10
Millimeter-Wave MIMO Transceivers: Theory, Design and Implementation
Akbar M. Sayeed and John Brady

10.1 Introduction

Wireless communication technology is approaching a spectrum crunch with the proliferation of data-hungry applications enabled by broadband mobile wireless devices, such as smartphones, laptops, and tablets \[1, 2, 3, 4\]. On the one hand, the data-hungry applications will soon outgrow the Megabits/sec speeds offered by current networks and require aggregate data transport capability of tens or hundreds of Gigabits/sec (Gbps) \[5, 6, 7, 8\]. On the other hand, the wireless electromagnetic spectrum is essentially limited, due to technological, physical, and regulatory constraints \[5, 9\]. This has led to new approaches for efficient use of the available spectrum, such as advanced interference management techniques \[10, 11\], multi-antenna technology \[12, 13, 14, 15, 16\], cognitive radio \[17, 18, 19\], and the current industry approach of small-cell technology to maximize the spatial reuse of the limited spectrum \[4, 20\]. However, despite these advances there is a growing realization that current wireless networks, operating at frequencies below 5GHz, will not be able to meet the growing bandwidth requirements – there simply is not enough physical spectrum \[5, 6, 21, 7, 9, 8\].

Two technological trends offer new synergistic opportunities for meeting the exploding bandwidth requirements in 5G wireless. First, millimeter-wave (mmW) operating in the 30-300GHz band offer orders of magnitude larger chunks of physical spectrum \[22, 23, 5, 6, 21, 24, 9\]. The second trend involves multi-antenna MIMO (multiple input, multiple output) transceivers that exploit the spatial dimension to significantly enhance the capacity and reliability compared to traditional single-antenna systems \[13, 12, 14, 15, 16\]. MIMO systems represent a particularly promising opportunity at mmW frequencies: high-dimensional MIMO operation \[25, 26, 27\] is possible with physically compact antennas due to small wavelengths. The large number of MIMO degrees of freedom can be exploited for a number of critical capabilities, including \[5, 6, 28, 25, 29, 26, 30, 31, 32\]:

- higher antenna or beamforming gain for enhanced power efficiency;
- higher spatial multiplexing gain for enhanced spectral efficiency; and
highly directional communication with narrow beams for enhanced spatial reuse.

However, despite the intensive MIMO research in the last fifteen years [12, 13, 33, 14, 15, 34, 35, 16], conventional state-of-the-art approaches [36, 37, 38, 39, 40, 41, 42] fall significantly short of harnessing these opportunities because of their failure to address fundamental performance-complexity tradeoffs inherent to high-dimensional MIMO systems. In particular, the hardware complexity of the mmW beamforming front-end and the software complexity of the back-end digital signal processing challenge the current paradigm and result in significant technology gaps that require a fresh look at the design of the high-dimensional spatial analog-digital interface.

In this chapter, we present a framework for the design, analysis, and practical implementation of a new MIMO transceiver architecture – Continuous Aperture Phased MIMO (CAP-MIMO) – that combines the directivity gain of traditional antennas, the beam-steering capability of phased arrays, and the spatial multiplexing gain of MIMO systems to realize the multi-Gbps capacity potential of mmW technology as well as the unprecedented operational functionality of dynamic multi-beam steering and data multiplexing [25, 29, 26]. It is based on the concept of beamspace MIMO communication – multiplexing data into multiple orthogonal spatial beams – for optimally exploiting the spatial antenna dimension [25, 26, 43, 35]. The CAP-MIMO transceiver leverages beamspace MIMO theory through two key elements for realizing the full potential of mmW technology: i) a lens-based front-end antenna for analog beamforming, and ii) a multi-beam selection architecture that enables joint hardware-software optimization of transceiver complexity. As a result, CAP-MIMO promises significant advantages over competing technologies, including:

- Significant improvements in capacity, power efficiency, and functionality.
- Optimum performance with the lowest transceiver complexity - number of transmit/receive (T/R) modules, and digital signal processing (DSP) complexity.
- Electronic multi-beam steering and data multiplexing (MBDM) capability.

CAP-MIMO also offers a broad application footprint spanning point-to-point (P2P) and point-to-multipoint (P2MP) network operation in Line-of-sight (LoS) and/or multipath (MP) propagation conditions [25, 26, 30, 44, 45].

10.1.1 mmW MIMO Technology: Background and Promise

Advances in device, integrated circuit, and antenna technology are enabling mmW wireless communication [21, 46, 47, 38, 24, 9]. Recently, mmW wireless backhaul systems [11] (20-90GHz) have emerged as a promising alternative to traditional fiber-based solutions for connecting a local enterprise network to the wired backbone. Emerging mmW systems (60GHz) are also being envisioned for delivering multi-Gbps speeds in indoor applications (e.g., HDTV) and at smaller scales [49, 47, 50, 51].

1) For example, Bridgewave Communications (http://www.bridgewave.com) and Siklu Communications (http://www.siklu.com).
Figure 10.1 (a): Antenna gain vs. frequency. (b): Antenna beampatterns for a 6” antenna at 3GHz vs. 30GHz. (c): Potential spectral efficiency gains due to spatial multiplexing at 80GHz vs. 3GHz with a 6” antenna.

The current trend for increasing the spectral efficiency of cellular wireless (below 5GHz) is to increase spatial reuse of spectrum through smaller cell sizes. However, this approach is inherently limited by the available spectrum and introduces new challenges, such as making the inter-cell interference more pronounced. Fig. 10.1 illustrates the potential of beamspace CAP-MIMO transceivers for providing a strong complement to small-cell technology by exploiting the large number of spatial degrees of freedom at mmW frequencies. For a given antenna size, the large antenna directivity gains shown in Fig. 10.1(a) more than offset the increased path loss and atmospheric absorption at mmW frequencies compared to lower frequencies. The extremely narrow beamwidths at mmW frequencies, illustrated in Fig. 10.1(b), offer dramatically enhanced spatial reuse through dense spatial multiplexing: the spectrum resources can be reused across distinct beams. The resulting spatial multiplexing gain promises unprecedented gains in spectral efficiency illustrated in Fig. 10.1(c) that shows idealized spectral efficiency (bits/s/Hz) upper bounds for downlink communication from an access point (AP) with a 6”x6” antenna: while at 3GHz a maximum of 9 users can be spatially multiplexed, at 80GHz orders-of-magnitude improvements are possible in signal-to-noise ratio (SNR) due to antenna gain, and in spectral efficiency due to spatial multiplexing gain. These gains in power and spectral efficiency, coupled with larger available bandwidths, promise very large gains in overall network throughput. Indeed, 100-10,000 Gbps aggregate rates, over 100-300 users, seem attainable with 1-10GHz of available mmW bandwidth. In contrast, much lower aggregates rates of 30-50Gbps, over a maximum of 9 users, are possible with 300-500MHz of bandwidth at 3GHz. Furthermore, the required total transmit SNR at 3GHz is about 30dB higher than that at 80GHz.

Unleashing this potential of mmW technology for 5G wireless requires the critical functionality of electronic multi-beamforming and data multiplexing (MBDM).

