IEICE TRANS. COMMUN., VOL.E99–B, NO.3 MARCH 2016

INVITED SURVEY PAPER

Massive MIMO Technologies and Challenges towards 5G

Haralabos PAPADOPOULOS[†], **Chenwei WANG**[†], **Ozgun BURSALIOGLU**[†], **Xiaolin HOU**[††], ***Nonmembers***, ***and*** **Yoshihisa KISHIYAMA**[†††a)], ***Member***

SUMMARY Massive MIMO is widely recognized as an essential technology for 5G. Together with newly allocated spectrum (bandwidth) and network densification (small cells), it is expected to play a key role in coping with the ongoing explosion in data-traffic demand and services. Compared to 4G MIMO technologies, massive MIMO can offer large gains in cell spectral efficiency, which, in combination with small cells and additional bandwidth, can translate into vast gains in throughput per unit area. We briefly overview the most promising TDD and FDD operation modes for massive MIMO, and discuss their potential benefits and challenges considering operation over different tiers and frequency bands. TDD operation is naturally suited to massive MIMO and can offer "massive MIMO" gains, with simple in-cell processing, low overheads and low end-to-end latencies. We also briefly describe some important massive MIMO activities towards 5G, including standardization efforts, system development and experimental trials.

key words: *5G, massive MIMO, TDD, FDD*

1. Introduction

The past few decades we have witnessed an explosion in wireless devices and communication modes and services. All these have been enabled by ever evolving wireless networks enabled by novel wireless technologies in combination with technology standardization efforts. MIMO (Multiple-Input-Multiple-Output) has been in the forefront of enabling wireless system technologies, both in theory and practice. MIMO requires equipping base stations (BS), and possibly, user terminals (UTs) with multiple antennas or antenna elements. A broad range of MIMO techniques have been developed over the years that can provide substantial benefits with respect to their SISO (Single-Input-Single-Output) counterparts by exploiting transmission and/or reception from multiple antennas or antenna elements. At one end, these include diversity benefits, resulting in improved communication link reliability. At the other end, they include multiplexing gains by simultaneously allowing transmission of multiple user streams over the same transmission resources, thereby increasing the cell spectral efficiency. As a result, MIMO technologies have been thoroughly studied and have found application in 3G and 4G systems. For instance, LTE-Advanced allows spatial multiplexing up to eight layers for both FDD (frequency-division duplexing) and TDD (time-division duplexing).

The prevalence of smart devices and laptops, and the proliferation of social networking, on-line gaming and multimedia on-demand services have led to an explosion in data-traffic in cellular networks, with demands expected to keep on increasing exponentially. Indeed, as pointed out by Cisco's visual networking index forecast, the global mobile data traffic is expected to increase approximately 10-fold from 2013 to 2018 [3]. Compared to 4G, the 5G system would have to deliver substantially higher user peak rates and system spectral efficiencies per unit area, with significantly lower end-to-end latencies, enabling at the same time massive device connectivity, energy savings and cost reductions.

Massive MIMO is widely viewed as one of the foundational 5G technologies and is expected to play a key role in achieving the 5G targets. Indeed, it is widely acknowledged that the 5G goals will be met by a combination of network densification (small cells), additional bandwidth at lower and/or higher frequency bands, and spectral efficiency increases that are to be realized by the use of massive MIMO.

Massive MIMO, originally introduced in the seminal work of Marzetta [4], [45], also known as "Large-Scale Antenna Systems", "Full-Dimension MIMO", "Very Large MIMO", makes a clean break with current practice through the use of a large number of antennas at the BSs. Initially the term massive MIMO was synonymous with a regime where the number of antennas grows to infinity. Strictly speaking, there is no clear-cut breakpoint between the conventional and the massive MIMO regime. In general, however, the massive MIMO regime is usually associated with settings where the number of antennas or antenna elements at the BS is several times more than the number of simultaneously served streams [5]–[7].

Compared to conventional MIMO, the excess antennas offered by massive MIMO help focus energy into smaller regions, incurring thereby low interference levels, so that all UTs can harvest large improvements in throughput and radiated energy efficiency. Other benefits of massive MIMO include cell range extension, especially in higher frequency bands, higher affinity for Multi-user MIMO (MU-MIMO), near-optimal performance with simple transceivers, analysis simplification of UT scheduling owing to channel hard-

Manuscript received October 7, 2015.
Manuscript revised December 14, 2015.

[†]The authors are with DOCOMO Innovations, Inc., Palo Alto, CA 94304, USA.
[††]The author is with DOCOMO Beijing Labs, Beijing, 100190, China.
[†††]The author is with NTT DOCOMO, Inc., Yokosuka-shi, 239-0847 Japan.
a) E-mail: kishiyama@nttdocomo.com

DOI: 10.1587/transcom.2015EBI0002

ening and robustness to channel small-scale fading [13], and, equally important, substantial gains in edge-user performance [6].

Furthermore, massive MIMO becomes especially attractive at higher frequency bands, including millimeter wave (mmWave) bands. Indeed, massive MIMO can be viewed as an enabling technology at higher frequency bands, as large numbers of operational antennas can be readily in packed with a small footprint (even at UTs). These large arrays can allow higher power directivity and can compensate for the harsh propagation conditions at these bands.

Similar to conventional MIMO, and even more so with massive MIMO, one of the most important challenges is low-overhead acquisition of channel state information (CSI) at the BS regarding the channels between the massive BS antenna array and the UTs it serves. Indeed with massive MIMO, the CSI acquisition problem is accentuated by the need to learn a massive number of UT-BS antenna-pair channels. The overheads of the traditional FDD-based CSI acquisition approach employed by LTE systems (relying on the use of downlink reference signaling and subsequent CSI feedback) can become overwhelming in the massive MIMO regime. It was recognized by Marzetta [4] that TDD is the more natural mode of operation in the massive MIMO regime, as a single uplink pilot per UT allows learning the channel between the massive array and the user, whether for uplink or downlink data transmission.

In this paper, we present an overview of massive MIMO technology including cutting-edge research and development in both academia and industry. This survey paper is organized as follows. We first introduce the fundamental principles and properties of massive MIMO in Sect. 2. In Sect. 3, we provide an overview of the state-of-the art of massive MIMO in industry including IEEE and 3GPP standards. In Sect. 4, we briefly describe some important ongoing activities towards 5G massive MIMO technologies, and, in particular, focus on the undergoing projects led by NTT DOCOMO. Finally, we conclude this paper in Sect. 5.

2. Overview of Massive MIMO

BSs with large antenna arrays are expected to play a key role in coping with the exploding traffic in future wireless networks. Indeed, they promise large-spectral efficiency increases, and when used in combination with small cells and new frequency bands, they can provide large throughput gains per unit area required to meet the exploding traffic demands.

MIMO technologies exploit the presence of multiple antennas at each BS and/or UTs to provide significant spectral efficiency gains. In the context of communicating a single stream, MIMO enables beam-forming gains via appropriate spatial filtering at the transmitter (precoding) and/or the receiver. These beamforming gains can be increased with more antenna elements and result in increased effective receiver signal-to-interference plus noise ratio (SINR), and in turn higher spectral efficiencies. More importantly, MIMO can provide multiplexing gains. Considering, for example, the downlink (DL) of a cellular system, a BS with multiple antennas can send multiple streams simultaneously to multiple (single or multi-antenna) UTs via MU-MIMO, in such a way that each UT "sees" and decodes only its desired streams.

Massive MIMO takes these attributes to the extreme. Communication of many high-rate streams (many more than in conventional systems) can be achieved by leveraging large-scale antenna arrays at the BS, whereby the number of antenna elements at the BS is potentially much larger than the total number of antennas at the UTs. Indeed, as described henceforth in this survey, inserting additional antennas at each site does not incur any additional overheads with the preferred CSI acquisition modes. As a result, the use of large antenna arrays can result in very large beamforming gains at each receiver for the desired stream while at the same time it effectively suppresses interference from all other transmitted streams.

In addition to the significant performance improvements that massive MIMO promises with respect to conventional MIMO, there are also a number of additional properties that are unique to massive MIMO. Some of these properties that massive MIMO offer are highlighted in the document and include the following:

- *Large spectral efficiencies*: Massive MIMO can simultaneously create multiple very sharp beams. As a result, many high-rate streams can be multiplexed over the same transmission resources.
- *Near-optimal performance with simple transceivers*: Massive arrays allow achieving good performance even with a rudimentary precoding technology such as conjugate beamforming, which ignores multiuser interference in designing user beams.
- *Much greener technology*: The massive array allows the creation of sharp beams, directing the radiated power to the desired UT. As a result, the radiated power per bit delivered can be significantly lowered.
- *Simpler hybrid-ARQ and scheduling*: despite the fact that each user's channel changes according to its channel's coherence time and bandwidth, the rates that can be delivered to the user can be accurately predicted a priori, without knowing the channel small scale fading.
- *Substantially improved edge-user performance*: while the desired signal at the cell-edge receiver harvests a "massive MIMO" beamforming gain, the interfering signals coming from the nearby BSs do not have a beamforming gain. As a result, massive MIMO can offer high SINRs to cell-edge users, in stark contrast to conventional MIMO systems, where cell-edge users experience low SINR.
- *Flexible off-loading and load balancing*: Due to the fact that the large beamforming gain is only experienced by the desired signal, the network can serve an edge-user at sufficiently high SINR from *any* nearby BS. This allows additional flexibility in balancing the

traffic load across the network and can provide large improvements in user experience.

As pointed out in the Introduction, CSI is needed at the BS transmitter (CSIT) regarding the channels between the UTs and the massive array at the BS, so as to enable the BS to create sharp beams in the direction of UTs, and achieve beamforming and multiplexing gains. The CSI acquisition mechanism is an essential piece of massive MIMO as the CSI-acquisition overhead/CSI quality trade-offs directly impact the effective massive MIMO performance.

Massive MIMO was originally presented as a macro-cell technology by training in the UL, whether for UL or DL transmission. Indeed, as advocated in [4], operating a massive MIMO macro-cell deployment with reuse-7 and rudimentary precoders and schedulers can in principle yield a 10-fold improvement in system performance with respect to the (at the time) state-of-the-art LTE, even in the presence of mobility. DL operation based on UL training leverages radio channel reciprocity, i.e., the fact that the UL and DL radio channels between any pair of antenna elements are effectively the same, when the two channels are measured within the coherence time and coherence bandwidth of the channel. Consequently, UL training based massive MIMO is naturally suited for TDD systems. Training in the UL, whether for UL or DL data transmission, has been well recognized as the preferred training-overhead channel-acquisition method, as a single pilot from a UT antenna allows training an unlimited number of nearby infrastructure antennas. Many works since [4] have exploited this operation to spatially multiplex large numbers of users on the same time and frequency resources and provide very large aggregate spectral-efficiencies with low overheads [6]–[9], [12], [14], [29], [30]. Section 2.1 presents an overview of massive MIMO schemes with UL-pilot based channel acquisition.

Massive MIMO has also been considered in the context of channel acquisition systems of the form deployed in FDD-based LTE, which rely on DL training and UL feedback. Section 2.2 presents an overview of DL MIMO schemes based on DL training and UL feedback that are specifically tuned to massive MIMO. Other aspects, including operation over OFDM and practical implementation considerations are briefly considered in Sect. 2.3.

Since the introduction of massive MIMO in [4], many works have focused on deployments involving massive MIMO macro cells. This is partly due to the form factor required for deploying massive arrays at frequencies in the order of 2 GHz. Another factor was that massive MIMO macro cells were originally envisioned as a competitive cost-effective alternative to network densification and small cells [4], [6], [7]. When used in conjunction with small cells, however, massive MIMO promises to deliver orders of magnitude higher aggregate spectral efficiencies per unit area. As bands at higher carrier frequencies are expected to become available, it is clear that more and more antenna elements can be packaged into a small area, making it possible to have small cells with large antenna-arrays. A brief overview of some of the challenges and benefits of massive MIMO small cells are considered in Sect. 2.4.

2.1 Uplink Pilot Based Massive MIMO

In this section we present an overview of massive MIMO with uplink pilot based channel acquisition. One key operational assumption is that the channels are learned by UL pilots whether for UL or DL transmission [4]. A single UL pilot transmitted by a UT trains all nearby BS antennas, regardless the number of nearby BSs and antennas. This allows letting the number of BS antennas $M \to \infty$ *without* increasing overheads. As shown in [4] and subsequent works, significant cell and cell-edge throughput benefits can be harvested in both the UL and the DL with very simple operation.

Although the focus is on the DL of cellular networks, we also briefly describe selected counterparts for the UL. We make an effort to restate the massive MIMO operational aspects and rate analysis of key existing works using as much as possible a common notation. As a result, the operation and achievable-rate formulas presented here can be obtained, albeit at the cost of some effort, by particularizing the results found in several papers (in particular, see [4]–[7]).

