Chapter 1

Energy-efficient design for doubly Massive MIMO millimeter wave wireless systems

Stefano Buzzi and Carmen D'Andrea [1]

1.1 Introduction

Future wireless networks are expected to provide huge performance improvements compared to currently available systems [1]. Fifth generation (5G) and beyond wireless systems will support three different service classes, namely enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable and low-latency communications (URLLC) [2]. The eMBB class will support stable connections with very high peak data rates; the mMTC will support a massive number of Internet of Things (IoT) devices, which are only sporadically active and send small data payloads, while. finally, URLLC will support low-latency transmissions of small payloads with very high reliability from a limited set of terminals, which are active according to patterns typically specifed by outside events. Among the main factors that are at the foundation of 5G and beyond networks we find: (a) the reduction in the size of the radio-cells, so that a larger area spectral efficiency can be achieved; (b) the use of large-scale antenna arrays at the base stations (BSs), i.e., Massive multiple-input-multiple-output (MIMO) systems [3]; and (c) the use of carrier frequencies in the range 10-100 GHz, a.k.a. millimeter waves (mm-waves)[2] [4]. Factor (a) is actually a trend that has been observed for some decades, since the size of the radio cells has been progressively reduced over time from one generation of cellular networks to the next one. Factor (b) has been considered starting from the fourth generation of wireless networks, and indeed the latest Third Generation Partnership Project (3GPP) LTE releases already include the possibility to equip BSs with antenna arrays of up to 64 elements; this trend continues in the 5G New Radio (5G-NR) standard, where fully-digital beamforming structures are adopted. Finally, factor (c) is instead a more recent technology, since for decades mmWaves have been considered as being unsuited for cellular wireless networks.

[1]The authors are with University of Cassino and Southern Latium, Italy and Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT); their work has been supported by Ministero dell'Istruzione dell'Universita' e della Ricerca (MIUR) though the program "Dipartimenti di Eccellenza 2018-2022" and through the Progetti di Ricerca di Interesse Nazionale (PRIN) 2017 "LiquidEDGE" project.

[2]Even though mm-waves is a term that historically refers to the range 30-300 GHz, in the recent literature about future wireless networks the term is used to refer to frequency above-6 GHz.

1.1.1 State of the art

Focusing on the Massive MIMO technology, most of the research and experimental work has traditionally considered its use at conventional cellular frequencies (e.g. sub-6 GHz). Such a range of frequencies is generally denoted as microwave. In the recent past, the combination of the Massive MIMO concept with the use of mm-wave frequency bands has started being considered [5], [6]. For conventional sub-6 GHz cellular systems it has been shown that equipping a BS with a very large (> 100) number of antennas [3, 7, 8] significantly increases the network capacity, mainly due to the capability of serving several users on the same frequency slot with nearly orthogonal vector channels. In traditional Massive MIMO literature, while the number of antennas at the BS grows large, the user device is usually assumed to have only one or very few antennas. When moving to mm-wave, however, the wavelength gets reduced, and, at least in principle, a large number of antennas can be mounted not only on the BS, but also on the user device. As an example, at a carrier frequency of 30 GHz the wavelength is 1 cm, and for a planar antenna array with half-wavelength spacing, more than 180 antennas can be placed in an area as large as a standard credit card. This leads to the concept of *doubly Massive MIMO system* [6, 9], that is defined as a wireless communication system where the number of antennas grows large at both the transmitter and the receiver.

While there are certainly a number of serious practical constraints – e.g., large power consumption, low efficiency of power amplifiers, hardware complexity, analog-to-digital conversion and beamformer implementation – that currently prevent the feasibility of the use of large scale antenna arrays at both sides of the communication links, it is on the other hand believed that these are just technological issues that will be solved or worked around in the near future. On the other hand, since in Massive MIMO systems hardware complexity and energy consumption issues may be problematic, fully digital (FD) beamforming, which requires one radio-frequency (RF) chain for each antenna element, is very challenging; as a consequence, recent research efforts have been devoted towards devising suboptimal, lower complexity, beamforming structures [10]. Hybrid (HY) beamforming structures have been proposed, with a limited number (much smaller than the number of antenna elements) of RF chains. The paper [11], as an instance, analyzes the achievable rate for a MU-MIMO system with HY pre-coding and limited feedback; it is therein shown that, for the case of single-path (i.e., rank-one) channels, HY pre-coding structures achieve a spectral efficiency very close to that of a FD beamformer. In [12], it is shown that a HY beamformer with twice as many RF chains as transmitted data streams may exactly mimic an FD beamformer; the analysis neglects energy efficiency issues, assumes perfect channel state information (CSI) and is limited to either a single-user MIMO system or a MU-MIMO system with single-antenna receivers. The paper [13] proposes a new low-complexity post-coding structure, based on switches rather than on analog phase shifters; the performance of this new structure is evaluated in a rather simple scenario, i.e., single-user MIMO system with a limited number of transmit and receive antennas. The paper [14] focuses on sub-6 GHz frequencies and introduces a novel post-coding structure made of fixed (rather than tunable) phase shifters and of switches, under the assumption that the receiver is

equipped with a large array, while the transmitters have only one antenna. In [15], the authors considers five different low-complexity decoding structures, all based on the use of phase shifters and switches, and provide an analysis of the achievable spectral efficiency along with estimates of the energy consumption of the proposed structures. The paper [16] considers the issue of energy efficiency maximization in a downlink Massive MIMO mm-wave systems by deriving an energy-efficient HY beamformer; however, the paper considers the case in which the user terminals are equipped with just one antenna, and this is a key assumption that is exploited to solve the considered optimization problems. Recently, paper [17] shows that the FD beamforming with zero-forcing (ZF) beamforming at the BS outperforms low-complexity structures both in terms of spectral and energy efficiencies. This trend is also confirmed in [18–20] where the FD beamforming is again considered as a possible candidate technology in order to achieve high energy-efficiency at mmWave frequencies.

1.1.2 Chapter organization

This chapter focuses on the analysis of doubly Massive MIMO mm-wave systems, proposing several beamforming structures and comparing them from the point of view of the energy efficiency. In particular, a detailed description of doubly Massive MIMO systems at mm-waves is given highlighting the differences with respect to massive MIMO systems operating at microwave frequencies. Some relevant use cases where the use of doubly Massive MIMO systems may turn out to be extremely useful are then discussed. Afterwards, building upon mm-wave channel models available in the current literature, the clustered channel model is described. Two main performance measures are defined, i.e., the achievable spectral efficiency (SE) and the global energy efficiency (EE), that is defined as the ratio between the SE and the total power consumption of the network. Beamforming structures are then detailed for FD, HY and analog implementation, with emphasis on the power consumption of each structure. An asymptotic analysis is performed in the limit of large number of transmit and receive antennas, and SE expressions for the FD and analog beamforming (AB) structures in this regime are derived. Using asymptotic formulas for the large number of antennas regime, low-complexity power allocation strategies aimed at the EE maximization are derived. In particular two different techniques are described, one relying on the asymptotic expressions in the interference-free case, and one relying on the interference-limited case. In both cases, the EE-maximing power allocation is obtained by means of fractional programming techniques [21]. Numerical results are finally presented to corroborate the analysis and validate the theoretical findings.

This chapter is organized as follows. Next section contains the description of doubly Massive MIMO systems and the discussion of some relevant use cases. Section 1.3 reports the system model, the transceiver processing and the definition of the performance measures. Section 1.4 details the considered beamforming structures, while Section 1.5 derives the asymptotic SE expressions for the digital and analog beamforming structures. Section 1.6 contains the EE-maximizing power allocation strategies based on the derived asymptotic expressions, while Section 1.7 contains

the discussion about the obtained numerical results. Finally, concluding remarks are given in Section 1.8.

1.1.3 Notation

The following notation is used. The transpose, the inverse and the conjugate transpose of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^{-1}$ and $\mathbf{A}^H$, respectively. The square root of the matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{1/2}$. The trace and the determinant of the matrix $\mathbf{A}$ are denoted as $\mathrm{tr}(\mathbf{A})$ and $|\mathbf{A}|$, respectively. The magnitude of the complex scalar a is denoted as $|a|$. The N-dimensional identity matrix is denoted as $\mathbf{I}_N$, the $(N \times M)$-dimensional matrix with all zero entries is denoted as $\mathbf{0}_{N\times M}$. The matrix containing the rows of $\mathbf{A}$ from the ℓ-th to the ℓ'-th and the columns from the m-th to the m' is denoted as $\mathbf{A}_{(\ell:\ell',m:m')}$. The diagonal matrix obtained from scalars $a_1,\ldots,a_N$ is denoted by $\mathrm{diag}(a_1,\ldots,a_N)$. The statistical expectation operator is denoted as $\mathbb{E}[\cdot]$; $\mathscr{CN}\left(\mu,\sigma^2\right)$ denotes a complex circularly symmetric Gaussian random variable with mean μ and variance σ^2 and $\mathscr{U}(a,b)$ denotes a random variable uniformly distributed in the range $[a,b]$.

1.2 Doubly Massive MIMO systems

The idea of using large scale antenna array was originally launched by Marzetta [7] with reference to BSs. The paper showed that in the limit of a large number of BS antennas small-scale fading effects vanish by virtue of channel hardening and favorable propagation effects. The former effect makes a fading channel behave as deterministic and the latter makes the directions of two users channels asymptotically orthogonal. Consequently, plain channel-matched (CM) beamforming at the BS permits serving several users on the same time-frequency resource slot with (ideally) no interference, and the only left impairment is imperfect channel estimates due to the fact that orthogonal pilots are limited and they must be re-used throughout the network (this is the so-called *pilot contamination*). Recent research has shown that pilot contamination represents a fundamental limit of Massive MIMO system only if maximum ratio combining/precoding is considered [22], and that advanced signal processing techniques can be used to mitigate this effect [23]. Reference [7] considered a system where user devices were equipped with just one antenna. Further studies have extended the Massive MIMO idea at microwave frequencies to the case in which the user devices have multiple antennas, but this number is obviously limited to few units. Indeed, at microwave frequencies the wavelength is in the order of several centimeters, and it is thus difficult to pack many antennas on small-sized user devices. At such frequencies, thus, Massive MIMO just refers to BSs. Things are instead different at mm-waves, wherein multiple antennas are necessary first and foremost to compensate for the increased path loss with respect to conventional sub-6 GHz frequencies [24, 25]. The feasibility of communicating at a high rate in line-of-sight (LOS), benefiting from the wide available bandwidth, also over long distances has been exploited using high-gain directional antennas. Instead of deploying a huge array at one side of the link, the same signal-to noise ratio (SNR) can be achieved

by deploying substantially smaller arrays at both sides. The beamforming gains are multiplied together, so, considering the downlink, instead of having 1000 antennas at the transmitter to serve single-antenna receivers, we can have 100 antennas at the transmitter and 10 antennas at the receivers. This also opens the door to explore systems with massive arrays at both sides. This consideration leads to the concept of doubly Massive MIMO system, firstly introduced in [6, 17], wherein the number of antennas grows large at both sides of the communication link.