2) The TX SNR values in Fig. 10.1(c) ignore propagation/absorption losses, which would be higher at 80GHz. However, for a given antenna size, the significantly higher antenna gains at 80GHz versus 3GHz (Fig. 10.1(a)) more than compensate for such losses. While the actual SNR values will depend on the link characteristics (e.g., the length), the relative SNR differences are valid.
and associated MIMO communication techniques. This, in turn, presents significant challenges in terms of transceiver complexity for conventional MIMO designs due to the high dimension of the spatial signal space. Systems with fewer but widely spaced antennas have been proposed to reduce complexity in P2P links [36][37][38][39][42], but they suffer from significantly lower array gain and grating beams that also increase interference and compromise security [25][29]. Antenna selection is another sub-optimal mechanism for reducing complexity [54][55][56][57]. The CAP-MIMO transceiver architecture optimally reduces complexity through beam selection, can deliver significantly superior performance, and is squarely aimed at addressing the challenges in realizing the potential of mmW technology. The underlying beamspace MIMO theory also provides a unifying framework for comparing and analyzing other state-of-the-art architectures for mmW MIMO, as overviewed in Sec. 10.2. It builds on cutting-edge research on MIMO theory, transceiver design and wireless channel modeling performed in the Wireless Communication and Sensing Laboratory (http://dune.ece.wisc.edu) over the last 10+ years [43][35][58][59][60][61][62][63][64][25][26][65]. These works show the optimality of beamspace MIMO communication and suggest powerful strategies for practical system design.

10.1.2 Organization

The next section provides an overview of three main transceiver architectures for realizing mmW MIMO communication. All three architectures can be designed, analyzed and compared in the common framework of beamspace MIMO communication that is developed in subsequent sections. Sec. 10.3 develops the key ideas in the context of single-user point-to-point (P2P) links. Sec. 10.4 extends the framework to point-to-multipoint (P2MP) network settings. The focus is on the spatial dimension and on one-dimensional antenna arrays. Extensions to 2D arrays and time- and frequency-selective channels are briefly discussed in Sec. 10.5. Overall the beamspace framework developed in this chapter enables the optimization of fundamental performance-complexity tradeoffs inherent to high-dimensional mmW MIMO systems. The CAP-MIMO transceiver architecture suggests a novel practical realization of beamspace MIMO concepts. The basic theory developed in this chapter applies to systems equipped with uniform linear arrays (ULAs), or uniform planar arrays (UPAs), as well as continuous aperture antennas (e.g., the lens antenna used in CAP-MIMO) via the concept of critical spatial sampling.

10.2 Overview of mmW MIMO Transceiver Architectures

MIMO techniques have been the focus of intense research since the mid-1990s to enhance the capacity of wireless communication systems without additional bandwidth or power [12][13][33]. The key MIMO operation for capacity enhancement is spatial multiplexing: transmission of simultaneous data streams to a single receiver in a P2P
link or multiple receivers in a P2MP configuration. In contrast to the assumption of
rich multipath in traditional MIMO [12, 13, 33, 43], the highly directional nature of
propagation at mmW frequencies makes LoS (and sparse multipath) propagation the
dominant mode of communication [5, 44]. Thus, the beamspace approach to optimal
MIMO communication, first proposed in [43] for multipath channels, is particularly
relevant at mmW frequencies [25, 26]. The following observations and insights un-
derlie the CAP-MIMO transceiver architecture for approaching optimal performance
and MBDM functionality with the lowest transceiver complexity [25, 29, 26, 43]:

- **Optimality of Beamspace Communication**: Orthogonal spatial beams serve as
  approximate channel eigen-modes for optimal spatial communication.

- **High Dimensional Signal Space**: For a given antenna size, the dimension of the
  signal space, \( n \), increases quadratically with frequency; e.g., for a 6”x6” antenna,
  \( n \approx 9 \) at 3GHz and \( n \approx 6000 \) at 80GHz. The antenna gain and the maximum
  multiplexing gain are proportional to \( n \); see Fig. 10.1. This leads to an extremely
  high, \( \mathcal{O}(n) \), transceiver complexity in conventional MIMO; see Fig. 10.2(a).

- **Low Dimensional Communication Subspace**: The multiplexing gain, \( p \), or the
  number of spatial data streams, is typically much smaller than \( n \). In P2P links, this
  is due to the expected channel sparsity in beamspace [5, 28, 44]. In P2MP links,
  \( p \approx K \), the number of mobile stations (MSs) [5, 30].

- **Analog Beamforming and Beam Selection**: Enables near-optimal performance
  with the lowest, \( \mathcal{O}(p) \), transceiver complexity by providing direct beamspace ac-
  cess to the \( p \) channel modes through the beam selector network; see Fig. 10.2(c).

Fig. 10.2 shows three main mmW MIMO transceiver architectures. Given the narrow
beams and highly directional nature of propagation at mmW frequencies, all
architectures are based on beamspace MIMO communication. The differences are
in the hardware implementation. Each transceiver has four components: i) digital
signal processor (DSP), ii) beam selector, iii) transceiver hardware consisting
of multiple Analog-to-Digital-Convertor (ADC) and Digital-to-Analog-Convertor
(DAC) modules and T/R (transmit/receive) chains, and iv) beamforming mecha-
nism. Each T/R chain includes mixers, filters, and amplifiers.

A Conventional MIMO Transceiver, shown in Fig. 10.2(a), uses an \( n \)-element ar-
ray with half-wavelength spaced antennas and employs baseband digital beamform-
ing to map the \( p \) data streams into the \( n \) antenna signals. Each antenna element is
driven by a dedicated ADC/DAC module and a T/R chain. The conventional MIMO
transceiver has full flexibility but suffers from a prohibitively high \( \mathcal{O}(n) \) complexity
of the front-end hardware (T/R chains and ADCs/DACs), regardless of the number
of data streams \( p \), as well as a corresponding \( \mathcal{O}(n) \) computational DSP complexity.

A Phased-array-based Transceiver, illustrated in Fig. 10.2(b), has been proposed
for reducing the complexity of the conventional architecture [48, 31, 66]. In this
architecture, each data stream is associated with a network of \( n \) mmW phase shifters
to map it into a particular beam direction. Thus, the mapping of the \( p \) data streams
is accomplished in passband via the overall network of \( np \) phase-shifting elements.
Compared to the conventional architecture, the number of ADC/DAC modules and
T/R chains can be reduced from \( n \) to \( p \) via beam selection. But it still suffers from a
Figure 10.2 Three main mmW MIMO transceiver architectures. (a) A **conventional MIMO transceiver** that uses an $n$-element antenna array and baseband **digital beamforming**. (b) A **phased array-based transceiver** that uses a network of $n$ mmW phase shifters, one for each of the $p$ data streams, to drive the $n$-element antenna array for generating $p$ beams. (c) The **CAP-MIMO transceiver** that uses a lens antenna for **analog beamforming** to directly map the $p$ data streams to $O(p)$ beams via the **mmW beam selector**.

The high complexity of the $np$-element mmW phase shifting network, especially as $p$ gets larger. The design of this phase-shifting network becomes even more challenging when fully utilizing the wider bandwidths (1-10GHz) available in the mmW band.

**A CAP-MIMO Transceiver**, shown in Fig. 10.2(c), uses a **continuous aperture lens-based front-end** for mmW **analog beamforming**. Unlike the other two architectures, CAP-MIMO samples the spatial dimension in beamspace via an array of feed antennas arranged on the focal surface of the lens antenna. With a properly designed front-end, different feed antennas excite orthogonal spatial beams that span the coverage area. The number of ADC/DAC modules and T/R chains tracks the number of data streams $p$, as in the phased array-based transceiver. However, the mapping of the $p$ data streams into corresponding beams is accomplished via the **mmW beam selector** that maps the mmW signal for a particular data stream into feed antenna(s) representing the corresponding beam. A **wideband lens antenna** can be designed in a number of efficient ways, including a **discrete lens array** (DLA) for lower frequencies or a dielectric lens at higher frequencies [26].

