Liu*, and Jeroen Hoebeke*

* imec - IDLab, Department of Information Technology at Ghent University, Ghent, Belgium
† Télécom SudParis, Institut Mines-Télécom, Institut Polytechique de Paris, France
‡ CONNECT - Trinity College Dublin, Ireland
§ Chair for Communication Technology (ComTec), University of Kassel, Kassel, Germany
¶ Centre Suisse d'Electronique et de Microtechnique S.A.,Jaquet Droz 1, CH-2007, Neuchatel, Switzerland
‖ Hermes Partnership, Zwijndrecht, Netherlands

Abstract—The wireless industry is driven by key stakeholders that follow a holistic approach of "one-system-fits-all" that leads to moving network functionality of meeting stringent End-to-End (E2E) communication requirements towards the core and cloud infrastructures. This trend is limiting smaller and new players for bringing in new and novel solutions. For meeting these E2E requirements, tenants and end-users need to be active players for bringing their needs and innovations. Driving E2E communication not only in terms of quality of service (QoS) but also overall carbon footprint and spectrum efficiency from one specific community may lead to undesirable simplifications and a higher level of abstraction of other network segments may lead to sub-optimal operations. Based on this, the paper presents a paradigm shift that will enlarge the role of wireless innovation at academia, Small and Medium-sized Enterprises (SME)'s, industries and start-ups while taking into account decentralized mandate-driven intelligence in E2E communications.

Index Terms—End-to-end Communications, Network Orchestration, Decentralized Intelligence, Mandate-driven View.

I. INTRODUCTION

Nowadays, network and service providers, as well as telecom-equipment manufacturers, are expanding their market beyond consumers by targeting the professional market, especially industry, to generate new revenue flows [1]. Key stakeholders realize this expansion using a holistic "one-system-fits-all" philosophy for serving industry verticals, leading to a central core specification with a multitude of interoperable options, that will survive or not on the market [2]. This approach is gradually leading to very complex systems, making it very hard for tenders to select the best options, and turning impossible for new and smaller players to bring in their innovations. To satisfy the future demands of multiple tenants, e.g., private and public service providers, as well as application developers on one side; and end-users, both consumers and verticals, on the other side, we see the need for open architectures, systems and solutions. The tenants and end-users must become active players of this evolution to bring in their own requirements and innovations.

In addition, there is a clear trend of moving radio and network functionality towards the core and cloud infrastructures; this dependency on the core and the cloud may be a limiting factor for smaller deployments and dynamic environments.

E2E communication involves different network segments, each having its own characteristics, community and domain-related expertise. Driving E2E solutions from one specific community may lead to undesirable simplifications and high-level abstractions of other segments, resulting in sub-optimal network performance [3].

Based on this assessment, in this paper, we propose a paradigm shift for broadening and reestablishing the role of innovative wireless research at academia, SME's, industries and startups in E2E communications. We target the professional market for integrating private network and services providers in a new standardization and specification paradigm fostering cooperation, interfaces and openness.

II. THE PARADIGM SHIFT

Our proposed paradigm shift is driven solely by the E2E performance between communicating applications, where E2E networks and services are dynamically composed through cooperation between the most appropriate combination of network segments and network services. Additional criteria for our envisioned evolution comprises E2E system optimizations towards higher resource efficiency, in particular spectrum and energy resources, as well as higher security and improved overall cost efficiency. Our vision promotes evolved standardization processes and focuses on Expressive Application Program Interface (xAPI), expressing user and device-specific profiles, requirements, services and intentions that are both technology and Operating System (OS) agnostic. This approach leads to absolute freedom for implementation and distributed optimization, and hence, maximizes the role of academia, SME, industries and startups in the innovation cycle of E2E communication ecosystems.

Our mandate-driven E2E vision is illustrated in Figure 1. It is inspired by several limitations identified in current network designs and based on several considerations, which are detailed in the following paragraphs.

Consideration 1: *Current network designs fall short in considering the end devices as a full part of the network.*
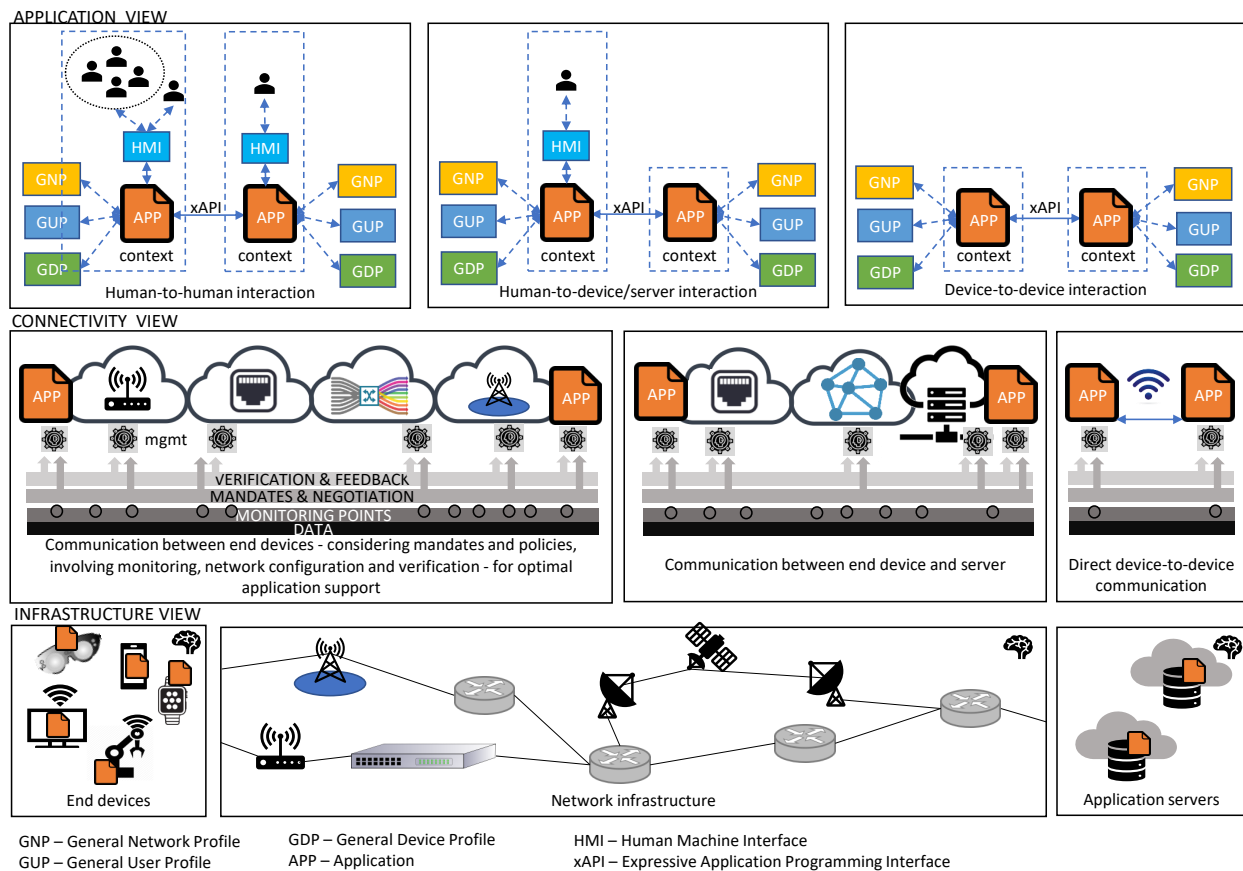
Fig. 1. Mandate-driven E2E vision.

E2E communication happens between applications on end-devices and/or application servers. Devices can be very diverse: from tiny sensor devices to high-end servers and storage devices in data centers, from traditional smartphones, laptops and desktops to more sophisticated devices, e.g., AR/VR headsets. Some devices may interact with humans through user Human-Machine Interfaces (HMIs), e.g., keyboards, touch-screens, or more sophisticated interactions via speech, gestures or signatures (face, eyes) recognised by cameras. Other devices, referred to as 'machines', directly interact with other devices without human intervention. Independent of the type of applications and end devices, they should be considered as a first-class citizen in future networks and take up a more prominent role. In order to realize this inclusion of applications in the overall service establishment, application to network interfaces and APIs should evolve into agile services themselves. These network APIs must become an integral part of future network design.