We consider a cellular network with $\mathcal{J}$ BSs serving a user population $\mathcal{K}$. We let M_j denote the number of antenna elements at BS j. Without loss of generality, we consider single-antenna UTs for all the users. It is convenient to first express the basic operation and main results in terms of a slotted system, whereby the transmission resources are partitioned into slots, or resource blocks, each with N channel uses[†]. We consider a fixed but arbitrary such slot, and let $\mathcal{K}_j$ denote the set of active users in cell j, that is, the set of users scheduled for data transmission in the given slot by BS j. At first we assume a block-fading model where the channel between users and BSs remain constant over time-frequency coherence slots of N channel uses, and let $\mathbf{h}_{k,j}$ denote the channel between the single antenna UT of active user k (in the network) and the M_j dimensional array at BS j. The channel vector $\mathbf{h}_{k,j}$ captures both large and small scale fading. In this treatment we present an overview of the key techniques, many of which are based on different channel models for $\mathbf{h}_{k,j}$. In the context of downlink transmission, these techniques rely on radio channel reciprocity, according to which the DL and UL channels between any two pairs of antennas are the same. Consequently, $\mathbf{h}_{k,j}$ represents both the UL and the DL channels between UT k and the M_j dimensional array at BS j.

2.1.1 Data Transmission Phase

One use of the (complex discrete-time baseband) channel

[†]For instance, in the context of LTE language, a resource block corresponds to a slot, and the time-frequency elements within a resource block are N channel uses within a slot.

observed at the k-th user receive antenna during the downlink data transmission phase can be represented as

$$y_k = \sum_{j \in \mathcal{J}} \mathbf{h}_{k,j}^{\mathsf{T}} \mathbf{x}_j + z_k, \tag{1}$$

where $\mathbf{x}_j \in \mathbb{C}^{M_j}$ is the transmit signal vector of BS j, and $z_k \sim \mathcal{CN}(0, N_0)$ is the additive Gaussian noise sample at receiver k. The transmitted signals are constrained by $\mathbb{E}\left[\|\mathbf{x}_j\|^2\right] \leqslant P_j$, where P_j denotes the transmit power of BS j. BS j sends $S_j = |\mathcal{K}_j|$ downlink data streams, one per active user (i.e., one stream for each $k \in \mathcal{K}_j$. With linear MU-MIMO precoding, BS j forms its transmitted signal as

$$\mathbf{x}_j = \sum_{k \in \mathcal{K}_j} \mathbf{v}_{k,j} u_k, \tag{2}$$

where $\{u_k\}$ are mutually uncorrelated zero-mean data symbols with the same per-symbol average energy for all $k \in \mathcal{K}_j$ (we assume equal power per stream at each BS). We also use $\nu_j = S_i/M_j$ to denote the spatial load of BS j, that is, the number of downlink streams per BS antenna.

The precoding vectors $\{\mathbf{v}_{k,j} : k \in \mathcal{K}_j\}$ are computed by BS j as a function of the CSI obtained at BS j based on the uplink pilot transmissions. We focus on two of the most popular and well-studied MU-MIMO downlink precoding methods in the context of massive MIMO: conjugate beamforming (CBF) and linear zero-forced beamforming (ZFBF) [4]–[7], [45]. However, we remark several other techniques have been also considered in the literature, including methods for multi-site precoding, e.g., [8]–[10].

Regarding UL transmission, one use of the (complex discrete-time baseband) channel observed at the j-th BS array can be represented as

$$\mathbf{y}_j^{\mathrm{ul}} = \sum_{\ell \in \mathcal{J}} \sum_{k \in \mathcal{K}_\ell} \mathbf{h}_{k,j} u_k^{\mathrm{ul}} + \mathbf{z}_j^{\mathrm{ul}}, \tag{3}$$

where u_k^{ul} is the transmitted information bearing signal stream from UT k, power constrained to $\mathbb{E}\left[\|u_k^{ul}\|^2\right] \leq P_{\mathrm{ul}}$ for some power level P_{ul}, $\mathcal{K}_\ell$ denotes the set of indices of active UTs in cell l, and $\mathbf{z}_j^{\mathrm{ul}} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{M_j})$ represents the additive noise. For decoding the stream of user $k \in \mathcal{K}_j$, BS j filters $\mathbf{y}_j^{ul}$ with an appropriately designed linear filter $\mathbf{v}_k$,

$$y_k^{\mathrm{ul}} = \mathbf{v}_h^{\mathsf{H}} \mathbf{y}_j^{\mathrm{ul}}. \tag{4}$$

The linear filter $\mathbf{v}_k$ again is chosen based on CSI collected from UL sounding. The most widely common types of linear front-ends are maximum ratio combining (MRC), the decorrelator and its regularized generalization.

2.1.2 UL Pilot Transmission and Channel Estimation

A subset Q out of the N channel uses within the generic slot is allocated for UL pilot transmission [4]. CSI is obtained by letting the active users in each cell send their uplink pilot signals on the first Q channel uses in each slot. Then, data transmission (downlink or uplink) takes place in the remaining N-Q channel uses. Each of the active users in the network is assigned one column of the $Q \times Q$ pilot matrix

$$\mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_Q \end{bmatrix}. \tag{5}$$

The matrix $\mathbf{\Phi}$ is a scaled unitary matrix, i.e., $\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi} = Q\alpha^{\mathrm{ul}}\mathbf{I}$, where α^{ul} denotes the transmit power per channel use at a UT in the UL. We index the set of mutually orthogonal pilot signals $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_Q$ by the set $\mathcal{Q} = \{1, 2, \ldots, Q\}$. We let $q(k)$ denote the pilot index of active user k, i.e., active user k uses pilot $\boldsymbol{\phi}_{q(k)}$.

Pilot signals are distributed across the BSs such that BS j is given a subset $\mathcal{Q}^{(j)} \subseteq \mathcal{Q}$ of size $|\mathcal{Q}^{(j)}| = S_j$ of mutually orthogonal pilots (we thus necessarily have $S_j \leqslant Q$). In particular, if $k \in \mathcal{K}_j$, then $q(k) \in \mathcal{Q}^{(j)}$. Also, we let $\mathcal{J}^{(q)} \subseteq \mathcal{J}$ denote the set of BSs which make use of pilot signal q, i.e., $\mathcal{J}^{(q)} = \{j \in \mathcal{J} : q \in \mathcal{Q}^{(j)}\}$. The pilot signal allocation, defined equivalently by the ensembles of sets $\{\mathcal{Q}^{(j)} : j \in \mathcal{J}\}$ or $\{\mathcal{J}^{(q)} : q \in \mathcal{Q}\}$, can be optimized in some suitable way depending on the topology of the network (see for example [6] for a thorough analysis of optimized pilot reuse schemes). The pilot allocations of the type considered here encompass all allocations considered in [4], [6], [7].

The uplink signal block received at BS j during the uplink training phase is given by

$$\mathbf{Y}_j^{\mathrm{ul}} = \sum_{\ell \in \mathcal{J}} \sum_{k \in \mathcal{K}_\ell} \mathbf{h}_{k,j} \boldsymbol{\phi}_{q(k)}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{ul}}, \tag{6}$$

where $\mathbf{Z}_j^{\mathrm{ul}} \in \mathbb{C}^{M_j \times Q}$ with i.i.d. elements $\sim \mathcal{CN}(0, N_0)$.

We shall consider two types of closely related estimators. The first type is based on the CSI estimation approach given in [4]: BS j obtains the estimate of the downlink channel vector for user $k \in \mathcal{K}_j$ up to a real positive scaling factor and some bias additive terms known as *pilot contamination* by projecting $\mathbf{Y}_j^{ul}$ along the pilot signal vector $\boldsymbol{\phi}_{q(k)}$.

Pilot contamination is due to the following reason. Since N is limited by the channel coherence time and bandwidth, Q cannot be arbitrarily large. Hence, the Q mutually orthogonal pilot signals must be reused by several BSs. In particular, the CSI estimate for user $k \in \mathcal{S}_j$, is given by

$$\tilde{\mathbf{h}}_{k,j} = \frac{\mathbf{Y}_j^{\mathrm{ul}} \boldsymbol{\phi}_{q(k)}}{\|\boldsymbol{\phi}_{q(k)}\|^2} = \sum_{n : q(n) = q(k)} \mathbf{h}_{n,j} + \tilde{\mathbf{z}}_{k,j}^{\mathrm{ul}}, \tag{7}$$

which contains the linear combination of the channels from all users k using the same pilot signal $q(k)$ (where users $k \in \mathcal{S}_j$ and the other $k' \neq k$ are active in other cells) to BS j. We also let $\tilde{\mathbf{z}}_{k,j}^{\mathrm{ul}} = \mathbf{Z}_j^{\mathrm{ul}} \boldsymbol{\phi}_{q(k)} \|\boldsymbol{\phi}_{q(k)}\|^{-2}$ denote the projected noise vector with i.i.d. components $\sim \mathcal{CN}(0, \rho_{\mathrm{p}}^{-1})$ where $\rho_{\mathrm{p}} \triangleq \frac{Q\alpha^{\mathrm{ul}}}{N_0}$.

In computing user SINRs and achievable rates, it is convenient to consider the well-known MMSE channel estimates $\hat{\mathbf{h}}_{k,j}$ (whether or not j is the serving BS of user k), based on $\mathbf{Y}_j^{ul}$ and exploit the estimate-error decomposition

$$\mathbf{h}_{k,j} = \hat{\mathbf{h}}_{k,j} + \mathbf{e}_{k,j} \tag{8}$$

with $\hat{\mathbf{h}}_{k,j}$ and $\mathbf{e}_{k,j}$ uncorrelated.

For instance, in the case that $\mathbf{h}_{k,j} \sim \mathcal{CN}(\mathbf{0}, g_{k,j}\mathbf{I})$, with $g_{k,j}$ denoting the large scale gain between UT k and BS j, the MMSE estimator is given by

$$\hat{\mathbf{h}}_{k,j} = w_{k,j}\tilde{\mathbf{h}}_{k,j} \tag{9}$$

with

$$w_{k,j} = \frac{g_{k,j}}{\rho_{\mathrm{p}}^{-1} + G_j^{(q(k))}}, \tag{10}$$

where

$$G_j^{(q_o)} = \sum_{n:q(n)=q_o} g_{n,j}. \tag{11}$$

We also have $\mathbb{E}\left[\hat{\mathbf{h}}_{k,j}\hat{\mathbf{h}}_{k,j}^{\mathsf{H}}\right] = \xi_{k,j}\mathbf{I}$, $\mathbb{E}\left[\mathbf{e}_{k,j}\mathbf{e}_{k,j}^{\mathsf{H}}\right] = \epsilon_{k,j}\mathbf{I}$. It will also prove convenient to express the channel estimate and estimation error power as follows:

$$\xi_{k,j} = \frac{g_{k,j}}{1+\gamma_{k,j}^{-1}} = \frac{g_{k,j}^2}{\rho_{\mathrm{p}}^{-1} + G_j^{(q(k))}}, \tag{12}$$

$$\epsilon_{k,j} = g_{k,j} - \xi_{k,j} = \frac{g_{k,j}}{1+\gamma_{k,j}}, \tag{13}$$

where

$$\gamma_{k,j} = \frac{g_{k,j}}{\rho_{\mathrm{p}}^{-1} + \sum_{n:q(n)=q(k);n\neq k} g_{n.j}}. \tag{14}$$

2.1.3 Marzetta's Insight: Infinite Number of Antennas

Marzetta's seminal work [4] considered the symmetric setting $P_j = P$, $M_j = M$, and $S_j = S$. He considered system operation with S fixed and $M \to \infty$ or equivalently $v_j = v = S/M \to 0$. The only assumption made about the user channels is that the inner products between the channels of different users to the same BS satisfy:

$$\frac{1}{M}\mathbf{h}_{k,j}^{\mathsf{H}}\mathbf{h}_{n,j} \to q_{k,j}\delta_{k-n} \tag{15}$$

as $M \Rightarrow \infty$. The implication is that the magnitude of the inner product between the channels of any two users to the same BS grows sublinearly with M, that is, it grows slower than the norm of the individual user channels. By using only this very broad assumption, the massive MIMO operation limits were derived for the most rudimentary transceivers: CBF in the downlink and MRC in the uplink.

In the DL, Marzetta's CBF precoding vector for active user $k \in \mathcal{K}_j$ at BS j is given by

$$\mathbf{v}_{k,j}^{\text{CBF-M}} = \sqrt{\alpha}\tilde{\mathbf{h}}_{k,j}^{*} \tag{16}$$

where the notation "*" represents the complex conjugate of each element of its channel vector. The scalar parameter α signifies a power scaling selected commonly across all BSs, chosen so that $\mathbb{E}\left[\|\mathbf{x}_j\|^2\right]$ does not violate the power constraint $P_j = P$ at BS j for all j. Substituting $\mathbf{x}_j$ from **(2)** in **(1)** with $\mathbf{v}_{k,j}$ given by $\mathbf{v}_{k,j}^{\text{CBF}}$ in (16) and considering the signal and interference terms with the spatial load $v \to 0$ (i.e., S fixed and $M \to \infty$), Marzetta derived the following expression for the SINR of user $k \in \mathcal{K}_j$:

$$\mathsf{SINR}_{k,j}^{\infty} = \lim_{M\to\infty} \mathsf{SINR}_{k,j} = \frac{g_{k,j}^2}{\sum_{\ell\in\mathcal{J}^{(q(k))};\ell\neq j} g_{n,\ell}^2}. \tag{17}$$

Inspection of (17) reveals that, in the limit of infinite number of BS antennas, the only impact to the user's SINR in the DL comes from pilot contamination, that is, from the fact that its pilot is reused in other nearby cells in the network. The effect of pilot contamination at user k is directly determined by the sum of the large-scale gains between the user and the set of BSs that operate cells where the pilot of UT k is reused by another active user.

