Favorable Propagation Studies for Massive MIMO Systems

by

Elhamsadat Anarakifirooz

Thesis submitted to the University of Ottawa

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

Faculty of Engineering

University of Ottawa

Abstract

Massive MIMO (multiple-input multiple-output) is a key technology for 5G/6G networks. Its main advantages lies in enhanced spectral/energy efficiency and simplified processing in multi-user scenarios. These benefits are attributed to a phenomenon known as favorable propagation (FP). In this thesis, we study the asymptotic FP (as the number N of antennas increases without bound) for various scenarios and array configurations. In particular, we establish the asymptotic FP for uniform circular/cylindrical arrays under fixed element spacing. To do so, a novel technique is developed based on a Bessel series expansion. The impact of grating lobes (GL) on the asymptotic FP of uniform linear/planar arrays is analyzed. A novel design of non-uniform linear arrays, based on subarray structure, is proposed to eliminate the impact of GL. This approach is robust in the frequency domain and is applicable to wideband systems. The impact of location and phase errors on favorable propagation is studied. It is shown that the asymptotic FP holds for perturbed arrays if and only if it holds for the unperturbed ones, for any i.i.d distribution of finite variance. While errors have negligible asymptotic effect, they significantly affect the rate of convergence with N: it slows down from $1/N^2$ (no errors) to $1/N$ (with errors), so that more antennas are needed in the later case to attain high SINR. Next, we consider non-asymptotic scenario and to reduce the array complexity, minimize the number of antennas subject to SINR constraints. Despite the non-convex nature of resulting optimization problems, globally optimal closed-form solutions are obtained. The number of antennas can be reduced by almost 50% if variable antenna spacing is allowed compared to the fixed spacing of half a wavelength.

In memory of my father, Mohammad

To

my beloved mother, Zahra,

my lovely husband, Ali,

and my sweet little son, Kourosh.

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my supervisor, Professor Sergey Loyka, for his unrelenting support, invaluable guidance, knowledge and constructive suggestions throughout my studies. It was a great privilege and honor to work and study under his guidance.

A heartfelt gratitude and love goes to my mother, Zahra, for providing me with abundant support and continuous encouragement throughout my years of study, and to my late father, Mohammad, for his constant motivation to pursue my PhD, even though never saw this adventure. I am also very grateful to my sisters and brother, for their unwavering support and inspiration.

Last, but not least, I wish to express my infinite appreciation and love to my husband, Ali. Without his endless patience and unfailing support, completing this thesis would not have been possible.

Table of Contents

Abbreviations

5G	**F**ifth **G**eneration
6G	**S**ixth **G**eneration
AF	**A**rray **F**actor
AoA	**A**ngle-**o**f-**A**rrival
ADC	**A**nalog-to-**D**igital **C**onverter
BS	**B**ase **S**tation
C2D	**C**ompact **2**-**D**
CCDF	**C**omplementary **C**umulative **D**istribution **F**unction
CSI	**C**hannel **S**tate **I**nformation
DAC	**D**igital-to **A**nalog **C**onverter
DOF	**D**egree **O**f **F**reedom
DPC	**D**irty-**P**aper **C**oding
FP	**F**avorable **P**ropagation
FR	**F**orbidden **R**egion
GL	**G**rating **L**obe
i.i.d	**I**ndependent and **I**dentically **D**istributed

INR	**I**nterference to **N**oise **R**atio
ISR	**I**nterference to **S**ignal **R**atio
IUI	**I**nter-**U**ser **I**nterference
LA	**L**arge **A**rray
LB	**L**ower **B**ound
LOS	**L**ine-**O**f-**S**ight
MC	**M**onte-**C**arlo
MF	**M**atched **F**ilter
MIMO	**M**ultiple-**I**nput, **M**ultiple-**O**utput
mMIMO	massive **M**ultiple-**I**nput, **M**ultiple-**O**utput
MMSE	**M**inimum **M**ean **S**quare **E**rror
MSE	**M**in **S**quare **E**rror
MF	**M**atched **F**iltering
MRC	**M**aximum **R**atio **C**ombining
MRT	**M**aximum **R**atio **T**ransmission
NLOS	**N**on-**L**ine-**O**f-**S**ight
NULA	**N**on-**U**niform **L**inear **A**rrays
RF	**R**adio **F**requency
RZF	**R**egularized **Z**ero-**F**orcing
SA	**S**ub**A**rray
SE	**S**pectral **E**fficiency
SIC	**S**uccessive **I**nterference **C**ancellation

SINR	**S**ignal-to-**I**nterference-plus-**N**oise **R**atio
SIR	**S**ignal-to-**I**nterference **R**atio
SNR	**S**ignal-to-**N**oise **R**atio
UCA	**U**niform **C**ircular **A**rrays
UCLA	**U**niform **C**yLindrical **A**rrays
UE	**U**ser **E**quipment
ULA	**U**niform **L**inear **A**rrays
UR-LOS	**U**niform **R**andom **L**ine-**O**f-**S**ight
UPA	**U**niform **P**lanar **A**rrays
UR-LOS	**U**niform **R**andom **L**ine-**O**f-**S**ight
VLA	**V**ery **L**arge **A**rray
ZF	**Z**ero-**F**orcing

List of Symbols

$\mathbf{A}$	Bold capitals like $\mathbf{A}$ denote the matrices
$\mathbf{a}$	Bold lower case like $\mathbf{a}$ denote vectors
C	The channel capacity, p. 12
κ	The condition number, p. 17
$\sigma_{\max}$	Maximum singular value of a matrix, p. 17
$\sigma_{\min}$	Minimum singular value of a matrix, p. 17
$y(t)$	The received signal vector at time t, p. 30
$\boldsymbol{h}_i$	The channel vector of User i, p. 30
g_i	The large-scale propagation path loss from Tx-to-Rx for User i, p. 30
$x_i'(t)$	The transmitted signal of User i at time t, p. 30
$x_i(t)$	Scaled version of the transmitted signal of User i at time t, p. 30
M	The number of users in a multiuser system, p. 30
N	The number of receiver antennas, p. 30
$\boldsymbol{\xi}(t)$	Zero-mean white Gaussian circularly-symmetric noise vector at the receiver with variance σ_0^2, p. 30
$\boldsymbol{w}_i$	The beamforming vector used for detecting User i, p. 31

$\alpha_{1i,N}$ IUI power "leakage" factor of User i, p. 31

P_{xi} The transmit power of User i, p. 31

σ_{xi}^2 Scaled version of the transmit power of User i, p. 31

γ_i The single-user Rx SNR for User i, p. 31

α_N The total IUI power "leakage" factor , p. 31

$\boldsymbol{H}^{N\times M}$ The channel matrix of M single-antenna users communicating to an N-element base station, p. 34

R_N The circle radius of a N-element uniform circular array, p. 39

d The antenna spacing in an array of antenna elements, p. 39

ϕ_i The angle of arrival of User i, p. 39

J_m Bessel function of 1st kind and order m, p. 40

$\boldsymbol{h}_{im}$ The channel vector of UCA m for User i in a uniform cylindrical array, p. 47

N_c The number of UCAs located on top of each other in a uniform cylindrical array, p. 47

ψ_{ni} The phase shift of User i received signal at element n with respect to 1st element, p. 54

α_{siN} The IUI leakage factor for a single subarray in a non-uniform linear array, p. 58

α_{bi} The IUI leakage factor for a block array in a non-uniform linear array, p. 58

α_{iN} The overall IUI leakage factor of User i in a non-uniform linear array, p. 58

N_b The number of blocks in a non-uniform linear array, p. 56

D The spacing between the start of two neighboring subarrays in a non-uniform linear array, p. 56

ΔD The subarray spacing in a non-uniform linear array, p. 57

$\boldsymbol{h}_{si}$ The channel vector of User i to subarray in a non-uniform linear array, p. 57

$\boldsymbol{h}_{bi}$ The channel vector of User i to block array in a non-uniform linear array, p. 57

p A positive integer coprime with the number of blocks N_b in a non-uniform linear array, p. 60

p_k A divisor of the number of blocks N_b in a non-uniform linear array, p. 61

$K(\lambda)$ The number of grating lobes in a ULA array pattern at wavelength λ, p. 63

α'_{iN} The inter-user interference leakage term under directional elements, p. 66

$\boldsymbol{p}_n$ The actual location of n-th antenna array element, p. 72

$\boldsymbol{p}_n^0$ The nominal location of n-th antenna array element, p. 72

$\Delta\boldsymbol{p}_n$ The random offset of n-th antenna array element, p. 72

$\boldsymbol{u}_i$ The unit direction vector for User i, p. 72

$\boldsymbol{h}_i^0$ User i nominal channel vector, p. 72

$\Delta\phi_n$ The beamforming phase error of n-th antenna array element, p. 72

Ψ_{in}^0 The phase differences of i-th and 1st users' signals for the n-th element of nominal array, p. 73

$\Delta\psi_{in}$ The uncompensated phase shifts of User i signal at the n-th antenna array element due to element location errors, p. 73

$\alpha^0_{1i,N}$ The IUI leakage factor of the nominal array, p. 73

$\sigma^2_{|\alpha_N|^2}$ The variance of $|\alpha_N|^2$, p. 81

σ^2_p The variance of the antenna location error, p. 82

σ^2_ϕ The variance of the antenna phaseshifts error, p. 82

γ A threshold SINR, p. 94

θ_0 The AoA of the main user, p. 95

$\Delta\theta$ The minimum angular distance of the main user and interfering users, p. 95

α^{up}_{iN} The upper bound of the IUI leakage factor, p. 96

Chapter 1

Introduction

1.1 Motivation

Mobile communication technologies undergo a dramatic technological revolution roughly every ten years. In addition, each generation brings about significant improvements in performance [1]. These swift changes are because of the massive explosion of data demand over the last decades, primarily driven by video, in social media and video-on-demand services, with around a 70 percent share of traffic in 2022 [2], owing to the availability of high resolution videos, the increase in users viewing time and interest in streaming and sharing videos, and the advent of smartphones with 4K video capabilities. This is expected to increase by around 30 percent annually until the end of 2028, when it is forecast to account for 80 percent of global mobile data traffic [1][2]. There appears to be no saturation with this trend, and indeed, the demand for content is expected to grow substantially, outpacing forecasts [1].

In the years ahead, the annual growth in mobile data traffic has been forecast to be 32% from 2020 to 2028 [2]. Thus, it stands to reason that wireless data connectivity will continue to grow in the future, leading to the important question of how the current wireless technologies should evolve in order to meet both the growing demand and the

quality expected from the services offered. It is therefore important to enlist the support of both industrial and academic researchers to design revolutionary new wireless network technologies. Fifth generation (5G) technology promises not only to provide an incredible data rate, but to also make major strides in terms of latency, massive connectivity, network reliability, and energy efficiency [3].

With the dramatic increase in wireless data traffic and emerging applications such as the Internet of Everything, Holographic Telepresence, collaborative robots, and space and deep-sea tourism, current cellular networks (even 5G) cannot keep up with the rising technical demands [4][5]. The limitations are evident in terms of data-rates, latency, reliability, availability, processing, connection density, and global coverage, involving ground-level, underwater, and space aspects. Sixth generation (6G) mobile networks are expected to offer the high technical standards for new spectrum and energy-efficient transmission techniques necessary to meet these requirements [5].

MIMO (multiple-input, multiple-output) technology, widely recognized as a way to considerably improve spectral efficiency (SE), is an innovation in wireless communication systems where both transmitter and receiver are equipped with multiple antennas. Through spatial multiplexing, the technology exploits the spatial dimension, thereby increasing the data rate almost linearly as the number of antennas increases [6][7].

Massive MIMO (mMIMO, also called Large-Scale Antenna Systems) is one of the key enabling technologies of 5G and beyond that has drawn enormous attention in both academic research and the wireless industry in recent years. To demonstrate its increasing acceptance, the significant increase in the number of publications that have included "massive MIMO" technology since 2010 is illustrated in Fig. 1.1. This state-of-the-art technology is a promising candidate for meeting the ever-increasing demand for large wireless throughput. Each base station (BS) in a cellular mMIMO system is equipped with an array with a large number of co-located antennas, and serves a cell with a large number of terminals [7] which typically (and stipulated throughout this study) have a single antenna

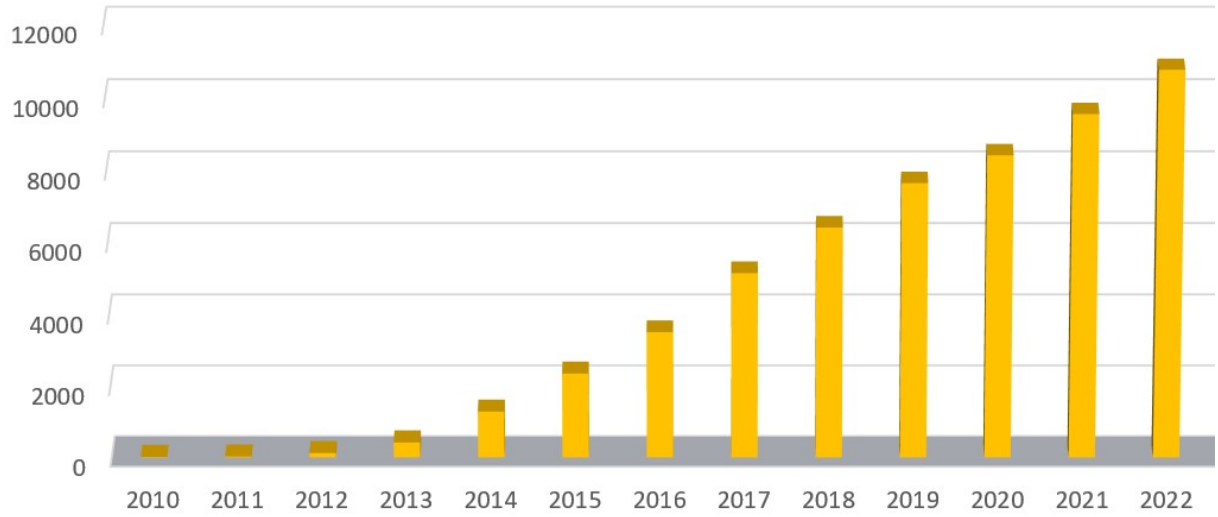


Figure 1.1: Yearly numbers of scholarly publications including the term "massive MIMO", as returned by Google Scholar.

each. Fig. 1.2 illustrates the uplink operation of the Massive MIMO system. M single-antenna users transmit data streams at the same time and over the same frequencies, and the N- element antenna array at the base station receives the sum of the data streams, modified by their respective propagation channels. The decoding unit untangles the received signals to produce the individual data streams [8]. As a key technology for 5/6G and beyond, mmWave/THz systems allow a large number of antennas to be located in a small space due to the extremely small wavelengths involved, which allows for extremely high-gain antennas, but where a line-of-sight (LOS) environment is essential to maintain a proper signal-to-noise-ratio (SNR) (since significant LOS blockage results in large SNR loss, resulting in link outage) [9]. As the number of antennas increases, it is possible to focus energy into ever smaller regions of space, resulting in large increases in energy efficiency and throughput. Other benefits of massive MIMO include the extensive use of inexpensive low-power components, reduced latency, simplification of the MAC layer, and robustness against intentional jamming [10].

Marzetta [11] demonstrated that simple linear processing is nearly optimal when there is a large number of BS antennas relative to the number of users. More specifically, even with simple maximum-ratio combining (MRC) in the uplink or maximum-ratio transmission

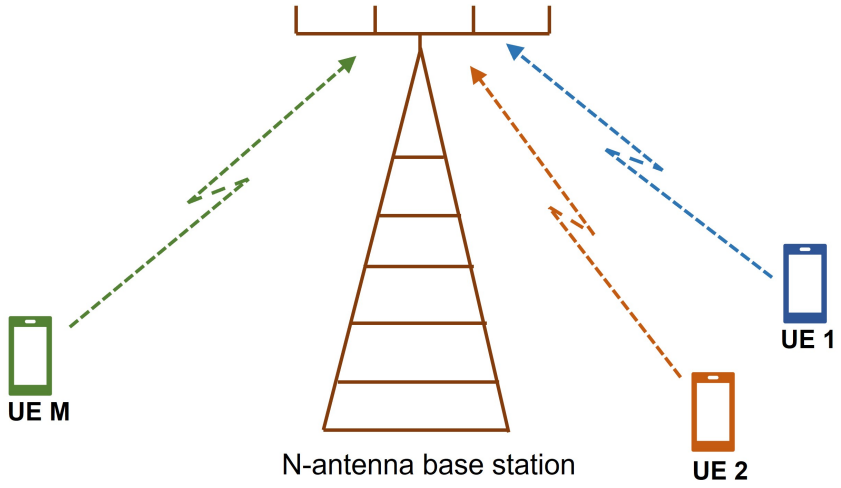


Figure 1.2: Uplink operation of a Massive MIMO link. Users transmit data streams that share the same time/frequency resources, and the antenna array processes the signals to recover the individual streams.

(MRT) in the downlink, small-scale fading and intracell interference gradually disappear as the number of BS antennas increases. Using such a large array of antennas is instrumental in order to obtain high spectral and energy efficiency in a cell through channel hardening and favorable propagation (FP) [7][8].

The significance of channel hardening is that the beamforming generated by a large number of antennas cancels the effects of small-scale fading and frequency dependence, and hence transforms a multi-antenna fading channel into an almost deterministic one [12]. Consequently, resource allocation is simplified, as it is not necessary to adapt the power allocation to the small-scale fading variations [3].

Under favorable propagation, different users' channels become asymptotically orthogonal as the number of antenna elements increases [8], and therefore inter-user interference disappears asymptotically. This property makes it easier for the BS to mitigate interference between these users, thus improving spectral efficiency, and linear processing techniques such as MRC become optimal [8][3].

The study of favorable propagation in LOS channels is of particular importance for the following reasons:

i) As an extreme opposite to independent identically distributed (i.i.d) fading channels, it is particularly challenging for the FP to hold in this environment. This is because the law of large numbers can be exploited in the i.i.d. fading channels, simplifying the analysis considerably, whereas it can not be exploited in the LOS channels, which require a different approach.

ii) There are many LOS application scenarios, including cellular, WiFi, wireless backhaul, short-range, UAV, and satellite systems [13]-[16].

iii) In contrast to below-6 GHz systems, emerging mmWave and THz systems (key technologies for 5/6G, usually employing mMIMO) are critically dependent on the LOS environment. Due to the relatively small wavelengths involved, this permits the co-location of massive numbers of antennas within a small space [1][7][17][18][19].

1.2 Objectives and Scope

The major focus of our study is to explore the favorable propagation for different antenna array geometries. Since linear processing techniques are optimal under the FP, we consider matched filtering to detect the main user signal, and the advantages over other linear techniques will be discussed later in this thesis. Aggregated inter-user interference leakage is used as our main FP performance metric in order to determine whether it diminishes as the number of antennas grows unbounded. While having an infinite number of antennas is impossible in practice, this asymptotic study has an important practical implication: it reflects that one can make the IUI as small as desired by using a sufficiently large number of antennas.

There are a number of gaps in the existing studies on favorable propagation in LOS channels, including whether FP holds for uniform circular and cylindrical arrays with fixed element spacing, the effects of grating lobes, implementation inaccuracies, and analysis under a finite number of antennas. Our objective is to investigate these gaps and provide analytical results along with practical design guides.

1.3 Thesis Contributions

In this thesis, we analyze favorable propagation, as the key condition of massive MIMO, for various array geometries. The main contributions of this thesis are summarized as follows:

- We rigorously demonstrate that FP holds for uniform circular arrays with (any) fixed antenna spacing and under LOS propagation conditions for any finite number of users. The method used to prove the FP for uniform linear and planar arrays is not applicable for UCAs, and hence we propose a new method based on Bessel series expansion. These results are extended to cylindrical arrays, and also for variable numbers of users and their angles-of-arrival [J3] (this reference numbering is from the list in Section 1.5).