### 10.3 Point-to-Point Single User Systems

This section develops the beamspace MIMO system model, including channel models for line of sight (LoS) and multipath propagation environments, in a single-user
point-to-point setting. The design of optimum and low-complexity beamspace MIMO transceivers is also discussed along with numerical results on their performance. For simplicity, we consider systems with 1D ULAs.

### 10.3.1 Sampled MIMO System Representation

Consider a linear antenna of length $L$. If the aperture is sampled with critical spacing, $d = \frac{\lambda}{2}$ where $\lambda$ is the wavelength, there is no loss of information and the sampled points on the aperture are equivalent to an $n$-dimensional ULA, where $n = \lfloor \frac{2L}{\lambda} \rfloor$ is the maximum number of spatial modes supported by the ULA [25, 67]. A MIMO system with ULAs at the transmitter and the receiver can be modeled as

$$ r = Hx + w $$  \hspace{1cm} (10.1)

where $H$ is the $n_R \times n_T$ aperture domain channel matrix representing coupling between the transmitter and receiver ULA elements, $x$ is the $n_T$-dimensional transmitted signal vector, $r$ is the $n_R$-dimensional received signal vector, and $w \sim \mathcal{N}(0, I)$ represents the Gaussian noise vector.

### 10.3.2 Beamspace MIMO System Representation

Beamspace MIMO system representation is obtained from Eq. (10.1) via fixed beamforming at the transmitter and the receiver. Each column of the beamforming matrix, $U_n$, is a steering/response vector for a specified angle [43, 25, 35]. For a critically spaced ULA, a plane wave in the direction of angle $\phi \in [-\pi/2, \pi/2]$ (see Fig. 10.3) corresponds to a spatial frequency, $\theta \in [-1/2, 1/2]$, given by

$$ \theta = \frac{d}{\lambda} \sin(\phi) = 0.5 \sin(\phi) , $$  \hspace{1cm} (10.2)

and the corresponding array steering/response (column) vector is given by [43, 25]

$$ a_n(\theta) = \left[ e^{-j2\pi \theta i} \right]_{i \in \mathcal{I}(n)} ; \mathcal{I}(n) = \left\{ i - \frac{(n - 1)}{2} : i = 0, \cdots, n-1 \right\} $$  \hspace{1cm} (10.3)

where $\mathcal{I}(n)$ is a symmetric set of indices, centered around 0, for a given $n$. The columns of beamforming matrix $U_n$ correspond to $n$ fixed spatial frequencies/angles with uniform spacing $\Delta \theta_o = \frac{1}{n}$:

$$ U_n = \frac{1}{\sqrt{n}} \left[ a_n(\Delta \theta_o i) \right]_{i \in \mathcal{I}(n)} , \Delta \theta_o = \frac{1}{n} = \frac{\lambda}{2L} , $$  \hspace{1cm} (10.4)

which represent $n$ orthogonal beams that cover the entire spatial horizon ($-\pi/2 \leq \phi \leq \pi/2$) and form a basis for the $n$-dimensional spatial signal space [43, 25].
In fact, $U_n$ is a unitary discrete Fourier transform (DFT) matrix: $U_n^H U_n = I$. The beamspace system representation is obtained from Eq. (10.1) as

$$r_b = H_b x_b + w_b, \quad H_b = U_{n_R}^H H U_{n_T}$$

where $x_b = U_{n_T}^H x$, $r_b = U_{n_R}^H r$, and $w_b = U_{n_R}^H w$ are the transmitted, received, and noise signal vectors, respectively, in beamspace. Since $U_{n_T}$ and $U_{n_R}$ are unitary DFT matrices, $H_b$ is a 2D DFT of $H$ and thus a completely equivalent representation of $H$ [43, 25, 35].

10.3.3 Channel Modeling

Due to the highly directional nature of propagation at mmW frequencies, LoS propagation is expected to be the dominant mode, with some additional sparse (single-bounce) multipath components possible in urban environments [5, 32]. For LoS channels, $H$ can be represented in terms of the array response vectors [25, 26]. As illustrated in Fig. 10.3(a), the $n_T$ columns of $H$ can be constructed via the receiver array response vectors corresponding to the spatial frequencies, $\theta_{R,\ell} = 0.5 \sin(\phi_{R,\ell})$, induced by the different transmitter ULA elements [25, 26]:

$$H = [a_{n_R}(\theta_{R,\ell})]_{\ell \in I(n_T)}, \quad \theta_{R,\ell} = \Delta \theta_{ch} \ell \approx \frac{\lambda}{4R} \ell. \quad (10.6)$$

The rank $p_{\text{los}}$ of the LoS channel matrix is typically much smaller than $\min(n_R, n_T)$ if the link length $R$ is large compared to the antenna lengths $L_T$ and $L_R$. Given the above construction of $H$, the rank can be accurately estimated by the number of orthogonal transmit beams that span the receiver aperture [25, 26]:

$$p_{\text{los}} \approx \frac{\Delta \theta_{\text{max},R}}{\Delta \theta_{o,R}} + 1 = \frac{\Delta \theta_{\text{max},T}}{\Delta \theta_{o,T}} + 1 \approx \frac{L_T L_R}{R\lambda} + 1 \quad (10.7)$$

where $\Delta \theta_{\text{max},R} = \max_{\ell} \theta_{R,\ell} - \min_{\ell} \theta_{R,\ell} \approx \frac{n_R \lambda}{4R}$ is the range of spatial frequencies induced by the transmitter elements at the receiver, and similarly $\Delta \theta_{\text{max},T} \approx \frac{n_T \lambda}{4R}$ is the maximum range of spatial frequencies induced by the receiver elements at the transmitter. The orthogonal steering vectors also serve as approximate eigenfunctions of $H$: only an approximately $p_{\text{los}} \times p_{\text{los}}$ sub-matrix of $H_b$ is non-zero and approximately diagonal (see Fig. 10.4(a)).
A multipath channel can be modeled as \[ H = \sum_{i=0}^{N_p} \beta_i \alpha_{nR}(\theta_{R,i}) \alpha_{nT}^H(\theta_{T,i}) \] (10.8)

where \( N_p \) denotes the number of paths and \( \beta_i, \theta_{R,i} \) and \( \theta_{T,i} \) represent the complex amplitude, angle of arrival (AoA), and angle of departure (AoD) of the \( i \)-th path, respectively. The \( i = 0 \) path is the LoS path with \( \beta_0 = 1 \) and \( \theta_{T,0} = \theta_{R,0} = 0 \). For the other paths, \( \beta_i \) can be modeled as \( \beta_i = |\beta_i| \exp(-j\psi_i) \) where \( |\beta_i|^2 \) represents path loss (between -5 and -10dB \[68,32\]), and \( \psi_i \) is uniformly distributed in \([0, 2\pi]\).

