

Received March 16, 2020, accepted March 26, 2020, date of publication March 31, 2020, date of current version May 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984548

Taxonomy and Performance Evaluation of Hybrid Beamforming for 5G and Beyond Systems

MOHAMED RIHAN^{1,3}, (Member, IEEE), TAREK ABED SOLIMAN³, CHEN XU¹,
LEI HUANG², (Senior Member, IEEE), AND MOAWAD I. DESSOUKY³

¹School of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China

²School of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

³Department of Electronics and Communication Engineering, Faculty of Electronic Engineering, Menoufia University, Menoufia 21974, Egypt

Corresponding author: Lei Huang (lhuang@szu.edu.cn)

This work was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 61925108 and Grant 61801302, and in part by the National Natural Science Foundation of China under Grant U1713217.

ABSTRACT Increasing demand for higher data-rate wireless connectivity with lower latency is fueling the explorations of millimeter-wave (mmWave) spectrum and massive MIMO communications. Both technologies are recognized as the key enablers of 5G and beyond 5G (B5G) networks. Hybrid beamforming is one of the most promising energy and cost-effective approaches to realize mmWave massive MIMO communications with lower complexity and smaller training overhead. With the motivation of giving more insights and in-deep technical recommendations to B5G network designers regarding hybrid beamforming, we present a hybrid beamforming taxonomy in terms of channel state information (CSI) availability, frequency bandwidth, architecture complexity, analog beamformer components, number of users, connectivity to RF chains, and the digital and analog beamforming design. Furthermore, we provide a comprehensive survey on the state-of-the-art use-cases for each classification followed by identification of the future challenges and open research issues.

INDEX TERMS Hybrid beamforming (HBF), energy efficiency (EE), millimeter wave (mmWave), hardware complexity, massive-MIMO, analog beamforming (ABF), and digital beamforming (DBF).

I. INTRODUCTION

Fifth-generation (5G) wireless networks have recently gained considerable attention from academia and telecommunication industry as the cornerstone of future communication networks and smart societies. Recent reports reveal that the wireless communication traffic is doubling annually and overtaking the wired communication traffic [1]–[3]. Moreover, the demand for wireless broadband and content-rich services has also grown as a result of the emerging bandwidth-hungry applications such as cloud gaming, vehicle-to-Everything (V2X) communications, industrial automation, remote health services, augmented reality, hologram services, smart city applications, and smart homes [4], [5]. The evolution of the internet of things (IoT) is also fueling the need to support massive connectivity of devices with ultra-reliable and ultra-low-latency communications (URLLC) to enable delay-sensitive and mission-critical

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

services that require very low end-to-end (E2E) delay such as tactile internet, remote control of medical or industrial robots, and real-time traffic control [6]. It is predicted that roughly 50 billion connected devices will be served by 5G networks by 2020 [6] with an average of 6 devices per individual.

Additionally, besides its ability to support a large number and different types of communicating devices, 5G is expected to provide many services with different traffic features such as different quality of service levels (e.g. latency and data rates), varying mobility levels, different types of data (e.g. Internet protocol (IP) and non-IP data), and multiple traffic models (e.g., burst traffic, high throughput traffic, delay-sensitive traffic, and non-real-time traffic). The analysis of these specifications implies that performance guidelines for spectral efficiency, latency, system capacity, and data rates will require new supporting technologies to address such communications. As summarized in Table. 1, it is shown in the literature [7] that these specifications can be realized through the adoption of three main technologies: **network densification** to increase area spectral efficiency (i.e. more

TABLE 1. 5G performance criteria and its corresponding enabling technologies [8].

Parameter	Value	Enabling technology
Link latency	<1ms	New multiplexing and precoding methods, shorter transmit time interval (TTI).
E2E latency (from device to core)	<1ms	New network architectures, HetNets.
Connection density	100x LTE	New multiplexing and precoding methods.
Area capacity	1 Tbits/s/Km ²	Employ higher frequency bands and beamforming, mmWave and massive MIMO.
System spectral efficiency	10 bits/s/Hz/cell	New multiplexing and precoding methods, massive MIMO.
Downlink peak throughput	10 Gbits/s	Beamforming, mmWave, massive MIMO.
Energy efficiency	>90% improvement over LTE	New multiplexing and precoding methods, new control channel structures.

nodes per unit area per Hz), **utilizing the mmWave band** to increase the bandwidth (i.e. more Hz), and **employing massive MIMO** to improve spectral efficiency (i.e. more bits/s/Hz/node). The adoption of these three technologies will also provide a substantial accumulative gain in the area spectral efficiency measured in (bits/s/unit area).

With this type of massive and dense communications, energy efficiency will need to be improved in future wireless networks. Developing energy-efficient solutions while sustaining the required technical and performance specifications is inevitable, since almost 70% of the electricity consumed by telecommunication operators is actually consumed in the radio part [9]. To satisfy these specifications within the future generation networks and concurrently keep the energy-efficiency, hybrid beamforming (also known as hybrid precoding) for massive MIMO mmWave is regarded as an essential component of the 5G and beyond (B5G) wireless networks. A complete list of abbreviations that will be used throughout the paper is summarized in Table. 2.

II. CORNER STONE TECHNOLOGIES FOR ENABLING 5G AND BEYOND NETWORKS

The two most important technologies which are widely accepted as principle candidates to meet the 5G network requirements are the mmWave communication and massive MIMO which are somehow related to one another. In the following sub-sections, a detailed explanation to the concepts of mmWave communication and massive MIMO will be manipulated.

A. MILLIMETER WAVE COMMUNICATIONS

While 5G standards are still in progress, the aim for higher data rates, lower latency, and energy efficiency are obvious, and with higher data rates come with the need for wider spectrum bands. Spectral efficiency and bandwidth are the most prominent fields to be investigated to meet this increasing demand. Nowadays, most wireless systems are operating in the 300 MHz to 5 GHz band [8], but the available spectrum in these bands and up to a value of 6 GHz are not sufficiently enough to meet the 5G requirements and has become nearly fully utilized. Also, because the technologies used in the physical layer (i.e. the hardware) have reached Shannon capacity [10], the only alternative is to explore the system

bandwidth, in particular, the mmWave band extending from 3 GHz to 300 GHz. The availability of free spectrum at these bands is higher than those at sub 6 GHz bands, therefore significant efforts are exerted on utilizing these spectrum bands.

The small wavelengths at these high-frequency bands enable implementations with much more antenna elements per system within very small form factors. However, mmWave bands also have harsh propagation conditions, including heavy path-loss, high atmospheric and rain absorptions, low diffraction around obstacles and penetration through objects [7]. For example, gas absorption can cause a 60 GHz signal to attenuate 10 dB/km, while a traditional 700 MHz signal undergoes attenuation of only 0.01 dB/km. It is recognized that in specific bands (e.g. 35 GHz, 94 GHz, 140 GHz, and 220 GHz), mmWave propagation undergoes comparatively small attenuation, therefore, a long-haul connection can be realized in these particular mmWave bands. But, in other bands (e.g. 60 GHz, 120 GHz, 180 GHz), mmWave propagation exhibits a severe attenuation up to 15 dB/km [11]. The mmWave signals also undergo poor diffraction when confronting blockages because of its short wavelengths [12]. These two characteristics reduce the transmission range of mmWave and cause a frequent drops to the links.

The major difficulties in mmWave systems are spatial management, link margin enhancement, interference control, and blockage. Semiconductors and RF integrated circuits developments for mmWave are growing [13], their costs and power consumption gradually declining, and the other propagation impediments are now considered easier to bypass over time with the continuous and focused effort [14]–[17]. Besides, intelligent large array designs and the use of spatial signal-processing techniques, including beamforming, have been widely exploited to extend the coverage of mmWave networks and overcome path-loss and undesirable interference sources [16], [18]. For example, the link margin can be improved by performing beamforming in large antenna arrays to generate high directional beams. The mmWave short wavelength enables the implementation of a large number of antennas in a small form factor.

Although high mobility management is not anticipated in mmWave networks, the support of low mobility devices

TABLE 2. List of Abbreviations.

Abbreviation	Term
5G	Fifth generation
MSE	Mean square error
ADC	Analog to digital converter
MU-MIMO	Multi-user MIMO
AoA	Angle of arrival
Nbs	Number of base-stations
BD	Block diagonalization
NP hardness	Non-deterministic polynomial-time Hardness
B-MIMO	Beamspace MIMO
Ns	Number of data streams
BS	Base-station
OFDM	Orthogonal frequency division multiplexing
CAP-MIMO	Continuous aperture phased MIMO
OMP	Orthogonal matching pursuit
CoMP	Coordinated multi-point
OSA	Overlapped sub-array
CSI	Channel state information
PAPR	Peak to average power ratio
DAC	Digital to analog converter
PS	Phase shifter
DoA	Direction of arrival
RF	Radio frequency
DoF	Degrees-of-freedom
RMS	Root mean square
E2E	End-to-end
SDMA-OFDMA	Space-division multiple-access and orthogonal Frequency-division multiple-access
FDD	Frequency division duplexing
SINR	Signal-to-interference-plus noise ratio
FeICIC	Further enhanced inter-cell interference coordination
SMG	Spatial multiplexing gain
GP	Gradient pursuit
SNR	Signal to noise ratio
HBF	Hybrid beamforming
SU-MIMO	Single user multiple input multiple output
IoT	Internet-of-things
SVD	Singular value decomposition
IP	Internet protocol
T	Number of channel uses in a time-frequency block
JSDM	Joint spatial division multiplexing
TDD	Time division duplexing
K	Number of single-antenna users
TTI	Transmit time interval
LS	Least square
ULA	Uniform linear array
MDs	Mobile devices
URLLC	Ultra-reliable low latency communication
MF	Matched filtering
V2X	Vehicle-to-everything
MIMO	Multiple Input multiple output
VLSI	Very large scale integration
MMSE	Minimum mean square error
WSMSE	Weighted-sum mean-square error
MPCs	Multipath components
ZF	Zero-forcing
EGT	Equal gain transmission

is necessary. Using directional antenna arrays in mmWave bands requires accurate beam alignment since the dominant multi-path components (MPCs) are limited. Moreover, link acquisition and establishment have to be quick and adaptive based on device locations. Thus, beam steering capability is inevitable in mmWave networks, and exploiting the array gain to offset the pathloss effect should be considered carefully. The baseband effective channel can also be utilized efficiently through employing directional arrays. For instance,

the channel delay spread can be reduced and consequently alleviating the inter-symbol interference and enhancing the system performance [19]. However, since mmWave systems with large bandwidth require short symbol duration, complex equalization methods are expected [20]. Unlike the current microwave-based architectures, hybrid beamforming (HBF) architecture turns out to be more suitable for mmWave characteristics [21]. The HBF architectures and algorithms in the sub 6 GHz bands can, in theory, be utilized at the mmWave spectrum. In practice, however, mmWave bands exhibit different propagation characteristics and require specific RF hardware. Therefore, innovative HBF methods considering those practical differences are necessary to realize 5G systems. Fortunately, the sparse nature of the mmWave channels can be exploited for channel estimation and beam training optimization by reducing the calculation complexity.

Regardless of the potential advantages offered by the mmWave system, it is obvious that it is a power-limited system as a result of its high blocking and pathloss. Also, it is considered as interference-limited system because of co-channel interference. Therefore, it is essential for the mmWave 5G system to develop an optimal beamforming design to focus the beams into the desired direction to reduce the pathloss and avoid co-channel interference is essential to enable mmWave in 5G systems. Fortunately, mmWave has small wavelengths that allow the implementation of a massive number of beamforming antenna elements in a practical form factor to generate directional beams that can serve the maximum number of mobile devices (MDs). Significant efforts have been focused on different beamforming techniques to improve the massive MIMO system energy and spectral efficiencies. The works in [22], [23] exploited massive MIMO deployment in mmWave systems and examined the integration of both technologies in the 5G networks. It is shown that the mmWave massive MIMO represents as a great potential for 5G networks.

B. MASSIVE MIMO

Massive MIMO increases the data rates and capacities of the traditional MIMO systems and has emerged as an enabling technology for the 5G networks [7], [24]. The combination of massive MIMO and mmWave technologies, in particular, regarded as a fundamental component in 5G to improve spectral efficiency and defeat the bandwidth constraints [15], [25]. Multiple antennas can be leveraged within the transmitter and/or the receiver to achieve multiplexing gain, in which a simultaneous transmission of parallel data streams is transmitted over multiple antennas to improve bit-error-rate (BER), diversity gain where a redundant data streams are sent using space-time coding to reduce BER, or antenna gain by increasing the signal-to-noise ratio (SNR) at the receiver and to eliminate co-channel interference in a multi-user systems.

With proper antenna arrangements, antenna array directional beams can also increase SNR, reduce the root-mean-squared (RMS) delay spread due to multi-path scattering at the receiver, and enhance the Rician factor gain [26].