- A gap in the existing studies of favorable propagation for uniform linear arrays is identified which is due to the existence of grating lobes (GLs) in the array pattern. Our results demonstrate that the presence of GLs leads to violations of the FP condition even at distinct angles of arrival for different users. Using these results, a novel non-uniform linear array design is proposed that cancels grating lobes and restores favorable propagation. This design is based on a sub-array structure and is consistent with the popular hybrid beamforming paradigm. In addition, we demonstrate that the proposed design is robust in the frequency domain and can be used in wideband or ultra-wideband systems [J2][C3].

- Next, we analyze the FP property under element location and beamforming phase errors, which are due to implementation inaccuracies, component aging, environmental effects, etc., and both Gaussian and non-Gaussian error distributions are considered. Under some mild conditions, the FP property is shown to hold for an arbitrary array geometry under random i.i.d. errors as long as it holds for a nominal array (i.e., one without errors). In other words, even with a large number of antennas, small errors

do not impact the FP. The negative impact of random errors includes slowing down the convergence of IUI to its asymptotic low value. Some practical guidelines are also provided on the degree of accuracy needed to minimize the impact of random errors [J1][C3].

- The non-asymptotic regime (finite number of antenna elements) of massive MIMO (mMIMO) systems with uniform linear arrays is investigated. Our objective is to find the minimum number of antenna elements required to satisfy the design constraints (optimizing the design subject to SINR). This will reduce the complexity of implementation. In spite of the non-convex nature of the constrained optimization problems, a novel analytical approach is proposed, and several globally optimal solutions are obtained. In addition to being robust, the proposed designs do not require precise knowledge of interfering users' directions of arrival, thereby preserving other important properties [C1][C2].

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows:

Chapter 2

In Chapter 2, an overview of the literature on favorable propagation in massive MIMO systems for different channel conditions and antenna geometries is presented. Both theory and measurement results in the literature shows that massive MIMO provides an approximately orthogonal channel in many realistic propagation scenarios and linear processing at the receiver will provide nearly optimal performance. A number of metrics are discussed for quantification of the FP. Various measurement-based and theoretical studies investigating the impact of implementation errors and inaccuracies on MIMO and massive MIMO performance are also reviewed.

Chapter 3

In Chapter 3, a model of a fixed MIMO channel with M independent single-antenna users transmitting simultaneously to an access point is given, on which the analyses in this thesis are based. Using linear processing at the receiver for detection of the main user, signal-to-interference-plus-noise ratio (SINR) is derived as a function of the main user SNR and an aggregated IUI of the interfering users (our primary metric for evaluating the FP). While our model applies directly to the uplink (users-to-BS), the obtained SINR also applies to the downlink (BS-to-users) under some mild conditions. A comparison between various linear processing techniques, including matched filtering, zero-forcing and minimum-mean-square error is presented in terms of computational complexity, their performance under the FP and robustness. The definition of asymptotic FP is also provided.

Chapter 4

A novel approach to demonstrating the favorable propagation for both two and three dimensional uniform circular arrays (UCA) and uniform cylindrical arrays (UCLA) with fixed antenna spacing under LOS propagation conditions is presented in Chapter 4. As opposed to uniform linear/planar arrays, UCA/UCLA arrays have no closed-form expressions for the FP metric, which makes the analysis more challenging. New tools to analyze the FP of these arrays are developed based on the generating function of Bessel coefficients [J3].

Chapter 5

In Chapter 5, a gap in the existing studies of favorable propagation for ULAs is identified. The current studies are based on an implicit assumption that the array pattern does not contain grating lobes (GLs), a phenomenon that typically occurs when antenna spacing exceeds half a wavelength. However, we show that GLs lead to FP not holding in certain directions, and propose a novel design for a subarray-based non-uniform antenna array for

which FP holds, where the subarrays are uniform linear arrays with (any) fixed antenna spacing. Using this design, all GLs are canceled and the FP is guaranteed. A detailed analysis of the required number and spacing of subarrays is also provided [J2][C4].

Chapter 6

In Chapter 6, the robustness of the FP in mMIMO is analyzed by explicitly considering random errors in element locations and beamforming phases. It is shown that, for an arbitrary array geometry, the FP holds for the perturbed array under random i.i.d. errors as long as it holds for the nominal (no-error) array. A negative impact of errors on the FP is to slow down the convergence of the IUI to zero, requiring more antenna elements to achieve the same low level of IUI compared to nominal antenna arrays. The asymptotic distribution of the inter-user interference under random errors and an approximation of its variance are obtained, which can be used for system design [J1][C3].

Chapter 7

Chapter 7 considers uniform linear arrays with finite numbers of elements and optimizes their designs in order to reduce complexity and cost. Our objective is to minimize the number of antenna elements subject to SINR constraints. Due to the non-convex nature of this optimization problem, it cannot be solved with standard convex optimization tools, and hence new tools are needed. Several novel inequalities are developed along with partitioning the feasible set into several subsets and individual optimization is done over each subset in order to overcome the nonconvexity of the problem. With this approach, we are able to achieve globally optimal solutions in spite of the non-convex structure of the underlying optimization problem [C1][C2].

1.5 Publications

The results of this thesis have been presented in:

[J1] E. Anarakifirooz and S. Loyka, "The Robustness of Favorable Propagation in Massive MIMO to Location and Phase Errors", IEEE Transactions on Wireless Communication, doi: 10.1109/TWC.2023.3278490, Apr. 2023.

[J2] E. Anarakifirooz and S. Loyka, "Structural Design of Non-Uniform Linear Arrays for Favorable Propagation in Massive MIMO", IEEE Communications Letters, vol. 27, no. 1, pp. 367-371, Jan. 2023.

[J3] E. Anarakifirooz and S. Loyka, "Favorable Propagation for Massive MIMO With Circular and Cylindrical Antenna Arrays", IEEE Wireless Communications Letters, vol. 11, no. 3, pp. 458-462, Mar. 2022.

[C1] E. Anarakifirooz and S. Loyka, "Globally-Optimal Designs of Uniform Linear Arrays for Multi-User Massive MIMO", IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Toronto, ON, Canada, Sep. 2023.

[C2] E. Anarakifirooz and S. Loyka, "Optimizing Uniform Linear Arrays for Massive MIMO Applications", Biennial Symposium on Communications (BSC), Montreal, QC, Canada, pp. 65-70, Jul. 2023.

[C3] E. Anarakifirooz and S. Loyka, "Robustness of Massive MIMO to Location and Phase Errors", IEEE Information Theory Workshop (IWT), Saint-Malo, France, pp. 440-444, Apr. 2023.

[C4] E. Anarakifirooz and S. Loyka, "Favorable Propagation for Wideband Massive MIMO with Non-Uniform Linear Arrays", 17th Canadian Workshop on Information Theory (CWIT), Ottawa, ON, Canada, pp. 92-96, Jun. 2022.

Chapter 2

Literature review

To put this work into perspective, we provide here an overview of the existing literature on MIMO, Massive MIMO, and one of its key properties, favorable propagation. The main focus of study is on the FP and the theoretical and experimental results for various antenna array geometries and propagation environments.

2.1 From Point-to-Point to Multi-user MIMO

The performance of wireless systems is restricted by the limited capacity of the channel between any two locations. A method to improve the overall performance of wireless networks is to enhance the spectral efficiency (SE)[1], and this can be done through the use of multiple antennas, which is also known as multiple-input, multiple-output (MIMO). MIMO technology has been mainly developed in the last two decades, but its basic idea dates back more than a century; the authors of [20] proposed the use of an antenna array for directional beamforming so that scarce frequency spectrum could be utilized more efficiently for transoceanic communications.

Point-to-point MIMO emerged in 1965 [21], where the capacity of a memoryless Gaussian vector channel with constrained input signal power was obtained. Next, it was re-

[1]Number of bits per second that can be transmitted in a unit of bandwidth [bit/second/Hertz]

discovered in the context of wireless multi-antenna systems in the late 1990s [22]-[24], where both the transmitter and the receiver are equipped with N_t and N_r-element antenna arrays, respectively. A period of extensive research for MIMO system followed in the 2000s [6][25]-[27].

In a MIMO system with the normalized channel matrix $\boldsymbol{H}^{N_t \times N_r}$, the channel capacity C is [22][23]

$$C = \log \det \left(\boldsymbol{I} + \frac{\gamma}{N_t} \boldsymbol{H} \boldsymbol{H}^{+} \right) \tag{2.1}$$

where $\boldsymbol{I}$ is the identity matrix and γ is the SNR at the receiver [22]. At large γ, the capacity scales linearly with the multiplexing gain r and logarithmically with SNR γ [28]:

$$C(\gamma) \approx r \log(\gamma), \quad r = \lim_{\gamma \to \infty} \frac{C(\gamma)}{\log(\gamma)} \tag{2.2}$$

It is demonstrated in [28] that the multiplexing gain equals the channel matrix rank:

$$r = \text{Rank}(\boldsymbol{H}) \leq \min(N_t, N_r) \tag{2.3}$$

and the upper bound in (2.3) is achieved when the channel matrix is full rank. In this case, $C(\gamma) \approx \min(N_t, N_r) \log(\gamma)$, and so the channel capacity scales linearly with $\min(N_t, N_r)$ and logarithmically with SNR. However, even with large arrays at both ends, three factors seriously limit the usefulness of point-to-point MIMO. First, the receiver equipment is complex, as it requires advanced digital processing to separate data streams. Second, the propagation environment must support $\min(N_t, N_r)$ independent streams, which is often not realistic in practice when compact arrays are used [7]. Third, for cellular systems, the channel capacity scales slowly near the cell edge, where SNR tends to be low because of path loss [7].

In multiuser MIMO (MU-MIMO), as the next stage of development, a single transmitter equipped with an array of antennas serves multiple users with the same time-frequency resources [31]-[33]. In this case, the capacity region, which is the set of all simultaneously achievable rates for different users and also shows the fundamental limit of communication

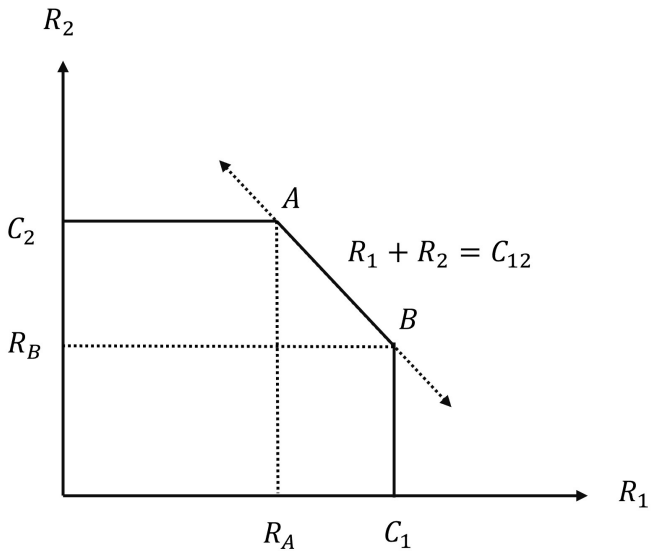


Figure 2.1: The capacity region for two single-antenna users ($i = 1, 2$) that meets three constraints: the rate of individual users R_i is limited by their link's point-to-point capacity C_i and the total throughput $R_1 + R_2$ cannot exceed the sum capacity C_{12} [28].

achievable by any scheme, can be used as the performance metric.

To provide a schematic of the capacity region, let us consider a simple scenario in which two single-antenna users transmit a signal to a multi-antenna base station. The pentagon shown in Fig. 2.1 represents the capacity region for these two users. Note that any rates which are inside of the region are achievable [28], and for User i, the achievable rate R_i is limited by its capacity C_i:

$$C_i = \log(1 + |\boldsymbol{h}_i|^2 \gamma_i) \tag{2.4}$$

where $\boldsymbol{h}_i$ and γ_i are the channel and the SNR of User i, respectively. Points A and B can be approached arbitrarily closely by using successive interference cancellation (SIC)[2] and MMSE[3] processing [28]. For example, if User 1 is detected first, the signal from User 2 is

[2] A capacity optimal decoding technique where after one user is decoded, its signal is subtracted from the aggregate received signal before the next user is decoded [28].

[3] The signal of uncancelled user (User 2) is treated as a colored noise and the MMSE filter is a capacity-optimal demodulator in this case [28, Section 8.3.3]

treated as interference, and so the maximum achievable rate for User 1 is:

$$R_A = \log(1 + \gamma_1 \boldsymbol{h}_1^+ (\boldsymbol{I} + \gamma_2 \boldsymbol{h}_2 \boldsymbol{h}_2^+)^{-1} \boldsymbol{h}_1) \tag{2.5}$$

and since after the detection of User 1, their signal is stripped away from the total received signal, the User 2 channel simplifies to a single-user AWGN channel, and therefore its capacity is C_2. For point B, the detection order changes. There is a trade off region (the border AB in Fig. 2.1) which is attained using time-sharing, and C_{12} is the sum rate capacity [28]:

$$C_{12} = \log\det(\boldsymbol{I} + \sum_i \boldsymbol{h}_i \boldsymbol{h}_i^+ \gamma_i) \tag{2.6}$$

Note that the above results can be readily extended to an arbitrary number of users [28, p. 430]. The capacity region can be used to derive other scalar performance measures, e.g.,

i) The symmetric capacity: the maximum common rate at which all the users can simultaneously and reliably communicate.

ii) The sum capacity: the maximum total throughput that can be achieved.

Considering the uplink channel $\boldsymbol{H}^{MN_t \times N_r}$, where M transmitters each equipped with N_t antennas are sending their signal to a receiver with N_r antennas, the multiplexing gain satisfies $r = \text{Rank}(\boldsymbol{H}) \leq \min(MN_t, N_r)$. For a full-rank channel matrix, if $N_r \geq MN_t$, the multiplexing gain for the MU-MIMO is $r = MN_t$, which is M times larger than that of the point-to-point MIMO ($M = 1$), where $r = N_t$. Hence, multi-user systems have larger overall spectral efficiency compared to a single-user one.

2.2 Massive MIMO

The data rates demand for wireless systems will continue to grow in the future [29], and due to current technology limitations, new and more efficient system designs need to be

developed in order to support these demands. One of the most viable solutions is the deployment of large-scale multiple antenna wireless systems with hundreds of low-power antennas, which offers higher data rates, increased link reliability, and greater energy/spectral efficiency [29][30].

Massive MIMO, which is a scalable version of Multi-user MIMO, was originally proposed in [11][34], where each base station is equipped with a large number of antennas and all users occupy the full time-frequency resources concurrently. There are significant differences between Massive MIMO and conventional Multiuser MIMO. In massive MIMO systems, the number of BS antenna elements is typically much larger than the number of users, and simple linear signal processing is used on both uplinks and downlinks and is near-optimal. This occurs due to a key property of massive MIMO known as "favorable propagation", whereby different users' channel vectors become orthogonal to each other as the number of BS antenna elements increases [7][11]. Under this condition, the pentagon representing the capacity region in Fig. 2.1 becomes a rectangle (the largest possible capacity region under the per-user capacity constraint), $R_i < C_i$, see Fig. 2.2. This implies that every user can achieve their own maximum rate even with the presence of other users, so that favorable propagation results in the best possible capacity region.

To estimate the uplink channel, terminals send a pilot to the BS, with the pilots being mutually orthogonal and unique to each transmitting antenna. Pilot resources are limited, so the same pilot must be reused in several cells. The result is a phenomenon called pilot contamination, which degrades channel estimation quality and causes coherent interference to persist even if the number of antennas is increased. In this case, the interference from these users is increased by a factor N (number of BS antenna elements), under maximum ratio (MR) combining/precoding and i.i.d. Rayleigh fading assumptions, meaning that pilot contamination creates a finite SE limit as N grows unbounded [7][35]. It was argued in [36] that by increasing the number of antennas, the capacity increases without bound, even under pilot contamination, provided that multicell MMSE precoding/combining is used; however, this requires the transmitting power to grow unbounded along with the

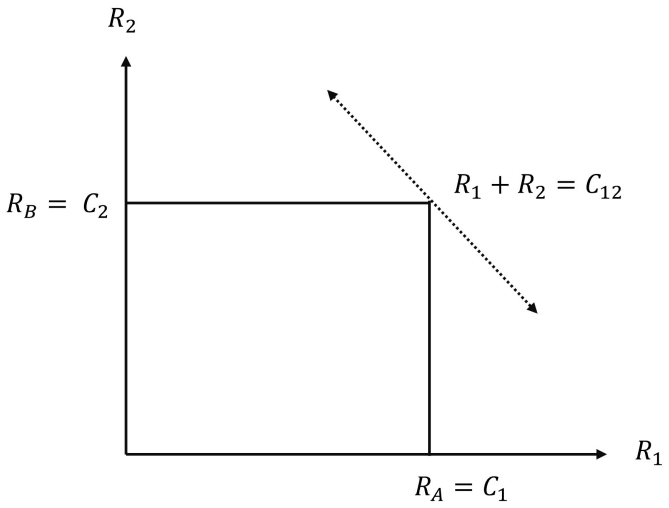


Figure 2.2: The capacity region for two single-antenna users ($i = 1, 2$) under favorable propagation. Note that the capacity region becomes a full rectangle: $R_A = C_1$, $R_B = C_2$ and $C_{12} = C_1 + C_2$, which is its maximum possible value.

number of antennas as well to respect the law of energy conservation.

Millimeter-wave (mmWave) is one of the key technologies of 5G and beyond wireless networks [38]. One possible way to improve capacity is to have a greater bandwidth. It is therefore essential to explore the high-frequency mmWave band ranging from 3 GHz to 300 GHz. However, there are two issues with having large bandwidth in this band. First, larger bandwidth, with a fixed transmitting power, decreases the spectral efficiency because the noise power rises, thereby lowering SNR, and second, the path loss is high at these frequencies [37]. Hence, massive MIMO is critical, as it exploits the beamforming gain to achieve sufficient SNR [37][40]. At mmWave frequencies, the antenna elements dimensions and inter-element spacing become small due to their dependence on wavelength, and therefore a large number of antenna elements can be packed into a limited physical space, thereby enabling massive MIMO antenna arrays at the BSs [10][38].

2.3 Favorable Propagation

Exact favorable propagation holds when the users' channel vectors are orthogonal to each other and, by increasing the number of antennas without bound, asymptotically favorable propagation holds if they become asymptotically orthogonal. This leads to elimination of the inter-user interference, and thereby a substantial simplification in signal processing. In this case, linear precoding and decoding techniques, such as matched filtering (MF) or conjugate beamforming, become nearly optimal [7][8][46]. The FP property has been studied both theoretically [7][46] -[49] and experimentally [50]-[53]. Geometrical configuration of the antenna array, propagation conditions and user locations are three key factors affecting the asymptotic orthogonality of users.

2.3.1 Measures of Favorable Propagation

A channel offers exact FP if different users' channel vectors are pair-wise orthogonal [7], i.e.

$$\boldsymbol{h}_i^{+}\boldsymbol{h}_j = 0, \quad \text{if } i \neq j \tag{2.7}$$

where $\boldsymbol{h}_i$ is the channel vector of User i (see (3.2) for a detailed channel model). However, this condition can hardly be satisfied exactly in practice but only approximately, where it is said the channel offers "approximately favorable propagation", $\boldsymbol{h}_i^{+}\boldsymbol{h}_j \approx 0$. Also, under some assumptions regarding the propagation environment, the channel is said to offer "asymptotically favorable propagation" if $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$ become asymptotically orthogonal [7][46],

$$\frac{1}{N}\boldsymbol{h}_i^{+}\boldsymbol{h}_j \to 0, \quad \text{as } N \to \infty \tag{2.8}$$

which corresponds to the interference to signal ratio (ISR) approaching zero.

