Millimeter-Wave Massive MIMO Communication for Future Wireless Systems: A Survey

Sherif Adeshina Busari, *Student Member, IEEE*, Kazi Mohammed Saidul Huq, *Senior Member, IEEE*, Shahid Mumtaz, *Senior Member, IEEE*, Linglong Dai, *Senior Member, IEEE*, and Jonathan Rodriguez, *Senior Member, IEEE*

***Abstract*—Several enabling technologies are being explored for the fifth-generation (5G) mobile system era. The aim is to evolve a cellular network that remarkably pushes forward the limits of legacy mobile systems across all dimensions of performance metrics. One dominant technology that consistently features in the list of the 5G enablers is the millimeter-wave (mmWave) massive multiple-input-multiple-output (massive MIMO) system. It shows potentials to significantly raise user throughput, enhance spectral and energy efficiencies and increase the capacity of mobile networks using the joint capabilities of the huge available bandwidth in the mmWave frequency bands and high multiplexing gains achievable with massive antenna arrays. In this survey, we present the preliminary outcomes of extensive research on mmWave massive MIMO (as research on this subject is still in the exploratory phase) and highlight emerging trends together with their respective benefits, challenges, and proposed solutions. The survey spans broad areas in the field of wireless communications, and the objective is to point out current trends, evolving research issues and future directions on mmWave massive MIMO as a technology that will open up new frontiers of services and applications for next-generation cellular networks.**

***Index Terms*—5G, channel estimation, channel feedback, channel measurement, channel modeling, massive MIMO, mmWave, precoding, propagation, wireless fronthaul.**

I. Introduction

AS A TRADITION since the early 1980's, operators and regulators of mobile wireless communication systems roll out a new generation of cellular networks almost every decade. The year 2020 is expected to herald a new dawn with the introduction and possible commercial deployment of the fifth-generation (5G) cellular networks that will significantly outperform prior generations (i.e., from the first generation (1G) to the fourth generation (4G)). Widespread adoption of 5G networks is anticipated by 2025. The 5G era is foreseen to usher in next-generation mobile networks (NGMNs) that will deliver super-speed connectivity and much higher data rates with more robust reliability, higher spectral efficiency and lower energy consumption than today's legacy systems. This quest is motivated by a mix of economic demands (mobile traffic growth, cost, energy, etc.) and socio-technical concerns (health, environment, technological advances, etc.) which render current standards and systems unsustainable [1], [2].

Specifically, the ambitious goals set for 5G networks, as compared to the 4G Long Term Evolution Advanced (LTE-A) systems include: 1000x higher mobile data traffic per geographical area, 100x higher typical user data, 100x more connected devices, 10x lower network energy consumption and 5x reduced end-to-end latency [3], [4]. A quantitative comparison of 4G performance metrics and the corresponding 5G targets is shown in Table I [1], [2], [5].

However, there is a major problem: 5G must support much higher data rates (100-1000x legacy networks), yet current systems are not very far from the Shannon Limit (albeit, treating other-cell interference (OCI) as noise). Towards this end, the research community identifies three plausible ways to get several orders of magnitude throughput gain: (i) extreme densification of infrastructure, (ii) large quantities of new bandwidth, and (iii) many more antennas, allowing a throughput gain in the spatial dimension. These methods are complementary in many respects. Large swathes of bandwidth require going to higher frequencies, especially the promising millimeter-wave (mmWave) spectrum with carrier frequencies of 30-300 GHz. These high frequencies require many antennas to overcome the path losses in such an environment (mainly from such small antennas) and enable precisely that because half-wavelength dipole antennas are so small at such frequencies. Furthermore, higher frequencies need smaller cells to overcome blocking and pathloss, while the same channel difficulties (i.e., pathloss and blocking) cause the interference due to densification to decay quickly [9].

The amalgam of these three features, ultra-dense networks (UDNs), mmWave spectrum and massive multiple-input-multiple-output (massive MIMO), which fortunately exhibit symbiotic relationship as illustrated in Fig. 1, produces the heterogeneous network (HetNet) architecture and the mmWave massive MIMO paradigm that have emerged as key subjects of research trends for next-generation cellular networks.

Manuscript received June 20, 2017; revised October 31, 2017; accepted December 21, 2017. Date of publication December 28, 2017; date of current version May 22, 2018. The work of S. A. Busari and K. M. S. Huq was supported by the Fundação para a Ciência e a Tecnologia, Portugal, through Ph.D. Grants under Reference PD/BD/113823/2015 and through Post-Doctoral Grants under Reference SFRH/BPD/110104/2015. The work of L. Dai was supported by the National Natural Science Foundation of China for Outstanding Young Scholars under Grant 61722109. *(Corresponding author: Sherif Adeshina Busari.)*

S. A. Busari, K. M. S. Huq, S. Mumtaz, and J. Rodriguez are with the Instituto de Telecomunicações, 3810-193 Aveiro, Portugal (e-mail: sherifbusari@av.it.pt; kazi.saidul@av.it.pt; smumtaz@av.it.pt; jonathan@av.it.pt).

L. Dai is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: daill@tsinghua.edu.cn).

Digital Object Identifier 10.1109/COMST.2017.2787460

TABLE I
PERFORMANCE COMPARISON OF 4G AND 5G NETWORKS

Performance Metrics	4G	5G
Peak data rate (Gbps)	1	20
User experienced data rate (Mbps)	10	100
Connection density (devices/km^2)	10^5	10^6
Mobility support (kmph)	350	500
Area traffic capacity (Mbit/s/m^2)	0.1	10
Latency (ms)	10	1
Reliability (%)	99	99.99
Positioning accuracy (m)	1	0.01
Spectral efficiency (bps/Hz)	3	10
Network energy efficiency (J/bit)[1]	1	0.01

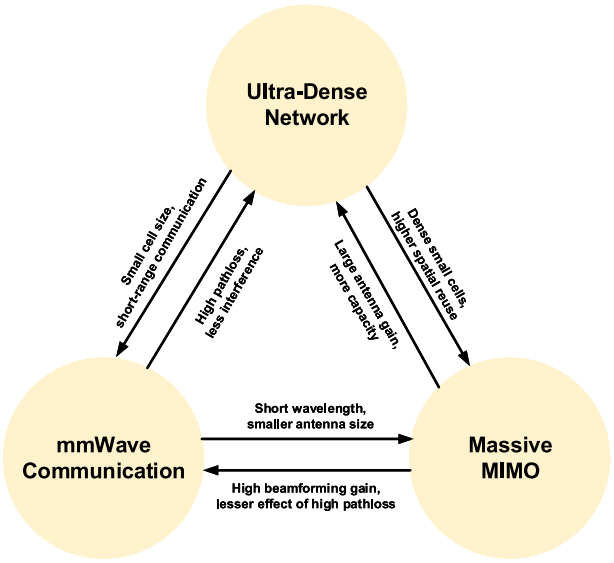

Fig. 1. The symbiotic cycle of the three prominent 5G enablers.

MmWave massive MIMO springs as a technology that combines the prospects of the huge available mmWave bandwidth on the one hand, and the expected gains from massive MIMO antenna arrays on the other. When the mmWave massive MIMO technology is used in the HetNet topology, next-generation mobile networks can be projected to reap the benefits of the three enablers on a very large scale, and thereby support a plethora of high-speed services and bandwidth-hungry applications not hitherto possible.

Although the potential of mmWave Massive MIMO is exciting, the challenges that evolve span the broad fields of communication engineering and allied disciplines. With rapid urbanization and an accelerating number of cellular-enabled devices per person, the density and physical environment these mmWave massive MIMO-based networks will face are new, and the challenges are immense and barely understood. However, the trends are becoming more widely recognized, and models, measurements, analyses and evaluation methodologies are being undertaken and developed for real-life deployment of a hyper-dense urban mmWave massive MIMO-based cellular system [207]–[209]. This survey, therefore, attempts to explore the emerging trends, evolving challenges and the proposed solutions for mmWave massive MIMO, in the march towards the 5G era. As a prelude, we present an overview of the three enablers for mmWave massive MIMO HetNets.

UDN refers to the hyper-dense deployment of small cell base stations (BSs) within the coverage areas of the macrocell BSs. It has been identified as the single most effective way to increase network capacity based on its potentials to significantly raise throughput, increase spectral and energy efficiencies as well as enhance seamless coverage for cellular networks. Small cells are classified as metro, micro, pico or femtocells (in decreasing order of capabilities) based on power range, coverage distance and the number of concurrent users to be served. The rationale behind them is to get users physically close to their serving BSs [6].

The overlay of small cells on traditional macrocells leads to a multi-tier HetNet where the host macrocell BSs handle more efficiently control plane signaling (e.g., resource allocation, synchronization, mobility management, etc.) while the small cell BSs provide high-capacity and spectrally-efficient data plane services to the users [3], [226]. This HetNet topology has the potential to deliver on all fronts: it increases network capacity based on increased cell density and high spatial and frequency reuse; enhances spectral efficiency based on improved average signal to interference and noise ratio (SINR) with tighter interference control; and improves energy efficiency based on reduced transmission power and lower pathloss resulting from smaller cell radii or shorter distance between the small cells and their User Equipment (UEs) [3], [6], [9]–[13].

The mmWave frequency band (i.e., the extremely high frequency (EHF) range of the electromagnetic spectrum representing 30-300 GHz and corresponding to wavelength 1-10 mm) has an abundant bandwidth of up to 252 GHz. With a reasonable assumption of 40% availability over time, these mmWave bands will possibly open up ∼100 GHz new spectrum for mobile broadband applications [84]. Out of this, about 23 GHz bandwidth is being identified for mmWave cellular in the 30-100 GHz bands, excluding the 57-64 GHz oxygen absorption band which is best suited for indoor fixed wireless communications (i.e., the unlicensed 60 GHz band) [163]. As a key enabler for the multi-gigabit-per-second (Gbps) wireless access in NGMNs, mmWave wireless connectivity offers extremely high data rates to support many applications such as short-range communications, vehicular networks, and wireless in-band fronthauling/backhauling, among others.

Massive MIMO, the third of the triad, is a technology which scales up the number of antenna elements by several orders of magnitude more than the number used in conventional MIMO systems [18]. This is with the aim of reaping the benefits of MIMO on a much larger scale [21]. It has the potential to increase the capacity of mobile networks in manifolds through aggressive spatial multiplexing while simultaneously improving the radiated energy efficiency. Using the excess degrees of freedom (DoF) resulting from its large number of antennas (100 or more), massive MIMO can harness the available space resources to improve spectral efficiency, and with the

[1]Normalized with 4G value.

aid of beamforming, it can suppress interference by directing energy to desired terminals only. This avoids fading dips and thereby reduces the latency on the air interface [22], [23]. When deployed in the mmWave regime, the corresponding small sizes of massive MIMO antennas drastically reduce cost and power, not only by using low-cost low-power components but also by eliminating expensive and bulky components (such as large coaxial cables) and high-power radio frequency (RF) amplifiers at the front-end [17], [21].

The three technological enablers are the key ingredients for the expected orders of magnitude throughput gain, and their combined impact is capable of realizing the anticipated 1000-fold increase in capacity for 5G networks relative to 4G legacy systems [6]–[9].

A. Contributions of the Survey

MmWave massive MIMO evolves from multiple constituent parts, and so the technology is expected to inherit concepts and benefits of prior technologies such as MIMO, conventional massive MIMO, and mmWave communications, in such a way as to adopt and/or adapt the techniques, as well as engineer new ones. This convergence of different technologies necessitates the design of novel strategies and procedures to incorporate the new characteristics, address the increasing complexities and resolve the emergent challenges, across all the layers and components in the network.

In this survey, we attempt a holistic overview of the concepts and techniques on the subject from cross-layer design perspectives. The aim, however, is not to describe in detail the theory behind every associated technique but to present the state-of-the-art (SOTA) techniques being explored for mmWave massive MIMO systems and how they have evolved over time. It, therefore, serves to describe mmWave massive MIMO and its interplay with legacy technologies, with particular highlights on the areas of convergence and emerging trends.

To the authors' best knowledge, there has been no survey available in the scientific arena which has discussed mmWave Massive MIMO communication jointly, although there are quite a few surveys available such as [29], [155], [210], [215], etc., which deliberated these two technologies separately. These surveys typically overview mmWave MIMO or microwave (μWave) massive MIMO, with mmWave massive MIMO cursorily presented as outlook or future works. However, in order to achieve the 5G goals, these two technologies should be studied together. This survey, therefore, presents concepts and techniques specific to mmWave massive MIMO, with the overview of prior or legacy systems for evolutionary trend purposes only. It thus acts as glue between these two promising technologies for 5G networks and beyond.

B. Paper Organization

The rest of the paper is organized as follows. In Section II, we present how the wireless communication technologies have evolved from the single-input-single-output (SISO) to massive MIMO and transitioned from employing μWave to mmWave frequencies. As well, we discuss the main benefits and challenges along the evolutionary line. In Section III, we introduce the candidate architecture for mmWave massive MIMO and discuss the propagation characteristics in this new domain. Section IV then uses a single-cell system model in the downlink to describe the three main communication system blocks (transmitter, channel, and receiver) with respect to their associated techniques and processes such as precoding, antenna array, channel estimation and feedback, and channel modeling among others. This model is based on the mmWave massive MIMO paradigm, with highlights on how the techniques differ from those of the earlier systems. In Section V, we present the emerging trends, research challenges and proposed solutions for mmWave massive MIMO in the main areas of propagation, transceiver architecture, waveform, multiple access schemes, resource allocation, and wireless fronthaul design. Overview of standardization activities and health and safety issues are presented as well. The conclusion then follows in Section VI.

C. Notation

Boldface lower and upper case symbols represent vectors and matrices, respectively. The transpose, Hermitian transpose, and trace operators are denoted by $(\cdot)^T$, $(\cdot)^*$ and $tr(\cdot)$, respectively, while the deteminant operator and the norm of a vector are denoted by $|\cdot|$ and $\|\cdot\|$, respectively. $M \gg N$ means M is much greater than N. The list of abbreviations used is provided in Table X towards the end of this article.

II. Evolution of Cellular Network Technologies

Wireless communication has today reached several significant milestones, and the literature is replete with technologies that have enabled the progress so far. Over time, mobile networks' technologies have continued to evolve, pushing legacy systems towards their theoretical limits and motivating research for next-generation networks with better performance capabilities in terms of reliability, latency, throughput, cost, energy and spectral efficiency, among others.

Cellular networks have undergone remarkable transitions: from SISO at μWave frequencies to the latest legacy system with massive MIMO at μWave frequencies (hereafter referred to as conventional massive MIMO). However, the projections of explosive growth in mobile traffic, unprecedented increase in connected wireless devices and proliferation of increasingly smart applications and broadband services have motivated research for the development of 5G mobile networks. These networks are expected to deliver super-speed connectivity and high data rates, provide seamless coverage, support diverse use cases and satisfy a wide range of performance requirements for which the legacy cellular networks have reached their theoretical limits. This has prompted the research community to consider mmWave massive MIMO with a view to exploiting the joint capabilities of mmWave communication and massive MIMO. We present a brief overview of the road in this direction.

A. From μWave to mmWave Cellular Networks

Up till now, the operation of cellular networks has been mainly limited to the congested sub-6 GHz μWave frequency

TABLE II
COMPARISON OF AVAILABLE BANDWIDTH AT μWAVE AND MMWAVE FREQUENCIES[2]

Band	Generation	Frequency (GHz)	Wavelength (m)	Frequency Range (GHz)	Bandwidth (GHz)	Available Bandwidth (GHz)[3]
μWave	2G	0.8	0.3750	0.791 - 0.862	1.117[4]	2.5
		0.9	0.3333	0.880 - 0.959		
	3G	1.8	0.1667	1.710 - 1.880		
	4G	2.1	0.1429	1.920 - 2.169		
		2.6	0.1154	2.500 - 2.690		
	5G	0.6	0.500	0.470 – 0.694		
		0.7	0.4286	0.694 - 0.790		
		1.5	0.2000	1.427 – 1.518		
		3.5	0.0857	3.300 - 3.800		
		4.7	0.0638	4.500 – 4.990		
		5.6	0.0536	5.500 – 5.700		
mmWave	5G	23	0.0130	22.55 - 23.55	1.0	23
		28	0.0107	27.50 - 31.23	1.3	
		38	0.0079	38.6 - 40.0	1.4	
		40	0.0075	40.5 - 42.5	2.0	
		46	0.0065	45.5 - 46.9	1.4	
		47	0.0064	47.2 - 48.2	1.0	
		49	0.0061	48.2 - 50.2	2.0	
		73	0.0041	71 - 76	5.0	
		83	0.0036	81 - 86	5.0	
		93	0.0032	92 - 95	2.9	

bands. Though these bands have favorable propagation characteristics, the total available bandwidth of about 1–2 GHz is grossly insufficient to support the foreseen traffic demand of next-generation mobile services and applications. This challenge pushes for the exploration of the under-utilized higher frequency bands with an abundant amount of bandwidth (first in the the 30-100 GHz mmWave band where approximately 10 times that available at μWave bands can be exploited) as shown in Table II [9], [10], [16]–[18].

When compared to the congested sub-3 GHz μWave bands used by 2G-4G cellular networks hitherto, and the additional television white space (TVWS) and other sub-6 GHz μWave frequencies approved by the International Telecommunication Union's World Radio Conference (ITU-WRC) 2015 (as shown in Table II), the mmWave bands offer several advantages in terms of larger bandwidths which translate directly to higher capacity and data rates, and smaller wavelengths enabling massive MIMO and adaptive beamforming techniques. The relatively closer spectral allocations in the mmWave bands lead to a more homogeneous propagation, unlike the disjointed spectrum in legacy networks.

On the other hand, mmWave signals are prone to higher path loss, higher penetration loss, severe atmospheric absorption and more attenuation due to rain, when compared with μWave signals. In addition, they are vulnerable to blockages by objects. Thus, directional communication is employed in mmWave systems to counter the severe propagation losses and avoid interference thereby ensuring higher throughput and better energy efficiency [14]–[17].

However, recent research studies and channel measurement campaigns carried out at different mmWave frequencies (e.g., 28, 38, 60 and 73 GHz) have revealed that mmWave signals can gainfully exploit these challenges by employing adaptive beamforming techniques to suppress interference and use relay stations to circumvent obstacles thereby avoiding blockages [18], [19], [25], [26], as shown in Fig. 2. More so, for the 50–200 m cell size envisaged for mmWave small cells, the expected 1.4 dB attenuation (i.e., 7 dB/km) due to heavy rainfalls has a minimal effect [16]. The high path loss limits inter-cell interference (ICI) and allows more frequency re-use, which by extension improves the overall system capacity [8].

[2]Excluding the International Scientific and Medical (ISM) band and the wireless fidelity (WiFi) frequencies at 2.4, 5 and 60 GHz.

[3]Subject to regional availability.

[4]Existing allocated IMT spectrum.

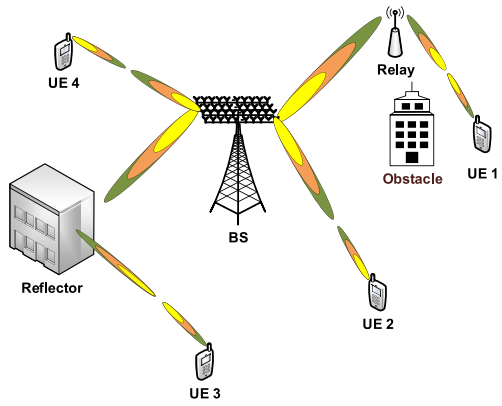

Fig. 2. Directional communications in 5G HetNets.

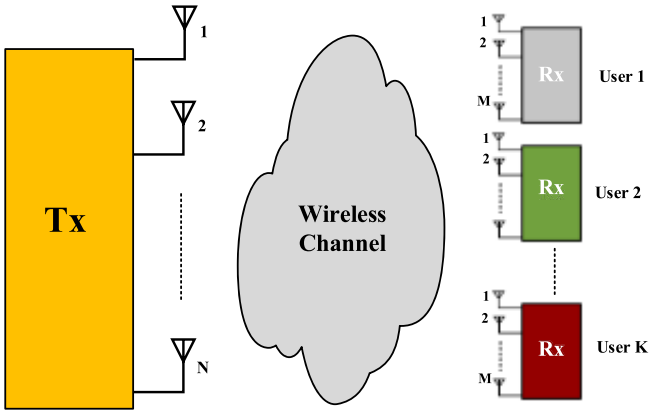

Fig. 3. Wireless communication system model.

In addition, the huge spectrum offered by the mmWave band will enable both the access (BS-UE) and fronthaul/backhaul links to support much higher capacity than present 4G networks [16]. In fact, the supposed drawback of short-distance or short-range mmWave communication perfectly fits into the UDN trend and opens up new avenues for short-range applications such as the potential use in data centers.

B. From SISO to MIMO to Massive MIMO

SISO systems employ single antennas at both the transmitters (i.e., BSs) and the receivers (i.e., UEs), while MIMO systems use multiple antennas at both. MIMO offers higher capacity and reliability than SISO systems as its channels have considerable advantages over SISO channels in terms of multiplexing, diversity and array gains [29], [30]. The diversity gains of MIMO scale with the number of independent channels between the transmitter and the receiver while the maximum multiplexing gain is the lesser of the number of antennas at the transmitter and receiver units. However, maximum multiplexing and diversity gains cannot be simultaneously reaped from MIMO systems as a fundamental trade-off exists between both. The best of the two gains cannot be achieved concurrently [30]. Massive MIMO, on the other hand, uses a much larger number of antennas than those used in conventional MIMO systems.

In evaluating the performance of cellular systems with the different antenna configurations (i.e., SISO to massive MIMO), we consider a single cell, downlink wireless communication system model shown in Fig. 3 under the following four scenarios: SISO, single-user MIMO (SU-MIMO), multi-user MIMO (MU-MIMO) and massive MIMO. The set-up has one transmitter (Tx) with N transmit antennas and k users, with each user having a receiver (Rx) equipped with M receive antennas.