1.2.1 Differences with Massive MIMO at microwave frequencies

An understanding of the electromagnetic propagation is crucial when considering Massive MIMO systems at mm-wave and microwave frequencies. The propagation channels build on the same physics, but basic phenomena such as diffraction, attenuation, and Fresnel zones are substantially different [25]. Now, while at microwave the use of hybrid beamformer brings an unavoidable performance degradation, at mm-waves something different happens in the limiting regime of large number of antennas by virtue of the different propagation mechanisms. In particular, in [24] it has been shown that the use of large-scale antenna arrays at mm-wave has not an as beneficial impact on the system multiplexing capabilities as it has at microwave frequencies. In fact, the *clustered* structure of the mm-wave channel makes the purely analog (beam-steering) beamforming optimal in a single-user scenario. Otherwise stated, in a single-user link, the channel eigendirections associated to the largest eigenvalues are just the beam-steering vectors corresponding to the arrival and departure angles and associated with the predominant scatterers. This suggests that pre-coding and post-coding simply requires pointing a beam towards the predominant scatterer at the transmitter and at the receiver, respectively. Additionally, the availability of doubly Massive MIMO wireless links enables the generation of very narrow beams, resulting in reduced co-channel interference to other users using the same time-frequency resources. Another key difference between Massive MIMO at mm-Waves with respect to microwave is the fact that the computational complexity of channel estimation weakly depends on the number of antennas, especially for the case in which analog (beam-steering) beamforming strategies are used [24, 25].

1.2.2 Use cases

The use of doubly Massive MIMO systems at mm-wave in the traditional vision of cellular communication may be unfeasible due to practical constraints that prevent the use of large antenna arrays on mobile handheld devices. However, different and realistic use-cases for doubly massive MIMO systems can be already envisioned. Firstly, considering currently available technology, doubly Massive MIMO mm-wave systems can be used for the *vehicle-to-vehicle (V2V)* and *vehicle-to-infrastructure (V2I)* communications. For example the authors of the paper [26] consider Massive MIMO access points (APs) mounted on street lampposts spaced at very short intervals and they compare the performance obtained with mm-wave and microwave frequencies. The use of large antenna arrays at both sides of the communications can be applied, again with currently available technology, in order to

implement the wireless backhaul in a cellular network. Since mm-wave is also suitable for access links, an *integrated access and backhaul (IAB)* architecture for the 5G cellular networks, in which the same infrastructure and spectral resources will be used for both access and backhaul links [27], can be realized.
Considering now emerging technologies that are not available in today's networks but have significant potential for future sixth-generation (6G) systems, the role of the mm-waves and consequently of the doubly Massive MIMO systems will be crucial. The market demands of 2030 and beyond will introduce new applications, with more stringent requirements, in terms of ultra-high reliability, capacity, energy efficiency, and low latency, which may saturate the capacity of traditional technologies for wireless systems and 6G will contribute to fill this gap [28]. Focusing on the capacity requirements, reference [26] confirms that mm-wave Massive MIMO can deliver Gbps data rates for next-generation wireless networks. The availability of such data rates will allow the implementation of *augmented reality (AR) and virtual reality (VR)* applications. AR and VR over wireless will become a key application in various use cases including, but not limited to, education, training, gaming, workspace communication, and general entertainment. VR/AR applications will require large amount of data-rate and extremely low-latency, so that the use of mm-wave can be crucial both for large bandwidths but also for the high gain obtained from large arrays at both the sides of the communication links. Other 6G applications that require high data rate and very low latency are *holographic telepresence* and *eHealth*: these applications can benefit from mm-wave and doubly Massive MIMO systems. Another interesting application of the doubly Massive MIMO systems can be in the improving the *indoor coverage*. Network infrastructures operating in the mm-wave spectrum will hardly provide indoor connectivity as high-frequency radio signals cannot easily penetrate solid material. One solution to improve the indoor coverage using mm-wave can be the use of cell-free Massive MIMO systems [29, 30] that can be used in order to guarantee a good coverage for indoor users; some results on cell-free mm-wave Massive MIMO systems can indeed be found in [31, 32].

1.3 System model

We focus on the downlink of a doubly Massive MIMO system at mm-wave. The parameter N_{T} denotes the number of antennas at the transmitter, and N_{R} denotes the number of antennas at the receiver[3].

[3]For the sake of simplicity all the receivers are assumed to have the same number of antennas; however, this hypothesis can be easily relaxed.

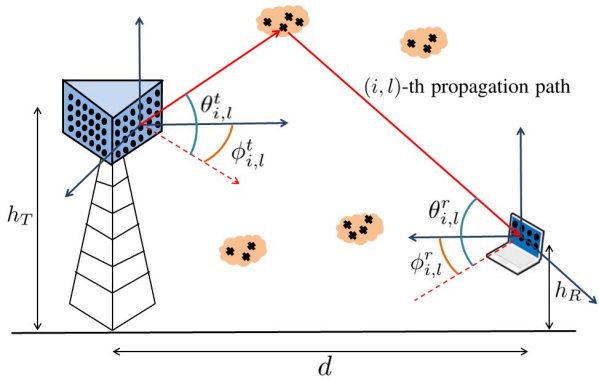


Figure 1.1 The considered reference scenario.

1.3.1 The clustered channel model

The popular narrowband clustered mm-wave channel model is assumed to hold [33–36]. The considered reference scenario for the clustered channel model is reported in Fig. 1.1. The baseband equivalent of the propagation channel between the BS and the generic receiver[4] is thus represented by an $(N_{\rm R} \times N_{\rm T})$-dimensional matrix expressed as[5]:

$$\mathbf{H} = \gamma \sum_{i=1}^{N_{\rm cl}} \sum_{l=1}^{N_{{\rm ray},i}} \alpha_{i,l} \sqrt{L(r_{i,l})} \mathbf{a}_{\rm R}(\phi^{\rm R}_{i,l}) \mathbf{a}^H_{\rm T}(\phi^{\rm T}_{i,l}) + \mathbf{H}_{\rm LOS} \,. \tag{1.1}$$

In Eq. (1.1), it is implicitly assumed that the propagation environment is made of $N_{\rm cl}$ scattering clusters, each of which contributes with $N_{{\rm ray},i}$ propagation paths, $i = 1, \ldots, N_{\rm cl}$, plus a possibly present LOS component. The parameters $\phi^{\rm R}_{i,l}$ and $\phi^{\rm T}_{i,l}$ denote the angles of arrival and departure of the l^{th} ray in the i^{th} scattering cluster, respectively. The quantities $\alpha_{i,l}$ and $L(r_{i,l})$ are the complex path gain and the attenuation associated to the (i,l)-th propagation path. The complex gain $\alpha_{i,l} \sim \mathscr{CN}(0, \sigma^2_{\alpha,i})$, with $\sigma^2_{\alpha,i} = 1$ [33]. The vectors $\mathbf{a}_{\rm R}(\phi^{\rm R}_{i,l})$ and $\mathbf{a}_{\rm T}(\phi^{\rm T}_{i,l})$ represent the normalized receive and transmit array responses evaluated at the corresponding angles of arrival and departure; for an uniform linear array (ULA) with half-wavelength inter-element spacing it holds

$$\mathbf{a}_{\rm T}(\phi^{\rm T}_{i,l}) = \frac{1}{\sqrt{N_{\rm T}}} \left[1 \; e^{-j\pi \sin\phi^{\rm T}_{i,l}} \; \ldots \; e^{-j\pi(N_{\rm T}-1)\sin\phi^{\rm T}_{i,l}} \right]^T . \tag{1.2}$$

[4]For ease of notation, we omit, for the moment, the subscript "k" to denote the BS to the k-th user channel matrix.

[5]For the sake of ease of explanation, a frequency flat channel model us considered here. A wideband channel model as in [11, 37] could however be also adopted.

A similar expression can be also given for $\mathbf{a}_{\mathrm{R}}(\phi_{i,l}^{\mathrm{R}})$. Finally,

$$\gamma = \sqrt{\frac{N_{\mathrm{R}} N_{\mathrm{T}}}{\sum_{i=1}^{N_{\mathrm{cl}}} N_{\mathrm{ray},i}}} \tag{1.3}$$

is a normalization factor ensuring that the received signal power scales linearly with the product $N_{\mathrm{R}} N_{\mathrm{T}}$. With regard to the attenuation of the (i,l)-th path, we considered the results of [38] for four different use-case scenarios: Urban Microcellular (UMi) Open-Square, UMi Street-Canyon, Indoor Hotspot (InH) Office, and InH Shopping Mall. Following [38], the attenuation of the (i,l)-th path is written in logarithmic units as

$$L(r_{i,l}) = -20\log_{10}\left(\frac{4\pi}{\lambda}\right) - 10n\left[1 - b + \frac{bc}{\lambda f_0}\right]\log_{10}\left(r_{i,l}\right) - X_{\sigma}\ , \tag{1.4}$$

with n the path loss exponent, X_{σ} the zero-mean, σ^2-variance Gaussian-distributed shadow fading term in logarithmic units, b a system parameter, and f_0 a fixed reference frequency, the centroid of all the frequencies represented by the path loss model. The values for all these parameters for each use-case scenario are reported in Table 1.1.