Massive MIMO can also minimize fading effects and the transmission energy using the beamforming gain [27]. Besides, massive MIMO is necessary for mmWave frequencies because it utilizes beamforming gain to improve the link margin. Consequently, the massive MIMO system facilitates additional access to mmWave bands and help increasing the spectral efficiency [22], [28].

The use of massive MIMO holds the potential for higher array gain. Additionally, increasing the number of antennas at the transmitter and receiver, enables higher multiplexing and diversity gains and yield a channel matrix with desirable characteristics [28]. The precoding and combining techniques are utilized to achieve this gain. It was shown that linear precoding and combining techniques such as zero-forcing (ZF) and matched filtering (MF) can provide optimal performance by exploiting the favorable propagation characteristics [24]. However, these techniques need a dedicated radio frequency (RF) chain for each antenna. In conventional MIMO, every antenna is attached to a separate digital base-band. For each antenna, this structure requires dedicated filters, digital-to-analog converter (DAC) or an analog-to-digital converter (ADC), and amplifiers. These series of components that connect the antennas to the baseband is called radio frequency (RF) chain. Thus, using different digital beamforming (DBF) methods at the baseband, the precoding and combining can be performed with full control over the amplitude and phase of the signal at and from each antenna. Unfortunately, massive MIMO also brings some challenges [24]. Considering the massive number of antennas, a large computational load, and more cost are inevitable since RF components for mmWave frequencies are expensive and power hungry [27]. Traditional DBF will lead to higher implementation complexity, especially with mmWave communication systems [18]. Consequently, power consumption and cost are the limiting factors for using DBF in massive MIMO mmWave systems. It is important to develop efficient and cost-effective architecture design that can achieve the potential gain using massive antenna arrays and small number of RF chains.

Massive MIMO technology also introduces new obstacles such as pilot contamination which is a result of inter-cell or intra-cell interference during pilot transmissions from the mobile devices (MDs) to the base-station (BS) to handle the interference problem. Precise CSI estimation is necessary for wireless systems. In the case of multi-user massive MIMO (MU- massive MIMO) systems, the CSI is significantly important, where it enables interference-free transmission of multiple streams and at the same time eliminates colored the interference between users. To estimate the CSI, training sequences or pilots are used. The performance of massive MIMO systems is affected by the CSI availability at the BS. For the BS to obtain the CSI in a massive MIMO system, a time-division duplex (TDD) mode is usually considered. With this mode, the BS estimates the uplink channel and applies the same channel parameters in the downlink. In the frequency-division duplex (FDD) mode, the estimation poses a big challenge because the channel estimation time increases

with the transmit antenna number and accordingly increases the signaling overhead, where the MDs have to feedback the estimated channel parameters to the BS. Besides, high mobility decreases the channel coherence time. Consequently, estimating the channel in the downlink might not be feasible. In contrast, the TDD mode allows lower complexity designs as both the uplink and downlink transmissions use the same frequency at different time slots. CSI availability at the BS is necessary to apply the beamforming to increase the system energy and spectral efficiency. In this context, [29] provides analysis and classification for the pilot contamination problem and demonstrates the main factors that affect the performance of the massive MIMO system using TDD mode such as the need for feedback transmission. However, the study did not consider pilot contamination and estimation errors in HBF architectures. The survey in [30] provides classification and analysis for massive MIMO systems and applications as well as channel measurements and models. However, The leveraging of mmWave technology and the necessity of employing HBF to decrease the system complexity and increase the energy efficiency did not give considerable attention. Motivated by the shortcomings of the study given in [30], this work mainly aims to fill the gaps in the previous studies regarding hybrid precoding and gives a detailed exploration for hardware architectures, methods of deployment, frequency bandwidth, and CSI availability, etc.

III. BEAMFORMING IN 5G NETWORKS

Through beamforming, the signals produced by an antenna array are directed to a specific angular direction [31]. Specifically, beamforming transmits redundant symbols over each transmit antenna with a weighting factor. At the receiver, the received signals are coherently combined using separate weighting factors to increase the SNR. The increase in SNR in large antenna array systems is known as the beamforming gain, while diversity gain is considered as the change in error probability slope resulting from the beamforming gain [32]. In massive MIMO systems, beamforming leverages smart antennas to transmit and receive the signals. Smart antennas are arrays that employ signal processing algorithms to detect spatial signal identifiers; e.g. the direction of arrival (DoA); and use these identifiers to estimate the beamforming vectors which are used to recognize and trace the desired signal transmitted from MDs.

It is favored to have independent weighting control over each element in an active antenna-array to realize the beamforming control and flexibility. This, however, requires a dedicated RF chain for each element. Such a requirement in a massive MIMO system is limited by power, cost, and space. In beamforming, concentrating the radio signal to a narrow beam helps to overcome the effect of reduced propagation associated with very high-frequency carriers in mmWave. However, beamforming antennas at BSs must follow the mobile equipment for the device to remain within the beam, and cost would likely be significantly increased if much more antennas are added to support large numbers of users per cell.

Besides, both horizontal and vertical orientations need to be taken into account when designing the beams.

There are many efforts in the literature to classify beamforming methods based on different aspects. Authors in [33] classified beamforming methods based on their physical characteristics into two classes, namely, switched and adaptive beamforming. They also classified the methods based on the type of antenna arrays to linear arrays, circular arrays, and rectangular arrays. Recently, authors in [34], [35] divided the beamforming methods into ABF, DBF, and HBF. The ABF has the advantage of using low cost phase shifters for massive MIMO, while DBF has the advantage of providing more precise and fast results to acquire user signals, but it also adds more complexity and cost to the design. Therefore, DBF is not suitable for massive MIMO systems. In contrast, HBF is developed to acquire the benefits of ABF and DBF for massive MIMO. Another factor that could be used to classify beamforming methods is the signal bandwidth (i.e. wideband or narrow-band). Using a wideband mmWave instead of the standard narrowband can enhance the beamforming performance. Besides, because of the small wavelength and sharp beamwidth, the antenna array can be extremely small, allowing more antenna elements to be implemented in small form factors while limiting the distance between the BS and MD to a few hundred meters.

A. NARROWBAND VERSUS. WIDEBAND BEAMFORMING

Based on the signal bandwidth, beamforming can be classified into two classes, narrowband and wideband beamforming techniques. Beamforming with narrowband signals can be performed by the instant linear combination of the received array signals. On the other hand, wideband signals require additional processing (e.g. tapped delay lines and sensor delay lines). Current standard wireless technologies are primarily focused on narrowband beamforming, however, wideband beamforming is also growing as an essential part of the 5G network due to the adoption of mmWave which can enhance the beamforming capabilities. Besides, because of the small wavelength and the sharp beam-width, the antenna array can be extremely small, allowing more antenna elements to be implemented in small form factors. Therefore, the mmWave wideband beamforming in 5G can offer remarkably high speeds that approaches the maximum achievable capacities.

B. SWITCHED VERSUS ADAPTIVE ARRAY BEAMFORMING

Beamforming can also be classified into switched or adaptive array systems. The switched method employs a determined or fixed beamforming network that generates predefined beams. The Butler matrix is a well-known solution for switched beamforming which consists of phase shifters, crossovers, and hybrid couplers [36]. To determine the proper beam to acquire the desired signal from a specific MD, the switched system needs a switching network. However, the determined beams might not always point to the desired

direction. Besides, several MDs are usually served by the same beam. These issues have been addressed in [37]–[39].

Adaptive array systems in contrast with switched systems can form a unique beam for each MD by applying weight vectors to the detected signals through adaptive array processors to control phase changes between the antenna array elements and their amplitude spreading. This method can form a precise beam and direct the main-lobe towards the preferred MD and produce the null toward the interfering MD. However, the adaptive system requires that the BS updates the MD locations as an estimated DoA of the received signals, while in practice, the DoA estimates of a large number of MDs can be a challenging task. In adaptive array systems, beamforming algorithms are classified into two types; non-blind and blind adaptive algorithms. Non-blind algorithms require prior information of the transmitted signal by using training signals to determine the beam direction. Blind algorithms, on the other hand, require no prior knowledge. In practice, it is easier to implement a switched beamforming system than an adaptive beamforming system. However, adaptive beamforming can decrease the interference between different MDs and achieve better energy efficiency [40], [41]. In general, both methods have their advantages and disadvantages [42], which demonstrate that although adaptive beamforming is hard to implement, most of the massive MIMO studies favor this design over switched beamforming due to its reliability and applicability.

C. HARD AND SOFT ANTENNA SELECTION

To reduce the number of RF chains in massive MIMO systems, hard and soft antenna selection methods are proposed in [43]. A network of switches is used with the hard selection approach to connect the RF chains to the antennas and, based on the design objective (e.g. spectral efficiency maximization), the best antennas set is chosen. Exhaustive search is employed with different combinations of the selected antennas to reach the maximum performance, which introduces high computational complexity. To circumvent this issues, sub-optimal methods to maximize the spectral efficiency based on convex optimization are explored in [44]–[47]. When the number of antennas is significantly larger than the number of RF chains, considerable beamforming gains cannot be realized because of the array gain loss, which prohibits the application of hard antenna selection approaches. In contrast, the RF chains are connected to the antennas through a network of phase shifters in the soft antenna selection [43], [48], [49], which offers better flexibility.

D. ANALOG, DIGITAL, AND HYBRID BEAMFORMING

Another classification of beamforming techniques is the classification into analog, digital, and hybrid beamforming techniques. ABF is the oldest spatial filter which is introduced 50 years ago. Analog beamforming controls the transmitted signal phase through low cost phase shifters while employing RF switches to steer the beams. Analog or RF beamforming, as shown in Fig. 1, are typically

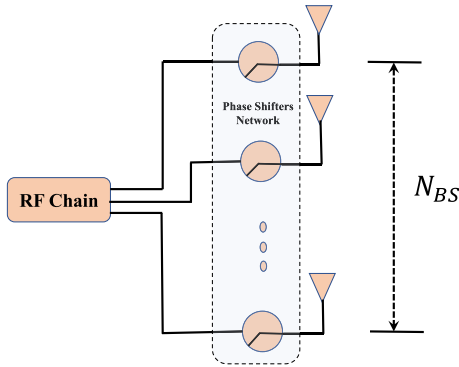


FIGURE 1. ABF at the basestation where all antenna elements share a single RF chain through a phase shifter in which N_{BS} is the number of antennas at base station.

implemented using phase shifters where all the antenna elements share a single RF chain and the beamforming matrix weights are constrained with the phase shifters constant-amplitude. The ABF has been extensively studied in the literature and extensively explored in mmWave MIMO systems in particular [34], [50]–[53]. However, most of the approaches introduced in literature provide insufficient antenna gain and moderate performance. In particular, these efforts have not exploited with the fully connected mmWave massive MIMO channel structure. ABF offers lower complexity compared to DBF, but its performance still smaller due to the lack of amplitude control. A solution for this problem has been proposed in [45], [51] by utilizing simple analog switches to design the antenna subset selection. However, this design can only provide limited array gain and lower performance in correlated channels.

Based on mmWave realistic channels structure, a low complexity design is presented in [54]. A clustered channel model is formulated to exploit the limited scattering at high frequency and antenna correlation for single-user precoding in practical transceiver architectures. ABF has been used for short-range mmWave systems such as in IEEE 802.15 standard [55]. However, the phase shifters used in ABF impose the constant modulus constraint to the system. With each Beamforming criteria, the beamforming design problem is formulated to an optimization problem. Solving such optimization problem may be a challenging task. The solution becomes more challenging with the consideration of phase shifters' practical constraints which adds large computational load to the system based on the value of quantization resolution.

In contrast to ABF, the DBF employs a digital signal processor to perform beamforming, providing a higher degree of freedom (DoF) that allows more flexibility in implementing effective beamforming algorithms. As shown in Fig. 2, each antenna element in DBF systems requires a separate RF chain and additional operations such as DoA estimation, adaptive steering of its beams and nulls to improve the signal-to-interference-plus-noise ratio (SINR), and programmable

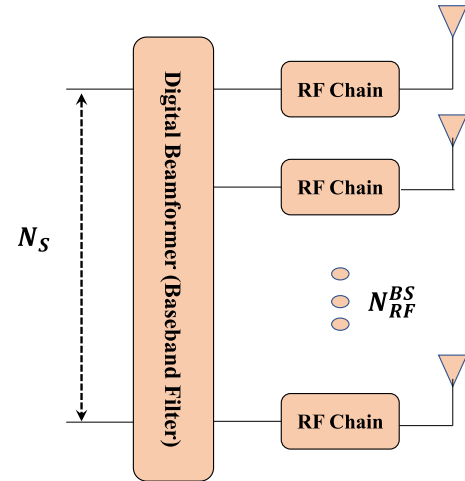


FIGURE 2. DBF at the base station where each antenna element requires a separate RF chain, where N_S is the number of data streams, and N_{RF}^{BS} is the number of RF chains at the base station.