Two other measurement metrics, usually used in experimental studies, are presented in

[46], and are the condition number and the distance from FP. The channel condition number is often used as an indication of the degree of mutual orthogonality among users' channel vectors [7][51]. The condition number is defined as $\kappa = \sigma_{\max}/\sigma_{\min}$, where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values of the channel matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_M]$, where M is the number of users. In particular, if $\kappa = 1$, favorable propagation holds[4] [7].

$$\kappa = 1 \quad \Rightarrow \quad \boldsymbol{h}_i^+ \boldsymbol{h}_j = 0, \quad i \neq j \tag{2.9}$$

However, the converse is not true unless all the channels have the same norm (gain),

$$\begin{cases} \boldsymbol{h}_i^+ \boldsymbol{h}_j = 0 \\ |\boldsymbol{h}_i| = |\boldsymbol{h}_1| \end{cases} \quad \Rightarrow \quad \kappa = \frac{\sigma_{\max}}{\sigma_{\min}} = 1 \tag{2.10}$$

Hence, when the channel vectors have the same norm, one can use this metric to measure how favorable the propagation is. While this is a simple method, it has two drawbacks: i) it only has an operational meaning when all the channel vectors have the same norm; otherwise ii) it disregards all singular values other than $\sigma_{\max}$ and $\sigma_{\min}$ [46].

A slightly different approach to measure the channel orthogonality is via the correlation coefficient of any two distinct channel vectors [52]:

$$c_{i,j} = \frac{|\boldsymbol{h}_i^+ \boldsymbol{h}_j|^2}{|\boldsymbol{h}_i|^2 |\boldsymbol{h}_j|^2} \tag{2.11}$$

where $c_{i,j} = 0$ means that $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$ are orthogonal to each other so that there is no inter-user interference. Note that (2.11) is equivalent to (2.8) with a different normalization. The correlation coefficient $c_{i,j}$ is used as a measurement metric in [52] to analyze the FP property for ULA (vertical and horizontal) and UPA antenna arrays; see Fig. 2.3. It is also demonstrated in [52] that in both indoor and outdoor settings, the correlation coefficient becomes smaller as the number of BS antennas grows (this observation is also clear from

[4]The converse is not true: favorable propagation does not imply a condition number of one. This is because users may experience different channel gains.

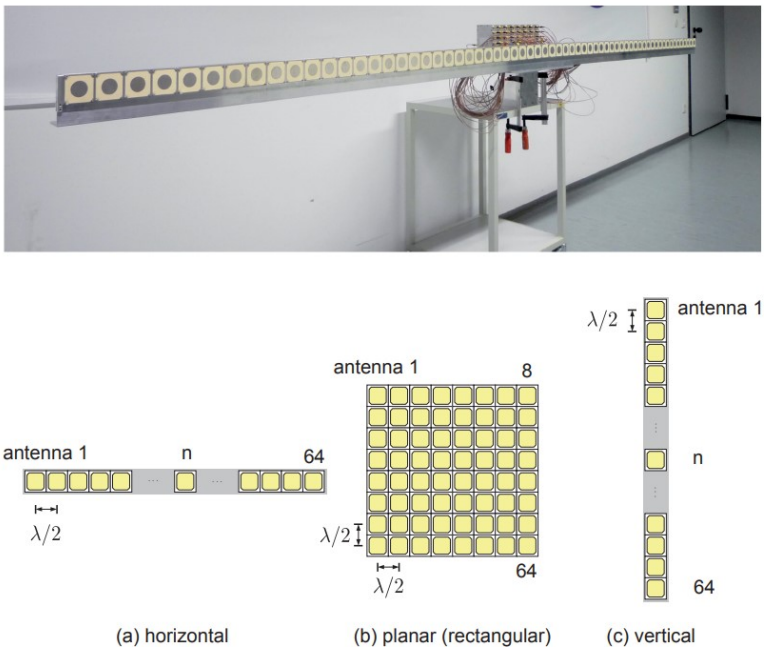


Figure 2.3: Photo of 64-element antenna array in horizontal arrangement (top), and overview of geometries used in the field measurements (bottom) [52].

the measurements in [50]). For example, in the indoor case, $c_{i,j} \approx 0.1$ with 10 BS antennas, and decreases to 0.02 with 64 BS antennas. Based on [52], the horizontal antenna array yields the smallest correlation coefficient value, and the vertical antenna array yields the largest because users are usually located in the horizontal plane, so that a large antenna array is needed with a small beamwidth in that plane to discriminate between users.

Another metric, termed "distance from FP", is based on the channel capacity. It quantifies how far a given channel matrix $\boldsymbol{H}$ is from favorable propagation by measuring the ratio of its actual sum-capacity C_{sum} and its largest possible value C_{max} ($C_{sum} \leq C_{max}$) [7]:

$$\Delta_C \triangleq \frac{C_{sum}}{C_{max}} \tag{2.12}$$

where $\Delta_C = 1$ if favorable propagation holds, since C_{sum} achieves its largest possible value C_{max} ($C_{sum} = C_{max}$) when all the channel vectors are mutually orthogonal, i.e.

$\boldsymbol{H}^{+}\boldsymbol{H}$ is diagonal[5] (equivalent to the exact FP) [7]. Δ_C is used as a performance metric in [54] to evaluate the favorable propagation for a ULA with up to 64 elements under LOS propagation. It is demonstrated that by increasing the number of antennas, the gap between the sum rate capacity and its upper bound decreases, confirming the improvement in channel orthogonality.

Different metrics, including sum capacity, condition number, and angle to interference factor, are used in [55] to quantify the degree of user separability in the spatial domain. A cylindrical array with up to 128 antenna elements is used to serve up to 18 closely located users under a LOS channel. Using the sum rate metric, channel orthogonality is improved by increasing the number of antenna elements from $N = 9$ to $N = 128$ and the SNR loss, which is the horizontal SNR gap between the ideal AWGN channel and the realistic one, decreases substantially by increasing N. Using the condition number metric, [55] shows that adding BS antennas helps in decreasing the channel matrix condition number, thereby improving users' spatial separability. Finally, for a fixed number of users, it is shown that by increasing N, the angular distance between the main user channel and all other interfering users approaches 90°. Hence, all three metrics confirm the N-related channel orthogonality improvement.

2.3.2 Favorable Propagation in Different Channels

It has been shown that by using the law of large numbers, the FP holds in i.i.d. fading channels [7][22][46][47]; however, the i.i.d. fading assumption neglects the impact of antenna array geometry and is justified provided that: (i) the multipath is rich enough (without a single dominant component) and (ii) the antenna spacing is large enough. If either of these conditions is violated, e.g., if there is a dominant LOS component, then the i.i.d. assumption no longer holds.

In [56], the correlation coefficient (see (2.11)) is considered in order to compare the per-

[5] This can be proved using the Hadamard inequality [7]

formance of a massive MIMO system under two extreme propagation models, i.e., involving i.i.d. Rayleigh fading and LOS. It is shown that as the number of antenna elements, N, increases, the correlation between a pair of LOS channel vectors decreases at least as fast as $\log(N)/N$, and in the i.i.d. Rayleigh fading channel, it decreases as $1/\sqrt{N}$. That is, it is shown that for large N, the channel correlation is much smaller for LOS than for the i.i.d. Rayleigh channel.

In fact, the LOS environment situation is the extreme opposite of that related to i.i.d. fading, and is considered to be "particularly difficult" for users' channel vector orthogonality and mMIMO system performance [51]; the law of large numbers is not applicable in this case. Real-world channels are expected to be somewhere in between these two extremes [7][56]. mMIMO in a LOS environment was studied in [7][57], and the FP was shown to hold for uniform linear, planar, circular, and cylindrical arrays (ULA, UPA, UCA and UCLA) with fixed element spacing and distinct angles-of-arrival (AoA) of different users or for uniformly-random users.

Two extreme scenarios, i.i.d. Rayleigh fading and uniform random line-of-sight (UR-LOS), which represent rich and no-scattering environments respectively, are examined in [46], as real-world scenarios are likely to occur somewhere between these two cases. It is demonstrated that both environments offer approximate FP with finite N. Meanwhile, correlated fading, usually seen in practical deployments, has also been investigated in both theoretical and measurement-based studies [48][50].

In [58], evaluation of the pairwise orthogonality of channels is performed in a real environment using cylindrical and planar patch arrays equipped with 128 and 64 antennas, respectively (see Fig. 2.4), and it is shown that while user orthogonality improves by increasing the number of antennas, there is little improvement beyond 20 antennas. A similar result is obtained in [50], where the BS is equipped with a 7-element vertical ULA rotating around a horizontal circle, so that a 112-element "virtual" cylindrical array is formed. To do measurements for $N \leq 112$, N antennas are selected at random, and the

Figure 2.4: A uniform cylindrical array with 128 dual-polarized antennas corresponding to 64 elements (left) and a uniform planar array with 64 dual-polarized antennas corresponding to 32 elements (right) [58].

measured results are averaged over different choices. In this study, the correlation between users' channels is shown to decrease at a rate of $1/N$ until about 10 elements after which it decreases at a much slower rate. These results agree with the analysis performed in [48] for channels with correlated fading.

In [59], it was demonstrated that the FP conditions are satisfied for a majority of correlated Ricean channels (which is a more general propagation model than for i.i.d. fading and LOS); however, it does not hold if the LOS channel vectors are parallel or nearly parallel. This was demonstrated by analyzing the mean value of the inter-user interference power. In these unlikely scenarios, it has been shown that the IUI decreases by using any standard scheduling scheme to drop highly-correlated (parallel/nearly parallel) users, and so the FP is restored [56][59].

The FP for a uniform planar array (UPA) with $N = 32$ to 256 antenna elements and element spacing of half a wavelength ($\lambda/2$) was studied empirically in [60] by measuring the sum rate and comparing it with the capacity of an i.i.d. channel under both LOS and NLOS channel propagation conditions. The measurements were performed for two different scenarios. For the first scenario, a base station with up to 256 antenna elements serves 2 single-antenna users. Increasing the number of BS elements from 8 to 32 increases the sum rate in both LOS/NLOS channels quickly, but then slowly until $N = 64$, after which

it saturates to a certain value. This can be explained by the correlation between users' channel vectors induced by their close positioning and the existence of shared scatterers. The second scenario evaluates the impact of increasing the number of receiver antennas while keeping N fixed at SNR= 20 dB. It is shown that by increasing the number of users' antennas from 2 to 16, the capacity improves from 22.3 to 50 bps/Hz. Both scenarios confirm the improvement in channel orthogonalities by increasing the number of antennas.

2.3.3 Favorable Propagation for Different Antenna Array Geometries

Antenna array geometry is known to have a significant impact on massive MIMO system performance, especially when scattering is not rich enough. This impact was studied in [46], where it was shown that FP holds asymptotically for uniform linear arrays (ULA) with fixed antenna spacing, for both fixed and uniformly-random users' angles-of-arrival (AoA) under line-of-sight (LOS) propagation. It is proved in [48] that FP holds under LOS propagation for ULA and UPA, but not for uniform circular arrays (UCA), but there are two issues with this. First, the comparison is unfair because ULA and UPA are analyzed under fixed antenna spacing (i.e., their array size grows linearly with N), while UCA is analyzed under fixed array size. Secondly, the analysis performed for ULA implicitly assumes there are no grating lobes (GL), and the reported results do not hold if GLs are present in the antenna array pattern. More details relating to this issue can be found in Chapter 5.

In [49], space-constrained massive arrays are studied in which the inter-element spacing decreases as the number of antennas increases. This study focused on the uniform random LOS channels and showed that as the number of antennas increases within a fixed aperture length, the interference leakage factor does not tend to zero, which indicates that there will be inter-user interference, and so maximum ratio transmission is not optimal.

As indicated in [48], asymptotic FP holds for uniform linear and planar arrays with

fixed inter-element spacing; however, whether FP holds for uniform circular and cylindrical arrays (which are important array geometries in providing service in all directions) under fixed antenna spacing was not determined until the present thesis. In some studies, an increase in the number of antennas in circular arrays has been shown experimentally to improve user orthogonality [50][58]. It is important to mention that the method in [48] to prove FP for linear and planar arrays does not apply to UCA with fixed antenna spacing, and therefore, a new approach is needed.

The ability of massive MIMO to distinguish between spatially close users under LOS propagation in a real-world environment is studied in [76]. A cylindrical base station (BS) with 64 dual-polarized antennas installed on a 20m tall building was chosen to serve 8 single-antenna users who were each confined to a 5m diameter circles and asked to walk around slowly and randomly. The singular value spread of the channel matrix was used to measure the degree of orthogonality. The measurements demonstrated that even close users can be separated in a massive MIMO system. The degree of spatial user separability is quantified in [55]. To this end, three scalar metrics: the achievable sum-rates, the condition number of the channel matrix, and the angle to interference factor[6], were used. The measurement results demonstrated that by using ZF or regularized zero-forcing (RZF) precoding, a ratio of three or four antennas per user is required to spatially separate concurrent users. However, this result was established empirically, and with no analysis provided.

2.3.4 Impact of Grating Lobes

Having greater element spacing under a fixed number of elements (i.e., fixed complexity/cost) is desirable, since it increases the array spatial resolution (or, equivalently, decreases its beamwidth), meaning that it is able to distinguish nearby users and hence cancel inter-user interference. However, a major drawback of having a large antenna spacing (half a wavelength or more for beam steering cases) for ULA is the presence of grating lobes

[6] Angle between the main user channel vector and a subspace spanned by the columns of a matrix which contains the interfering user channel vectors

in some directions in the antenna pattern, depending on the antenna spacing (d) [61][62]. It can further shown that FP does not hold if the direction of other users coincides with that of grating lobes. The larger the antenna spacing, the more grating lobes will appear [61], and so more directions should be eliminated for the other users to maintain FP (see Chapter 5 for more details).

A number of traditional antenna array designs (i.e., not massive MIMO) have been proposed to suppress grating lobes partially or completely; see e.g., [63]-[65]. However, these designs do not target FP explicitly and it is far from clear as to whether the FP condition is satisfied with these designs (note that the absence of GLs does *not* guarantee the FP). In particular, it is impossible to verify the FP condition for numerically-based designs, such as in [63][64], and due to the complicated analytical structure and the absence of analytically tractable expressions, it is very difficult, if possible at all, to verify the FP for amplitude-tapering designs such as in [65].

Non-uniform linear array (NULA) antennas are much more appealing in massive MIMO systems (especially in mm Wave systems). Considering the high cost, implementation complexity, and power consumption arising from high numbers of RF chains (analog-to-digital converters (ADC)/digital-to-analog converters (DAC), mixers, etc.) for a massive number of antennas, hybrid subarray implementation is shown to be a beneficial strategy to tackle this problem [44][66]-[68].

2.3.5 The Impact of Aperture Size

Antenna arrays with larger aperture sizes offer better performance in massive MIMO systems [72], since larger apertures have higher angular resolutions, which helps to better separate users spatially. For a fixed number of antennas and element spacing, linear arrays are shown to perform better in separating users compared to cylindrical arrays, as they have larger apertures and so provide better angular resolutions [73][74]. This comparison between linear and cylindrical arrays is, however, misleading because despite having a

lower spatial resolution than linear arrays, cylindrical arrays are more compact and provide full coverage, which is preferred from an antenna design perspective due to certain implementation limitations.

The effect of aperture size is investigated in [53][72][75], where the use of three different 64-element arrays with different shapes and apertures are examined. Due to practical reasons, the arrays were composed of 8-element ULA subarrays which were placed in horizontal lines 2 and 6 meters long, named very large array (VLA) and large array (LA), respectively, as well as vertically to form a square array named compact 2-D (C2D). It is shown in [72] that an array's physical characteristics, including physical size and geometry, are limiting factors in providing angular resolvability, with the largest aperture having the best performance in terms of user discrimination (i.e., inter-user interference is relatively small). Investigations of channel hardening and user orthogonality with the above three antenna arrays were performed in [53], and it was demonstrated that the VLA showed the best performance for channel hardening and C2D the worst, since by increasing the array aperture, the channel vectors become more orthogonal.

Orthogonality of channel vectors is measured in [53] using the sum rate at the receiver. In the LOS scenario, the position of elements, users and the aperture size are shown to be key in determining orthogonality. It is demonstrated that by adding subarrays (SAs), grating lobes appear in the antenna pattern, so the orthogonality depends on the user positions with respect to the main and grating lobes [53][75].

Table. 2.1 provides an illustrative summary of the above measurement-based studies which used different metrics and antenna array geometries to evaluate the FP performance. All the studies confirmed that under LOS/NLOS propagation channels and distinct users' AoAs, the pairwise orthogonality of different users channel vectors improves with increased N.

Table 2.1: Comparison of FP measurement-based studies

Ref.	Array geometry	Performance metric	Propagation channel	Number of antennas
[50]	cylindrical	correlation coefficient/ condition number	LOS/NLOS	112
[52]	planar/linear	correlation coefficient	LOS	64
[54]	linear	Distance to capacity	LOS	64
[55]	cylindrical	condition number/angle to interference factor	LOS	128
[58]	cylindrical/planar	correlation coefficient	LOS/NLOS	128
[60]	planar	condition number	LOS/NLOS	256

2.3.6 Favorable Propagation Under Array Perturbations

The above studies assume perfect channel state information, perfect element location or array calibration, no inaccuracies in beamforming weights etc. In practice, however, such perfect settings are hardly possible, as implementation inaccuracies and tolerances always exist.

In massive MIMO, base stations have many antennas deployed on co-located or distributed arrays. In general, hardware costs and circuit power consumption scale linearly with the number of BS antennas N [77]. Therefore, low-cost hardware and low-power antenna branches are key to deploying large arrays at a reasonable price. So, hardware imperfections are likely to be greater; in particular, non-idealities are carrier-frequency and sampling-rate offset, phase-noise, IQ-imbalance, and amplifier nonlinearities. Low-cost and power-efficient A/D converters yield higher levels of quantization noise. With low-cost phase locked loops or even free-running oscillators at each antenna, phase noise may become a limiting factor [10]. Due to the law of large numbers, massive MIMO averages out noise, fading, and hardware imperfections when signals from a large number of antennas are combined. It is also robust to failure of one or a few of its antenna unit(s)

due the same property that makes it resilient to fading [10][77][78].

The impact of implementation inaccuracies/errors on traditional antenna arrays has been studied, and robust beamforming strategies have been proposed, including the well-known diagonal loading technique and its modifications [61][79]-[81]. While some studies advocate a random error model [61], others use deterministic normed or bounded uncertainty models and worst-case performance optimization [79]-[81]. A general conclusion is that with proper design, small errors do not have a catastrophic impact on performance. Likewise, the impact of channel uncertainty on traditional MIMO system performance has also been studied in the classical setting using information-theoretic tools (compound channel capacity), and it was shown that small normed uncertainties do not have a large impact under proper transmitter and receiver design regimes [82][83]. However, the above results were established in the traditional settings (i.e., non-mMIMO), and it remains unclear whether they still hold in an mMIMO setting and apply to the FP property as well (especially because increasing the number of antennas to very large values has the potential to "amplify" small per-element errors and generate a large aggregate effect, thereby destroying the FP property).

Residual transmit-RF (Tx-RF) impairments, which is the remaining impairments after calibration or predistortion methods, are modeled in [84] as aggregates of all transmitter impairments. While only some of the impairments in the transmission system are reported to have Gaussian distribution [85], the MIMO-OFDM measurement results performed in [84] confirm that the sum of all such residual impairments can be accurately modeled as additive Gaussian noise. It is shown that under this perturbation model, the channel capacity in the presence of Tx-noise is upper-bounded when SNR $\rightarrow \infty$. Meanwhile, the real-world measurements performed in [84] confirmed that the Tx-RF impairments can be accurately represented by the proposed i.i.d. Gaussian Tx-noise model. The introduced additive error model in [84] is also used in [78] to obtain the capacity bounds in the high SNR regime. It is shown that when there are distortions from physical transceiver impairments, the capacity of physical MIMO channels is bounded for any channel distribution

and SNR.

The additive error model in [84] is extended in [86], where it incorporates general transceiver hardware impairments at both BS equipped with large antenna arrays as well as single-antenna user equipment (UE). The Gaussian distributions of impairments were verified experimentally in [87, Fig. 4.13]. A generalized uplink massive MIMO system with arbitrary array configurations is considered in [77]. The proposed system model jointly considers the impact of multiplicative phase-drifts, additive distortion noise, noise amplification, and inter-carrier interference. The user achievable rate using the MRC technique is derived and it is shown that only phase-drifts limit the achievable rates as $N \to \infty$.