Regarding the LOS component, denoting by $\phi_{\mathrm{LOS}}^{\mathrm{R}}$, $\phi_{\mathrm{LOS}}^{\mathrm{T}}$ the arrival and departure angles corresponding to the LOS link, it is assumed that

$$\mathbf{H}_{\mathrm{LOS}} = I_{\mathrm{LOS}}(d)\sqrt{N_{\mathrm{R}} N_{\mathrm{T}} L(d)}\, e^{j\theta}\, \mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{LOS}}^{\mathrm{R}})\mathbf{a}_{\mathrm{T}}^{H}(\phi_{\mathrm{LOS}}^{\mathrm{T}})\ . \tag{1.5}$$

In the above equation, $\theta \sim \mathscr{U}(0, 2\pi)$, while $I_{\mathrm{LOS}}(d)$ is a random variate indicating if a LOS link exists between transmitter and receiver, with $\widetilde{P}$ the probability that $I_{\mathrm{LOS}}(d) = 1$. Denoting by $\widetilde{P}$ the probability that $I_{\mathrm{LOS}}(d) = 1$, i.e., a LOS link exists, we use the results in [38, 39]; for the UMi scenarios, we have:

$$\widetilde{P} = \min\left(\frac{20}{d}, 1\right)\left(1 - e^{-\frac{d}{39}}\right) + e^{-\frac{d}{39}}\ , \tag{1.6}$$

while for the InH scenarios we have:

$$\widetilde{P} = \begin{cases} 1 & d \leq 1.2, \\ e^{-\left(\frac{d-1.2}{4.7}\right)} & 1.2 < d \leq 6.5, \\ 0.32e^{-\left(\frac{d-6.5}{32.6}\right)} & d \geq 6.5. \end{cases} \tag{1.7}$$

A detailed description of all the parameters needed for the generation of sample realizations for the channel model of Eq. (1.1) is reported in [40].

1.3.2 Transmitter and receiver processing

Assume that M denotes the number of streams sent to each user in each signalling interval[6], and $\mathbf{x}_k$ is the M-dimensional vector of the data symbols intended for the

[6] Otherwise stated, the BS transmits in each time-frequency slot MK data symbols.

Table 1.1 Parameters for Path Loss Model [38]

Scenario	Model Parameters
UMi Street Canyon LOS	$n = 1.98$, $\sigma = 3.1$ dB , $b = 0$
UMi Street Canyon NLOS	$n = 3.19$, $\sigma = 8.2$ dB , $b = 0$
UMi Open Square LOS	$n = 1.85$, $\sigma = 4.2$ dB , $b = 0$
UMi Open Square NLOS	$n = 2.89$, $\sigma = 7.1$ dB , $b = 0$
InH Indoor Office LOS	$n = 1.73$, $\sigma = 3.02$ dB , $b = 0$
InH Indoor Office NLOS	$n = 3.19$, $\sigma = 8.29$ dB $b = 0.06$, $f_0 = 24.2$ GHz
InH Shopping Mall LOS	$n = 1.73$, $\sigma = 2.01$ dB , $b = 0$
InH Shopping Mall NLOS	$n = 2.59$, $\sigma = 7.40$ dB $b = 0.01$, $f_0 = 39.5$ GHz

k-th user, with $\mathbb{E}\left[\mathbf{x}_k \mathbf{x}_k^H\right] = \mathbf{I}_M$; the discrete-time signal transmitted by the BS can be expressed as the N_{T}-dimensional vector

$$\mathbf{s} = \sum_{\ell=1}^{K} \mathbf{Q}_\ell \mathbf{P}_\ell^{1/2} \mathbf{x}_\ell \,, \tag{1.8}$$

with $\mathbf{Q}_\ell$ the $(N_{\mathrm{T}} \times M)$-dimensional pre-coding matrix for the ℓ-th user and $\mathbf{P}_k^{1/2} = \mathrm{diag}\left(\sqrt{p_{\ell,1}}, \ldots, \sqrt{p_{\ell,M}}\right)$, with $p_{\ell,q}$ is the downlink transmit power for the ℓ-th user in the q-th stream. The signal received by the generic k-th user is expressed as the following N_{R}-dimensional vector

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s} + \mathbf{n}_k \,, \tag{1.9}$$

with $\mathbf{H}_k$ representing the clustered channel (modeled as in Eq. (1.1)) from the BS to the k-th user and $\mathbf{n}_k$ is the N_{R}-dimensional additive white Gaussian noise with zero-mean independent and identically distributed (i.i.d.) entries with variance σ_n^2. Denoting by $\mathbf{D}_k$ the $(N_{\mathrm{R}} \times M)$-dimensional post-coding matrix at the k-th user device, the following M-dimensional vector is finally obtained:

$$\mathbf{r}_k = \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_k \mathbf{P}_k^{1/2} \mathbf{x}_k + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_\ell \mathbf{P}_\ell^{1/2} \mathbf{x}_\ell + \mathbf{D}_k^H \mathbf{w}_k \,. \tag{1.10}$$

1.3.3 Performance measures

Two performance measures will be considered: the SE and the EE. The SE is measured in [bit/s/Hz], while the EE is measured in [bit/Joule] [41]. Assuming Gaussian data symbols in (1.10), the SE is [42][7]

$$\mathrm{SE} = \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \mathbf{R}_{\mathrm{D},k}^{-1} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_k \mathbf{P}_k \mathbf{Q}_k^H \mathbf{H}_k^H \mathbf{D}_k \right| , \tag{1.11}$$

wherein $\mathbf{R}_{\mathrm{D},k}$ is the covariance matrix of the overall disturbance seen on the downlink by the k-th user receiver, i.e.,

$$\mathbf{R}_{\mathrm{D},k} = \sigma_n^2 \mathbf{D}_k^H \mathbf{D}_k + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_\ell \mathbf{P}_\ell \mathbf{Q}_\ell^H \mathbf{H}_k^H \mathbf{D}_k \, . \tag{1.12}$$

The EE is defined as

$$\mathrm{EE} = \frac{W \mathrm{SE}}{\eta \sum_{k=1}^{K} \sum_{q=1}^{M} p_{k,q} + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} , \tag{1.13}$$

where W is the system bandwidth, $P_{\mathrm{TX,c}}$ is the amount of power consumed by the transmitter circuitry, $P_{\mathrm{RX,c}}$ is the amount of power consumed by the receiver circuitry, and $\eta > 1$ is a scalar coefficient modeling the power amplifier inefficiency [44]. The cost due to signal processing at both the transmitter and the receiver is contained in the terms $P_{\mathrm{TX,c}}$ and $P_{\mathrm{RX,c}}$, which will be specified later for each beamforming structure.

1.4 Beamforming structures

In the following, some beamforming pre-coding and post-coding structures are described, along with details on their power consumption.

1.4.1 Channel-matched, fully-digital (CM-FD) beamforming

Let $\mathbf{H}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{V}_k^H$ denote the singular-value-decomposition (SVD) of the matrix $\mathbf{H}_k$, and assume, without loss of generality, that the diagonal entries of $\mathbf{\Lambda}_k$ are sorted in descending order [45]. The k-th user pre-coding and post-coding matrices are chosen as the columns of the matrices $\mathbf{V}_k$ and $\mathbf{U}_k$, respectively, corresponding to the M largest entries in the eigenvalue matrix $\mathbf{\Lambda}_k$, $\forall k = 1, \ldots, K$, i.e.,

$$\begin{aligned} \mathbf{Q}_k^{\mathrm{CM-FD}} &= [\mathbf{v}_{k,1} \; \mathbf{v}_{k,2} \; \ldots \; \mathbf{v}_{k,M}] \, , \\ \mathbf{D}_k^{\mathrm{CM-FD}} &= [\mathbf{u}_{k,1} \; \mathbf{u}_{k,2} \; \ldots \; \mathbf{u}_{k,M}] \, , \end{aligned} \tag{1.14}$$

where the column vectors $\mathbf{u}_{k,i}$ and $\mathbf{v}_{k,i}$ denote the i-th column of the matrices $\mathbf{U}_k$ and $\mathbf{V}_k$, respectively. In the perfect CSI case, the CM-FD beamforming is optimal in

[7] This expression represents the achievable SE only under the assumption of perfect CSI, that we assume to hold in this chapter. Under the assumption of imperfect CSI, instead, it is possible to derive SE bounds, as reported in [43].

the interference-free case, and tends to be optimal in the case in which the number of antennas at the transmitter grows. The considered FD pre-coding architecture requires a baseband digital precoder that adapts the M data streams to the N_{T} transmit antennas; then, for each antenna there is a digital-to-analog-converter (DAC), an RF chain and a power amplifier (PA). At the receiver, a low noise amplifier (LNA), a RF chain, an analog-to-digital converter (ADC) is required for each antenna, plus a baseband digital combiner that combines the N_{R} outputs of ADC to obtain the soft estimate of the M trasmitted symbols. The amount of power consumed by the transmitter circuitry can thus be expressed as

$$P_{\mathrm{TX,c}} = N_{\mathrm{T}}\left(P_{\mathrm{RFC}} + P_{\mathrm{DAC}} + P_{\mathrm{PA}}\right) + P_{\mathrm{BB}}\ , \tag{1.15}$$

and the amount of power consumed by the receiver circuitry can be expressed as

$$P_{\mathrm{RX,c}} = N_{\mathrm{R}}\left(P_{\mathrm{RFC}} + P_{\mathrm{ADC}} + P_{\mathrm{LNA}}\right) + P_{\mathrm{BB}}\ . \tag{1.16}$$

In the above equations, $P_{\mathrm{RFC}} = 40$ mW [15] is the power consumed by the single RF chain, $P_{\mathrm{DAC}} = 110$ mW [46] is the power consumed by each DAC, $P_{\mathrm{ADC}} = 200$ mW [15] is the power consumed by each ADC, $P_{\mathrm{PA}} = 16$ mW [47] is the power consumed by each PA, $P_{\mathrm{LNA}} = 30$ mW [15] is the power consumed by each LNA, and P_{BB} is the amount of power consumed by each baseband precoder/combiner; assuming a CMOS implementation we have a power consumption of 243 mW [48]. The values of the power consumed by the each ADC present high variability in literature [15]. A conservative value is chosen since the literature does not refer to commercial products and these values might be too optimistic with respect to the final products. The values of the power consumption of each device considered in this section are summarized in Table 1.2.