Consideration 2: *There is a gap between applications/users and the networks they are using, preventing the sharing of information about policies, capabilities, etc.*

User involvement and participation in setting policies, rules and preferences should be improved in 6th Generation (6G). Indeed, access and distribution of data should not be determined by the directives of users' terminals OSs or single companies, but by users based on a democratic consensus. Users should be able to express consent, preferences and opt-in and out with service providers and systems providing means to verify and audit the implementation of the user expressed privacy and personal data management [4]. The system must be able to dynamically create and setup new profiles based on context to drive the end-to-end service life cycle management.

Consideration 3: *Currently, there are no mechanisms for applications to communicate their fine-grained requirements to the underlying networks, and conversely, for these networks to monitor the status/health/requirements of their applications in real time*

E2E applications can be started manually by users, or triggered autonomously by events or context. When an application starts running, it should express its communication requirements, i.e., its specific QoS and traffic demands, to the underlying network in a technology-agnostic way. An application should not care by which network interface or technology it is served, as long its requirements are fulfilled and its preferences expressed in profiles are respected. Depending on the application needs and context, different types of network, e.g., wired, wireless, public, private, cellular, wide, local, proximity, and different spectrum bands, can serve

the application. Despite huge standardization efforts for the sake of interoperability, we have witnessed an explosion of new releases and technology standards. The belief that a single technology or operator can manage all today or future communications needs is unrealistic.

Consideration 4: *Networks behave as isolated systems, focusing on optimizing their own operation.*

An E2E connection between applications may be a single link i.e., in case of device-to-device communications, but it will most probably consist of multiple wired/wireless segments crossing different operator or administrative domains. E2E connections may change dynamically to keep fulfilling its requirements, while coping with changing contexts. This should happen transparently to the applications without loosing connection. Context information (location, time, weather condition, presence of other devices/persons in the environment, activity, mood) is also expressed in a generic way, which can be seen as a dynamic profile information that can be accessed to provide context-aware services to applications. Such dynamic information is also protected and secured according to the generic user/device/network profiles.

Consideration 5: *The granularity in supporting application requirements is too limited considering the diverse types of applications that will run over future networks.*

A single network is capable to serve many devices and very different applications requirements in terms of QoS requirements. The application to application communications and networking QoS should be defined in the E2E sense and should be expressed independently of the services and underlying networks that realize the E2E connection between applications. This application level QoS requirements need to be decomposed in QoS objectives for the segments that will collectively lead to the application desired QoS. Realizing the E2E QoS should go beyond typical aggregated capabilities, e.g., overall capacity and number of supported users/devices, by enabling applications to express requirements per flow throughput, different flavours of latency (in-time, on-time, cyclic), reliability, coverage, availability.

Consideration 6: *Applications do not have the ability to verify in detail whether guarantees have been met. Today's toolbox is limited to coarse SLAs and basic end-to-end measurements. Crucial data for end-to-end network management is missing.*

Future applications should have capabilities for verifying their E2E guarantees in real-time, for finding out how each involved network segment or even network component was able to deal with them. In turn, future networks must be supportive in making the required info available and consistency in the format of such information across networks should be considered. Based on the obtained insights, future applications can provide fine-grained feedback to the network, forming the basis for advanced, evidence-based network management.

III. MULTI-DISCIPLINARY RESEARCH CHALLENGES

In order to realise the E2E vision, several multi-disciplinary research challenges must be addressed and are detailed in the following subsections.

A. Profiles

Profiles are indispensable part of the E2E paradigm, which leverages communication entities including users, applications, networks, etc. to interact with other entities while respecting privacy, societal, and other related aspects. Users should have Generic User Profiles (GUPs) that keep track of personal data (age, gender, interests, contact info, subscriptions, home/family/professional/social related info) and preferences (security, privacy, consent, generic application preferences, priorities of services/applications). Construction of a user profile is as much as possible automated with minimal user interaction based on generic expressions independent from type of device, OSs, applications and networks the user interacts with and without the need for deep legal or technical knowledge from the user. This approach eliminates the duplication of profiles and preferences across different devices, OSs and applications. Moreover, each device possesses a Generic Device Profile (GDP), concerning, e.g., ownership (single person, company, community), users and Access-Control Lists (ACLs). Finally, E2E networks, and their network segments, have their own Generic Network Profiles (GNPs), containing information related to, e.g., business, security and protection policies; compliance with regulations, e.g., General Data Protection Regulation (GDPR) and limitations with respect to spectrum usage. The GNPs and preferences must be respected by the OS of every node attached to the network, as well as every application and network service, for ensuring that any data exchange is properly protected and secured, and not violating policies and regulations.

B. Social-technical Logical and Design Processes

In order to meet future objectives in privacy protection social, human and legal sciences have to be included at early design stages. These social and technical aspects are extracted from the profiles developed for the involved entities. 6G should provide privacy protection by design with the direct involvement of the end-users in the privacy protection configuration, control and management process. In essence, user intent and consent should drive the composition of privacy protection and security services irrespective of terminals and hosting infrastructures serving end-users. Actually, the same concept could be extended to all the stakeholders where not just dynamic SLAs could be established but also all security and privacy and property protection can be adjusted on situation, context and environment conditions.

C. Integration of End-users and Applications in the Network Design Process

The main focus of wireless system design today is still on increasing the capacity (mostly expressed in aggregated throughput and the number of simultaneously supported users

or devices) at the physical level. An enormous R&D effort is spent to push the physical boundaries, e.g., by increasing modulation rates (up to 4096-QAM), by introducing massive (distributed) Multiple-Input Multiple-Output (MIMO), moving to the higher frequency bands (60 GHz and beyond), extremely small cells. However, just throwing in more capacity when targeting higher throughout applications, e.g., ultra high resolution (8K) and multi-view video needed for AR/VR applications, does not solve the strict E2E reliability and latency requirements for most other consumer and professional applications that do not require such ultrahigh data rates and can also rely on less complex and less costly wireless infrastructures and lower (sub-6 GHz) spectrum bands.

Today, wireless networks are operating as black boxes towards the applications. Current QoS mechanisms, defined at the network level, e.g., Type of Service (TOS) classes in IP networks, Access Categories (AC) in Wi-Fi, or network slices in 5G based on high-level Service Level Agreements (SLAs), only offer aggregated network services and still cannot guarantee E2E QoS demands by individual applications. Until today, there exist no scalable mechanisms for an individual application to express its individual application-level QoS desires to the network. In our vision, we believe that:

- applications must become an integral part of network management to produce the appropriate E2E network configurations based on the expression of the applications' requirements (also called mandates);
- networks should become more measurable, giving applications insights into how their data is treated and performance guarantees are met;
- end-users' preferences and consent have to be reflected automatically in the way the networks are configured, controlled, and managed, so applications degrade gracefully in decreasing or limited radio coverage and battery depletion.

Radically new network protocol designs are needed that allow (i) to extend the notion of management, control and data plane separation to the application level, (ii) applications to make their mandates known to the networking stack not only in the originating node, but across the entire E2E journey of a packet to the destination node, and (iii) include measurability as an inherent part of every data transfer.

In generating the mandates, applications can take into consideration of the profiles. For instance, based on user preferences and the device capabilities, e.g., camera resolution, processing capacity, etc., and the available network information, a conference call application can compose which end-to-end treatment it requires regarding continuous throughput, latency, jitter, etc. Once active, in-band telemetry can be exploited to continuously follow-up on whether QoS is met, across different network segments [5]. In case of context changes, adaptations can be made.

Hence, applications must exchange key information to produce better network configurations and applications adaptations based on profile, preferences and context information. This assumes that users, devices, network and applications should reveal the minimal required information available at each level without compromising or conflicting with stakeholders' interests by respecting their nondisclosure requirements.

D. New Terminal Designs

End devices like user equipments (UEs) should conceptually be an integral part of the overall 6G system and offer reconfigurability as predicted in [6]. This should facilitate the following innovations:

- Bring the terminal into the design process, with more open air interfaces to the terminal/ end-device, reprogrammability, and a radio application store in which one can just download the next air interface innovation.
- Radios should be more flexible, software definable, compared with nowadays where each radio technology has a specific radio interface and transceiver design.
- Personalized, context-aware and graceful degradation of terminal (radio and associated applications), due to spectrum or any other computing resources, e.g., memory, processing or battery resources.