Chapter 3

System Model

Let us consider a frequency-flat[1] Gaussian MIMO channel based on the standard baseband discrete-time model (after matched filtering and sampling) described in [7][11][41][46][47][49], where M independent single-antenna users transmit data simultaneously to an N-antenna base station (BS):

$$\boldsymbol{y}(t) = \boldsymbol{h}_1\sqrt{g_1}x_1'(t) + \sum_{i=2}^{M}\boldsymbol{h}_i\sqrt{g_i}x_i'(t) + \boldsymbol{\xi}(t) \tag{3.1}$$

$$= \boldsymbol{h}_1 x_1(t) + \sum_{i=2}^{M}\boldsymbol{h}_i x_i(t) + \boldsymbol{\xi}(t) \tag{3.2}$$

where $\sqrt{g_i}$ models the large-scale propagation path loss from Tx-to-Rx for User i, $\boldsymbol{h}_i = [h_{1i}...h_{Ni}]^T \in \mathbb{C}^{N\times 1}$ models the small-scale variations which are assumed to be fixed and deterministic, $x_i'(t)$ is the transmitted signal of User i and $x_i(t) = \sqrt{g_i}x_i'(t)$ is its scaled version (additional details of this model can be found in [7, pp. 30-32][11][41][47][56]). Meanwhile, $\boldsymbol{y}(t)$ is the received signal vector at discrete time t, $\boldsymbol{\xi}(t)$ is a zero-mean white Gaussian circularly-symmetric noise vector of variance σ_0^2 per dimension, and $x_i(t)$ is the transmitted signal of User i, $i = 1...M$. Further, $|\boldsymbol{h}|$, $\boldsymbol{h}'$, and $\boldsymbol{h}^+$ denote the Euclidean norm (length), transposition and Hermitian conjugation, respectively, of vector $\boldsymbol{h}$. These results can also be extended to frequency-selective channels using an OFDM-type approach.

[1] The delay spread of the channel is much smaller than the symbol duration [28]

To detect the signal $x_1(t)$ of User 1 (the main user), linear beamforming $\boldsymbol{w}_1^+\boldsymbol{y}(t)$ is used by the BS [7], where $\boldsymbol{w}_1$ is the beamforming vector used for detecting User 1

$$\boldsymbol{w}_1^+\boldsymbol{y}(t) = \boldsymbol{w}_1^+\boldsymbol{h}_1 x_1(t) + \boldsymbol{w}_1^+ \sum_{i=2}^{M} \boldsymbol{h}_i x_i(t) + \boldsymbol{w}_1^+ \boldsymbol{\xi}(t) \tag{3.3}$$

$\boldsymbol{w}_1^+\boldsymbol{h}_1 x_1(t)$ is the desired signal and the other users' contribution $\boldsymbol{w}_1^+ \sum_{i=2}^{M} \boldsymbol{h}_i x_i(t)$ is treated as interference. Since $x_1(t), .., x_M(t)$ are independent of each other and also of $\boldsymbol{\xi}(t)$, the main user SINR can be expressed as follows:

$$\text{SINR} = \frac{|\boldsymbol{w}_1^+\boldsymbol{h}_1|^2\sigma_{x_1}^2}{\sum_{i=2}^{M}|\boldsymbol{w}_1^+\boldsymbol{h}_i|^2\sigma_{x_i}^2 + |\boldsymbol{w}_1|^2\sigma_0^2} \tag{3.4}$$

$$= \frac{|\alpha_{11,N}|^2\gamma_1}{\sum_{i=2}^{M}|\alpha_{1i,N}|^2\gamma_i + 1} \tag{3.5}$$

$$\leq |\alpha_{11,N}|^2\gamma_1 \leq \gamma_1 \tag{3.6}$$

where $|\alpha_{1i,N}|^2$, $i = 2...M$ is the IUI power "leakage" factor of User i to the main user:

$$|\alpha_{1i,N}|^2 = \frac{|\boldsymbol{w}_1^+\boldsymbol{h}_i|^2}{|\boldsymbol{w}_1|^2|\boldsymbol{h}_i|^2}, \quad i = 1...M \tag{3.7}$$

In the case of no interference, $\alpha_{1i,N} = 0$. γ_i is the single-user Rx SNR for User i:

$$\gamma_i = \frac{|\boldsymbol{h}_i|^2 g_i P_{xi}}{\sigma_0^2} = \frac{|\boldsymbol{h}_i|^2\sigma_{xi}^2}{\sigma_0^2} \tag{3.8}$$

and $|\alpha_{1i,N}|^2\gamma_i$ is the "leaked" IUI power measured in σ_0^2. $P_{xi} = \mathbb{E}\{|x_i'(t)|^2\}$ is the transmit power of User i, and $\mathbb{E}\{\cdot\}$ is the statistical expectation. $\sigma_{xi}^2 = \mathbb{E}\{|x_i(t)|^2\} = g_i P_{xi}$ is the scaled version of the transmit power at the receiver, which includes the large-scale path loss but not small-scale variations. Hence, the propagation path loss is absorbed into the single-user SNR γ_i. Note that $|\boldsymbol{w}_1^+\boldsymbol{h}_1|^2$ is the beamforming power gain for User 1 (with respect to single isotropic antenna) and therefore $|\alpha_{11,N}|^2$ is the normalized channel power gain. In the case of no power/SNR loss for the main user, $|\alpha_{11,N}|^2 = 1$.

While the model in (3.2) applies directly to the uplink (UL, users-to-BS), the User 1

SINR in (3.4) also applies to the downlink (DL, BS-to-users) if:

i) Equal amount of channel state information is available in both cases.

ii) The channel is reciprocal, i.e., $\boldsymbol{h}_i$ is the same for UL and DL, and MRT/MRC (with the beamforming weigh $\boldsymbol{w}_i = \boldsymbol{h}_i/|\boldsymbol{h}_i|$) is used in both cases.

To see this, note the following. For the uplink, $x_i(t)$ is transmitted by User i and $\boldsymbol{y}(t)$ is received by the BS, where the BS performs the beamforming $\boldsymbol{w}_1^+\boldsymbol{y}(t)$ to detect User 1, as in (3.1) - (3.3). For the downlink, the transmitter first precodes the signal of each user, $x_i(t)$, by the weight $\boldsymbol{w}_i$ (transmit beamforming) and then transmits the composite signal of all the users $\sum_i \boldsymbol{w}_i x_i(t)$. Hence, User 1 receives:

$$y_1(t) = \boldsymbol{h}_1^+\boldsymbol{w}_1 x_1(t) + \sum_i \boldsymbol{h}_1^+\boldsymbol{w}_i x_i(t) + \xi(t) \tag{3.9}$$

and so the downlink SINR of User 1 is:

$$\text{SINR}_{DL} = \frac{|\boldsymbol{h}_1^+\boldsymbol{w}_1|^2\sigma_{x_1}^2}{\sum_{i=2}^M |\boldsymbol{h}_1^+\boldsymbol{w}_i|^2\sigma_{x_i}^2 + \sigma_0^2} \tag{3.10}$$

$$= \frac{\gamma_1}{\sum_{i=2}^M |\alpha_{1i,N}|^2\gamma_i + 1} \tag{3.11}$$

$$= \text{SINR} \tag{3.12}$$

where $\alpha_{1i,N}$ is as defined in (3.7) with the beamforming weight $\boldsymbol{w}_1 = \boldsymbol{h}_1/|\boldsymbol{h}_1|$, and γ_i is the interference to noise ratio (INR) of User i at the User 1 receiver,

$$\gamma_i = \frac{|\boldsymbol{h}_1|^2\sigma_{x_i}^2}{\sigma_0^2} \tag{3.13}$$

Observe that (3.11) equals the UL SINR in (3.4) if all the conditions above hold, which is a reflection of the well-known reciprocity principle in electromagnetics and antenna theory (i.e., antenna pattern is the same for transmission and reception) [7]. As a consequence, under these conditions, if the FP holds for the uplink, it will also hold for the downlink, and vice-versa.

3.1 Favorable Propagation

In this section, we consider the system model introduced above and particularize the favorable propagation definition for that model. To decode the main user signal, we utilize matched filter (MF) beamforming (also known as maximum ratio combining, which maximizes the single-user SNR) tuned to the User 1 channel, $\boldsymbol{w}_1 = \boldsymbol{h}_1/|\boldsymbol{h}_1|$ (detailed discussion for this choice is provided in Section 3.2). The exact FP corresponds to the total IUI power being zero [7], i.e.,

$$\alpha_N^2 = \sum_{i=2}^{M} |\alpha_{1i,N}|^2 = 0 \quad \Leftrightarrow \quad \alpha_{1i,N} = 0, \quad \forall i > 1 \tag{3.14}$$

so that the first upper bound in (3.4) is reached,

$$\text{SINR} = |\alpha_{11,N}|^2 \gamma_1 \tag{3.15}$$

This favorable condition can be approached in certain scenarios by increasing the number of antennas without bound, $N \to \infty$, which is referred to as "asymptotically favorable propagation" [7][48][56]. Since the total IUI is rarely exactly zero when the number of antennas is finite, we further omit the term "asymptotically" and use "favorable propagation" (FP), where $N \to \infty$ is included in the definition:

$$\lim_{N\to\infty} \alpha_N^2 = 0 \quad \Leftrightarrow \quad \lim_{N\to\infty} |\alpha_{1i,N}| = 0, \quad \forall i > 1 \tag{3.16}$$

From the limit definition, IUI becomes negligible in this case and the SINR approaches its maximum (single-user value) as N increases.

While $N \to \infty$ is not possible in practice, an important practical implication of (3.16), which follows from the limit definition, is that the IUI power can be made as small as desired provided the number of antennas is large enough. From the limit definition, it follows that for any $\epsilon > 0$, there exists a number N_ϵ such that, for any $N \geq N_\epsilon$, the

normalized IUI does not exceed the threshold ϵ,

$$\forall \epsilon > 0, \quad \exists N_\epsilon : \quad \forall N \geq N_\epsilon, \quad \alpha_N \leq \epsilon \tag{3.17}$$

It should be pointed out that $|\alpha_{1i,N}|$ is an important metric of channel orthogonality that is often used in experimental studies [46][49][52].

3.2 SINR of Different Beamformers Under the FP

There are several different beamforming techniques, such as matched filtering (MF, also known as MRC/MRT), zero forcing (ZF) and minimum mean square error (MMSE). Match filtering or conjugate beamforming is an accepted approach for massive MIMO systems in the literature [7][10][11][48][42]. Attractive features of MF compared to other beamforming techniques include [10]:

1. Computational simplicity
2. Distributed nature
3. Robustness

The computational complexities for the three linear beamforming methods MF, ZF, and MMSE are shown in Tables 3.1 and 3.2, where $\boldsymbol{H} = [\boldsymbol{h}_1..\boldsymbol{h}_M] \in \mathbb{C}^{N \times M}$ is the channel matrix of all M single antenna users communicating to an N-element base station. The complexity metric is the total number of flops[2].

Additionally, the distributed nature of MF, where the beamforming vector to detect User 1 is $\boldsymbol{h}_1$, means that it does not require knowledge of the channel state information (CSI) of other users to detect a single user, unlike other linear methods (see Table. 3.2) . Moreover, the MF beamforming technique has greater robustness and a decentralized architecture, which makes it preferable over other methods. Adding or removing antennas

[2] A flop is defined as one addition, subtraction, multiplication, or division of two floating-point numbers [111, p. 662].

Table 3.1: Comparison of linear beamformers computational complexity for detection of all M users

	Beamforming matrix	Computational complexity	$M = 10$, $N = 100$	$M = 100$, $N = 1000$
MF	$\boldsymbol{H}$	MN	10^3	10^5
ZF	$\boldsymbol{H}(\boldsymbol{H}^+\boldsymbol{H})^{-1}$	$3NM^2/2$	$1.5 \cdot 10^4$	$1.5 \cdot 10^7$
MMSE	$\boldsymbol{H}(\boldsymbol{H}^+\boldsymbol{H} + \sigma_x^{-2}\boldsymbol{I}_M)^{-1}$	$3NM^2/2$	$1.5 \cdot 10^4$	$1.5 \cdot 10^7$

Table 3.2: Comparison of linear beamformers computational complexity for detection of User 1 alone

	Beamforming vector	Computational complexity	$M = 10$, $N = 100$	$M = 100$, $N = 1000$
MF	$\boldsymbol{h}_1$	N	10^2	10^3
ZF	$[\boldsymbol{H}(\boldsymbol{H}^+\boldsymbol{H})^{-1}]_1$	$3NM^2/2$	$1.5 \cdot 10^4$	$1.5 \cdot 10^7$
MMSE	$[\boldsymbol{H}(\boldsymbol{H}^+\boldsymbol{H}+\sigma_x^{-2}\boldsymbol{I}_M)^{-1}]_1$	$3NM^2/2$	$1.5 \cdot 10^4$	$1.5 \cdot 10^7$

does not affect the operation of any other antennas, since each service antenna is responsible for its own channel estimation and signal processing [10].

In general, however, matched filtering has inferior performance compared to MMSE or ZF in canceling IUI. SIC is best of all, but it is nonlinear and therefore much more complicated to implement [10][28]. If the FP holds, the performance of MF approaches that of the other techniques (including that of SIC) as the number of antennas increases and the full capacity region is achieved. To illustrate this, let us consider the exact FP condition for simplicity, and observe that the channel vectors of different users become orthogonal to each other (the channel Gramm matrix $\boldsymbol{H}^+\boldsymbol{H}$ becomes diagonal) so that there is no IUI to cancel:

$$\boldsymbol{h}_i^+\boldsymbol{h}_j = 0, \quad i \neq j \quad \leftrightarrow \quad \boldsymbol{H}^+\boldsymbol{H} = \boldsymbol{D} \tag{3.18}$$

where $\boldsymbol{D} = \text{diag}\{|\boldsymbol{h}_i|^2\}$. Hence, using the MF to detect User k ($\boldsymbol{w}_k = \boldsymbol{h}_k$), (3.3) extends

to:

$$\boldsymbol{w}_k^+ \boldsymbol{y}(t) = \boldsymbol{h}_k^+ \boldsymbol{h}_k x_k(t) + \underbrace{\boldsymbol{h}_k^+ \sum_{\substack{i=1 \\ i \neq k}}^{M} \boldsymbol{h}_i x_i(t)}_{\text{IUI}} + \boldsymbol{h}_k^+ \boldsymbol{\xi}(t) \tag{3.19}$$

$$= |\boldsymbol{h}_k|^2 x_k(t) + \boldsymbol{h}_k^+ \boldsymbol{\xi}(t) \tag{3.20}$$

Therefore, under the exact FP, there is no IUI to cancel in detecting User k and the MF receiver becomes identical to the SIC detector (see [28] for more details on SIC).

Chapter 4

Favorable propagation for uniform circular/cylindrical arrays

4.1 Introduction

Previous studies have examined favorable propagation in certain settings, but there are some challenging settings where the FP has not been previously explored. In this chapter, we analyze the FP for a finite number of users with fixed and distinct AoAs and address the open question of whether the FP holds for uniform circular arrays with any fixed element spacings, d.

Next, a uniform cylindrical array (UCLA) is considered, which is a popular array geometry for practical implementations [50], [51], [52], [90], [91], but for which the FP has not yet been analyzed. It should be emphasized that the analysis of UCA/UCLA is more challenging than that of ULA/UPA, since there is no closed-form expression for the array patterns in circular arrays, unlike the linear/planar ones. Hence, new tools are needed to analyze the FP for these arrays.

We note that the integral sum representation method, which was used in [48] to prove the FP for fixed-size UCAs, cannot be applied to UCAs with fixed element spacing d, since

in the latter case, the summation does not converge to an integral. Moreover, since our model is deterministic, the law of large numbers, which is used for fading channels, cannot be applied. To this end, we will use the generating function of Bessel coefficients, carefully bounding each term in the series, and rigorously prove that the FP holds asymptotically for UCAs and UCLAs with fixed antenna spacing under LOS propagation, see Theorem 1 and Propositions 1-3. Based on the asymptotic analysis, a condition on the number of antennas to closely approach the FP is derived.

4.2 Channel Model

Consider the system model defined in (3.2), where an N-element BS serves M single-antenna users simultaneously. The channel vectors are assumed to be fixed, as in a quasi-static channel model during a block time when the delay requirement for the system is less than the channel coherence time [28]. This assumption corresponds to scenarios such as LOS-dominated propagation, as in mmWave systems [17][19]. By applying matched filter beamforming with the beamforming weight $\boldsymbol{w}_1 = \boldsymbol{h}_1/|\boldsymbol{h}_1|$ to detect User 1, the SINR defined in (3.4) simplifies to:

$$\text{SINR} = \frac{\gamma_1}{\sum_{i=2}^{M} |\alpha_{iN}|^2 \gamma_i + 1}, \quad \alpha_{iN} = \frac{\boldsymbol{h}_1^+ \boldsymbol{h}_i}{N} \tag{4.1}$$

So, we use the IUI "leakage", α_{iN}, in (4.1) to evaluate the FP property of the UCA in both 2-D and 3-D cases and also uniform cylindrical arrays.

4.3 FP for Uniform Circular Arrays

In this section, we first analyze the FP for a finite number of users with fixed and distinct AoAs and demonstrate that favorable propagation does hold for uniform circular arrays with any fixed element spacing $d > 0$ under LOS propagation. Next, we will briefly discuss

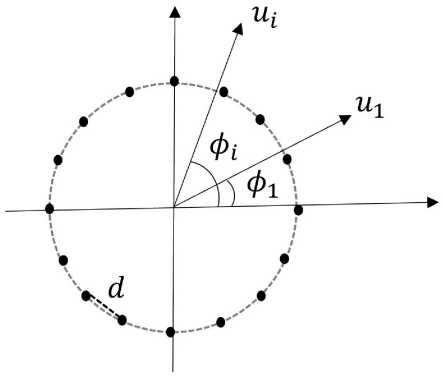


Figure 4.1: A planar (2-D) case: users are located in the UCA plane.

the FP in the case of an unbounded number of users and identify a few scenarios for which the FP does not hold.

4.3.1 Planar 2-D Case

Since the analysis is rather involved, we begin with the planar (2-D) case, when users are located in the plane of the UCA (see Fig. 4.1) and then extend this to the full 3-D case.

In this 2-D case, under LOS propagation in the far field, the normalized channel vectors can be expressed as[1]

$$\begin{aligned} \boldsymbol{h}_i &= [e^{j\psi_{0i}}, ..., e^{j\psi_{(N-1)i}}]', \ i = 1...M \\ \psi_{ni} &= 2\pi R_N \cos(\phi_i - \varphi_n), \quad n = 0...N-1 \end{aligned} \tag{4.2}$$

where $\varphi_n = 2\pi n/N$, $R_N = d|2\sin(\varphi_1/2)|^{-1}$ is the circle radius, d is the element spacing (normalized to the wavelength), ϕ_i is the angle-of-arrival (AoA) of User i. The FP property for this array geometry is established below.

Theorem 1. *Favorable propagation holds for the UCA under LOS propagation, (any) fixed*

[1]We omit here the common phase shift term $e^{j\varphi_i}$ since it does not affect FP; the LOS path loss is absorbed into the Rx signal power/SNR.

element spacing $d > 0$ and any fixed number of users $M < \infty$ with distinct AoAs, i.e.

$$\phi_1 \neq \phi_i, \quad i = 2...M, \quad \Leftrightarrow \quad \lim_{N\to\infty} |\alpha_N| = 0 \tag{4.3}$$

in the planar 2-D case.

Proof. Assume, without loss of generality (due to rotational symmetry), that the AoA of User 1 and User i are 0 and $\phi_i \neq 0, i \geq 2$ respectively. Thus, α_{iN} can be represented as follows:

$$\alpha_{iN} = \frac{\boldsymbol{h}_1^+ \boldsymbol{h}_i}{N} = \frac{1}{N}\sum_{n=0}^{N-1} e^{j2\pi R_N[\cos(\varphi_n - \phi_i) - \cos(\varphi_n)]} \tag{4.4}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} e^{jz_{Ni}\sin(\varphi_n - \phi_i/2)}, \; z_{Ni} = 4\pi R_N \sin(\phi_i/2) \tag{4.5}$$

where $\varphi_n = 2\pi n/N$. It is noteworthy that this sum does not converge to an integral since $|z_{Ni}| \to \infty$ as $N \to \infty$ and hence the method of [48] cannot be applied.