1.4.2 Partial zero-forcing, fully digital (PZF-FD) beamforming

Zero-forcing pre-coding nulls interference at the receiver through the constraint that the k-th user pre-coding be such that the product $\mathbf{H}_\ell \mathbf{Q}_k = \mathbf{0}_{N_T \times M}$ for all $\ell \neq k$. In order to avoid a too severe noise enhancement, a partial zero-forcing approach is adopted here, namely the columns of the pre-coding matrix $\mathbf{Q}_k$ are required to be orthogonal to the M (the number of transmitted data-streams to each user) right eigenvectors of the channel $\mathbf{H}_\ell$ corresponding to the largest eigenvalues of $\mathbf{H}_\ell$, for all $\ell \neq k$. In this way, the precoder orthogonalizes only to a $M(K-1)$-dimensional subspace and nulls the most significant part of the interference. Formally, the precoder $\mathbf{Q}_k^{\mathrm{PZF-FD}}$ is obtained as the projection of the CM-FD precoder $\mathbf{Q}_k^{\mathrm{CM-FD}}$ onto the orthogonal complement of the subspace spanned by the M dominant right eigenvectors of the channel matrices $\mathbf{H}_1, \ldots, \mathbf{H}_{k-1}, \mathbf{H}_{k+1}, \ldots, \mathbf{H}_K$. Otherwise stated,

$$\mathbf{Q}_k^{\mathrm{PZF-FD}} = \sqrt{M}\frac{\left(\mathbf{I}_{N_T} - \widetilde{\mathbf{V}}_k \overline{\mathbf{V}}_k^H\right)\mathbf{Q}_k^{\mathrm{CM-FD}}}{\left\|\left(\mathbf{I}_{N_T} - \widetilde{\mathbf{V}}_k \widetilde{\mathbf{V}}_k^H\right)\mathbf{Q}_k^{\mathrm{CM-FD}}\right\|_F}\ , \tag{1.17}$$

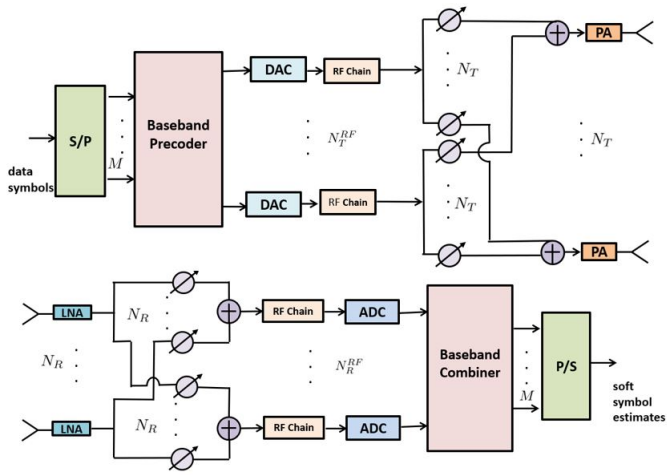


Figure 1.2 Block-scheme of a transceiver with HY digital/analog beamforming.

where $\widetilde{\mathbf{V}}_k$ is a matrix containing M dominant eigenvectors associated to the non-zero eigenvalues of the matrix

$$\overline{\mathbf{V}}_k = [\mathbf{V}_1(:,1:M), \mathbf{V}_2(:,1:M), \ldots, \mathbf{V}_{k-1}(:,1:M), \mathbf{V}_{k+1}(:,1:M), \ldots, \mathbf{V}_K(:,1:M)]$$

and the factor $\sqrt{M}$ allow us to maintain the same Frobenius norm of the precoders. The post-coding matrix is such that $\left(\mathbf{D}_k^{\text{PZF-FD}}\right)^H = \left(\mathbf{H}_k \mathbf{Q}_k^{\text{PZF-FD}}\right)^+$. Since the PZF-FD beamforming requires a FD post-coding, its power consumption is the same as that of the CM-FD beamformer.

1.4.3 Channel-matched, hybrid (CM-HY) beamforming

In order to avoid the use of the same number of RF chains as the number of antennas, HY beamforming architectures have been proposed in literature; in particular, denoting by N_{T}^{RF} and N_{R}^{RF} the number of RF chains available at the transmitter and at the receiver, respectively, the k-th user pre-coding and post-coding matrices are decomposed as follows:

$$\mathbf{Q}_k^{\text{CM-HY}} = \mathbf{Q}_k^{\text{RF}} \mathbf{Q}_k^{\text{BB}} \ , \quad \mathbf{D}_k^{\text{CM-HY}} = \mathbf{D}_k^{\text{RF}} \mathbf{D}_k^{\text{BB}} \ . \tag{1.18}$$

In the above decomposition, the matrices $\mathbf{Q}_k^{\text{RF}}$ and $\mathbf{D}_k^{\text{RF}}$ have dimensions $(N_{\text{T}} \times N_{\text{T}}^{\text{RF}})$ and $(N_{\text{R}} \times N_{\text{R}}^{\text{RF}})$, respectively, and their entries are constrained to have constant norm (i.e., they are implemented through a network of phase-shifters[8]); the matrices $\mathbf{Q}_k^{\text{BB}}$ and $\mathbf{D}_k^{\text{BB}}$, instead, have dimension $(N_{\text{T}}^{\text{RF}} \times M)$ and $(N_{\text{R}}^{\text{RF}} \times M)$, respectively, and their entries are unconstrained complex numbers. A block-scheme of the architecture of the HY transceiver is depicted in Fig. 1.2.

[8]The case of quantized phase-shifts is also considered in the literature, but it is neglected here for the sake of simplicity.

Algorithm 1 Block Coordinate Descent for Subspace Decomposition Algorithm for Hybrid Beamforming

1: Initialize $I_{\max}$ and set $i = 0$
2: Set arbitrary $\mathbf{Q}_{\mathrm{RF},0}$ and $\mathbf{D}_{\mathrm{RF},0}$
3: **repeat**
4: Update $\mathbf{Q}_{\mathrm{BB,i+1}} = \left(\mathbf{Q}_{\mathrm{RF,i}}^H \mathbf{Q}_{\mathrm{RF,i}}\right)^{-1} \mathbf{Q}_{\mathrm{RF,i}}^H \mathbf{Q}^{\mathrm{opt}}$
and $\mathbf{D}_{\mathrm{BB,i+1}} = \left(\mathbf{D}_{\mathrm{RF,i}}^H \mathbf{D}_{\mathrm{RF,i}}\right)^{-1} \mathbf{D}_{\mathrm{RF,i}}^H \mathbf{D}^{\mathrm{opt}}$
5: Set $\phi_i = \mathbf{Q}^{\mathrm{opt}} \mathbf{Q}_{\mathrm{BB,i+1}}^H \left(\mathbf{Q}_{\mathrm{BB,i+1}} \mathbf{Q}_{\mathrm{BB,i+1}}^H\right)^{-1}$
and $\psi_i = \mathbf{D}^{\mathrm{opt}} \mathbf{D}_{\mathrm{BB,i+1}}^H \left(\mathbf{D}_{\mathrm{BB,i+1}} \mathbf{D}_{\mathrm{BB,i+1}}^H\right)^{-1}$
6: Update $\mathbf{Q}_{\mathrm{RF,i}} = \frac{1}{\sqrt{N_{\mathrm{T}}}} e^{j\phi_i}$
and $\mathbf{D}_{\mathrm{RF,i}} = \frac{1}{\sqrt{N_{\mathrm{R}}}} e^{j\psi_i}$
7: Set $i = i + 1$
8: **until** convergence or $i = I_{\max}$

Now, designing an HY beamformer is tantamount to finding expressions for the matrices $\mathbf{Q}_k^{\mathrm{RF}}, \mathbf{Q}_k^{\mathrm{BB}}, \mathbf{D}_k^{\mathrm{RF}}$, and $\mathbf{D}_k^{\mathrm{BB}}$, so that the FD beamforming matrices are approximated with the decomposition reported in Eq. (1.18). For the CM-HY beamforming, the desired beamformers are the CM-FD matrices, and their approximation is here realized by using the block coordinate descent for subspace decomposition (BCD-SD) algorithm [49, 50]. Briefly, a block coordinate descent is an optimization algorithm that successively minimizes along coordinate directions to find the minimum of a function. At each iteration, the algorithm determines a coordinate block via a coordinate selection rule, then minimizes over the corresponding coordinate hyperplane while keeping fixed all other coordinates blocks. A line search along the coordinate direction can be performed at the current iteration to determine the appropriate step size [51]. The BCD-SD algorithm here used is briefly reported in Algorithm 1, with $\mathbf{Q}^{\mathrm{opt}} = \mathbf{Q}^{\mathrm{CM-FD}}$ and $\mathbf{D}^{\mathrm{opt}} = \mathbf{D}^{\mathrm{CM-FD}}$.
The amount of power consumed by the transmitter circuitry is [15]:

$$P_{\mathrm{TX,c}} = N_{\mathrm{T}}^{\mathrm{RF}} \left(P_{\mathrm{RFC}} + P_{\mathrm{DAC}} + N_{\mathrm{T}} P_{\mathrm{PS}}\right) + N_{\mathrm{T}} P_{\mathrm{PA}} + P_{\mathrm{BB}} \,, \tag{1.19}$$

and the amount of power consumed by the receiver circuitry is:

$$P_{\mathrm{RX,c}} = N_{\mathrm{R}}^{\mathrm{RF}} \left(P_{\mathrm{RFC}} + P_{\mathrm{ADC}} + N_{\mathrm{R}} P_{\mathrm{PS}}\right) + N_{\mathrm{R}} P_{\mathrm{LNA}} + P_{\mathrm{BB}} \,. \tag{1.20}$$

Numerical values for the above quantities have already been given, except that for P_{PS}, the power consumed by each phase shifters, that is assumed to be 19.5 mW as in [52].

1.4.4 Partial zero-forcing, hybrid (PZF-HY) beamforming

Similarly to what has been described in the previous subsection, also the PZF beamformers may be approximated through HY architectures. In this case, expressions for

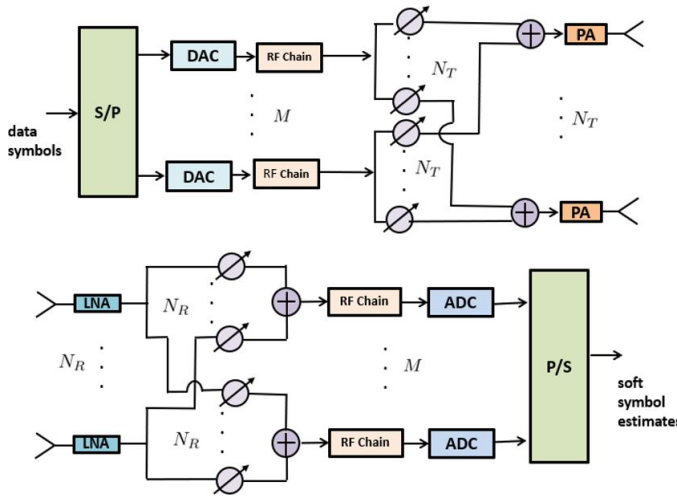


Figure 1.3 Block-scheme of a transceiver with AB.

the matrices $\mathbf{Q}_k^{\mathrm{RF}}, \mathbf{Q}_k^{\mathrm{BB}}, \mathbf{D}_k^{\mathrm{RF}}$, and $\mathbf{D}_k^{\mathrm{BB}}$ are to be found, so that the PZF-FD beamforming matrices are approximated as closely as possible. Also in this case the BCD-SD algorithm can be used and the hybrid beamformers are evaluated following Algorithm 1, with $\mathbf{Q}^{\mathrm{opt}} = \mathbf{Q}^{\mathrm{PZF-FD}}$ and $\mathbf{D}^{\mathrm{opt}} = \mathbf{D}^{\mathrm{PZF-FD}}$.
The amount of power consumed by the transmitter circuitry of th PZF-HY beamformers is the same as that consumed by the CM-HY ones.