E. End-to-End Slice Management and Orchestration Involving Multiple Segments and Domains

Our proposed mandate-driven network ecosystem involves the combination of multiple network segments and administrative domains for realizing the E2E communication between humans and/or devices. However, each network segment and administrative domain may possess different purposes, e.g., mobile coverage, point-to-point connection, and indoor offloading; and may be subject to different constraints, e.g., resources, SLAs, and policies. As a consequence, the composition of E2E services across such diverse infrastructure components easily becomes overly complex or even intractable. Therefore, we must adopt a decomposition of E2E network management, allowing each separate network segment in every administrative domain to be independently managed by a separate specialized orchestrator, which reduces the overall complexity and simplifies the design of resource allocation directives [3].

Our paradigm shift encompasses the introduction of a hyperstrator, either centralized or distributed, for leveraging and coordinating multiple specialized orchestrators to deploy E2E services [3], as shown in Figure 2. Each specialized orchestrator can be designed by domain experts to include Artificial Intelligence (AI) and Machine Learning (ML) solutions to gain deeper insights in the operation of its respective network segment, enabling optimal configuration and resource allocation, assisted by the novel monitoring and auditing capabilities. The hyperstrator ensures a cohesive allocation of resources across network segments for guaranteeing a consistent E2E QoS for the E2E services. This functionality requires aligning network functions and services descriptors (e.g., tenant slices, service function chains, radio and network functions) proposed by various communities to ensure interoperability and cooperation of multiple orchestrators. In addition, there is the challenge
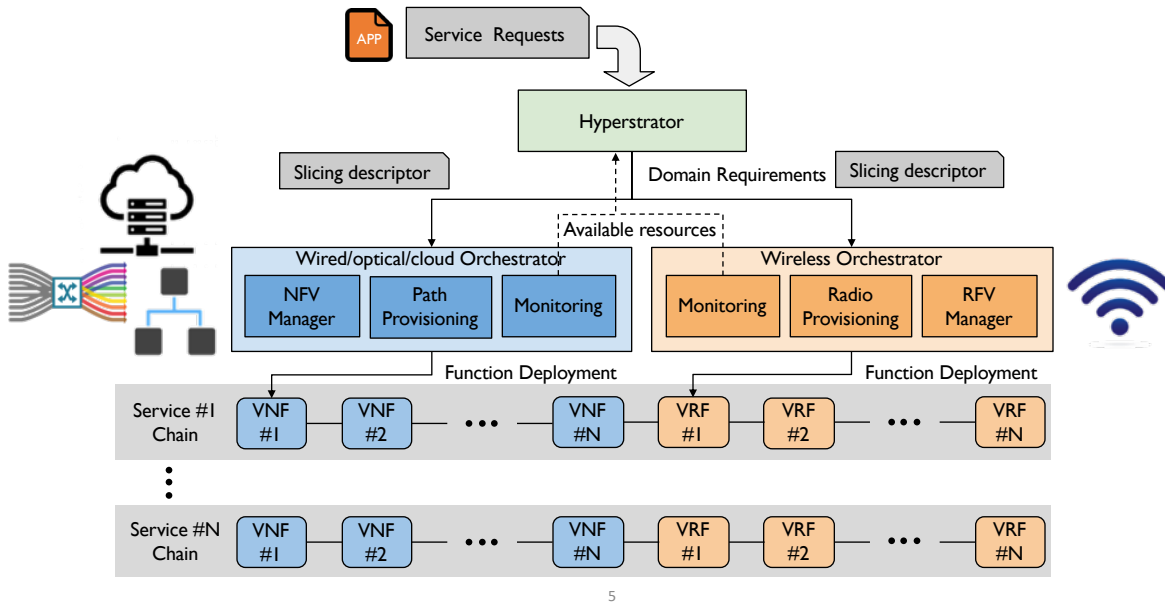
Fig. 2. Decomposition of the E2E network management using a hyperstrator and specialized orchestrators for managing different network segments.

of operating in the presence of heterogeneous interfaces be them east-west or north-south. These interfaces should become an integral part of the service composition and should be seen as service components involved in the overall process. The introduction of generic and semantically-rich interfaces for description and discovery of services, components, and resources, can enable a more agile and automated composition and deployment using knowledge-based reasoning and inference.

Realizing effective cooperation between network segments and administrative domains towards a global E2E efficiency is an interoperability and compatibility challenge, which must be specifically addressed in future open and cooperative systems. In addition, there is also the design of a language that can express the both the diverse set of assets, e.g., network resources and components, and constraints, e.g., policies and QoS requirements, while sharing the minimal amount of data. For instance, at the highest orchestration level, the hyperstrator, does not need be aware about how individual wireless resource blocks are allocated, however, it must know that available wireless link capacity, throughput and reliability. Harnessing the intelligence in the different levels, e.g., device, network segment, and E2E network is a major concern to design a cohesive, stable, reliable and resilient system.

F. Context and Purpose Aware Composition and Deployment of Services

In order to achieve maximum flexibility and efficiency, both devices and the network infrastructure must cooperate for selecting the best technology and configuration, e.g., air interfaces (Visible Light Communication (VLC), WiFi, Device-to-Device (D2D) and mobile), spectrum bands (licensed or unlicensed spectrum), and bearers (wireless or wired), and

network segment, based on the context and purpose of the deployment. The context may include relevant information beyond the GUP, e.g., location, environment and speed of the user, but also the requirements of the desired services. Based on the context, the mandate-driven networking ecosystem should infer the necessary network resources and functionality for meeting the services' requirements, and autonomously select the most appropriate technologies and configuration support these services and applications. With such capabilities integrated into the 6G network design, we can reduce or eliminate need for human interaction for configuring and optimizing the network infrastructure to support new types of services. In addition, it also fosters innovation, creating space for SMEs, industry and start-ups to develop better automation and decision tools for achieving higher efficiency and performance.

G. Machine Learning and Artificial Intelligence for Addressing System Complexity

The surge in the constant changing and dynamic relationships of various components, e.g., node, links, flows, and functions, makes it difficult to optimally configure the wireless networks. Traditionally, it is done by network experts that use their extensive knowledge of network topology, mobility, coverage, etc., in devising policies for configuring the network. In 6G, these network elements and their dynamic behaviour grow tremendously along with a huge number of network configuration options can lead to varying impact on different performance objectives. Such dynamicity and extreme level of configurability become a difficult challenge for network experts unless AI and ML are introduced to assist the experts in configuring and managing the networks in an optimal and autonomous manner. In addition, the wireless networks are

becoming highly distributed and it becomes essential to take decisions locally. This is because centralized management poses formidable complexity in the network due to increased information exchange overhead. ML or AI can be used for enabling intelligence locally and in real-time, leverages on the monitoring information from in-band telemetry. For instance, [7] uses a non-cooperative game for performing resource allocation in a decentralized way in ultra-dense small cell networks, where game theory is essential to enable AI in multi-agent environments. In such multi-agent environments, a deep reinforcement learning technique can be employed in which an agent (which can be a node, a base station, or an access point) can take network configuration decisions while meeting certain design objectives locally without having any interaction with a centralized controller. The latter can nevertheless, adjust and steer, the local decisions, though the injection of new policies, rules and eventually drive in a non compelling manner the distributed, local and autonomous agents (to enhance learning and the related exploration and exploitation processes).

Decentralized learning has marked advantages over centralized such as spectrum efficiency, privacy, low latency, high scalability, energy consumption, etc. Decentralized learning leverages to maintain the privacy of users by not sending raw information to the edge. We call this decentralized learning as *on-device learning* and it becomes feasible now because modern smart phones and sensors are equipped with high end computational units such as A11 bionic chip in IPhone X [8], Snapdragon neural processors (e.g., Snapdragon 855 [9]), and Raspberry Pi 4 [10], which paves the way towards on-device learning. In order to realize decentralized learning, a new ML paradigm has emerged named as Federated learning (FL) – a concept introduced by Google in 2016 [11]. It leverages decoupling of data processing and computation at the central point and uses on-power device capabilities for model training from locally acquired datasets. FL is in its infancy and a lot fundamental challenges need to be solved including different sizes and distribution datasets, issues in model weight sharing, scheduling, on-device constraints, communication channel impairments, etc.

IV. CONCLUSION

The growing complexity of current and future networks calls for the direct involvement of tenants and end-users in the overall design of E2E communication. Because E2E solutions driven by one specific community moving every functionality towards cloud infrastructures, when actually many decisions can be taken locally, limit overall system efficiency and slow down the introduction of innovations by small players. Based on this analysis and observation, we propose to broaden research and innovation from the current telecom-centric view with limited actors, to a E2E mandate and purpose-driven view steered by the needs of professional markets also targeting private networks involving multiple actors and academia. Further, the proposed paradigm shift is technology and actor neutral and can be implemented independently of any specific technology generation and business relationships.