For UL data transmission, the vector filter $\mathbf{v}$ in **(4)** used at BS j to project the received vector $\mathbf{y}_j^{\text{ul}}$ in **(3)** is given by

$$\mathbf{v}^{\text{MRC}} \propto \tilde{\mathbf{h}}_{k,j}^{*}. \tag{18}$$

Marzetta derived the following expression for the SINR of user $k \in \mathcal{K}_j$:

$$\overline{\mathsf{SINR}}_{k,j}^{\infty} = \lim_{M\to\infty} \mathsf{SINR}_{k,j}^{\text{ul}} = \frac{g_{k,j}^2}{\sum_{n:q(n)=q(k)} g_{n,j}^2}. \tag{19}$$

Inspection of (19) reveals that, in the limit of infinite number of BS antennas, the only impact to the user's SINR in the UL is also from pilot contamination. The effect of pilot contamination at user k is *different* in the UL: it is captured by the sum of the large scale gains between the user's serving BS and the set of active users in nearby cells reusing the pilot $q(k)$ of user k. Clearly UL/DL duality does not hold in this infinite-number-of-antennas massive MIMO regime.

Marzetta used Shannon-type rate expressions [see, e.g., (21)] based on the limiting SINRs in (17), (19) to study the achievable user-rates and their CDFs in scenarios comprising of regular hexagonal cellular macro-cell layouts, with randomly distributed users and where the large-scale gains include distance-based path-loss and shadowing. From these expressions[†], it is clear that the main dimensionality bottleneck of a massive MIMO cellular system is the number of orthogonal uplink pilot signals Q that can be sent during a fading coherence block. Since the latter is limited by physical quantities, such as the Doppler bandwidth and the coherence bandwidth of the underlying fading random process, and since channel estimation and downlink pre-coded data transmission must occur in the same fading coherence block, the training dimension directly appears as an overhead cost that limits the overall system spectral efficiency. Marzetta justified his $M \to \infty$ approach, by noting that, while the number of users that can be served by spatial multiplexing by each BS is limited by the number of available orthogonal pilots, the number of BS antennas is limited only by technology constraints (and can be made virtually unlimited if the cost of hardware is not taken into account).

[†]See, e.g., (21) with $\mathrm{SINR}_{k,j}$ given by (17).

Simulations in [4] also revealed the extent of pilot contamination in the $M \to \infty$ regime. When all Q pilots are reused in every cell, the impact of pilot contamination on the cell-edge user rates is found to be substantial. To combat pilot contamination conventional frequency reuse-F was advocated, whereby only one every F BS is active in any given resource block (each assigning all its pilots to Q users). Indeed, reuse-7 can substantially improve the edge-user rates, at the cost, however, of a 3-fold reduction in cell throughput.

While Marzetta's seminal work [4] makes some key observations, it turns out to be somewhat simplistic. Some of the resulting drawbacks of the approach in [4] are:

- The analysis considers the regime of infinite BS antennas per user (i.e., $M \to \infty$ with S fixed). As a result, it over-emphasizes the role of pilot contamination as the sole system bottleneck and leads to the misleading conclusion that CBF is as good as any other linear technique, such as ZFBF, in its ability of eliminating completely the intra-cell multiuser interference.
- The analysis does not capture the impact of spatial correlation in the user channels. As it turns out, the user-channel spatial correlation can have substantial implications in the actual system design, operation and performance.

2.1.4 Achievable Rates for Large but Finite Numbers of Antennas

Motivated by some of the shortcomings of the analysis and assumptions in the seminal work [4], tools were subsequently developed in [5]–[7] that enable accurate performance analysis in more realistic multi-cell settings, where the number of antennas at each BS is large but finite. In particular, these works are able to accurately predict the performance for a given set of (M_j, S_i), as the limit of a fictitious setting, where the number of BS antennas and active users at cell j grow to infinity but at a fixed ratio $1/v_j = M_j/S_j$. In this way, the results of [4] are recovered as the special case for this ratio going to 0. More important, this type of analysis shed light on a much more general regime of antennas versus user scaling. It is able to analyze the system when the number of antennas per user is large but finite, which can be regarded as a "practically relevant" massive MIMO regime. In this overview we focus on the downlink expressions. Similar expressions can be readily derived for the uplink channels.

In all cases the achievable rate (over the N channel uses in the given resource block) of active user k served by BS j based on the observation **(1)** is captured via a standard achievable bound (that is based on the worst-case uncorrelated noise) [11], [12]

$$R_{k,j} = \left(1 - \frac{Q}{N}\right)\log_2(1 + \mathsf{SINR}_{k,j}) \tag{20}$$

where

$$\mathsf{SINR}_{k,j} = \frac{\mathbb{E}\left[|\text{useful signal term}|^2 | \{\hat{\mathbf{h}}_{k,\ell}; \ell \in \mathcal{J}\}\right]}{\mathbb{E}\left[|\text{noise + interference term}|^2 | \{\hat{\mathbf{h}}_{k,\ell}; \ell \in \mathcal{J}\}\right]}. \tag{21}$$

The "useful signal term" and the "noise + interference term" are with respect to the signal model **(1)**. In particular, the "useful signal term" is given by $\mathbf{h}_{k,j}^{\mathsf{T}}\mathbf{v}_{k,j}u_k$ while the remaining terms in y_k comprise the "noise + interference term".

The value of $\mathsf{SINR}_{k,j}$ depends on the precoding method and system parameters, such as M_ℓ and S_ℓ, that is, the number of BS antennas and active users at cell l. Using [6], it is shown in [13] that for large M_j and a given set of $\{\nu_\ell = S_\ell/M_\ell\}$, the SINR at active user k served by BS j via CBF according to (16) is closely approximated by the *deterministic* quantity

$$\mathsf{SINR}_{k,j}^{\text{CBF-M}} = \frac{\frac{\rho_j}{\nu_j} q_{k,j}^2}{\eta + \sum_{\ell \in \mathcal{J}} \rho_\ell g_{k,\ell} + \sum_{\ell \in \mathcal{J}^{(q(k))}: \ell \neq j} \frac{\rho_\ell}{\nu_\ell} g_{k,\ell}^2}, \tag{22}$$

where $\rho_j \triangleq P_j/N_0$, and P_j denotes the transmit power constraint at BS j within the given slot. The quantity $\eta \geqslant 1$ arises from using an appropriate normalization value of α in (16) that is common to all BSs and is chosen so the transmit power constraint is not violated at any BS. We note that Marzetta's expression (17) is recovered in the limit $\nu_\ell = \nu \to \infty$ and with the SNR choice $\rho_j = \rho$.

It is also straightforward [13] to derive the corresponding expression with Marzetta's version of ZFBF. In his ZFBF version, the beam for user k in cell j is created as the conjugate of the projection of the user k channel estimate $\tilde{\mathbf{h}}_{k,j}$ in the orthogonal subspace of the channel estimates of the other active users in cell j. Specifically, let $\tilde{\mathbf{H}}_{k,j}$ denote the $M_j \times (S_j - 1)$ matrix whose columns comprise the channel estimates of the form (7) of all active users in cell j except user k. Let also $\mathbf{U}_{k,j}$ denote a basis for the orthogonal complement of $\tilde{\mathbf{H}}_{k,j}$, i.e., $\mathbf{U}_{k,j}$ is a unitary $M_j \times (S_j - 1)$ matrix satisfying $\mathbf{U}_{k,j}^{\mathsf{H}}\mathbf{U}_{k,j} = \mathbf{I}$ and $\tilde{\mathbf{H}}_{k,j}^{\mathsf{T}}\mathbf{U}_{k,j} = \mathbf{0}$. Then, the associated ZFBF precoder is given by

$$\mathbf{v}_{k,j}^{\text{ZF-M}} = \sqrt{\alpha}\mathbf{U}_{k,j}^{\mathsf{H}}\tilde{\mathbf{h}}_{k,j}^{*}. \tag{23}$$

Again, much like (16), the scalar parameter α signifies a power scaling selected commonly across all BSs, and is chosen so that $\mathbb{E}\left[\|\mathbf{x}_j\|^2\right]$ does not violate the power constraint $P_j = P$ at BS j for all j.

Particularizing the analysis in [6], the SINR at user k receiver served by BS j for large M_j and $\nu_j = S_j/M_j$, is closely approximated by the deterministic quantity [13]:

$$\mathsf{SINR}_{k,j}^{\text{ZF-M}} = \frac{\frac{(1-\nu_j)\rho_j}{\nu_j} g_{k,j}^2}{1 + \rho_{\text{p}}^{-1}\rho_j + \sum_{\substack{\ell \in \mathcal{J} \\ \ell \neq j}} q_{k,\ell}\rho_\ell + \sum_{\substack{\ell \in \mathcal{J}^{(q(k))} \\ \ell \neq j}} \frac{(1-\nu_\ell)\rho_\ell}{\nu_\ell} g_{k,\ell}^2}. \tag{24}$$

Comparing (24) with (22), we notice that the effect of ZFBF consists of decreasing the beamforming gain and the pilot contamination effect by the quantity $(1-\nu_j)$, due to ZF precoding. The use of ZFBF, however, also results in reduced intra-cell interference, from $g_{k,j}\rho_j$ to $\rho_{\rm p}^{-1}\rho_j$. In the case of ideal channel estimation for instance, the intra-cell interference with ZFBF is exactly zero.

It is worth making a few additional remarks regarding (22) and (24). First, simple inspection reveals that in the limit of $\nu_j \to 0$ for all j, the SINRs with ZFBF and CBF are both pilot-contamination limited and in fact coincide with (17). Although this confirms the CBF/ZFBF equivalence in the regime of "sufficiently large" numbers of antennas and finite number of users per BS [4], this limiting regime is reached with M being *many* orders of magnitude larger than S. In general, in the practically large-antenna array regime, ZFBF outperforms CBF significantly. Indeed, as shown in [6], in the practical settings involving 10 s and 100 s of antennas, ZFBF can achieve the same performance as CBF using just a small fraction of the antennas. Furthermore, although it is been eluded in [4] that CBF is preferable due to its lower complexity, it turns out the comparison is not simple. Indeed when ZFBF and CBF are both designed in their sweet spots of operation, the computational burden of ZFBF can be similar or even lower than that of CBF[†] [14].

Simple inspection of (22) and (24) also reveals that in each case the SINR of user k is *decoupled* from the channels of the other active users sharing the same pilot in nearby cells. In stark contrast to conventional MIMO transmission, the rate of active user k with massive MIMO can be predicted *a priori*, without prior knowledge of the other active users in its cell and/or other nearby cells and without knowledge of the small-scale fading channels. In fact, the expressions (22) and (24) can be readily used to predict the rate that user k would receive from different cells without actually scheduling the user for transmission. This fact has been exploited in [13] to develop methods for optimal load-balancing association and user scheduling.

It is also noted that (22) and (24) were derived based on a precoder operation according to which the power allocated to each user stream is effectively *proportional* to the users' large-scale channel gains. SINR expressions can also be derived for the equal-power per stream case, which is much more attractive in practice[††] than proportional power allocation.

Particularizing the analysis in [6] to the equal-power per stream version of (16), i.e., to the case that $\|\mathbf{v}_{k,j}\|^2\mathbb{E}\left[|u_k|^2\right] = \rho_j/S_j$ for all active users in cell j, the SINR at user k served by BS j is closely approximated by the deterministic quantity:

$$\mathsf{SINR}_{k,j}^{\rm CBF} = \frac{\frac{\rho_j}{\nu_j}\xi_{k,j}}{1+\sum_{\ell\in\mathcal{J}}\rho_\ell g_{k,\ell} + \sum_{\ell\in\mathcal{J}^{(q(k))};\ell\neq j}\frac{\rho_\ell}{\nu_\ell}\xi_\ell}, \tag{25}$$

where $\xi_{k,j}$ is given by (12). Similarly for the equal-power per stream version of (23), the SINR at user k served by BS j is closely approximated by the quantity:

$$\mathsf{SINR}_{k,j}^{\rm ZF} = \frac{\frac{1-\nu_j}{\nu_j}\rho_j\xi_{k,j}}{1+\epsilon_{k,j}\rho_j + \sum_{\substack{\ell\in\mathcal{J}\\ \ell\neq j}} g_{k,\ell}\rho_\ell + \sum_{\substack{\ell\in\mathcal{J}^{(q(k))}\\ \ell\neq j}}\frac{1-\nu_\ell}{\nu_\ell}\xi_{k,\ell}\rho_\ell}. \tag{26}$$

where $\epsilon_{k,j}$ is given by (13).