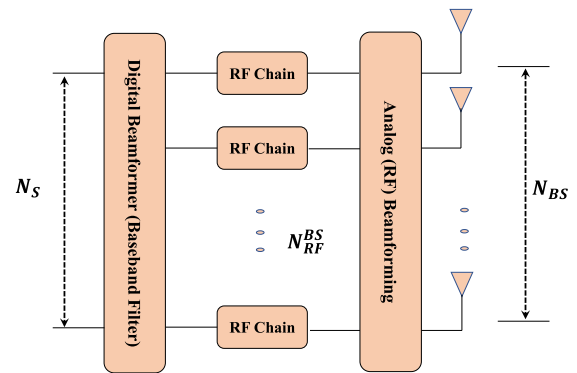


FIGURE 3. HBF at the base station, where ABF connects a large number of antennas and the DBF connects a small number of RF chains, where N_S is the number of data streams, N_{RF}^{BS} is the number of RF chains at the base station, and N_{BS} is the number of antennas at base station.

control of antenna radiation patterns. In general, to achieve optimal MIMO channel capacity, the availability of full CSI at the transceiver is crucial and digital processing at both ends is required. This, however, demands a dedicated RF chain for each antenna to deliver higher DoF. However, using a large number of RF chains will produce a complex architecture with high-power consumption due to its complexity and high-power consumption of its mixed-signal circuits. Moreover, due to the large number of antenna elements in massive MIMO, DBF can also be expensive to realize. Hence, the fully DBF solution is not viable for implementation for massive MIMO systems at mmWave frequencies.

In conclusion, by using low cost phase shifters, ABF is easier to implement and more cost-effective when compared to DBF but achieves lower performance due to the lack of phase shifter amplitude control. As shown in Fig. 4, the fully DBF outperforms the ABF and achieves higher spectral efficiency. To realize an optimal trade-off between the cost of the analog and the performance of the DBF, the HBF architecture

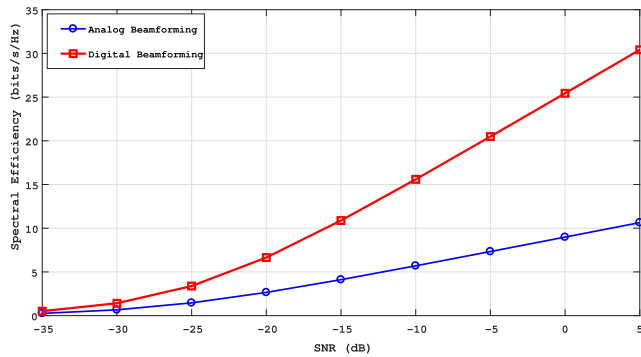


FIGURE 4. The spectral efficiency of the ABF and DBF, DBF achieves higher spectral efficiency with more complexity and less energy efficiency when compared to analog beamforming.

has been proposed [56]–[58]. HBF is broadly employed in massive MIMO systems where baseband signals are generated in the DBF while the analog part accounts for the RF chains by decreasing the quantity of ADCs and DACs, which in turn enhance power amplifier performance or modifies the structure of the mixers to reduce the costs. In this structure, as shown in Fig. 3, a network of phase shifters connects a small number of RF chains to a large number of antennas.

It is envisioned that HBF will achieve the gains of analog and DBF. In this context, HBF architecture has appeared as a promising solution for the next generation of mmWave massive MIMO systems [59]. Accordingly, HBF architecture has a critical value when implementing massive MIMO systems because of its energy and cost-efficiency [20]. In mmWave massive MIMO, the analog and digital signal components for each antenna make it impractical to perform all the signal processing tasks with full-digital architecture because of the high cost and energy consumption [28], [48]. This therefore motivates us to develop different HBF architectures.

IV. HYBRID BEAMFORMING

Hybrid ABF and DBF was first suggested in [43] under the term “soft antenna selection”. The hybrid architecture utilizes the digital precoding and analog beamforming methods to balance the cost and performance of both approaches. The HBF can be seen as a spatial filter that has the ability to strengthen the desired signal components as well as reduce the impact of undesired signal components. The huge interest in HBF is driven by the fact that the number of RF chains is only lower-limited by the number of transmitted data streams, while the diversity and beamforming gains are limited by the number of antenna elements [32].

Massive MIMO arrays offer a large number of DoFs, which improves the wireless system performance by reducing the channel fading effect. In DBF, every antenna element has to be connected to at least one RF chain which in turn leads to a huge complexity and cost with mmWave massive MIMO systems [18]. In contrast, the HBF employs analog phase shifters with lower RF chains which leads to low complexity and cost-effective system with virtually the same

performance [35], [60]–[69]. With HBF, the beamforming process could be implemented within both the analog and digital domains. In the digital domain, the transmitting signal is first processed using a digital precoder in the digital domain, since no RF chains are required in this part, the signal dimension is not high and low-dimension precoder could be used. In the second part, traditional analog beamforming, which is typically built based on RF phase shifters, is employed to direct the digital output to the antenna arrays with low complexity. At the receiver, the same procedures are performed in a reverse order. In general, HBF adjusts an efficient trade-off between the spectral, energy efficiencies, and hardware complexity to take the potential advantages of both ABF and DBF [48], [70]. The main benefits of using HBF architectures can be summarized as follows:

- HBF enables practical and efficient use of the massive MIMO mmWave systems with great capacity improvements. Unlike HBF, full-DBF is more complex and expensive and analog beamforming suffers from imprecision and interference problems [49].
- HBF minimizes the hardware cost since it employs fewer RF chains at the transceivers compared to fully digital architecture with the same number of antennas [57].
- HBF is more energy-efficient than legacy MIMO systems operating in traditional microwave bands. HBF allows the realization of the large antenna arrays in mmWave massive MIMO which minimizes the consumed power in the transceivers. In the uplink, HBF can also achieve more energy efficiency because it reduces the power consumption for each MD without affecting the performance [71], [72].
- HBF designs leverage additional DoFs compared to analog beamforming by adding the extra digital precoding stage to support multi-stream and multi-user transmission.

V. HYBRID BEAMFORMING CLASSIFICATION

There are only a few studies that demonstrate the hardware features and classification of HBF architectures. Authors in [25] classified HBF architectures at the BS side based on the CSI availability (i.e. full or estimated), the frequency bandwidth (i.e. wideband or narrowband), and the architecture complexity (i.e. full complexity, reduced complexity, and switched). Authors in [28] classified HBF architecture based on the components that build the used ABF, to switches-based HBF (S-HBF), phase shifters based HBF (PS-HBF), and lens antenna arrays based HBF (LNA-HBF). In the S-HBF architecture, the sparsity of the mmWave channel is exploited, where only a subset of antennas is chosen instead of optimizing all the quantized phase values. The PS-HBF experiences quantization errors because of the phase shifters finite step and consumes more power, however, it also can reduce the residual interference between data streams. In the LNA-HBF architecture, an LNA is used in the ABF stage instead of the switches and phase shifters, and the continuous

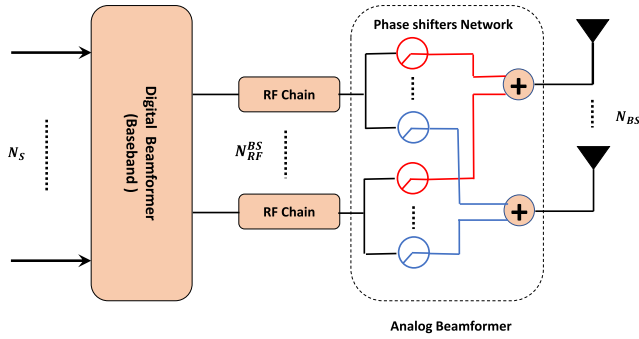


FIGURE 5. HBF fully-connected architecture at the BS, all RF chains are connected to all the antennas; large number of phase shifters is needed to fully map all the RF chains to the antennas, where N_S is the number of data streams, N_{RF}^{BS} is the number of RF chains at the BS, N_{BS} is the number of antennas at BS.

aperture of the array directs the transmitted beams. This study [28], concentrated on the analog stage of the HBF without addressing the DBF stage. As highlighted above, the HBF design problem can be considered based on several criteria in different situations. For instance, the spectral efficiency maximization can be examined in SU or MU scenarios, narrowband or wideband channels, with full or estimated CSI, and by using joint or separate optimization design. In the following sections, a comprehensive classification of HBF and a survey of the existing works in each class are presented.

A. FULLY-CONNECTED AND PARTIALLY-CONNECTED HYBRID BEAMFORMING ARCHITECTURES

Based on the connectivity of phase shifters network, PS-HBF architecture is commonly divided into two configurations, fully-connected and partially-connected architectures. Fig. 5 shows the fully-connected architecture, all RF chains are connected to all the antennas. In this architecture, a large number of phase shifters are employed to fully map all the RF chains to all antenna elements. In the partially-connected architecture shown in Fig. 6, every RF chain is connected to a subset of antennas and each antenna of this subset is attached to a phase shifter. This architecture has many advantages compared to the fully-connected architecture such as lower complexity and easy of implementation. Furthermore, the hardware cost is also decreased as fewer RF chains are needed. On the other hand, since RF chains are attached to fewer antenna elements, several undesired issues arise, such as the lower spectral efficiency, weaker directivity, wider beam-width, and other chains interference. Despite these limitations, MIMO systems help to reduce the interference in the partially-connected architectures. Furthermore, partially-connected architectures with its remarkable low complexity and accordingly low power consumption make it more practical for mobile stations implementation in the uplink. Fully-connected architecture offers better spectral efficiency and higher beamforming gain, but it also requires more power and it is harder to realize because of the higher number of RF chains and its inter-connections [73], [74]. Therefore,

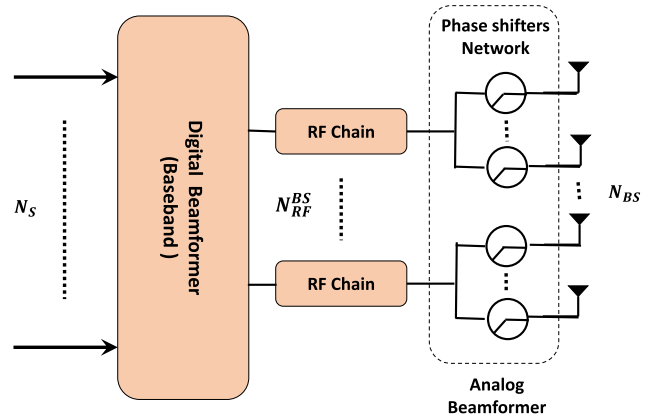


FIGURE 6. HBF partially-connected architecture, every RF chain is connected to only a subset of antennas and every antenna is attached to a phase shifter where N_S is the number of data streams, N_{RF}^{BS} is the number of RF chains at the base station, N_{BS} is the number of antennas at BS.

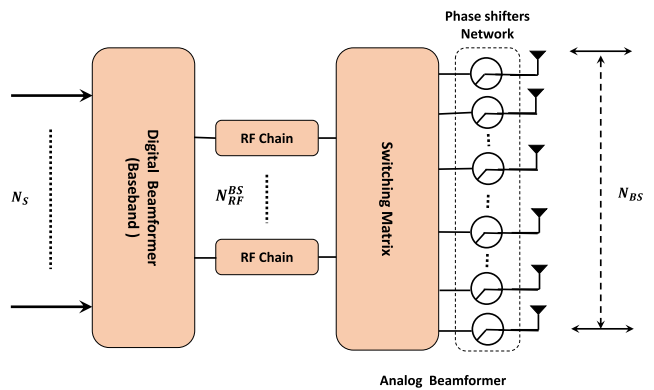


FIGURE 7. Partially-connected dynamic architecture, dynamically adapt the average channel statistics by using switches and phase shifters and employing greedy algorithm with low complexity instead of exhaustive antenna search where N_S is the number of data streams, N_{RF}^{BS} is the number of RF chains at the BS, and N_{BS} is the number of transmit antennas.

partially-connected architecture is more practical and cost-effective, however, fully-connected architecture is still regularly considered in academic works.

The purpose of all HBF architectures is to minimize the signal processing and hardware complexity while achieving near-optimal performance. As discussed earlier in the previous section, the fewer RF chains in partially-connected architectures provides better energy efficiency with a remarkable decrease in spectral efficiency, therefore the collaboration of such architecture and massive MIMO, where spectral and energy efficiency are maximized, is anticipated to deliver significant energy and spectral efficiency improvement for 5G and beyond network.