REFERENCES

[1] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5g mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[2] M. Iwamura, "Ngmn view on 5g architecture," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. IEEE, 2015, pp. 1–5.

[3] J. F. Santos, J. van de Belt, W. Liu, V. Kotzsch, G. Fettweis, I. Seskar, S. Pollin, I. Moerman, L. A. DaSilva, and J. Marquez-Barja, "Orchestration next-generation services through end-to-end networks slicing," 2018.

[4] E. Mansour, A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, and T. Berners-Lee, "A demonstration of the solid platform for social web applications," in *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 223–226.

[5] J. Haxhibeqiri, I. Moerman, and J. Hoebeke, "Low overhead, fine-grained end-to-end monitoring of wireless networks using in-band telemetry," in *CNSM2019, the 15th International Conference on Network and Service Management*, 2019, pp. 1–5.

[6] "Intel and softbank beware. open source is coming to the chip business," https://www-bloomberg-com.cdn.ampproject.org/c/s/www.bloomberg.com/amp/news/articles/2020-01-22/open-source-transformed-software-the-chip-industry-is-next, accessed: 2020-02-13.

[7] A. Shahid, V. Maglogiannis, I. Ahmed, K. S. Kim, E. De Poorter, and I. Moerman, "Energy-efficient resource allocation for ultra-dense licensed and unlicensed dual-access small cell networks," *IEEE Transactions on Mobile Computing*, 2019.

[8] Apple Inc., "The future is here: iphone x," Available at https://www.apple.com/newsroom/2017/09/the-future-is-here-iphone-x/.

[9] Qualcomm, "Snapdragon 855 mobile platform," Available at https://www.qualcomm.com/products/snapdragon-855-mobile-platform.

[10] Raspberry Pi 4, Available at https://www.raspberrypi.org/products/raspberry-pi-4-model-b/.

[11] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan *et al.*, "Towards federated learning at scale: System design," *arXiv preprint arXiv:1902.01046*, 2019.

GOLD PATRONS

ERICSSON

HUAWEI

KEYSIGHT TECHNOLOGIES

NOKIA Bell Labs

ZTE

SILVER PATRONS

ROHDE & SCHWARZ

VDI Virginia Diodes, Inc.

BRONZE PATRONS

interdigital

wipro

Dual-Polarized 28-GHz Air-Filled SIW Phased Antenna Array for Next-Generation Cellular Systems

Kamil Yavuz Kapusuz, *Student Member, IEEE*, Sam Lemey, *Member, IEEE*, and Hendrik Rogier, *Senior Member, IEEE*

Abstract—A high-performance dual-polarized eight-element air-filled substrate-integrated-waveguide (AFSIW) cavity-backed patch antenna array is presented. The antenna operates in the [26.5-29.5] GHz band and provides a stable high data-rate wireless communication link between end-user terminals and access points in next-generation cellular systems. Its topology is carefully selected to maximize the performance of the array. In addition, by combining the AFSIW technology with a new antenna architecture, a low-profile, low-cost, stable, and high-performance array design is guaranteed. A prototype was fabricated and validated, demonstrating a wide active impedance bandwidth over $\pm 35^o$ scanning range and low-cross polarization level within the entire frequency band.

Index Terms—Air-filled substrate-integrated-waveguide (AF-SIW), aperture-coupled antenna, cavity-backed patch antenna, dual-polarized, 5G wireless communication, phased antenna array.

I. INTRODUCTION

THE fifth-generation (5G) wireless communication systems will improve multiple key aspects of wireless communication to support the ever increasing demand for reliable and low-latency high-data rate mobile data traffic between end-user terminals and access points [1]. To address these stringent 5G goals, the next generation of wireless communication systems will exploit the large bandwidths available at millimeter wave (mmWave) bands [2], while adopting multi-antenna systems to leverage spatial multiplexing via highly directive communication links [3] and featuring ultra-dense deployment of small-cells [4].

In such a scenario, multiple short-range base stations (BSs) will be densely deployed using large-scale multi-antenna systems to precisely focus pencil-beams to end users. This will improve the cell border coverage and enhance the spectral efficiency and frequency reuse factor of the network. In addition, these pencil beams overcome the extremely challenging propagation conditions at the mmWave bands [5]. However, to achieve a broad market acceptance, these large-scale antenna systems need to be cost-effective, non-obstructive and low-weight. Moreover, they should provide beam scanning ability. Hence, there is a stringent need for innovative phased antenna array architectures that support adaptive steering of directive beams.

In this paper, a wideband dual-polarized eight-element air-filled substrate-integrated-waveguide (AFSIW) cavity-backed patch antenna array covering the [26.5-29.5] GHz 5G band is presented. By an AFSIW implementation of the cavity, a low-loss low-cost antenna is obtained with highly efficient stable radiation characteristics and excellent antenna/platform isolation. A wideband antenna performance is achieved by exciting two hybrid modes. To exploit polarization diversity for reliable and low-latency high-data rate mobile data traffic between end-user terminals and access points, the antenna array consists of eight dual-polarized antenna elements that are rotated by 45°. A 70° scanning range for both polarizations is achieved. To the authors' best knowledge, this is the first paper that describes an innovative highly-efficient low-cost dual-polarized AFSIW phased antenna array structure capable of steering highly directive beam for 5G cellular systems.

In recent years, multiple phased antenna arrays [6]–[9] have been proposed for 5G wireless communication systems. A single-polarized 4 x 8 antenna array composed of two 4 x 4 multilayer arrays is presented in [6]. About 25° beam steering is achieved in a bandwidth of [22 - 32] GHz with a gain of approximately 19 dBi. In [7], a multi-beam antenna composed of four single-polarized 1 x 4 multilayer SIW slot antenna arrays is designed. The antenna array can produce sixteen predefined sector beams with a gain of 14.5 dBi at 28 GHz. A single-polarized 64-element phased-array formed by a 2 x 2 transmit/receive beamformer chip is presented in [8] at 28 GHz. The phased array can scan up to $\pm 50^o$ in azimuth (H-plane) and $\pm 25^o$ in elevation (E-plane). An overview of some existing multibeam antenna technologies for 5G wireless communication is presented in [9].

This paper is organized as follows. In Section II, the topology and operation principle of the stand-alone antenna element is discussed. Moreover, the manufacturing process is explained in full detail, including a thorough corner analysis. Section III covers the design and construction of the 8 x 1 phased antenna array. In Section IV, a prototype of the antenna is validated. Conclusions and an outlook on future research are presented in Section V.

II. STAND-ALONE AFSIW ANTENNA ELEMENT

A. Antenna Design Tradeoffs

Large-scale highly-directive mmWave phased antenna arrays will play a key role in next-generation cellular mobile broadband communication systems. To guarantee peak performance of such a complex antenna system, the individual antenna element needs to be designed with utmost care and a challenging set of technical and economical requirements is imposed. Obviously, it should exhibit high and stable performance over a broad bandwidth in terms of radiation efficiency, half-power beamwidth and front-to-back ratio. Next, the proposed antenna topology needs to be compact to guarantee a broad grating-lobe free scanning range. To facilitate integration with active electronics, it must be low-profile, while providing high isolation between adjacent antenna elements. Moreover, fabrication cost and complexity have to be kept to the bare minimum to pave the way for deep market penetration. These tradeoffs make the design of an efficient mmWave-band antenna system without expensive multi-layer RF laminates and high-precision fabrication a real challenge.

The proposed wideband dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna reconciles these conflicting demands in terms of performance, cost and footprint. In particular, following design requirements are imposed. The reflection coefficient with respect to 50 Ω of each port should remain below -10 dB in the entire [26.5-29.5] GHz band, while guaranteeing an intra-element port-to-port isolation larger than 25 dB. In addition, the antenna element should exhibit a cross-polarization discrimination larger than 15 dB, a front-to-back ratio larger than 15 dB and a half-power beamwidth of more than 70° over the entire frequency band of operation. Finally, a radiation efficiency of more than 80% is required.

B. Principle of Operation

The proposed wideband dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna is depicted in Fig. 1. A stacked AFSIW-based antenna technology is exploited to overcome the drawbacks of microstrip antennas (low efficiency and narrow bandwidth) while preserving the advantages of planar technology, being cost-effective fabrication, light weight and compact integration with monolithic microwave integrated components (MMICs), such as phase shifters and amplifiers. In addition, judiciously positioned rows of vias guarantee maximal suppresion of surface waves and excellent antenna/platform isolation.