Fig. 10.3(b) shows a simple physical model of a sparse multipath channel: the buildings alongside the road create single-bounce multipath propagation paths. We assume that the link length \( R \) is large enough so that \( |\theta_{R,i}| \in [\Delta \theta_{o,R} - \Delta \theta_{o,R}/4, \Delta \theta_{o,R} + \Delta \theta_{o,R}/4] \) and \( |\theta_{T,i}| \in [\Delta \theta_{o,T} - \Delta \theta_{o,T}/4, \Delta \theta_{o,T} + \Delta \theta_{o,T}/4] \), where \( \Delta \theta_{o,T} = 1/n_T \) and \( \Delta \theta_{o,R} = 1/n_R \). This leads to approximately \( p_{mp} = 3 \) orthogonal beams that couple the the transmitter and the receiver, resulting in an \( H \) with approximate rank \( p_{mp} = 3 \); \( H_b \) only has a \( p_{mp} \times p_{mp} \) non-zero sub-matrix that is approximately diagonal (see Fig. 10.4(d)).

10.3.4

**Beam Selection: Low-Dimensional Beamspace MIMO Channel**

A key feature of mm-wave MIMO is that while the system dimension \( n \) is high, the dimension of the communication subspace, \( p \ll n \), is typically much smaller. Efficient access to the communication subspace is critical from a practical viewpoint. We now outline a framework for beamspace MIMO transceiver design for achieving near-optimal performance with complexity that tracks the low dimension of the communication subspace.

Let \( \sigma_c^2 = \text{tr}(H_b^H H_b) = \text{tr}(H^H H) \) denote the channel power. For a given channel realization, the low-dimensional communication subspace is captured by the singular value decomposition (SVD) of \( H \): \( H = U \Lambda V^H \), where \( \Lambda \) is a diagonal matrix of (ordered) singular values: \( \lambda_1 \geq \lambda_2 \cdots \lambda_{\min(n_T,n_R)} \). We define the effective channel rank, \( p_{\text{eff}} \), as the number of singular values that capture most of channel power: \( \sum_{i=1}^{p_{\text{eff}}} \lambda_i^2 \geq \eta \sigma_c^2 \), for some \( \eta \) close to 1 (e.g., 0.8 or 0.9). Optimal communication over the \( p_{\text{eff}} \)-dimensional communication subspace is achieved through the corresponding right and left singular vectors in \( V \) and \( U \) \[25\].

Beamspace MIMO naturally enables access to the low-dimensional communication subspace through the Fourier basis vectors that serve as approximate singular vectors for the spatial channel \[43,25\]. The channel power is concentrated in a low-dimensional sub-matrix, \( H_b \), of \( H_b \) whose entries capture most of the channel power. Let \( \Sigma_{T,b} = H_b^H H_b \) and \( \Sigma_{R,b} = H_b H_b^H \) denote the transmit and
receive beamspace correlation matrices \(\tilde{H}_b\). We define \(\tilde{H}_b\) as
\[
\tilde{H}_b = [H_b(i,j)]_{i,j} \in M_R, j \in M_T
\]
(10.9)
\[
M_T = \{i : \Sigma_{T,b}(i,i) > \gamma_T \max_i \Sigma_{T,b}(i,i)\}
\]
(10.10)
\[
M_R = \{i : \Sigma_{R,b}(i,i) > \gamma_R \max_i \Sigma_{R,b}(i,i)\}
\]
(10.11)
where the thresholds \(\gamma_T, \gamma_R \in (0, 1)\) are chosen so that
\[
\sum_{i \in M_T} \Sigma_{T,b}(i,i) \approx \sum_{i \in M_R} \Sigma_{R,b}(i,i) \geq \eta b \sigma_c^2,
\]
(10.12)
for some \(\eta_b\) close to 1. \(M_T\) and \(M_R\) denote the transmit and receive sparsity masks representing the beams selected for communication. The overall channel sparsity mask is given by
\[
\mathcal{M} = \{(i,j) : i \in M_R, j \in M_T\}.
\]
(10.13)
Finally, define \(p_{\text{eff},T} = |M_T|, p_{\text{eff},R} = |M_R|\) and \(p_{\text{eff},b} = \min(p_{\text{eff},T}, p_{\text{eff},R})\). The low-complexity beamspace MIMO transceivers access the low-dimensional communicationsubspace by selecting the \(p_{\text{eff},T}\) transmit beams in \(M_T\) and \(p_{\text{eff},R}\) receive beams in \(M_R\). For example, in the CAP-MIMO architecture, shown in Fig. 10.2(c), this corresponds to activating the corresponding feed antennas via the beam selector. By choosing the thresholds appropriately, \(p_{\text{eff},b} \approx p_{\text{eff}}\) and because of Eq. (10.12) the resulting performance is near optimum. The above discussion applies to deterministic LoS channels. For random multipath channels, \(M_T\) and \(M_R\) are determined using statistical channel covariance matrices, \(\Sigma_{T,b} = E[H_b^H H_b]\) and \(\Sigma_{R,b} = E[H_b H_b^H]\) with \(\sigma_c^2 = \text{tr}(\Sigma_{T,b}) = \text{tr}(\Sigma_{R,b})\).

### 10.3.5 Optimal Transceiver

The performance benchmark is provided by the SVD transceiver in which independent data streams are communicated over channel singular vectors to eliminate interference. The transmitted signal in Eq. (10.1) is precoded as \(x = V x_e\) and the received signal is transformed as \(r_e = U^H r\) to result in non-interfering eigenchannels: \(r_e = \Lambda x_e + w_e\), where \(w_e \sim \mathcal{N}(0, I)\). The capacity-achieving transmitted signal \(x_e\) consists of independent Gaussian signals: \(x_e \sim \mathcal{N}(0, \Lambda_s)\) with \(\Lambda_s = \text{diag} (\rho_1, \cdots, \rho_n)\) representing the allocation of total transmit power \(\rho\) over different channels, \(\rho = E[\|x\|^2] = \sum_i \rho_i\). For a given channel realization, the conditional link capacity is
\[
C(\rho|H) = \max_{\rho_i : \sum \rho_i = \rho} \sum_{i=1}^n \log (1 + \text{SNR}_i(H))
\]
(10.14)
which represents optimal power allocation via water-filling [69], with \(\text{SNR}_i(H) = \rho_i \lambda_i^2\). The power would be mainly allocated to the \(p_{\text{eff}}\) dominant channels. For stochastic multipath channels, the ergodic capacity is obtained by averaging over the channel statistics: \(C(\rho) = E[C(\rho|H)]\).
10.3.6 Beamspace MIMO Transceivers

The low-dimensional beamspace MIMO (B-MIMO) transceivers operate on the $p_{\text{eff},R} \times p_{\text{eff},T}$ sub-system induced by $\tilde{H}_b$: $\tilde{r}_b = \tilde{H}_b \tilde{x}_b + \tilde{w}_b$. We focus on the class of linear transceivers which use a precoding matrix $G_b$ at the transmitter ($\tilde{x}_b = G_b s_b$) and a filter matrix $F_b$ at the receiver ($y_b = F_b^H \tilde{r}_b$) to yield

$$ y_b = F_b^H \tilde{H}_b G_b s_b + z_b ; \text{ where } z_b = F_b^H \tilde{w}_b \sim \mathcal{CN}(0, F_b^H F_b). \quad (10.15) $$

The optimal low-dimensional transceiver is now determined by the SVD of $\tilde{H}_b$, $\tilde{H}_b = \tilde{U}_b \tilde{\Lambda}_b \tilde{V}_b^H$, and choosing $G_b = \tilde{V}_b$ and $F_b = \tilde{U}_b$ to create $p_{\text{eff},b} \approx p_{\text{eff}}$ non-interfering channels. The optimal transmitted signal $s_b$ is again an independent Gaussian vector and, for a given $\tilde{H}_b$, the system capacity is

$$ C(\rho | \tilde{H}_b) = \max_{p_{\text{eff},b} \leq p_{\text{eff}}} \sum_{i=t}^{p_{\text{eff}}} \log \left( 1 + \text{SNR}_i(\tilde{H}_b) \right), \quad (10.16) $$

with $\text{SNR}_i(\tilde{H}_b) = \rho_i \tilde{\lambda}_i$, and the ergodic capacity $C(\rho) = E[C(\rho | \tilde{H}_b)]$. By increasing $p_{\text{eff},R}$ and $p_{\text{eff},T}$ - that is, by including progressively larger number of dominant beams by choosing $\gamma_R$ and $\gamma_T$ appropriately - the performance of SVD-based B-MIMO transceiver can be made arbitrarily close to the optimal transceiver.