Dynamic partially-connected architecture is shown in Fig. 7, and its main goal is to dynamically adapt the average channel statistics by using switches and phase shifters, then employing a low complexity greedy algorithm instead of an exhaustive antenna search [75]. Its spectral efficiency is higher than the partially-connected architecture but still lower

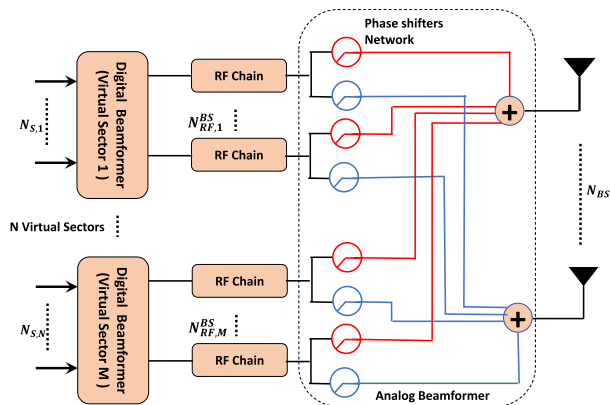


FIGURE 8. Fully-connected virtual sectorization, a special type of fully-connected architecture in which a separate digital beamformer is connected to every virtual sector where $N_{S,i}$, is the number of data streams for the i^{th} virtual sector with $i = \{1, \dots, N\}$, $N_{RF,j}^{BS}$ is the number of BS RF chains at the j^{th} sector with $j = \{1, \dots, M\}$, and N_{BS} is the number of BS antennas.

than the fully-connected one and it also consumes less power than ABF. However, the use of switches in the mmWave causes high insertion losses [76]. The virtual sectorization architecture in Fig. 8, is a special type of fully-connected architecture in which a separate digital beamformer is connected to every virtual sector. This architecture has the same spectral efficiency as that achieved with the fully-connected architecture.

B. HYBRID BEAMFORMING IN SINGLE-USER AND MULTI-USER SCENARIOS

1) SINGLE-USER SCENARIO

The sparsity nature of mmWave channels offers a great advantage over conventional sub 6 GHz channels, which lies in the available fewer spatial DoFs. The sparsity can be exploited to simplify both the channel estimation and beam training procedures. In SU-MIMO systems, the simplest form of HBF exploits the channel sparsity and concentrates the array gain to a limited number of multipaths in the ABF stage, while multiplexing data streams and allocating powers in the DBF stage. But, it turns out that such a hybrid architecture is only asymptotically optimal in conventional MIMO systems [77]. For massive MIMO with large array size, optimal HBF architectures and algorithms are still not fully understood. Also, Hardware and computational complexity reductions are highly investigated due to the unique features of mmWave massive MIMO systems. Several HBF methods have been proposed for mmWave SU-MIMO channels. Among the proposed approaches for SU-MIMO mmWave systems are the codebook-based beamforming, spatially sparse precoding, and antenna and beam selection methods.

- **Codebook-based beamforming:** Instead of estimating the large channel matrix directly at the receiver, the codebook-based beamforming approach employs a pre-defined set of beams to perform downlink training

and only feeds back the selected beam index to the transmitter. With large antenna array systems, in particular, with a fully-connected hybrid architecture, the beam search over a large space can be complex and the feedback process imposes a large overhead. The mmWave sparsity can be exploited in the design of the codebook to reduce the beam search complexity and the feedback overhead. Each codeword is created based on the orthogonal matching pursuit (OMP) algorithm to minimize the mean square error (MSE) with a pre-defined number of RF chains equal to the number of the antenna beam patterns.

- **Spatially Sparse Precoding:** This technique can attain the same performance level as the fully DBF does. With mmWave channels of a sparse nature, the number of dominant multipaths components is small, the optimal precoder can be achieved by using a finite number of antenna elements [48]. The multi-path sparsity restricts the ABF to a set of array response vectors, and the baseband precoder optimization can be formulated into a matrix reconstruction with a cardinality constraint on the number of RF chains. Analog combiner near-optimal solution can then be determined using sparse approximation techniques (e.g. OMP) [48].

2) MULTI-USER SCENARIOS

Recently, HBF is heavily considered in mmWave MU-MIMO systems. The hybrid architecture at the BS can help it to multiplex and transmit data streams to multiple MDs equipped with single antenna or an array of antennas. When several MDs, each with a single RF chain and many antennas are considered, the strongest beam pair is selected and this in turn helps the ABF and ZF digital precoding to alleviate the inter-user interference. In this scenario, the hybrid architecture significantly outperforms the analog beam steering approach. HBF based on beam selection and Beam-space-MIMO (B-MIMO) approaches can also be applied to MU-MIMO systems with linear baseband digital precoders. Despite of its effectiveness and advantages compared to fully-digital precoders in mmWave channels, the use of HBF with mmWave MU-MIMO systems exploits many new challenges which need further research efforts. Among such challenges, we can mention the multi-user scheduling, 2D and 3D designs of lens antenna arrays. Another important aspect that should be explored is the hardware imperfections of RF transceivers which degrade the spectral efficiency in many ways. For example, it is harder to precisely create the desired transmit signals when high beamforming gain is required. That is due to the fact that non-linear distortion at the receiver depends on the instant channel gain, which in turn affects the SNR. Transceiver imperfections are more obvious at mmWaves frequencies, and subsequently the spectral efficiency and SNR of hybrid precoders and combiners do not grow linearly with the number of RF chains. Knowledge of the statistical characteristics of transceiver imperfections at mmWaves frequencies is required to fully recognize the spectral efficiency

scalability in large MIMO systems. Furthermore, with SU transmission, the channel capacity can be achieved with the transmission design based on singular value decomposition (SVD) and water-filling. However, in MU transmission, SVD calculation at both sides becomes more complicated due to the lack of collaboration between MDs. Usually, dirty paper coding is employed to achieve the channel capacity, but its complexity diminish its realization in practice. Accordingly, low complexity sub-optimal linear algorithms, such as, ZF are still in use and its ability to achieve optimal sum-capacity for massive MIMO also has been demonstrated [24].

C. HYBRID BEAMFORMING WITH PERFECT AND ESTIMATED CSI

The performances of Beamforming algorithms rely mainly on the availability and quality of CSI at the transceivers. In practice, the resources assigned to the process of channel estimation are limited with both time division duplex (TDD) and frequency division duplex (FDD) modes, which in turn degrades the estimation accuracy and affects the beamforming performance. Moreover, the channel estimation process becomes considerably difficult within massive MIMO systems and many research efforts have been devoted to minimize the overhead and complexity of such a process. In FDD mode, a series of training pilots are broadcasted from the transmitter to the receiver to estimate the channel and then feedback the channel parameters to the transmitter. In this situation, the estimation error depends mainly on the noise and the limited number of feedback bits. In massive MIMO systems with large number of antennas, the estimation and signaling time may exceed the channel coherence time, which in turn lead to the generation of interference. In a TDD mode, it is necessary to calibrate the uplink and downlink circuits. Also, because the number of orthogonal pilots is related to the number of users, users sharing same pilots in different cells can interfere with each others.

In HBF architectures with RF chains of numbers significantly smaller than the number of antennas, the aforementioned problems might become more challenging. Specifically, reducing channel estimation signaling is a key design objective in HBF which limits the achievable spectral efficiency. Another challenge may arise due to the rapid channel variations and its accompanied estimation delay, which in turn can affect the performance, particularly with high mobility of the MDs in outdoor conditions. Additional difficulties in the design of the channel estimation algorithms can also appear in practice. For instance, beamformer designs that are based on the antenna array geometry, can not be achieved in practice due to the neglect of user fingers effect on the array gain [78]. Based on the above discussion, we have classified the works that concentrate on the effects of channel estimation on the HBF, as follows:

- **HBF Based on Perfect CSI:** Finding the ABF and DBF matrices, which are the two parts of the HBF, is a challenging task even when full CSI is available at the

transmitter [48]. The difficulty stems from several facts, e.g., (1) hybrid beamformers and combiners at each end are coupled, which results in a non-convex optimization problem; (2) Using phase shifters to design the analog precoder and combiner inflicts additional constraints; (3) When phase shifters with finite-resolution are used, the optimal analog beamformers can be obtained from a discrete finite set and results in a non-deterministic polynomial-time hardness (NP-hardness) problem.

- **HBF Based on Estimated CSI:** The added overhead to estimate the CSI is a significant challenge in HBF designs, which depends on whether the system employs TDD or FDD. Massive MIMO downlinks with a digital structure allow a spatial multiplexing gain (SMG) of $M(1 - M/T)$, where T is the number of channel uses in a time-frequency block, and $M = \min(N_{BS}, K, T/2)$. Here, N_{BS} , K , and N_s stands for the number of antennas at the BS, the number of single-antenna users, and number of data streams, respectively. In massive MIMO with full CSI, the highest possible SMG is constrained by the channel coherence block size due to the large values of K and N_{BS} . To solve this issue, low dimensional CSI may be used to reduce the overhead. With the employment of FDD mode, the overhead increases because of the required uplink feedback for each antenna [25], [79]. One way to minimize the training overhead is to use low resolution CSI. In this context, several papers addressing the HBF designs can be divided into two stages. Based on the estimated CSI only the ABF is performed in the first stage, while in the second stage, DBF is performed based on the full CSI [48].

D. HYBRID BEAMFORMING WITH NARROWBAND AND WIDEBAND CHANNELS

With narrowband transmission, the channel does not change over the coherence bandwidth, and the signal bandwidth should be less than or equal to the channel coherence bandwidth. Basically, an optimal design of DBF depends on the evaluation of SVD of the channel matrix. Specifically, the right singular vectors and water-filling approach are used to evaluate the optimal precoder while the left singular vectors are used to evaluate the combiner. In practice, the assumption that the signal bandwidth is less than or equal to the channel coherence bandwidth in narrowband systems is very hard and can limit the system data rate. Therefore, wideband beamforming systems are proposed to get rid of this technical limitation. With wideband transmission, the channel is decomposed into sub-bands where the channel remains constant over each sub-band, and accordingly narrowband beamforming algorithms can then be applied over each sub-band based on the performance requirements. The augmentation of OFDM and MIMO technologies [80] represents one of the standard wideband beamforming algorithms which requires a full-digital scheme to control the phase and amplitude of the signal for each sub-carrier. However, in HBF, the ABF stage decreases the DoFs of the wideband system.

The phase shifter network produces the same phase shift for all sub-bands, making the wideband HBF problem more challenging [56], [81]–[83].

E. HYBRID BEAMFORMING BASED ON SWITCHES, PHASE SHIFTERS, AND LENS ANTENNA ARRAY

In HBF, the limited number of the RF chains is mapped to a massive number of antennas via an analog network of phase shifters and/or switches. The proposed hybrid architectures for mmWave massive MIMO systems are mostly based on phased arrays. However, in mmWave systems, the phase shifter network can be complex and power consuming [84]–[86]. Besides, in practice, the finite phase shifters used in phased arrays are not sufficiently precise for accurate beam steering. Furthermore, increasing the quantization bits to improve the accuracy incurs more complexity and power consumption. Additionally, employing passive instead of active phase shifters to reduce the power consumption is not feasible due to the higher insertion losses [84]. On the other hand, switches-based architectures can achieve lower power consumption with the same number of RF chains [76], [87]. However, switches-based architectures offer lower array gain when compared to the phase shifters which in turn leads to a reduction in the spectral efficiency. If equal power consumption is considered, switches-based architecture can use more RF chains to improve spectral efficiency. In general, the switches network is lower, but also offers lower spectral efficiency in comparison with the phase shifters structures [43].

The HBF architecture can also be realized using continuous aperture phased MIMO (CAP-MIMO) at the transmitter and receiver. In this architecture, lens antenna arrays are adopted instead of the switches or phase shifters to realize the beamspace MIMO (B-MIMO) [88]. A large lens antenna array is triggered through the array feed, known as the beam selector, which controls the angles of the focused beams generated from the antenna lens. Similar to the spatially sparse precoding, with the help of a limited number of RF chains to select a sub-set of antennas, the CAP-MIMO is able to leverage high gain low-dimensional beam-space, which results from the utilization of the sparse nature of the multi-path mmWave channel.

F. HYBRID BEAMFORMING USING JOINT OR SEPARATE DESIGN FOR DIGITAL AND ANALOG BEAMFORMERS

HBF has two distinguished characteristics inherited from both ABF and DBF, which are: 1) the low number of RF chains that help to reduce the digital processing complexity, and 2) the constant amplitude of phase shifters which help to reduce the complexity of the analog processing and allows a phase-only control. These two characteristics enable us to considerably decrease the overall complexity of massive MIMO mmWave system. It is a challenging task to maximize the spectral efficiency and determine the optimum matrices of both DBF and ABF stages under the aforementioned characteristics/constraints, which

can be addressed by two different types of methods. The first one is to design the analog and digital precoders and combiners with the adoption of joint optimization techniques. The second type adopts two-steps approach, that is in the first step, the analog precoder and combiner are optimized based on a specific criteria, and then the digital precoder and combiner are determined in the second step to improve the performance [93], [94].