To overcome this difficulty, we employ a different approach based on the generating function of the Bessel coefficients [101, sec. 2.22]:

$$e^{jx\sin(\phi_i)} = \sum_{m=-\infty}^{+\infty} e^{jm\phi_i} J_m(x) \tag{4.6}$$

where J_m is Bessel function of 1st kind and order m, so that α_{iN} in (4.4) can be expressed as follows:

$$\alpha_{iN} = \frac{1}{N}\sum_{m=-\infty}^{+\infty} e^{-j\phi_i m/2} J_m(z_{Ni}) \sum_{n=0}^{N-1} e^{j2\pi mn/N} = \sum_{k=-\infty}^{+\infty} e^{-jkN\phi_i/2} J_{kN}(z_{Ni}) \tag{4.7}$$

where (4.7) follows from below:

$$\sum_{m=0}^{N-1} e^{jmn\frac{2\pi}{N}} = \begin{cases} N & \text{if } n = kN \\ 0 & \text{if } n \neq kN \end{cases} \qquad k = 0, \pm 1, \pm 2, \cdots \tag{4.8}$$

One should note that interchanging the order of sums in (4.7) needs to be evaluated since

the second sum is an infinite series (see proof in Appendix A.1).

So, by substituting (4.8) to (4.7), α_{iN} is simplified to the following expression:

$$\alpha_{iN} = \sum_{k=-\infty}^{+\infty} e^{-j\frac{kN\phi_i}{2}} J_{kN}(z_{Ni}) \tag{4.9}$$

There is no simple analytical expression for this series, and the complex coefficients further complicate the determination of the limit of $|\alpha_{iN}|$. However, to address this issue, we can investigate the upper bound of α_{iN} and claim that if the upper bound tends to zero with increasing N, the limit of α_N would also approach zero due to the limited number of users. We use the following upper bound based on (4.7) and utilize the symmetry property $J_{-k} = (-1)^k J_k$, to facilitate the evaluation of the limit:

$$|\alpha_{iN}| \leq 2\sum_{k=0}^{k_1-1} |J_{kN}(z_{Ni})| + 2\sum_{k=k_1}^{\infty} |J_{kN}(z_{Ni})| \tag{4.10}$$

$$= A_{1Ni} + A_{2Ni} \tag{4.11}$$

where we set $k_1 = \lceil e\pi d \rceil$ (the reason of this choice will be justified later on), $\lceil x \rceil$ is the smallest integer greater or equal to x; A_{1Ni}, A_{2Ni} denote 1st and 2nd summation terms. Next, we obtain an upper bounds on A_{1Ni}, A_{2Ni} and demonstrate that they converge to 0 as $N \to \infty$. To upper bound A_{1Ni}, we use:

$$z_{Ni} = 4\pi R_N \sin(\phi_i/2) = \frac{2Nd\sin(\phi_i/2)}{\mathrm{sinc}(1/N)} \tag{4.12}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$. Exploiting the monotonicity of $\mathrm{sinc}(1/N)$ for $N \geq 2$, one obtains:

$$2/\pi \leq \mathrm{sinc}(1/N) \leq 1 \tag{4.13}$$

so, $|z_{Ni}|$ is bounded by:

$$2Nd|\sin(\phi_i/2)| \leq |z_{Ni}| \leq N\pi d \tag{4.14}$$

Combining (4.14) with the following upper bound [102][103]

$$|J_k(x)| \leq |x|^{-1/3} \tag{4.15}$$

one obtains

$$|J_{kN}(z_{Ni})| \leq |z_{Ni}|^{-1/3} \leq |2Nd\sin(\phi_i/2)|^{-1/3} \tag{4.16}$$

and an upper bound on A_{1Ni} follows:

$$A_{1Ni} \leq 2\sum_{k=0}^{k_1-1} |2Nd\sin(\phi_i/2)|^{-1/3} = 2k_1|2Nd\sin(\phi_i/2)|^{-1/3} \tag{4.17}$$

Next, we obtain an upper bound on A_{2Ni}. To this end, we will need the following technical Lemma, which presents a novel upper bound on Bessel functions.

Lemma 1. *If* $|x| \leq 1$, *then*

$$|J_n(nx)| \leq |x.e/2|^n \tag{4.18}$$

Proof. Using the following inequalities:

$$|J_n(nx)| \leq \left| \frac{x\exp\left\{\sqrt{(1-x^2)}\right\}}{\left\{1+\sqrt{(1-x^2)}\right\}} \right|^n \leqslant x^n \cdot \max_{|x|\leq 1} \left\{ \frac{e^{\sqrt{1-x^2}}}{1+\sqrt{1-x^2}} \right\}^n \leqslant \left(\frac{xe}{2}\right)^n \tag{4.19}$$

where the first inequality in (4.19) is Kapteyn's inequality [101, sec. 8.7] and the last inequality is due to the following:

$$\max_{|t|\leq 1} \frac{e^{\sqrt{1-t^2}}}{1+\sqrt{1-t^2}} = e/2 \tag{4.20}$$

□

Using (4.18) and (4.14), one obtains:

$$A_{2Ni} \leqslant 2\sum_{k=k_1}^{\infty} \left(\frac{e\pi d}{2k}\right)^{kN} \leq 2\sum_{k=k_1}^{\infty} q^{Nk} \leq 2\sum_{k=k_1}^{\infty} (1/2)^{Nk} = \frac{2(1/2)^{Nk_1}}{1-(1/2)^N} \tag{4.21}$$

where $q = e\pi d/2k_1 \leq 1/2$, from which proper k is obtained. And finally, using (4.21) and (4.17), one obtains:

$$\lim_{N\to\infty} |\alpha_{iN}| \leq \lim_{N\to\infty} (A_{1Ni} + A_{2Ni}) \tag{4.22}$$

$$\leq \lim_{N\to\infty} \frac{2\lceil e\pi d\rceil}{|2Nd\sin(\phi_i/2)|^{1/3}} + \lim_{N\to\infty} \frac{2(1/2)^{N\lceil e\pi d\rceil}}{1-(1/2)^N} = 0 \tag{4.23}$$

and therefore $\lim_{N\to\infty} |\alpha_{iN}| = 0$, which implies $\lim_{N\to\infty} |\alpha_N| = 0$, meaning that the FP holds for uniform circular arrays. □

It should be pointed out that Theorem 1 establishes that the FP holds point-wise, i.e. for any given User (which can be taken to be User 1, without loss of generality) under distinct AoAs.

As the number of users M increases unboundedly with N, the FP no longer holds in general, but it may still hold in some special cases. To see this, consider $M = N$ users with uniformly-distributed AoAs so that $\phi_i = 2\pi(i-1)/N,\ i = 1...N$. Next, using (4.9)-(4.12), one obtains:

$$\lim_{N\to\infty} \alpha_N \geq \lim_{N\to\infty} |\alpha_{2N}| = |J_0(2\pi d)| \tag{4.24}$$

and $J_0(2\pi d) \neq 0$ unless $2\pi d$ is a null of $J_0(x)$, where J_0 is Bessel function of the 1st kind and order 0. For example, $J_0(2\pi d) = J_0(\pi) \approx -0.3$ if $d = 1/2$, so there is no FP here. (4.24) also holds for arbitrary AoAs distribution provided that the closest user to User 1 is at $2\pi/N$. Furthermore, the FP may be violated even if M is fixed and finite, but as N grows, some AoAs will approach arbitrarily close to the AoA of user 1. For example, let $M = 2$ and $\phi_2 = 2\pi/N$ so that

$$\lim_{N\to\infty} \alpha_N = |J_0(2\pi d)| \neq 0 \tag{4.25}$$

i.e. the FP does not hold.

The behavior of α_N and $|\alpha_{12,N}|$ for $M = 10$ and 100 as the number N of antennas is

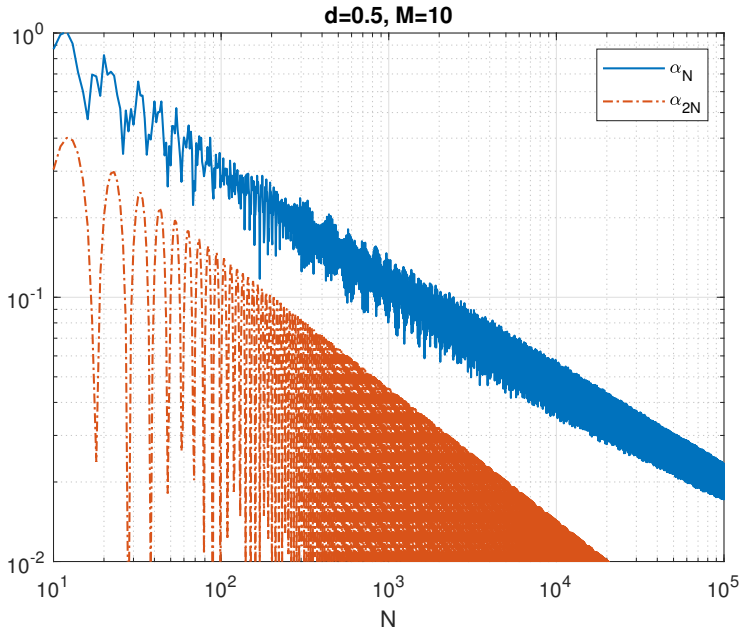


Figure 4.2: IUI factors $|\alpha_{2N}|$ and $|\alpha_N|$ vs. N for $M = 10$, $d = 0.5$; uniform user AoAs $\phi_i = 2\pi(i-1)/M$, $i = 1...M$. Note that the envelope of α is monotonically decreasing with N, i.e., the FP property holds, and that larger M calls for larger N to attain low $|\alpha|$.

increasing is demonstrated in Fig. 4.2 and Fig. 4.3, respectively. Here, $|\alpha_{12,N}|$ corresponds to the 2-user case with $\phi_1 = 0 and \phi_2 = 2\pi/M$. These simulation results are representative, since similar behaviours also hold for other values of M and d, and they are consistent with our analytical results since the respective limit of α_N is zero when the FP holds, so arbitrary low α_N can be achieved for a large enough N.

The results presented in this section are discussed in [94].

4.3.2 Extension to 3-D Case

Next, we consider the full 3-D case, see Fig. 4.4.

Proposition 1. *Favorable propagation holds for the UCA in the full 3-D case under LOS propagation, for (any) fixed element spacing $d > 0$, any fixed $M < \infty$, and any distinct and fixed AoAs if and only if neither of the following holds for any $i \geq 2$:*

1. $\theta_{1,i} = 0$ *or* π

2. $\phi_1 = \phi_i$ *and* $\theta_1 = \theta_i$ *or* $\theta_1 = \pi - \theta_i$

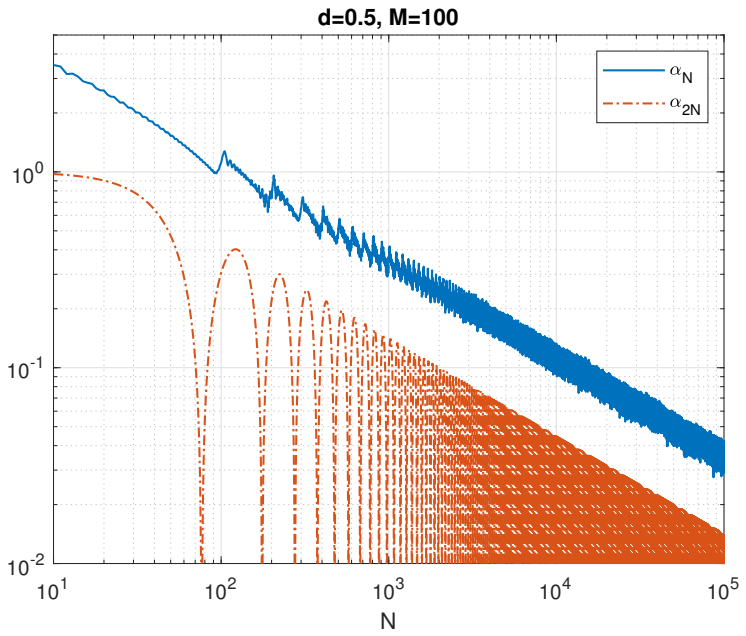


Figure 4.3: The same setting as in Fig. 4.2 but with $M = 100$ users. Note that larger N is needed to achieve the same low α compared to Fig. 4.2

.

Proof. This general case is proved with using the Theorem 1 proof with appropriate modifications for 3-D environments. In particular, the channel vector $\boldsymbol{h}_i$ of User i can be expressed as

$$\begin{aligned} \boldsymbol{h}_i &= [e^{j\psi_{0i}}, ..., e^{j\psi_{(N-1)i}}]', \ i = 1...M \\ \psi_{ni} &= 2\pi R_N \sin(\theta_i)\cos(\phi_i - \varphi_n), \ n = 0...N-1 \end{aligned} \tag{4.26}$$

where $\varphi_n = 2\pi n/N$; α_{iN} can be expressed as

$$\begin{aligned} \alpha_{iN} &= \frac{1}{N}\sum_{n=0}^{N-1} e^{j2\pi R_N(\delta_{1i}\cos\varphi_n + \delta_{2i}\sin\varphi_n)} \\ &= \frac{1}{N}\sum_{n=0}^{N-1} e^{j2\pi R_N \delta_i \sin(\varphi_n + \beta_i)} \\ &= J_0(z_{Ni}) + \sum_{k=1}^{\infty}(e^{jkN\beta_i} + (-1)^{kN}e^{-jkN\beta_i})J_{kN}(z_{Ni}) \end{aligned} \tag{4.27}$$

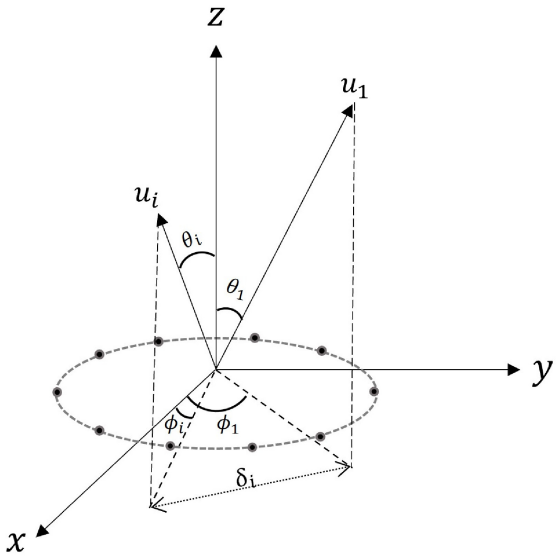


Figure 4.4: The full 3-D case geometry: users are not confined to the UCA plane.

where $z_{Ni} = 2\pi R_N \delta_i$, $\delta_i = \sqrt{\delta_{1i}^2 + \delta_{2i}^2}$, $\beta_i = \arg(\delta_{2i} + j\delta_{1i})$,

$$\begin{aligned} \delta_{1i} &= \sin\theta_i \cos\phi_i - \sin\theta_1 \cos\phi_1 \\ \delta_{2i} &= \sin\theta_i \sin\phi_i - \sin\theta_1 \sin\phi_1 \end{aligned} \tag{4.28}$$

Now, $|\alpha_{iN}|$ can be upper bounded as follows:

$$\begin{aligned} |\alpha_{iN}| &\leq 2\sum_{k=0}^{k_1-1} |J_{kN}(z_{Ni})| + 2\sum_{k=k_1}^{\infty} |J_{kN}(z_{Ni})| \\ &= A_{1N,i} + A_{2N,i} \end{aligned} \tag{4.29}$$

Using (4.17) and (4.21), the desired result follows if $\delta_i \neq 0$, which gives the conditions of Proposition 1. □

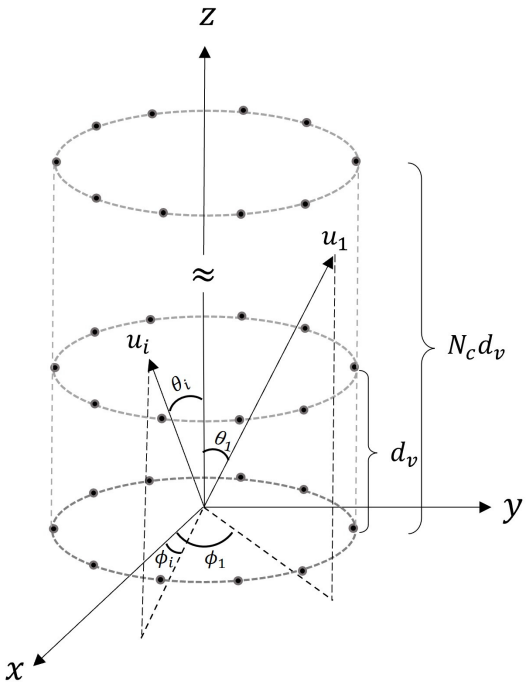


Figure 4.5: Uniform cylindrical array (UCLA) comprised of N_c UCAs located on top of each other.

4.4 FP for Uniform Cylindrical Arrays

In this section, we consider a uniform cylindrical array (UCLA) of $(N_c \times N)$ elements consisting of N_c UCAs located on top of each other, each with N elements, as shown in Fig. 4.5.

The aggregate channel vector $\boldsymbol{h}_i$ of User i can be expressed as follows:

$$\boldsymbol{h}_i = [\boldsymbol{h}'_{i1}, ..., \boldsymbol{h}'_{iN_c}]' = \boldsymbol{h}_{ih} \otimes \boldsymbol{h}_{iv} \tag{4.30}$$

where $\boldsymbol{h}_{im}$ is the channel vector of UCA m for user i, $\otimes$ denotes Kronecker product, $\boldsymbol{h}_{ih} = \boldsymbol{h}_{i1}$ is the 1st UCA streeting vector, $\boldsymbol{h}_{iv}$ is the steering vector of vertical N_c-element ULA whose entries h_{ivm} are the respective phase shift terms (which account for phase shifts

of UCA m with respect to 1st UCA):

$$h_{ivm} = \exp\{j2\pi(m-1)d_v \cos(\theta_i)\},\ m = 1...N_c \tag{4.31}$$

The interference leakage factors from i-th to main user are denoted as follows:

$$\alpha_i = \frac{1}{N_c N}\boldsymbol{h}_1^+\boldsymbol{h}_i,\ \alpha_{ih} = \frac{1}{N}\boldsymbol{h}_{1h}^+\boldsymbol{h}_{ih},\ \alpha_{iv} = \frac{1}{N_c}\boldsymbol{h}_{1v}^+\boldsymbol{h}_{iv}$$

i.e. α_h and α_v represent the respective factors for a single UCA and ULA (located in horizontal and vertical planes, respectively). The following Proposition is instrumental in establishing the FP for the UCLA.

Proposition 2. *Let $\boldsymbol{h}_i$ have Kronecker structure as in (4.30). Then, α_i can be expressed as follows:*

$$\alpha_i = \alpha_{ih}\alpha_{iv} \tag{4.32}$$

and, furthermore,

$$|\alpha_i| \leq \min\{|\alpha_{ih}|, |\alpha_{iv}|\} \tag{4.33}$$

Proof. Observe the following:

$$\begin{aligned}
\alpha_i &= \boldsymbol{h}_1^+\boldsymbol{h}_i/(N_c N) \\
&= (\boldsymbol{h}_{1h} \otimes \boldsymbol{h}_{1v})^+(\boldsymbol{h}_{ih} \otimes \boldsymbol{h}_{iv})/(N_c N) \\
&= (\boldsymbol{h}_{1h}^+ \otimes \boldsymbol{h}_{1v}^+)(\boldsymbol{h}_{ih} \otimes \boldsymbol{h}_{iv})/(N_c N) \\
&= (N_c^{-1}\boldsymbol{h}_{1v}^+\boldsymbol{h}_{iv})(N^{-1}\boldsymbol{h}_{1h}^+\boldsymbol{h}_{ih}) = \alpha_{ih}\alpha_{iv}
\end{aligned} \tag{4.34}$$

where 3rd and 4th equalities are due to the properties of Kronecker products [95]. The inequality in (4.33) follows from $|\alpha_{h,v}| \leq 1$. □

Note that this result applies to any array satisfying the Kronecker product property in

(4.30), not just cylindrical (e.g. it also applies to planar arrays) and applies to Kronecker products for any number of vectors, not just two.