1.4.5 Fully analog beam-steering beamforming (AB)

Fully analog beamforming requires that the entries of the pre-coding and post-coding matrices have a constant norm, the block-scheme of a transceiver with AB and M RF chains at both sides of the communication link is reported in Fig. 1.3. Here, it is considered an even simpler structure by introducing a further constraint and assuming that the columns of matrices $\mathbf{Q}_k$ and $\mathbf{D}_k$ are unit-norm beam-steering vectors, i.e., the generic column of an N-dimensional beamformer is

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}}[1 \; e^{-jkd\sin\phi} \; \ldots \; e^{-jkd(N-1)\sin\phi}] \, . \tag{1.21}$$

Focusing on the generic k-th user, the columns of the matrix $\mathbf{Q}_k^{\mathrm{AB}}$ are chosen as the array responses corresponding to the departure angles in the channel model (1.1) associated to the M dominant paths. A similar choice is made for $\mathbf{D}_k^{\mathrm{AB}}$, whose columns contain the array responses corresponding to the M arrival angles associated to the M dominant paths. In order to avoid self-interference, a further constraint is added in the choice of the dominant paths to ensure that the angles of departure (arrival) of the selected paths are spaced of at least 5 deg. Note that for large values of N_{T} and N_{R} the array responses of the transmitter and receiver, corresponding to the departure and arrival angles associated to the dominant propagation path, become concident

Table 1.2 Power consumption of each device [17]

Name	Value	Device	Reference
P_{RFC}	40 mW	RF chain	[15]
P_{DAC}	110 mW	DAC	[46]
P_{ADC}	200 mW	ADC	[15]
P_{PA}	16 mW	PA	[15]
P_{LNA}	30 mW	LNA	[15]
P_{BB}	243 mW	Baseband beamformer	[48]
P_{PS}	19.5 mW	Phase shifter	[52]
P_{element}	27 mW	Element of the phased array	[47]
η	2.66	PA inefficiency	[53]

with dominant right and left singular vectors of the channel. This implies that the AB beamforming structure (1.21) tends to become optimal. The amount of power consumed by the transmitter circuitry is:

$$P_{\text{TX,c}} = N_{\text{T}}^{\text{RF}} \left(P_{\text{RFC}} + N_{\text{T}} P_{\text{element}} + P_{\text{DAC}} \right) , \tag{1.22}$$

and the amount of power consumed by the receiver circuitry is:

$$P_{\text{RX,c}} = N_{\text{R}}^{\text{RF}} \left(P_{\text{RFC}} + N_{\text{R}} P_{\text{element}} + P_{\text{ADC}} \right) , \tag{1.23}$$

where $P_{\text{element}} = 27$ mW [47] is the power consumed by each element of the phased array.

1.5 Asymptotic SE analysis

1.5.1 CM-FD beamforming

In the large number of antennas regime, making the assumption that the set of arrival and departure angles across clusters and users are different with probability 1, it readily follows from the SVD expression of the channel that $\mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_\ell \to \mathbf{\Lambda}_{k,M} \mathbf{V}_{k,M} \mathbf{Q}_\ell$, whenever $k \neq \ell$, where $\mathbf{\Lambda}_{k,M}$ is an $(M \times M)-$dimensional diagonal matrix containing the M largest eigenvalues (denoted by $\lambda_{k,1}, \ldots, \lambda_{k,M}$) of the channel matrix $\mathbf{H}_k$ and $\mathbf{V}_{k,M}$ is an $(N_{\text{T}} \times M)$-dimensional matrix containing the columns of $\mathbf{V}_k$ associated to the eigenvalues in $\mathbf{\Lambda}_{k,M}$. Using the above limiting values, the asymptotic SE in Eq. (1.11) can be expressed as

$$\text{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \left(\sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{\Lambda}_{k,M} \mathbf{V}_{k,M}^H \mathbf{Q}_\ell \mathbf{P}_\ell \mathbf{Q}_\ell^H \mathbf{V}_{k,M} \mathbf{\Lambda}_{k,M}^H \right)^{-1} \mathbf{\Lambda}_{k,M} \mathbf{P}_k \mathbf{\Lambda}_{k,M}^H \right| . \tag{1.24}$$

1.5.2 PZF-FD beamforming

For PZF-FD beamforming, using a similar approach as in Eq. (1.24) and the definition of the PZF-FD in Eq. (1.17), we can note that the product $\mathbf{V}_{k,M}^H \mathbf{Q}_\ell$ is an all-zero

matrix whenever $k \neq \ell$. As a consequence, $\mathbf{R}_{\mathrm{D},k} \sim \sigma_n^2 \mathbf{I}_M$, and the asymptotic SE can be shown to be written as[9]

$$\mathrm{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \frac{1}{\sigma_n^2} \mathbf{\Lambda}_{k,M} \mathbf{P}_k \mathbf{\Lambda}_{k,M}^H \right| . \tag{1.25}$$

In order to explicitly show the dependence of the Eq. (1.25) on the number of antennas, note that the squared moduli of the eigenvalues $\lambda_{k,i}$ depend linearly on the product $N_{\mathrm{T}} N_{\mathrm{R}}$. Otherwise stated, the following holds:

$$\lambda_{k,q} = \sqrt{N_{\mathrm{T}} N_{\mathrm{R}}}\, \widetilde{\lambda}_{k,q} \, , \forall \, k = 1, \ldots, K, \; q = 1, \ldots, M \, , \tag{1.26}$$

with $\widetilde{\lambda}_{k,q}$ normalized eigenvalues independent of the number of transmit and receive antennas. Using this definition Eq. (1.25) can be written as

$$\mathrm{SE} \sim \sum_{k=1}^{K} \sum_{q=1}^{M} \log_2 \left(1 + N_{\mathrm{T}} N_{\mathrm{R}} p_{k,q} \frac{|\widetilde{\lambda}_{k,q}|^2}{\sigma_n^2} \right) . \tag{1.27}$$

1.5.3 Analog beamforming

The case of AB pre-coding and post-coding is now considered. As a preliminary step to the analysis, it is convenient to recall that the ULA response in Eq. (1.21) is a unit-norm vector, and that the inner product between two ULA responses of length P and corresponding to incidence angles ϕ_1 and ϕ_2 is written as

$$f_P(\phi_1, \phi_2) \triangleq \mathbf{a}^H(\phi_1)\mathbf{a}(\phi_2) = \frac{1}{P} \frac{1 - e^{jkd(\sin\phi_1 - \sin\phi_2)P}}{1 - e^{jkd(\sin\phi_1 - \sin\phi_2)}} \, . \tag{1.28}$$

The above inner product, that is denoted by $f_P(\phi_1, \phi_2)$, has a magnitude that, for large P, vanishes as $1/P$, whenever $\phi_1 \neq \phi_2$. Let us now write the channel matrix for k-th user as

$$\mathbf{H}_k = \gamma_k \sum_{i=1}^{N} \alpha_{k,i} \mathbf{a}_{\mathrm{R}}(\phi_{i,k}^{\mathrm{R}}) \mathbf{a}_{\mathrm{T}}^H(\phi_{i,k}^{\mathrm{T}}) = \gamma_k \mathbf{A}_{k,\mathrm{R}} \mathbf{L}_k \mathbf{A}_{k,\mathrm{T}}^H \, , \tag{1.29}$$

namely the path-loss term has been lumped into the coefficients $\alpha_{\cdot,\cdot}$, and the summation over the clusters and the rays has been compressed in just one summation, with $N = N_{\mathrm{cl}} N_{\mathrm{ray}}$. Additionally, $\mathbf{A}_{k,\mathrm{R}}$ is an $(N_{\mathrm{R}} \times N)$-dimensional matrix containing on its columns the vectors $\mathbf{a}_{\mathrm{R}}(\phi_{1,k}^{\mathrm{R}}), \ldots, \mathbf{a}_{\mathrm{R}}(\phi_{N,k}^{\mathrm{R}})$, $\mathbf{L}_k = \mathrm{diag}(\alpha_{1,k}, \ldots, \alpha_{N,k})$, and $\mathbf{A}_{k,\mathrm{T}}$ is an $(N_{\mathrm{T}} \times N)$-dimensional matrix containing on its columns the vectors $\mathbf{a}_{\mathrm{T}}(\phi_{1,k}^{\mathrm{T}}), \ldots, \mathbf{a}_{\mathrm{T}}(\phi_{N,k}^{\mathrm{T}})$[10]. It is also assumed, with no loss of generality, that the paths are sorted in decreasing magnitude order, i.e., $|\alpha_{1,k}| \geq |\alpha_{2,k}| \geq \ldots \geq |\alpha_{N,k}|$. In the following analysis, it is assumed that there are no collisions between arrival and departure angles across users, an assumption that is usually verified unless there are very close users.

[9] This is an asymptotic expression since the noise enhancement effect is neglected (that is a decreasing function of N_{T}) induced by the nulling of the interference.

[10] In order to avoid an heavy notation, it is here dropped the dependence of the matrices $\mathbf{A}_{k,\mathrm{R}}$ and $\mathbf{A}_{k,\mathrm{T}}$ on the propagation paths arrival and departure angles, respectively.