Again, it is worth comparing against the Marzetta's $M_j \to \infty$ regime, by considering the symmetric case where SNR_j and ν_j are identical for all BS, and where $\nu_j = \nu = S/M \to 0$. In the limit $\nu \to 0$, the CBF SINR expression (25) also coincides with the ZFBF SINR expression (26). Also, they both converge to the well-known massive MIMO expression (17) in [4] only in the case that the associated large-scale gains $G_\ell^{(q(k))}$ in (11) are the same for all l. Again, although in the limit of $\nu_j \to 0$ for all j the SINRs with ZFBF and CBF coincide [4], in the practical regime involving nonzero spatial loads (finite M_j's), ZFBF yields substantial performance benefits [6]. Indeed comparison of (25) with (26) reveals while the beamforming gain (in both the signal and pilot-contamination terms) is reduced by a factor of $(1-\nu_j)$ with ZFBF, the intra-cell interference with ZFBF can be substantially lower.

Finally it is worth noting that the expressions (25) and (26) are very general and can be used to capture a variety of different operations in a cellular network. For example, in addition to allowing the treatment of heterogeneous networks they allow capturing a number of different conventional interference-control operations. As an example, consider a hexagonal macro-cellular network where each BS has M antennas and Q out of N resource elements in a resource block have been allocated for training. Expressions (25) and (26) can be used to predict the achievable user rates with conventional reuse but also with soft reuse by means of appropriate selection of the $\{\rho_j\}$ and the $\{S_j\}$ sets. Consider for example conventional reuse with factor F and an average transmit power per resource element constraint of Po. In any given resource block only a fraction $1/F$ of BSs are active. Hence, when BS j is active $\nu_j = S/M$ and $\rho_j = FP_o/N_o$, while when BS j is inactive $\nu_i = \rho_j = 0$. The expressions also allow determining the performance in the case of frequency reuse one operation, with pilot reuse F. Assuming in this case that $Q = FS$, BS j is given a subset of S out of Q pilots from $\mathbf{\Phi}$ in (5) while $\nu_j = S/M$ and $\rho_j = P_o/N_o$.

[†]In the regime of high energy efficiency of an optimized massive MIMO system, the total computational burden of CBF may exceed that of ZFBF due to the larger number of UTs that CBF typically serves, and extra computations of the pseudo-inverse for ZFBF typically do not dominate the total computational burden.

[††]Indeed, as is well-known, equal-power allocation becomes asymptotically optimal (with respect to achievable sum-rates) at high SNR [15]. Proportional power allocation is even less preferable from a user-fairness perspective, as less power is allocated to cell-edge users, further reducing their SINRs.

2.1.5 Reducing the Impact of Pilot Contamination

Although pilot contamination can be mitigated by means of conventional frequency reuse or pilot reuse, this comes at a significant expense in cell-edge throughput. As a result several works have considered the problem of dealing with the impact of pilot contamination. One class of approaches involves constructing better user-channel estimators [16]–[18].

The work in [16] exploits the eigenvalue decomposition of the sample covariance matrix of the received UL data signal $\mathbf{y}_j^{\mathrm{UL}}$ at BS j from (3). Assuming independent user channels, with $\mathbf{h}_{k,j} \sim \mathcal{CN}(\mathbf{0}, g_{k,j}\mathbf{I})$, and the $g_{k,j}$ is known (and in the limit $M \to \infty$), each user channel can be (identified) estimated (up to a complex scalar) as one of the eigenvectors of the sample covariance matrix of the received UL data signal $\mathbf{y}_j^{\mathrm{UL}}$. Knowledge of $g_{k,j}$ allows the BS to compute the eigenvalue that is associated with the eigenvector (that is proportional to the channel) of UT k and thus match eigenvectors with UT channels. For a wide range of system parameters, the in-cell user channels can be estimated with greater accuracy than with conventional linear channel estimation methods, which are based on observation of UL pilot transmissions.

Similarly to [16], the work in [17] starts with the singular value decomposition of the received signal matrix. By exploiting results from random matrix theory for large matrices, it is shown that pilot contamination is not a fundamental limitation for massive MIMO systems, if a relative receive-power margin between in-cell and out-of-cell users can be provided, e.g., by means of path loss. In that case, the array gain can be utilized to have the accuracy of channel estimation grow unboundedly with the number of antennas at polynomial complexity. In particular, the support of the eigenvalues associated with the in-cell UTs separates (i.e., it is disjoint) from the support of the eigenvalues associated with the out-of-cell UTs and noise. As a result, the signal subspace can be readily identified. The received signal is subsequently projected onto an (almost) interference-free subspace, allowing accurate estimation of the in-cell UT channels.

Using Bayesian channel estimation, an improved user-channel estimation method is proposed in [18] that leverages knowledge of the spatial correlation of the channels of in-cell and out-of-cell users, and under certain conditions can effectively eliminate the impact of pilot contamination. In particular, [18] advocates the use of covariance-aware pilot assignment (which requires some coordination among BSs), exploiting the fact that, in realistic scenarios involving channels with finite angular spread, the in-cell and out-of-cell user signals are received at the BS with finite-rank covariance matrices. The gains provided by this approach are shown to depend on system parameters such as the typical angle spread measured at the BS and the number of BS antennas. Performance close to the interference-free channel estimation scenario is possible even for moderate numbers of BS antennas.

An alternative method for combating pilot contamination has been considered in [6], [8], which relies on simultaneously serving users across the network at similar geographic regions with respect to their serving BS. To achieve this, transmission resources are a priori split among a set of different "colored"' user-bins across the network so that in any given slot each BS independently schedules users coming from the same-color user-bin (or region). This allows the possibility of partial pilot signal reuse, where the same set of orthogonal pilot signals is reused in, e.g., the center of each cell, while users at the boundary (which are more prone to inter-cell interference) use extended pilot sets that are orthogonal across different sectors or cells, thus providing an option to limit the pilot contamination effect. Each BS bins each user so that users with geographically similar locations are (and are scheduled) in the same color bin. Consequently, although each BS schedules its user population from the appropriate colored bin independently of other BSs, the precoder method, the multiplexing gains, and the pilot reuse factor can be independently optimized for each bin color (or region). Subsequently, SINR expressions were obtained assuming large but finite numbers of antennas for CBF, ZFBF, and a variety of CoMP-type precoders, which are similar to (25) and (26) [6]. By selecting the transmission scheme for each bin that maximizes the bin spectral efficiency, it was shown that significant benefits can be harvested with respect to the single-bin CBF approach in [4] both in terms of (total) cell throughput and cell-edge throughput, thereby effectively limiting the impact of pilot contamination.

2.1.6 Extensions to Spatially Correlated Channels and CoMP

Several works in the literature have considered additional types of operation such as forms of coordinated multipoint transmission (CoMP) and corresponding performance analysis tools [5]–[7], [10]. Some of these tools also allow predicting the performance of cellular massive MIMO with spatially correlated channel models [7].

In terms of CoMP, [6] considers the option of local CoMP, where cooperation is restricted to clusters of nearby BSs. An accurate analysis of the large-system asymptotic performance is conducted for both CBF and various forms of ZFBF where the ZF constraints are imposed to eliminate intra-cell interference or also some fraction of the inter-cell interference, by zero-forcing to users at the edge of adjacent cells. What makes the closed-form analysis possible in this case is the symmetry in the scheduling according to which active user sets are chosen at the cell edges for CoMP transmission. [10] considers an alternative, simpler, distributed form of CoMP, according to which each user is served by multipoint equal-power transmission from nearby BSs based on BS precoders that are constructed locally at each BS, based on local CSI.

In general, however, where the cooperating BSs jointly design their precoders for an arbitrary cell-edge user set,

there are no closed form expressions for the user SINRs (and corresponding achievable rates). In these cases, the user SINRs and rates can be closely approximated as the solution to a fixed-point equation [5], [7]. Although the solution is not in closed form, there exist very simple iterative algorithms, which quickly converge to the solution of the fixed-point equation, yielding the desired user SINRs of the large-scale MIMO system.

The method of deterministic equivalent is very powerful. Not only it allows dealing with a broad range of linear CoMP schemes, but it can also determine the user achievable rates taking into account the user-channel spatial correlation. In particular, it assumes a channel model whereby the channel of active user k to BS j in the given slot is a Gaussian random vector with arbitrary covariance matrix $\mathbf{R}_{k,j}$. Letting $r = r_{k,j}$ denote the number of non-zero eigenvalues of $\mathbf{R}_{k,j}$ the channel of active user k to BS j in the given slot can be expressed as

$$\mathbf{h}_{k,j} = \mathbf{U}_{k,j}\mathbf{\Lambda}_{k,j}^{1/2}\mathbf{a}_{k,j} \tag{27}$$

where $\mathbf{a}_{k,j}$ is an $r \times 1$ vector with i.i.d. $\mathcal{CN}(0, 1)$ entries, $\mathbf{\Lambda}_{k,j}$ an $r \times r$ diagonal matrix comprising the non-zero eigenvalues of $\mathbf{R}_{k,j}$ and $\mathbf{U}_{k,j}$ is an $M_j \times r$ unitary matrix comprising the eigenvectors associated with the r non-zero eigenvalues of $\mathbf{R}_{k,j}$.

In order to determine the effect of pilot contamination in the regime of large but finite numbers of antennas, [7] considered in detail one simple example comprising a cell in the center interfered by a set of surrounding BSs and where $\mathbf{U}_{k,\ell} = \mathbf{U}_k$ for each BS l and the r non-zero eigenvalues of $\mathbf{R}_{k,j}$ are all equal to 1 if l is the user's serving BS, and are equal to an attenuated value $0 < \beta < 1$ if l is an interfering BS. The important aspect of this example is that r captures the richness of the scattering environment (degrees of freedom in the user channels), which in practice may be different from M, the number of BS antennas. As shown in [7], the (intra and inter cell) interference experienced by a user (excluding pilot contamination and thermal noise) mainly depends on the ratio r/S (multiplexing gain per user) and does not directly depend on M. Thus, using additional antennas at the BS can only reduce interference if the environment provides sufficient scattering so that increasing M also increases the richness of the scattering environment.

2.2 DL Massive MIMO Based on DL Training and CSI Feedback

In this section we focus on DL massive MIMO operation based on DL training and CSI feedback. In contrast to DL MIMO transmission based on UL training, where the CSI at the BS is acquired instantly (thereby enabling low-latency MIMO operation), with DL training, CSI is obtained first at the UTs, and subsequently fed back to the BS via an UL transmission. Although this is the MIMO operation used in LTE FDD systems, supporting it in the context of massive MIMO operation, especially at higher carrier frequencies, possesses some key challenges. The inherent delay in the CSI acquisition operation (due to the need for feedback) coupled with the fact that accurate CSI is needed for MU MIMO operation, implies that massive MIMO based on DL training can tolerate lower Doppler that its UL-training counterpart. Equally important, unlike UL training whereby a single UL pilot trains all nearby BS antennas (no matter how many), the overheads do not scale as well with DL training. Indeed, the number of channel dimensions that can be estimated between each of the active UTs and the BS antenna array are *finite*, no matter how large an antenna array is used at the BS. Finally, effective feedback mechanisms are needed for massive MIMO, which can provide a sufficiently accurate CSI estimate at the BS for enabling MU MIMO operation, while keeping the UL feedback transmission overheads acceptable.

Thankfully, the user-channel spatial correlation can be exploited for improving the efficiency of the DL training operation and of the subsequent feedback (in terms of incurred overheads). Here we briefly describe Joint Spatial Division and Multiplexing (JSDM), which is a systematic approach that exploits the structure of the user-channel correlation in order to enable DL MU-MIMO with large BS antenna arrays and reduced CSI acquisition overheads [20]–[22].

We focus on the simplest and most natural application of JSDM, which arises in macro deployments whereby the BS is above the surrounding buildings and structures, and where a uniform linear array is used at the BS[†] [20]. In this case, the transmit-antenna correlation can be approximated via a one-ring model based on which the transmit covariance for user k, and its eigen-space decomposition yielding the representation (27), becomes a function of the user's Angle of Arrival (AoA) distribution.

As shown in Fig. 1, with JSDM the active users are (scheduled and) partitioned in groups, such that users in the same group have nearly similar AoAs while different groups

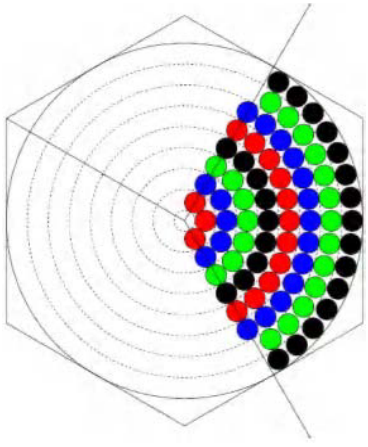


Fig. 1 An example of JSDM layout with 3D beamforming (Fig. 9 in [20]). The concentric regions are separated by the vertical beamforming. The circles indicate user groups. Same-color groups are served simultaneously using JSDM.