Next, We are ready to establish the FP property for the UCLA.

Proposition 3. *Favorable propagation holds for the UCLA in the full 3-D case under LOS propagation for (any) fixed element spacings $d, d_v > 0$, any fixed number of users $M < \infty$ with distinct and fixed AoAs in any of the two cases:*

- *Case 1. $N \to \infty$ and, for all $i \geq 2$, neither of the following holds :*

  (a) $\theta_{1,i} = 0$ or π;

  (b) $\phi_1 = \phi_i$ and $\theta_1 = \theta_i$ or $\theta_1 = \pi - \theta_i$,

- *Case 2. $N_c \to \infty$, $\theta_1 \neq \theta_i$ for any $i \geq 2$, $d_v < \lambda/2$.*

Proof. Consider case 1. Using (4.30) and (4.33), one obtains:

$$\lim_{N\to\infty} |\alpha_i| \leq \lim_{N\to\infty} |\alpha_{ih}| = 0 \tag{4.35}$$

which proves the desired result. The equality in (4.35) is due to Proposition 1. Case 2 follows in a similar approach from the FP property for the ULA. □

Based on the above result, we conclude that, in order to achieve favorable propagation, one should

- expand UCLA horizontally ($N \to \infty$) if $\phi_1 \neq \phi_i$ or if $\theta_1 \neq \theta_i$ and $\theta_1 + \theta_i \neq \pi$, or
- expand UCLA vertically ($N_c \to \infty$) if $\theta_1 \neq \theta_i$, $d_v < \lambda/2$.

Clearly, simultaneous expansion (both vertically and horizontally) is also possible to achieve the FP.

4.5 Conclusion

Favorable propagation is rigorously demonstrated in LOS environments for uniform circular and cylindrical arrays. This result holds for any fixed antenna spacing and a finite number M of users with fixed locations[2] and distinct angles of arrival. Under these conditions, with a large enough number of antennas, inter-user interference can be made as low as desired. To approach the FP closely, a condition on N is derived based on the asymptotic analysis. The FP does not hold in general if either M grows unbounded with N or if the AoAs are allowed to approach each other. However, it might hold in some special cases.

[2]Further, it can be shown that favorable propagation still holds even when users are not at a fixed location but do not approach the main user too closely.

Chapter 5

Structural design of non-uniform linear array for favorable propagation

5.1 Introduction

Extensive research on Favorable propagation for ULAs has demonstrated that the FP holds asymptotically as the number of antenna elements increases without bound. This holds true for a fixed antenna spacing and LOS propagation, as long as users have distinct angles of arrival (AoA) [7][46][48]-[53]. In this chapter, we show that the aforementioned conclusion relies on an implicit assumption, not explicitly mentioned in the previous studies: the absence of grating lobes (GLs) in the array pattern. If GLs are present and align with the AoA of some users, the conclusion fails.

It is desired to have greater element spacing under fixed number of elements (i.e., fixed complexity and cost), since in this case, the array spatial resolution increases or, equivalently, its beamwidth decreases. This results in decreased inter-user interference as the antenna array can better distinguish nearby users. Nevertheless, this has the major drawback of generating grating lobes in the array pattern for greater spacings [61][62]. It is therefore impossible to distinguish different users appearing in the main beam and GL

directions, which results in significant IUI. This has a negative consequence: the FP no longer holds, even under distinct AoAs. The greater the antenna spacing, the more grating lobes appear so that more user directions should be banned to maintain the FP.

To address these issues, we present an analysis of the effect of grating lobes on favorable propagation and propose a non-uniform linear array (NULA) design that effectively cancels grating lobes, thereby guaranteeing the FP for any element spacing and any distinct AoAs. Therefore, under fixed complexity, spatial resolution is increased, whereas in the standard approach, spatial resolution is increased by increasing the number of antennas. We propose a block-partitioned NULA design, where each block (subarray) is an ULA but the overall array is not uniform. The number of subarrays (blocks) and their spacing should be carefully selected so that, asymptotically, grating lobes are canceled by nulls in the block array factor, restoring FP for any distinct AoAs. To find the proper number and spacing of subarrays, a rigorous analysis is presented that incorporates some tools from number theory and, to the best of our knowledge, has not been applied to antenna arrays or massive MIMO literature before. It should be noted that while the actual number of grating lobes and their directions do depend on the main beam direction (i.e., beam steering) (see e.g., [61][62]), the design we propose is independent of it. Therefore, it can accommodate beam steering, and grating lobes are canceled for any direction of the main beam. This block-partitioned design is suitable for many practical applications, where elements are grouped in subarrays and hybrid beamforming is used to reduce its complexity and cost [96].

As discussed in the literature review, large bandwidth is essential for 5G and beyond. The design of antenna arrays for these systems is generally more complicated, as highlighted in [61]. To address the challenges posed by large bandwidths, we have extended our antenna array design from narrowband to wideband systems.

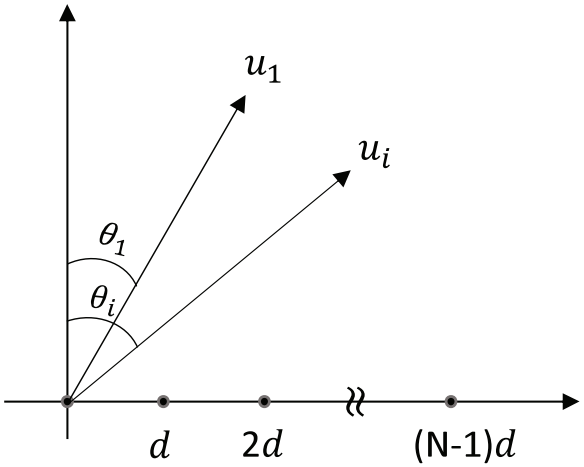


Figure 5.1: An illustration of ULA(N, d), where u_i is a unitary vector directed to User i with AoA θ_i, all measured from broadside, and the main user is at θ_1; $-\pi/2 \leq \theta_1,\ \theta_i \leq \pi/2$. When decoding the 1st user's signal, the i-th user becomes a source of interference.

5.2 Channel Model and Favorable Propagation for ULAs

Consider a Gaussian MIMO channel defined in (3.2), where an N-element BS serving M single-antenna users simultaneously. The channel is frequency-flat, implying that the delay spread is much smaller than the symbol duration. Using the linear processing with matched filter beamforming $\boldsymbol{w} = \boldsymbol{h}_1/|\boldsymbol{h}_1|$ to decode the main user, the IUI leakage factor, α_{iN}, in (3.4) is simplified to $\alpha_{iN} = \boldsymbol{h}_1^+ \boldsymbol{h}_i/N$. If all users have the same SNR ($\gamma_i = \gamma_1$), then the SINR simplifies to:

$$\text{SINR} = (\alpha_N^2 + \gamma_1^{-1})^{-1} \leq \gamma_1 \tag{5.1}$$

In the far-field, the normalized channel vector of User i for an N-element ULA with omnidirectional elements (e.g., vertical dipoles for a horizontal array as shown in Fig. 5.1)

is [7]:

$$\boldsymbol{h}_i = [e^{j\psi_{0i}}, \cdots, e^{j\psi_{(N-1)i}}]', \quad i = 1, \cdots, M \tag{5.2}$$
$$\psi_{ni} = 2\pi n d \sin(\theta_i), \quad n = 0, \cdots, N-1$$

where d is the element spacing (measured in wavelengths), θ_i is the AoA of User $i = 1...M$ signal, $-\pi/2 \leq \theta_i \leq \pi/2$; ψ_{ni} is the phase shift at element n with respect to 1st element, see Fig. 5.1. For simplicity, omnidirectional elements are assumed here, and the results can be readily extended to directional elements as well (see Section 5.5). All users are assumed to be in the front half-plane, $-\pi/2 \leq \theta_i \leq \pi/2$, and there is no backward radiation. For further reference, we use ULA(N, d) to denote a uniform linear array with N omnidirectional elements and element spacing d. Using the system model in (3.2), the inter-user interference leakage from User i to the main user ($i = 1$) can be expressed as follows:

$$\alpha_{iN} = \frac{1}{N}\sum_{n=0}^{N-1} e^{jn\Delta\psi_i} = \frac{\sin(N\Delta\psi_i/2)}{N\sin(\Delta\psi_i/2)} e^{j(N-1)\Delta\psi_i/2} \tag{5.3}$$

where $\Delta\psi_i = 2\pi d(\sin\theta_i - \sin\theta_1)$. Further notice that $\lim_{N\to\infty}\alpha_{iN} = 0$, i.e., the FP holds, provided that $\sin(\Delta\psi_i/2) \neq 0$. The latter condition may be violated even if $\theta_i \neq \theta_1$ (distinct AoAs), e.g. if $d = 1$, $\theta_1 = 0$, $\theta_i = \pm 90^o \neq \theta_1$ so that $\Delta\psi_i = \pm 2\pi$, $\sin(\Delta\psi_i/2) = 0$ and hence $|\alpha_{iN}| = 1$ for any N. This represents a grating lobe in the array pattern, see e.g. [61][62]. In general, GL directions ϕ_k correspond to zero denominator in (5.3), i.e.,

$$\sin(\Delta\psi_i/2) = 0 \quad \Leftrightarrow \quad \Delta\psi_i = 2\pi k \quad \Leftrightarrow \quad \alpha_{iN} = 1 \tag{5.4}$$

where for a given θ_1 and $\theta_i = \phi_k$, it is equivalent to

$$\sin(\phi_k) = \sin(\theta_1) + k/d \tag{5.5}$$

where k is the GL index, $k = \pm 1, \pm 2, ...$ Since $|\sin\phi_i| \leq 1$, there exist no grating lobes if

$$d(1 + |\sin\theta_1|) < 1 \tag{5.6}$$

In this case, the results in [7][48][46] do hold for any distinct θ_i, but they may fail to hold if (5.6) is not satisfied. Indeed, using (5.3), it follows that, under distinct AoAs $\theta_i \neq \theta_1$,

$$\lim_{N\to\infty} |\alpha_{iN}| = \begin{cases} 1, \text{if } d(1 + |\sin\theta_1|) \geq 1 \ \& \ \theta_i = \phi_k \\ 0, \ \text{otherwise.} \end{cases} \tag{5.7}$$

Note that there is a dichotomy here: the limit is either zero or one. The latter case gives the conditions when the FP fails to hold: $d(1 + |\sin\theta_1|) \geq 1$ is the condition for the GL existence, and $\theta_i = \phi_k$ is the condition for i-th user AoA to coincide with k-th GL direction.

To determine the number of grating lobes, observe that $|\sin(\phi_k)| \leq 1$ and use (5.5) to obtain the range of k:

$$k \geq k_{min} = -\lfloor d(1 + \sin(\theta_1))\rfloor \geq -\lfloor 2d\rfloor \tag{5.8}$$

$$k \leq k_{max} = \lfloor d(1 - \sin(\theta_1))\rfloor \leq \lfloor 2d\rfloor \tag{5.9}$$

where $\lfloor\cdot\rfloor$ is the floor function, so that $I_k = \{\{k_{min}, \cdots, k_{max}\} - \{0\}\}$ is the GL index set, $k = 0$ represents the main beam and hence is excluded; I_k is an empty set if there are no GLs. Thus, the number K of GLs does not exceed $\lfloor 2d\rfloor$: $K = k_{max} - k_{min} \leq \lfloor 2d\rfloor$, and there are no grating lobes if $d < 1/2$, for any θ_1.

Examples: To illustrate cases where the FP fails due to the presence of GLs, let $d = 0.55$, $\theta_1 = 60^o$ and $\sin\theta_i = \sin\theta_1 - 1/d$ so that $\theta_i \approx -72^o \neq \theta_1$ (i.e., distinct AoAs), but $\Delta\psi_i = -2\pi$ and the denominators in (5.3) is zero and $|\alpha_{iN}| = 1$ for any N, i.e., the FP fails to hold even though the AoAs are distinct. This can be explained via the array factor shown in Fig. 5.2, where the main beam is at $\theta_1 = 60^o$ to follow User 1 while the grating lobe appears at -72^o, so that, if another user is at the latter direction, it cannot be discriminated from the 1st user and hence the FP fails to hold. Note from (5.7) that

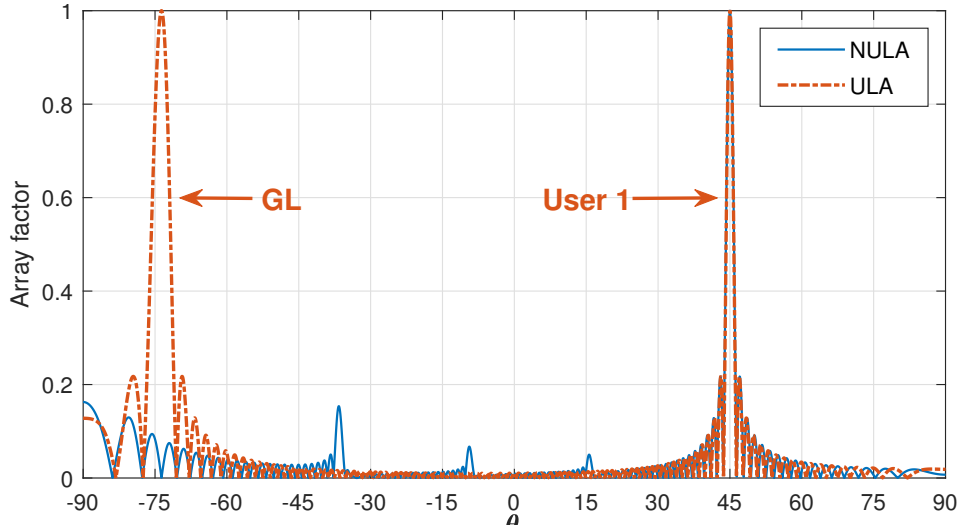


Figure 5.2: Array factor of a ULA with $N = 75$, $d = 0.55$. While the main beam is at $\theta_1 = 60^o$, note the presence of a grating lobe at $\theta_2 = -72^o$. The proposed NULA design with $N_b = 25,\ N = 3,\ p = 21$ partially cancels this grating lobe, see Theorem 2.

the FP may fail to hold even if $d = 1/2$ e.g. if $\theta_1 = 90^o$, $\theta_i = \phi_{-1} = -90^o \neq \theta_1$ so that $\Delta\psi_i = -2\pi$, $\sin(\Delta\psi_i/2) = 0$ and $|\alpha_{iN}| = 1$ for any N (but the FP always holds if $d < 1/2$).

Even if the main user is at broadside, i.e., $\theta_1 = 0^o$, grating lobes appear if $d \geq 1$ and the FP fails to hold, even under distinct AoAs[1], e.g. if $d = 1,\ \theta_1 = 0^o,\ \theta_i = \pm\pi/2$. In general, the larger d, the more grating lobes emerge [61][62] and favorable propagation fails to hold if users' AoAs, being distinct from each other, coincide with the GL directions. Eq. (5.7) gives precise condition for this to happen.

5.3 Non-Uniform Linear Array Design for FP

Following the above analysis of grating lobes and their impact on the FP, this section introduces a nonuniform linear array design. This design effectively eliminates all GLs and guarantees the FP for any distinct AoAs and any element spacing d. The designed non-uniform array has a block-partitioned structure and is composed of N_b subarrays (blocks),

[1]since the ULA made of isotropic or omnidirectional elements is not able to discriminate users' signals coming from opposite directions, i.e., θ_1 and $180^o - \theta_1$, the FP also fails to hold under distinct but opposite AoAs, for any element spacing. However, in practice this is not that important since backward radiation is usually eliminated due to an element pattern or an array design including a ground plane. Thus, it is not considered here; instead, we assume that $-\pi/2 \leq \theta_1,\ \theta_i \leq \pi/2$.

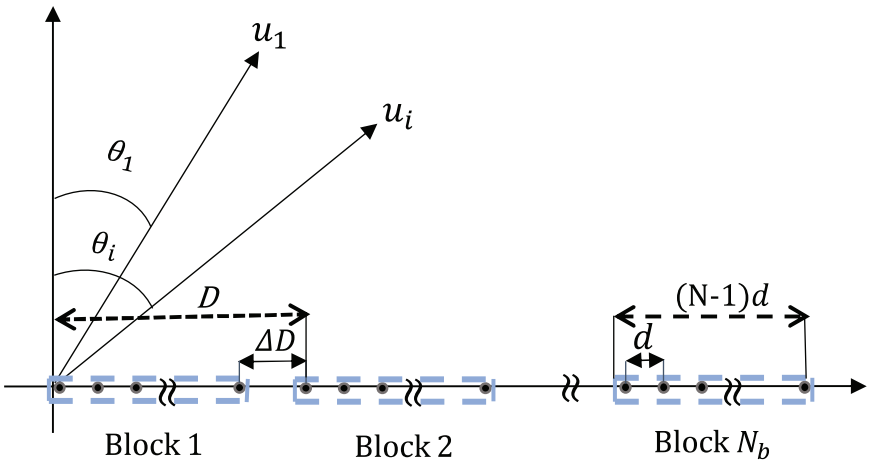


Figure 5.3: Block-partitioned NULA of N_b subarrays (blocks) ULA(N, d) with subarray spacing ΔD; $D = (N-1)d + \Delta D$, and $(N-1)D$ is the subarray length. The AoAs are distinct, $\theta_1 \neq \theta_i$, and $-\pi/2 \leq \theta_1,\ \theta_i \leq \pi/2$.

each of which is a ULA(N, d) and is arranged in the ULA(N_b, D) block-wise pattern. The overall design pattern is a product of the subarray factor of ULA(N, d) and the block array factor of ULA(N_b, D), where each block is replaced with an omnidirectional element, see e.g. [61][62]. To achieve this, we need to determine the appropriate values for N_b and the subarray spacing ΔD in order to cancel the grating lobes (GLs) in the subarray factor using nulls in the block array factor. This is crucial for ensuring favorable propagation with any distinct AoAs and any d.

Due to the symmetry of the structure, the overall channel vector $\boldsymbol{h}_i$ of the NULA for User i can be expressed as follows:

$$\boldsymbol{h}_i = \boldsymbol{h}_{si} \otimes \boldsymbol{h}_{bi} \tag{5.10}$$

where $\otimes$ denotes Kronecker product; $\boldsymbol{h}_{si}$ and $\boldsymbol{h}_{bi}$ represents the channel vector of the subarray ULA(N, d) and of the block array ULA(N_b, D), respectively, where $\boldsymbol{h}_{si}$ is as in (5.2) and $\boldsymbol{h}_{bi}$ is

$$\boldsymbol{h}_{bi} = [e^{j\psi_{b,0i}}, \cdots, e^{j\psi_{b,(N_b-1)i}}]', \quad i = 1, \cdots, M \tag{5.11}$$

$$\psi_{b,ni} = 2\pi n D \sin(\theta_i), \quad n = 0, \cdots, N_b - 1$$

For further use, the interference leakage terms of the NULA are defined as follows:

$$\alpha_{iN} = \frac{\boldsymbol{h}_1^+ \boldsymbol{h}_i}{N_b N}, \ \alpha_{siN} = \frac{\boldsymbol{h}_{s1}^+ \boldsymbol{h}_{si}}{N}, \ \alpha_{bi} = \frac{\boldsymbol{h}_{b1}^+ \boldsymbol{h}_{bi}}{N_b} \tag{5.12}$$

where α_{siN} and α_{bi} represent the respective terms for a single subarray ULA(N, d) and the block array ULA(N_b, D) while α_{iN} represents the overall leakage.