1) SISO [$N = 1,\ M = 1$ and $K = 1$]

SISO is composed of a BS (Tx) and a UE (Rx), each with a single antenna. Given that the received signal $\boldsymbol{y}$ is generally modeled as

$$\boldsymbol{y} = \boldsymbol{H}x + \mathbf{n} \tag{1}$$

where $\boldsymbol{H}$, $\boldsymbol{x}$, and $\boldsymbol{n}$ represent the channel matrix, transmit signal vector and the noise vector respectively, and $\boldsymbol{n}$ is assumed to be the additive white Gaussian noise (AWGN) following a complex normal distribution $\mathcal{CN}(0,\ \sigma)$ with zero mean and σ standard deviation. For the SISO scenario, the channel matrix in (1) becomes one-dimensional and scalar. The received signal $\boldsymbol{y}$ reduces to

$$\boldsymbol{y} = h\boldsymbol{x} + \boldsymbol{n} \tag{2}$$

The achievable capacity (bits/s/Hz) of the single link can thus be expressed as

$$C_{SISO} = \log_2\ (1 + \gamma) = \log_2\left(1 + h^2 \frac{P_t}{{\sigma_n}^2}\right) \tag{3}$$

where γ is signal-to-noise ratio (SNR), P_t is the transmit (signal) power, ${\sigma_n}^2$ is the noise power and h is the channel coefficient.

2) Single User MIMO [$N > 1,\ M > 1$ and $K = 1$]

In MIMO antenna array systems, both the transmitters and the receivers are equipped with multiple antennas. This configuration leads to significant increase in the data rates and mean spectral efficiencies of wireless systems without any increase in the SNR or the bandwidth of such systems, as required by the Shannon Capacity theorem on the theoretical limit for SISO systems. In MIMO, the additional increase in capacity comes from spatial multiplexing through multi-stream transmissions from the multiple antennas [6], [27].

While the channel capacity of SISO systems increases logarithmically with an increase in system's SNR, that of MIMO systems increases linearly with increasing number of antennas (i.e., scales with the smaller of the number of transmit or receive antenna when the channel is full rank). The gains can be limited (i.e., not exactly linear increase) when the channel is not full rank [214]. This increase is however at the expense of the additional cost of deployment of multiple antennas, space constraints (particularly for mobile terminals) and increased signal processing complexities [31].

For the SU-MIMO system with multi-antenna transmitter and receivers, but where only one active user is served or scheduled in a transmission time interval (TTI), the received signal vector, $\mathbf{y}_m \in C^{M\times 1}$, can be expressed as

$$\mathbf{y}_m = \sqrt{\rho}\mathbf{H}_{m,n}\mathbf{x}_n + \mathbf{n}_m \tag{4}$$

where $n = \{1, 2, \ldots, N\}$ transmit antennas and $m = \{1, 2, \ldots, M\}$ receive antennas. $\mathbf{x}_n \in C^{N\times 1}$ (transmit signal vector), $\mathbf{n}_m \in C^{M\times 1}$ (noise and interference vector), $\mathbf{H}_{m,n} \in C^{M\times N}$ (assumed narrow-band time-invariant channel with deterministic and constant channel matrix) and ρ is a scalar representing the normalized transmit power (i.e., the total power of the transmit signal sum to unity, $E\{\|x_n\|^2\} = 1$).

Assuming independent and identically distributed (i.i.d) Gaussian transmit signals, zero-mean circularly symmetric complex Gaussian noise with an identity covariance matrix I and perfect channel state information (CSI) at the receiver, the instantaneous achievable rate (bits/s/Hz) is given by

$$C_{SU-MIMO} = \log_2\left|\left(\mathbf{I} + \frac{\rho}{\mathrm{N}}\mathbf{H}\mathbf{H}^*\right)\right| \tag{5}$$

where N is the number of transmit antennas and the *m,n* channel subscripts have been dropped. The expression in (5) is bounded by

$$\log_2\,(1+\rho M) \le C_{SU\text{-}MIMO} \le \min(N,\ M)\log_2\left(1 + \frac{\rho\max(N,M)}{N}\right) \tag{6}$$

3) Multi-User MIMO [$N > 1,\ M > 1,\ M_k = 1$ and $K > 1$]

MIMO systems have two basic configurations: SU-MIMO and MU-MIMO [29], [32]. MU-MIMO offers greater advantages [21] which include the following:

- SU-MIMO transmissions dedicate all time-frequency resources to a single terminal, employing transmit diversity, spatial multiplexing, and beamforming techniques. However, MU-MIMO exploits multi-user diversity in the spatial domain by allocating a time-frequency resource to multiple users, which results in significant gains over SU-MIMO, particularly when the channels are spatially correlated [32].
- MU-MIMO BS antennas simultaneously serve many users, where relatively cheap single-antenna devices can be employed at user terminals while expensive equipment is only needed at the BS thereby bringing down cost.
- MU-MIMO system is less sensitive to the propagation environment than SU-MIMO system, and so a rich scattering is generally not required.

For this scenario, the BS transmits simultaneously to multiple users each with a single antenna. The received signal vector, $\mathbf{y}_k \in C^{K\times 1}$, can be expressed as

$$\mathbf{y}_k = \sqrt{\rho}\mathbf{H}_{k,n}\mathbf{x}_n + \mathbf{n}_k \tag{7}$$

$\mathbf{x}_n \in C^{N\times 1}$ (transmit signal vector), $\mathbf{n}_k \in C^{K\times 1}$ (noise and interference vector) and $\mathbf{H}_{k,n} \in C^{K\times N}$ (channel matrix). The achievable capacity (bits/s/Hz) is given as

$$C_{MU-MIMO} = \max_{\mathbf{P}} \log_2\ \left|\left(\mathbf{I} + \rho\mathbf{H}\mathbf{P}\mathbf{H}^*\right)\right| \tag{8}$$

where $\mathbf{P}$ is a positive diagonal matrix with power allocations $\mathbf{P} = \{p_1, p_2, \ldots, p_K\}$ which maximizes the sum transmission rate. Here, the *k,n* channel subscripts have been dropped.

Multiple antenna system implementations have shifted to MU-MIMO in recent years due to its benefits, which have made it the candidate for several wireless standards [21], [29]. However, despite its great significant advantages over SISO and SU-MIMO antenna systems, MU-MIMO has been identified as a non-scalable technology, and as a sequel, massive MIMO is evolving to scale up the benefits of MIMO significantly. Unlike MU-MIMO which has roughly the same number of terminals and service antennas, massive MIMO has an excess of service antennas over active terminals which can be used for enhancements such as beamforming, in order to bring about improved throughput and energy efficiency. Its excess antennas, therefore, increase its scalability [21].

Overall, MIMO is a smart technology aimed at improving the performance of wireless communication links [27]. Compared to the SISO systems, multiple antenna systems have been shown from studies and commercial deployment scenarios in different wireless standards, such as the IEEE 802.11 (WiFi), IEEE 802.16 (WiMAX), the third generation (3G) Universal Mobile Telecommunications System (UMTS) and the High Speed Packet Access (HSPA) family series as well as the LTE, to offer significant improvements in the performance of cellular systems, with respect to both capacity and reliability [28], [29].

4) Massive MIMO [$N \gg M, N \to \infty$ *or* $M \gg N, M \to \infty$]

Massive MIMO is also known as Large Scale Antenna Systems (LSAS), Full Dimension MIMO (FD-MIMO), Very Large MIMO and Hyper MIMO. It employs an antenna array system with a few hundred BS antennas simultaneously serving many tens of user terminals in the same time-frequency resource [21], [32]. It has the potential to enormously improve spectral efficiency by using its large number of BS transmit antennas to exploit spatial domain DoF for high-resolution beamforming and for providing diversity and compensating pathloss, thereby improving energy efficiency, data rates, and link reliability [18], [33].

When the number of antennas grows large such that $N \gg M$ and $N \to \infty$, the achievable rate for MIMO in (5) becomes

$$C_{massive\ MIMO} \approx \mathrm{M}\log_2(1+\rho) \tag{9}$$

and when $M \gg N$ and $M \to \infty$, the achievable rate in (5) approximates to

$$C_{massive\ MIMO} \approx \mathrm{N}\log_2\left(1 + \frac{\rho M}{N}\right) \tag{10}$$

Eqns. (9) and (10) assume that the row or column vectors of the channel $\boldsymbol{H}$ are asymptotically orthogonal and demonstrate the advantages of massive MIMO, where the capacity grows linearly with the number of the employed antenna at the BS or the UE, as the case may be. It should, however, be noted that the preceding equations (1)–(10) give only a superficial analysis of SISO, MIMO, and massive MIMO systems, aimed only at an overview. For in-depth analyses and derivations, the reader is referred to [23], [29], and [164] and relevant references therein.

Massive MIMO employs spatial multiplexing, requires CSI for both uplink and downlink and depends on phase coherent signals from all the antennas at the BS. It is an enabling technology for enhancing the energy and spectral efficiencies, reliability, security and robustness of future broadband networks, both fixed and mobile. However, despite these obvious benefits, massive MIMO implementation is

TABLE III
[5]SUMMARY OF BENEFITS AND CHALLENGES FOR ANTENNA TECHNOLOGIES

Antenna Technology	Diversity Gain	Multiplexing Gain	Array Gain	mmWave Bandwidth	Computational Complexity	Channel Estimation challenge	Pilot Contamination issue
SISO	×	×	×	×	×	×	×
SU-MIMO	✓	✓✓	✓✓	×	××	××	××
MU-MIMO	✓✓	✓✓✓	✓✓	×	×××	×××	×××
Massive MIMO	✓✓✓✓	✓✓✓✓	✓✓✓✓	×	××××	××××	××××

faced with some challenges which have been subjects of research studies. These challenges, among others, include the following [17], [18], [21], [32]:

- Need for simple, linear and real-time techniques and hardware for optimized processing of the vast amounts of generated baseband data, at associated internal power consumption.
- Need for new and realistic characterization and modeling of radio channels taking cognizance of the number, geometry, and distribution of the antennas.
- Need for accurate CSI acquisition and feedback mechanisms and techniques to combat pilot contamination and effects of hardware impairments due to the use of low-cost, low-power components.
- Need for the development of commercial and scalable prototypes and deployment scenarios to engineer the heterogeneous network solutions for future mobile systems.

C. Summary and Open Issues

The summary of the basic benefits and challenges of the different antenna technologies discussed in this section is presented in Table III. While the benefits in terms of diversity, multiplexing and array gains have improved along the evolutionary line, new challenges in terms of computational and signal processing complexities, channel estimation issues, pilot contamination problems, etc, have increased as well. In addressing these challenges, many techniques and concepts are being explored in order to maximize the system benefits and optimize the cost-benefit tradeoffs.

For the legacy massive MIMO, as an example, the near-optimal linear precoders such as the matched filter (MF) and Zero Forcing (ZF) precoders have been proposed [22], [23]. They have lower computational complexities and better implementation feasibility than non-linear precoders (such as dirty-paper-coding (DPC) [46], vector perturbation (VP) [191] and lattice-aided methods [192]), without significant performance loss. In combating pilot contamination, some of the techniques that have been proposed for massive MIMO include: protocol-based methods [22], [41], [193], [194], blind methods [42], [196], [197], precoding-based methods [36], [39], [123], [195], and angle of arrival (AoA)-based methods [38], [198], [199], etc.

Also, only a few works have incorporated the use of mmWave frequencies for cellular communication. As a result, it has remained an open issue and is now aggressively being considered for 5G in the mmWave massive MIMO technology. How the cost-benefit tradeoffs of the technology will play out for improved system performance is the task for 5G, and thus an open issue.

III. DAWN OF MMWAVE MASSIVE MIMO

MmWave massive MIMO is a promising candidate technology for exploring new frontiers for next-generation cellular systems, starting with 5G networks. It benefits from the combination of large available bandwidth (in mmWave frequency bands) and high antenna gains (achievable with massive MIMO antenna arrays). With enhanced energy and spectral efficiencies, increased reliability, compactness, flexibility, and improved overall system capacity, mmWave massive MIMO is expected to break away from today's technological shackles, address the challenges of the explosively-growing mobile data demand and open up new scenarios for future applications.

For mmWave massive MIMO systems, maximum benefits can be achieved when different transmit-receive antenna pairs experience independently-fading channel coefficients. This is realizable when the antenna elements' spacing is at least 0.5λ, where λ is the wavelength of the signal. Since λ reduces with increasing carrier frequency, a higher number of elements in antenna arrays of same physical dimension can be realized at mmWave than at μWave frequencies. The realization of this optimal performance is dependent on the availability of accurate CSI [8], [26], [91], [92].

At mmWave frequencies, the dimensions of antenna elements (as well as the inter-antenna spacing) become incredibly small (due to their dependence on wavelength). It thus becomes possible to pack a large number of antenna elements in a physically-limited space, thereby enabling massive MIMO antenna array, not only at the BSs but also at the UEs [175]–[177]. However, the maximum numbers of antennas under consideration by 3GPP (for example at 70 GHz) are 1024 and 64 for the BSs and UEs, respectively. As for the RF chains, the maximum numbers are 32 and 8 for the BSs and UEs, respectively [210], [223].

[5](✓ means benefit, × means challenge and the number/amount of the symbols signifies normalized quantity relative to SISO).

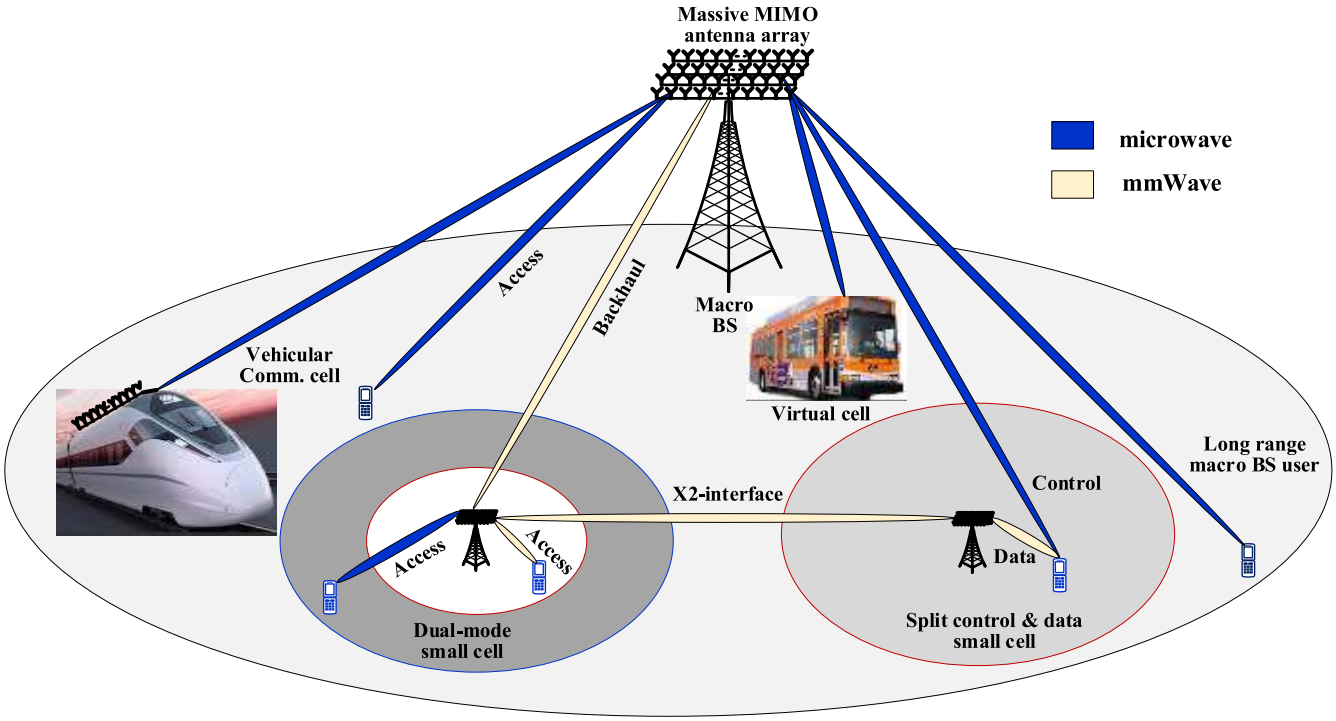

Fig. 4. Candidate 5G network architecture based on mmWave massive MIMO.

With low radiation power (due to small antenna size) and high propagation attenuation, it becomes necessary to use highly-directional, steerable, configurable or smart antenna arrays for mmWave massive MIMO in order to ensure high received signal power for successful detection [177]. In this section, we give an overview of the architecture for mmWave massive MIMO and the propagation characteristics in this new domain.

A. Architecture

A candidate architecture for the mmWave massive MIMO-based 5G network is shown in Fig. 4. The architecture is a multi-tier cellular HetNet composed of the macrocell and small cell BSs, all with massive MIMO, μWave, and mmWave communication capabilities. It features several scenarios that are subjects of ongoing research in realizing the 5G goals. Some of these scenarios are highlighted as follows:

- Split control and data plane framework where the control signals are handled by the long-range μWave massive MIMO macrocell BS (for efficient mobility and other control signaling) while data signals are handled by the mmWave massive MIMO small cells for high capacity.
- Dual-mode or dual-band small cells where close-by users are served by mmWave access links while farther-away users are served at μWave frequency band, thereby serving as a dynamic cell and emulating the "cell breathing" concept from legacy networks.
- In-band backhauling where both the access and backhaul links are on the same mmWave frequency band to reduce cost and optimize spectrum utilization.
- Vehicular communication where the μWave macrocell BSs serve the long-range and highly mobile users.
- Virtual cells where a user, particularly cell-edge user, chooses its serving BS without being constrained to be served only by the closest BS as in the legacy user association and handover approach based on coverage zone.

This architecture, therefore, brings to fore many opportunities in terms of possible scenarios, use cases, and applications. Some of these possibilities have been considered lately for legacy systems. However, the mmWave massive MIMO-based 5G networks show greater potentials in realizing them.

The realization of the architecture in Fig. 4 is being pursued for 5G through multi-disciplinary, cross-layer research. Inter-operator resource management (i.e., sharing of spectrum, access, and infrastructure) in mmWave networks is considered in [224] and [225]. Resource allocation and user scheduling in joint μWave-mmWave dual-mode BSs using novel-context-awareness principles have been proposed by Semiari *et al.* [153], [222]. Fronthaul design options in mmWave massive MIMO-based 5G networks have been discussed in [140] while the energy and spectral efficiency analysis of wireless backhauling in mmWave-based small cell architectures are provided in [218] and [219]. These cross-layer design proposals are further overviewed in Section V-D while the relevant references therein provide detailed discussions.

B. Propagation Characteristics

Marked differences exist in the propagation characteristics of mmWave massive MIMO networks and those of legacy/conventional cellular systems. At mmWave frequencies,

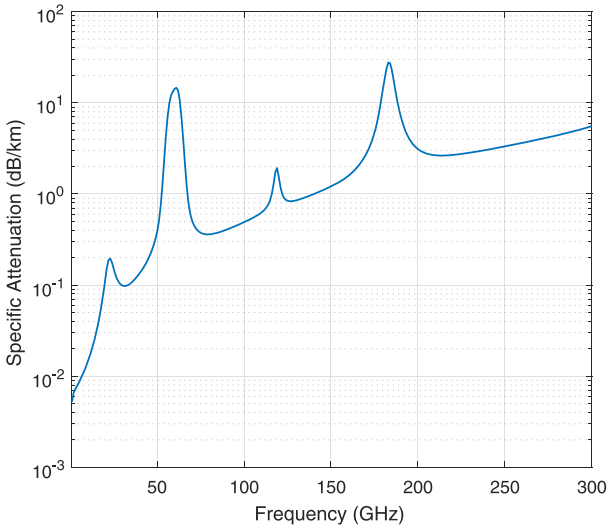

Fig. 5. Atmospheric and molecular absorption at mmWave frequencies.

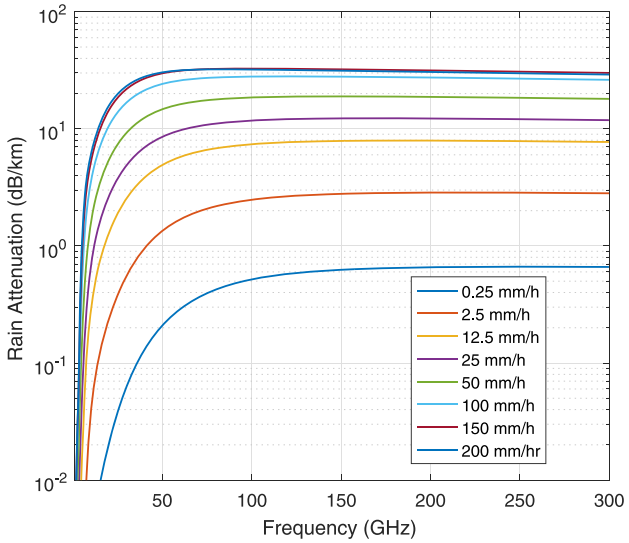

Fig. 6. Rain attenuation at mmWave frequencies.

and as the number of BS antennas goes to infinity, the channel characteristics generally become deterministic, different users experience asymptotic channel orthogonality, and fewer user terminals can be supported due to reduced coverage area [2].

Also, signals propagated at mmWave frequencies experience higher pathloss (with increasing frequency) [84] and have lesser penetrating power through solids and buildings, thereby making them significantly more prone to the effects of shadowing, diffraction and blockage, as the wavelength is typically less than the physical dimensions of the obstacles [80], [85], [86]. In addition, mmWave signals suffer more attenuation due to rain [87], have increased susceptibility to atmospheric absorption [88] and experience higher attenuation due to foliage [89]. The attenuations due to atmospheric/molecular absorption and rain as a function of carrier frequency are presented in Figs. 5 and 6, respectively [210].

As shown in Fig. 5, the specific attenuation due to atmospheric and molecular absorption have peaks around 60 and 180 GHz. Hosseini *et al.* [150] using air composition and atmospheric data, have shown that the peaks are due to the high absorption coefficient of oxygen (O_2) and water vapor (H_2O molecules) at 60 GHz and 180 GHz, respectively. These two frequency bands are thus best suited for short distance indoor applications, and the unlicensed 60 GHz mmWave WiFi has already taken the lead in this direction through its standardization.

Further, the experienced attenuation and molecular noise at mmWave frequencies (excluding the 60 GHz band) vary with the time of the day, and the season of the year, being more pronounced during the night than the day, and more during winter than in summer, due to the combined effect of the fall in temperature and corresponding rise in humidity [150]. Though the impact of these attenuation effects limits communication coverage and link quality, the impact is however minimal for the average small cell sizes of 50-200 m envisaged for future cellular networks. More so, beamforming will further be employed to increase the array gains and improve the SNR, in order to counter the effects of the comparatively higher pathloss at mmWave frequencies [16], [210].