The stacked AFSIW antenna is composed of three layers: the radiating layer (Layer 1), the cavity layer (Layer 2) and the feed layer (Layer 3). The thickness and dielectric properties of each substrate are carefully chosen to achieve the required impedance bandwidth and a very high efficiency, while maintaining a small footprint and keeping back radiation and cost low. To obtain a wide impedance bandwidth and high efficiency, a circular radiating patch is printed on a thin low-permittivity substrate, forming Layer 1, and an air-filled cavity is introduced underneath the patch in Layer 2. To protect the patch from detrimental environmental effects, it is printed on the bottom side of its dielectric substrate. Next, the cavity
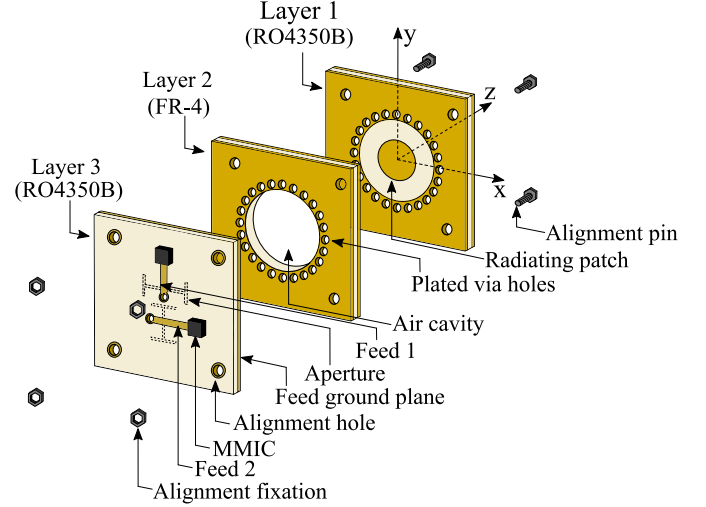

Fig. 1. Exploded view of the stand-alone AFSIW cavity-backed patch antenna element.

sidewalls are realized in Layer 2 by rows of plated via holes, at locations that are very close to the edges of the air-cavity. This avoids that the electromagnetic fields penetrate substrate Layer 2 and minimizes mutual coupling between adjacent antenna elements. A small dielectric supporting region inside the cavity is preserved to guarantee sufficient mechanical strength. The design procedure described in [10] was leveraged to minimize loss and to obtain nearly substrate-independent behavior. As a result, a low-cost substrate material may be used to implement Layer 2 without compromising antenna performance. As shown in Fig. 1, the circular cavity-backed patch is fed by microstrip aperture coupling. Two perpendicular H-shaped apertures are used to maximize bandwidth and to excite two orthogonal linear polarizations, which can be used to generate circular polarizations in future designs. Leveraging aperture coupling also guarantees shielding between the antenna and the feed network, and contributes to radiation pattern purity. Back radiation is minimized by implementing the feed network on a thin low-loss dielectric substrate (Layer 3) and exploiting short-ended microstrip lines to excite the H-shaped apertures. Tuning is accomplished by trimming the short-circuited stub of the microstrip feed line and by adjusting the dimensions of the aperture. As shown in Fig. 1, MMICs can be compactly integrated on Layer 3 at the back of the antenna element.

The stand-alone antenna element was designed in two steps to ensure dual-linear polarization over a broad bandwidth. First, a single-feed single-polarized aperture-coupled AFSIW cavity-backed patch antenna topology was conceived such that two modes are excited at two neighbouring frequencies. The mode at the lower frequency combines a weak $TM_{010,\text{ cavity}}$ mode and a strong $TE_{11,\text{ patch}}$ mode [Fig. 2(a)]. The mode at the higher frequency consists of strong $TM_{010,\text{ cavity}}$ and $TE_{11,\text{ patch}}$ modes [Fig. 2(b)]. By judiciously distributing both resonant frequencies over the frequency band of operation, the impedance bandwidth is increased significantly and the envisioned band is covered. To generate a second polarization, in the second and final step, a second feed line with H-spaped aperture is inserted by rotating the first feed line and aperture, and shifting it along the axis, as shown in Fig. 1. To provide
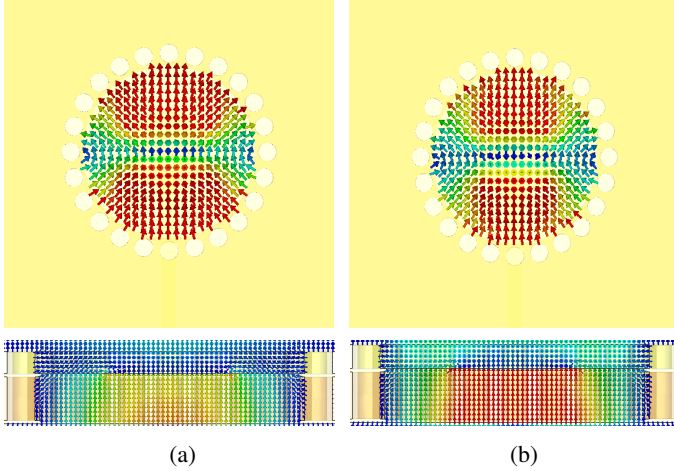
Fig. 2. Simulated hybrid mode profiles: electric field [V/m] at (a) 27.5 GHz and (b) 29.8 GHz.

the same wideband performance for this second feed at second polarization state, the same modes are excited. At both resonant frequencies, a large electric field occurs across the antenna aperture, ensuring strong radiation in the hemisphere along the positive $z-$direction and weak radiation along the negative $z-$direction.

C. Manufacturing Process

The antenna has been optimized for high-performance mass production without using expensive multi-layer/thick RF PCB laminates. Therefore, in our design, only standard single layer PCB laminates were used to realize a cost-effective, broadband and modular solution. In particular, all layers (Layers 1-3) were fabricated by means of standard single layer PCB manufacturing technology at Eurocircuits [11]. The conductive parts of each layer were realized using a standard photolithography technique. A 0.508-mm-RO4350B substrate is exploited to construct Layer 1 and Layer 3 [Fig. 1]. Layer 2 is implemented in low-cost 1-mm-FR4 substrate. The air-filled cavity is implemented by milling out a cylinder with diameter $2b$, after which plated via holes (diameter d) are inserted around this air-filled region [Fig. 3(b)]. Finally, all layers are held together using 1-mm-diameter alignment pins and fixations [Fig. 1], to guarantee sufficient electrical contact and to provide correct alignment between layers. This allows a modular design, in which each part can be replaced later on with a different version.

Prior to the manufacturing, the antenna was optimized by using the frequency domain solver of CST Microwave Studio. Utmost care was taken to ensure that all design requirements, put forward in Section II.A, are met with sufficient margin, to account for variations in the material parameters and in-accuracies of the fabrication process. The antenna dimensions resulting from an extensive optimization process are shown in Fig. 3. The simulated magnitude of the reflection coefficients [$|S_{11}|$ and $|S_{22}|$] are shown in Fig. 4(a). In the simulation, the S-parameters are normalized to 50 Ω impedance. Both exhibit two resonances distributed over the desired frequency band. The stand-alone antenna element features an impedance bandwidth of 6.5 GHz (or 23%) and 5.70 GHz (or 20%) for the vertical and
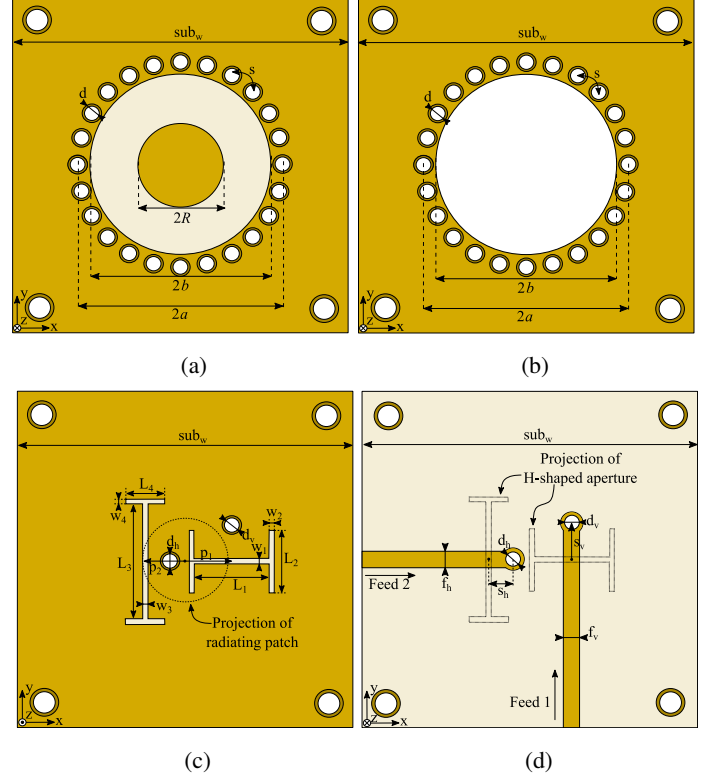