[†]The approach readily extends to 2D antenna arrays.

are sufficiently well separated in the AoA domain. JSDM forms the MU-MIMO downlink precoder by concatenating a pre-beamforming matrix, which depends only on the channel second-order statistics, with a classical multiuser precoder, based on the instantaneous knowledge of the resulting lower-dimensionality "effective" channels. Focusing on precoder operation in a single generic cell j and assuming that the active users in the cell have been partition in G groups the precoding signal **(2)** in JSDM can be expressed in the form

$$\mathbf{x}_j = \sum_{g=1}^{G} \mathbf{V}_g \mathbf{u}_g = \sum_{g=1}^{G} \mathbf{B}_g \mathbf{C}_g \mathbf{u}_g, \tag{28}$$

where $\mathbf{u}_g$ comprises the u_k's of the users in group g. $\mathbf{B}_g$ is a (tall) $M \times b_g$ pre-beamforming matrix used for group g that is constructed based on second order statistics and it is common for all users in group g. C_g is a classical $b_g \times s_g$ MU-MIMO matrix for group g determined based on instantaneous CSI, with s_g denoting the size of user group g and where $s_g \leqslant r_g$. Given that users in the same group have nearly similar AoAs the channels of the s_g active users in group g from (27) can be compactly re-expressed as

$$\mathbf{H}_g = \mathbf{U}_g \mathbf{\Lambda}_g^{1/2} \mathbf{A}_g, \tag{29}$$

where the s_g columns of $\mathbf{H}_g$ are the channels of the users in group g, and $\mathbf{U}_g \mathbf{\Lambda}_g$ are from the eigen-decomposition of the (common) transmit covariance matrix of group g.

Substituting the expression for $\mathbf{X}_j$ in **(28)** into **(1)**, we can compactly represent the received signals of the users in group g as

$$\mathbf{y}_g = \check{\mathbf{H}}_g^{\mathsf{T}} \mathbf{C}_g \mathbf{u}_g \mathbf{u}_g + \sum_{g' \neq g} \mathbf{H}_g^{\mathsf{T}} \mathbf{B}_{g'} \mathbf{C}_{g'} \mathbf{u}_{g'} + \mathbf{n}_g, \tag{30}$$

where

$$\check{\mathbf{H}}_g = \mathbf{B}_g^{\mathsf{T}} \mathbf{H}_g \tag{31}$$

denotes the "effective channel" of user group g and $\mathbf{n}_g$ comprises all the noise and out-of-cell interference terms experienced by the active users in group g. The key design feature of JSDM is that the user groups are scheduled with non-overlapping AoA distributions. This allows designing the pre-beamforming vectors so that

$$\mathbf{H}_g^{\mathsf{T}} \mathbf{B}_{g'} \approx 0, \quad \text{for all } g' \neq g \tag{32}$$

thereby enabling transmission to group g that is free from inter-group interference. Indeed, considering the special case where $\mathbf{U}_g^{\mathsf{H}} \mathbf{U}_{g'} = 0$ (any two user-groups have AoA distributions with non-overlapping spatial support), choosing $\mathbf{B}_g = \mathbf{U}_g^*$ yields transmission free from inter-group interference. Hence knowledge of the group effective matrix $\check{\mathbf{H}}_g$ (reduced CSIT) suffices for MU-MIMO transmission based on instantaneous CSI on $\{\check{\mathbf{H}}_g\}$.

Equally important, this CSI can be obtained via low-overhead DL training. To illustrate this, consider the case where there are $s_g = s$ active users per group, $r_g = r$, and $b \geqslant r$ so that **(32)** holds. In this case the BS requires only b pilot dimensions (resource elements allocated to pilots) to allow all sG active users to estimate their effective channels, i.e., to allow each active user in group g to estimate the column of $\check{\mathbf{H}}_g$ corresponding to its channel. Indeed, let the n-th pilot $\boldsymbol{\phi}_n$ (i.e., the pilot across the BS array over the n-th resource element) comprise the sum of the n-th columns of the pre-beamforming matrices $\{\mathbf{B}_g\}_{g=1}^{G}$, i.e.,

$$\boldsymbol{\phi}_n \propto \sum_{g=1}^{G} \mathbf{B}_g \tilde{\mathbf{e}}_n$$

where $\tilde{\mathbf{e}}_n$ is n-th column of a $b \times b$ identity matrix. Due to **(32)**, this pilot transmission allows the users in the g-th group to estimate the n-th column of $\check{\mathbf{H}}_g$ (one entry each). Hence, b pilot transmissions are needed to estimate the effective channel of each active user. Numerical examples given in [20] reveal that an order of magnitude savings can be harvested with such approaches, both in training and feedback dimensions.

It is also worth noting that, for large-scale ULAs, the pre-beamforming matrices $\mathbf{B}_g$ can be obtained by selecting blocks of columns of a unitary Discrete Fourier Transform (DFT) matrix. Indeed by exploiting the method of deterministic equivalent [7] and the Toeplitz eigen-subspace approximation from [23], DFT pre-beamforming is shown to achieve very good cell and cell-edge throughput performance with a large reduction in CSI acquisition overheads, requiring only coarse knowledge of each active group's AoA distribution. The approach is readily extendable to the case of 2-dimensional ULAs and 3D beamforming, whereby fixed beams are also created in the elevation angle direction, in addition to the azimuth angle (planar) direction, with the beamforming matrix taking on the appealing form of a Kronecker product. In this way, angularly separated groups of users in different angular regions in a sector, and at different distances from the BS can be served simultaneously.

2.3 Aspects on Practical Operation and Implementations

2.3.1 MIMO Operation over OFDM

The achievable rate characterization in the previous sections can be readily translated into a practical OFDM system. To illustrate their mapping we consider here one simplified operational scenario from [4] involving cellular downlink transmission and TDD system operation (based on UL training) of the form described in Sect. 2.1. Similar expressions can be also readily derived for its FDD counterparts in Sect. 2.2. Given an OFDM symbol of L_{total} subcarriers with subcarrier spacing Δf and a circular prefix length of $T_{\text{CP}} = L_{\text{CP}}/L_{\text{total}}\Delta f$ seconds, each OFDM symbol has duration $T_s = (L_{\text{total}} + L_{\text{CP}})/L_{\text{total}}\Delta f$, where "$L_{\text{total}}\Delta f$" is the sampling frequency. By assuming the user channels do not exceed the circular prefix, T_{CP}, this implies a coherence bandwidth of L_{total}/L OFDM subcarriers, with $L = L_{\text{CP}}$. Hence knowing the $(L+1)$-tap time-domain impulse response of

a (vector) user-channel suffices to determine all the OFDM user channel coefficients over the L_{total} tones for any OFDM symbol.

The transmission resources on the OFDM plane are partitioned in resource blocks (RBs) or slots. By assuming a user-channel coherence time of $\boldsymbol{I}$ OFDM symbols, each RB comprises $N' = (L_{\text{total}}/L)I$ time-frequency resource elements (REs) or channel uses, spanning a contiguous set of L tones and a contiguous set of $\boldsymbol{I}$ OFDM symbols. In particular, assuming a maximum (accommodated) mobile speed of v, $\boldsymbol{I}$ can be found as the largest integer for which user channel coherence time is given by [53]

$$I L_{\text{total}} + L_{\text{CP}} \leqslant \frac{3}{4\sqrt{\pi} f_o T_{\text{CP}}} \frac{c}{v}$$

with f_o denoting the carrier frequency and c denoting the speed of light.

Over the set of RBs spanning a contiguous set of $\boldsymbol{I}$ OFDM symbols each active BS either serves (independently of other BSs) a subset of users in each cell, or is silent. Whether or not a BS is active over such a set of RBs spanning $\boldsymbol{I}$ contiguous OFDM symbols depends on the frequency reuse pattern; all BSs are active on all RBs with frequency reuse 1, while for instance with frequency reuse F, one every F BS is active on all the RBs spanning these $\boldsymbol{I}$ OFDM symbols.

The UL training - DL data transmission cycle for a set of RBs spanning a contiguous set of OFDM symbols (over all subcarriers) over which a BS is active is as follows. First, $\tau < I$ OFDM symbols are entirely used for UL pilots by the UTs in the cell. The next OFDM symbol is also considered as overhead and serves for allowing the BS to perform the required processing (estimating the user channels, precoding operation, performing other hardware related tasks), and also allowing a guard band for switching between UL and DL transmission mode. Finally the next $I-\tau+1$ OFDM symbols are used by the active BS for MU-MIMO data transmission to its S UTs.

It can be readily verified that within each resource block, it is equivalent to the block fading model, where we have $N = (L_{\text{total}}/L)(I-1)$ channel uses, $Q = \tau(L_{\text{total}}/L)$, and the achievable rates are given by (20) with $\mathbf{R}_{k,j}$ scaled by $I/(I+1)$. By taking into account the numerology of the OFDM system, these user achievable rates can be readily exploited to obtain cell-spectral efficiency and cell-throughput expressions [4].

2.3.2 Simple Resource-Efficient In-Cell Processing

As noted earlier, the massive antenna array can offer "massive MIMO" spectral efficiency gains, with simplified operation with respect to their conventional MIMO counterparts. Even simple precoders with rudimentary power allocation mechanisms can yield large gains with respect to conventional MIMO systems. Inspection of the SINR expressions (22) and (24) also reveals that the effective SINR of an active user does not depend on the identity and the channels of the other active users simultaneously served by the BS (and in fact they can be predicted a priori), in sharp contrast to conventional MIMO. This allows very simple near optimal scheduling, near-optimal user re-association and admission control, subject to a broad class of system fairness criteria [13].

Furthermore, massive MIMO also allows reductions in radiated power both in the UL and DL to achieve a desired SINR level [4]. Many works have consider the potential for power savings on the uplink and of very large MU-MIMO systems as a function of the number of BS antennas, active UTs and precoder method, e.g. [7], [19], [51], [52].

2.3.3 Aspects of Practical Implementations

Many challenges have to be overcome in practice in order to develop cost-effective reliable massive MIMO implementations that can harvest the gains predicted by theory.

One challenge that arises in exploiting the reciprocity-based DL MU-MIMO operation in Sect. 2.1 is that, while the UL/DL radio channels are reciprocal, the transmitter and receiver hardware are not, and introduce time-varying distortions in the UL and the DL channels [47]. In systems without built-in self-calibration capability, these non-reciprocal effects of the receiver and transmitter hardware must be compensated explicitly via a TDD reciprocity calibration protocol. TDD reciprocity was enabled in [47], via a calibration technique based on exchanging pilots between the BS and UTs. An alternative calibration method, referred to as "Argos," was presented in [29] as part of an SDR implementation of a TDD reciprocity-based massive MIMO BS. A key advantage of Argos is that it does not require UT participation. It enables calibration via signaling exchanges and processing involving the BS and a reference antenna. Multi-stage calibration methods are developed in [49], which allow substantially increases in the size of the array that can be calibrated subject to a given calibration overhead, thereby allowing calibration of massive arrays with small overheads.

Enabling reciprocity-based CoMP MU-MIMO transmission with low-cost hardware is even more challenging, as there is a need for synchronization of non-collocated antenna elements driven by different clocks, and the calibration schemes for collocated antennas do not suffice. In [48] calibration methods were presented, which offer substantial calibration improvements with respect to Argos. Also, [48] develops signaling protocols and methods for synchronization, which together with the calibration methods enable reciprocity-based CoMP MU-MIMO transmission with low-cost hardware.

2.4 Massive MIMO Small Cells

Early on in the quest for 5G system designs, it became evident that there are two main approaches for substantially improving the network spectral efficiency per unit area (with respect to 4G systems): massive MIMO arrays and small cell densification.

In massive MIMO, each BS is equipped with a large-scale antenna array in linear, patched, cylindrical, or other preferred shapes, and the antenna elements are deployed centrally. In contrast, in small cell networks, there are a number of micro cells with limited number of antennas. These small cells can be used either to increase the cell coverage, particularly the cell edge areas, or to divide a regular cell into many small cells so that the BS in each small cell can serve a small number of UTs. Thus, they seem to result in two opposing ideas: concentrating the antennas to form massive MIMO BSs, and distributing the antennas to form disjoint small cells. Is this perception true?

While there have been indeed many works in the literature studying and even comparing the performance of massive MIMO and small cells, in 5G networks, these two ideas are essentially not opposite but can be symbiotic for two reasons. First, although densification of small cells shortens the distance between the UTs and interfering BSs, the resulting less inter-cell interference pass-loss could be combated by the massive antenna array gain offered by massive MIMO in its own cell; Second, if we explore higher frequency bands, e.g., mmWave bands above 30 GHz, to avoid overcrowded frequency band allocation in regular bands, the increased inter-cell interference caused by small cell densification could also be neutralized by increased pass-loss and, if necessary, appropriate interference management.

To study the performance gain of massive MIMO in small cells, we can directly apply the math that we develop in prior sections. Specifically, the UL-training operation and expressions derived in Sect. 2.1 are not particular to macro deployments but can also be directly applied to massive MIMO small cells. One such approach is described in [13]. By assuming a pilot reuse factor F, the pilots set $\mathbf{\Phi}$ in (5) are split into $F \geq 4$ disjoint groups. Using a coloring assignment of F colors (one color per set of pilots), the colors (colored sets of pilots) are assigned to cells across the network so that no two neighboring cells are assigned the same color (Otherwise, they would have to share the same pilot set). Then the rate performance over the small cells (with local user scheduling) can be directly inferred from the SINR expressions in (22) and (24). Inspection for instance, of (24), reveals that each BS can predict the peak rate it can provide to its users without requiring knowledge of the scheduling decisions at the other BSs. This decoupling effect can be readily used at each BS for scheduling decisions and by the network for user re-association decisions and admission control [13].

In the remainder of this section we briefly consider other important aspects of massive MIMO small cells, including macro-assisted operation, performance evaluations in hot spots with cellular and CoMP transmission, and operation over higher frequencies.