Overall, losses in mmWave systems are higher than those of μWave systems. However, the smaller wavelength (which enables massive antenna arrays) and the huge available bandwidth in the bands can compensate for the losses to maintain and even drastically boost performance gains with respect to spectral and energy efficiencies, provided evolving computational complexity, signal processing, and other implementation issues are addressed [90].

In order to unleash the potentials of mmWave massive MIMO technology, a broad range of challenges spanning the length and breadth of communications theory and engineering has to be addressed. The challenges arise due to the differences in the architecture and propagation characteristics of mmWave massive MIMO networks when compared to prior systems.

Among others, the areas of the challenges include: channel modeling, antenna and radio frequency (RF) transceiver architecture design, waveforms and multiple access schemes, information theoretic issues, channel estimation techniques, modulation and energy efficiency issues, medium access control (MAC) layer design, interference management, backhaul transmissions, mobility management, health and safety issues, system-level modeling, experimental demonstrations, tests and characterization, standardization and business models [2].

C. Summary and Open Issues

The propagation characteristics at mmWave frequencies differ from those of μWave frequencies. This has necessitated changes in the architecture and applications of cellular networks, as evident in the candidate architecture shown in Fig. 4. In Table IV, we present a summary of the fundamental differences between conventional (μWave) massive MIMO and mmWave massive MIMO. The comparison in Table IV shows the challenges which have to be addressed or exploited in order to realize the anticipated benefits of mmWave massive MIMO networks.

In terms of benefits, the larger bandwidth available in the mmWave bands, when compared the μWave bands, enables new applications such as wireless fronthauling, the higher path

TABLE IV
COMPARISON OF μWAVE AND MMWAVE MASSIVE MIMO PROPAGATION PROPERTIES

Properties	μWave massive MIMO	mmWave massive MIMO
[6]Received power	$P_{xy}(d) \approx P_x G_{xy} \chi \left(\frac{\lambda}{4\pi}\right)^2 d^{-\alpha}$	$P_{xy}(d) \approx P_x G_{xy} \chi(x,y,B) \left(\frac{\lambda}{4\pi}\right)^2 d^{-\alpha(x,y,B)}$
Gain (G_{xy})	Larger than that of mmWave	At least two orders of magnitude smaller than in μWave.
Pathloss (PL)	Lower pathloss compared to mmWave at the same BS-UE separation distance (d). At fc = 2 GHz and for d = 500 m typical of macrocells, maximum PL of 93.4, 136.8 and 169.3 dB can be expected in LOS, NLOS and O2I scenarios, respectively [217].	Higher pathloss compared to μWave at the same BS-UE separation distance (d). At fc = 28 GHz and for d = 100 m typical of small cells, maximum PL of 103.3, 123.8 and 154.2 dB can be expected in LOS, NLOS and O2I scenarios, respectively [220].
Shadowing (χ)	Independent random variable, small and independent of blockage, non-line-of-sight (NLOS) propagation. Typical values are 4, 6 and 7 for LOS, NLOS, and O2I, respectively [217]	Large, dependent on other random variables and mainly caused by blockage, line-of-sight (LOS)/near- LOS propagation. Typical values are 3.1, 7.8 and 9 for LOS, NLOS, and O2I, respectively [220]
Interference	Dominated by a few nearby ones, leads to background interference floor for large number of interferers and is distance-dependent	Strongly attenuated by randomly-aligned antenna gain patterns and blockage; Not really distance-dependent, assumes an ON-OFF type of behavior.
SINR	Changes slowly from cell center to cell edge	Undergoes extremely random rapid fluctuations; assumes an ON-OFF depending on the beam steering efficiency, blockage, and random beam alignment.
Antenna array	Singly-massive, only the BSs have massive antenna arrays	Doubly-massive, both the BSs and the UEs have massive antenna array capability.
Signal processing	Moderately-complex, particularly CSI acquisition and feedback	Highly-complex due very large amount of CSI.
Handover	Usually done at cell edges based on signal strength for load balancing considerations	Occurs more frequently because of blockage, beam alignment, and high network density.

loss favors high-rate, short-range communications while the shorter wavelength allows the antennas at both the BSs and UEs to go large (i.e., massive) due to the dramatic reduction in antenna sizes. However, new challenges evolve. The combination of the huge bandwidth and massive antenna arrays translate to a lot of computations and signal processing due to the large amount of CSI that have to be processed for channel estimation and feedback, precoding, etc. Also, mmWave signaling will lead to frequent handover which is undesirable for ultra-dense small cell networks.

With the mmWave massive MIMO paradigm, the different challenges which surface in the different blocks of the communication system are reviewed in this survey as summarized in the roadmap of Fig. 7, followed by a discussion of the concepts and techniques being explored to address them in the following Sections IV and V.

IV. MMWAVE MASSIVE MIMO SYSTEM MODEL

Consider a single cell network shown in Fig. 8. The system is composed of the transmitter, the wireless channel and the receivers (UEs). Using this system set-up, we survey the various blocks and processes making up the three principal components of the system based on mmWave massive MIMO.

A. Transmitter Elements, Processes, and Techniques

In this subsection, we review the emerging techniques for the transmitter block of the mmWave massive MIMO communication network, with respect to antenna array design, precoding techniques, and channel estimation processes.

1) Antenna Array: In terms of transmission schemes, beamforming, spatial multiplexing or a combination of both are being used in MIMO systems to improve the performance of wireless links, in terms of achievable throughput. With beamforming, the phases and/or amplitudes of transmitted signals can be controlled, according to the channel environment and desired application [90]. An extensive survey on beamforming types and architectures can be found in [14], [15], [44], [51], and [93]–[96] and a quick overview on comparison of digital and hybrid beamforming with simulation-based link level performance can be found in [90] and [97].

There are three types of antenna array architecture (including the RF chain) that have evolved over time: fully-digital,

[6]x and y refer to the coordinates of the transmitter and receiver, respectively; d is their separation distance.

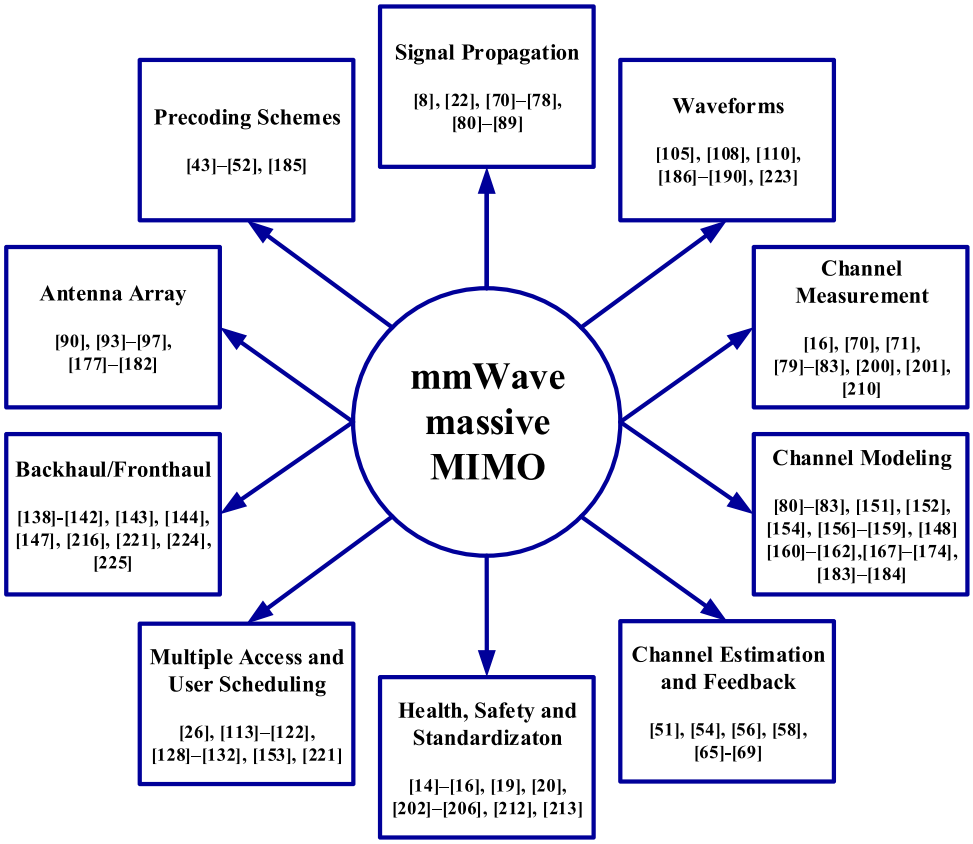

Fig. 7. Survey roadmap.

fully-analog, and hybrid analog-digital architecture. A fully-digital implementation employs dedicated RF front-end and digital baseband per antenna, which for the mmWave massive MIMO is prohibitively costly and practically infeasible due to tight space constraints. The fully-analog array, on the other hand, uses only one RF chain with multiple analog phase shifters (PS). It has simple hardware structure but suffers from poor system performance. Also, it has low antenna gain, as only the phases of the signals, but not their amplitudes can be controlled. The more feasible and practical approach, according to research trends, is the massive hybrid array which consists of multiple analog sub-arrays with their own respective digital chains [177]–[180].

In the massive hybrid array architecture, antenna elements are grouped into analog sub-arrays. Only one PS is dedicated to a single antenna element; all other components are shared by all antenna elements in each sub-array. Each sub-array is fed with only one digital input (in the transmitter) and outputs only one digital signal (at the receiver), and all digital signals from all the sub-arrays are jointly processed in a digital processor. Overall, this hybrid structure, shown in Fig. 9, significantly reduces the cost, number of required hardware components and system complexity, and the performance is roughly comparable with the optimal (but costly and unfeasible) fully-digital architecture [211]. High-performance prototypes of the hybrid arrays are already being built and tested. Examples are the Commonwealth Scientific

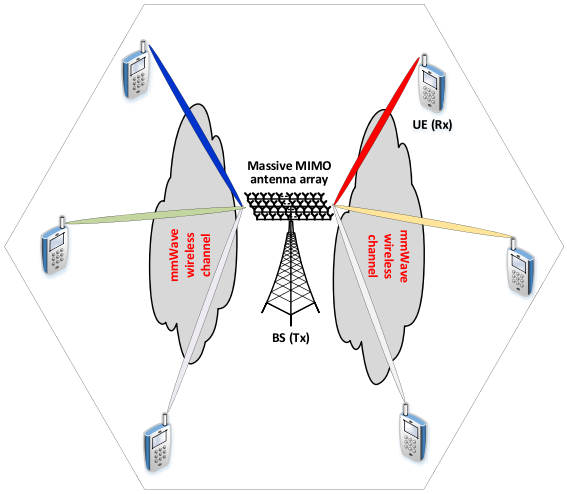

Fig. 8. MmWave massive MIMO system model.

and Industrial Research Organization (CSIRO) [181] and Samsung's [182] prototypes.

In general, the prohibitively high cost and power consumption discourage the use of digital beamforming for mmWave massive MIMO systems. Hybrid beamforming, with less RF chain than the number of antennas, is feasible for mmWave massive MIMO and exhibits only a negligible performance loss when compared to the digital beamformer, which though

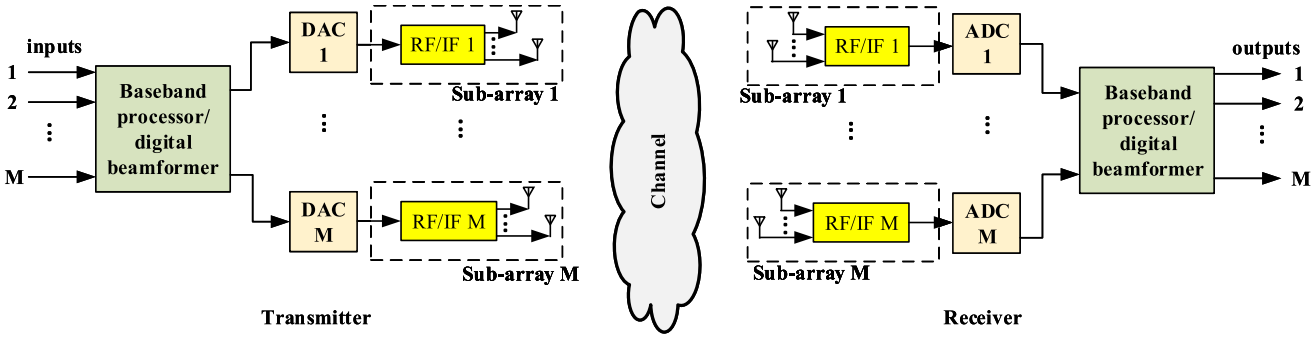

Fig. 9. Architecture of hybrid antenna array system.

has optimal performance for conventional MIMO systems but practically infeasible for mmWave massive MIMO systems.

Spatial multiplexing can be used to drastically boost the system capacity of mmWave massive MIMO channels as the number of BS and UE antennas is much larger than in conventional systems. It increases transmission throughput by subdividing outgoing signals into multiple streams where each stream is transmitted simultaneously and in parallel on the same channel through different antennas. At μWave frequencies, signals are easily diffracted by physical objects thereby leading to several reflected signals from the different scatters. Such rich-scattering environment favors independent and parallel data streams which increase spatial multiplexing gains [90].

However, mmWave channels are specular and have low rank, particularly for line of sight (LOS). They tend to have a lower number of time clusters and spatial lobes (or generally less multipath components) compared to μWave. They are thus incapable of exploiting all available DoF (i.e., only a couple of spatial streams can be supported), thereby limiting achievable multiplexing gains [95], [98], [99]. Also, sufficient decorrelation between different closely-spaced antennas needed by the channels for optimal spatial multiplexing is usually unrealizable in most practical systems such as mmWave indoor [90]. Works such as [98] explored spatial multiplexing for single-user mmWave systems and [100] for multi-user mmWave applications.

In terms of structure, antenna arrays are typically designed as either uniform linear array (ULA) or uniform planar array (UPA) as shown in Fig. 10. However, UPAs are of more interest for mmWave massive MIMO channels as they yield smaller antenna array dimensions, thereby facilitating the packing of more antenna elements in a reasonably-sized array and enabling beamforming in the elevation domain (also known as 3D beamforming). The array response vector for ULA is expressed as

$$a_{ULA}(\phi) = \frac{1}{\sqrt{N}}\left[1,\ e^{j\frac{2\pi}{\lambda}d\sin\phi},\ \ldots,\ e^{j(N-1)\frac{2\pi}{\lambda}d\sin\phi}\right]^T \quad (11)$$

For UPAs, the array response vector is

$$a_{UPA}(\phi,\theta) = \frac{1}{\sqrt{N}}\left[\begin{array}{c}1,\ldots,e^{j\frac{2\pi}{\lambda}d(\mathrm{x}\sin(\phi)\sin(\theta)+\mathrm{y}\cos(\theta))},\ldots,\\ e^{j\frac{2\pi}{\lambda}d((\mathrm{W}_1-1)\sin(\phi)\sin(\theta)+(\mathrm{W}_2-1)\cos(\theta))}\end{array}\right]^T \quad (12)$$

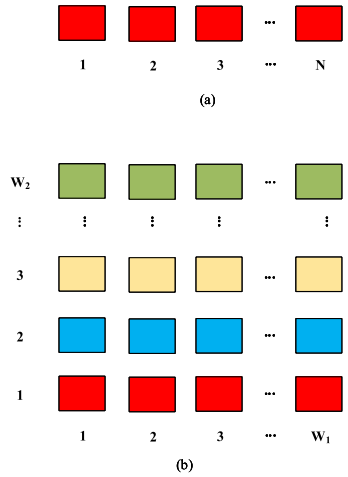

Fig. 10. Antenna array structure (a) ULA (b) UPA ($N = W_1W_2$).

where λ denotes the wavelength, d is the inter-element spacing, ϕ is the azimuth angle, θ is the elevation (or zenith) angle and N is the number of antenna elements [43].

In the case of UPAs, $N = W_1W_2$ where W_1 and W_2 represent the number of elements on the horizontal and vertical, respectively, with $0 \leq x \leq W_1 - 1$ and $0 \leq y \leq W_2 - 1$.

Antenna array elements can be arranged in interleaved or localized mode. The two types of array configurations in hybrid uniform rectangular arrays are shown in Fig. 11. Each square represents an antenna element with squares of the same color representing antenna elements in the same analog subarray. According to [179], localized arrays have higher performance and better support for systems with relatively larger angles of arrival (AoAs). Interleaved arrays have narrower beamwidth which is advantageous but is harder to implement due to space constraints.

Precoding techniques, which we review in the next subsection, are determined by the type of antenna array architecture employed.

2) Precoding Techniques: The design of precoding schemes is highly essential for mmWave massive MIMO cellular systems. Precoders optimize the performance of mobile networks using the concept of interference cancellation in advance, by controlling the phases and/or amplitudes of original signals. This process is also known as beamforming,

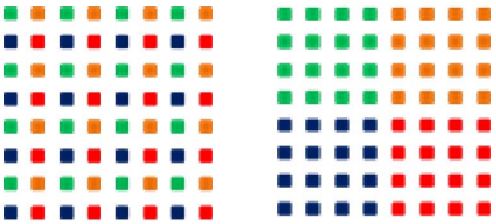

Fig. 11. Antenna array (a) interleaved (b) localized.

and it aims at transmitting pencil-shaped beams that point directly at intended terminals with minimal or no interference at unintended ones. Precoding schemes can generally be classified into three: digital precoding, analog beamforming, and hybrid (analog-digital) precoding specified to the three architectures introduced in Section IV-A1 above. While analog beamforming can only be employed for single-stream single-user systems, both digital and hybrid precoding schemes can be used for single-user as well as multi-user systems [43]. We present an overview of the three schemes in this sub-section.

(i) *Analog beamforming:* This is used to control the phases of signals with the single data stream in order to realize optimal antenna array gain and effective SNR. With perfect knowledge of the CSI available at both the BS and the UE, the analog beamformer employs N antennas at the BS with only one RF chain to send a single data stream to a terminal (i.e., UE) with M antennas and only one RF chain too.

This concept is also called beam steering and aims at the design of the analog beamformer vector $\boldsymbol{f}$ and analog combiner $\boldsymbol{w}$ which maximizes the effective SNR, using the conditions in (13).

$$\begin{aligned} &\left(\mathbf{w}^{opt},\ \mathbf{f}^{opt} = \arg\max\left|\mathbf{w}^{*}\mathbf{H}\mathbf{f}\right|^{2}\right) \\ &\text{subject to: } w_i = \sqrt{N^{-1}}e^{j\varphi_i},\quad \forall\ i, \\ &\qquad\qquad f_l = \sqrt{M^{-1}}e^{j\varphi_l}, \forall l \end{aligned} \tag{13}$$

However, perfect CSI is unrealistic in practical systems, thus necessitating beam training where both the UE and the BS collaborate in selecting the best beamformer (at the BS end) and combiner (at the user end) pair ($|\boldsymbol{f}|-|\boldsymbol{w}|$ pair) from pre-defined codebooks in order to optimize performance [43]. For mmWave massive MIMO systems, the codebook sizes could be very large due to a large number of antennas, together with the accompanying huge overheads. In solving this, a systematic beam training scheme was proposed by Wang *et al.* [44] which reduces the overhead and limits the potentially exhaustive codebooks search using a hierarchical approach.

Similarly, Cordeiro *et al.* [45] proposed a single-sided beam training scheme using a two-step approach, which was adopted by the IEEE 802.11ad standard, where the combiner is first fixed to search for the best precoder exhaustively, and later the found best precoder is fixed to search for the best combiner exhaustively. Though analog beamforming has simple hardware requirement (only one RF chain), it suffers severe performance loss as only the phases of transmitting signals can be controlled. More so, an extension of the scheme to multi-user systems seem not trivial [43], and so it cannot be used for mmWave massive MIMO networks.

(ii) *Digital precoding:* This can control both the phases and amplitudes of signals to be transmitted and can be employed in both single-user and multi-user systems. Digital precoding schemes are classified as either linear or non-linear.

Single-user, linear precoding schemes employ N antennas at the BS to transmit M data streams to a user with M antennas ($M < N$). The ($N \times M$) precoder D uses N RF chains. Examples of such linear precoders include the Matched Filter (MF), Zero Forcing (ZF) and the Wiener Filter (WF) precoder, in increasing order of complexity and performance. The precoder models are expressed in (14)–(16), respectively [43], where H is the $M \times N$ channel matrix with normalized power, P_r represents the average received power, and ${\sigma_n}^2$ denotes the noise power.

$$\mathbf{D}_{\boldsymbol{MF}} = \sqrt{\frac{\mathrm{M}}{\mathrm{tr}(\mathbf{F}\mathbf{F}^{*})}}\boldsymbol{F},\ \mathbf{F} = \mathbf{H}^{*} \tag{14}$$

$$\mathbf{D}_{\boldsymbol{ZF}} = \sqrt{\frac{\mathrm{M}}{\mathrm{tr}(\mathbf{F}\mathbf{F}^{*})}}\boldsymbol{F},\ \mathbf{F} = \mathbf{H}^{*}(\mathbf{H}\mathbf{H}^{*})^{-\mathbf{1}} \tag{15}$$

$$\mathbf{D}_{\boldsymbol{WF}} = \sqrt{\frac{\mathrm{M}}{\mathrm{tr}(\mathbf{F}\mathbf{F}^{*})}}\boldsymbol{F},\ \mathbf{F} = \mathbf{H}^{*}\left(\mathbf{H}\mathbf{H}^{*} + \frac{{\sigma_n}^2 M}{P_r}\mathbf{I}\right)^{-\mathbf{1}} \tag{16}$$

On the other hand, multi-user systems employing digital precoders simultaneously transmit to K mobile terminals, where each terminal is equipped with M antennas and the BS equipped with N antennas, N RF chains and ($N \times MK$) precoders, such that ($MK = N$) and the kth user has ($N \times M$) digital precoder $\mathbf{D}_{\mathrm{k}} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{\mathrm{K}}]$ with total transmit power constraint $\|\mathbf{D}_{\mathrm{k}}\|_F = M$. The received signal by the kth terminal is thus expressed as:

$$y_k = \mathbf{H}_k \sum_{n=1}^{K} \boldsymbol{D_n x}_n + \boldsymbol{n}_k \tag{17}$$

where $\boldsymbol{x}_n$ of size ($M \times 1$) is the original signal vector before precoding with normalized power. $\mathbf{H}_k$ which is of size ($M \times N$) is the mmWave massive MIMO channel matrix between the BS and the kth UE, and $\boldsymbol{n}_k$ is an AWGN vector with entries following i.i.d distribution $\mathcal{CN}(0,\ {\sigma_n}^2)$. From (17), the terms $\mathbf{H}_k\boldsymbol{D_n x}_n$ for $n \neq k$ are interferences to the kth UE. For Block Diagonization (BD) precoder design [185], $\boldsymbol{D_n}$ is chosen to satisfy the condition $\boldsymbol{H}_k\boldsymbol{D_n} = \mathbf{0}$.