The optimal SVD transceiver (antenna or beam domain) requires channel state information (CSI) at both the transmitter and receiver, which is impractical in many situations. Thus, we present two simple sub-optimal B-MIMO transceivers for $\tilde{H}_b$ that only require CSI at the receiver. For the sub-optimal receivers we assume that $p_{\text{eff},b} = \min(p_{\text{eff},T}, p_{\text{eff},R}) = p_{\text{eff},T} \leq p_{\text{eff},R}$. In both transceivers $G_b = I$ and the transmitted signal $s_b$ in Eq. (10.15) is an independent Gaussian vector with equal power allocation over the $p_{\text{eff},b}$ beams: $s_b \sim \mathcal{CN}(0, \rho I / p_{\text{eff},b})$. The transceivers differ in their choice of $F_b$ for suppressing interference at the receiver. In the matched filter (MF) transceiver, $F_{b,MF} = \tilde{H}_b$ and Eq. (10.15) becomes $y_b = \tilde{H}_b^H \tilde{H}_b s_b + z_b$. Since $\tilde{H}_b^H \tilde{H}_b$ is diagonally dominant, the interference is limited but still present. In the Minimum Mean Squared Error (MMSE) receiver, $F_b$ is chosen to minimize the MSE at the receiver to further suppress interference

$$ F_{b,\text{MMSE}} = \arg \min_{F_b} E \left[ \left\| F_b^H \tilde{r}_b - s_b \right\|^2 \right] \quad (10.17) $$

$$ F_{b,\text{MMSE}} = Q_b \tilde{H}_b^H \left( \tilde{H}_b Q_b \tilde{H}_b^H + I \right)^{-1} \quad (10.18) $$

where $Q_b = E[s_d s_d^H]$ is the covariance matrix of $s_d$ which equals $Q_b = \frac{\rho}{p_{\text{eff},b}} I$ under independent and equal-power signaling. For both sub-optimal transceivers, the conditional capacity is

$$ C(\rho | \tilde{H}_b) = \sum_i \log \left( 1 + \text{SNR}_i(\tilde{H}_b) \right) \quad (10.19) $$
where the interference is treated as noise and the signal-to-interference-and-noise-ratio (SINR) for the $i^{\text{th}}$ data stream is

$$
\text{SINR}_i(\tilde{H}_b) = \frac{\left| f_{b,i}^H \tilde{H}_b g_{b,i} \right|^2 \rho_i}{\sum_{j \neq i} \left| f_{b,i}^H \tilde{H}_b g_{b,j} \right|^2 \rho_i + \left\| f_{b,i} \right\|^2},
$$

(10.20)

where $\rho_i = \rho / p_{eff,b}$, $f_{b,i}$ is the $i^{\text{th}}$ column of $F_b$, $g_{b,i}$ is the $i^{\text{th}}$ column of $G_b$.

For stochastic channels, the ergodic capacity is $C(\rho) = E[C(\rho|\tilde{H}_b)]$.

10.3.7 Numerical Results

We present numerical results for beamspace MIMO transceivers at $f_c = 80$GHz for both LoS and multipath channels. Fig. 10.4(a)-(c) correspond to a LoS channel with antennas of length $L = 0.6m$ and link length $R = 100m$ which yields $n = 326$ and $p_{los} = 2$. Fig. 10.4(a) shows a contour plot of $|H_b|^2$. Using $p_{eff} = 2$ channel singular values yields $\eta = 0.98$. Using $\gamma_T = \gamma_R = 0.1$ results in a 2x2 $\tilde{H}_b$ with $\eta_b = 0.8$. As shown in Fig. 10.4(b), $H_b$ captures the two dominant channel singular values. Fig. 10.4(c) shows the capacity of the different transceivers for the LoS channel. The capacity of the low-dimensional $p \times p$ SVD system based on $\tilde{H}_b$ is nearly identical to that of the $n \times n$ SVD system based on $H_b$. Furthermore, the low-dimensional beamspace MMSE transceiver closely approximates the performance of SVD transceiver by effectively suppressing interference. The interference is not negligible as evident from the performance degradation in the MF transceiver at higher SNRs.

Fig. 10.4(d)-(f) correspond to a multipath channel with $L = 0.15m$ and $R = 200m$, yielding $n = 81$. In addition to the LoS path, there are additional $N_p = 10$ single-bounce multipath components, as illustrated in Fig. 10.3(b), which result in a total of $p_{mp} = 3$, as discussed in Sec. 10.3.3. Fig. 10.4(d) shows a contour plot of $E[|H_b|^2]$ from which three dominant diagonal entries are evident. Three dominant channel singular values result in $\eta = 0.98$. Using $\gamma_T = \gamma_R = 0.1$ yields a 3x3 $\tilde{H}_b$. As shown in Fig. 10.4(e), the eigenvalues of $E[H_b^H \tilde{H}_b]$ are nearly identical to the eigenvalues of $E[\tilde{H}_b^H \tilde{H}_b]$ and capture 95% of the channel power ($\eta_b = 0.95$). Again, despite the dimensionality reduction from $n = 81$ to $p = 3$, the capacities of the low dimensional and full dimensional SVD systems are nearly identical. Furthermore, the low-complexity MMSE transceiver closely approximates the SVD transceiver, while the MF transceiver degrades at high SNR due to residual interference. The results for the multipath channel were obtained by averaging over 1000 channel realizations.
10.4 Point-to-Multipoint Multiuser Systems

We now consider a point-to-multipoint multiuser MIMO (MU-MIMO) link in which an access point (AP) equipped with an $n$-element ULA communicates with $K$ single-antenna MSs. We focus on the more challenging scenario of downlink communication - the uplink problem is well-studied \cite{34} and can be formulated easily along the lines discussed here.