Fig. 3. Disassembled dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna. (a) Bottom view of Layer 1. (b) Bottom view Layer 2. (c) Top view of Layer 3. (d) Bottom view of Layer 3. Optimized dimensions: sub_w = 10.52 mm, a = 3.2 mm, b = 2.8 mm, R = 1.31 mm, L_1 = 2.3 mm, L_2 = 1.9 mm, L_3 = 3.4 mm, L_4 = 1.2 mm, w_1 = w_2 = 0.15 mm, w_3 = w_4 = 0.2 mm, p_1 = 1.3 mm, p_2 = 1.15 mm, f_v = f_h = 0.5 mm, s_v = 1.2 mm, s_h = 0.65 mm, d_v = d_h = 0.5 mm, d = 0.6 mm, s = 0.84 mm, $h_1 = h_3$ = 0.508 mm, h_2 = 1 mm. Substrate parameters at 28.0 GHz: RO4350B (ϵ_r = 3.66, tanδ = 0.0037), FR4 (ϵ_r = 4.3, tanδ = 0.024).

horizontal polarization, respectively. Moreover, over the entire frequency band of operation, an isolation larger than 25 dB is guaranteed between both feeds, as shown in Fig. 4(a).

Fig. 4(b) shows the maximum gain of the antenna in the hemisphere along the positive and negative z-axis for each polarization. On average, for both polarizations the gain of the antenna equals 7 dBi and the front-to-back ratio is higher than 17 dB over the frequency band of interest.

As shown in Fig. 4(c), the total antenna efficiency of the ultra-wideband dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna is larger than 85% over the entire impedance bandwidth, with a maximum of 96%.

D. Analysis of Antenna Parameters

One of the key challenges of antenna design at high frequencies is coping with the manufacturing accuracy and tolerances, such as misalignment between layers and variations in antenna dimensions. Manufacturing errors need to be carefully considered from the early design stages to obtain a robust design. Hence, a systematic analysis of the antenna parameters was performed to understand the effects of manufacturing tolerances that are considered relevant for the topology.
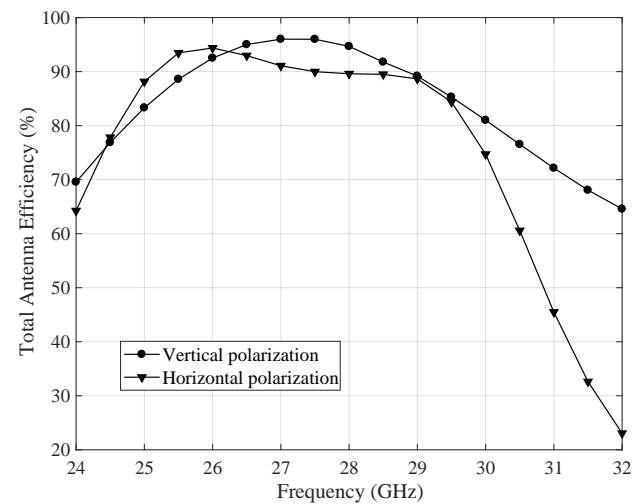
In particular, several parameter sweeps were carried out to assess the influence of etching tolerances, misalignment, and
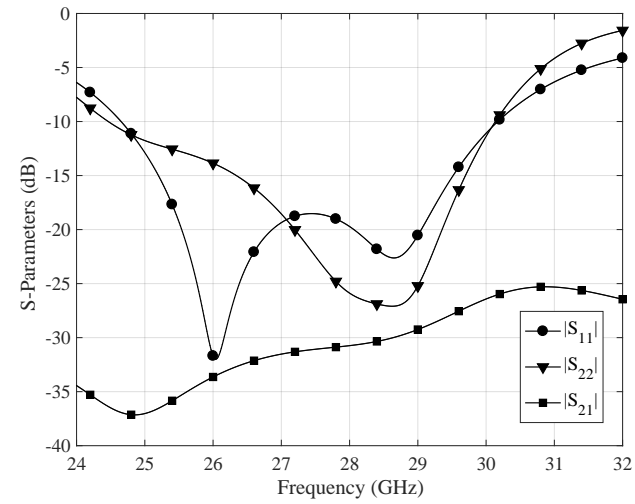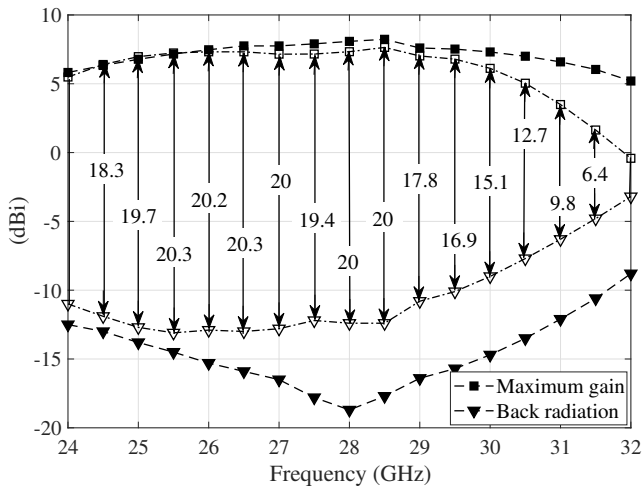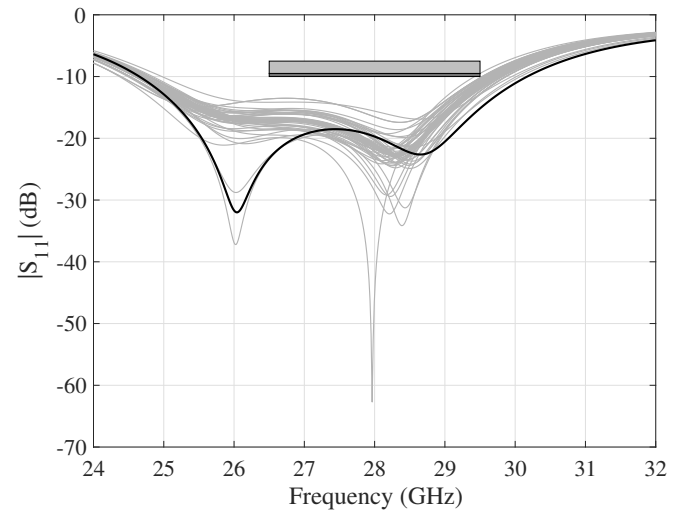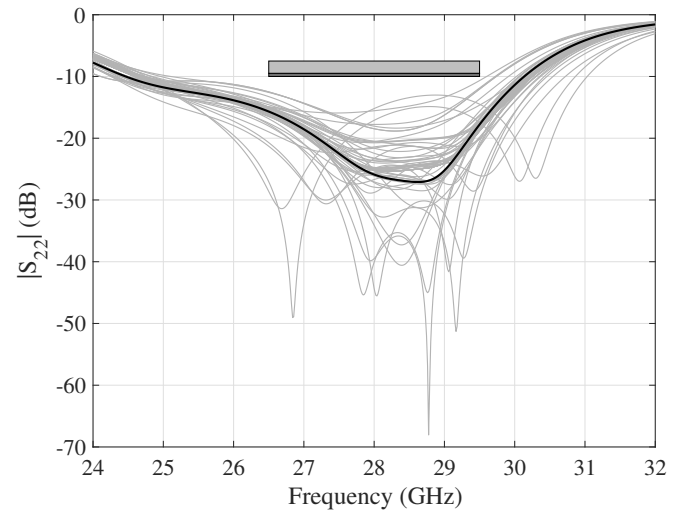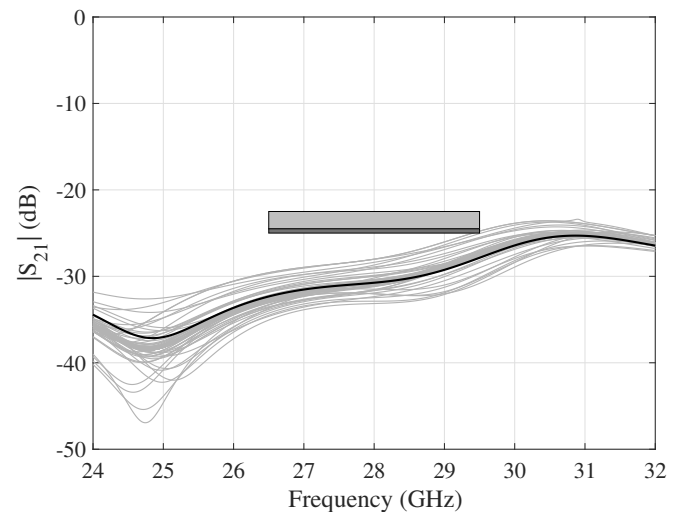
Fig. 4. Simulated stand-alone wideband dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna performance. (a) S-parameters. (b) Maximum gain and back radiation: vertical polarization (dashed line) and horizontal polarizations (dash-dotted line). (c) Total antenna efficieny.