The analog post-coding and pre-coding matrices are written as

$$\begin{aligned}\mathbf{D}_k &= [\mathbf{a}_{\mathrm{R}}(\phi_{1,k}^{\mathrm{R}}),\ldots,\mathbf{a}_{\mathrm{R}}(\phi_{M,k}^{\mathrm{R}})]\,,\\ \mathbf{Q}_k &= [\mathbf{a}_{\mathrm{T}}(\phi_{1,k}^{\mathrm{T}}),\ldots,\mathbf{a}_{\mathrm{T}}(\phi_{M,k}^{\mathrm{T}})]\,,\end{aligned} \tag{1.30}$$

$\forall k$, and they are actually submatrices of $\mathbf{A}_{k,\mathrm{R}}$ and $\mathbf{A}_{k,\mathrm{T}}$, respectively. Define now the following $(M\times N)$-dimensional matrices: $\mathbf{F}_{k,\ell,\mathrm{R}} \triangleq \mathbf{D}_k^H\mathbf{A}_{\ell,\mathrm{R}}$ and $\mathbf{F}_{k,\ell,\mathrm{T}} \triangleq \mathbf{Q}_k^H\mathbf{A}_{\ell,\mathrm{T}}$. Note that the (m,n)-th entry of the matrix $\mathbf{F}_{k,\ell,\mathrm{T}}$ is $f_{N_{\mathrm{T}}}\left(\phi_{m,k}^{\mathrm{T}},\phi_{n,\ell}^{\mathrm{T}}\right)$, while the (m,n)-th entry of the matrix $\mathbf{F}_{k,\ell,\mathrm{R}}$ is $f_{N_{\mathrm{R}}}\left(\phi_{m,k}^{\mathrm{R}},\phi_{n,\ell}^{\mathrm{R}}\right)$. Equipped with this notation, the SE in (1.11) can be now expressed as follows:

$$\mathrm{SE} = \sum_{k=1}^{K}\log_2\left|\mathbf{I}_M + \gamma_k^2\mathbf{R}_{\mathrm{D},k}^{-1}\mathbf{F}_{k,k,\mathrm{R}}\mathbf{L}_k\mathbf{F}_{k,k,\mathrm{T}}^H\mathbf{P}_k\mathbf{F}_{k,k,\mathrm{T}}\mathbf{L}_k^H\mathbf{F}_{k,k,\mathrm{R}}^H\right|\,, \tag{1.31}$$

with

$$\mathbf{R}_{\mathrm{D},k} = \sigma_n^2\mathbf{D}_k^H\mathbf{D}_k + \gamma_k^2\sum_{\substack{\ell=1\\ \ell\neq k}}^{K}\mathbf{F}_{k,k,\mathrm{R}}\mathbf{L}_k\mathbf{F}_{\ell,k,\mathrm{T}}^H\mathbf{P}_\ell\mathbf{F}_{\ell,k,\mathrm{T}}\mathbf{L}_k^H\mathbf{F}_{k,k,\mathrm{R}}^H\,. \tag{1.32}$$

In order to have an asymptotic expression of Eq. (1.31) for a large number of antennas, it can be noted that the $(M\times N)$-dimensional matrix $\mathbf{F}_{k,\ell,\mathrm{R}}$ is such that (a) for $k\neq\ell$ all its entries have a norm that for large N_{R} vanishes as $1/N_{\mathrm{R}}$; while (b) for $k=\ell$ the M entries on the main diagonal are equal to 1 while all the remaining terms again vanish in norm as $1/N_{\mathrm{R}}$. A similar statement also applies to the matrix $\mathbf{F}_{k,\ell,\mathrm{T}}$, of course with entries vanishing as $1/N_{\mathrm{T}}$. Accordingly, the following asymptotic formulas can be proven.

(a) $N_{\mathrm{T}}\to+\infty$**, finite** N_{R}: in this case the system becomes free from the interference between different users and it is obtained

$$\mathrm{SE}\sim\sum_{k=1}^{K}\log_2\left|\mathbf{I}_M + \frac{\gamma_k^2}{\sigma_n^2}\left(\mathbf{D}_k^H\mathbf{D}_k\right)^{-1}\mathbf{F}_{k,k,\mathrm{R}}\mathbf{L}_k\widetilde{\mathbf{P}}_{k,N}\mathbf{L}_k^H\mathbf{F}_{k,k,\mathrm{R}}^H\right|\,, \tag{1.33}$$

where $\widetilde{\mathbf{P}}_{k,N}$ is a N-dimensional diagonal matrix defined as

$$\widetilde{\mathbf{P}}_{k,N} = \mathrm{diag}\left(p_{k,1},\ldots,p_{k,M},\underbrace{0,\ldots,0}_{N-M}\right)\,. \tag{1.34}$$

Note that in this case there is interference between different streams intended for the same user.

(b) $N_{\mathrm{R}}\to+\infty$**, finite** N_{T}: it holds now

$$\mathrm{SE}\sim\sum_{k=1}^{K}\log_2\left|\mathbf{I}_M + \gamma_k^2\mathbf{R}_{\mathrm{D},k}^{-1}\left(\mathbf{L}_k\mathbf{F}_{k,k,\mathrm{T}}^H\mathbf{P}_k\mathbf{F}_{k,k,\mathrm{T}}\mathbf{L}_k^H\right)_{(1:M,1:M)}\right|\,, \tag{1.35}$$

with $\mathbf{R}_{\mathrm{D},k} = \sigma_n^2\mathbf{I}_M + \gamma_k^2\sum_{\substack{\ell=1\\ \ell\neq k}}^{K}\left(\mathbf{L}_k\mathbf{F}_{\ell,k,\mathrm{T}}^H\mathbf{P}_\ell\mathbf{F}_{\ell,k,\mathrm{T}}\mathbf{L}_k^H\right)_{(1:M,1:M)}$.

(c) $N_{\mathrm{R}}, N_{\mathrm{T}} \to \infty$: finally it holds

$$\mathrm{SE} \sim \sum_{k=1}^{K} \sum_{q=1}^{M} \log_2 \left(1 + p_{k,q} \frac{\gamma_k^2 |\alpha_{k,q}|^2}{\sigma_n^2} \right) . \tag{1.36}$$

It is easily seen that the above expression coincides with Eq. (1.25), i.e., the system becomes free from the interference between different users and streams.

1.6 EE maximizing power allocation

In this Section the problem of the downlink power allocation maximizing the asymptotic EE is presented. In particular, two different situations are separately analyzed: the interference-free case and the interference-limited case. The former problem is addressed in order to obtain a simple and low-complexity solution, and then the latter problem is discussed and solved using an alternating optimization technique.

1.6.1 Interference-free case

In the interference-free case the SE of one user is not corrupted by other co-channel communications. From the asymptotic SE analysis, we can note that Eqs. (1.27) and (1.36) fall in this category. We thus focus on the following optimization problem

$$\begin{cases} \max\limits_{\mathbf{p}} \dfrac{\sum\limits_{k=1}^{K} \sum\limits_{q=1}^{M} W \log_2 \left(1 + p_{k,q} g_{k,q} \right)}{\eta \sum\limits_{k=1}^{K} \sum\limits_{q=1}^{M} p_{k,q} + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} \\ \text{s.t.} \sum\limits_{k=1}^{K} \sum\limits_{q=1}^{K} p_{k,q} \leq P_{\max} \\ p_{k,q} \geq 0 \, , \forall \, k = 1, \ldots, K, \, q = 1, \ldots, M \, , \end{cases} \tag{1.37}$$

where $\mathbf{p} = [p_{1,1}, \ldots, p_{1,M}, \ldots, p_{K,1}, \ldots, p_{K,M}]^T$. Notice that, considering Eq. (1.27), i.e., PZF-FD beamforming, $g_{k,q} = N_{\mathrm{T}} N_{\mathrm{R}} \frac{|\tilde{\lambda}_{k,q}|^2}{\sigma_n^2}$, and considering Eq. (1.36), i.e., AB, $g_{k,q} = \frac{\gamma_k^2 |\alpha_{k,q}|^2}{\sigma_n^2}$. The objective function in Problem (1.37) is the ratio of a concave function (with respect to the optimization variables) over a linear one, and, thus, Dinkelbach's algorithm in [54] may be readily applied to maximize the ratio [21,55]. Denoting as $N(\mathbf{p})$ and $D(\mathbf{p})$ as

$$N(\mathbf{p}) = \sum_{k=1}^{K} \sum_{q=1}^{M} W \log_2 \left(1 + p_{k,q} g_{k,q} \right) , \tag{1.38}$$

$$D(\mathbf{p}) = \eta \sum_{k=1}^{K} \sum_{q=1}^{M} p_{k,q} + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}} \tag{1.39}$$

the Dinkelbach's procedure can be summarized as in Algorithm 2.

Algorithm 2 Dinkelbach's procedure [54]

1: Set $\varepsilon = 0$, $\pi = 0$ and FLAG=0
2: **repeat**
3: Update $\mathbf{p}$ by solving the following concave maximization:
$$\max_{\mathbf{p}} N(\mathbf{p}) - \pi D(\mathbf{p}) \tag{1.40}$$
4: **if** $N(\mathbf{p}) - \pi D(\mathbf{p}) < \varepsilon$ **then**
5: FLAG=1
6: **else**
7: Set $\pi = \dfrac{N(\mathbf{p})}{D(\mathbf{p})}$
8: **end if**
9: **until** FLAG=1

Given the structure of the objective in Problem (1.37), we can note that standard techniques can be used to solve the concave maximization in Problem (1.40). The solution of the concave maximization in Problem (1.40) can be computed starting from the following KKT conditions [56]

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}p_{k,q}} \left(N(\mathbf{p}) - \pi D(\mathbf{p})\right) - \lambda + \mu_{k,q} = 0, \ \forall k,q \\
&\sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \leq P_{\max} \\
&\lambda \left(P_{\max} - \sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \right) = 0 \\
&\lambda >= 0 \\
&p_{k,q} \geq 0, \ \forall k,q \\
&\mu_{k,q} p_{k,q} = 0, \ \forall k,q \\
&\mu_{k,q} \geq 0, \ \forall k,q
\end{aligned} \tag{1.41}$$

where λ and $\mu_{k,q}$ are the Lagrange multipliers associated to the constraints on the maximum power radiated by the transmitter and on the minimum power level of the transmitter to the k-th user on the q-th streams, respectively. After standard algebraic manipulations we obtain the following waterfilling-like problem

$$\begin{cases}
p_{k,q} = \max\left\{0, \dfrac{W/\ln 2}{\pi\eta + \lambda} - \dfrac{1}{g_{k,q}}\right\} \\
\displaystyle\sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \leq P_{\max}
\end{cases} \tag{1.42}$$

and the optimal value of the non-negative Lagrange multiplier λ can be derived by bisection search. The optimality and convergence conditions of the algorithm to solve Problem (1.37) follow from the ones of Dinkelbach's procedure in Algorithm 2, see [21, 54], further details are omitted for the sake of brevity.