VI. HYBRID BEAMFORMING CHALLENGES

With the mmWave massive MIMO system, HBF designs attempt to solve several key challenges as follows:

A. CHANNEL ESTIMATION

The CSI is required to achieve optimal beamforming gains. Channel estimation and beamforming for HBF are more challenging compared to fully digital systems. This is due to the fact that the number of the RF chains is significantly smaller than the number of antennas, and the constraints imposed by the phase shifters turns the design process to non-convex optimization problems that are very hard to solve. Therefore, channel estimation process for HBF based system can be considered as an optimization problem, in which the optimal beams for the transmitter and receiver are known. In the design of channel estimation techniques for HBF based systems, it is required to minimize the channel estimation overhead in the system which limits the spectral efficiency. Compressive sensing methods can be employed to exploit the mmWave channel sparsity in the angular domain to reduce the number of estimated channel parameters and the complexity of the process. This in turn will result in a significant decrease in the dimensions of the channel matrix and the complexity of the precoder and combiner designs. Additionally, in high MDs mobility scenarios, rapid channel variations lead to an estimation delay which it turn can affect the performance.

B. PHASE SHIFTERS CONSTANT MODULUS AND FINITE RESOLUTION

Another challenging point with designing hybrid beamformers arises when the finite resolution phase shifters of constant modulus characteristics are considered, which transform the HBF optimization into a difficult non-convex and combinatorial problem. Besides, since the hybrid beamformer performance depends on the joint optimization of analog and digital beamformers, HBF design schemes are different from those in the conventional MIMO systems beamforming. Actually, the conventional systems consider either fully-ABF or fully-DBF approaches. With digital phase shifters, the optimization problem becomes combinatorial problem with a large search space, especially with massive MIMO systems. The use of a predefined set of phases or codebooks can reduce the search space and complexity. Unfortunately, the codebooks can fit perfectly with only special sets of channel structure, such as sparse channels.

C. OPTIMAL NUMBER OF RF CHAINS

A significant challenge with HBF architecture designs is to determine the optimal number of RF chains while considering multi-stream transmission with large number of antenna elements. The main design goal is to achieve DBF optimal performance considering the hardware complexity and power consumption limitations.

D. PRECODER AND COMBINER DESIGN

Joint design of digital and analog precoders usually aims to maximize the spectral efficiency under both angle quantization and constant amplitude/modulus constraints. It is useful to design algorithms that allow parallel hardware architecture to achieve effective signal processing and very-large-scale integration (VLSI) implementation. Also, it poses a challenge to put forward an optimal joint design of precoder and combiner, and develop algorithms that are able to minimize the inter-cell interference as well as the out-of-cell interference in MU-MIMO systems.

E. BEAM TRAINING AND FEEDBACK

The hybrid precoders and combiners design is typically constrained with the constant-gain analog phase shifters and the low-resolution quantized phase control. Therefore, there is a need to develop optimal beamforming algorithms and codebook designs. Additionally, optimal reference signal for HBF can also decrease the training and feedback overhead, and accordingly helps to achieve a power-efficient and low-complexity system. Different solutions have been considered to circumvent this issue. For instance, user localization is employed in the literature to keep knowing the channel conditions, and different equalization techniques are exploited to maintain the adaptation with dynamically varying channel conditions.

VII. RECENT STUDIES REGARDING HYBRID BEAMFORMING IN THE LITERATURE

In this section, we will classify HBF techniques based on the design criteria. Up to the best of our knowledge, this is the first time to address the HBF taxonomy, which will help the designers of 5G and beyond networks to decide on which HBF approach they should deploy their networks.

A. HBF WITH FULLY AND PARTIALLY-CONNECTED ARCHITECTURES

1) FULLY-CONNECTED ARCHITECTURE

Despite the fact that fully-connected HBF architecture is harder to implement compared with the partially-connected architecture, it is still in use due to its higher spectral efficiency. In [18], an optimization framework is presented to obtain the optimum number of RF chains as well as the number of antenna elements in each transceiver in HBF-based mmWave system, which in turn proves the advantages and disadvantages of fully-connected HBF approach. Specifically, the authors determined the optimal configurations that

maintain an efficient tradeoff between the spectral and energy efficiencies with different number of antennas and RF chains when the static energy consumption is considered. The results could serve as a reference to determine different configuration parameters based on the application requirements. Additionally, the authors in [43] have introduced an antenna selection approach in fully-connected HBF architecture to avoid the channel fading effects suffered with conventional antenna selection approaches. To achieve this, the proposed method requires a variable phase shifter to control the number of RF chains to be involved in the operation. Despite this approach slightly increases the hardware complexity, it is able to achieve higher multiplexing and diversity gains. In [73] the authors attempted to achieve optimum energy efficiency at the BS by reducing the number RF chains and baseband energy consumption. An energy-efficient algorithm is proposed to evaluate the optimum number of RF chains and an efficient tradeoff between the hardware cost and energy efficiency has been achieved. The results illustrated that the joint optimization of the hardware cost and energy efficiency has led to a 170% improvement. The work in [89] has also optimized the energy efficiency through designing a HBF for the downlink of mmWave MU-massive-MIMO system. The proposed architecture employs a two-stage approach. In the first stage, the ABF is employed to improve the link margin and to choose the optimum beam to maximize the desired user power while minimizing the interference of all other users. In the second stage, DBF utilizes a zero-gradient strategy that improves the spatial multiplexing gain and maximizes the energy efficiency of the desired user. Authors in [68], [90] have studied the fully-connected HBF architecture at the uplink and downlink of a massive MIMO system. They considered a finite-resolution phase shifter and a limited RF chain number. Furthermore, they proposed a heuristic algorithm for optimizing the system sum-rate which has been demonstrated that the achievable rate can be increased when the RF chains number is either greater than or equal to the data streams number. Moreover, in [91] a fully connected HBF architecture is also considered for downlink, and its effect on both the data rate and coverage performance in the case of MU-MIMO, SU with spatial multiplexing, and SU with analog beamforming, have been investigated.

The concept of codebook-enabled HBF has been proposed in [83] for mmWave systems with limited feedback between transmitters and receivers. The studies in [48], [60], [88], [92] have proposed the adoption of orthogonal matching pursuit (OMP) and gradient pursuit (GP) algorithms for the fully-connected HBF architectures to achieve higher performance based on the excellent channel estimations. As shown in Fig. 9, the OMP-based HBF approach has provided a comparable performance to that provided by the fully DBF in terms of the spectral efficiency. However, since it is considered as a sparsity constrained matrix completion problem, the HBF design needs to maintain a good tradeoff between the performance loss and the simplicity. Moreover, extra overheads are also incurred to acquire the array response vectors

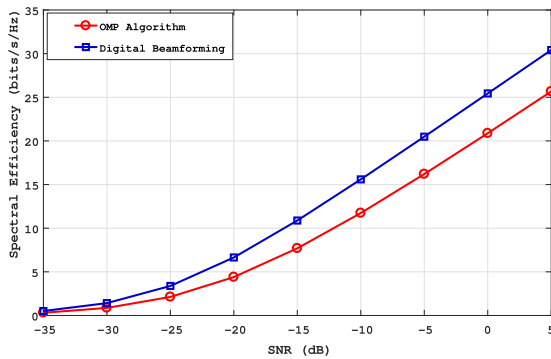


FIGURE 9. HBF based on OMP algorithm versus fully DBF. $N_{RF} = N_S = 3$, number of clusters = 5, number of rays in each cluster = 10, number of transmit antennas = 144, number of receive antennas = 36.

information to exploit the sparsity. In [68] a fully-connected HBF is considered for both the uplink and downlink of mmWave massive MIMO system. The authors demonstrated that the performance of the HBF-based system can equivalently be achieved with fully DBF-based architecture using a fewer number of RF chains. However, it has to send at least twice the number of data streams. In [95], a practical MU massive MIMO HBF system is investigated with ZF precoding which is used in the DBF stage and the beam selection process is executed within the ABF stage. The proposed system showed that the HBF with more RF chains can offer better performance than the fully-digital architecture. In [78], two low-resolution ADCs models are introduced while employing fully-connected HBF architectures in both the uplink and downlink. The two models exploit different channel features, different antenna configurations, and different hardware limitations. A single comparator is also used to implement a 1-bit ADC in the second model to reduce the power consumption in the power-limited mobile devices (MDs), while the first model achieves better performance in the backhaul connections. Finally, an iterative algorithm is proposed in [96] to implement the HBF architecture. In this study, only the discrete set of phase changes can be provided by the phase shifters to improve the spectral efficiency and enable low-cost implementation of the ABF using practical finite resolution phase shifters.

2) PARTIALLY-CONNECTED ARCHITECTURE

In [56], [97], a partially-connected HBF architecture is designed, where the transmitter and receiver are equipped with many antenna sub-arrays, with independent phase shifters, that are used to design the beam steering. In this design, the transmitted signal is adaptively modified based on the mmWave channel characteristics. The partially-connected HBF architecture is also investigated in [98] where the authors have introduced an efficient solution to the peak-to-average-power-ratio (PAPR) problem in massive MIMO-OFDM system under total average transmission power constraints. To realize the optimum achievable rate under transmit power constraints, joint optimization of

the digital and analog beamforming matrices is proposed in [99]. With this architecture, each element of the uniform linear array (ULA) is attached to a different RF chain. The proposed solution provides an enhancement over the beam-steering method with various antenna configurations. Another partially-connected HBF architecture is also proposed in [100] where a low complexity iterative algorithm is used first to design the DBF stage followed by the ABF stage. In [101], the phased arrays based ABF stage is used such that the HBF can efficiently combine or distribute signal energy in sparse mmWave channels while the digital RF chains perform the multiplexing for more flexibility. The study reveals that HBF with phase shifters network achieves better performance with narrowband signals, while the more complex tapped-delay beam steering approach is more suitable for wideband signals. The authors in [102], present a partially-connected HBF that is based on maximizing the spectral efficiency for downlink multi-user MIMO (MU-MIMO) mmWave systems while full CSI is available at both the users and BS. Finally, the authors in [75] proposed a dynamic partially-connected HBF architecture that uses the channel statistics as a criteria for sub-array selection. To achieve optimal sub-array selection, exhaustive search approach is replaced by a greedy algorithm with low complexity which achieves nearly the same spectral efficiency level.

In [103], an iterative hybrid precoding algorithm that is based on the concept of successive interference cancellation (SIC), is proposed for the partially-connected architecture. The authors assumed a diagonal digital precoding matrix that allocates the power to different data streams. Moreover, they assumed an equal number of RF chains and data streams, which makes the beamforming gains for the analog precoder to dominate the gain achieved with the whole HBF architecture. As shown in Fig. 10, the SIC algorithm outperforms the ABF. However, it only provides a sub-optimal solution compared with the fully DBF and HBF. In [104], the authors presented a HBF approach by applying a semi-definite relaxation algorithm and developed an alternating minimization approach, namely SDR-AltMin, to evaluate the optimum values for both the digital and analog precoders. As shown in Fig. 10, the SDR-AltMin algorithm offers a remarkable enhancement compared to ABF. Also, the SDR-AltMin algorithm outperforms the SIC-based algorithm. This is due to the fact that the digital precoder in the semi-definite relaxation is able to fully exploit the capabilities of DBF in steering the transmitted signals in contrary with the SIC-based algorithm which uses the digital precoder to only allocate the power to the data streams.