The following proposition is instrumental in establishing the FP for the block-partitioned NULA.

Proposition 4. *Let $\boldsymbol{h}_i$ have Kronecker structure as in (5.10). Then, α_{iN} can be expressed and bounded as follows:*

$$\alpha_{iN} = \alpha_{siN} \alpha_{bi} \tag{5.13}$$

$$|\alpha_{iN}| \leq \min\{|\alpha_{siN}|, |\alpha_{bi}|\} \tag{5.14}$$

Proof. Observe the following:

$$\begin{aligned} \alpha_{iN} &= (N_b N)^{-1} \boldsymbol{h}_1^+ \boldsymbol{h}_i \\ &= (N_b N)^{-1} (\boldsymbol{h}_{s1} \otimes \boldsymbol{h}_{b1})^+ (\boldsymbol{h}_{si} \otimes \boldsymbol{h}_{bi}) \\ &= (N_b N)^{-1} (\boldsymbol{h}_{s1}^+ \otimes \boldsymbol{h}_{b1}^+)(\boldsymbol{h}_{si} \otimes \boldsymbol{h}_{bi}) \\ &= (N^{-1} \boldsymbol{h}_{s1}^+ \boldsymbol{h}_{si})(N_b^{-1} \boldsymbol{h}_{b1}^+ \boldsymbol{h}_{bi}) = \alpha_{siN} \alpha_{bi} \end{aligned} \tag{5.15}$$

where 3rd and 4th equalities are due to the properties of Kronecker products [95]. The inequality in (5.14) follows from $|\alpha_{siN}|, \ |\alpha_{bi}| \leq 1$. □

Thus, the impact of subarray and block array factors α_{siN}, α_{bi} on the overall IUI leakage factor α_{iN} is factorized, which simplifies the analysis considerably. In particular, using (5.14), $|\alpha_{iN}| \to 0$ if either $|\alpha_{siN}| \to 0$ or $|\alpha_{bi}| \to 0$. This can be exploited to cancel grating lobe's effect on the FP. To this end, let us consider the asymptotic ("massive") regime where $N \to \infty$ while N_b is fixed (constant), under distinct AoAs.

Proposition 5. *If N_b is fixed and $\theta_1 \neq \theta_i$, then the following asymptotic relationship holds for the block-portioned NULA:*

$$\lim_{N\to\infty} |\alpha_{iN}| = \begin{cases} |\alpha_{bi}(\phi_k)|, & \exists k \in I_k : \ \theta_i = \phi_k \\ 0, & \text{otherwise}\,. \end{cases} \tag{5.16}$$

Proof. Using (5.14),

$$\lim_{N\to\infty} |\alpha_{iN}| = 0 \text{ if } \theta_i \neq \theta_1 \ \& \ \theta_i \neq \phi_k \ \forall k \in I_k \tag{5.17}$$

since, from (5.7), $\lim_{N\to\infty} |\alpha_{siN}| = 0$ in this case. On the other hand, if $\theta_i = \phi_k$ for some $k \in I_k$, then $|\alpha_{iN}| = |\alpha_{bi}(\phi_k)|$ since $|\alpha_{siN}| = 1$ in this case. Note from (5.12) that

$$\begin{aligned} \alpha_{bi}(\phi_k) &= \frac{1}{N_b} \sum_{n=0}^{N_b-1} e^{j2\pi n D(\sin(\phi_k)-\sin(\theta_1))} \\ &= \frac{1}{N_b} \frac{\sin(\pi N_b k \Delta D/d)}{\sin(\pi k \Delta D/d)} e^{j\pi N_b k \Delta D/d} \end{aligned} \tag{5.18}$$

where the last equality is from (5.5) and $D = (N-1)d + \Delta D$. Thus, $\alpha_{bi}(\phi_k)$ is independent of N and this proves the 1st case in (5.16). □

From (5.16), the FP is guaranteed under distinct AoAs if users do not align with grating lobes (or if GLs do not exist), $\theta_i \neq \phi_k \ \forall \ k \in I_k$. If some users do align, then the following equivalence holds:

$$\lim_{N\to\infty} \alpha_{iN} = 0 \quad \Leftrightarrow \quad \alpha_{bi}(\phi_k) = 0 \ \forall k \in I_k \tag{5.19}$$

i.e., grating lobes are canceled and the FP holds under *any* distinct AoAs if $\alpha_{bi}(\phi_k) = 0 \ \forall k \in I_k$. The latter can be achieved by exploiting the NULA structure and choosing appropriate values of N_b and ΔD as shown below.

To this end, we need the following concepts from number theory [97, p. 231]:

- *Greatest common divisor* of two integer m and n, $\gcd(m, n)$: the largest positive integer that devides m and n without remainder; e.g. $\gcd(15, 12) = 3$.

- *Coprime (relative prime)*: two numbers n and m are coprime if $\gcd(n,m)=1$ (no common divisors); e.g. $\gcd(4,5)=1$, so, 4 and 5 are coprime. If $\gcd(n,m)=1$ and $n>1$, then m/n is not integer.

The following theorem presents the NULA design to cancel all GLs and to achieve the FP under all distinct AoAs. The beam steering is subject to $|\theta_1|\leq\theta_{max}$ constraint, where $\theta_{max}\leq\pi/2$ is a given maximum steering angle. Note that $\theta_{max}=\pi/2$ corresponds to no constraint on steering.

Theorem 2. *In LOS environment, the FP holds asymptotically for the NULA comprised of N_b subarrays* ULA(N,d)*, as in Fig. 5.3, with any fixed element spacing $d>0$ and any distinct users' AoAs, $\theta_1\neq\theta_i$, $|\theta_1|\leq\theta_{max}$, if:*

$$\text{(a)}\;\; N_b>\lfloor d(1+\sin\theta_{max})\rfloor \;\text{ and }\; \text{(b)}\;\; \Delta D=pd/N_b, \tag{5.20}$$

where p is a positive integer coprime with N_b, i.e., $\gcd(p,N_b)=1$*, and there is no backward radiation.*

Proof. From (5.16), the FP holds if $\alpha_{bi}(\phi_k)=0\;\forall k\in I_k$. To ensure this, we use (5.18) and find a proper ΔD so that

$$\text{(i)}\;\; \sin(\pi N_b k\Delta D/d)=0\;\;\&\;\;\text{(ii)}\;\; \sin(\pi k\Delta D/d)\neq 0 \tag{5.21}$$

for all $k\in I_k$. (i) is equivalent to $N_b\Delta D/d$ being an integer:

$$N_b\Delta D/d=p \quad\Rightarrow\quad \Delta D=pd/N_b,\quad p=1,2,\cdots \tag{5.22}$$

However, one has to ensure (ii) as well with the following implication:

$$\sin(\pi k\Delta D/d)\neq 0 \;\Leftrightarrow \frac{k\Delta D}{d}=\frac{pk}{N_b}\neq n' \quad\forall k\in I_k \tag{5.23}$$

where $n'\in\{\pm1,\pm2,\cdots\}$ and the equality follows from (5.22). To this end, we show that (5.20) imply $pk/N_b\neq n'\;\forall k\in I_k$. We need the following technical Lemmas.

Lemma 2. *[97, p. 231, fact 5]*

If a and b are integers with $\gcd(a,b) = d$, then $\gcd(a/d, b/d) = 1$.

Lemma 3. *If a and b are integers with $\gcd(a,b) = 1$ and c is a divisor of a, i.e., $\gcd(a,c) = c$, then $\gcd(a/c, b) = 1$*

Lemma 4. *[97, p. 231, fact 9]*

If $a, b,$ and c are integers with $\gcd(a,b) = \gcd(a,c) = 1$, then $\gcd(a, bc) = 1$.

Now, assume that (5.20) holds and let p_k be the greatest common divisor of N_b and $|k|$:

$$\gcd(N_b, |k|) = p_k \tag{5.24}$$

Using (5.20)(a) and $|k| \leq \lfloor d(1 + |\sin\theta_1|)\rfloor$, which follows from (5.8) and (5.9), one obtains

$$p_k \leq |k| \leq \lfloor d(1 + |\sin\theta_1|)\rfloor < N_b \tag{5.25}$$

since $|\theta_1| \leq \theta_{max} \leq \pi/2$, which implies $N_b > |k|\ \forall k \in I_k$. Using (5.24) and Lemma 1, one obtains

$$\gcd(N_b, |k|) = p_k \quad \Rightarrow \quad \gcd(N_b/p_k, |k|/p_k) = 1 \tag{5.26}$$

Using $\gcd(N_b, p) = 1$ and Lemma 3,

$$\gcd(N_b, p) = 1 \quad \Rightarrow \quad \gcd(N_b/p_k, p) = 1 \tag{5.27}$$

where p_k is a divisor of N_b. Next, using (5.26), (5.27), and Lemma 4, one obtains:

$$\gcd(N_b/p_k, p|k|/p_k) = 1 \tag{5.28}$$

which means that N_b/p_k and $p|k|/p_k$ are co-prime numbers and therefore their ratio is not an integer,

$$\frac{p|k|/p_k}{N_b/p_k} = \frac{p|k|}{N_b} \neq |n'| \tag{5.29}$$

This proves (5.23) and hence Theorem 2. The no backward radiation condition is required to eliminate the opposite AoAs θ_1 and $\pi - \theta_1$, as they cannot be discriminated by a NULA made of isotropic elements. It is always satisfied if $-\pi/2 \leq \theta_1,\ \theta_i \leq \pi/2$, as assumed here. □

Intuitively, the condition on the number of subarrays N_b, given in (5.20)(a), is needed to make sure that there are enough nulls in the block array factor to cancel all GLs. The block array factor is the array factor of a ULA(N_b, D), where each subarray is replaced by an isotropic element. The condition (5.20)(b) on ΔD is necessary to ensure that nulls of the block array align with the subarray GLs and cancel them. Note that if p is not coprime with N_b, then some nulls may not exist.

To verify Theorem 2, we apply it to the example in Fig. 5.2. The number of blocks should satisfy $N_b > \lfloor 2d \rfloor = 1$, so that the setting $N_b = 25, p = 22$, where p is coprime with N_b and $\Delta D = pd/N_b$, meets all the conditions of Theorem 2. The results are illustrated in Fig. 5.2, where the only grating lobe is partially canceled even with finite $N_b N = 100$. It should be noted that this selection of $N_b, p, \Delta D$ ensures cancellation of all the grating lobes for any beam steering, not just $\theta_1 = 60^o$.

A large d requires a large number of subarrays, N_b, but the same N_b and p can accommodate various d; under this condition, ΔD is proportional to d. Note that the above designs are independent of θ_1, i.e., they are not affected by the beam steering, although the actual directions and number of GLs do depend upon it. Due to this, beam steering can be accomplished without requiring changes to the array geometry, thus providing the flexibility required for the design. Using the superposition principle, all of the results above hold true for multipath and blocked-LOS channels, as long as different users have distinct AoAs, with each user being able to have multiple AoAs to represent multiple paths now. Since N_b and p in (5.20) are not unique, they can be further optimized to improve the performance for finite N.

5.4 Extension to Wideband Channels

While the above analysis and design are valid for single carrier frequency, which is typical for antenna arrays, it remains to be seen whether they are still applicable to wideband or ultra-wideband systems. These systems are typical in both 5G and 6G systems, which utilize multiple sub-carriers, such as wideband OFDM or bandwidth aggregation. In this section, we demonstrate that the proposed NULA design is highly robust in the frequency domain, making it suitable for wideband and ultra-wideband channels.

Let's consider a LOS channel operating in the frequency range $[f_1, f_2]$, which corresponds to the wavelength interval $[\lambda_2, \lambda_1]$. Here, d and ΔD represent physical distances in meters, and $\mathrm{ULA}(N, d)$ denotes an N-element ULA with a physical element spacing of d. We impose no bandwidth limitation. The following proposition extends Theorem 2 to a wideband setting.

Proposition 6. *Consider a LOS channel operating over a (wide) frequency range* $[f_1, f_2]$ *under the conditions of Theorem 2. Then, favorable propagation holds asymptotically over the entire frequency range for a NULA comprised of* N_b *subarrays* $\mathrm{ULA}(N, d)$ *with subarray spacing* ΔD *if:*

$$\text{(a)}\ \ N_b > \lfloor 2d/\lambda_2 \rfloor \ \text{ and } \ \text{(b)}\ \ \Delta D/d = p/N_b, \tag{5.30}$$

where d *is the physical element spacing and* p *is a positive integer coprime with* N_b*.*

Proof. Let $K(\lambda)$ be the number of grating lobes at wavelength $\lambda \in [\lambda_2, \lambda_1]$. Using (5.9),

$$K(\lambda) \leq \lfloor 2d/\lambda \rfloor \leq \lfloor 2d/\lambda_2 \rfloor, \quad \forall \lambda \geq \lambda_2 \tag{5.31}$$

In particular, if $d < \lambda_2/2$, there are no grating lobes for any $\lambda \geq \lambda_2$. This is remarkable robustness property as it also applies to wide and ultra-wideband channels. Now, (5.30)(a) ensures that there are enough nulls in the block array factor to cancel all grating lobes.

Condition (5.30)(b) ensures that those nulls do align with GL directions. It follows directly from (5.20)(b) since it is frequency-independent (both sides can be multiplied by λ to obtain physical distances). □

Comparing (5.30) to (5.20), note that N_b should to be set based on the highest frequency or shortest wavelength so that it works for the whole frequency range; there is no any specific limitation related to bandwidth here.

To illustrate Proposition 6 for a finite number of elements, we apply it to *Example 1* above with $d = \lambda_2/2, \theta_1 = 90°$ (this corresponds to an endfire array, which are often used in practice, see e.g. [61, Sec. 2.5][62, Sec. 9.2][98][99, Sec. 6.3.2]), and set $f_1 = 0.8f_2$. It follows from (5.30) that $N_b > \lfloor 2d/\lambda_2 \rfloor = 1$ so that setting $N_b = 5, p = 4$ (coprime with N_b), $\Delta D = pd/N_b = 2\lambda_2/5$ satisfies the conditions of Proposition 6. Fig. 5.4 shows that this design partially cancels the GL even with finite $NN_b = 25$. It should be pointed out that this selection of $N_b, p, \Delta D$ guarantees GL cancellation for any main beam direction θ_1 (not only $\theta_1 = 90^o$) and any $f \leq f_2$, i.e., for a wideband or ultra-wideband channel. Fig. 5.4 suggests that even the single carrier frequency design at f_1 can be significantly improved (better GL cancellation) by designing at a slightly higher frequency f_2 instead, which corresponds to using a smaller d.

5.5 Extension to Directional Elements

It is straightforward to extend the above results to directional elements (rather than omni-directional ones), as is typical in practice. Real radiators, and thus the total array factor, are not isotropic. This is due to the multiplication theory, where the total array factor, $F(\theta)$, is the multiplication of the array factor with isotropic elements, $F_a(\theta)$, and the beam pattern of one non-isotropic element, $F_e(\theta)$, so $F(\theta) = F_a(\theta) \cdot F_e(\theta)$ [61, section 2.8, p. 76][99, p. 386]. However, element patterns, which usually have $F_e(\theta) = \cos\theta$ pattern such as a short dipole element [99, p. 165] (other weakly-directional elements, e.g. microstrip

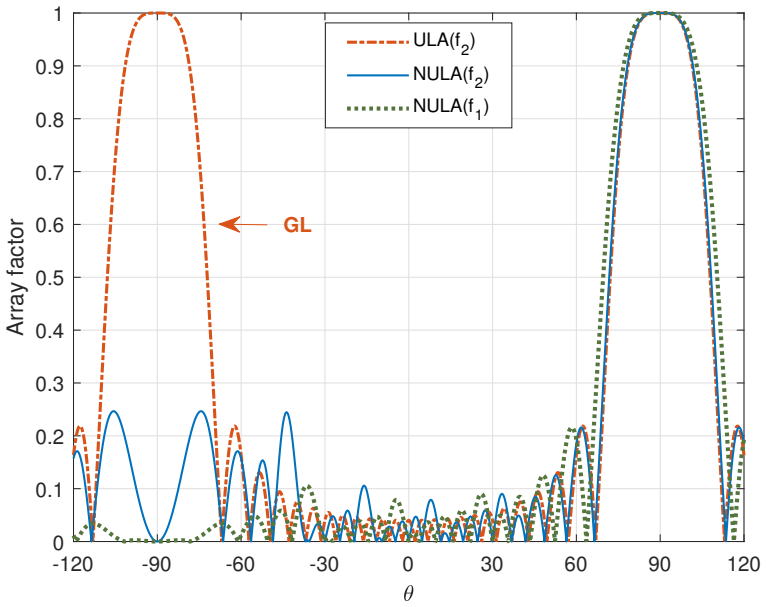


Figure 5.4: Array factors of ULA$(25, d)$ and the NULAs at corresponding frequencies, with $f_1 = 0.8f_2$, and $d = \lambda_2/2$. While the main beam is at $\theta_1 = 90^o$, note the presence of a grating lobe in the ULA factor at $\theta_2 = -90^o$. The proposed NULA design with $N = 5,\ N_b = 5, p = 4$ partially cancels this grating lobe over the entire frequency range, with better cancellation at f_1.

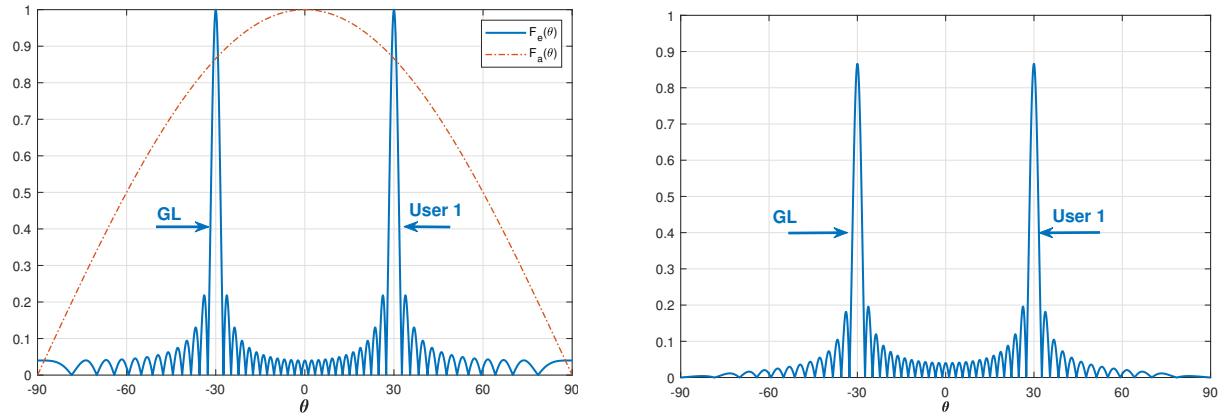
patches, have similar patterns [62]), have a limitted impact on the FP. This is because they are not highly directional and usually have a wide pattern with nulls at endfire directions ($\theta = \pm 90°$). So, they help in cancelling grating lobes at $\pm 90^o$ but not in other directions. For example, refer to Fig. 5.5, where the element does not have an isotropic pattern and follows a $\cos(\theta)$ pattern.

Also, the grating lobes appear within the scanning range of $\pm\theta_0$, if the element spacing is grater than below threshold:

$$d \geq \frac{1}{1 + \sin(\theta_0)} \tag{5.32}$$

For example, for the scanning range of $\theta_0 \pm 60^o$, grating lobes will appear if $d \geq 0.54$, see Fig. 5.6.

Hence, for an array with a large number of elements, the main directivity is attributed to the array factor [99, p. 386]. That is why the existing literature also adopts the same



(a) (b)

Figure 5.5: An array with non-isotropic elements. Part (a) shows the pattern of the array factor $F_a(\theta)$ and the element factor $F_e(\theta)$ seperately, while part (b) demonstrates the total non-isotropic array factor. Notice the relatively small contribution of the element pattern in decreasing the grating lobe magnitude in (b).

approach and assumes that the elements are isotropic [7][46][48]. See Fig. 5.7 for an example in which the elements are not isotropic. It can be observed that non-isotropic elements can decrease the grating lobe effect, but their contribution is very limited. The grating lobe remains fixed for ULA's as the total number of elements increases, though we can adjust the magnitude of NULA to a desired value by having enough elements.