1.6.2 Interference-limited case

The interference-limited case includes the asymptotic SE expressions in Eqs. (1.24), (1.33) and (1.35). We can rewrite these expressions in the following common form

$$\begin{aligned}\mathrm{SE} &= \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \left(\sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right)^{-1} \mathbf{A}_{k,k} \mathbf{P}_k \mathbf{A}_{k,k}^H \right| \\ &= \sum_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|,\end{aligned} \tag{1.43}$$

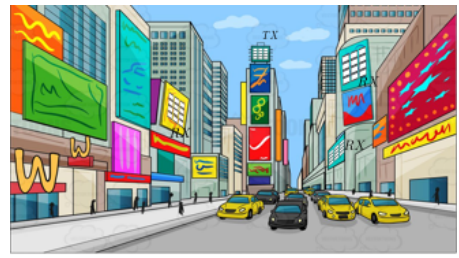
where $\mathbf{A}_{k,\ell}, \forall k, \ell$ are properly defined matrices. Given the SE expression in Eq. (1.43), the EE maximizing power allocation is obtained by solving the following optimization problem

$$\begin{cases} \max\limits_{\mathbf{P}_1, \ldots \mathbf{P}_K} & \dfrac{\sum\limits_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum\limits_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum\limits_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|}{\eta \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} \\ \text{s.t.} & \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) \leq P_{\max} \\ & p_{k,q} \geq 0 \, , \forall \, k = 1, \ldots, K, \, q = 1, \ldots, M \, , \end{cases} \tag{1.44}$$

Problem (1.44) has non-concave objective function, which makes its solution challenging. Additionally, the large number of optimization variables, would still pose a significant complexity challenge. To solve (1.44) the framework of alternating optimization can be used, see [57, Section 2.7] for the details. In particular, each subproblem, one for each user, can be written as

$$\begin{cases} \max\limits_{\mathbf{P}_k} & \dfrac{\sum\limits_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum\limits_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum\limits_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|}{\eta \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} \\ \text{s.t.} & \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) \leq P_{\max} \\ & p_{k,q} \geq 0 \, , \forall \, k = 1, \ldots, K, \, q = 1, \ldots, M \, , \end{cases} \tag{1.45}$$

Since the logarithm of the identity plus an Hermitian positive semidefinite matrix is a matrix-concave function [58], the numerator of the objective function in (1.45) is concave in $\mathbf{P}_k$. Regarding the denominator, the sum of the traces of $\mathbf{P}_k \, \forall k$ is a linear function and consequently Problem can be solved with affordable complexity by means of Dinkelbach's procedure [21] in Algorithm 2.



(a)

(b)

Figure 1.4 In (a) scenario 1: Outdoor street canyon environment, and in (b) scenario 2: Indoor shopping mall environment.

1.7 Numerical results

In the simulation setup, a communication bandwidth of $W = 1$ GHz centered over the carrier frequency $f_c = 28$ GHz is considered. We assume a single-cell doubly Massive MIMO system with $K = 10$ users. We consider two different scenarios: the former is an outdoor scenario, denoted as "scenario 1" and the latter is an indoor scenario, denoted as "scenario 2". In scenario 1, it is considered the downlink communication between one BS connected via wired backhaul to the core network, with K BSs served via wireless backhaul. In scenario 2, the downlink communication between an indoor BS with K robots in a shopping mall is assumed. A qualitative representation of the considered scenarios is reported in Fig. 1.4. The parameters for the generation of the matrix channels are the ones reported in Section 1.3.1 for the "UMi Street Canyon NLOS" in scenario 1 and for "InH Shopping Mall NLOS" in scenario 2. The noise power $\sigma_n^2 = FN_0W$, with $F = 5$ dB the receiver noise figure and $N_0 = -174$ dBm/Hz. A detailed description of the simulation parameters is reported in Table 1.3. The shown results come from an average over 500 independent realizations of users' locations and propagation channels.

Table 1.3 System parameters

Description	**Value**
BS position (scenario 1)	Horizontal: origin of the reference system. Vertical: 10 m
BS position (scenario 2)	Horizontal: origin of the reference system. Vertical: 5 m
Users distribution (scenario 1)	Horizontal: uniform, azimuth in $[-\frac{\pi}{3}, \frac{\pi}{3}]$, fixed distance at 100 m. Vertical: 10 m.
Users distribution (scenario 2)	Horizontal: uniform, azimuth in $[-\frac{\pi}{3}, \frac{\pi}{3}]$ deg, distance in $[5, 100]$ m. Vertical: 1.65 m.
Carrier frequency	$f_c = 28$ GHz
Bandwidth	$W = 1$ GHz
BS antenna array	ULA with $\lambda/2$ spacing
User antennas	ULA with $\lambda/2$ spacing
N_0	-174 dBm/Hz
Noise figure	5 dB

We start considering the uniform power allocation (Uni), i.e, the M-dimensional diagonal matrix containing the power allocation is

$$\mathbf{P}_\ell = \text{diag}\left(\frac{P_T}{KM}, \ldots, \frac{P_T}{KM}\right) \quad \forall \ell = 1, \ldots, K, \tag{1.46}$$

where P_T is the total transmit power. In all the presented results the hybrid and the analog beamformers have been realized using a number of RF chains equal to KM at the transmitter and equal to M at the receiver, i.e., $N_{\text{T}}^{\text{RF}} = KM$ and $N_{\text{R}}^{\text{RF}} = M$.

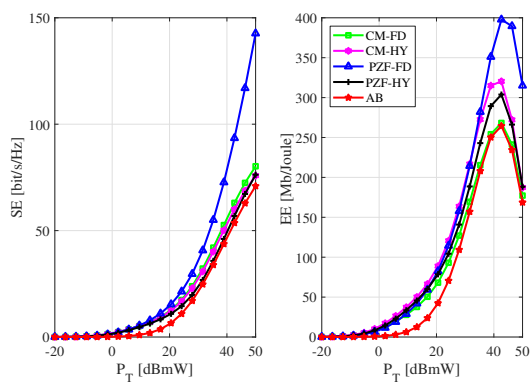


Figure 1.5 SE and EE versus transmit power in scenario 1. Parameters: Uni, $N_T = 64$, $N_R = 64$, $K = 10$ and $M = 4$.

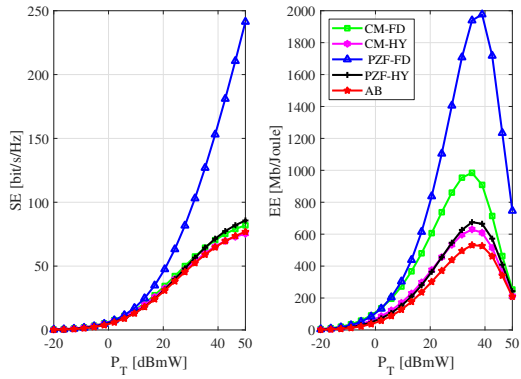


Figure 1.6 SE and EE versus transmit power in scenario 2. Parameters: Uni, $N_T = 64$*,* $N_R = 16$*,* $K = 10$ *and* $M = 4$*.*

Figs. 1.5 and 1.6 report the downlink system SE and EE versus the transmit power, for the case of multiplexing order $M = 4$ and uniform power allocation for the two scenarios in Fig. 1.4. In scenario 1, $N_T = N_R = 64$, a symmetrical communication is considered, while in scenario 2, $N_T = 64$ and $N_R = 16$, an asymmetrical one is assumed. This choices are based on the use cases associated with the considered scenarios: in scenario 1, the transmitter and the receivers are BSs that communicates for wireless backhaul while in scenario 2 the communication is between an indoor BS and mobile users, in particular robots, so it reasonable to assume that the number of antennas at the users is lower with respect to the one at the transmitter. Inspecting the figures, in both scenarios, it is seen that the best performing beamforming structure is the PZF-FD, both in terms of SE and of EE. This means that the increase in system SE given by the PZF-FD beamforming structure, that nulls the interference and gives good performance in terms of simoultaneously transmission, is not well compensated by the reduction of the power consumption in the PZF-HY structure. In particular in Fig. 1.5, we can see that the performance in terms of SE of the CM-FD, CM-HY, PZF-HY and AB are very close, while the performance of the PZF-FD are considerably better; similar results are presented also in [17, 20, 59]. Focusing on the EE performance we can see that the reduction of the power consumption in the hybrid and analog structures gives an increase in terms of EE, while the reduction of power consumption in the PZF-HY structure can not compensate the gap in terms of spectral efficiency with the PZF-FD beamforming. Results show here a trend that has already been found elsewhere (e.g., in [55]); in particular, while the SE grows with the transmit power (at least in the considered range of values), the EE exhibits instead a maximum around 40 dBmW in both scenarios. This behavior is explained by the fact that for large values of the transmit power, the numerator in the EE grows at a slower rate than the denominator of the EE, and so the EE itself decreases. From an energy-efficient perspective, increasing the transmit power beyond the EE-optimal point leads to moderate improvements in the system throughput at the price of a much higher increase in the consumed power. Additionally, we can

see that there is a region where both SE and EE grows, similar results can be also obsersevd elsewhere in literature, as an example the reader can be referred to [22].

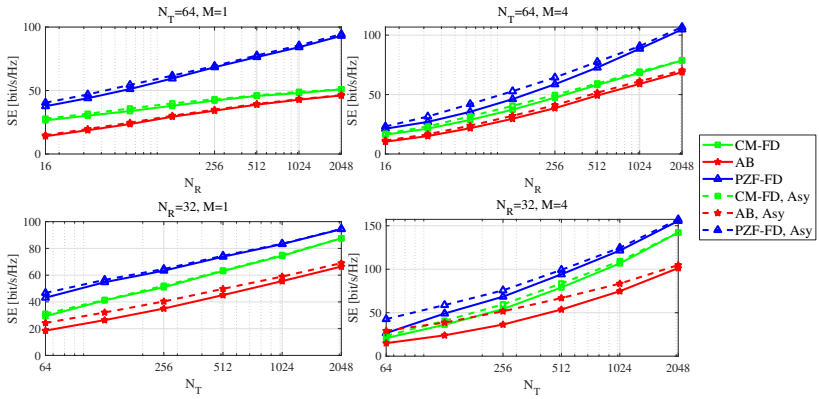


Figure 1.7 SE versus number of antennas in scenario 1, Uni. Validation of the derived asymptotic SE formulas.

Fig. 1.7 is devoted to the validation of the derived asymptotic formulas in the large number of antennas regime for scenario 1 in Fig. 1.4. In particular, "CM-FD, Asy" refers to Eq. (1.24), "PZF-FD, Asy" refers to Eq. (1.27) and finally "AB, Asy" refers to Eq. (1.33) for increasing N_T and to (1.35) for increasing N_R. Similar results are also obtained in scenario 2 and they are here omitted due to the lack of space. Results fully confirm the effectiveness of the found asymptotic formulas, for both the single-stream ad multiple streams cases. Given the effectiveness of the asymptotic formulas in terms of SE, we now consider the performance of the power allocation based on these expressions and detailed in Section 1.6.