2.4.1 Macro-assisted Small Cell Deployment with Directional Antenna Array

Massive MIMO deployment in small cells is promising for boosting the data throughput per unit area, by combining the high network density and the large spectral efficiency. As shown in Fig. 2, a macro assisted small cell architecture (also known as a Phantom cell) is considered, where the macro BS utilizes lower frequency bands to ensure seamless coverage and the small cells utilize higher frequency bands to increase the data throughput.

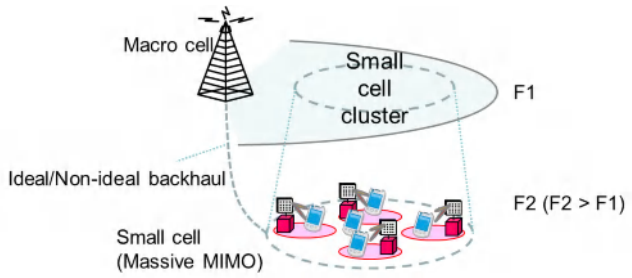


Fig. 2 Massive MIMO in macro-assisted small cell.

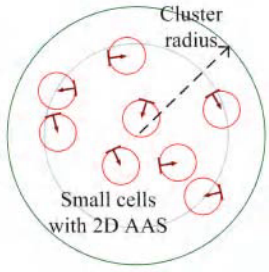


Fig. 3 Clustered small cell deployment with 2D AAS.

A clustered small cell deployment model was defined in 3GPP [33], where the small cell BS with 2D AAS will be placed at the edge of small cells, with the boresight of antenna array pointing to the small cell center, as illustrated in Fig. 3. With the directional antenna array as well as the narrow beam generated by massive MIMO, the signals radiated by the serving BS can be transmitted to the intended users more efficiently with reduced out-of-cell interference compared to that in the small cell deployment with omnidirectional antenna, thereby enabling densification benefits.

2.4.2 Massive MIMO Small Cells in Hot Spots

To illustrate the performance gains that can be offered by massive MIMO and small cell densification, we evaluate the throughput of the network in Shinjuku, a small 120 m × 150 m hotspot area in Tokyo based on the ray-tracing channel data. While the reader can read the full paper in [46], we briefly summarize some observations in this section.

Figure 4 shows the map of the Shinjuku area that we studied, with dots representing BS. In particular, each set of commonly colored dots corresponds to a particular deployment comprising a fixed (and different in each case) number of BSs, with the number ranging from 3 BSs all the way to 12 small cell BSs. In our simulations, we assume that the frequency of the carrier is 3.5 GHz and the available band-

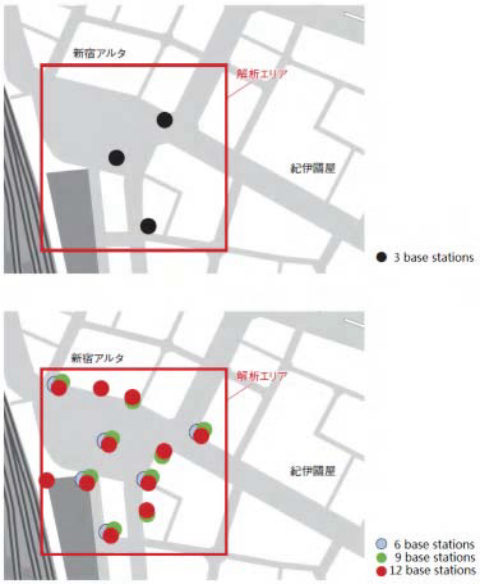


Fig. 4 BSs deployment in Shinjuku, Tokyo (Fig. 12 in [46]).

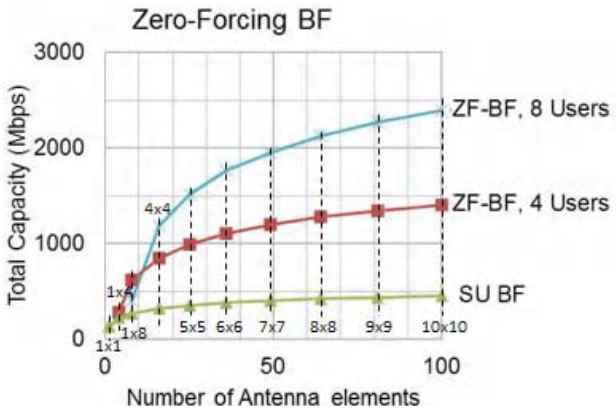


Fig. 5 Network throughput by using ZFBF at BSs regarding the number of antennas at each BS (Fig. 13(b) in [46]).

width is 20 MHz.

By assuming that 6 BSs are turned on, Fig. 5 shows the network throughout performance regarding the number of antennas equipped at each BS for three schemes. In particular, the green curve represents the performance of single-user beamforming at each BS where the BS only serves one user at each time. Note that in the case that the number of antennas equals to 1 this transmission mode reduces to SISO point-to-point transmission. The red and blue curves represent the performance of MU-MIMO where each BS serves 4 and 8 users respectively via ZFBF and equal power allocation per stream. Each UT is associated to the BS from which it has experiences the highest received signal strength. Each BS serves it users via round robin scheduling. It can be easily seen that the throughput increases substantially when the number of antennas grows, and the throughput gain is even larger when MU-MIMO is incorporated.

Finally, Fig. 6 demonstrates the throughput performance when each BS is equipped with 25 antennas, but the number of BSs grows from 3, 6, 9 to 12, i.e., a small cells densification process. In particular, the red curve represents the case of single-antenna everywhere; the blue curve denotes the case of single-user beamforming; and the remaining two curves represent the MU-MIMO ZFBF scheme where each BS serves 4 and 8 users every time, respectively. Observation of Fig. 6 reveals that, besides the performance gains offered by massive MIMO, the small cell densification process also brings prominent throughput gains.

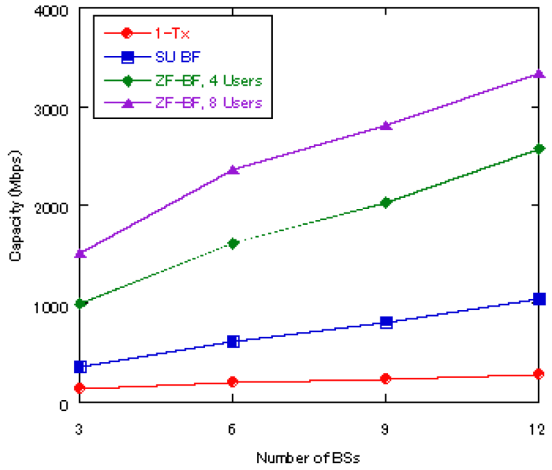


Fig. 6 Network throughput by using ZFBF at BSs regarding the number of BSs in the network (Fig. 14(a) in [46]).

2.4.3 High Frequencies

Massive MIMO small cells promise large spectral efficiencies per unit area and are naturally one of the most promising applications of massive MIMO. With bandwidth expected to become available at higher and higher carrier frequencies [24], it will become increasingly possible to implement massive MIMO in relatively small BSs and within a reasonable form factor. At 3.5 GHz, a small-cell BS implementation that uses, for instance, a 20 cm × 20 cm antenna patch will only be able to support about 25 antenna elements spaced at half wavelength (about 4.3 cm). In contrast, at 30 GHz, where the half-wavelength is 5 mm, a total of 1600 antennas can be supported within the same a 20 cm × 20 cm antenna patch.

As bandwidth at higher frequencies becomes available, it will allow increasing the beamforming gains provided by small cell BSs, thus enabling cellular operations to provide "very dense" massive MIMO type operation and benefits.

However, operation at higher frequencies comes with additional challenges, such as the path-loss decay which is well-known with the square of the wavelength. In fact, many other challenges arise in deploying massive MIMO small cells at higher frequencies. For instance, as the equation in (28) suggests, for the same UT mobility channels decorrelate 10 times faster at 30 GHz than they do at 3 GHz. As a result, TDD operation with reciprocity-based training be-

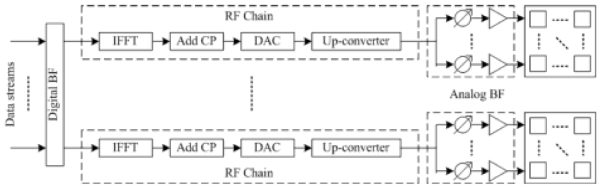


Fig. 7 Hybrid beamforming structure.

comes even more attractive.

Massive MIMO is attractive for high frequency bands, where much more antenna elements can be accommodated within a compact array and the large beamforming gain provided by massive MIMO can effectively compensate the severe path-loss in high frequency bands. Although the full digital array structure can maximize the beamforming flexibility, its cost may be too high due to one separate RF chain for each antenna element. Also, it is impractical to accommodate so many RF chains behind a compact array. Hybrid beamforming provides a feasible solution for massive MIMO deployment in high frequency bands [31], where the antenna array includes multiple analog subarrays and each subarray consisting of multiple antenna elements is connected to one RF chain, as illustrated in Fig. 7. With hybrid beamforming, the beam direction is jointly controlled by analog beamforming at the RF front-end and digital beamforming at the baseband. However, common to both full digital beamforming and hybrid beamforming, common channels and signals cannot benefit from user-specific beamforming gain. Therefore, the careful design of these channels and signals is important in order to ensure coverage [32].

2.4.4 Distributed MIMO

Once the network layout and deployment are determined, another issue to restrict the network performance is how to deal with the interference, because most UTs especially the cell edge UTs are not noise limited but interference limited. One approach to mitigate the interference is called Cooperative Multi-Point (CoMP) transmission where a small number of BSs first share the data intended for a user through backhaul and exchange the CSI, and then they simultaneously transmit the same data to that user via precoding over those BSs. Since all that those BSs form a new virtual BS with all their antenna elements virtually co-located, CoMP is essentially a type of centralized MIMO at the cost of sharing user data and all the CSI used for precoding.

Contrary to centralized MIMO, another approach to mitigate interference is called distributed MIMO. We developed a very simple scheme where a small number of BSs first share only the data intended for a user through backhaul to form a cluster, and then these BSs simultaneously transmit the same data to that user but the precoding is independently calculated at each BS, i.e., the precoding at each BS depends on the CSI of that cell only. By doing so, each user is able to harvest not only the local beamforming gain regarding each BS, but also the cluster beamforming gain at the cost of losing a bit multiplexing gain [10]. Compared to centralized MIMO, this type of distributed MIMO requires much less overhead for precoding, which is translated to improvements in network throughputs and especially the cell-edge user rates in certain network deployment.

3. MIMO Standardization

3.1 IEEE 802 (11ac and 11ad)

There are two main standardization efforts exploiting MIMO within IEEE 802: Very High Throughput (VHT) for frequencies below 6 GHz (except 2.4 GHz) and Directional Multi-gigabit (DMG) for mmWave (frequencies above 45 GHz). 802.11ac and its extension 802.11ax belong to VHT PHY while 802.11ad and its extension (currently under development) are for DMG PHY. For 802.11ac and its extensions, MU-MIMO has played and continues to play an important role: 802.11 ac has adopted MU-MIMO in downlink to enhance throughputs, while the next generation 802.11ax is considering the standardization of MU-MIMO in the uplink.

In 802.11ac MU-MIMO transmission, an access point (AP) can serve up to 8 parallel streams in total to at most 4 users, where at most 4 streams per user are allowed. It is worth noting, however, 802.11ac (and the 802.11ax) is not considering MIMO operation in the "massive MIMO" regime, as the number of AP antennas is of the same order as the number of streams served by the AP.

MU-MIMO in these dimensions can significantly increase downlink throughput in wireless systems, but these large performance gains are possible only if the MU-MIMO beamformer has a very accurate estimate of the channel to each of the MU-MIMO beamformees, similar to the CSI requirement of massive MIMO (as explained in Sect. 2). In 802.11ac, the channel estimate is provided to the AP by an explicit feedback mechanism, which comprises DL training, followed by an uplink feedback.

The feedback mechanism in 802.11ac was developed after careful investigation of the trade-offs between the feedback overhead and the feedback quality. In this process, very efficient descriptions were devised for compressing beamforming information. The explicit feedback consists of a unitary version (instead of the full version) of the channel matrix. This allows a lossless reduction in overhead by a factor of up to 2 [28]. If desired, instead of the full rank version, a reduced-rank version of the channel can also be fed back [28]. Feedback for MU-MIMO in 11ac is similar to the single-user (SU) case. However, the number of quantization bits that can be provided for MU feedback is higher than SU, in order to account for the fact that higher quality CSI is required for enabling MU-MIMO transmission.

For any subcarrier where channel feedback is desired, the channel estimate matrix is first decomposed to the product of an orthonormal matrix and a real diagonal matrix. The orthonormal matrix is then represented as a series of

Givens rotations [27], and the parameters of the rotations are uniformly quantized [28]. An example to illustrate Givens rotations for a single antenna beamformee and a 3-antenna beamformer is provided in [28]. In this example for a single subcarrier 34 feedback bits is required.