Fig. 5. Simulated S-parameters for variations on the antenna parameters. Assessing the influence of etching tolerances [± 25 μm], misalignment [± 100 μm], variations in material permittivity [± 0.1], and thickness [± 50 μm] on antenna performance: final design (full black line) and variations (full gray lines). (a) $|S_{11}|$. (b) $|S_{22}|$. (c) $|S_{21}|$.
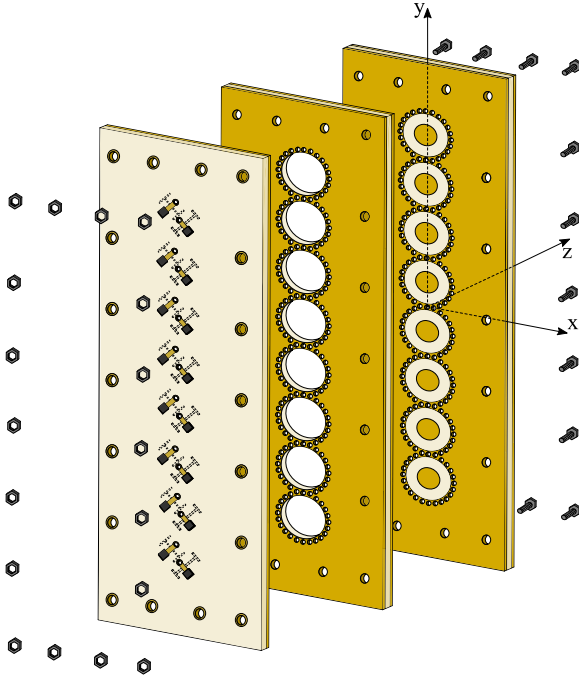
Fig. 6. Exploded view of the eight element aperture-coupled AFSIW cavity-backed patch antenna array.

variations in material parameters (dielectric permittivity and thickness). Worst-case variations in the etching of the different metalizations and in the alignment between the different metalization layers in the assembly are estimated to be ± 25 μm and ± 100 μm, respectively. The S-parameters were analyzed for misalignments of ± 100 μm in the $x-$ and $y-$directions, as well as along the first bisection.

Before analyzing the effect of variations in dielectric material parameters, the RO4350B and FR4 laminates' electrical characteristics at 28 GHz were extracted by applying the matrix-pencil two-line method [12]. In the tolerance analysis, a deviation of 0.1 on the dielectric permittivity, being 3% for RO4350B and 2.3% for FR4, is considered. Finally, the influence of a dielectric thickness variation of ± 50 μm on the S-parameters is analyzed.

Fig. 5 displays the simulated S-parameters for all previously discussed parameter sweeps. Based on this extensive parameter analysis, it can be concluded that the proposed design is sufficiently robust against fabrication errors and material variations.

III. ANTENNA ARRAY

Fig. 6 shows the exploded view of the proposed linear array. The array comprises eight dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna elements. In the array, elements are rotated over 45° to achieve a higher signal-to-noise-ratio [13] and to provide a larger physical space for the compact integration of all required active MMICs.

The proposed array is designed for the frequency range of 26.5 to 29.5 GHz. An inter-element spacing of 6.4 mm (0.63λ at 29.5 GHz) was chosen to avoid grating lobes when scanning from -35° to $+35^o$. The full array was designed in the frequency domain solver of CST Microwave Studio to validate the active reflection coefficient of the antenna element and to observe overall scanning performance of the array.
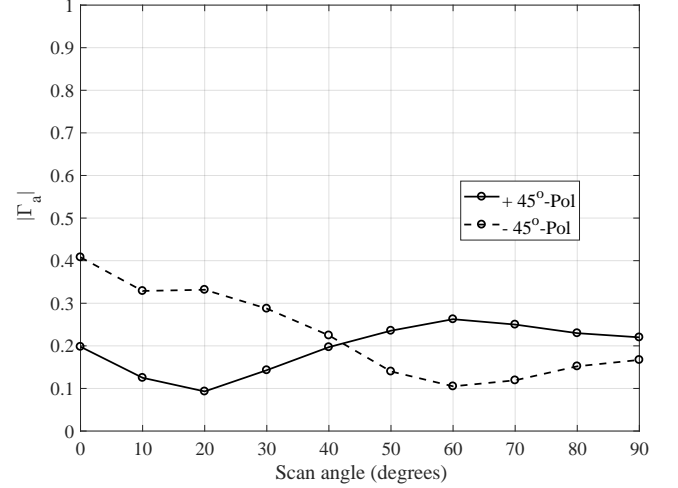

Fig. 7. Simulated active reflection coefficient ($|\Gamma_a|$) for the antenna at 28 GHz.
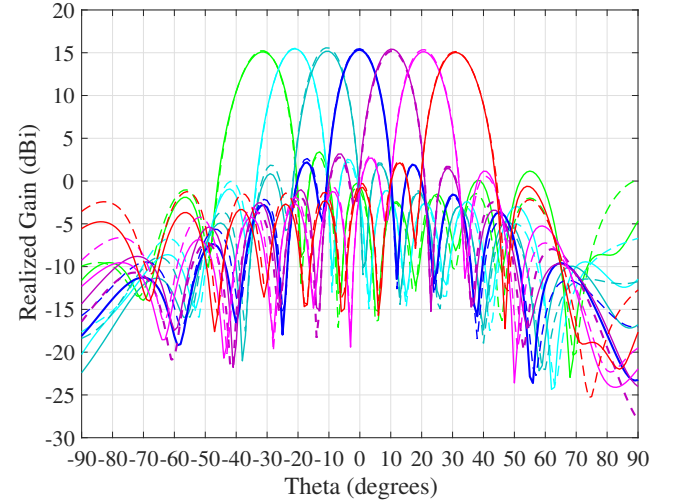

Fig. 8. The simulated radiation patterns of 8 element array at 28 GHz in the YZ-plane. Solid and dashed lines refer to $+45^o$ and -45° polarizations, respectively.

Fig. 7 shows the active reflection coefficient of one of the center antenna elements in the linear array for $\pm 45^o$ polarizations at 28 GHz, when the antenna array is steered in the YZ-plane. No scan blindness was found for both polarizations. For both polarizations, the scanning performance of the array in the YZ-plane was examined. A uniform amplitude distribution is imposed. Fig. 8 shows the radiation patterns at 28 GHz for both polarizations, when scanning in the YZ-plane. A broadside array gain of 15.4 dBi is achieved without gain reduction when the array scans up to $\pm 35^o$. Moreover, the radiation patterns for both polarizations demonstrate a grating-lobe-free beam scanning range up to a $\pm 35^o$. The half-power beamwidth amounts to 10.1^o and 77° in the YZ-plane and XZ-plane, respectively. In this scanning range, the level of the side lobes remain below -12 dB.