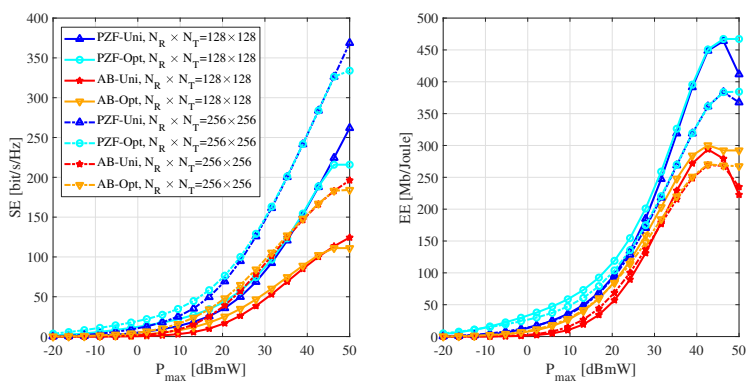


Figure 1.8 SE and EE versus $P_{\max}$, comparison between EE-maximizing and Uni in scenario 1 with $M = 4$ and two antenna configurations.

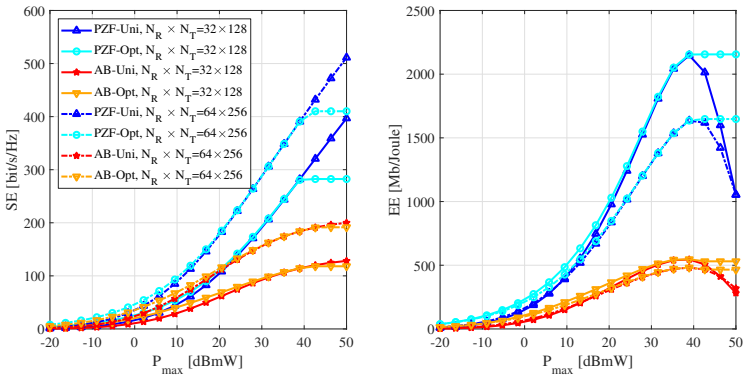


Figure 1.9 SE and EE versus $P_{\max}$, comparison between EE-maximizing and Uni in scenario 2 with $M = 4$ and two antenna configurations.

Figs. 1.8 and 1.9 report the performance of the EE-maximizing power allocation detailed in Section 1.6 for the case of PZF-FD and AB beamforming. We report the performance for the two antenna configurations in both the scenarios and we denote by "Opt" the power allocation obtained with the procedure in Section 1.6.1. Inspecting the figures, we can see that for low values of $P_{\max}$ the Uni and the EE-maximizing power allocation provide similar performance, both in terms of SE and EE, since the radiated power consumption and also the cochannel interference is small compared to the noise power. For larger values of $P_{\max}$, around 40 dBmW in all the shown results, instead, the two power allocations lead to different performance. In this regime, the Uni increases the SE at the price of heavy degradation in the system energy efficiency. On the other hand, the EE-maximizing power allocation exhibits a floor as $P_{\max}$ increases, since it does not use the excess available power to further increase the rate.

1.8 Conclusions

This chapter has focused on the analysis of doubly Massive MIMO mm-wave systems and it has presented some relevant use cases, focusing on energy-efficiency issues. In particular a detailed description of doubly Massive MIMO system has been given highlighting the differences with respect to Massive MIMO at microwave. Some examples of use cases of doubly Massive MIMO systems have been detailed. Asymptotic formulas for the large number of antennas regime and low-complexity EE-maximizing power allocation strategies have been derived. In particular two different techniques are described, the first one used the asymptotic expressions in the interference-free case, the second one used the expressions with interference. In both cases the EE-maximizing power allocation is obtained by means of fractional programming techniques. The obtained results have revealed that, using some of the most recent available data on the energy consumption of transceiver components, fully digital architectures, especially in the multiple-streams transmission, are su-

perior not only in terms of achievable rate, but also in terms of energy efficiency. In particular, among fully digital implementations, the PZF architecture has been shown to provide the best performance, while the AB structure can be considered for its extremely low complexity. Of course the provided results and the relative ranking among the considered structures in terms of energy efficiency is likely to change in the future as technology progresses and devices with reduced power consumption appear on the scene, even though it may be expected that in the long run fully digital architectures will be fully competitive, in terms of hardware complexity and energy consumption [17, 20, 25], with hybrid alternatives.

Bibliography

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] 3GPP, "Study on new radio (NR) access technology physical layer aspects," 3GPP, Tech. Rep. TR. 38.802, Mar. 2017.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] 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, May 2013.

[5] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[6] S. Buzzi and C. D'Andrea, "Doubly massive mmwave MIMO systems: Using very large antenna arrays at both transmitter and receiver," in *Proc. of 2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–6.

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

[8] X. Ge, R. Zi, H. Wang, J. Zhang, and M. Jo, "Multi-user massive MIMO communication systems based on irregular antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5287–5301, Aug. 2016.

[9] S. Buzzi, "Doubly-massive MIMO systems at mmWave frequencies: Opportunities and research challenges," in *IEEE WCNC'2016 Workshop on Green and Sustainable 5G Wireless Networks*, Doha, Qatar, Apr. 2016, keynote talk.

[10] S.Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[11] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[12] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, Jan. 2016.

[13] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmwave: Phase shifters or switches?" in *2015 Information Theory and Applications Workshop (ITA)*, Feb. 2015, pp. 90–97.

[14] A. Alkhateeb, Y.-H. Nam, J. Zhang, and R. W. Heath, "Massive MIMO combining with switches," *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 232–235, Jan. 2016.

[15] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247 – 267, Mar. 2016.

[16] R. Zi, X. Ge, J. Thompson, C. X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 758–771, Apr. 2016.

[17] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmwave systems," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 385–396, Jun. 2018.

[18] S. Dutta, C. N. Barati, A. Dhananjay, D. A. Ramirez, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmwave," *arXiv preprint arXiv:1901.08693*, Jan. 2019.

[19] O. Orhan, H. Nikopour, J. Nam, N. Naderializadeh, and S. Talwar, "A power efficient fully digital beamforming architecture for mmwave communications," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, Apr. 2019, pp. 1–6.

[20] A. Pizzo and L. Sanguinetti, "Optimal design of energy-efficient millimeter wave hybrid transceivers for wireless backhaul," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017, pp. 1–8.

[21] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[22] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[23] L. Sanguinetti, E. Bjornson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[24] S. Buzzi and C. D'Andrea, "Massive MIMO 5G cellular networks: mm-wave vs. μ-wave frequencies," *ZTE Communications*, vol. 15, no. S1, pp. 41 – 49, Jun. 2017, invited paper.

[25] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmwave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, Apr. 2019.

[26] S. A. Busari, M. A. Khan, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Millimetre-wave massive MIMO for cellular vehicle-to-infrastructure communication," *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 983–990, Feb. 2019.

[27] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in millimeter wave integrated access and backhaul for 5G," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8195–8210, Dec. 2018.

[28] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Towards 6G networks: Use cases and technologies," *arXiv preprint arXiv:1903.12216*, Mar. 2019.

[29] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[30] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Communications Letters*, vol. 6, 2017.

[31] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, Mar. 2019.

[32] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44 596–44 612, Apr. 2019.

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

[34] S. Haghighatshoar and G. Caire, "Enhancing the estimation of mm-Wave large array channels by exploiting spatio-temporal correlation and sparse scattering," in *Proc. of WSA 2016; 20th International ITG Workshop on Smart Antennas*, Mar. 2016, pp. 1–7.

[35] S. Buzzi, C. D'Andrea, T. Foggi, A. Ugolini, and G. Colavolpe, "Single-carrier modulation versus OFDM for millimeter-wave wireless MIMO," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1335–1348, Mar. 2018.

[36] J. Lee, G. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 3326–3331.

[37] K. Venugopal, N. Gonzalez-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 330–333, Jun. 2017.

[38] "5G Channel Model for bands up to 100 GHz," http://www.5gworkshops.com/5GCM.html, 2015.

[39] K. Haneda, J. Zhang, L. Tan *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–7.

[40] S. Buzzi and C. D'Andrea, "On clustered statistical MIMO millimeter wave channel simulation," *arXiv preprint arXiv:1604.00648*, May 2016.

[41] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, Apr. 2016.

[42] F. Negro, S. P. Shenoy, I. Ghauri, and D. T. Slock, "On the MIMO interference channel," *Proc. of Information Theory and Applications Workshop 2010*, pp. 1–9, Feb. 2010.

[43] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3258–3268, May 2018.

[44] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, Feb. 2015.

[45] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2012, vol. 3.

[46] A. V. den Bosch, M. A. F. Borremans, M. S. J. Steyaert, and W. Sansen, "A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 315–323, Mar. 2001.

[47] L. Kong, "Energy-efficient 60GHz phased-array design for multi-Gb/s communication systems," Ph.D. dissertation, EEECS Department University of California, Berkeley, Dec. 2014.

[48] Y.-Y. Lee and C.-H. W. . Y.-H. Huang, "A hybrid RF/Baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 305 – 317, Nov. 2014.

[49] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[50] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimetre-wave MIMO systems," in *Proc. of 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, Jun. 2015, pp. 395–399.

[51] S. J. Wright, "Coordinate descent algorithms," *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, 2015.

[52] Y. Yu, P. G. M. Baltus, A. de Graauw, E. van der Heijden, C. S. Vaucher, and A. H. M. van Roermund, "A 60 GHz phase shifter integrated with LNA and PA in 65 nm CMOS for phased array systems," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1697–1709, Sep. 2010.

[53] D. Y. Lie, J. C. Mayeda, Y. Li, and J. Lopez, "A review of 5G power amplifier design at cm-wave and mm-wave frequencies," *Wireless Communications and Mobile Computing*, vol. 2018, Jul. 2018.

[54] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[55] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 1–14, Jan. 2015.

[56] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, 2004.

[57] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[58] E. Jorswieck and H. Boche, "Majorization and matrix-monotone functions in wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 6, pp. 553–701, 2007.

[59] S. Buzzi and C. D'Andrea, "Are mmwave low-complexity beamforming structures energy-efficient? analysis of the downlink MU-MIMO," in *Proc. of 2016 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.