3) FULLY-CONNECTED AND PARTIALLY-CONNECTED COMPARISON

The fully-connected HBF architecture provides better performance compared to the partially-connected counterpart in terms of the spectral efficiency for any combination of transceiver antennas in the ABF and DBF approaches. However, the energy efficiency of the fully-connected HBF

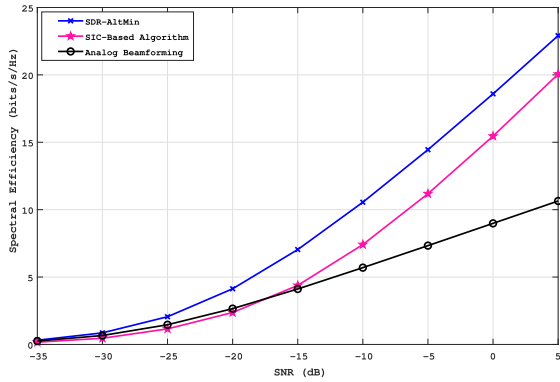


FIGURE 10. The SDR-AltMin algorithm offers remarkable enhancement over ABF. Also, the SDR-AltMin algorithm outperforms the SIC-based algorithm by exploiting the digital precoder functions. The value of $N_{RF} = N_S = 3$, number of clusters = 5, number of rays in each cluster = 10, number of transmit antennas = 144, number of receive antennas = 36.

architecture is lower than that provided with the partially-connected counterpart. The spectral efficiency deterioration caused by applying the partially-connected architecture can be balanced by increasing the number of antenna elements. Accordingly, adjusting the connectivity should be done to maintain an optimum tradeoff between the performance loss and the number of phase shifters (i.e. cost and complexity) in the HBF architecture. The works in [18], [25], [75] provided a comprehensive comparison between fully-connected and partially-connected architectures. Moreover, in [104], an efficient alternating minimization based algorithms are proposed to design the HBF with either full and partially-connected architectures while formulating the HBF design as a matrix factorization problem. The study reveals that the complex fully-connected architecture can achieve the same performance as that given by the fully DBF if the number of RF chains is larger than the number of data streams. Additionally, the partially-connected HBF architecture achieves better spectral efficiency compared with the fully connected counterpart. Moreover, the study also investigated the performance of the HBF design with both narrowband and wideband channels and also with partial CSI. In the partially-connected case, an alternating minimization algorithm based on semi-definite relaxation (SDR-AltMin) is used to design the analog and digital precoders stages. In the fully-connected architecture, the authors proposed to implement the analog precoder stage by taking into consideration the unity modulus constraint and carrying out the optimization over Riemannian manifold to maintain an accurate solution. Furthermore, the authors have proposed an alternating minimization algorithm based on manifold optimization, namely MO-AltMin, for designing the analog precoding stage. Regarding the digital precoding, the authors developed also an alternating minimization algorithm based on phase extraction. It turns out that, the partially-connected HBF architecture achieves higher energy efficiency while the fully-connected architecture provides higher spectral efficiency. On the other hand, the authors in [105] have introduced heuristic algorithms

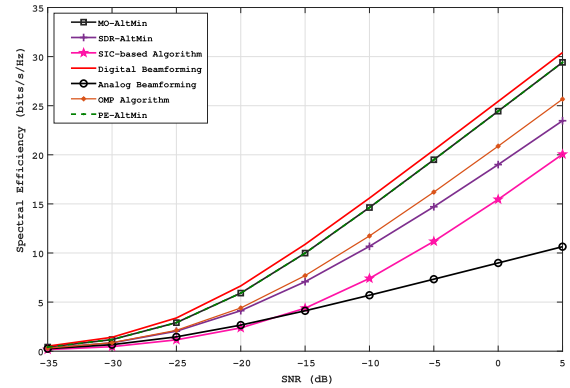


FIGURE 11. Spectral efficiency in (bits/s/Hz) vs SNR in (dB) for different beamforming approaches with $N_{RF} = N_S = 3$, number of transmit antennas = 144, number of receive antennas = 36.

to implement both the digital and analog precoding stages for the HBF. The proposed approach has achieved a near-optimum spectral efficiency for fully-connected architecture with significantly smaller number of RF chains. Another important contribution, that is worth mentioning here, is the work presented in [106]. In this work, an overlapped subarray (OSA) design is proposed for fully and partially-connected HBF architectures. Specifically, a unified low-rank sparse recovery algorithm is used to design the HBF for mmWave MU massive MIMO downlink system. It turns out that, this strategy maintains an optimal trade-off between the achievable performance and system complexity.

In Fig. 11, the spectral efficiency is plotted versus SNR for different HBF approaches with fully ABF and DBF approaches as the baselines. We have assumed an equal number of RF chains and data streams (i.e., $N_{RF}^T = N_{RF}^R = N_S$) as the worst scenario). This is due to the fact that, in practical systems, the number of RF chains cannot be smaller than the number of data streams. In this scenario, for fully-connected architecture, the OMP algorithm performs significantly lower than the optimal fully DBF in terms of the spectral efficiency due to the fact that the OMP algorithm performs poorly when N_{RF} equals N_S . In contrast, the manifold optimization algorithm based on alternating minimization (i.e., MO-AltMin) [104] delivers near-optimal performance over the entire SNR range although the recommended N_{RF} number is not employed (i.e., $N_{RF} \geq 2N_S$). The partially-connected architecture achieves considerable performance gain over the fully ABF with the alternating minimization based SDR algorithm (SDR-AltMin) [104]. However, the achievable spectral efficiency is much higher with higher SNR levels. The SIC-based approach in [103] is also included in the comparison as a benchmark for the partially-connected architecture. As shown in Fig. 11, the SDR-AltMin algorithm outperforms the SIC-based algorithm since the SDR-AltMin algorithm fully exploits the digital precoder in contrast with the SIC-based technique which only utilizes the digital part to distribute the power to the available data-streams.

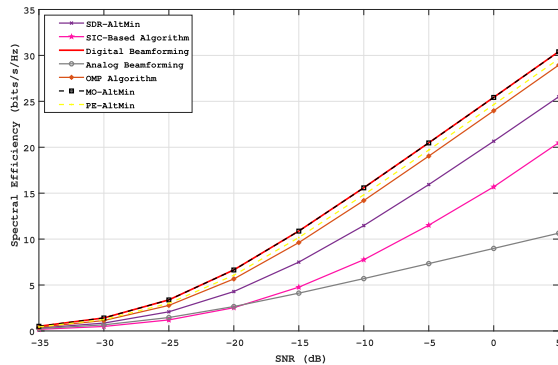


FIGURE 12. Spectral efficiency in (bits/s/Hz) vs SNR in (dB) for different classifications of beamforming with $N_{RF} = 6$, $N_S = 3$, number of transmit antennas = 144, number of receive antennas = 36.

In Fig. 12, the spectral efficiency versus SNR is provided for different beamforming approaches, assuming the optimal values of N_S and N_{RF} (i.e., $N_{RF} \geq 2N_S$). It is noticed that the performances of SDR, OMP, PE-AltMin, and MO-AltMin approaches are improved due to the increase in N_{RF} . The performance gap between OMP and alternating-based algorithms shrinks and the three iterative algorithms achieve a near-optimal performance over the entire SNR range.

B. HYBRID BEAMFORMING BASED ON FULL CSI AND ESTIMATED CSI

In the literature, it is observed that the presented strategies for implementing HBF are generally based on either estimating the CSI or assuming the availability of full CSI at all the network nodes. In the following two sub-sections, we explain the concepts behind both the full and estimated CSI and their impact on the design of HBF.

1) FULL CHANNEL STATE INFORMATION

The two main approaches to achieve practical near-optimal solutions and overcome the challenges discussed in section VI are addressed as follows:

- 1) Optimal beamformer approximation: In the fully digital structure, optimal beamforming for SU-MIMO can be obtained through evaluating the SVD of the channel matrix. Another method based on eigen-value decomposition is proposed in [43]. A different approach is also addressed in [48] which obtains near-optimal hybrid beamformers by reducing the Euclidean distance to the fully digital solution. In mmWave sparse channels, reducing the Euclidean distance produces a quasi-optimal solution. In the traditional non-sparse channels, analog and digital beamformers can be evaluated using alternating optimization based algorithms.
- 2) Decoupling the design of the analog and digital beamformers: the coupling between analog and digital beamformers, as well as the coupling between the transceivers, are major challenges in HBF. To reduce the complexity, decoupling the beamformers is a critical issue, which can be achieved sequentially. For

example, we can cancel the effect of the combiner on the precoder by assuming a fully digital MMSE-based receiver to increase the sum rate for SU-MIMO, and then assuming a unitary-based fully digital precoder to decouple the analog and digital precoders. Finally, the precoder matrix is optimized column-by-column by forcing the phase-only constraint on each antenna and a closed-form formulation of the digital precoder can then be obtained. Several different decoupling approaches have also been investigated in [25].

2) ESTIMATED CHANNEL STATE INFORMATION

Using an estimated CSI for the analog part in HBF is initially proposed in [107], where a near-optimal beamformer for SU-MIMO systems is designed. The authors in [79] also developed a joint spatial division multiplexing (JSDM) design for MU-MIMO system with HBF employed at the BS which serves single-antenna MD. In this scheme, the MDs sharing the same transmit channel covariance are grouped together in a cluster to reduce the training and feedback overheads. But, the proposed JSDM approach cannot deal efficiently with the interference between MDs within the same cluster. In [108], the JSDM design was extended to overcome the interference problem and increased the average sum-rate by proposing a modified MMSE algorithm.

The majority of the channel estimation methods in HBF systems exploit the sparsity of the mmWave channels in the angular domain [34], [48], [52], [61], [64], [109]–[113]. This is because, the sparsity in the angular domain allows the channels to be modeled by AoA, AoD, and the gain of each path. The sparsity feature also facilitates the use of compressive sensing methods in the channel estimation to estimate the elements of channel with a small number of measurements, resulting in a reduced channel matrix dimension [110]. In general, the channel estimation methods can be classified into open-loop or closed-loop, depending on whether a feedback connection is used or not [111]. In the closed-loop method, the receiver determines the best elements from a predefined codebook and feeds them back to the transmitter [34], [52]. In this case, the selection of the codebook can significantly affect system performance. Additionally, large codebook size allows the formation of precise and sharp beams while increasing the feedback overhead. Therefore, the impact of the RF codebooks has been investigated by many studies in the literature [34], [48], [52], [64], [109].

The studies in [48], [64] show that a simple codebook can be designed by consistently sampling the beam steering space. But, the downside of this scheme is that a resolution of 6 – 7 bits for each phase shifter is required for acceptable performance. Since high-resolution phase shifters suffer from high insertion losses, the codebooks with low-resolution phase shifters are favored in practice. In [112], an efficient hybrid codebook is proposed by employing OMP which needs an analog phase shifters and its achieved performance is based on the number of the RF chains. In [61], the closed-loop multi-resolution channel estimation is proposed where

wide beams with low directivity gains are used at both the transmitter and receiver at the beginning of the channel estimation. Subsequently, a multi-step power allocation is employed where a high transmit power is used for wide-beam and slightly reduced as the beam gets sharper. The main drawback of the multi-resolution channel estimator is the fact that the number of training steps scales up with the number of the multi-path components and the initial high value of power required at the initial operation. In [113], the authors have extended the channel estimation method in [61] to one-sided search which reduces the initial transmit power. Also, the proposed method is based on the ping-pong iterations which eliminated the need for a separate feedback connection. In [111], another open-loop channel estimation method for HBF is proposed, which is based on the multi-grid OMP. Additionally, the authors have proven that the proposed algorithm performs better than the least-squares method with much lower complexity. A different approach has been proposed in [107] to minimize the channel estimation delay and signaling overhead in SU-MIMO HBF by changing the analog beamformer according to the varying parameters of the channel. Moreover, in [79], [114], a single-antenna MU with a BS that employs HBF is considered, where the authors propose a joint spatial division multiplexing method that minimizes the downlink training and uplink feedback overhead. Specifically, the user channels are arranged into groups based on their covariance matrices, and the analog beamformer is devised such that the inter-group interference is minimized, and eventually, the digital precoder separates the user signals in each group.

C. HYBRID BEAMFORMING WITH NARROWBAND AND WIDEBAND CHANNELS

1) NARROWBAND CHANNELS

The HBF with a fully-connected architecture for narrowband SU system is initially investigated in [43]. The study has proven that, initially, when the number of RF chains is twice the number of data-streams, fully digital beamforming performance can be reached by the HBF design. Then, the optimal analog beamforming is designed based on the phase shift corresponding to the elements of the right and left singular vectors of the strongest singular value of the channel matrix. The study showed that when many symbols are transmitted over both uncorrelated and correlated channels, the same performance can be achieved with HBF and fully DBF approaches, respectively. The study in [77], has demonstrated that a beam steering method towards the AoA and AoD can achieve the same performance achieved by the fully-DBF when the number of transceiver antennas is very large. This results from the convergence between the the singular vectors of the channel matrix and the steering vectors towards the AoAs and AoDs. In [48], [54], the joint optimization design of the digital and analog beamformers is proposed based on the matching pursuit method exploiting the mmWave channel sparsity. Initially, the channel singular vectors are calculated,

and then, the hybrid beamformer is determined by reducing Euclidean distance between the singular vectors and hybrid beamformer weights. Since the calculations of the singular vectors require intensive computations, the computations can cause critical delays in practice. Furthermore, in [48], it was proven that the spectral efficiency depends significantly on the number of RF chains in the system and the multi-path components of the channel. The studies in [68], [90], [115] considered the iterative HBF algorithms and achieved near-optimal performance in both sparse and rich scattering channels. Also, studying the tradeoff between beamforming and multiplexing gains is investigated in [116] using iterative precoding design. The proposed algorithm has achieved the channel capacity at low-SNR and full multiplexing gain at high-SNR. Moreover, the design has improved the spectral efficiency by deciding whether to use an additional sub-array to increase the SNR or to transmit a new stream. By exploiting the sparsity of the mmWave channel, the authors in [69] have proposed a codebook-based partially-connected HBF design. However, the codebook-based approaches are only applicable for specific channels and fixed array geometry, meanwhile, a new codebook should be designed when changing either the channel or the geometry.