Examples of other linear, multi-user digital precoders include the optimal Dirty Paper Coding (DPC) [46] and the near optimal Tomlinson-Harashima (TH) precoding [47], both of which have excellent performance but high computational complexity [43]. Generally, digital precoders have better performance than analog beamformers, as they are able to control both the phases and amplitudes of transmit signals. However, with their use of one dedicated RF chain per antenna, they, unfortunately, have higher energy consumption and prohibitive hardware cost, which make them impractical for mmWave massive MIMO systems.

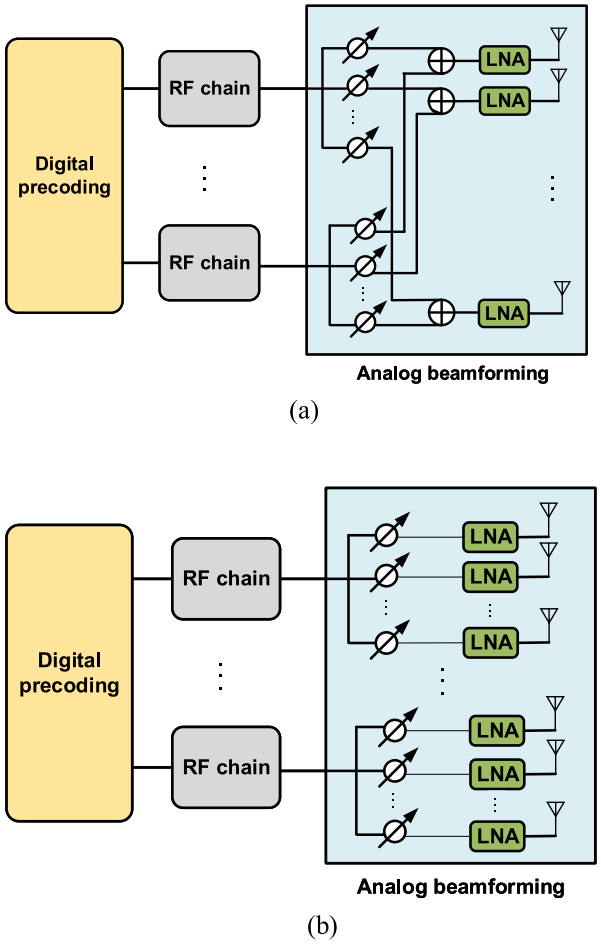

Fig. 12. Hybrid (transmit) precoding for single-user mmWave massive MIMO system (a) Fully-connected architecture (b) Sub-connected architecture.

(iii) *Hybrid precoding:* This is a promising scheme for mmWave massive MIMO cellular networks as it offers a significant reduction in the number of required RF chains and the associated energy consumption and cost (when compared with digital precoding), yet achieving near-optimal performance. It realizes this hybrid configuration by employing a small-size digital precoder with a small number of RF chains to cancel interference in the first step, and a large-size analog beamformer with a large number of only phase shifters (i.e., without RF chains) in the second step to increase the antenna array gain [43].

For a single-user system, hybrid precoders can be grouped into two structural classes: (i) the spatially-sparse hybrid precoder (a fully-connected architecture where all BS antennas are connected to each RF chain via phase shifters) and (ii) the successive interference cancellation (SIC)-based hybrid precoder (a sub-connected architecture where only a subset of BS antennas are connected to each RF chain), as shown in Figs. 12 (a) and (b), respectively [43].

For a multi-user system, Alkhateeb *et al.* [48] proposed a two-stage hybrid precoder that employs the BS analog precoder and the UE analog combiner to jointly maximize the desired signal power of each user in the first stage, and in the second stage employs BS digital precoder to manage multi-user interference. For a multi-user hybrid precoding scheme, consider a mmWave massive MIMO system with N BS antennas and $\mathrm{N_{RF}}$ RF chains such that $N_{RF} \leq N$, and K terminals each with M antennas and only one RF chain. For a fully connected hybrid precoder system, the BS employs a $K \times K$ digital precoder in the baseband $\mathbf{D_k} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D_K}]$ followed by $N \times K$ analog precoder $\mathbf{A_k} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A_K}]$, such that the transmitted signal becomes $\mathbf{x} = \mathbf{ADs}$. The received signal vector r_k observed by the k_{th} terminal after precoding can then be expressed as

$$r_k = \mathbf{H}_k \sum_{n=1}^{K} \boldsymbol{A_n D_n s}_n + \boldsymbol{n}_k \tag{18}$$

After being combined with the analog combiner w_k, where w_k has similar constraints as the analog precoder A_k, the signal y_k becomes

$$y_k = {w_k}^H r_k = {w_k}^H \mathbf{H}_k \sum_{n=1}^{K} \boldsymbol{A_n D_n s}_n + {w_k}^H \boldsymbol{n}_k \tag{19}$$

Full mathematical models, simulation results and comparative analyses of the above-mentioned three types of precoders are presented in [43]. The authors established that hybrid precoders outperform analog beamformers in terms of achievable rate (bps/Hz) and approaches the optimal performance of digital precoders, particularly as the number of BS antennas get large, which is good for mmWave massive MIMO systems.

Other hybrid precoding schemes such as the minimal Euclidean hybrid precoder [49], mean-squared error (MSE)-based precoding [50] and more advanced hybrid precoding technologies or circuitry such as the 1-bit Analog to Digital Converters (ADCs) [51] and beamspace MIMO [52] have recently been proposed, with a view to reducing the energy consumption and hardware cost of precoders. This is in order to make them realistic and practical for mmWave massive MIMO systems while keeping complexity modest. The indicated references provide further discussions on them, and a comparison of the three precoding techniques is summarized in Table V.

3) Channel Estimation: The availability of accurate CSI at the transmitter is key to downlink precoding operations in large-scale antenna systems for achieving optimal system performance. This CSI is obtained via uplink pilots in TDD-based systems with the assumption of channel reciprocity or fed back to the BSs by the UEs in FDD-based systems [22], [63]. In this subsection, we discuss channel estimation in TDD systems, while CSI feedback in FDD systems is discussed in Section IV-C.

TDD protocol exploiting channel reciprocity is popular for large antenna systems, where uplink pilots are used for channel estimation and subsequent downlink precoding and transmission. For mmWave massive MIMO systems, the knowledge of the CSI at the BS is very important in realizing spatial multiplexing [21], and is critical for high system performance [34]. Since both uplink and downlink in TDD systems operate on same frequency bands, the use of uplink pilots whose demand for time-frequency resources is only

TABLE V
COMPARISON OF PRECODING TECHNIQUES

Features	Analog beamforming	Digital Precoding	Hybrid (Analog-Digital) Precoding
Number of streams	Single-stream	Multi-stream	Multi-stream
Number of users	Single-user	Multi-user	Multi-user
Signal control capability	Phase control only	Phase and amplitude control	Phase and amplitude control
Hardware requirement	Least; one RF chain only	Highest; number of RF chains equal number of transmit antennas	Intermediate; number of RF chains less than number of transmit antennas
Energy consumption	Least	Highest	Intermediate
Cost	Least	Highest	Intermediate
Performance	Least	Optimal	Near-Optimal
Suitability for mmWave massive MIMO	Unsuitable; no amplitude control, no multi-user	Impractical; prohibitive cost and high energy consumption	Practical and realistic

dependent on the number of receive antennas is both feasible and realistic for channel estimation. With TDD, CSI is obtained using uplink training of pilots, as the downlink channel matrix is assumed a conjugate of the uplink channel matrix [21], [29], [35].

Uplink pilots should ideally be orthogonal from one terminal to the other. However, in multi-cellular MIMO systems, a dearth of orthogonal pilot sequences can result. This shortage becomes more pronounced in massive MIMO systems due to the large number of users, and much more difficult in mmWave massive MIMO due to shorter coherence interval. According to [21], the maximum number of distinct orthogonal pilot signals that can be assigned to users in each cell is given by (20).

$$Pilots_{max} = \frac{Coherence\ interval\ (T_{coh})}{Channel\ delay\ spread\ (DS)} \tag{20}$$

With a limited number of orthogonal pilot sequences, users in neighboring cells employ pilots which no longer remain orthogonal to those within the cell, thereby leading to the effect termed pilot contamination [21], [29]. This corrupts the CSI obtained at BS and reduces achievable throughput [34], [36]. Users in neighboring cells re-use pilots which are on the same frequency thereby generating directed interference. And unlike intra-cell interference, the generated directed interference does not disappear as the number of transmit antennas increases [29], which thus degrades system performance [13], [21], [34].

Pilot contamination is a challenge for mmWave massive MIMO systems as it presents a limit on the upper bound of the achievable sum-rate throughput. For avoiding intracell pilot contamination, orthogonal pilot transmission is employed, but it limits the number of users that can be scheduled in a cell within a coherence interval. This also leads to significant overhead in the uplink, as amply demonstrated by Marzetta [22], Fernandes *et al.* [41], and Jose *et al.* [135].

In orthogonal pilot assignment systems, the transmission time of each coherence interval of N symbols is divided into four overlapping phases: channel training, uplink data, processing time and downlink data as illustrated in Fig. 13. Here, though the uplink and downlink data are transmitted in superimposed, non-orthogonal mode and are separated by Space Division Multiple Access (SDMA), the pilot symbols must be transmitted in an orthogonal fashion and only limited K users corresponding to the number of allowable pilots can be scheduled. Under this setting, the number of active users that can be served is set to $\frac{N}{2}$ and the total number of data symbols that can be sent therefore is $[\frac{N}{2}] \times [\frac{N}{2}] = \frac{N^2}{4}$ [137].

One particularly notable and promising technique that is being proposed for mmWave massive MIMO system is the use of semi-orthogonal multiple access (SOMA) scheme, as against the conventional orthogonal pilot allocation and access scheme used in Orthogonal Frequency Division Multiplexing Access (OFDMA). SOMA proves to combat the problem of pilot contamination and eases resource allocation as more users can be scheduled within a coherence interval.

Khormuji [137] developed a novel uplink transmission mode employing SOMA, using a single cell multi-user massive MIMO scenario (extensible to mmWave massive MIMO regime). The SOMA approach not only extends the number of users that can be scheduled but also increases the cell's sum-rate throughput. In this access strategy, $K = N - 1$ users can be scheduled in each coherence bandwidth. The pilots are transmitted over orthogonal time-frequency resources thereby avoiding interference and pilot contamination, while the uplink data are transmitted such that some users appear orthogonal and others appear non-orthogonal (and hence the name semi-orthogonal), as illustrated in Fig. 14.

In [137], simulation results and comparative analyses of the achievable sum-rate by the conventional TDD, SOMA and its generalized form (GSOMA) are presented. The results show that SOMA can provide 70–100% gain in throughput as compared to the conventional TDD, with higher complexity of

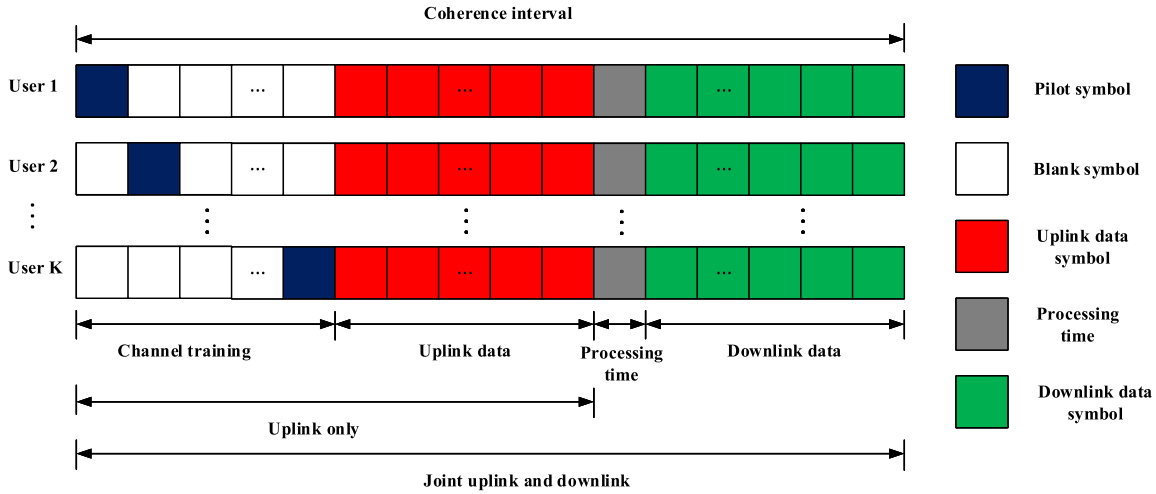

Fig. 13. TDD transmission protocol with orthogonal pilots.

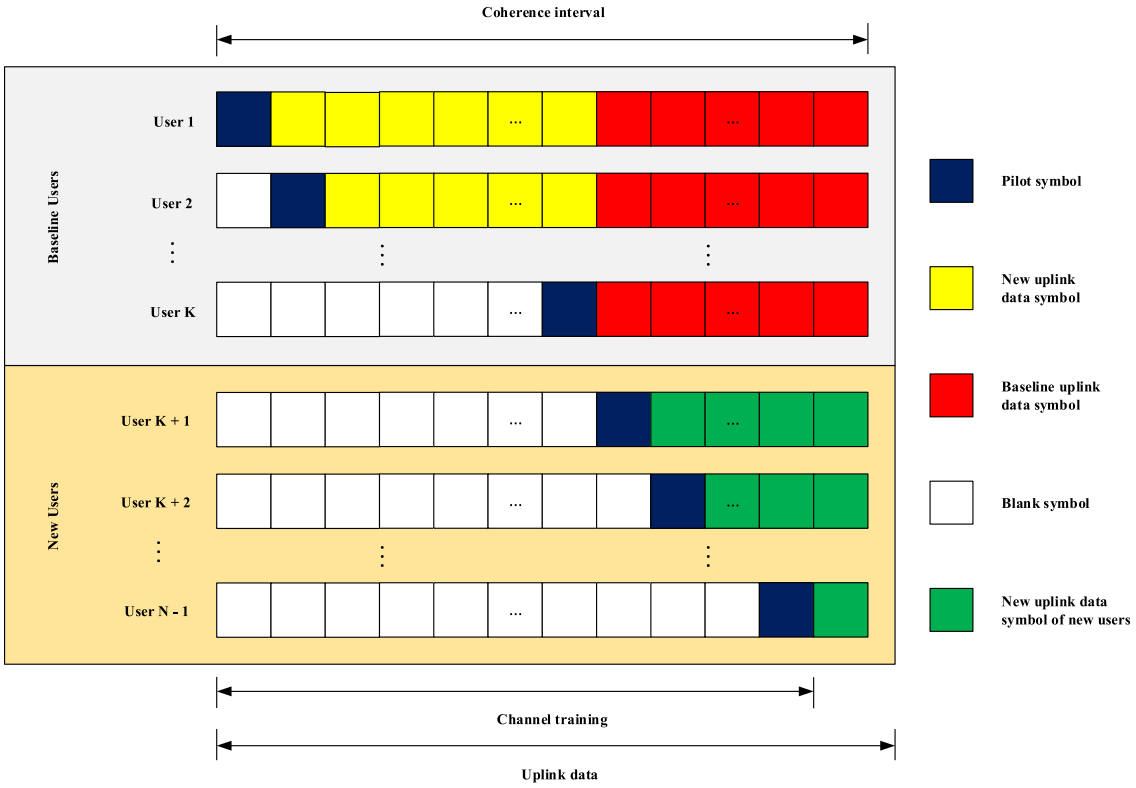

Fig. 14. TDD transmission with SOMA.

the receiver being the tradeoff cost for SOMA's enhanced performance.

The received superimposed noisy baseband signals/data are processed using sequential filtering and SIC at the receiver, employing minimum mean squared error (MMSE) channel estimator followed by spatial MF (see [137] for a full analysis of the model). The scheme is thus capable of receiving $\frac{N(N-1)}{2}$ error-free symbols from serving $K = N - 1$ users (upper bound). The SOMA scheme [137] achieves a throughput nearly twice as that of the baseline conventional TDD [22]. To ensure fairness among the $K = N - 1$ users such that each user achieves roughly the same average throughput, the ordering of the users (shown in Fig. 14) can be reshuffled at every coherence interval.

In mitigating the effects of pilot contamination in mmWave massive MIMO systems, several approaches have been proposed in [37]–[42] which largely classify into: pilot contamination precoding which takes network structure into account, blind techniques which avoids the use of pilots altogether or optimal pilot allocation/coordination techniques which adapt the use of pilot sequences or use less aggressive frequency reuse factors for pilots thereby making mutually contaminating cells far apart.

Accurate CSI is highly critical for the full exploitation of the huge potential benefits of mmWave massive MIMO systems [14], [53], but its acquisition via channel estimation is more challenging in mmWave massive MIMO systems than in conventional massive MIMO due to the following [54]:

- Doubly massive system with limited massive MIMO (dozens of the antenna) at the UE, which combined with the hundreds of antennas at the BS, result in high pilot overhead, even for TDD-based systems employing channel reciprocity
- Demand for both CSI at the Transmitter (CSIT) and CSI at the Receiver (CSIR), for precoding in the downlink and combining in the uplink, respectively, since both the BS and MS employ massive MIMO as against conventional massive MIMO where only the BS goes massive
- Special hardware constraints arising from low-cost, small-size and low-energy consumption considerations of the very large number of components (ADCs, Digital to Analog Converters (DACs), synthesizers, mixers and other hardware in the RF chains) for mmWave massive MIMO systems
- Low SNR before beamforming resulting from increased thermal noise with increased available bandwidth and increased Doppler shift due to the high carrier frequency and the blockage effect at mmWave frequencies

Inspired by the above challenges, cost and energy-efficient SOTA transceiver architectures that exploit the sparsity of mmWave channels and the low-rank property of mmWave massive MIMO channels have been proposed. This is to simplify channel estimation by using the concept of compressive sensing (CS). These transceivers use RF chains that are much smaller than the number of BS and UE antennas. Examples are the transceiver structure with analog phase shift networks (cascade of analog RF beamformer and digital baseband precoder), and the transceiver with low resolution (1-bit) ADC at the receiver. These transceivers exhibit negligible performance loss for mmWave massive MIMO when compared with conventional MIMO systems [14], [51], [54]. With respect to sparsity exploited with the concept of CS, channel estimation schemes can be classified as either CS-based or non-CS-based, as hereunder discussed.

(i) *CS-based Channel Estimation Schemes:* CS theory posits that an original signal which exhibits sparsity in some transformation domains can be recovered from far fewer samples than those required by the classical Shannon-Nyquist theorem [55]. Most real-world signals are inherently redundant or correlated, and therefore sparse. As a result, the effective information rates of continuous signals are much smaller than their bandwidth. Similarly, the number of effective degrees of freedom of corresponding discrete signals is usually much smaller than their dimensions [54].

Channel estimation (CE) in mmWave massive MIMO systems can thus be realized by transforming channel measurement signals into sparse matrices, compressing the sparse signals into signals with far lower dimensions than real mmWave massive MIMO channel estimates and finally recovering the original signals from the compressed signals. These three stages form the fundamentals of CS theory and CS-based channel estimation schemes [54], [55].

In mathematical terms, compressive sensing problems are formulated as follows. Consider a sparse signal x, which has a sparsity level k (i.e., x has k non-zero elements), and a measurement matrix Φ which converts x into the measurement signal y,

$$y = \Phi x \tag{21}$$

where $x \in \mathbb{C}^{N\times 1}$, $\Phi \in \mathbb{C}^{M\times N}$, $y \in \mathbb{C}^{M\times 1}$ and $M \ll N$.

If x does not exhibit sparsity itself, but sparsity surfaces in some transformation domains, where s is now the sparse signal with sparsity level k and Ψ is the transform matrix, the original signal, x, can now be expressed as

$$x = \Psi s \tag{22}$$

The measurement signal y then becomes

$$y = \Phi x = \Phi\Psi s = \Theta\ s \tag{23}$$

where $s \in \mathbb{C}^{N\times 1}$, $\Psi \in \mathbb{C}^{N\times N}$ and $\Theta = \Phi\Psi$.

The design of Θ should compress the dimensions of measurements while minimizing the information loss and ensure the reliable reconstruction of s from the limited measurement signal y. By exploiting the sparsity of mmWave massive MIMO channels, CS theory estimates large-sized channels from small-sized measurements, via the careful design of the transformation matrix which transforms the channel vectors into sparse signals. The first step is realized by formulating channel estimation as a CS problem followed by the reconstruction of the original sparse signals of high dimensions from the measurements of low dimension.

Classical orthogonal matching pursuit (OMP) algorithm is an example of sparse channel estimation schemes from which most other SOTA channel estimators derive their basis, due to its low complexity and good performance [51], [56], [57]. Multi-grid OMP (MG-OMP) is the improved version with its adaptive capabilities [56]. Both OMP and MG-OMP are based on the hybrid MIMO transceiver architecture. Maximal Likelihood (ML) estimators represent another set of CE schemes exploiting sparsity. Examples are the expectation-maximization (EM) algorithm, and its combination with the generalized approximate message passing (GAMP) called EM-GAMP. Their architecture is based on the 1-bit (low-resolution) ADCs at the receiver. The 1-bit ADC configuration has simpler architecture and consumes less energy, with only a slight loss in performance when compared with the other estimators [58].

(ii) *Non-CS-based Channel Estimation Schemes:* Parametric CE schemes such as the super-resolution sparse channel estimator and its multi-user multi-stream hybrid beamforming/combining counterpart (Heuristic algorithm) [59] use a limited set of channel parameters (path gains, AoAs and angles of departure (AoDs)) to fully estimate the channel. In addition, they employ continuous AoAs and AoDs as against discrete sets of AoA and AoD used by CS-based schemes. They also use architectures with reduced number of RF chains to realize good system performance that approaches

the optimal performance of fully-digital precoding/combining schemes [54].