The received signal at the $i$th MS is given by $r_i = h_i^H x + w_i$, where $x$ is the $n$-dimensional transmitted signal, $h_i$ is the $n$-dimensional channel vector, and $w_i \sim \mathcal{N}(0, \sigma^2)$ is additive white Gaussian noise. Stacking the signals for all MSs in a $K$-dimensional vector $r = [r_1, \cdots, r_K]^T$ we get the antenna domain system equation

$$r = H^H x + w, \quad H = [h_1, \cdots, h_K]$$

(10.21)

where $H$ is the $n \times K$ channel matrix that characterizes the system. Our focus is on the design of the linear precoding matrix $G = [g_1, g_2, \cdots, g_K]$ for the transmitted signal, $x = Gs = \sum_{i=1}^{K} g_i s_i$, where $s$ is the $K$ dimensional vector of independent symbols for different MSs. The overall system equation becomes

$$r = H^H Gs + w, \quad E[\|x\|^2] = \text{tr}(G\Lambda_s G^H) \leq \rho$$

(10.22)

where the second equality represents the constraint on total transmit power, $\rho$, and $\Lambda_s = E[ss^H]$ denotes the diagonal correlation matrix of $s$. 
10.4.1 Channel Model

The channel matrix $H$ governs the performance of the MU-MIMO link. Due to the highly directional and quasi-optical nature of propagation at mmW frequencies, LoS propagation is the predominant mode of propagation, with possibly a sparse set of single-bounce multipath components [32, 68]. We assume that LoS paths exist for all MSs. Let $\theta_{k,0}$, $k = 1, \cdots, K$, denote the LoS directions (spatial frequencies) for the $K$ MSs. Then the LoS channel for the $k$th MS is

$$h_k = \beta_{k,0} a_n(\theta_k, 0)$$

where $\{\theta_{k,i}\}$ denote the path angles and $\{\beta_{k,i}\}$ represent the complex path losses associated with the different paths for the $k^{th}$ MS. The amplitudes $|\beta_{k,i}|$ for multipath components are typically 5 to 10dB weaker than the LoS component $|\beta_{k,0}|$ [32]. For numerical results, we focus on purely LoS channels with $\theta_{k,0} = \theta_k$, $|\beta_{k,0}| = 1$, and $\beta_{k,i} = 0$ for $i \neq 0$ for all MSs.

10.4.2 Beamspace System Model

The beamspace MIMO system representation is obtained from Eq. (10.21) via fixed beamforming at the transmitter using the beamforming matrix $U_n$ defined in Eq. (10.4), and by using the beamspace representation of the precoding matrix $G = U_n G_b$ in Eq. (10.22)

$$r = H_b^H G_b s_b + w, \quad H_b = U_n^H H = [h_{b,1}, \cdots, h_{b,K}]$$

where $s_b = s$ represents the beamspace symbol vector, and $G_b$ is the beamspace precoder. $x_b = G_b s_b$ represents the precoded beamspace transmit signal vector. Since $U_n$ is a unitary matrix, the beamspace channel matrix $H_b$ is a completely equivalent representation of $H$.

Figure 10.5 (a) Contour plot of $|H_b^H|^2$ for a ULA with $n = 81$, representing the beamspace channel vectors (rows) for 20 MSs randomly distributed between $\pm 90^\circ$. (b) Illustration of beamspace channel sparsity masks $M_k$ and $M$ for the $H_b$ in (a).
10.4.3

Beam Selection: Low Dimensional Channel

The most important property of $H_b$ is that it has a sparse structure representing the directions of the different MSs, as illustrated in Fig. 10.5(a) for LoS links. The $k^{th}$ column $h_{b,k} = U_n^H h_k$ (the rows in Fig. 10.5(a)) is the beamspace representation of the $k^{th}$ MS channel and has a few dominant entries near the true LoS direction $\theta_k$ of the MS. This sparse nature of the beamspace channel is exploited for designing reduced-complexity beamspace precoders that deliver near-optimal performance through the concept of beam selection.

We define the following sets of beam indices -- channel sparsity masks – that represent the dominant beams selected for transmission at the AP (see Fig. 10.5(b)):

$$M_k = \left\{ i : |h_{b,k}(i)|^2 \geq \gamma_k \max_i |h_{b,k}(i)|^2 \right\}, \quad M = \bigcup_{k=1,\ldots,K} M_k$$ (10.25)

where $M_k$ is the sparsity mask for the $k^{th}$ MS, determined by the threshold $\gamma_k \in (0, 1)$, and $M$ is the overall beamspace sparsity mask representing the beams activated by the AP. This beam selection is equivalent to selecting a subset of $p = |M|$ rows of $H_b$ resulting in the following low-dimensional system equation

$$r = \tilde{H}_b^H \tilde{G}_b s_b + w, \quad \tilde{H}_b = [H_b(\ell,:)]_{\ell \in M}.$$ (10.26)

where $\tilde{H}_b$ is the $p \times K$ beamspace channel matrix corresponding to the selected beams, and $\tilde{G}_b$ is the corresponding $p \times K$ precoder matrix, where $p \leq n$.

For a given $H$, the total multiuser channel power is defined as $\sigma_c^2 = \text{tr}(HH^H) = \text{tr}(H_bH_b^H)$, which under the simple LoS model is $\sigma_c^2 = n \sum_{k=1}^K |\beta_k|^2 = nK$. The beam selection thresholds $\{\gamma_k\}$ can be chosen so that $M_k$ captures a significant fraction $\eta_k$ of the power of $h_{b,k}$ (e.g., $\eta_k \geq 0.9$). This in turn implies that the fraction $\eta$ of the total channel power captured by $\tilde{H}_b$ is at least $\min_{k=1,\ldots,K} \eta_k$.

Conversely, the sparsity masks $M_k$ can be chosen to select the $m$ dominant (strongest) beams for each MS. This choice implicitly defines $\gamma_k$ as the ratio between the power of the $m^{th}$ strongest beam to the power of the strongest beam for the $k^{th}$ user. For the simple LoS channel model this corresponds to selecting the $m$ orthogonal beams closest to the true LoS direction of the MS $\theta_k$. For numerical results, we use a 2-beam mask for complexity reduction (see Fig. 10.5(b)).

The expected value of $\eta$ for the 2-beam mask can be lower bounded as

$$E[\eta] \geq \frac{2}{n} \int_0^{\Delta \theta_o} f_n^2(\delta) + f_n^2(\delta - \Delta \theta_o) d\delta$$ (10.27)

where $f_n(\theta) = \sin(n \pi \theta) / \sin(\pi \theta)$ is the Dirichlet sinc function.
10.4.4
Multiuser Beamspace MIMO Precoders

Generally, achieving true sum capacity requires dirty paper coding, which suffers from high complexity [34]. We thus focus on simple linear precoders. There are three main types of linear MU-MIMO precoders: the matched filter (MF), zero-forcing (ZF), and Wiener filter (WF). For the full-dimensional antenna domain system Eq. (10.22), the three precoders are given by [70, 35, 71]

\[ G = \alpha F = \alpha [f_1, f_2, \cdots, f_K], \quad \alpha = \sqrt{\frac{\rho}{\text{tr}(F\Lambda_s F^H)}} \] (10.28)

\[ F_{MF} = H, \quad F_{ZF} = H(H^H H)^{-1} \] (10.29)

\[ F_{WF} = (HH^H + \zeta I)^{-1}H, \quad \zeta = \frac{\text{tr}(\Sigma_w)}{\rho} = \frac{\sigma^2 K}{\rho} \] (10.30)

where \( \Sigma_w = E[ww^H] \) and the precoder matrix \( G \) is an \( n \times K \) matrix. In beamspace, the equivalent full-dimensional precoder \( G_b \) can be obtained via the above equations by replacing \( H \) with \( H_b \). Similarly, the reduced-complexity B-MIMO precoder matrix \( \tilde{G}_b (p \times K) \) is obtained via Eqs. (10.28)-(10.30) by replacing \( H \) with \( \tilde{H}_b \). As we demonstrate in the numerical results section, the reduced-complexity B-MIMO precoder can deliver the performance of the full-dimensional precoder with a complexity that tracks the number of MSs \( K \). The computational complexity of the full dimensional precoders is driven by the \( n \times K \) matrix \( H \) for the determination of \( n \times K \) \( G \) as evident from Eqs. (10.28)-(10.30). However, the computational complexity of the low-dimensional B-MIMO system is driven by the \( p \times K \) matrix \( \tilde{H}_b \), and is thus significantly lower.