2) WIDEBAND CHANNELS

For wideband mmWave systems, a codebook-based HBF for SU scenario is proposed in [83]. The proposed method applies Gram-Schmidt orthogonalization for designing the HBF matrices. Additionally, the authors of [56], have applied exhaustive search over a specific codebook for designing the beamforming vectors. Furthermore, a channel estimation algorithm for a single-stream transmission in mmWave MIMO-OFDM system with the partially-connected HBF is also developed. In [81], the space-division multiple-access and orthogonal frequency-division multiple-access (SDMA-OFDMA) schemes are leveraged for downlink transmission in MU massive MIMO to maximize the sum-rate. Nevertheless, it calls for more efforts to develop low complexity algorithms for HBF in wideband channels. A combination of frequency scheduling and HBF is expected to be essential in the future research [25], [28].

D. HYBRID BEAMFORMING BASED ON SWITCHES, PHASE SHIFTERS, AND LENS ARRAY ARCHITECTURES

There are many studies investigating the energy and spectral efficiency of the HBF implemented in both phase shifter or switches [74], [76], [87], [117]. Most of these studies tried to determine several key design issues to achieve optimal energy and spectral efficiency, including, the most energy-efficient structure, the optimum number of antennas to realize maximum energy efficiency, and the performance of either phase shifters or switches in terms of energy efficiency. The energy and spectral efficiency of phase shifters based HBF has been intensively studied in [43], [48], [74], [76], [87], [117], [118]. In [118], the insertion losses of the phase shifters based HBF schemes have been assessed

under the specified constraints of both the energy and spectral efficiency.

The works in [63], [119] proposed an HBF architecture where a combination of switches and phase shifters are employed. A phase shifters selection method has been proposed in [63] to switch-off half of the phase shifters without degrading the spectral efficiency and consequently reduces the energy consumption significantly. Based on the same concept, the authors in [119], have developed a new design by employing a switching network to reduce the active number of phase shifters while preserving the same spectral efficiency. It has proven that the proposed design is also applicable in the sparse scattering and correlated Rayleigh fading channels. However, both studies only investigated the HBF spectral efficiency without giving much attention to the energy efficiency. In [120], the energy efficiency with distinct structures of switches and phase shifters are investigated for the HBF-based massive MIMO systems. In [63], [119], the authors have derived a closed-form expression for the energy efficiency as a function of the spectral efficiency, enabling a concurrent examination of spectral and energy performance for each structure. This in turn leads to an efficient design of HBF that adjusts smart trade-off between the spectral and energy efficiencies. A beamspace HBF and digital precoder acquisition algorithms that based on compressive sensing for SU mmWave MIMO system are proposed in [121]. The beamspace algorithm has showed a 99.4% complexity reduction compared to the digital precoder acquisition algorithm. However, this study lacks any contributions for MU scenario. The study in [88] integrates the hybrid transceiver with beamspace MIMO to create continuous aperture phased MIMO (CAP-MIMO), which is able to achieve a near-optimal performance. Also, in [122], beam selection and beamspace MIMO concepts are integrated to achieve a near-optimal performance. Additionally, the beam selection concept in [122], has been formed based on different criteria such as, system capacity, path loss, receiver SINR, and minimum error rate. Moreover, to investigate the advantages of beamspace MIMO in terms of capacity improvement, the number of RF chains, spectral and energy efficiencies are also considered in [123]–[125].

In HBF, the digital stage includes the power-greedy DAC and ADC at the transmitter and receiver, respectively. The analog stage consists of a network of phase shifters or switches. In practice, the phase shifters network is implemented using finite resolution DAC, ADC, and phase shifters. Additionally, the higher mmWave frequencies inflict high power consumption due to the required high bit-resolution and sampling rate [123]. In [124], the HBF complexity for ADCs resolution is investigated. The study reveals that low-resolution ADC (e.g. 1-bit ADC) has gained considerable attention due to its energy efficiency. However, it also allows rate loss and requires long training sequences for channel estimation. Nevertheless, this study did not address other hardware components such as switches or phase shifters within different HBF architectures. In [78], fully-digital and

HBF with 1-bit ADC resolution have been addressed. The authors have showed that utilizing low-resolution ADC is able to achieve comparable performance at low and medium SNR values. Because of the low power consumption, low-resolution ADCs can also be employed at the downlink in the power-limited MDs while using the HBF at the uplink. In [125], the authors have shown that employing 5-6 bits ADC utilization can achieve almost the same performance level as the infinite ADC assuming full-resolution phase shifters in the uplink of MU massive MIMO systems. The study in [126] also shows that employing a 4-5 bits finite resolution HBF achieves better performance in terms of spectral and energy efficiency in comparison with the 1-bit ADC full-digital beamforming. In [127], the authors have proved that the 1 – 2 bits resolution fully DBF outperforms the HBF at low SNR, while 3 – 5 bits resolution HBF achieves the optimal spectral efficiency and power consumption for a wide range of SNR. Finally, the authors in [128] have studied the trade-off between spectral and energy efficiencies for analog, digital, and hybrid precoding/combining designs. It turns out that, at low SNR, fully analog precoding/combining delivers the best spectral and energy efficiencies. When low-bits ADC fully DBF is compared to HBF, the fully DBF achieves better spectral efficiency.

VIII. OPEN RESEARCH ISSUES

Although different solutions to the HBF design in massive MIMO mmWave systems have been explored, the following issues have not been sufficiently addressed.

- **Dual-band small cells:** Small cells with dual access can work in licensed and unlicensed bands and anticipated as a fundamental part of 5G networks. So far, it is not clear how traffic should be directed from the licensed band to the unlicensed band and how such integration influences the HBF design [129].
- **Antenna selection optimization:** The absence of full CSI is a challenge in the HBF context. More investigations are needed for the RF chains and throughput optimization in the presence of CSI errors and mmWave wideband channels [35], [61]. Besides, more efforts are also needed in the optimal selection of antenna arrays [130].
- **Handover and mobility management:** Mobility control and handover are critical issues in HBF mmWave systems due to the frequent hand-over in mmWave band imposed by its blockage prone nature [131].
- **Interference management:** Current conventional interference management methods for omni-directional transmission are not suitable for directional beamforming transmission. HBF also affects the interference levels because of its different architecture. The effect for handling these issues have not been sufficiently investigated [132].
- **Beamforming for green communications:** The research on 5G and beyond networks is mainly focused

on satisfying the huge traffic demand and provides seamless connectivity. It is also important to consider the design of energy-efficient systems with minimal footprint [133], [134]. The joint design of energy and spectral efficiencies attracts the most attention [135]. To further leverage the energy efficiency offered by the hybrid beamforming architecture, many studies attempted to determine the optimal tradeoff between the numbers of RF chains and antenna elements [136]. The goal is to optimize system parameters such that the energy efficiency for a specified spectral efficiency can be maximized. The joint design of the digital and analog domains for energy-efficiency and low-complexity, is an interesting point to consider in future research.

- **Hybrid beamforming with switches or phase shifters:** In this paper, we presented different combinations of phase shifters and switches, and addressed the corresponding trade-off between energy and spectral efficiencies. With ideal assumptions, such as neglecting the phase shifters and switches insertion losses, this architecture can offer improved performance. However, further efforts are needed to investigate the feasibility of such architectures in practice and the tradeoff between energy and spectral efficiencies when the insertion losses are considered.
- **Analysis of the channel estimation in hybrid beamforming:** In practice, the channel varies over time and channel estimation should be repeated after the channel coherence time is passed. Investigating the tradeoffs between the acceptable estimation errors, the signaling overhead and data transmission within the channel coherence time is essential to realize the HBF.
- **Deep learning based hybrid beamforming algorithms:** Recently, many studies exploited the machine learning algorithms to solve many traditional communication problems, such as implementing autoencoders at both the transmitter and receiver, signal representation prediction, modulation classification, and channel estimation [137], [138]. Additionally, there are also some studies that manipulated the problem for HBF for simple scenarios [139]–[141]. More efforts should be exerted to solve the joint optimization problems of channel estimation, hybrid precoding, and scheduling. Also, employing deep learning based HBF algorithms in complex scenarios such as UAV-assisted communication scenarios, cell-free massive MIMO scenarios, and NOMA-based backscattering communication scenarios should also be given much more attention in the future.

IX. CONCLUSION

HBF systems have been presented a decade ago but they are given much attention in the past few years due to their essential role in enabling energy and cost-efficient massive MIMO and mmWave systems in 5G networks. HBF combined with massive MIMO and mmWave can solve many technical problems for the 5G network varying from the

capacity improvement to the energy efficiency. In this paper, we have presented a comprehensive classification for HBF based on different criterias and scenarios. We also surveyed the most recent works in each of these classifications. Based on the discussions presented in this work, it is evident that there is no single HBF architecture that can provide the best trade-off between performance, complexity, and cost. Consequently, to obtain the best performance out of the hybrid beamforming with the lowest cost and complexity, it is required to dynamically design the architecture according to the channel characteristics and the intended applications.

REFERENCES

- [1] A. Alexiou, "5G vision, enablers and challenges for the wireless future," *Wireless World Res. Forum*, vol. 2015, no. 16, pp. 1–12, Apr. 2015.
- [2] M. Carugi, *Key Features and Requirements of 5G/IMT-2020 Networks*, document 16, International Telecommunication Union, 2018.
- [3] *Series Y: Global Information Infrastructure, Internet Protocol Aspects, Next-Generation Networks, Internet of Things and Smart Cities Future networks*, document ITU-T Y Suppl.53, ITU-T, Telecommunication Standardization Sector OF ITU, 2018.
- [4] *5G: A Technology Vision*, Huawei, Shenzhen, China, 2014, pp. 1–16.
- [5] W. Haerick and M. Gupta, "5G and the factories of the future," 5G PPP, Brussels, Belgium, White Paper, 2014, pp. 1–31. Accessed: Nov. 2019. [Online]. Available: <https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf>
- [6] L. Ericsson. (2011). *More Than 50 Billion Connected Devices*. Accessed: Nov. 2019. [Online]. Available: https://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf
- [7] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [8] H. Ericsson. (2015). *The Road to 5G—Drivers, Applications, Requirements and Technical Development*. Qualcomm. Accessed: Nov. 2019. [Online]. Available: https://www.huawei.com/minisite/5g/img/GSA_the_Road_to_5G.pdf
- [9] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
- [10] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.
- [11] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.
- [12] A. Bhattacharjee, R. Bhattacharjee, and S. K. Bose, "Mitigation of beam blocking in mmWave indoor WPAN using dynamic control delegation based approach," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Nov. 2016, pp. 1–6.
- [13] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 63–69, Sep. 2014.
- [14] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [15] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [16] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [17] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [18] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

- [19] G. R. MacCartney, M. K. Samimi, and T. S. Rappaport, "Exploiting directionality for millimeter-wave wireless system improvement," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2416–2422.
- [20] J. G. Andrews, T. Bai, M. Kulkarni, A. Alkhateeb, A. Gupta, and R. W. Heath, Jr., "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Oct. 2017.
- [21] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [22] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.
- [23] T. E. Bogale and L. B. Le, "Massive MIMO and millimeter wave for 5G wireless HetNet: Potentials and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Feb. 2016.
- [24] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [25] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
- [26] H. Yang, M. H. A. J. Herben, I. J. A. G. Akkermans, and P. F. M. Smulders, "Impact analysis of directional antennas and multi-antenna beamformers on radio transmission," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1695–1707, May 2008.
- [27] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [28] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [29] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO—5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2nd Quart., 2016.
- [30] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1738–1760, 3rd Quart., 2015.
- [31] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [32] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.
- [33] A. K. Gotsis and N. J. Sahalos. *Beamforming in 3G and 4G Mobile Communications: The Switched-Beam Approach, Recent Developments in Mobile Communications—A Multidisciplinary Approach*. Accessed: Dec. 2019. [Online]. Available: <https://www.intechopen.com/books/recent-developments-in-mobile-communications-a-multidisciplinary-approach/beamforming-in-3g-and-4g-mobile-communications-the-switched-beam-approach>
- [34] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [35] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 4066–4071.
- [36] H. Ren, B. Arigong, M. Zhou, J. Ding, and H. Zhang, "A novel design of 4×4 Butler matrix with relatively flexible phase differences," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 1277–1280, Dec. 2016.
- [37] J. Bae, Y. S. Choi, J. S. Kim, and M. Y. Chung, "Architecture and performance evaluation of mmWave based 5G mobile communication system," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2014, pp. 847–851.
- [38] Z.-E. Huang and J.-Y. Pan, "Coordinative switch beamforming scheduler for guaranteed service with service area subsectorization in next generation cellular network," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1000–1005.
- [39] N. Tiwari and T. R. Rao, "A switched beam antenna array with Butler matrix network using substrate integrated waveguide technology for 60 GHz communications," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 2152–2157.
- [40] L. Huang, B. Zhang, and Z. Ye, "Robust adaptive beamforming using a new projection approach," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1181–1185.
- [41] X. Wu, Y. Cai, R. C. de Lamare, B. Champagne, and M. Zhao, "Adaptive blind widely linear CCM reduced-rank beamforming for large-scale antenna arrays," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 5–9.
- [42] E. Ali, M. Ismail, R. Nordin, and N. F. Abdulah, "Beamforming techniques for massive MIMO systems in 5G: Overview, classification, and trends for future research," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 6, pp. 753–772, Jun. 2017.
- [43] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.
- [44] S. Mahboob, R. Ruby, and V. C. M. Leung, "Transmit antenna selection for downlink transmission in a massively distributed antenna system using convex optimization," in *Proc. 7th Int. Conf. Broadband, Wireless Comput., Commun. Appl.*, Nov. 2012, pp. 228–233.
- [45] A. F. Molisch, M. Z. Win, C. Yang-Seok, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, Sep. 2005.
- [46] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, Sep. 2006.
- [47] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.
- [48] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [49] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.
- [50] J. Via, I. Santamaria, V. Elvira, and R. Eickhoff, "A general criterion for analog TX-RX beamforming under OFDM transmissions," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2155–2167, Apr. 2010.
- [51] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.
- [52] J. Wang, Z. Lan, C.-w. Pyo, T. Baykas, C.-s. Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
- [53] F. Gholam, J. Via, and I. Santamaria, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2373–2378, Apr. 2011.
- [54] O. E. Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3724–3729.
- [55] *IEEE Standard for Information Technology—Local and Metropolitan Area Networks—Specific requirements—Part 15.3: Amendment 2, Millimeter-Wave-Based Alternative Physical Layer Extension*, IEEE Standard 802.15.3c-2009, Oct. 2009, pp. 1–200.
- [56] C. Kim, T. Kim, and J.-Y. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 61–65.
- [57] T. E. Bogale, L. B. Le, A. Haghigat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.
- [58] D. Ying, F. W. Vook, T. A. Thomas, and D. J. Love, "Hybrid structure in massive MIMO: Achieving large sum rate with fewer RF chains," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1–6.
- [59] F. Khan, Z. Pi, and S. Rajagopal, "Millimeter-wave mobile broadband with large scale spatial processing for 5G mobile communication," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 1517–1523.
- [60] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