The problems with parametric channel estimators are that they require *a priori* information of the array manifold and estimate the entire full channel. Sub-space Estimation and Decomposition (SED)-based CEs, which have been employed in conventional MIMO systems can be extended to mmWave massive MIMO systems. They do not require *a priori* information and only estimate the eigenmodes of the dominated singular values which are effective and sufficient for transmission. However, these schemes employ multiple amplify and forward (AF) operations which inadvertently introduce much noise and degradation in the accuracy of the channel estimation process [54]. Analysis of the above parametric CE schemes are fully presented in [54].

The channel estimation schemes proposed for mmWave massive MIMO channels so far have practical limitations and are still subject of extensive research in some key areas. First, while optimal digital precoding and combining patterns have been proposed, the corresponding optimal analog beamforming and combining pattern design is still a subject of research. Second, the assumption of discrete AoA/AoD in sparse channel estimation algorithms leads to quantization error, as against continuous AoA/AoD in practice. And third, adaptive CS-based CE schemes would have to be developed for special scenarios (such as indoor office) where mmWave massive MIMO channels are non-sparse [54]. Along with these, modifications to CE schemes used in conventional systems [60]–[62] such as the widely-adopted/standardized codebook-based channel estimators may be helpful in the search for optimal CE schemes for mmWave massive MIMO systems [54].

In scaling up link capacity of mmWave massive MIMO systems, beamforming and spatial multiplexing are essential, and both depend on reliable CSI acquisition, whether for Frequency Division Duplexing (FDD) or Time Division Duplexing (TDD) mode. TDD-based CSI acquisition is more common in mmWave massive MIMO systems as it is more economical to send pilots from the UEs to the BSs, as the number of users is far lesser than the BS antennas. Extensive works on CSI acquisition are available in the literature. However, application to the mmWave band has to factor in the effects of less coherence time due to Doppler shifts [101], [102], higher CSI learning overhead particularly for mobile environment, higher probability of link outage due to beamforming mis-alignment and blockage [80] and loss in available DoF resulting from hybrid precoding employing lower number of RF chains than antennas [90]. Research efforts have identified increased training period as important for improved and accurate CSI acquisition [103], and the optimization of the training duration can be formulated as a concave maximization problem with global optimal solution obtainable via convex optimization tools such as CVX [104].

Besides CS and non-CS-based CE schemes, precoding coefficients can be designed without recourse to either channel estimation or feedback. In this scheme, each BS relies on its own capability of receive beamforming by using previous uplink signaling from the UE for partial CSI acquisition. Such signaling (e.g., random access request, acknowledgement of previous downlink transmission, etc.) do not contain downlink CSI information and this is without additional usage of transmission bandwidth or radio resources [65].

Similar to the concept of channel reciprocity used by TDD-based cellular systems, FDD systems can exploit the property that for a particular UE, AoAs at the BSs have strong correlations with the AoDs from that UE [67]. In this approach, downlink precoders/beamformers steer transmit power to desired directions such that target UEs receive the most power in a way that limits or minimizes interference from neighboring UEs. This is implemented using a beamspace division method where the angular domain space is divided into M mutually orthogonal beams. To reduce interference to other UEs due to large beamforming sidelobe power (if the number of BS antennas is small) or minimize the sensitivity to errors resulting from narrow beamwidth (for large size antenna system), these precoder designs are proposed to boost transmission energy to desired UEs and suppress interference caused to other UEs located at different locations.

Some of the other multi-user precoder designs with no CSI feedback being explored include the Diagonal Loading (DL) approach [68] from radar technology and the Classical Parks-McClellan (PM) algorithm [69] based on spatial-domain filter design approach. Channel models, simulations, and analysis of the proposed precoder designs in comparison with the beamspace division method and the ideal Block Diagonalization (BD) method are presented in [65]. Even without CSI feedback, these two proposed algorithms only exhibit slight performance degradation (in terms of sum rate (bps/Hz)) when compared to the ideal BD method which requires full, accurate CSI feedback [65].

B. Channel Measurement and Modeling

The mmWave massive MIMO channel is intrinsically an ultra-broadband channel with huge bandwidth and spatial multiplexing capability to significantly enhance wireless access and improve cell and user throughputs [8], [70]. On propagation mechanism, mmWave signals exhibit LOS or near-LOS propagation with absolute increase in pathloss with increasing frequency [72], specular reflection attenuation [73], diffuse scattering [74], [75], very high diffraction attenuation [76] and frequency dispersion effect, which with the prospect of huge bandwidth, allows the propagation to be considered as frequency-dependent [22], [77].

So far, there is no official or widely acknowledged mmWave massive MIMO channel either in academia or the industry. While research and standardization efforts are being undertaken for the mmWave massive MIMO channel models, distinctive features peculiar to large antenna arrays at mmWave frequencies have to be critically considered in the areas of propagation mechanisms for future applications, both in static and dynamic environments [71].

Under static conditions, mmWave massive MIMO channels exhibit frequency-selective fading which would need to be addressed by modulation or equalization [8]. At mmWave frequencies, the assumption of asymptotic pair-wise orthogonality between channel vectors under i.i.d Rayleigh

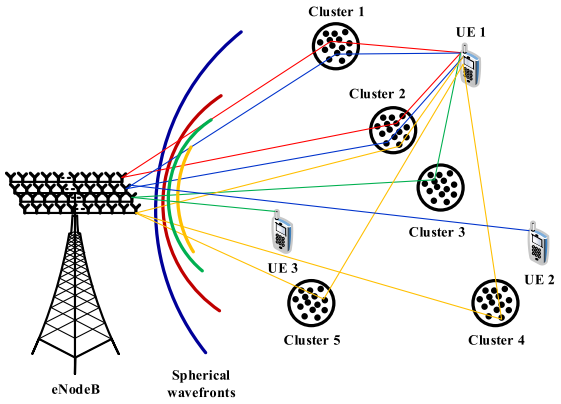

Fig. 15. Illustration of non-stationary and spherical wavefront phenomena for mmWave massive MIMO.

fading channel which is valid for conventional massive MIMO no longer holds. The number of independent multipath components (MPCs) becomes limited and so channel vectors exhibit correlated fading [8], [71].

For a mutual orthogonal channel, every pair of column vectors of the channel **H** satisfies the condition

$$h_i{}^* h_j = 0 \text{ for all } i \neq j \tag{24}$$

With the assumption of i.i.d. Rayleigh fading channel, the condition in (24) is asymptotically achieved with very large Tx antenna numbers (N), which by the law of large numbers gives

$$\frac{1}{N} h_i{}^* h_j \to 0 \text{ as } N \to \infty \text{ for } i \neq j \tag{25}$$

However, mmWave massive MIMO channels are neither i.i.d. nor is the number of Tx antennas (N) infinite. Therefore, the condition for mutual orthogonality in (24) cannot be satisfied in reality [71]. MmWave massive MIMO channel models, therefore, have to consider this non-orthogonality for propagation in realistic environments (refer to [70] for proof of non-orthogonality based on realistic outdoor propagation field tests). Similarly, the assumption of planar waves in conventional massive MIMO has to be replaced with a more appropriate spherical waves representation (as shown in Fig. 15). Also, spatial non-stationality which becomes more severe has to be considered in mmWave massive MIMO channel models [71], [72], [78]. Extensive indoor and outdoor channel measurement campaigns and simulations have been carried out by [16], [70], and [79], among others, to deduce these outcomes.

Also, for practical realization of such channels, the corresponding high computational rate required for CSI estimation and feedback in such high-mobility channels have to be factored into the design [8]. Field measurement (channel sounding) and mathematical analysis (channel modeling) are the two techniques employed for the development of channel models. The overview of both methods with respect to mmWave massive MIMO is given in the following sub-sections.

1) Channel Measurement: Channel measurement techniques are widely used for studying the properties of wireless channels. With field measurement results, the statistical and/or deterministic properties and profiles of the channels are extracted and modeled. Sounding techniques are being used extensively in wireless channel modeling, and more recently for the mmWave frequency bands. Field measurement campaigns by the New York University (NYU) Wireless team at 28 and 73 GHz in New York City (USA) [80], mmMAGIC group for mmWave frequencies 10-100 GHz in Berlin (Germany) used for the QuaDRiGa channel model implementation [81], [82], Zhao *et al.* [83] in Beijing (China) at 32 GHz, METIS at 50–70 GHz [200] and the 5GCM at 6–100 GHz [201] are quintessential examples of these efforts. Reference [210] provides an excellent overview of recent, cross-country mmWave channel measurement and modeling efforts between 2012 and 2017.

While most existing mmWave channel measurements were done for static radio channels, the world's first demonstration for the dynamic channel was carried out by Samsung at 28 GHz, 110 kmph speed for 1.2 Gbps transmission in 2014. Contrary to earlier thoughts, research and field trials point to the suitability of mmWave frequencies for mobile applications, particularly for pico and femto cells which typically do not involve high-velocity users. However, moving users and/or transceivers introduce some dynamics in the channels which necessitate that the models incorporate the correlation between mobility and channel dynamics [71].

For mmWave massive MIMO channels, technical capabilities of sounding equipment (both hardware and software) must surpass those of prior channels for improved accuracy of measurements and corresponding models. Efficient channel sounders must therefore be capable of measurements up to several tens of GHz and several GHz of bandwidth (mmWave) as well as tens to hundreds of antennas at both the BSs and UEs (massive MIMO), with corresponding high-performance related components in the RF chain (including power amplifiers, low noise amplifiers, up and down convertors and RF switches etc).

Similarly, sounding equipment should be capable of measurements in static and dynamic conditions, indoor and outdoor environments, as well as have capabilities for an efficient signal generation, data acquisition, and storage. The equipment should also guarantee accurate synchronization, calibration, sensitivity, resolution and flexibility for an extension while maintaining reasonable system cost [71]. High performance sounding equipment are important for mmWave massive MIMO channel modeling, simulation, test, and validation.

2) Channel Modeling: Usually, channel measurement activities are followed by analytical channel modeling to develop mathematical and analytical frameworks to compare and/or validate empirical data from field tests, and for simulation studies and performance evaluation of the communication systems. The aim of channel models is to estimate channel parameters considering all necessary factors (such as frequency, propagation environment, etc.). Most recent channel model efforts are adopting the Third Generation Partnership Project (3GPP) three-dimensional (3D) geometry-based stochastic channel model (SCM) approach. The model generally follows the flow chart in Fig. 16 to estimate channel parameters [165], [166].

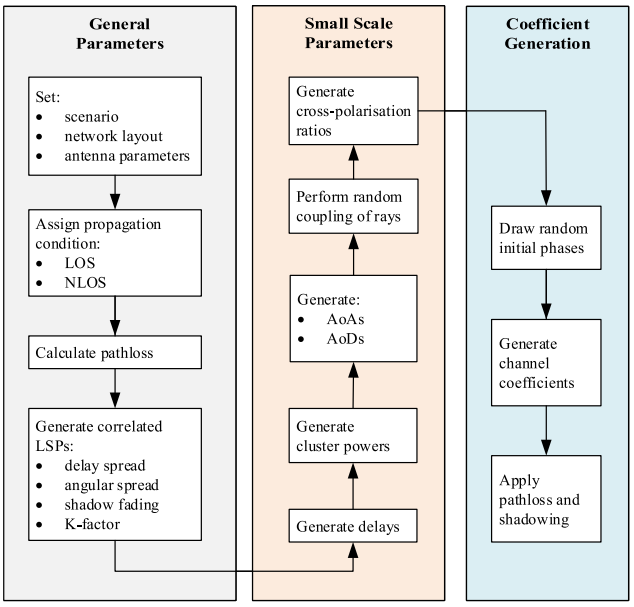

Fig. 16. Flow chart for the 3GPP 3D geometry-based SCM.

References [80]–[83] and [167]–[174] provide a good read on mmWave channel modeling using geometry-based SCM approach. Though two-dimensional (2D), a quintessential model for mmWave massive MIMO channel based on the spherical wavefront and birth-death process assumption, using a non-stationary wideband multifocal elliptic channel model approach has been proposed in [71].

Overall, a narrowband clustered channel representation (based on extended Saleh-Valenzuela model) is proposed for mmWave massive MIMO channel, as it allows accurate capturing of the characteristics of mmWave channels. Under this clustered model in (26), the discrete-time narrowband channel matrix $\boldsymbol{H}$ is assumed to be a sum of the contributions of L propagation paths [26], [92],

$$\mathbf{H} = \sqrt{\frac{NM}{L}} \sum_{l=1}^{L} \alpha_l \Lambda_r\left(\phi_l{}^r,\ \theta_l{}^r\right) \phi_t\left(\phi_l{}^t,\ \theta_l{}^t\right) \mathbf{a}_r\left(\phi_l{}^r,\ \theta_l{}^r\right) \times \mathbf{a}_t{}^*\left(\phi_l{}^t,\ \theta_l{}^t\right) \tag{26}$$

where N, M and L refer to the number of transmit antennas, receive antennas and paths respectively. α_l is the complex gain of the lth path, $\phi_l{}^r$, $\theta_l{}^r$ are the azimuth and elevation Angles of Departure (AoDs) respectively, while $\phi_l{}^r$, $\theta_l{}^r$ represent the azimuth and elevation Angles of Arrival (AoAs), respectively. The vectors $\mathbf{a}_r(\phi_l{}^r,\ \theta_l{}^r)$ and $\mathbf{a}_t{}^*(\phi_l{}^t,\ \theta_l{}^t)$ represent the normalized receiver and transmitter array response vectors at the azimuth (ϕ) and elevation (θ) angles respectively. $\Lambda_r(\phi_l{}^r,\ \theta_l{}^r)$ and $\Lambda_t(\phi_l{}^t,\ \theta_l{}^t)$ are the receiver and transmitter antenna element gain for the AoAs and AoDs respectively, which can be set to one within the range of the AoAs and AoDs, for simplicity and without loss of generality [183], [184].

System-level performance evaluation of cellular networks is usually performed via numerical simulations or mathematical analyses, both of which are widely regarded as difficult, time-consuming and demanding, with respect to ensuring high confidence levels, employing sufficient and representative data sets and parameters, and incorporating all network elements that can impact performance [148], [149]. In addressing these challenges, stochastic geometry modeling is emerging as a new and tractable mathematical tool for the analysis of current and next-generation cellular networks. Di Renzo *et al.* [151], [152], [156]–[158], Kutty and Sen [154], and Lu and Di Renzo [159] provide good literature reviews on the stochastic geometry approach.

Stochastic geometry models capture the inherent elements and impacting factors of the networks. It models the locations of the BSs using the concept of Poisson Point Process (PPP). Researchers are now turning their attention in the direction of this flexible mathematical tool (i.e., stochastic geometry employing PPP-based abstraction model) for the analysis and performance evaluation of mmWave cellular networks. This is with the aim of developing mathematical frameworks that model realistic mmWave propagation in a way that incorporates its characteristics (e.g., path loss models) and peculiarities (such as blockage models) and that distinguishes it from conventional systems [92], [160]–[162].

Stochastic geometry modeling for mmWave massive MIMO employs the three-state link model covering the LOS, NLOS, and Outage (OUT) links. According to [80], the probabilities of a link in any of the three states as a function of distance (d) is given by (27)

$$\begin{aligned} p_{OUT}(d) &= max\left\{0,\ 1 - \gamma_{OUT} e^{-\delta_{OUT} d}\right\} \\ p_{LOS}(d) &= (1 - p_{OUT}(d)) \gamma_{LOS} e^{-\delta_{LOS} d} \\ p_{NLOS}(d) &= (1 - p_{OUT}(d)) \left(1 - \gamma_{LOS} e^{-\delta_{LOS} d}\right) \end{aligned} \tag{27}$$

where (δ_{LOS}, γ_{LOS}) and (δ_{OUT}, γ_{OUT}) are parameters that depend on the propagation scenario and the carrier frequency. The LOS state implies that a direct path exists between a BS and a UE, NLOS means the direct path is blocked but there are still paths between the two elements (e.g., via reflections) while the OUT state occurs when the path loss between the BS and the UE is so high (assumed infinite) that no link can be established between them. With the assumption that the BSs are modeled as points of a homogeneous PPP, and the events of the three link states are independent, then each state can be modeled as an independent and nonhomogeneous PPP, such that $\Psi = \Psi_{LOS} \bigcup \Psi_{NLOS} \bigcup \Psi_{OUT}$ [148].

C. Receiver Processes and Techniques

The operations at the receiver block include detection of signals, feedback of CSI and combining operations, among others. We, however, limit the discussion to the downlink and so, we only present an overview of the proposed channel feedback techniques in this sub-section.

FDD-based systems use different frequency bands for the uplink and the downlink. While the uplink CSI estimation depends on the number of receive antennas, the downlink counterpart requires large amount of time-frequency resources which scales with the number of transmit antennas and becomes unrealistic and infeasible in large antenna systems [21], [29], [35]. Some of the techniques

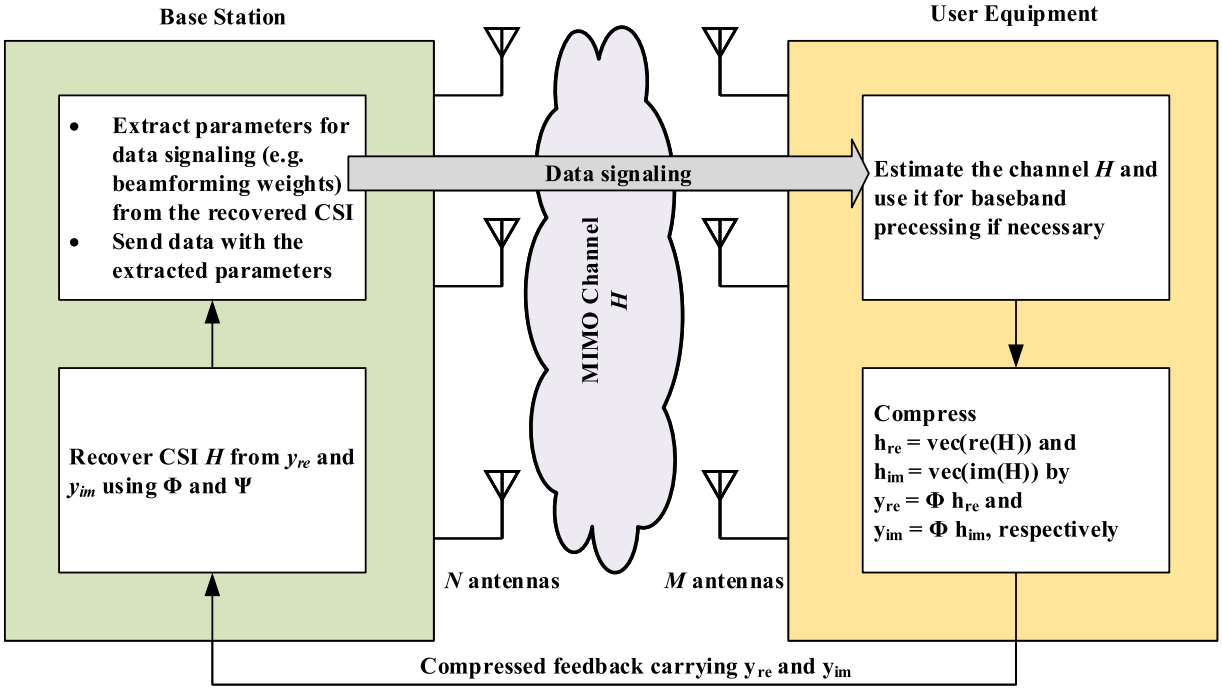

Fig. 17. CSI feedback framework based on compressive sensing.

being explored for the reduction of the burden of CSI feedbacks in mmWave massive MIMO systems include: compressive sensing-based CSI feedback and CSI feedback based on multi-stage beamforming [65].

(i) *Channel Feedback With Compressive Sensing:* CS, introduced in Section IV-B2, can also be adopted to reduce the feedback overheads required for CSI reporting by compressing the channel responses. With a sparsifying basis or sparsifying transformation matrix Ψ, the channel matrix $\mathbf{H}$ is compressed and encoded as a $K \times 1$ vector with a compression ratio $\eta = \frac{K}{M \times N}$ where K is smaller than $M \times N$. $\boldsymbol{S}$ is the sparse form of H, and the channel matrix $\mathbf{H}$ can be recovered by applying the inverse sparsifying transform: $\boldsymbol{H} = \Psi^T \boldsymbol{S}$. It is important to select a sparsifying transform that guarantees fewer non-zero elements in $\boldsymbol{S}$ to ensure higher accuracy in signal recovery [65].

A framework for CSI reporting based on CS approach is illustrated in Fig. 17. Examples of CS-based feedback mechanisms are the two-dimensional discrete cosine transformation (2D-DCT), the Karhunen-Loeve transformation (KLT) and the hybrid of both which evolves into a CSI feedback protocol with an adaptive sparsifying basis as discussed in [65].

Compact physical sizes resulting from the small antenna spacing of massive MIMO arrays dictate that some of the infrastructure nodes such as small cell BSs and remote radio heads (RRHs) have channel matrix entries with strong spatial correlation [66], and therefore a sparse representation in certain domains other than space [65]. This approach potentially reduces the required CSI feedback payload for massive MIMO systems. However, the assumption of high spatial correlation among the antennas makes the CS-based CSI feedback approach, which relies on sparse channel representation, invalid for mmWave massive MIMO systems [65].

(ii) *Channel Feedback With Angular-Domain Beamforming:* This approach explores the technique of multi-stage beamforming and feedback where UEs only perceive signal energy and estimate channels in a few beamforming directions. This reduces the dimensionality of the desired channel as the number of propagation paths in such a beamformed signal is much lower than in conventional systems where CSI of a large number of antennas are measured and reported [65]. While the two-stage beamforming comprises the beam selection and the beam-based CSI reporting phases, the three-stage beamforming has an additional intermediate phase where the user measures the transmitter correlation matrix and feeds back the eigenvectors to the BS. These two examples of multi-stage beamforming techniques are illustrated in Figs. 18 and 19, respectively.

Overall, both TDD and FDD modes have pros and cons. Though TDD simplifies CSI acquisition with the use of uplink pilots particularly for mmWave massive MIMO systems and is easily adopted for modeling and simulation studies, it does not account for ICIs specific to downlink transmission, it is prone to inaccuracies due to transmitter-receiver mismatch and as well as prone to pilot contamination effect which limits system performance [39], [64].