10.4.5
Numerical Results

We present numerical results to assess the sum capacity of the multiuser B-MIMO precoders. Let \( \rho \) denote the total transmit power, which equals the total transmit SNR for \( \sigma^2 = 1 \). We use the following idealistic upper bound for the sum capacity

\[ C_{ub}(\rho, K, n) = K \log_2 \left(1 + \rho \frac{n}{K}\right) \text{bits/s/Hz} \] (10.31)

which corresponds to \( K \) MSs with orthogonal channels (MS directions coincident with a subset of the fixed beams in \( U_n \)). The received SNR associated with each MS is given by \( \rho n / K \) reflecting the \( n \)-fold array beamforming gain of the AP antenna. For the general full-dimensional precoder in Eq. (10.28) we assess the conditional sum capacity for a given channel realization (random MS directions \( \{\theta_k\} \)) as

\[ C(\rho, G|H) = \sum_{i=1}^{K} \log_2 (1 + \text{SINR}_i(\rho, G|H)) \text{ bits/s/Hz} \] (10.32)
where the interference is treated as noise and the signal-to-interference-and-noise (SINR) ratio for the $i^{th}$ user is

$$\text{SINR}_i(\rho, G|H) = \frac{\rho \frac{|\alpha|^2}{K} |h_i^H f_i|^2}{\rho \frac{|\alpha|^2}{K} \sum_{m \neq i} |h_m^H f_i|^2 + \sigma^2}. \quad (10.33)$$

We can use the same relations for assessing the sum capacity of B-MIMO precoders as well by replacing $H$ with $H_b$ (full dimensional) or $\tilde{H}_b$ (low dimensional), and $G$ with $G_b$ or $\tilde{G}_b$. The ergodic sum capacity for a given precoder (determined by $G$) is given by $C(\rho, G) = E[C(\rho, G|H)]$, where the averaging is done over the random MS directions. Note that the precoder $G_b$ changes with the channel realization.

**Figure 10.6** (a) Capacity of three different B-MIMO precoders for downlink communication from an AP ($n = 81$) to $K = 40$ randomly distributed single-antenna MSs, and minimum MS separation of $\Delta \theta_{\text{min}} = 0$. (b) With minimum separation $\Delta \theta_{\text{min}} = \frac{\Delta \theta_o}{4}$.

**Figure 10.7** (a) Capacity of three different B-MIMO precoders for downlink communication from an AP ($n = 81$) to $K = 60$ randomly distributed single-antenna MSs, and minimum MS separation of $\Delta \theta_{\text{min}} = 0$. (b) With minimum separation $\Delta \theta_{\text{min}} = \frac{\Delta \theta_o}{4}$.

Fig. 10.6 and Fig. 10.7 show numerical ergodic sum capacity results for the B-MIMO precoders generated by averaging over 2000 channel realizations for an AP equipped with a ULA of dimension $n = 81$ (linear 6” antenna at 80GHz) communicating with $K = 40$ (Fig. 10.6) or $K = 60$ (Fig. 10.7) single-antenna MSs over LoS links. The idealized upper bound is also included for comparison. A 2-beam mask is used for complexity reduction, which from Eq. (10.27) captures at least 90 percent of
the channel power on average ($E[\eta] \geq 0.9$). The MSs are randomly located over the entire spatial horizon ($-0.5 \leq \theta \leq 0.5$). The curves in Fig. 10.6(a) and Fig. 10.7(a) were generated with no restrictions on the MS LoS directions $\{\theta_k\}$, while the curves in Fig. 10.6(b) and Fig. 10.7(b) were generated with a minimum MS separation of $\Delta \theta_{\text{min}} = \Delta \theta_o/4$.

These plots show that the simplest MF precoder performs well at lower SNR due to the approximate orthogonality of high-dimensional user channels as also noted in [35, 71, 34, 27]. However, there is always interference for finite $n$, resulting in performance loss at higher SNR. The ZF precoder completely eliminates interference, but significantly reduces the received signal power when the interference is high resulting in performance degradation. The WF precoder achieves the best performance in all cases by adapting to the operating SNR (see Eq. (10.30)). Most importantly, the reduced-complexity B-MIMO precoders ($\tilde{G}_b$) are able to closely approximate their full-dimensional counterparts ($G_b$).

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\Delta \theta_{\text{min}}$</th>
<th>Spectral Efficiency (bits/s/Hz)</th>
<th>Aggregate rate (Gbps)</th>
<th>Average per-user rate (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>134</td>
<td>670</td>
<td>33.5</td>
</tr>
<tr>
<td>20</td>
<td>$\Delta \theta_o/4$</td>
<td>159</td>
<td>795</td>
<td>39.8</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>192</td>
<td>960</td>
<td>24</td>
</tr>
<tr>
<td>40</td>
<td>$\Delta \theta_o/4$</td>
<td>243</td>
<td>1215</td>
<td>30.4</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>226</td>
<td>1130</td>
<td>18.8</td>
</tr>
<tr>
<td>60</td>
<td>$\Delta \theta_o/4$</td>
<td>283</td>
<td>1415</td>
<td>23.6</td>
</tr>
</tbody>
</table>

Table 10.1 Performance of the reduced-complexity B-MIMO WF precoders at an SNR of 20 dB with 5GHz of system bandwidth.

![Figure 10.8](a) $E[p]$ (average number of selected beams) using the 2-beam sparsity mask when $n = 81$. (b) Normalized capacity gap between the idealized upper bound and the WF precoder at an SNR of 30 dB.

Table 10.1 summarizes the performance of the reduced-complexity WF precoder when operating at an SNR of 20 dB with 5GHz bandwidth for $K = 20, 40, 60$ MSs and $\Delta \theta_{\text{min}} = 0$ or $\Delta \theta_o/4$. In the best case, the reduced-complexity WF precoder achieves an average per-user rate of 39.8 Gbps ($K = 20, \Delta \theta_{\text{min}} = \Delta \theta_o/4$). Even
in the worst interference case \( K = 60, \Delta \theta_{\text{min}} = 0 \), the reduced-complexity WF precoder supports an average per-user data rate of 18.8 Gbps. The aggregate sum rates range between 670-1415 Gbps with corresponding spectral efficiencies ranging between 134-283 bps/Hz.

The table shows that enforcing a minimum user separation increases the data rate. However this comes at the cost of increased complexity. Fig. 10.8(a) plots \( E[p] \) (average number of selected beams) for the 2-beam mask as a function of \( K \). Enforcing a minimum user separation requires the reduced-complexity precoders to select more beams on average. The maximum number of beams that the 2-beam mask can select (corresponding to the minimum complexity reduction) is \( p_{\text{max}} = \min(2K, n) \). Fig. 10.8(a) shows that for small \( K \), \( E[p] \approx p_{\text{max}} \). However, for larger \( K \), \( E[p] \) is generally much smaller than \( p_{\text{max}} \) with the largest gap when \( K \approx \frac{n}{2} \).