Applying the 802.11ac explicit feedback mechanism to massive MIMO deployments is, however, not practical. As we mentioned earlier, first, the number of feedback bits required by givens transformations becomes prohibitive for large antenna arrays. In addition, DL training overheads become prohibitive, as the overheads scale linearly with the number of transmit antennas.

Another standardization effort within IEEE 802 is for the PHY layer in mmWave by 802.11ad and its extensions. 802.11ad operates at roughly an order of magnitude higher carrier frequencies compared to 802.11ac/ax (and uses roughly an order of magnitude more bandwidth). As large-antenna arrays can be packed into a small design at mmWave, in 802.11ad channel training mechanisms are in place so as to enable large RX/TX beamforming gains to compensate for the harsh path-loss. Hence the envisioned operation for 802.11ad is in fact in the massive MIMO regime. IEEE 802 PHY layer standardization at these higher frequencies, however, is at its infancy. Indeed, 802.11ad only considers single-stream transmission, targeting static users in Line-Of-Sight (LOS) environments. This is because in these higher-frequency bands, channels decorrelate a lot faster and the path-loss is much harsher. Currently, extensions of 11ad are under consideration, called 802.11ay, which targets beam-steering and multi-stream operation in non-LOS (NLOS) environments.

3.2 3GPP

MIMO has been one of the key technologies since the first release of 3GPP LTE, i.e., LTE Rel. 8. Until LTE Rel. 12, the LTE specifications only focus on the one dimensional (1D) antenna array and the maximum number of supportable antenna ports is 8, which is far from being massive MIMO.

Before the standardization work for a larger (than 8 antenna ports) antenna array, 3GPP carried out two prerequisite study items (SIs), i.e., the RAN4 SI for active antenna array system (AAS) [34] and the RAN1 SI for three-dimensional (3D) channel model [35].

Contrary to that of the conventional passive antenna array, the radiation pattern of AAS may be dynamically adjusted during operation. The general AAS architecture includes three parts: the Transceiver Unit Array (TXRUA), the Radio Distribution Network (RDN), and the Antenna Array (AA), as shown in Fig. 8. This design provides enough flexibility to support different kinds of MIMO schemes. Specifically, the antenna array can be two-dimensional (2D) and, with proper TXRU connection, it becomes feasible to control the radiation pattern of 2D AAS in both horizontal and vertical dimensions.

3D channel model SI was initiated in 3GPP RAN1 to identify the potential scenarios applicable to elevation

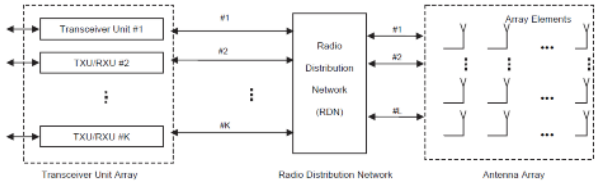


Fig. 8 General AAS architecture [34].

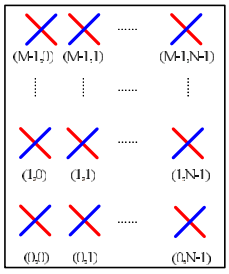
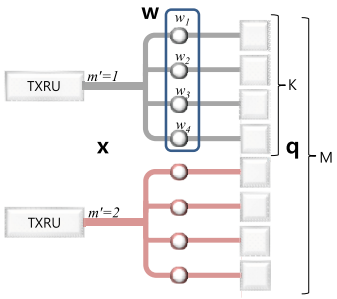


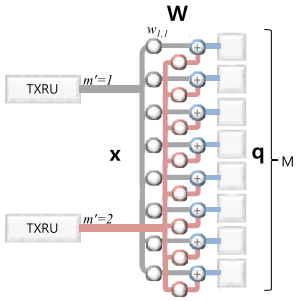
Fig. 9 The 2D planar antenna array structure with cross polarization [35].

beamforming and full-dimensional MIMO (EB/FD-MIMO) and the related evaluation methodology. Contrary to the conventional ITU channel model, the 3D channel model introduces 3 new features: 2D antenna array at the BS, 3D user distribution in both horizontal and vertical domains, and the zenith angle modeling, including both zenith angle of arrival (ZOA) and zenith angle of departure (ZOD). Three scenarios are identified, including 3D Urban Micro (3D-UMi), 3D Urban Macro (3D-UMa), and 3D Urban Macro cell with one high-rise per sector (3D-UMa-H). A 3GPP evaluation methodology needed for EB/FD-MIMO evaluation is also defined, including antenna modeling, path-loss modeling and fast fading model. A 2D planar antenna array structure with cross polarization is taken as the baseline, according to which antenna elements are uniformly placed in the vertical and horizontal direction as shown in Fig. 9, and where N denotes the column number (Note that each column contains two polarizations), and M denotes the row number. To align the implementation of 3D channel modeling in system level simulation, extensive calibration work has been done within the 3GPP community.

After the completion of 3D channel model SI (Study Item) within the time frame of 3GPP LTE Rel. 12, 3GPP RAN1 started the EB/FD-MIMO SI [36] at the beginning of 3GPP LTE Rel. 13 (2014 Q4). The EB/FD-MIMO SI aimed to investigate the performance benefit of standard enhancements targeting 2D antenna array operation with $\{8, 16, 32, 64\}$ TXRUs per transmission point. Two TXRU virtualization models are considered in this SI: ithe sub-array partition model (Fig. 10(a)); the full-connection model (Fig. 10(b)), showing how the connections between antenna elements and each TXRU are defined. Both homogeneous and heterogeneous deployment scenarios are considered in this SI, where EB/FD-MIMO can be applied to Macro and/or small cells. Also, both implementation-



(a) Sub-array partition model [36]



(b) Full-connection model [36]

Fig. 10 Two TXRU vitalization models.

based enhancement schemes and specification-based enhancement schemes for EB/FD-MIMO are thoroughly investigated. Based on this study, several major conclusions are drawn:

- Non-precoded, beamformed, and hybrid channel state information reference signal (CSI-RS) based schemes demonstrate significant throughput gain in realistic non-full buffer traffic models over the best baseline using implementation based enhancements in many scenarios;
- Non-codebook based CSI reporting is beneficial for EB/FD-MIMO;
- Sounding reference signal (SRS) enhancement is beneficial for EB/FD-MIMO;
- Demodulation reference signal (DMRS) enhancement is beneficial for EB/FD-MIMO.

The EB/FD-MIMO work item (WI) [37] aims to specify the potential enhancements identified during the EB/FD-MIMO SI, including extension of the maximum supportable number of antenna ports to 16, SRS and DMRS enhancement, codebook for 2D antenna arrays, and the related control signaling, etc. The EB/FD-MIMO WI was started in August 2015 and it is expected to be completed within the LTE Rel. 13 time frame.

4. Activities towards 5G Massive MIMO

4.1 5G Trials in DOCOMO

In December 2012, the first 10 Gbps packet transmission in the world was successfully tested in the field in outdoor mobile environments using 8×16 MIMO-OFDM transmission system with a 400 MHz bandwidth on an 11 GHz frequency band [38]. In the experiment, the transmission power per antenna was 25 dBm and 10 Gbps transmission was achieved by spatially multiplexing 8 streams using 64QAM, $R = 3/4$ for the modulation and coding scheme. Figure 11 depicts the throughput performance, which was calculated from the received signal measured in the outdoor transmission experiment by off-line processing over the experiment course in Ishigaki City, Japan. Between 10 m and 20 m, and 100 m and 120 m, a throughput greater than 10 Gbps is achieved successfully.

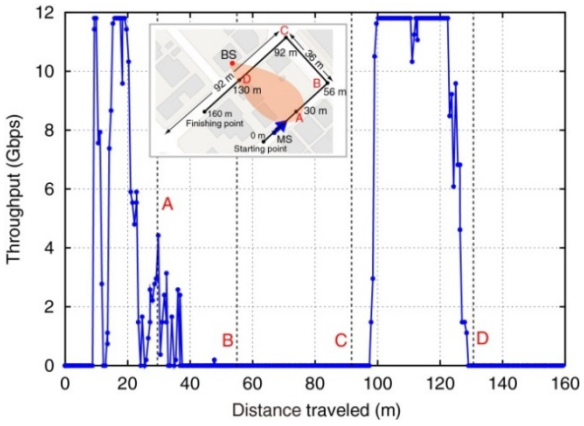


Fig. 11 Measured throughput performance.

Furthermore, experimental trials of emerging 5G mobile technologies for wide range of frequency bands are taking place in collaboration between NTT DOCOMO and world-leading mobile technology vendors [39]. Some related papers are introduced below.

In [40], [41], experimental trials for 5G wireless access at 15 GHz carrier frequency are presented. A picture of outdoor experiments, held in NTT DOCOMO R&D center in Japan, is given in Fig. 12. The trial system can achieve a maximum packet throughput pf more than 5 Gbps by applying 64QAM, 4-stream MIMO, and 400-MHz frequency bandwidth with OFDM waveforms. In [41], it is reported that throughputs over 5 Gbps were achieved in the outdoor environment by applying multi-point distributed MIMO transmissions to make spatial correlation lower even in a LOS environment.

In [42], [43], experimental trials for 3D beam tracking evaluations at a 70 GHz carrier frequency are presented. A picture of outdoor experiments, held in NTT DOCOMO R&D center in Japan, is shown in Fig. 13. As the figure illustrates, an extremely narrow 3-degree beam is applied to extend the coverage in a 70 GHz frequency band. The trial system can achieve a maximum packet throughput of more than 2 Gbps by applying 16QAM over an 1-GHz bandwidth with a single-carrier waveform. In [43], it is reported that a throughput of over 2 Gbps was achieved by applying beam tracking for low mobility users. In the trial, the best beam is adaptively selected from the predetermined candidate beams

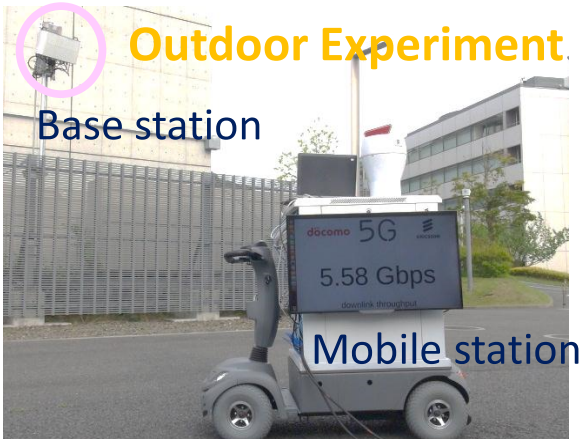


Fig. 12 Experimental trial for 5G wireless access using 15 GHz carrier frequency.

Fig. 13 Experimental trial for 3D beam tracking using the 70 GHz carrier frequency.

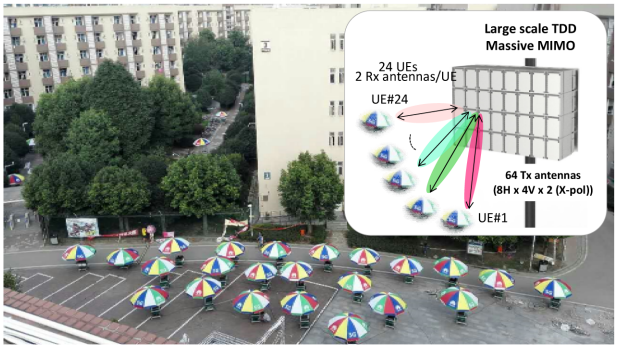


Fig. 14 Large-scale massive MIMO trial with 24 streams using below 6 GHz.

via the user feedback.

In [44], experimental trials for large-scale massive MIMO at carrier frequencies below 6 GHz are presented. A picture of outdoor experiments, held in Chengdu, China, is given in Fig. 14. As shown in the picture, the trial system can realize MU-MIMO transmission with 24 streams, where the BS has 64 transmitter antennas and each user device has 2 receiver antennas. In the trial, antenna weights for multi-user beamforming are generated via the (TDD) channel reciprocity based channel acquisition.

4.2 Channel State Acquisition at Higher Frequencies

Undoubtedly, one of the key challenges in enabling massive MIMO operation at higher frequencies is learning the user channels with acceptable overheads. As such, the usage scenarios and the massive MIMO operation and benefits will also be dictated by the corresponding channel characteristics at the bands where massive MIMO systems are deployed.

For example, increasing the carrier frequency from 3 GHz to 30 GHz results in severe increase in path-loss, and a 10-fold increase in Doppler (for the same speed) [53]. As the same time the channel becomes a lot sparser (i.e., it has a lot fewer dominant multipath components). Given that 5G deployments are expected to operate over bandwidths spanning a much broader frequency range than its predecessors, it is imperative that the design can be flexible enough to accommodate efficient operation over all these bands and scenarios. One approach involves an OFDM system with scalable numerology, so that the OFDM symbol duration and bandwidth can be flexibly scaled by integer factors with respect to a baseline design. Using the LTE numerology for instance as a baseline, such a design would scale the tone spacing by a suitable factor N (and the OFDM symbol duration down by the same factor) where N would be chosen to match the deployment scenarios and frequency bands.

It is also worth mentioning that at mmWave frequencies the channels become sparser in the domains of AOA, Angle of Departure (AOD), delay spread [24] and this sparsity should be exploited for efficient channel acquisition [50] and MU-MIMO precoding [22]. Clearly however, new accurate channel models are needed, which accurately model the required attributes of the massive MIMO channels for the scenarios envisioned, including carrier frequencies, bandwidths and form factors.