FDD systems, on the other hand, use dedicated feedback mechanism where UEs report measured channel information to the BSs, which is more accurate and less sensitive to latency, since the uplink and downlink transmissions operate at different frequencies. Accurate CSI is critical for improved systems performance, and decreased latency is important for mission-critical applications in next-generation networks. However, dedicated feedback in FDD systems consumes radio resources,

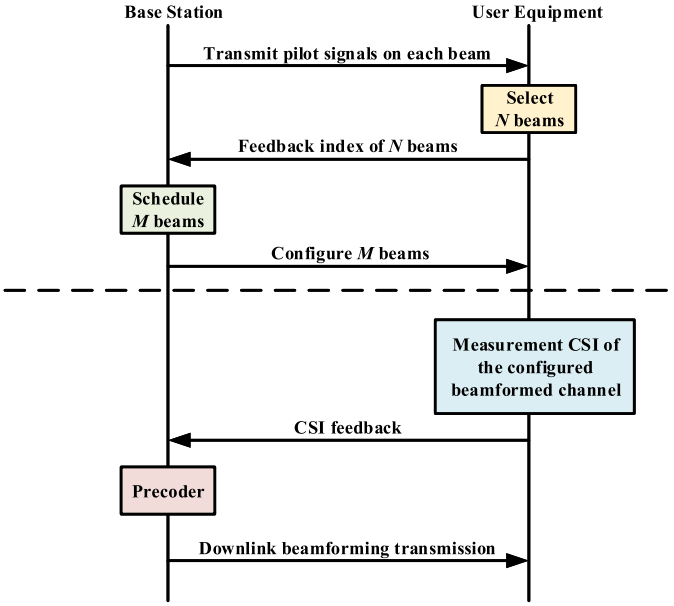

Fig. 18. Flow chart of two-stage beamforming for CSI acquisition.

particularly in mmWave massive MIMO systems, thereby resulting in massive overhead. It also requires the use of larger codebooks thus necessitating lower coding rates which inadvertently affects the reliability of the CSI feedbacks [65]. System performance metrics of interest will therefore dictate the duplexing mode to adopt and the required design tradeoffs.

D. Summary and Open Issues

The techniques for mmWave massive MIMO at the three communication blocks (i.e., transmitter, channel and receiver) have been reviewed in this section, based on the downlink set-up. MmWave signals have propagation features that are markedly different from those of μWave systems. These distinctive features have to be incorporated in the new channel models. Field measurement campaigns and analytical modeling, which are important for the simulation, test and validation of system performance, are being undertaken by various research groups and projects. Also, prototypes of hybrid antenna arrays are already being built and tested. In addition, various algorithms, schemes and techniques are being proposed for mmWave massive MIMO systems in the areas of precoding, channel estimation, channel feedback, among others. We present a summary of these techniques and concepts in Table VI.

For mmWave massive MIMO, there are still many issues that are open for discussion and that require finalizing. Many of the proposed concepts and techniques which have been overviewed here are still in the exploratory phase. While energy and spectral efficiency analysis of networks employing mmWave massive MIMO are ongoing, performance in real-life deployments still remains an open issue. More channel measurement campaigns in different parts of the world are also still needed for representative channel models.

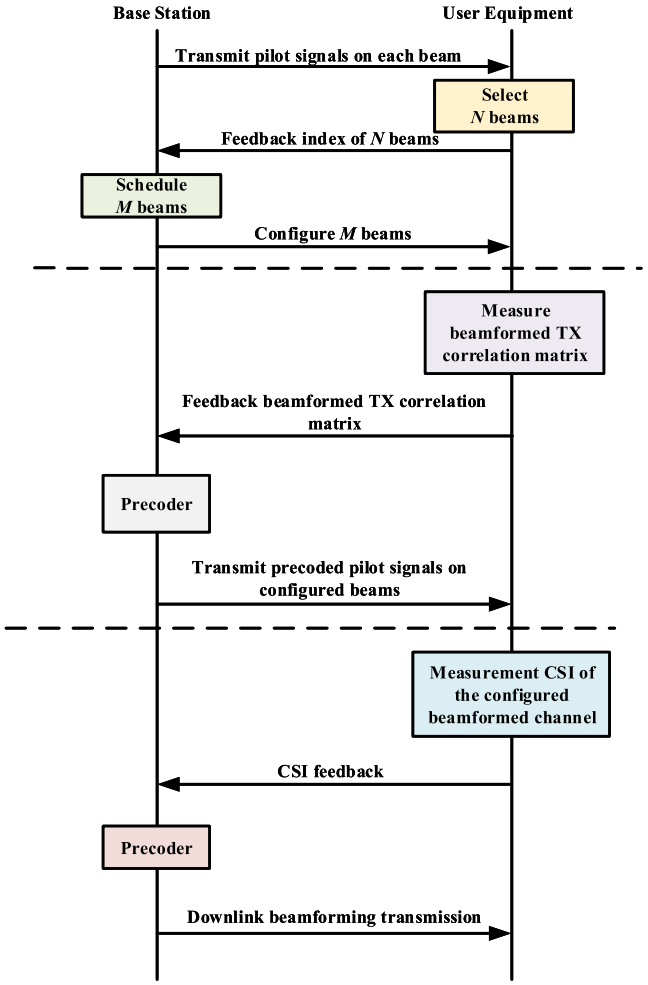

Fig. 19. Flow chart of two-stage beamforming for CSI acquisition.

V. Cross-Layer Design Trends and Challenges

In this section, we present an overview of cross-layer design considerations on the emerging trends and evolving techniques being explored towards realizing mmWave Massive MIMO, further to the major system blocks surveyed in Section IV. The considerations span the areas of signal waveforms, multiple access strategies, user scheduling, fronthaul design, standardization and health and safety issues. We highlight the trends and present the corresponding challenges and proposed solutions.

A. Waveform Design

The design of transmit signal waveforms is one of the determinants of link-level performance of wireless channels. The type of waveform influences system performance and link-level capacity with respect to channel estimation, signaling, beamforming, and realizable multiplexing gains [90], [91], [105].

Orthogonal Frequency Division Multiplexing (OFDM) has been chosen by a number of recent wireless standards including the 4G LTE/LTE-A, WiFi, worldwide interoperability for microwave access (WiMAX), digital video broadcasting – version 2 (DVB-2) etc., based on

TABLE VI
SUMMARY OF MMWAVE MASSIVE MIMO TECHNIQUES

Communication Block	Concept	Features/Techniques/Schemes/Algorithms and References	Remarks
Transmitter	Antenna Structure	• Interleaved or localized [43], [179]; • linear or planar [43], [179]; • Analog, digital, hybrid [177]-[180].	Localized arrays have better performance than interleaved, UPAs allow 3D beamforming which is impossible with ULAs, and hybrid antenna arrays are more practically feasible for mmWave massive MIMO.
	Precoding	Analog beamforming: • codebook-based beam training with exhaustive search [43]; • hierarchical codebook-based beam [44]; • two-step approach based on exhaustive search for best beamformer-combiner pair [45]	Analog beamformers have poor performance, digital precoders have optimal performance but costly while hybrid precoders have good performance and more practically feasible for mmWave massive from implementation and cost perspectives.
		Digital precoding: • Single-user linear precoder - MF, ZF and WF [43]; • multi-user linear precoder: DPC [46], TH [47], BD [185]	
		Hybrid Precoding: spatially-sparse [43]; SIC-based [43]; two-stage [48]; Minimal Euclidean [49]; MSE-based [50]; 1-bit ADC [51]; Beam space MIMO [52]	
	Channel Estimation	Compressive Sensing (CS)-based: orthogonal matching pursuit (OMP) [51], [56], [57]; Multi-grid OMP (MG-OMP) [56]; Expectation-maximization (EM) with the generalized approximate message passing (GAMP) called (EM-GAMP) [58]	There are still practical limitations: non-optimal analog beamforming and combining schemes for the hybrid precoders, quantization errors due to assumption of discrete AoAs and AoDs and application in special scenarios with non-sparse channels TTD-based channel feedback is simple due reciprocity, but it is prone to inaccuracies that can limit system performance.
		Non-CS-based: AoA-based super-resolution sparse channel estimator [59]; Sub-space Estimation and Decomposition (SED)-based [54];	
		TDD-based systems: uplink pilots due to channel reciprocity, with Semi-Orthogonal Multiple Access (SOMA) [137]	
Channel	Properties	LOS/near-LOS propagation [72]; specular attenuation [73]; diffuse scattering [74], [75]; high diffraction attenuation [76]; frequency dispersion effect [22], [77]; non-orthogonal [70]; frequency-selective fading [8]; correlated fading [8], [71]; non-statiality [71], [72], [78]; spherical wavefront [72].	mmWave channel models have to incorporate the effect of these properties, thereby discarding earlier assumptions such as planar waves and mutual orthogonality of conventional systems
	Sounder Features	Measurement capabilities: mmWave frequencies (up to 100 GHz), bandwidth (several GHz), static and dynamic condition, indoor and outdoor environment, efficient signal generation, processing and storage, moderate cost [71]	High performance sounding equipment are important for mmWave massive MIMO channel modeling, simulation, test and validation.
	Measurement Campaigns	• NYU Wireless (28 and 73 GHz) [80]; • QuaDRiGa/mmMAGIC (10-100 GHz) [81], [82]; • Zhao *et al.* (32 GHz) [83], • METIS at 50–70 GHz [200], • 5GCM at 6–100 GHz [201].	Measurement campaigns for 60 GHz has been excluded from the list. It is being used for indoor WLAN application and is considered unsuitable for outdoor cellular due to high oxygen and atmospheric absorption
	Models	• 2D mmWave model based on spherical wavefront and birth-death process [71]; • 3GPP 3D SCM [80]-[83], [167]-[174]; • Stochastic geometry [150]-[159].	Most models follow the 3GPP 3D SCM guidelines. The stochastic geometry analytical models capture the outage (OUT) state, alongside the LOS and NLOS
Receiver	Channel feedback	FDD-based systems: CS-based – 2D-DCT, Karhunen-Loeve transformation (KLT) and hybrid 2D-DC-KLT [65]; multi-stage beamforming [65], AoA-based [67]. Non-CSI-based: Diagonal Loading (DL) [68], Classical Parks-McClellan (PM) algorithm [69]	FDD-based feedback is more accurate and experiences less latency but consumes radio resources and has massive signaling overheads. A trade-off thus exists, depending on performance metrics of interest.

its numerous advantages over prior designs such as the time division duplexing (TDM), frequency division duplexing (FDM) and code division duplexing (CDM) of the earlier generations of cellular systems [108]. It transforms frequency-selective channels into multiple parallel flat-fading channels, uses easily-realizable fast Fourier transform (FFT)

TABLE VII
CANDIDATE WAVEFORMS FOR mmWAVE MASSIVE MIMO

Waveform	Description
GFDM	This generalized FDM has the flexibility to be configured as either OFDM or SC-FDM. This implies that the available bandwidth could be split into either several narrow-band sub-carriers or a smaller number of subcarriers of wide individual bandwidth. It is a block-based modulation approach, and its design reduces transceiver complexity and equalization and synchronization issues.
FBMC	The sub-carriers in FBMC are subjected to sidelobe suppression by passing them through a filter bank, which makes them robust to inherent synchronization challenges. It employs features training sequences for synchronization and channel estimation purposes, has flexible resource allocation in the time-frequency domains and is capable of delivering higher spectral efficiency than OFDM.
UFMC	This is a multi-carrier signal format that employs orthogonal sub-carriers to handle the problem of loss of orthogonality at the receiver end. It is a type of generalization of filtered OFDM and FBMC. However, unlike OFDM, it uses sub-band filters in place of cyclic prefixes. The short duration filters are more suitable for short burst communication than FBMC.
BFDM	Bi-orthogonal frequency-division multiplexing (BFDM) is based on the principle of replacing the orthogonality requirement (as in OFDM) of a set of a transmitter and receiver with bi-orthogonality, a relaxed form of orthogonality. In this scheme, the time-frequency representations of the transmitted and received pulses are to be pair-wise orthogonal and not individual. BFDM is more robust than OFDM to the frequency offset in transmission.

and inverse FFT (IFFT) operations (implementation-wise) and has low-complexity multiplexing capability for multi-user support [91], [106], [107].

However, the disadvantage of OFDM in terms of high peak to average power ratio (PAPR), which demands the use of high-quality and high-cost power amplifiers (PA) and results in severe degradation in bit error ratio (BER) performance [109], makes it unattractive for mmWave massive MIMO systems, and favors the use of single carrier transmission schemes [105], [110]. Though single-carrier FDM was adopted for the uplink of 4G systems, research trends point towards single- carrier-enabled waveform design approaches for mmWave systems too [108].

Other modulation techniques/waveforms are being considered for 5G due to the challenges of OFDM. The waveforms include filter bank multi-carrier (FBMC) [186], universal filtered multi-carrier (UFMC) [187], generalized FDM (GFDM) [188] and bi-orthogonal FDM (BFDM) [189].

These waveforms are candidate waveforms for mmWave massive MIMO due to their relaxed orthogonality and synchronization constraints. Table VII presents a description of the features of these candidate waveforms [190].

Nonetheless, 3GPP has decided to retain OFDM for the Phase 1 of the 5G New Radio (NR), with spectral confinement technologies such as filtering and windowing, thereby leading to filtered OFDM (f-OFDM) and windowed OFDM (w-OFDM), respectively. Use and combination of other waveforms are kept forward as a possible option for Phase 2 [223].

B. Access Scheme and User Scheduling

Another factor which influences link level performance of the mmWave massive MIMO system is the access strategy. To this end, research trends orient in the direction of separation architecture where the user plane and control plane are decoupled [26], [112]. In this network topology, critical control data (for mobility management, synchronization, resource allocation, etc.) are reliably transmitted on μWave links between the macrocell BSs and UEs while the small cells handle the high-capacity and spectrally-efficient data plane services on mmWave frequency bands. This layout leads to enhanced reliability in communication, higher cell capacity and improved user throughput [3], [13], [111].

Besides access links, the mmWave frequency bands are also being promoted for backhaul networks with a view to advancing deployment flexibility [26]. For the joint design of access and backhaul networks, several research solutions are being explored. The design and implementation challenges in this respect are discussed in [113], with scheduling and resource allocation challenges considered in [114]–[116].

MAC refers to the strategies or schemes with which a BS communicates with multiple users in the downlink and by which multiple users communicate with the BS in the uplink. MAC designs entail how communication resources (time, frequency and space) are assigned to the user(s) [117]. With the adoption of packet-switched networks for cellular systems since the late 1990s (i.e., from 2.5/3G systems), user scheduling and multi-user diversity have remained two key aspects of MAC designs [118].

User scheduling is the process of dynamically selecting some users for each communication resource block among many active users in the cell based on their channel conditions while multi-user diversity refers to the diversity gain realized from selecting a user or users with good channel conditions among many users with independent channel fading for each resource block, such that the overall system performance is optimized [91], [118], [119].

User scheduling can be in one, two or three dimensions of time, frequency, and space. The OFDM/OFDMA which was adopted as the physical layer transmission technology in 3GPP LTE allows two-dimensional time-frequency resource allocation. With the incorporation of MIMO technology, MIMO-OFDM networks allow user scheduling in three dimensions of time, frequency, and space. For each resource block, multiple data streams can be transmitted to either only one user (SU-MIMO) or to multiple users simultaneously (MU-MIMO) in the three dimensions, where multiple spatially-separated users can be scheduled simultaneously on the same time-frequency resource block based on SDMA using beamforming. MU-MIMO is capable of realizing full multiplexing gain thus achieving optimal sum-rate capacity of MIMO broadcast (BC) networks using smart scheduling algorithms, compared to SU-MIMO which is capable of realizing only a small fraction of the capacity, particularly when the number of BS transmit antennas is larger than receive antennas [120]–[122].

With the evolution of massive MIMO, user scheduling becomes more simplified due to the concept of channel hardening where small-scale channel fading vanishes. This

thus allows the entire bandwidth to be allocated to each scheduled user separated by MIMO in the spatial domain. The challenge here, however, is that large amount of signaling (CSI feedback) is required from each user due to the large dimension of channel vectors. Several scheduling algorithms and resource allocation techniques working with non-ideal or partial CSI have been proposed in [123]–[127] for massive MIMO cellular systems without significant loss in system performance. These new scheduling algorithms have largely been explored for the sub-6 GHz μWave frequency bands. Thus, they have to be revised for mmWave massive MIMO systems due to the remarkably different propagation characteristics in the μWave and mmWave bands. The works of Lee *et al.* [128]–[131] are quintessential in this respect. Beyond the basic schedulers such as the maximum rate, round robin (RR) and proportionally fair (PF) schedulers, more efficient scheduling algorithms such as the DPC and the greedy user selection method based on ZF beamforming and DPC (gZF-DP) [132] have been proposed and developed, and more recently the low-complexity random beamforming (RBF) [121] and semi-orthogonal user selection (SUS) [122] which are easier for practical implementation [117].

Using RBF as a case study, Lee and Sung [117] showed that scheduling based on partial CSI is not optimal for mmWave massive MIMO where the sum capacity is expected to scale with the number of BS antennas, as it requires the number of active users in the cell to increase exponentially, which is impractical under rich scattering environments. Given a constraint on the total amount of feedback, Lee and Sung [117] and Ravindran and Jindal [133] concluded that it is preferable to obtain accurate CSI feedback from a small number of users than to get coarse channel feedback from a large number of users.

Propagation in the mmWave band is quasi-optical and sparse with very few multiple paths and large path losses [95], [129], [134]. MmWave massive MIMO networks are expected to operate in small cells in sparse user regime. Thus the demands for an exponentially-increasing number of users for RBF scheduling in mmWave massive MIMO networks is thus infeasible. Therefore, efficient user scheduling algorithms that exploit wireless channel fading, consider channel sparsity, ensure low signaling overhead and work with partial CSI acquisition are essential to guarantee efficient resource allocation and optimal system performance for mmWave massive MIMO cellular networks [117].

Similarly, Semiari *et al.* [153] and Giatsoglou *et al.* [221] have proposed new resource allocation frameworks exploiting context-awareness for scheduling in dual-mode (μWave-mmWave) BSs. Here, scheduling is done per user application, and not per user. Multiple applications running on a single user device are allocated μWave or mmWave resources based on the specific delay requirement of the individual application. This is done not only to maximize the quality of service (QoS) of the BSs but also to enhance users' quality of experience (QoE). Simulation results show that the context-aware approach outperforms classical scheduling approaches employing CSI.

C. Fronthaul Design

Starting with 5G, the typical Radio Access Network (RAN) system architecture of mobile networks will be heterogeneous, comprised of macrocells and hyper-dense small cells (microcells, metrocells, picocells, femtocells, relays and device to device cells (D2D)) [2], [30], [221] which will adopt flexible, centralized processing for efficient radio resource management (RRM), cost-effective and performance-optimizing network in an architecture that is referred to as Cloud Radio Access Network (C-RAN).

Fronthaul design represents the family of potential connections among all the BSs and from the BS to the baseband units (BBU). Connections from the BBU pool to the core network called backhaul links are generally not considered here, as we limit the discussion to the access and fronthaul links in this work. Fronthaul links could be wired or wireless or a mix of both, and the choice is determined by cost and the expected QoS. As a critical factor with substantial impact on the performance of cellular networks [38], the fronthaul must provide high-capacity, high rate and low-latency connection between the small cells and the macrocells to support the explosive data traffic expected in 5G ultra-dense cellular networks [138]–[142].

The strict capacity and delay requirements for fronthaul design have so far favored the use of wired links (optical fibers). However, such deployments are not flexible, scalable nor cost-effective (representing over 50% of the total cost of ownership (TCO)) [139], and particularly not suitable for 5G HetNets [140]. This is because only spots with existing fiber access would be used or costly fiber access would have to be deployed. This challenge, therefore, is steering research in the direction of wireless fronthaul solutions for high-speed, high-capacity, easy-to-install, low-latency, flexible and cost-effective fronthaul technologies [138], [140].

A number of wireless fronthaul solutions are available as shown in Table VIII. MmWave massive MIMO shows promising prospects for both fronthaul and access links in future 5G cellular networks [143], [144], [216] due to its promising capabilities for supporting flexible beamforming, and providing high spatial multiplexing and diversity gains. It also shows potentials for lower cost (with regards to components), higher capacity, lesser interference (due to smaller cell sizes and narrower beams) and smaller antenna form factor [140].

The suitability assessment in Table VIII which favors mmWave fronthaul over μWave is principally due to the limited available bandwidth in the μWave bands. The available bandwidth in the TVWS sub-800 MHz band is extremely limited, that of the licensed sub-6 GHz is limited and expensive while the unlicensed sub-6 GHz is highly interference-prone. All these constraints limit performance in terms of the realizable fronthaul throughput thereby pushing consideration towards the mmWave bands where abundant bandwidth is available. However, the suitability of wireless fronthaul in the 60 GHz bands is classified as requiring dynamic interference avoidance techniques due to its unlicensed nature [140].

For practical implementation, a combination of both the V and E-bands is proposed for the fronthaul of next-generation 5G HetNets based on their promising characteristics as shown

TABLE VIII
CANDIDATE WIRELESS SOLUTIONS FOR MMWAVE MASSIVE MIMO FRONTHAUL LINKS

Characteristics \Frequency	Sub-800 MHz	Sub- 6 GHz (licensed)	Sub- 6 GHz (Unlicensed)	6 - 57 GHz	57 -67 GHz	71 - 76 GHz & 81 -86 GHz
Frequency band	TVWS (UHF)	μWave	μWave	mmWave	V-band (mmWave)	E-band (mmWave)
Network Topology	Point-to-Point (PtP), Point-to-Multi-Point (PtMP)	PtP, PtMP	PtP, PtMP, mesh	PtP, PtMP	PtP	PtP
Spectrum Licensing	Dynamic and lightly-licensed	Area licensed, unlicensed or Authorized Shared Access (ASA)	Area licensed, unlicensed or ASA	Link licensed or area licensed	Unlicensed	Lightly-licensed
Propagation	NLOS	NLOS	NLOS	LOS	LOS	LOS
Pros	Long link due to good propagation	Good propagation	Good propagation	Long link distance	Unlicensed, less interference, smaller equipment form factor and better spectrum re-use	Effective interference management
Cons	Limited bandwidth	Limited, narrow, crowded and expensive bandwidth	High interference	Larger antenna array	Atmospheric attenuation and oxygen absorption	Bigger antenna size requirements
Assessment	Unsuitable for 5G HetNet	Unsuitable for 5G HetNet	Insufficient for 5G HetNet, can be used as back-up	Suitable mostly for rural small-cell deployments	Suitable for 5G HetNet. Requires dynamic interference avoidance algorithms	Suitable for 5G HetNet

in Table IX. In all, the mmWave massive MIMO fronthaul design should support mesh network topology, be self-configuring and self-organizing (plug and play), support multi-hop links, adaptive beamforming algorithms, have immunity against interference and possibly operate in full-duplex mode, which is the optimal duplexing option for mesh fronthaul links [140]. However, since full duplexing is still in the exploratory phase, particularly for mmWave bands, TDD may be a better choice than FDD for fronthauling due the ease of finding clean spectrum and avoiding interference thereby, flexibility of adjusting slot ratio since fronthaul has asymmetric traffic and the suitability of using a single device across different countries and operators [140], [143]. This will also ease inter-operator resource sharing and management for overall network improvement [224], [225].