IV. EXPERIMENTAL VALIDATION

To validate the proposed array topology, a stand-alone antenna element, depicted in Fig. 9, is initially fabricated and validated. In the manufacturing process, the ground plane of
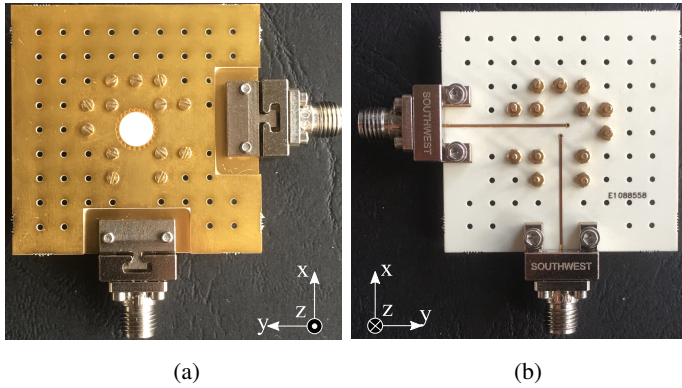
Fig. 9. Fabricated stand-alone antenna element. (a) Front-side view. (b) Back-side view.
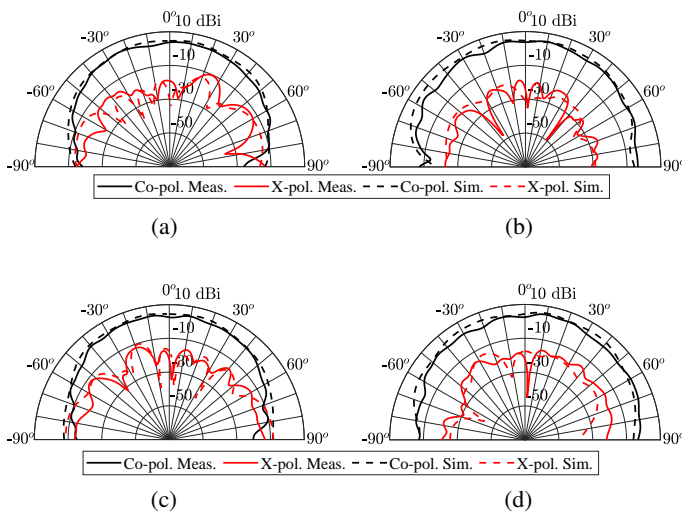

Fig. 10. Measured and simulated, including measurement connector, radiation patterns at 28 GHz. Vertical polarization: (a) XZ-plane and (b) YZ-plane. Horizontal polarization: (c) YZ-plane and (d) XZ-plane.

the antenna element is extended and an additional row of holes around the cavity is added to reduce the mechanical stress between the layers. Nevertheless, this is merely a precaution and does not have an impact on the antenna performance. The fabricated wideband dual-linearly polarized aperture-coupled AFSIW cavity-backed patch antenna prototype is validated by an Agilent E8364B PNA Microwave Network Analyzer and solder-free End-Launch connectors. The coaxial pins of the connectors are not soldered, but press fit onto the signal trace [see Fig. 9].

The radiation patterns for both orthogonal polarizations are measured in an anechoic chamber by means of an NSI MI near-field antenna measurement system. The measured and simulated gain patterns at 28 GHz, both in the XZ- and YZ-planes for both polarizations, agree well, as demonstrated in Fig. 10. A broadside gain of 5.6 dBi and a cross polarization of 15 dB are achieved. For the vertical polarization, the half-power beamwidths in the XZ- and YZ-planes equal approximately 90° and 111° whereas the half-power beamwidths approximately amount to 93° and 80° in the YZ-plane and XZ-plane for the horizontal polarization.

Furthermore, the complete array will be manufactured and measured to verify the design. The results of these studies will be presented at the conference.

V. CONCLUSION

A new low-cost and low-profile dual-polarized aperture-coupled AFSIW cavity-backed patch antenna topology has been developed as an array element. The antenna covers all worldwide-planned 28 GHz 5G bands. Its measured radiation patterns, with a low level of cross-polarization, correspond well with the simulation results. The array element has optimal electrical dimensions to be deployed in a phased antenna array for beamsteering up to $\pm35^o$. The active impedance variations with respect to steering angle and scan performance of 8 x 1 linear array have been presented. The array exhibits stable radiation patterns within the complete steering range.

REFERENCES

[1] F. Foglia Manzillo et al., "A multilayer LTCC solution for integrating 5G access point antenna modules," IEEE Transactions on Microwave Theory and Techniques, vol. 64, no. 7, pp. 2272–2283, Jul. 2016.
[2] H. Chu and Y. Guo, "A filtering dual-polarized antenna subarray targeting for base stations in millimeter-wave 5G wireless communications," IEEE Transactions on Components Packaging and Manufacturing Technology, vol. 7, no. 6, pp. 964–973, Jun. 2017.
[3] J. Lota, S. Sun, T. S. Rappaport, and A. Demosthenous, "5G uniform linear arrays with beamforming and spatial multiplexing at 28, 37, 64, and 71 GHz for outdoor urban communication: A two-level approach," IEEE Transactions on Vehicular Technology, vol. 66, no. 11, pp. 9972–9985, Nov. 2017.
[4] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "Survey of millimeter-wave communications (mmWave) for 5G: Opportunities and challenges," Wireless Netw., pp. 1–20, Apr. 2015.
[5] G. R. Maccartney, T. S. Rappaport, M. K. Samimi, and S. Sun, "Millimeter-wave omnidirectional path loss data for small cell 5G channel modeling," IEEE Access, vol. 3, pp. 1573–1580, 2015.
[6] C. Mao, S. Gao, and Y. Wang, "Broadband high-gain beam-scanning antenna array for millimeter-wave applications," IEEE Transactions on Antennas and Propagation, vol. 65, no. 9, pp. 4864–4868, Sep. 2017.
[7] J. Lian, Y. Ban, Q. Yang, B. Fu, Z. Yu, and L. Sun, "Planar millimeter-wave 2-D beam-scanning multibeam array antenna fed by compact siw beam-forming network," IEEE Transactions on Antennas and Propagation, vol. 66, no. 3, pp. 1299–1310, Mar. 2018.
[8] K. Kibaroglu, M. Sayginer, T. Phelps, and G. M. Rebeiz, "A 64-element 28-GHz phased-array transceiver with 52-dBm EIRP and 8–12-Gb/s 5G link at 300 meters without any calibration," IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 12, pp. 5796–5811, Dec. 2018.
[9] W. Hong et al., "Multibeam antenna technologies for 5G wireless communications," IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 6231–6249, Dec. 2017.
[10] K. Y. Kapusuz, S. Lemey, A. Petrocchi, P. Demeester, D. Schreurs, and H. Rogier, "Polarization reconfigurable air-filled substrate integrated waveguide cavity-backed slot antenna," IEEE Access, 2019.
[11] Eurocircuits. (2015). 'RF Pool'—All the Benefits of Pooling on RF Materials. [Online]. Available: http://www.eurocircuits.com/rf-pool-all-the-benefits-of-pooling-on-rf-materials.
[12] F. Declercq, H. Rogier, and C. Hertleer, "Permittivity and loss tangent characterization for garment antennas based on a new matrix-pencil two-line method," IEEE Trans. Antennas Propag., vol. 56, no. 8, pp. 2548–2554, Aug. 2008.
[13] Z. N. Chen and K.-M. Luk, Antennas for Base Stations in Wireless Communications. The McGraw-Hill, 2009, ch. 2, pp. 31–94.

2019 IEEE International Symposium on

Phased Array Systems and Technology

Revolutionary Developments in Phased Arrays

About the Symposium

Phased array systems continue to be a rapidly evolving technology with steady advances motivated by the challenges presented to modern military and commercial applications. This symposium will present the most recent advances in phased array technology and provide a unique opportunity for members of the international community to interact with colleagues in the field of Phased Array Systems and Technology.

15–18 October 2019 @ Westin Waltham Hotel, Greater Boston, Massachusetts, USA

Suggested Topics:

- System Architecture
- Aperture Design
- Antenna Elements
- Beamforming Techniques
- T/R Modules

- Signal Processing for Arrays
- Array Measurements
- Advanced Materials
- Packaging and Manufacturing Techniques

- 5G
- Metamaterials
- Radio Astronomy
- HF Arrays

See www.array2019.org for details

Special Sessions

Please provide suggestions for special sessions to the Technical Program Chair at info@array2019.org

General Paper Submission Procedures

All paper submissions will be peer reviewed and must be received in PDF format via the symposium web site on or before Friday, March 15, 2019. This is a firm deadline. Papers will not be accepted after this date. Papers must be in IEEE dual–column format and must be 2–pages (minimum) to 8–pages (maximum) in length including figures. Additional instructions are on the website.
www.array2019.org/call–for–papers/

Paper Template and Submission Procedures

Are available at www.array2019.org/call–for–papers/

Please note: Our submission process and dates have been streamlined – plan accordingly.

> **15 March 2019**
Full paper submission deadline

> **30 April 2019**
Author notification of paper acceptance

> **1 Sept. 2019**
Conference registration deadline for accepted authors

www.array2019.org