While the reduced-complexity WF precoder has the best performance, there is still residual interference between closely spaced MSs. This results in the idealized upper bound \( C_{\text{ub}} \) overestimating the sum capacity achieved by the system. However, as shown in Fig. 10.8(b), when the interference is limited \( (K \ll n \text{ and/or there is a minimum separation between the MSs}) \) the difference between \( C_{\text{ub}} \) and the sum capacity for the WF precoders is minimal.

The design and analysis of CAP-MIMO APs equipped with uniform planar arrays (UPAs) for small-cell applications is discussed in [45]. In particular, it is shown that using a UPA of dimension 2.3" × 11.5", the spectral efficiency of a CAP-MIMO transceiver with a 4-beam mask and servicing \( K = 100 \) MSs is 1067 bps/Hz at a transmit power of 20dBm. For a system with 5 GHz of bandwidth, this corresponds to an aggregate rate of 5335 Gbps or an average per user rate of about 53 Gbps. Currently, LTE Advanced using \( 8 \times 8 \) MIMO spatial multiplexing can provide a peak downlink rate of 3.3 Gbps over a 100 MHz bandwidth under ideal conditions [72]. Thus, the combination of the 50× increase in bandwidth, the increased array gain, and the dense spatial multiplexing of \( K = 100 \) MSs results in a more than 1000× increase in the aggregate downlink rate. On the other hand, since the 4-beam mask is used for complexity reduction, when using a system with an analog beamforming front-end, such as CAP-MIMO, there is only a 50× increase in transceiver hardware complexity (from 8 to 400 transceiver chains).

10.5 Extensions

For simplicity, we have focused on MIMO systems equipped with 1D ULAs in a frequency non-selective setting. In this section, we briefly discuss extensions to 2D uniform planar arrays (UPAs) and frequency selective channels.

Consider a critically sampled \( n \)-dimensional UPA with \( n = n_{\text{az}} \times n_{\text{el}} \) where \( n_{\text{az}} \) and \( n_{\text{el}} \) represent the number of critical samples in the azimuth and elevation planes. For a UPA, the 2D steering vector can be represented as an \( n \times 1 \) vector given by the
Kronecker product of the steering vectors for the azimuth and elevation angles [45]

\[ a_n(\theta_{az}, \theta_{el}) = a_{n_{az}}(\theta_{az}) \otimes a_{n_{el}}(\theta_{el}) \]  

(10.34)

where \( \theta_{az} \in [-1/2, 1/2] \) and \( \theta_{el} \in [-1/2, 1/2] \) are the azimuth and elevation spatial frequencies, and \( a_n(\theta) \) is the steering vector for a 1D ULA defined in Eq. (10.3). A multipath channel can then be developed using these steering vectors along the lines of Eq. (10.8) for P2P links and along the lines of Eq. (10.23) and Eq. (10.21) for P2MP links. The beamspace system representation is obtained considering \( n \) orthogonal spatial directions, \( n_{az} \) in azimuth and \( n_{el} \) in elevation with orthogonal spacings \( \Delta \theta_{az} = \frac{1}{n_{az}} \) and \( \Delta \theta_{el} = \frac{1}{n_{el}} \). The columns of the beamforming matrix, \( \mathbf{U}_n \), are steering vectors corresponding to \( n \) fixed spatial frequencies in azimuth and elevation [45]:

\[
\mathbf{U}_n = \frac{1}{\sqrt{n}} \left[ a_n(i \Delta \theta_{az}^i, \ell \Delta \theta_{el}^\ell) \right]_{i \in \mathcal{I}(n_{az}), \ell \in \mathcal{I}(n_{el})}
\]  

(10.35)

that represent \( n \) orthogonal beams covering the spatial horizon \( -\frac{\pi}{2} \leq \phi_{az} \leq \frac{\pi}{2}, -\frac{\pi}{2} \leq \phi_{el} \leq \frac{\pi}{2} \) and form a basis for the \( n \)-dimensional spatial signal space. In fact, \( \mathbf{U}_n \) can also be represented as a kronecker product of the beamforming matrices in azimuth and elevation: \( \mathbf{U}_n = \mathbf{U}_{n_{az}} \otimes \mathbf{U}_{n_{el}} ; \mathbf{U}_n^H \mathbf{U}_n = \mathbf{U}_{n_{az}}^H \mathbf{U}_{n_{az}} = \mathbf{I} \).

A time- and frequency-selective channel model can be obtained by including path delays and Doppler shifts [35, 73] into the models Eq. (10.8) and Eq. (10.23). A number of signaling strategies can be developed for the resulting time-varying wideband MIMO channel as discussed in [35]. However, new constraints relevant to mmW channels (e.g., sparsity) need to be incorporated into the models and signaling schemes.

### 10.6 Conclusion

In this chapter we have presented a framework for beamspace MIMO communication for optimizing the performance-complexity tradeoffs inherent to high-dimensional MIMO systems encountered at mmW frequencies. A key insight that drives complexity reduction is that the MIMO channel is expected to exhibit a sparse structure in beamspace due to the predominantly LoS and single-bounce modes of multipath propagation and the high dimension of the spatial signal space. MIMO system representation in the beamspace naturally reveals the channel sparsity. The concept of beamspace sparsity masks – that capture the dominant beams through power thresholding – is introduced to characterize the low-dimensional sparse channel subspace.

We have considered the design and analysis of both P2P single-user links and P2MP multiuser links. In particular, we have focussed on the development of low-complexity transceiver architectures that leverage the channel sparsity masks to design low-dimensional precoding schemes at the transmitter and the corresponding processing strategies at the receiver. By choosing the thresholding parameters of
the sparsity masks appropriately, the performance of the low-complexity beamspace MIMO transceivers can be made to approach the optimal performance arbitrarily closely. This performance-complexity optimization afforded by the beamspace MIMO transceivers is illustrated through representative numerical results.

The beamspace MIMO framework outlined in this chapter applies to the design and analysis of all leading architectures for high-dimensional MIMO systems overviewed in Sec. 10.2 including phased array-based systems and lens-based systems. In particular, the lens-based CAP-MIMO architecture is a natural candidate for realizing the performance-complexity tradeoffs afforded by beamspace MIMO theory. The CAP-MIMO architecture achieves the key operational functionality of electronic multi-beam steering and data multiplexity (MBDM) through the combination of front-end lens antenna, focal surface feed antennas, and the mmW beam selector network. The resulting transceiver architectures enable performance-complexity optimization from both hardware and computational perspectives.

Initial proof-of-concept demonstration of CAP-MIMO has been achieved via a 10GHz prototype link that can support four spatial channels at 1Gbps rate [26,74]. These results provide a compelling demonstration in a fixed P2P link and indicate that the hybrid analog-digital CAP-MIMO transceiver can potentially deliver near-optimal performance and unprecedented operational functionality with a dramatically lower complexity compared to competing mmW MIMO designs.

While the recent developments are promising, much work needs to be done to realize the potential of mmW MIMO in emerging 5G systems. The ideas outlined in this chapter provide a solid foundation for addressing the technical challenges and harnessing the tremendous opportunities offered by emerging mmW MIMO systems.
References

8. Nokia Solutions and Networks (2013), Looking ahead to 5G, *nsn.com/file/28771/nsn-5g-white-paper*
24

Magazine, pp. 104–112.
Signal Processing, 50 (10), 2563–2579.


pp. 345 – 364.