In addition, the fronthaul should meet the following market requirements: reduced impact on the TCO (resulting from easy deployment and low installation, licensing and/or leasing fees) [145], multi-Gbps throughput support [113], low "last mile" latency (< 5 ms for a single trip to support delay-critical applications and traffic classes [114]) and satisfactorily reliable link for the intercell/inter-site distance of up to 200 m for the HetNet small cells [145], [146]. Multiplexing both fronthaul and access links on the same frequency (mmWave) band, which is termed in-band wireless fronthaul, is being proposed for 5G due to the obvious cost and frequency re-use benefits. However, this is prone to the challenges of interference between the access and fronthaul links, and self-interference between transmit and receive signals [140].

To address these sources of interference, Ghauch *et al.* [147] proposed the TDM-based in-band scheme, which in itself suffers from large system latency and large storage/buffering requirement. This has thus led to the flexible hybrid beamforming design with the concept of beam division multiplexing (BDM) which is expected to be a more

TABLE IX
COMPARISON OF E-BAND AND V-BAND FOR MMWAVE MASSIVE MIMO FRONTHAUL

Characteristics	E-Band	V-Band
Regulation Implications	Lightly licensed. Simple registration with trivial license fee; avoid interference protection scheme	Unlicensed. No registration without license fee; dynamic interference avoidance algorithms are required to mitigate the mutual interference
Distance Implications	Several kilometers due to rain attenuation	Up to 500-700m due to both rain and oxygen attenuation
Form factor implications	Can be compact and dictated by typical one feet antenna due to the requirement of minimum antenna gain. Suitable for rooftops, towers, and masts	Can be tiny and blended on the street level, building walls, light poles, bus stations and traffic lights
Applications	Fiber extension for businesses; macro fronthaul, small-cell fronthaul in some particular cases; aggregation	Security (CCTV, traffic radars), WiFi fronthaul, small cell fronthaul, FTTH - Fiber extension to customer premises

TABLE X
ABBREVIATIONS

3GPP	Third Generation Partnership Project	**FDD**	Frequency Division Duplexing	**PtMP**	Point-to-Multi-Point
ADC	Analog to Digital Converters	**FDM**	Frequency Divison Multiplexing	**PtP**	Point-to-Point
AoA	Angle of Arrival	**FD-MIMO**	Full Dimension MIMO	**QoE**	Quality of Experience
AoD	Angle of Departure	**FFT**	fast Fourier transform	**RBF**	Random Beamforming
ASA	Azimuth Spread of Arrival / Authorized Shared Access	**GAMP**	Generalized Approximate Message Passing	**RF**	Radio Frequency
ASD	Azimuth Spread of Departure	**GDFM**	Generalized FDM	**RRH**	Remote Radio Head
AWGN	Additive white Gaussian noise	**HetNet**	Heterogeneous Network	**Rx**	Receiver
BBU	Base band unit	**ICI**	Inter-cell interference	**SCM**	Spatial/Stochastic Channel Model
BD	Block Diagonization	**IFFT**	inverse FFT	**SDMA**	Space Division Multiple Access
BDM	Beam Division Multiplexing	**ITU**	International Telecommunication Union	**SED**	Sub-space Estimation and Decomposition
BER	Bit Error Ratio	**KF**	K-factor	**SF**	Shadow Fading
BFDM	Bi-orthogonal FDM	**KLT**	Karhunen-Loeve transformation	**SIC**	Successive Interference Cancellation
BS	Base station	**LOS**	Line-of-sight	**SINR**	Signal to Interference and Noise Ratio
CDM	Code Division Multiplexing	**LSAS**	Large Scale Antenna Systems	**SISO**	Single-input-single-output
CE	Channel Estimation	**LTE-A**	Long Term Evolution - Advanced	**SNR**	Signal-to-Noise Ratio
(C)-RAN	(Cloud) Radio Access Network	**MAC**	Medium Access Control	**(G)SOMA**	Generalized Semi-Orthogonal Multiple Access
CS	Compressive Sensing	**MF**	Matched Filter	**SU-MIMO**	Single-user MIMO
CSI	Channel State Information	**(MG)-OMP**	Multi-Grid Orthogonal Matching Pursuit	**SUS**	Semi-orthogonal User Selection
CSIR	CSI at the Receiver	**MIMO**	Multiple-input-multiple-output	**TCO**	Total Cost of Ownership
CSIT	CSI at the Transmitter	**ML**	Maximum Likelihood	**TDD**	Time Division Duplexing
D2D	Device to Device	**MMSE**	Minimum Mean Square Error	**TH**	Tomlinson-Harashima
DAC	Digital to Analog Converter	**mmWave**	Millimeter-wave	**TVWS**	Television white space
DCT	Discrete Cosine Transformation	**MPC**	Multi-path Component	**Tx**	Transmitter
DL	Diagonal Loading	**MU-MIMO**	Multi-user MIMO	**UDN**	Ultra-dense network
DoF	Degrees of freedom	**NLOS**	Non-line-of-sight	**UE**	User Equipment
DPC	Dirty Paper Coding	**OFDMA**	Orthogonal Frequency Division Multiplexing Access	**UFMC**	Universal Filtered Multi-Carrier
DS	Delay spread	**OMP**	Orthogonal Matching Pursuit	**UHF**	Ultra-High Frequency
DVB-2	Digital Video Broadcasting – Version 2	**OUT**	Outage	**ULA**	Uniform Linear Array
EHF	Extremely High Frequency	**PA**	Power Amplifiers	**UPA**	Uniform Planar Array
EM	Expectation Maximization	**PAPR**	Peak to Average Power Ratio	**WF**	Wiener Filter
ESA	Elevation Spread of Arrival	**PM**	Parks-McClellan	**XPR**	Cross Polarization Ratio
ESD	Elevation Spread of Departure	**PPP**	Poisson Point Process	**ZF**	Zero Forcing
FBMC	Filter Bank Multi-Carrier	**PS**	Phase Shifters	**µWave**	Microwave

competitive solution for the in-band mmWave massive MIMO fronthaul due to its lower latency [143].

D. Health and Safety Issues

With the mmWave bands being promising candidates for future broadband mobile communication networks, it is essential to understand the impacts of mmWave radiation on the human body and potential health effects related to its exposure. In addition, the current safety rules regarding RF exposure do not specify limits above 100 GHz whereas spectrum use will inevitably move to these bands over time, hence the need for further investigations to codify safety metrics at these frequencies [2].

The mmWave band constitutes RF spectrum with carrier frequencies between 30-300 GHz. The photon energy in these bands ranges from 0.1 to 1.2 milli-electron volts (meV), and unlike ultraviolet, X-ray, and gamma radiation, mmWave radiation is non-ionizing and so cannot cause cancer. Therefore, the main safety concern is heating of the eyes and skin caused by the absorption of mmWave energy in the human body and constitutes the major biological effect that can be caused by the absorption of electromagnetic mmWave energy by tissues, cells, and biological fluid [2], [212].

Based on findings from mmWave radiation studies, Wu *et al.* [212] concluded that both the eyes and the skin (whose tissues would receive the most radiation) do not appear prone to damage from exposure experienced from mmWave communication technologies in the far-field, while more studies are required regarding exposure to communication devices (such as mobile devices with high-gain, adaptive and smart antennas) in the near-field.

Similarly, according to Wu *et al.* [213], more than 90% of the transmitted electromagnetic power is absorbed within the epidermis, and dermis layers and little power penetrates further into deeper tissues. However, heating of human tissue may extend deeper than the epidermis and dermis layers. Also, the steady state temperature elevations at different body locations may vary even when the intensities of electromagnetic wave radiations are the same. The authors concluded that power

density (PD) is not likely to be as useful as Specific Absorption Rate (SAR) for assessing safety, especially in the near-field, and proposed temperature-based technique an acceptable dosimetric quantity for demonstrating safety and setting exposure limits for mmWave radiations.

E. Standardization Activities

3GPP, as a standardization body, has already decided to include two core physical layer (PHY) technologies (mmWave and massive MIMO) in the 5G NR [223]. The joint use of these two technologies for future wireless systems is what is surveyed in this paper.

The attractive capabilities and high commercial potentials of mmWave communications have spurred several international activities aimed at standardizing it for wireless personal and local area networks (WPANs and WLANs) such as the IEEE 802.15.3c, IEEE 802.11ad, WirelessHD, WiGiG and IEEE 802.11ay [14]–[16]. The focus had been on the 60 GHz band for WiFi and WiGiG standards due to the earlier notion that mmWave bands are unsuitable for mobile communications. [105]. However, with the new findings about its suitability, several research projects (such as MiWEBA [202], [203] and MiWaveS [204], [205]) are ongoing in both the academia and the industry, with the aim of incorporating mmWave communications into cellular networks [19], [20]. Standards and regulatory bodies such as ITU, IEEE, ETSI, and 3GPP are already considering and concretizing mmWave spectrum standardization for mobile systems [205], and the WRC-2018 appears as the time and place where the decision and agreement will be finalized by ITU-R [105].

Massive MIMO, on the other hand, has been under consideration for standardization by 3GPP since LTE-A Rel. 12 [105]. The release work programme, which was largely completed in March 2015, considered four areas of significant enhancements and enablers: small cells and HetNets, multi-antennas (e.g., massive MIMO and elevation beamforming), proximity services and procedures for supporting diverse traffic types [206]. Further works on massive MIMO standardization featured in subsequent 3GPP LTE releases and enhancements will continue through releases 15 and 16, expected to be finalized by September 2018 and late 2019, respectively. Standardization of both mmWave communications and massive MIMO will open up new opportunities for next-generation cellular networks.

F. Summary and Open Issues

Many factors and design considerations affect the performance of systems and influence the trade-offs to make with respect to the metrics of interest and desired applications. These cross-layer design considerations include: the choice of modulation technique and signal waveform, multiple access scheme, user scheduling algorithm, fronthaul design, antenna array architecture, precoding schemes, health and safety issues, among others.

Though we have presented many candidate cross-layer design choices in this section, it is important to note that many of these schemes are still undergoing experimental demonstrations and rigorous test and validation procedures. While optimal designs are being concluded for some techniques, same is yet to be achieved for some others. Therefore, holistic performance characterization and evaluation of these mmWave massive MIMO techniques in real-life scenarios and real-time applications remain an open issue.

VI. Conclusion and Lessons Learned

MmWave massive MIMO technology represents an attempt to harness the promising prospects realizable with the huge available bandwidth in the mmWave frequency band and the high capacity gains of massive antenna array system. In this survey, we have presented an overview of the concepts and techniques being proposed for mmWave massive MIMO systems. In doing so, we highlighted the peculiarities of the emerging network technology as compared to prior systems from which it is evolving and enumerated key research challenges and future directions.

It is instructive to note that 5G and beyond-5G (B5G) systems are not only driven by technical advances and technological capabilities of delivering massive capacity and high data rates for the explosive mobile data demands projected for next-generation cellular networks. Instead, the 2020+ experience is expected to represent a balanced ecosystem where technical abilities synchronize well with socio-economic and environmental concerns. Increased efficiency, a higher level of safety, improved health, lower cost and limited energy and CO_2 footprints are among principal drivers for the B5G era, alongside the much-anticipated boost in network capacity, user throughput, spectral and energy efficiencies.

While research, demonstrations and field tests of candidate technologies and prototypes are ongoing across different projects in the academia and the industry, holistic system deployment and performance evaluation remain open issues, together with the new applications, standardization and business models that will follow. As we march towards 2020, research on mmWave massive MIMO will continue to mature, and new trends will emerge. Though there are many challenges to address, mmWave massive MIMO shows amazing prospects and potentials in realizing the 1000-fold capacity quest for 5G cellular networks. The technology will, no doubt, usher in new paradigms for next-generation mobile networks and open up new frontiers in cellular services and applications.

References

[1] "IMT vision-framework and overall objectives of the future development of IMT for 2020 and beyond," Int. Telecommun. Union, Geneva, Switzerland, ITU-Recommendation M.2083-0, Sep. 2015.

[2] S. Mumtaz, J. Rodriguez, and L. Dai, Eds., "Introduction to mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*. London, U.K.: Academic Press, 2017, pp. 1–18.

[3] F. B. Saghezchi *et al.*, "Drivers for 5G: The 'pervasive connected world,"' in *Fundamentals of 5G Mobile Networks*, J. Rodriguez, Ed. West Sussex, U.K.: Wiley, 2015, pp. 1–27.

[4] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, "Software defined mobile networks: Concept, survey, and research directions," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 126–133, Nov. 2015.

[5] "5G vision, requirements and enabling technologies," South Korea, 5G Forum, White Paper, Mar. 2016.

[6] S. Vahid, R. Tafazolli, and M. Filo, "Small cells for 5G mobile networks," in *Fundamentals of 5G Mobile Networks*, J. Rodriguez, Ed. West Sussex, U.K.: Wiley, 2015, pp. 63–104.

[7] F. Boccardi, R. W. Health, 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.

[8] A. L. Swindlehurst, E. Ayanoglu, P. Heydasri, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[9] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[10] Z. S. Bojkovic, M. R. Bakmaz, and B. M. Bakmaz, "Research challenges for 5G cellular architecture," in *Proc. 12th Int. Conf. Telecommun. Modern Satellite Cable Broadcast. Services (TELSIKS)*, 2015, pp. 215–222.

[11] D. Feng *et al.*, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 1st Quart., 2013.

[12] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 118–127, Jun. 2014.

[13] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Mar. 2016.

[14] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.

[15] S. Hur *et al.*, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[16] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[17] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[18] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 106–112, Apr. 2014.

[19] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[20] H. Shokri-Ghadikolaei, C. Fischione, P. Popovski, and M. Zorzi, "Design aspects of short-range millimeter-wave networks: A MAC layer perspective," *IEEE Netw.*, vol. 30, no. 3, pp. 88–96, May/Jun. 2016.

[21] 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.

[22] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[23] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[24] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[25] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[26] 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.

[27] S. Mattigiri and C. Warty, "A study of fundamental limitations of small antennas: MIMO approach," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, 2013, pp. 1–8.

[28] H. Inanoğlu, "Multiple-input multiple-output system capacity: Antenna and propagation aspects," *IEEE Antennas Propag. Mag.*, vol. 55, no. 1, pp. 253–273, Feb. 2013.

[29] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[30] L. G. Ordóñez, D. P. Palomar, and J. R. Fonollosa, "Fundamental diversity, multiplexing, and array gain tradeoff under different MIMO channel models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, 2011, pp. 3252–3255.

[31] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005, pp. 321–350.

[32] Q. Li *et al.*, "MIMO techniques in WiMAX and LTE: A feature overview," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 86–92, May 2010.

[33] S. Wang, Y. Xin, S. Chen, W. Zhang, and C. Wang, "Enhancing spectral efficiency for LTE-Advanced and beyond cellular networks [guest editorial]," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 8–9, Apr. 2014.

[34] S. A. R. Naqvi, S. A. Hassan, and Z. Mulk, "Encoding and decoding in mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 63–78.

[35] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[36] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2009, pp. 2184–2188.

[37] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," in *Proc. ITG Workshop Smart Antennas*, Stuttgart, Germany, Mar. 2013, pp. 1–6.

[38] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[39] A. Ashikhmin and T. L. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1137–1141.

[40] T. E. Bogale, L. B. Le, X. Wang, and L. Vandendorpe, "Pilot contamination mitigation for wideband massive MMO: Number of cells vs multipath," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–6.

[41] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[42] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process.*, Mar. 2012, pp. 3249–3252.

[43] X. Gao, L. Dai, Z. Gao, T. Xie, and Z. Wang, "Precoding for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 79–111.

[44] J. Wang, "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.

[45] C. Cordeiro, D. Akhmetov, and M. Park, "IEEE 802.11ad: Introduction and performance evaluation of the first multi-Gbps WiFi technology," in *Proc. ACM Int. Workshop mmWave Commun.*, 2010, pp. 3–8.

[46] M. H. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[47] U. D. Wessel, "Achievable rates for Tomlinson–Harashima precoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 824–831, Mar. 1998.

[48] 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.

[49] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4066–4071.

[50] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714–2720, Jun. 2015.

[51] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[52] 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.

[53] Z. Gao, L. Dai, and Z. Wang, "Channel estimation for mmWave massive MIMO based access and backhaul in ultra-dense network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–6.

[54] Z. Gao, L. Dai, C. Hu, X. Gao, and Z. Wang, "Channel estimation for mmWave massive MIMO systems," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, pp. 113–139, 2017.

[55] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, May 2012.

[56] 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 Globecom*, 2014, pp. 3326–3331.

[57] A. Alkhateeb, G. Leus and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, 2015, pp. 2909–2913.

[58] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. IEEE ACSSC*, 2014, pp. 957–961.

[59] Z. Gao *et al.*, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[60] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[61] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design for large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, Jun. 2014.

[62] Z. Gao, L. Dai, Z. Lu, C. Yuen, and Z. Wang, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1266–1269, Jul. 2014.

[63] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna TDD systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2102–2116, Jun. 2011.

[64] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. Future Netw. Mobile Summit*, Jun. 2010, pp. 1–10.

[65] P.-H. Kuo, B. Su, and C.-P. Yen, "Channel feedback for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 141–167.

[66] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2012, pp. 492–497.

[67] Y. Li, M. S. Rahman, and Y.-H. Nam, "Full-dimension MIMO cellular systems realizing potential of massive-MIMO," *IEEE Comsoc MMTC E-Lett.*, vol. 9, no. 6, pp. 6–10, Nov. 2014.

[68] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.

[69] T. Parks and J. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. 19, no. 2, pp. 189–194, Mar. 1972.

[70] S. L. H. Nguyen, K. Haneda, J. Jarvelainen, A. Karttunen, and J. Putkonen, "On the mutual orthogonality of millimeter-wave massive MIMO channels," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.

[71] B. Ai, K. Guan, G. Li, and S. Mumtaz, "mmWave massive MIMO channel modeling," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 169–194.

[72] K. Zheng, S. Ou, and X. Yin, "Massive MIMO channel models: A survey," *Int. J. Antennas Propag.*, vol. 2014, 2014, Art. no. 848071, doi: 10.1155/2014/848071.

[73] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[74] T. Kürner and S. Priebe, "Towards THz communications—Status in research, standardization and regulation," *J. Infrared Millimeter Terahertz Waves*, vol. 35, no. 1, pp. 53–62, Jan. 2014, doi: 10.1007/s10762-013-0014-3.

[75] R. Piesiewicz *et al.*, "Scattering analysis for the modeling of THz communication systems," *IEEE Trans. Antennas Propag.*, vol. 55, no. 11, pp. 3002–3009, Nov. 2007.

[76] S. Priebe, M. Jacob, and T. Kürner, "AoA, AoD and ToA characteristics of scattered multipath clusters for THz indoor channel modeling," in *Proc. 17th Eur. Wireless Conf. (EW)*, Vienna, Austria, Apr. 2011, pp. 1–9.

[77] M. Jacob *et al.*, "Diffraction in mm and sub-mm wave indoor propagation channels," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 833–844, Mar. 2012.

[78] I. F. Akyildiz, J. M. Jornet, and C. Han, "TeraNets: Ultra-broadband communication networks in the Terahertz band," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 130–135, Aug. 2014.

[79] S. Priebe, M. Jacob, and T. Kürner, "Calibrated broadband ray tracing for the simulation of wave propagation in mm and sub-mm wave indoor communication channels," in *Proc. 18th Eur. Wireless Conf. (EW)*, Poznań, Poland, Apr. 2012, pp. 1–10.

[80] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[81] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.

[82] S. Jaeckel, M. Peter, K. Sakaguchi, W. Keusgen, and J. Medbo, "5G channel models in mm-wave frequency bands," in *Proc. 22th Eur. Wireless Conf.*, Oulu, Finland, 2016, pp. 1–6.

[83] X. Zhao *et al.*, "Channel measurements, modeling, simulation and validation at 32 GHz in outdoor microcells for 5G radio systems," *IEEE Access*, vol. 5, pp. 1062–1072, 2017.

[84] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300GHz spectrum," in *Proc. IEEE Sarnoff Symp.*, Princeton, NJ, USA, May 2011, pp. 1–6.

[85] Z. Shi, R. Lu, J. Chen, and X. S. Shen, "Three-dimensional spatial multiplexing for directional millimeter-wave communications in multi-cubicle office environments," in *Proc. IEEE Glob. Telecommun. Conf.*, Atlanta, GA, USA, 2013, pp. 4384–4389.

[86] C. Kourogiorgas, S. Sagkriotis, and A. D. Panagopoulos, "Coverage and outage capacity evaluation in 5G millimeter wave cellular systems: Impact of rain attenuation," in *Proc. Eur. Conf. Antennas Propag. (EuCAP)*, Lisbon, Portugal, 2015, pp. 1–5.

[87] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[88] Z. Qingling and J. Li, "Rain attenuation in millimeter wave ranges," in *Proc. Int. Symp. Antennas Propag. EM Theory (ISAPE)*, Guilin, China, 2006, pp. 1–4.

[89] S. Joshi and S. Sancheti, "Foliage loss measurements of tropical trees at 35 GHz," in *Proc. IEEE Int. Conf. Recent Adv. Microw. Theory Appl.*, Jaipur, India, Nov. 2008, pp. 531–532.

[90] T. E. Bogale, X. Wang, and L. B. Le. "mmWave communication enabling techniques for 5G wireless systems: A link level perspective," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. San Diego, CA, USA: Acadamic Press, 2017, pp. 195–225.

[91] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[92] T. Bai, A. Alkhateeb, and R. W. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.

[93] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proc. 16th Int. ICST Conf. Commun. Netw.*, Harbin, China, 2011, pp. 361–365.

[94] Y. Tsang, A. Poon, and S. S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Glob. Telecommun. Conf. (GLOBECOM)*, Kathmandu, Nepal, 2011, pp. 1–6.

[95] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[96] S. Han, I. Chih-Lin, 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.