- [61] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [62] X. D. W. Ni and W. S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.
- [63] S. Payami, M. Ghorashi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.
- [64] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [65] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.
- [66] N. Song, H. Sun, and T. Yang, "Coordinated hybrid beamforming for millimeter wave multi-user massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [67] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.
- [68] F. Sahrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [69] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.
- [70] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, Jun. 2015.
- [71] H. Quoc Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [72] A. Alkhateeb, Y.-H. Nam, J. Zhang, and R. W. Heath, Jr., "Massive MIMO combining with switches," *IEEE Wireless Commun. Lett.*, vol. 5, no. 3, pp. 232–235, Jun. 2016.
- [73] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758–771, Apr. 2016.
- [74] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [75] S. Park, A. Alkhateeb, and R. W. Heath, Jr., "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [76] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 90–97.
- [77] O. E. Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2012, pp. 100–104.
- [78] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [79] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [80] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.
- [81] L. Kong, S. Han, and C. Yang, "Wideband hybrid precoder for massive MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 305–309.
- [82] S. Park and R. W. Heath, Jr., "Frequency selective hybrid precoding in millimeter wave OFDMA systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [83] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [84] A. S. Y. Poon and M. Taghivand, "Supporting and enabling circuits for antenna arrays in wireless communications," *Proc. IEEE*, vol. 100, no. 7, pp. 2207–2218, Jul. 2012.
- [85] Y. Zheng and C. E. Saavedra, "An ultra-compact CMOS variable phase shifter for 2.4-GHz ISM applications," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 6, pp. 1349–1354, Jun. 2008.
- [86] N. A. Talwalkar, C. P. Yue, H. Gan, and S. S. Wong, "Integrated CMOS transmit-receive switch using LC-tuned substrate bias for 2.4-GHz and 5.2-GHz applications," *IEEE J. Solid-State Circuits*, vol. 39, no. 6, pp. 863–870, Jun. 2004.
- [87] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [88] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [89] J. Jing, C. Xiaoxue, and X. Yongbin, "Energy-efficiency based downlink multi-user hybrid beamforming for millimeter wave massive MIMO system," *J. China Univ. Posts Telecommun.*, vol. 23, no. 4, pp. 53–62, Aug. 2016.
- [90] F. Sahrabi and W. Yu, "Hybrid beamforming with finite-resolution phase shifters for large-scale MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 136–140.
- [91] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.
- [92] A. Kaushik, J. Thompson, and M. Yaghoobi, "Sparse hybrid precoding and combining in millimeter wave MIMO systems," in *Proc. Radio Propag. Technol. for 5G*, Oct. 2016, pp. 1–7.
- [93] J. Palacios, D. De Donno, D. Giustiniano, and J. Widmer, "Speeding up mmWave beam training through low-complexity hybrid transceivers," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–7.
- [94] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2013, pp. 1–5.
- [95] Y. Ren, Y. Wang, C. Qi, and Y. Liu, "Multiple-beam selection with limited feedback for hybrid beamforming in massive MIMO systems," *IEEE Access*, vol. 5, pp. 13327–13335, 2017.
- [96] J.-C. Chen, "Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7604–7608, Aug. 2017.
- [97] J. Singh and S. Ramakrishna, "On the feasibility of beamforming in millimeter wave communication systems with multiple antenna arrays," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 3802–3808.
- [98] C. Kim, J.-S. Son, T. Kim, and J.-Y. Seol, "On the hybrid beamforming with shared array antenna for mmWave MIMO-OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 335–340.
- [99] Z. Xu, S. Han, Z. Pan, and C.-L. I, "Alternating beamforming methods for hybrid analog and digital MIMO transmission," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1595–1600.
- [100] G. Wang and G. Ascheid, "Hybrid beamforming under equal gain constraint for maximizing sum rate at 60 GHz," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [101] J. Zhang, X. Huang, V. Dyadyuk, and Y. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.
- [102] R. A. Stirling-Gallacher and M. S. Rahman, "Multi-user MIMO strategies for a millimeter wave communication system using hybrid beam-forming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2437–2443.
- [103] L. Dai, X. Gao, J. Quan, S. Han, and C.-L. I, "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1–6.
- [104] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

- [105] F. Sahrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.
- [106] N. Song, T. Yang, and H. Sun, "Overlapped subarray based hybrid beamforming for millimeter wave multiuser massive MIMO," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 550–554, May 2017.
- [107] P. Sudarshan, N. Mehta, A. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501–3511, Dec. 2006.
- [108] Z. Li, S. Han, and A. F. Molisch, "Hybrid beamforming design for millimeter-wave multi-user massive MIMO downlink," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
- [109] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Dec. 2011, pp. 253–257.
- [110] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.
- [111] J. Lee, G.-T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 3326–3331.
- [112] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1298–1303.
- [113] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Single-sided adaptive estimation of multi-path millimeter wave channels," in *Proc. IEEE 15th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 125–129.
- [114] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, Jun. 2014.
- [115] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3779–3784.
- [116] O. El Ayach, R. W. Heath, Jr., S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3476–3480.
- [117] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog–Digital transceivers for single- and multi-carrier large antenna array systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.
- [118] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog–digital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528–531, Oct. 2016.
- [119] S. Payami, M. Ghoraiishi, M. Dianati, and M. Sellathurai, "Hybrid beamforming with a reduced number of phase shifters for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4843–4851, Jun. 2018.
- [120] S. Payami, N. Mysore Balasubramanya, C. Masouros, and M. Sellathurai, "Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 13–16, Feb. 2019.
- [121] C.-H. Chen, C.-R. Tsai, Y.-H. Liu, W.-L. Hung, and A.-Y. Wu, "Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1412–1424, Mar. 2017.
- [122] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.
- [123] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [124] D. C. Araújo, T. Maksymyuk, A. L. F. de Almeida, T. Maciel, J. C. M. Mota, and M. Jo, "Massive MIMO: Survey and future research topics," *IET Commun.*, vol. 10, no. 15, pp. 1938–1946, Oct. 2016.
- [125] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.
- [126] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, Jr., "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.
- [127] K. Roth and J. A. Nossek, "Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2056–2068, Sep. 2017.
- [128] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.
- [129] J. Zhang, A. Wiesel, and M. Haardt, "Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3281–3285.
- [130] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, Jun. 2016.
- [131] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.
- [132] J. Zhang, M. Haardt, I. Solovychik, and A. Wiesel, "A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
- [133] C. Hu, J. Liu, X. Liao, Y. Liu, and J. Wang, "A novel equivalent baseband channel of hybrid beamforming in massive multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 764–767, Apr. 2018.
- [134] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.
- [135] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.
- [136] S. Han, C. Lin I, C. Rowell, Z. Xu, S. Wang, and Z. Pan, "Large scale antenna system with hybrid digital and analog beamforming structure," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2014, pp. 842–847.
- [137] J. Tao, Q. Wang, S. Luo, and J. Chen, "Constrained deep neural network based hybrid beamforming for millimeter wave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
- [138] A. M. Elbir and K. V. Mishra, "Deep learning design for joint antenna selection and hybrid beamforming in massive MIMO," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting*, Jul. 2019, pp. 1585–1586.
- [139] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, Jul. 2019.
- [140] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," 2019, *arXiv:1905.13212*. [Online]. Available: <http://arxiv.org/abs/1905.13212>
- [141] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," 2019, *arXiv:1907.00123*. [Online]. Available: <http://arxiv.org/abs/1907.00123>



MOHAMED RIHAN (Member, IEEE) received the B.Sc. degree (Hons.) in electronics and communication engineering from Menoufia University, Egypt, and the M.Sc. and Ph.D. degrees in electronics and communication engineering from the Egypt-Japan University of Science and Technology, in 2012 and 2015, respectively. From 2014 to 2015, he was a Researcher with the Department of Advanced Information Technology, Graduate School of ISEE, Kyushu University, Japan.

From 2016 to 2017, he was an Adjunct Professor with the Center of Photonics and Smart Materials, Zewail City of Science and Technology, Egypt. He is currently a Postdoctoral Research Fellow with the College of Information Engineering, Shenzhen University, China. He is also serving as an Assistant Professor with the Faculty of Electronic Engineering, Menoufia University. His research interests include massive MIMO, mmWave communications, interference alignment, resource allocation, cognitive heterogeneous networks, and applications of signal processing in wireless communications.



TAREK ABED SOLIMAN received the B.S. degree in electronic and communication engineering from the Delta Higher Institute of Engineering and Technology, Egypt, in 2012. He is currently pursuing the M.S. degree in electrical engineering with Menoufia University, Egypt. He is also working with Vodafone Egypt, for more than five years, as a Technical Project Coordinator. His research interests include massive MIMO, mmWave communications, hybrid beamforming, and network resource allocation for 5G networks.

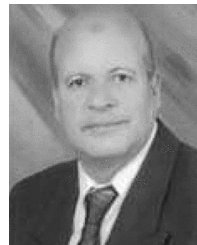


CHEN XU received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1986 and 1989, respectively, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, in 1992. In 1992, he joined Shenzhen University, Shenzhen, China, where he is currently a Professor. From 1999 to 2000, he was a Research Fellow with Kansai University, Suita, Japan. From 2002 to 2003, he was a Research Fellow with the University of Hawaii, Honolulu, HI, USA. His current research interests include image processing, intelligent computing, and wavelet analysis.



LEI HUANG (Senior Member, IEEE) was born in Guangdong, China. He received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

From 2005 to 2006, he was a Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong, and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong. From 2011 to 2014, he was a Professor with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School. Since November 2014, he has been with the College of Information Engineering, Shenzhen University, where he is currently a Distinguished Professor and the Director of the Shenzhen Key Laboratory of Advanced Navigation Technology. His research interests include array signal processing, statistical signal processing, sparse signal processing and their applications in radar, navigation, and wireless communications. He was elected an IET Fellow, in 2018. He is also on the Editorial Boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *Digital Signal Processing*, and *IET Signal Processing*. He is also an elected member of Sensor Array and Multichannel (SAM) Technical Committee (TC) of the IEEE Signal Processing Society.



MOAWAD I. DESSOUKY received the B.Sc. (Hons.) and M.Sc. degrees from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 1976 and 1981, respectively, and the Ph.D. from McMaster University, Canada, in 1986. He joined as a Teaching Staff with the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, in 1986. He has published more than 140 scientific articles in national and international conference proceedings and journals. He is currently the Head of the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University. His current research areas of interest include spectral estimation techniques, image enhancement, image restoration, super-resolution reconstruction of images, satellite communications, and spread spectrum techniques. He has received the most cited paper award from *Digital Signal Processing Journal*, in 2008.

...