5. Conclusion

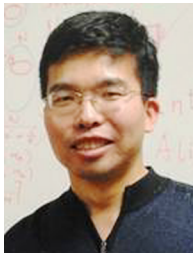
Massive MIMO has attracted significant attention in both academia and industry since its birth and is expected to become an essential 5G technology. In this survey, we briefly overviewed a number of aspects of massive MIMO, ranging from theory to standardization and practical implementations. In particular, we described massive MIMO uplink and downlink operation, and presented performance metrics that capture the impact of pilot contamination on performance. We also discussed the application of massive MIMO in practical OFDM systems, small cells, higher frequency bands, massive MIMO IEEE 802 and 3GPP standards. Finally, we highlighted ongoing activities towards 5G carried out by NTT DOCOMO. We believe that this survey covers the most important aspects of the cutting-edge massive MIMO technology, and sheds light on many open problems, such as channel state acquisition at higher frequencies and the system throughput performance.

References

[1] G.G. Raleigh and J.M. Cioffi, "Spatio-temporal coding for wireless communications," Proc. GLOBECOM'96. 1996 IEEE Global Telecommunications Conference, pp.1809–1814, 1996.
[2] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," Bell Labs Tech. J., vol.1, no.2, pp.41–59, 1996.
[3] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, Cisco, Feb. 2014. Available online: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf
[4] 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.
[5] 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, 2012.
[6] H. Huh, G. Caire, H.C. Papadopoulos, and S.A. Ramprashad, "Achieving "Massive MIMO" spectral efficiency with a not-so-large number of antennas," IEEE Trans. Wireless Commun., vol.11, no.9, pp.3226–3239, 2012.
[7] J. Hoydis, S.T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," IEEE J. Sel. Areas. Commun., vol.31, no.2, pp.160–171, Feb. 2013.
[8] H.C. Papadopoulos, G. Caire, and S.A. Ramprashad, "Achieving large spectral efficiencies from MU-MIMO with tens of antennas: Location-adaptive TDD MU-MIMO design and user scheduling," 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers, pp.636–643, 2010.
[9] S.K. Mohammed and E.G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," IEEE Trans. Commun., vol.61, no.3, pp.1059–1071, March 2013.
[10] Q. Ye, O.Y. Bursalioglu, and H.C. Papadopoulos, "Harmonized cellular and distributed massive MIMO: Load balancing and scheduling," arXiv preprint http://arxiv.org/abs/1509.07594, 2015.
[11] B. Hassibi and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?," IEEE Trans. Inf. Theory, vol.49, no.4, pp.951–963, April 2003.
[12] 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.
[13] D. Bethanabhotla, O.Y. Bursalioglu, H.C. Papadopoulos and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," arXiv preprint arXiv:1407.6731, 2014.
[14] H. Yang and T.L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," IEEE J. Sel. Areas. Commun., vol.31, no.2, pp.172–179, Feb. 2013.
[15] D. Tse and P. Viswanath, Fundamentals of wireless communication, Cambridge University Press, 2005.
[16] H.Q. Ngo and E.G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.3249–3252, March 2012.
[17] R.R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," IEEE J. Sel. Top. Signal Process., vol.8, no.5, pp.773–786, Oct. 2014.
[18] 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.
[19] H.Q. Ngo, E.G. Larsson, and T.L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," IEEE Trans. Commun., vol.61, no.4, pp.1436–1449, April 2013.
[20] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing — The large-scale array regime," IEEE Trans. Inf. Theory, vol.59, no.10, pp.6441–6463, Oct. 2013.
[21] 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. Top. Signal Process., vol.8, no.5, pp.876–890, Oct. 2014.
[22] A. Adhikary, E.A. Safadi, M.K. Samimi, R. Wang, G. Caire, T.S. Rappaport, and A.F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," IEEE J. Sel. Areas. Commun., vol.32, no.6, pp.1239–1255, June 2014.
[23] U. Grenander and G. Szego, Toeplitz forms and their applications, Chelsea Pub Co, 1984.
[24] T.S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," IEEE Access, vol.1, pp.335–349, 2013.
[25] J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C.K. Soong, and J.C. Zhang, "What will 5G be?," IEEE J. Sel. Areas. Commun., vol.32, no.6, pp.1065–1082, June 2014.
[26] I. Hwang, B. Song, and S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," IEEE Commun. Mag., vol.51, no.6, pp.20–27, June 2013.
[27] P. Eldad and R. Stacey, Next Generation Wireless LANS: 802.11n and 802.11ac, Cambridge University Press, 2013.
[28] R. Porat, E. Ojard, N. Jindal, M. Fischer, and V. Erceg, "Improved MU-MIMO performance for future 802.11 systems using differential feedback," Proc. 2013 Information Theory and Applications Workshop (ITA), pp.1–5, 2013.
[29] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: practical many-antenna base stations," Proc. 18th Annual International. Conference on Mobile Computer and Networking, Mobicom, pp. 53–64, 2012. [Online]. Available: http://doi.acm.org/10.1145/2348543.2348553
[30] X. Wang, C. Na, X. Hou, H. Jiang, and H. Kayama, "Wideband and high-order multi-user spatial multiplexing transmission with massive MIMO for future radio access," 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp.338–343, 2015.
[31] J. Zhang, X. Huang, V. Dyadyuk, and Y. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," IEEE Wireless Commun., vol.22, no.1, pp.79–87, 2015.
[32] X. Hou, H. Jiang, H. Kayama and W. Xi, "A novel hybrid beamforming transmission scheme for common channels and signals," accepted by APCC'15, Oct. 2015.
[33] 3GPP R1-144972, "Remaining details on heterogeneous network scenario using different frequency bands," NTT DOCOMO, 3GPP TSG RAN1 meeting #79, Nov. 2014.
[34] 3GPP Technical Report 37.840: "Study of radio frequency (RF) and electromagnetic compatibility (EMC) requirements for active antenna array system (AAS) base station," www.3gpp.org.
[35] 3GPP Technical Report 36.873: "Study on 3D channel model for LTE," www.3gpp.org.
[36] 3GPP Technical Report 36.897: "Study on elevation beamforming/full-dimension (FD) MIMO for LTE," www.3gpp.org.
[37] 3GPP RP-151028, "New WID proposal: Elevation beamforming/full-dimension (FD) MIMO for LTE," Samsung, 3GPP TSG RAN meeting #68, June 2015.
[38] S. Suyama, J. Shen, and Y. Oda, "10 Gbps outdoor transmission experiment for super high bit rate mobile communications," DOCOMO Technical Journal, vol.15, no.4, pp.22–28, April 2014.
[39] T. Nakamura, A. Benjebbour, Y. Kishiyama, S. Suyama, and T. Imai, "5G radio access: Requirements, concept and experimental trials," IEICE Trans. Commun., vol.E98-B, no.8, pp.1397–1406, Aug. 2015.
[40] S. Parkvall, J. Furuskog, E. Dahlman, Y. Kishiyama, A. Harada, and T. Nakamura "A trial system for 5G wireless access," VTC Fall 2015, pp.1–5, Sept. 2015.

[41] D. Kurita, K. Tateishi, A. Harada, Y. Kishiyama, S. Parkvall, E. Dahlman, and J. Furuskog "Field experiments on 5G radio access using multi-point transmission," accepted by IEEE Globecom Workshops 2015.
[42] M. Cudak, T. Kovarik, T.A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mm wave 5G cellular system," 2014 IEEE Globecom Workshops (GC Wkshps), pp.377–381, Dec. 2014.
[43] Y. Inoue, Y. Kishiyama, Y. Okumura, J. Kepler, and M. Cudak, "Experimental evaluation of downlink transmission and beam tracking performance for 5G mmW radio access in indoor shielded environment," Proc. 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp.862–866, 2015.
[44] A. Benjebbour et al., "Large scale experimental trial of 5G air interface," IEICE Gen. Conf. 2015, B-5-63, Sept. 2015.
[45] T.L. Marzetta, "How much training is required for multiuser MIMO?," Proc. 2006 Fortieth Asilomar Conference on Signals, Systems and Computers, pp.359–363, Nov. 2006.
[46] T.L. Marzetta, "5G Key Tehnology: Massive MIMO," Nikkei Electronics, pp.73–82, Nov. 2014.
[47] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," Proc. Future Network and Mobile Summit, 2010, pp.1–10, June 2010.
[48] R. Rogalin, O.Y. Bursalioglu, H. Papadopoulos, G. Caire, A.F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," IEEE Trans. Wireless Commun., vol.13, no.4, pp.1815–1831, April 2014.
[49] H. Papadopoulos, O.Y. Bursalioglu, and G. Caire, "Avalanche: Fast RF calibration of massive arrays," Proc. 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp.607–611, Dec. 2014.
[50] W.U. Bajwa, J. Haupt, A.M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," Proc. IEEE, vol.98, no.6, pp.1058–1076, June 2010.
[51] A. Adhikary, A. Ashikhmin, and T.L. Marzetta, "Uplink interference reduction in large scale antenna systems," Proc. 2014 IEEE International Symposium on Information Theory, pp.2529–2533, July 2014.
[52] H. Yang and T.L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," IEEE J. Sel. Areas. Commun., vol.31, no.2, pp.172–179, Feb. 2013.
[53] T.S. Rappaport, Wireless Communications, Prentice–Hall, New Jersey, 1996.

Haralabos Papadopoulos was born in Serres, Greece, in 1968. He received the S.B., S.M., and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, all in electrical engineering and computer science, in 1990, 1993, and 1998, respectively. Since December 2005, he has been with DOCOMO Innovations, Palo Alto, CA, working on physical-layer algorithms for wireless communication systems and architectures. From 1998 to 2005, he was on the faculty of the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, and held a joint appointment with the Institute of Systems Research. During his 1993–1995 summer visits to AT&T Bell Labs, Murray Hill, NJ, he worked on shared time-division duplexing systems and digital audio broadcasting. His research interests are in the areas of communications and signal processing, with emphasis on resource-efficient algorithms and architectures for wireless communication systems. Dr. Papadopoulos is the recipient of an NSF CAREER Award (2000), the G. Corcoran Award (2000) given by the University of Maryland, College Park, and the 1994 F. C. Hennie Award (1994) given by the MIT EECS department. He is also a coauthor of the VTC Fall 2009 Best Student Paper Award. He is a member of Eta Kappa Nu and Tau Beta Pi. He is also active in the industry and an inventor on several issued and pending patents.



Chenwei Wang received his Ph.D. in Electrical and Computer Engineering from the University of California, Irvine, USA, in December 2012. Prior to that, he received his B.Eng. in Information Engineering and M.S. in Communications and Information Systems from Beijing University of Posts and Telecommunications, China, in 2005 and 2008, respectively. From February 2013, Dr. Wang has been a Research Engineer in the Mobile Network Technology Group of DOCOMO Innovations Inc. (DoCoMo USA Labs before 2012), Palo Alto, CA, USA. His current research interests include the 5G key technologies, network information theory, wireless communications and networks. Before joining DOCOMO Innovations Inc., his industry experience includes intern positions in the Research, Technology and Platforms Group at Nokia Siemens Networks, Beijing, China in 2008, and the Network Innovation Lab of DoCoMo USA Labs, Palo Alto, CA, USA in 2010. He was a recipient of 2014 IEEE Signal Processing Society Young Author Best Paper Award, and also a recipient of the honor of IEEE Communications Letters Exemplary Reviewer in 2011 and 2014.

Ozgun Bursalioglu is a Research Engineer at Docomo Innovations Inc., working in the area of wireless communications on MIMO techniques and LTE-A enhancements since 2012. She graduated from the Ming Hsieh Department of Electrical Engineering, University of Southern California, in 2011. Her Ph.D. thesis is on joint source channel coding for multicast and multiple description coding scenarios using rateless codes. Previously she received M.S. and B.S. degrees from University of California, Riverside (2006) and Middle East Technical University (METU), Ankara, Turkey (2004), respectively. She received the best student paper award at the International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP), in 2006.

Xiaolin Hou received his B.S. and Ph.D. degrees from Beijing Univ. of Posts and Telecommunications (BUPT), Beijing, China, in 2000 and 2005, respectively. He joined DOCOMO Beijing Labs (DBL) in 2005 and now he is a vice principle researcher and research manager in DBL. He has been actively contributing to both academic research and 3GPP standardization in the field of mobile broadband communications. His current research interests focus on 5G key technologies, including dynamic time division duplexing, massive multiple-input multiple-output, and millimeter-wave communications.

Yoshihisa Kishiyama received his B.E., M.E., and Dr. Eng. degrees from Hokkaido University, Japan in 1998, 2000, and 2010, respectively. Since he joined NTT DOCOMO, INC. in 2020, he has been involved in research and standardization activities for 4G LTE/LTE-Advanced and 5G. He is currently a Senior Research Engineer of 5G Laboratory in NTT DOCOMO for 5G radio access network research group. His current research interests include 5G radio interface design and 5G experimental trials. In 2012, he received the International Telecommunication Union Association of Japan (ITU-AJ) Encouragement Award for his contributions to LTE international standardization.