[97] T. E. Bogale, L. B. Le, A. Haghighat, 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.

[98] E. Torkildson, C. Sheldon, U. Madhow, and M. Rodwell, "Millimeter-wave spatial multiplexing in an indoor environment," in *Proc. IEEE Glob. Telecommun. Conf.*, Honolulu, HI, USA, 2009, pp. 1–6.

[99] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, Jun. 2014.

[100] A. Arvanitis, G. Anagnostou, N. Moraitis, and P. Constantinou, "Capacity study of a multiple element antenna configuration in an indoor wireless channel at 60 GHz," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Dublin, Ireland, 2007, pp. 609–613.

[101] Z. Pi and F. Khan, "A millimeter-wave massive MIMO system for next generation mobile broadband," in *Proc. Asilomar Conf. Sign. Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 693–698.

[102] Robert W. Heath. (2014). *Comparing Massive MIMO and mmWave MIMO*. [Online]. Available: https://www.ieeectw.org/2014/slides/session3/Heath-CTW_v6.pdf

[103] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[104] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[105] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Area Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[106] A. Morello and V. Mignone, "DVB-S2: The second generation standard for satellite broad-band services," *Proc. IEEE*, vol. 94, no. 1, pp. 210–227, Jan. 2006.

[107] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.

[108] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[109] J.-C. Chen and C.-P. Li, "Tone reservation using near-optimal peak reduction tone set selection algorithm for PAPR reduction in OFDM systems," *IEEE Signal Process. Lett.*, vol. 17, no. 11, pp. 933–936, Nov. 2010.

[110] N. Benvenuto, R. Dinis, D. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come again," *Proc. IEEE*, vol. 98, no. 1, pp. 69–96, Jan. 2010.

[111] K. Zheng *et al.*, "10 Gb/s HetSNets with millimeter-wave communications: Access and networking-challenges and protocols," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 222–231, Jan. 2015.

[112] H. A. Ul Mustafa, M. A. Imran, M. Z. Shakir, A. Imran, and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 419–445, 1st Quart., 2016.

[113] A. Dehos, J. L. Gonzalez, A. De Domenico, D. Kténas, and L. Dussopt, "Millimeter-wave access and backhauling: The solution to the exponential data traffic increase in 5G mobile communications systems?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 88–95, Sep. 2014.

[114] D. R. Taori and A. Sridharan, "Point-to-multipoint in-band mmWave fronthaul for 5G networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 195–201, Jan. 2015.

[115] Y. Niu *et al.*, "Exploiting device-to-device communications in joint scheduling of access and backhaul for mmWave small cells," *IEEE J. Sel. Area Commun.*, vol. 33, no. 10, pp. 2052–2069, Oct. 2015.

[116] C. J. Bernardos, A. De Domenico, J. Ortin, P. Rost, and D. Wübben, "Challenges of designing jointly the backhaul and radio access network in a cloud-based mobile network," in *Proc. Future Netw. Mobile Summit.*, Lisbon, Portugal, Jul. 2013, pp. 1–10.

[117] G. Lee and Y. Sung, "MAC layer design for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds., London, U.K.: Academic Press, 2017, pp. 227–255.

[118] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, 2nd Quart., 2013.

[119] R. Knopp and P. A. Humblet. "Information capacity and power control in single-cell multi-user communication," in *Proc. ICC*, vol. 1. Seattle, WA, USA, 1995, pp. 331–335.

[120] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[121] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[122] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006

[123] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[124] G. Lee and Y. Sung. (2014). *A New Approach to User Scheduling in Massive Multi-User MIMO Broadcast Channels*. [Online]. Available: http://arxiv.org/pdf/1403.6931.pdf

[125] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[126] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, 2014, pp. 76–80.

[127] A. Tomasoni, G. Caire, M. Ferrari, and S. Bellini, "On the selection of semi-orthogonal users for zero-forcing beamforming," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, 2009, pp. 1100–1104.

[128] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1086–1100, Feb. 2016.

[129] G. Lee, Y. Sung, and M. Kountouris, "On the performance of random beamforming in sparse millimeter wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 560–575, Apr. 2016.

[130] G. Lee, Y. Sung, and M. Kountouris, "On the performance of randomly directional beamforming between line-of-sight and rich scattering channels," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, 2015, pp. 141–145.

[131] G. Lee, Y. Sung, and J. Seo, "How many users are needed for non-trivial performance of random beamforming in highly-directional mm-wave MIMO downlink?" in *Proc. IEEE Inf. Theory Workshop Fall (ITW)*, Jeju-do, South Korea, 2015, pp. 244–248.

[132] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[133] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3037–3046, Sep. 2012.

[134] S. Sun and T. S. Rappaport, "Wideband mmWave channels: Implications for design and implementation of adaptive beam antennas," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Tampa, FL, USA, 2014, pp. 1–4.

[135] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[136] E. Björnson, E. G. Larsson, and M. Debbah, "Optimizing multi-cell massive MIMO for spectral efficiency: How many users should be scheduled?" in *Proc. IEEE Glob. Conf. Signal Inf. Process. (GlobalSIP)*, Atlanta, GA, USA, 2014, pp. 612–616.

[137] M. N. Khormuji, "Enhanced multiple-access for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. San Diego, CA, USA: Academic Press, pp. 257–288, 2017.

[138] P. Rost *et al.*, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[139] N. Wang, E. Hossain, and V. K. Bhargava, "Backhauling 5G small cells: A radio resource management perspective," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 41–49, Oct. 2015.

[140] Z. Gao, L. Dai, X. Gao, M. Z. Shakir, and Z. Wang, "Fronthaul design for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. San Diego, CA, USA: Academic Press, 2017, pp. 289–312.

[141] S. Park, C.-B. Chae, and S. Bahk, "Large-scale antenna operation in heterogeneous cloud radio access networks: A partial centralization approach," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 32–40, Jun. 2015.

[142] Z. Zhang, X. Wang, K. Long, A. V. Vasilakos, and L. Hanzo, "Large-scale MIMO-based wireless backhaul in 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 58–66, Oct. 2015.

[143] Z. Gao *et al.*, "MmWave massive MIMO based wireless fronthaul for 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.

[144] R. J. Weiler *et al.*, "Enabling 5G fronthaul and access with millimeter-waves," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Bologna, Italy, Jun. 2014, pp. 1–5.

[145] *Small Cell Millimeter Wave Mesh Fronthaul*, InterDigital. Inc., Wilmington, DE, USA, Feb. 2013. [Online]. Available: http://www.interdigital.com/research_papers/2013_01_25_small_cell_millimeter_wave_mesh_fronthaul

[146] *Small Cell Millimeter Wave Mesh Fronthaul*. [Online]. Available: http://www.interdigital.com/research_papers/2013_01_25_small_cell_millimeter_wave_mesh_fronthaul

[147] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[148] W. Lu and M. Di Renzo, "mmWave cellular networks: Stochastic geometry modeling, analysis, and experimental validation," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. San Diego, CA, USA: Academic Press, 2017, pp. 313–341.

[149] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[150] S. A. Hosseini, E. Zarepour, M. Hassan, and C. T. Chou, "Analyzing diurnal variations of millimeter wave channels," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, San Francisco, CA, USA, 2016, pp. 377–382.

[151] M. Di Renzo, A. Guidotti, and G. E. Corazza, "Average rate of downlink heterogeneous cellular networks over generalized fading channels—A stochastic geometry approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 3050–3071, Jul. 2013.

[152] M. Di Renzo and W. Lu, "The equivalent-in-distribution (EiD)-based approach: On the analysis of cellular networks using stochastic geometry," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 761–764, May 2014.

[153] O. Semiari, W. Saad, and M. Bennis, "Context-aware scheduling of joint millimeter wave and microwave resources for dual-mode base stations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–6.

[154] 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.

[155] M. Di Renzo and W. Lu, "End-to-end error probability and diversity analysis of AF-based dual-hop cooperative relaying in a poisson field of interferers at the destination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 15–32, Jan. 2015.

[156] M. Di Renzo and W. Lu, "Stochastic geometry modeling and performance evaluation of MIMO cellular networks using the equivalent-in-distribution (EiD)-based approach," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 977–996, Mar. 2015.

[157] M. Di Renzo and W. Lu, "On the diversity order of selection combining dual-branch dual-hop AF relaying in a poisson field of interferers at the destination," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1620–1628, Apr. 2015.

[158] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.

[159] W. Lu and M. Di Renzo, "Stochastic geometry modeling and system-level analysis & optimization of relay-aided downlink cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4063–4085, Nov. 2015.

[160] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sep. 2014.

[161] T. Bai and R. W. Heath, Jr., "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[162] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[163] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300GHz spectrum," in *Proc. IEEE Sarnoff Symp.*, Princeton, NJ, USA, Mar. 2011, pp. 1–6. [Online]. Available: http://wcnc2011.ieee-wcnc.org/tut/t1.pdf

[164] D. Zhang, S. Mumtaz, and K. S. Huq, "SISO to mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 19–38.

[165] F. Ademaj, M. Taranetz, and M. Rupp, "Implementation, validation and application of the 3GPP 3D MIMO channel model in open source simulation tools," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, 2015, pp. 721–725.

[166] F. Ademaj, M. Taranetz, and M. Rupp, "3GPP 3D MIMO channel model: A holistic implementation guideline for open source simulation tools," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 55, pp. 1–14, Feb. 2016, doi: 10.1186/s13638-016-0549-9.

[167] N. A. Muhammad, P. Wang, Y. Li, and B. Vucetic, "Analytical model for outdoor millimeter wave channels using geometry-based stochastic approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 912–926, Feb. 2017.

[168] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.

[169] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, Nanjing, China, 2016, pp. 1–6.

[170] M. K. Samimi and T. S. Rappaport, "Local multipath model parameters for generating 5G millimeter-wave 3GPP-like channel impulse response," in *Proc. 10th Eur. Conf. Antennas Propag. (EuCAP)*, Davos, Switzerland, 2016, pp. 1–5.

[171] M. K. Samimi, S. Sun, and T. S. Rappaport, "MIMO channel modeling and capacity analysis for 5G millimeter-wave wireless systems," in *Proc. 10th Eur. Conf. Antennas Propag. (EuCAP)*, Davos, Switzerland, 2016, pp. 1–5.

[172] M. K. Samimi and T. S. Rappaport, "Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, San Diego, CA, USA, 2015, pp. 1–7.

[173] A. Torabi, S. A. Zekavat, and A. Al-Rasheed, "Millimeter wave directional channel modeling," in *Proc. IEEE Int. Conf. Wireless Space Extreme Environ. (WiSEE)*, Orlando, FL, USA, 2015, pp. 1–6.

[174] S. Hur *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 454–469, Apr. 2016.

[175] S. Buzzi and C. D'Andrea, "Doubly massive mmWave MIMO systems: Using very large antenna arrays at both transmitter and receiver," in *Proc. IEEE GLOBECOM*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[176] S. Buzzi and C. D'Andrea, "The doubly massive MIMO regime in mmWave communications," in *Proc. Tyrrhenian Int. Workshop Digit. Commun.*, Sep. 2016, pp. 1–28. [Online]. Available: http://tyrr2016.cnit.it/tyrr-content/uploads/The-doubly-massive-MIMO-regime-in-mmWave_DAndrea_TIW16.pdf

[177] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Hybrid antenna array for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. London, U.K.: Academic Press, 2017, pp. 39–61.

[178] X. Huang, Y. J. Guo, and J. D. Bunton, "A hybrid adaptive antenna array," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1770–1779, May 2010.

[179] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.

[180] Y. J. Guo, X. Huang, and V. Dyadyuk, "A hybrid adaptive antenna array for long-range mm-Wave communications [antenna applications corner]," *IEEE Antennas Propag. Mag.*, vol. 54, no. 2, pp. 271–282, Apr. 2012.

[181] V. Dyadyuk *et al.*, "Adaptive antenna arrays for adhoc millimeter-wave wireless communications," in *Advanced Trends in Wireless Communications*. Rijeka, Croatia: InTech, 2011.

[182] W. Roh *et al.*, "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.

[183] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[184] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "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.

[185] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[186] N. Van der Neut *et al.*, "PAPR reduction in FBMC systems using a smart gradient-project active constellation extension method," in *Proc. 21st Int. Conf. Telecommun. (ICT)*, Lisbon, Portugal, 2014, pp. 134–139.

[187] M. Mukherjee, L. Shu, V. Kumar, P. Kumar, and R. Matam, "Reduced out-of-band radiation-based filter optimization for UFMC systems in 5G," in *Proc. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Dubrovnik, Croatia, 2015, pp. 1150–1155.

[188] G. Fettweis, M. Krondorf, and S. Bittner, "GFDM—Generalized frequency division multiplexing," in *Proc. IEEE 69th Veh. Technol. Conf.*, Barcelona, Spain, 2009, pp. 1–4.

[189] G. Wunder, S. A. Gorgani, and S. S. Ahmed, "Waveform optimization using trapezoidal pulses for 5G random access with short message support," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun.*, Stockholm, Sweden, 2015, pp. 76–80.

[190] R. N. Mitra and D. P. Agrawal, "5G mobile technology: A survey," *ICT Exp.*, vol. 1, no. 3, pp. 132–137, Dec. 2015.

[191] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[192] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec. 2004.

[193] Y. Li, Y.-H. Nam, B. L. Ng, and J. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, 2012, pp. 4500–4504.

[194] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, 2010, pp. 1–5.

[195] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[196] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[197] L. Cottatellucci, R. R. Muller, and M. Vehkapera, "Analysis of pilot decontamination based on power control," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Dresden, Germany, 2013, pp. 1–5.

[198] M. Filippou, D. Gesbert, and H. Yin, "Decontaminating pilots in cognitive massive MIMO networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, 2012, pp. 816–820.

[199] H. Yin, D. Gesbert, M. C. Filippou, and Y. Liu, "Decontaminating pilots in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, 2013, pp. 3170–3175.

[200] "METIS channel models deliverable 1.4 version 3," document ICT-317669 METIS/D14, METIS, Sweden, Jul. 2015.

[201] A. Ghosh, Ed., *5G Channel Model for Bands up to 100 GHz*, Globecom, San Diego, CA, USA, Dec. 2015, accessed: Jan. 6, 2018. [Online]. Available: http://www.5gworkshops.com/5GCM.html

[202] *European 7th Framework Programme Project MiWEBA*. accessed: Jan. 6, 2018. [Online]. Available: http://www.miweba.eu

[203] K. Sakaguchi *et al.*, "Millimeter-wave evolution for 5G cellular networks," *IEICE Trans. Commun.*, vol. E98.B, no. 3, pp. 388–402, Mar. 2015.

[204] *European 7th Framework Programme Project MiWaveS*. accessed: Jan. 6, 2018. [Online]. Available: http://www.miwaves.eu

[205] V. Frascolla *et al.*, "MmWave use cases and prototyping: A way towards 5G standardization," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Paris, France, 2015, pp. 128–132.

[206] *Understanding 3GPP Release 12: Standards for HSPA+ and LTE Enhancements*, 4G Americas, Bellevue, WA, USA, Feb. 2015. [Online]. Available: http://www.5gamericas.org/files/6614/2359/0457/4G_Americas_-_3GPP_Release_12_Executive_Summary_-_February_2015.pdf

[207] "More than 50 billion connected devices," Stockholm, Sweden, Ericsson, White Paper, Feb. 2011.

[208] H.-K. Lee, D. M. Kim, Y. Hwang, S. M. Yu, and S.-L. Kim, "Feasibility of cognitive machine-to-machine communication using cellular bands," *IEEE Wireless Commun. Mag.*, vol. 20, no. 2, pp. 97–103, Apr. 2013.

[209] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[210] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[211] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "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.

[212] T. Wu, T. S. Rappaport, and C. M. Collins, "Safe for generations to come: Considerations of safety for millimeter waves in wireless communications," *IEEE Microw. Mag.*, vol. 16, no. 2, pp. 65–84, Mar. 2015.

[213] T. Wu, T. S. Rappaport, and C. M. Collins, "The human body and millimeter-wave wireless communication systems: Interactions and implications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., 2015, pp. 2423–2429.

[214] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. New York, NY, USA: Cambridge Univ. Press, 2009.

[215] 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, 2015.

[216] G. Kalfas, J. Vardakas, L. Alonso, C. Verikoukis, and N. Pleros, "Medium transparent MAC access schemes for seamless packetized fronthaul in mm-wave 5G picocellular networks," in *Proc. 19th Int. Conf. Transp. Opt. Netw. (ICTON)*, Girona, Spain, 2017, pp. 1–4.

[217] "Study on 3D channel model for LTE, v12.2.0," 3rd Gener. Partnership Project, Sophia Antipolis, France, Rep. 3GPP TR36.873, 2014.

[218] A. Mesodiakaki, F. Adelantado, L. Alonso, M. Di Renzo, and C. Verikoukis, "Energy- and spectrum-efficient user association in millimeter-wave backhaul small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1810–1821, Feb. 2017.

[219] A. Mesodiakaki *et al.*, "Energy impact of outdoor small cell backhaul in green heterogeneous networks," in *Proc. IEEE 19th Int. Workshop Comput. Aided Model. Design Commun. Links Netw. (CAMAD)*, Athens, Greece, 2014, pp. 11–15.

[220] "Study on channel model for frequency spectrum above 6 GHz, v14.2.0," 3rd Gener. Partnership Project, Sophia Antipolis, France, Rep. 3GPP TR38.900, 2016.

[221] N. Giatsoglou, K. Ntontin, E. Kartsakli, A. Antonopoulos, and C. Verikoukis, "D2D-aware device caching in mmWave-cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2025–2037, Sep. 2017.

[222] O. Semiari, W. Saad, and M. Bennis, "Joint millimeter wave and microwave resources allocation in cellular networks with dual-mode base stations," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4802–4816, Jul. 2017.

[223] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[224] M. Rebato, M. Mezzavilla, S. Rangan, and M. Zorzi, "Resource sharing in 5G mmWave cellular networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, San Francisco, CA, USA, 2016, pp. 271–276.

[225] O. Semiari, W. Saad, M. Bennis, and Z. Dawy, "Inter-operator resource management for millimeter wave multi-hop backhaul networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5258–5272, Aug. 2017.

[226] S. A. Busari, S. Mumtaz, K. M. S. Huq, J. Rodriguez, and H. Gacanin, "System-level performance evaluation for 5G mmWave cellular network," in *Proc. IEEE Glob. Telecommun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–7.

Sherif Adeshina Busari (S'14) received the B.Eng. and M.Eng. degrees in electrical and electronics engineering (communications option) from the Federal University of Technology Akure, Nigeria, in 2011 and 2015, respectively. He is currently pursuing the Ph.D. degree on the MAP-Tele doctoral program in Telecommunications (a joint program of the University of Minho, the University of Aveiro, and the University of Porto). He is currently a Researcher with the Instituto de Telecomunicações, Aveiro, Portugal. He is a Lecturer with the Federal University of Technology, Akure, Nigeria. His research interests include wireless channel modeling, millimeter wave communications, massive MIMO, radio resource management, and optimization for 5G and beyond-5G mobile networks.

Kazi Mohammed Saidul Huq (SM'17) received the B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology, Bangladesh, in 2003, the M.Sc. degree in electrical engineering from the Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree in electrical engineering from the University of Aveiro, Portugal, in 2014. He is a Senior Research Engineer with the Mobile Systems Group, Instituto de Telecomunicações, Aveiro. He is one of the founding members of the IEEE 1932.1 Standard. He has authored several publications, including papers in high-impact international conferences and journals, a book, and book chapters. His research activities include 5G paradigm, communication haul, D2D communication, energy-efficient wireless communication, higher frequency communication (mmWave, THz), radio resource management, and MAC layer scheduling.

Shahid Mumtaz (SM'16) received the M.Sc. degree in electrical and electronic engineering from the Blekinge Institute of Technology, Karlskrona, Sweden, and the Ph.D. degree in electrical and electronic engineering from the University of Aveiro. He was a Research Intern with Ericsson and Huawei Research Laboratories. He has over seven years of wireless industry experience and is currently a Senior Research Scientist and a Technical Manager with the Instituto de Telecomunicações Aveiro, 4TELL Group, Portugal. He has authored over 120 publications in international conferences, journals, and book chapters, as well as 3 book editorials. His research interests lie in the field of architectural enhancements to 3GPP networks (i.e., LTE-A user plane and control plane protocol stack, NAS, and EPC), 5G related technologies, green communications, cognitive radio, cooperative networking, radio resource management, cross-layer design, backhaul/fronthaul, heterogeneous networks, M2M and D2D communication, and baseband digital signal processing.

Linglong Dai (SM'14) received the B.S. degree from Zhejiang University in 2003, the M.S. degree (with Highest Hons.) from the China Academy of Telecommunications Technology in 2006, and the Ph.D. degree (with Highest Hons.) from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. He co-authored a book entitled *mmWave Massive MIMO: A Paradigm for 5G* (Academic Press, Elsevier, 2016). He has published over 50 IEEE journal papers and over 40 IEEE conference papers. He also holds 13 granted patents. His current research interests include massive MIMO, millimeter-wave communications, NOMA, sparse signal processing, and machine learning. He was a recipient of four conference Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, and the IEEE VTC 2017-Fall, the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the Second Prize of Science and Technology Award of China Institute of Communications in 2016, the IEEE COMMUNICATIONS LETTERS Exemplary Editor Award in 2017, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, and the IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award in 2017. He currently serves as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE COMMUNICATIONS LETTERS.

Jonathan Rodriguez (SM'13) received the master's degree in electronic and electrical engineering and the Ph.D. degree from the University of Surrey, U.K., in 1998 and 2004, respectively. In 2005, he became a Researcher with the Instituto de Telecomunicações, Portugal, and a Senior Researcher in 2008, where he established the 4TELL Research Group targeting next generation Mobile Systems with key interests on 5G, security, and antenna design. He has served as a Project Coordinator for major international research projects, which includes Eureka LOOP and FP7 C2POWER, whilst serving as a Technical Manager for FP7 COGEU and FP7 SALUS. He is currently leading the H2020-ETN SECRET project, a European Training Network on 5G communications. In 2015, he became an Invited Associate Professor with the University of Aveiro, Portugal, and the Honorary Visiting Researcher with the University of Bradford, U.K. Since 2017, he has been a Professor of mobile communications with the University of South Wales, U.K. He has authored over 400 scientific works, that includes nine book editorials. He has been a Chartered Engineer since 2013, and a fellow of the IET in 2015.