It is straightforward to see that $\alpha'_{iN} = \alpha_{iN} \cos\theta_1 \cos\theta_i$, where α'_{iN} is the inter-user interference leakage term under directional elements, and α_{iN} is that for isotropic elements. Clearly, $\alpha'_{iN} \to 0$ if and only if $\alpha_{iN} \to 0$, exept for $\cos\theta_i = 0$ (endfire directions only). Therefore, weakly-directional elements can reduce grating lobes to some extent but cannot eliminate them completely, and thus, they have no impact on the FP, except for endfire/backplane directions.

5.6 Conclusion

Our analysis demonstrates that the presence of grating lobes in a uniform linear array can lead to violations of the conditions for favorable propagation, even when users have distinct angles of arrival. A block-partitioned nonuniform linear array is proposed to cancel

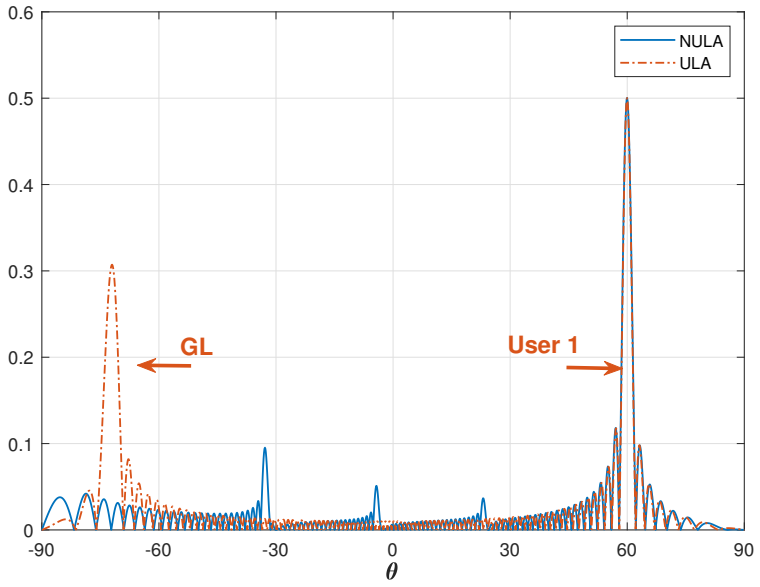


Figure 5.6: The pattern of a ULA and NULA with 100 non-isotropic elements and an element spacing of $d = 0.55$, where the NULA consists of 25 blocks with $p = 22$.

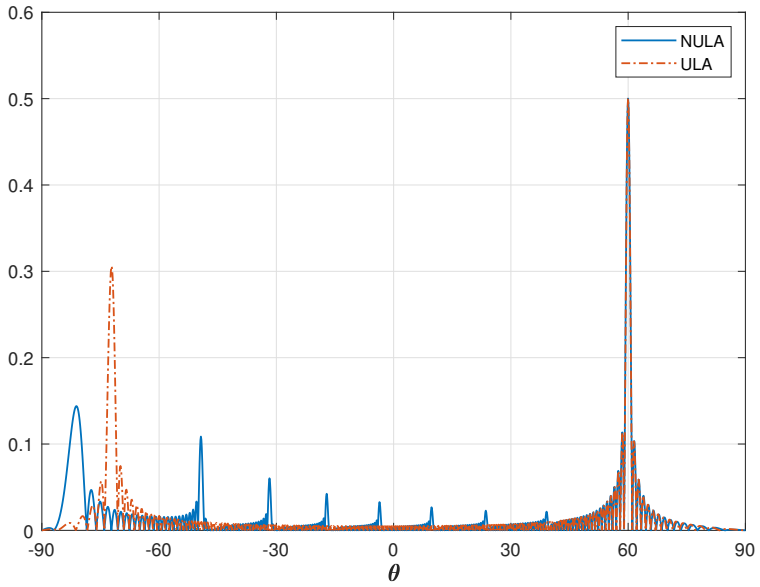


Figure 5.7: The pattern of a ULA and NULA with 200 non-isotropic elements and an element spacing of $d = 0.55$, where the NULA consists of 25 blocks with $p = 21$.

grating lobes and to restore favorable propagation for all distinct angles of arrival. Each block (subarray) in the NULA design is a ULA, but the overall array is non-uniform. The number and spacing of subarrays (blocks) are carefully selected so that, asymptotically, GLs are cancelled by nulls in the block array factor. The above analysis and design are shown to be remarkably robust in the frequency domain so that they can be applied to wide-band systems as well.

Chapter 6

Robustness of Favorable Propagation to Location and Phase Errors

6.1 Introduction

In favorable propagation studies [7][46][48][51][56][57], it is typically assumed that the beamforming weights are accurate, and that the channel information/estimation, as well as the knowledge of element locations or array calibration, are perfect. In practice, however, such a perfect setting is difficult to achieve due to implementation errors and tolerances. Thus, it remains to be seen whether the FP property still holds under these inaccuracies if the number of antennas increases without bound.

Robustness is crucial in practical scenarios, as it ensures that performance remains largely unaffected by minor errors or perturbations in element locations and beamforming weights. These errors often arise due to implementation inaccuracies, quantization effects, imperfect array calibration, component aging, environmental factors, etc. The significance of this becomes even more apparent in mmWave and THz systems, where smaller wavelengths exacerbate implementation errors. Phase noise and also quantization of phase shifts, on the other hand, pose significant performance issues in these systems. Therefore,

under these perturbations, it is unclear whether the FP property still holds in mMIMO settings. Since, due to the possibility of small per-element error aggregation when increasing the number of antennas to very large values, a large aggregate effect may also be generated, destroying the FP property.

In this chapter, we study the robustness of favorable propagation in mMIMO by taking into account random errors in element locations and beamforming phases. Both Gaussian and non-Gaussian error distributions are allowed and we answer the following three questions:

1. Does the FP hold under random errors? If so, under what conditions?
2. What is the IUI scaling with the number of antennas under random errors?
3. How accurate does a design need to be for the impact of random errors to be negligible?

Our results show that, for arbitrary array geometries, including ULA, UPA, UCA and UCLA as special cases, the FP holds under random independent errors as long as it holds for the nominal array (i.e., the one without errors). From a practical perspective, this means that, even if an array has a large number of elements, small inaccuracies do not "blow up". Hence, FP still holds and small IUIs are achievable if the number of antennas is large enough.

We also demonstrate that while random errors do not affect the FP asymptotically, they do have a pronounced negative effect on the convergence speed as the number of antenna elements increases. For example, in a ULA, the IUI power scales as N^{-2} for the nominal array (no errors), but it scales as N^{-1} for the perturbed one. This means that random errors slow down the convergence speed from N^{-2}, 20 dB per decade, to N^{-1} (10 dB per decade). Hence, under random errors, to achieve the same low IUI as the nominal one, more antennas are needed.

Next, we define a metric to determine small and large perturbation regimes and show that for a finite number of antennas, the impact of random errors is qualitatively different in these regimes. To achieve a negligible adverse impact, a quantitative condition is derived, which also determines the required implementation accuracy for a given number of antennas and IUI level.

6.2 System with Perturbations

Let us consider a Gaussian MIMO channel, as defined in (3.2), with an N-element BS serving M single-antenna users simultaneously. The base station array has perturbations in both the antenna locations and the beamforming phase. We consider linear processing (matched filter beamforming) weights $\boldsymbol{w}$ tuned to the nominal element locations, and thus, the SINR is obtained as in (3.4).

To determine the FP property in the case of random errors, we cannot rely on (3.14) anymore because $\alpha_{1i,N}$ becomes a random sequence, making the limits in (3.14) non-existent in the deterministic sense. Therefore, in the next section, we will extend the FP condition in (3.14) to account for random errors.

6.3 FP Under Location and Phase Errors

Let us consider an N-element antenna array with arbitrary geometry, including ULA, UPA, UCA, and UCLA as special cases. Random errors (perturbations) may occur in element locations as well as beamforming phases. In particular, the actual location vector $\boldsymbol{p}_n$ of n-th antenna array element is

$$\boldsymbol{p}_n = \boldsymbol{p}_n^0 + \Delta\boldsymbol{p}_n, \;\; n = 1...N \tag{6.1}$$

where $\boldsymbol{p}_n^0$ is the nominal location vector and $\Delta\boldsymbol{p}_n$ is its random offset, all measured in wavelengths. Here, we consider the LOS environment as in [7][46][48][49][56][57]. Under LOS propagation in the far field, the normalized channel vector entries for User i can be expressed as [7][61]

$$h_{in} = \exp(j2\pi \boldsymbol{u}_i^+ \boldsymbol{p}_n),\ i = 1..M, \quad n = 1...N, \tag{6.2}$$

where i and n are user and element indexes, $\boldsymbol{u}_i$ is the unit direction vector for User i. To detect user 1 signal, we use matched filtering with the beamforming weights $\boldsymbol{w}_1 = [w_1, ..., w_N]'$, tunned to User 1 nominal (rather than actual) channel vector $\boldsymbol{h}_1^0$:

$$\boldsymbol{h}_1^0 = [h_{11}^0, ..., h_{1N}^0]', \quad h_{1n}^0 = \exp(j2\pi \boldsymbol{u}_1^+ \boldsymbol{p}_n^0), \tag{6.3}$$

The beamforming weights are perturbed in phase, typically due to imperfect phase shifters and quantization errors, so that

$$w_n = \exp(j2\pi \boldsymbol{u}_1^+ \boldsymbol{p}_n^0 + j\Delta\phi_n) \tag{6.4}$$

where $\Delta\phi_n$ are beamforming phase errors. Both $\Delta\phi_n$ and $\Delta\boldsymbol{p}_n$ are modeled as zero-mean i.i.d. random vectors as widely-accepted in the literature [61][77][78][86]. The existing studies typically use Gaussian distribution to model these errors, as verified experimentally in [87, Fig. 4.13]. However, we consider both Gaussian and non-Gaussian distributions here, as distortion from non-linearities is generally modeled as non-Gaussian [86]. Using (3.7) and (6.1)-(6.4), $\alpha_{1i,N}$ can be expressed as

$$\alpha_{1i,N} = \frac{1}{N}\sum_{n=1}^{N} w_n^* h_{in} = \frac{1}{N}\sum_{n=1}^{N} e^{j\Psi_{in}}, \tag{6.5}$$

$$\Psi_{in} = \Psi_{in}^0 + \Delta\Psi_{in} \tag{6.6}$$

$$\Psi_{in}^0 = 2\pi(\boldsymbol{u}_i - \boldsymbol{u}_1)^+ \boldsymbol{p}_n^0 \tag{6.7}$$

$$\Delta\Psi_{in} = \Delta\psi_{in} - \Delta\phi_n \tag{6.8}$$

where Ψ_{in}^0 represents phase differences of i-th and 1st users' signals for the nominal array; $\Delta\psi_{in} = 2\pi \boldsymbol{u}_i^+ \Delta \boldsymbol{p}_n$ represents uncompensated phase shifts due to element location errors. The corresponding IUI leakage factor of the nominal array is

$$\alpha_{1i,N}^0 = \frac{1}{N} \sum_{n=1}^{N} e^{j\Psi_{in}^0} \tag{6.9}$$

To analyze favorable propagation under random errors, we cannot use the FP performance metric defined in (3.16) because $|\alpha_{1i,N}|^2$ is a random sequence indexed by N, and thus, the respective limits do not exist. Therefore, we have to use the notion of stochastic convergence. The following definition provides three different notions widely used in many areas of stochastic analysis and applications; this is a slight extension of the definitions in [106, p. 306][107, Ch. 6].

Definition 1. *A sequence $z_1, z_2, ...$ of random variables converges to a deterministic sequence $a_1, a_2, ...$ in the mean square sense (mse), $z_N \xrightarrow{mse} a_N$, if*

$$\lim_{N\to\infty} \mathbb{E}\{|z_N - a_N|^2\} = 0 \tag{6.10}$$

z_N converges in probability to a_N, $z_N \xrightarrow{\Pr} a_N$, if, for any $\epsilon > 0$,

$$\lim_{N\to\infty} \Pr\{|z_N - a_N| > \epsilon\} = 0 \tag{6.11}$$

z_N converges to a_N almost surely (a.s.) or with probability one, $z_N \xrightarrow{a.s.} a_N$, if

$$\Pr\left\{\lim_{N\to\infty} (z_N - a_N) = 0\right\} = 1 \tag{6.12}$$

where the probability applies to the set of all events where the limit exists and equals to 0.

We will use $\to$, without a superscript, to denote stochastic convergence in all three senses above as well as the regular (deterministic) convergence when all related sequences are deterministic. Note that convergence in the mean square sense and almost sure convergence imply convergence in probability but the converse is not true in general, see [106,

p. 310] and [107, p. 130-132] for examples and further discussion.

Some remarks are in ordered as to why these 3 modes of convergence are necessary.

- The mean square error is a well-established tool in many areas of communications, signal processing and stochastic control, including robust beamforming [81], which makes its use here appropriate. If MSE convergence holds, the random IUI z_N is ensured to converge to its mean value a_N.

- $\Pr\{|z_N - a_N| > \epsilon\}$ is the probability that the deviation of random IUI from its mean is not small. It is similar to outage probability, which is widely used in many wireless communication applications. It becomes the outage probability if a system is designed based on the mean IUI, but the random IUI varies significantly from it.

- The almost sure condition in (6.12) is needed to guarantee that $|z_N - a_N|$ cannot become arbitrary large, or have a large deviation, for infinitely-many N, while the condition in the probability convergence (6.11) does not guarantee this [108, p. 237]. Such a guarantee, that $|z_N - a_N|$ becomes and *stays* small as N increases, is provided by (6.12). It means that the set of all events where this does not hold, has a combined probability measure of zero, i.e., extremely unlikely to be encountered in the real world. This approach comes closest to deterministic convergence, ensuring that once mMIMO performance is acceptable for a given number N_0 of antennas, it will also be acceptable for any $N > N_0$.

As we will show below, our results hold for all three modes of convergence, which ensures that they are not influenced by particular definitions.

Next, after replacing the deterministic limits in the FP definition in (3.16) with stochastic convergence modes, we redefine the FP definition under random perturbations as

$$|\alpha_{1i,N}|^2 \to 0 \ \forall i > 1, \quad \text{as } N \to \infty \tag{6.13}$$

The following Theorem shows that the FP property holds for a perturbed array under the i.i.d. perturbations defined above if it holds for its nominal array. Therefore, random errors do not have catastrophic effects on the FP. This becomes "if and only if" (iff) under certain conditions and holds for all 3 convergence modes above. Hence, it is insensitive to a particular mode.

Theorem 3. *Under i.i.d. perturbations, the FP property of the channel in (6.1)-(6.4) can be characterized as follows:*

1. *If perturbations are Gaussian: the FP holds for a perturbed array if and only if it holds for a nominal (unperturbed) one, i.e., for any $i > 1$,*

$$|\alpha_{1i,N}|^2 \to 0 \quad \textit{iff} \quad \lim_{N\to\infty} |\alpha^0_{1i,N}|^2 = 0 \tag{6.14}$$

2. *If perturbations are non-Gaussian: (6.14) holds if $c_i \triangleq \mathbb{E}\{e^{j\Delta\Psi_{in}}\} \neq 0$; if $c_i = 0$, then the FP always holds for a perturbed array and 2nd condition in (6.14) is not necessary.*

3. *The following convergence holds in all considered cases for $i \geq 1$:*

$$|\alpha_{1i,N}|^2 \to \mathbb{E}\{|\alpha_{1i,N}|^2\} \tag{6.15}$$

$$= |c_i|^2|\alpha^0_{1i,N}|^2 + (1-|c_i|^2)N^{-1} \tag{6.16}$$

$$\to |c_i|^2|\alpha^0_{1i,N}|^2 \tag{6.17}$$

□

A proof will be presented in the following sections. First, we establish (6.15)-(6.17), from which (6.14) will follow. To this end, we will establish the claims above for the MSE convergence as a first step in section 6.3.1 and next, we establish the convergence in probability and almost sure convergence in sections 6.3.2 and 6.3.3, respectively. To simplify notations, we further use $\alpha_N = \alpha_{1i,N}$, $\alpha^0_N = \alpha^0_{1i,N}$ for $i > 1$.

6.3.1 Convergence in the MSE Sense

The definition in (6.10) with $z_N = |\alpha_N|^2$ and $a_N = \mathbb{E}\{|\alpha_N|^2\}$ is equivalent to

$$\begin{aligned} \mathrm{Var}\{|\alpha_N|^2\} &= \mathbb{E}\{(|\alpha_N|^2 - \mathbb{E}\{|\alpha_N|^2\})^2\} \\ &= \mathbb{E}\{|\alpha_N|^4\} - (\mathbb{E}\{|\alpha_N|^2\})^2 \to 0 \end{aligned} \tag{6.18}$$

Finding the variance in (6.18) is rather involved since the fourth moment analysis is complicated. To overcome this difficulty, we establish an upper bound and prove that it converges to zero as $N \to \infty$. To this end, the following Lemma shows that $\mathbb{E}\{|\alpha_N|^4\}$ can be "sandwiched" via $|\,\mathbb{E}\{\alpha_N\}|^4$.

Lemma 5. $\mathbb{E}\{|\alpha_N|^4\}$ *can be bounded as follows:*

$$|\,\mathbb{E}\{\alpha_N\}|^4 \leq \mathbb{E}\{|\alpha_N|^4\} \leq |\,\mathbb{E}\{\alpha_N\}|^4 + 12N^{-1} \tag{6.19}$$

Proof. See Appendix A.2. □

It follows from (6.19) that $\mathbb{E}\{|\alpha_N|^4\}$ and $|\,\mathbb{E}\{\alpha_N\}|^4$ behave similarly as $N \to \infty$:

$$\mathbb{E}\{|\alpha_N|^4\} \to 0 \;\text{ iff }\; |\,\mathbb{E}\{\alpha_N\}|^4 \to 0 \tag{6.20}$$

The next Lemma establishes the required upper bound.

Lemma 6. $\mathrm{Var}\{|\alpha_N|^2\}$ *can be upper bounded, for any* N*, as follows:*

$$\mathrm{Var}\{|\alpha_N|^2\} \leq 12N^{-1} \tag{6.21}$$

Proof. Since $|\cdot|^2$ is convex, it follows from Jensen's inequality [111] that

$$\mathbb{E}\{|\alpha_N|^2\} \geq |\,\mathbb{E}\{\alpha_N\}|^2 \tag{6.22}$$

Combining this with the upper bound in (6.19), one obtains

$$\begin{aligned}\mathrm{Var}\{|\alpha_N|^2\} &= \mathbb{E}\{|\alpha_N|^4\} - (\mathbb{E}\{|\alpha_N|^2\})^2 \\ &\leq |\,\mathbb{E}\{\alpha_N\}|^4 + 12N^{-1} - |\,\mathbb{E}\{\alpha_N\}|^4 = 12N^{-1} \end{aligned} \tag{6.23}$$

as required. □

Now, using (6.21) and (6.18),

$$\mathrm{Var}\{|\alpha_N|^2\} = \mathbb{E}\{(|\alpha_N|^2 - \mathbb{E}\{|\alpha_N|^2\})^2\} \leq 12N^{-1} \to 0 \tag{6.24}$$

which establishes the MSE convergence,

$$|\alpha_N|^2 \xrightarrow{mse} \mathbb{E}\{|\alpha_N|^2\} \tag{6.25}$$

meaning that (6.15) holds in the MSE sense. Next, we use the following technical Lemma to establish (6.16).

Lemma 7. *The following holds for $\alpha_N = \alpha_{1i,N}$ as defined in (6.5):*

$$\mathbb{E}\{\alpha_N\} = c_i \alpha_N^0, \quad c_i = \mathbb{E}\{e^{j\Delta\Psi_{in}}\} \tag{6.26}$$

$$\mathbb{E}\{|\alpha_N|^2\} = (1 - |c_i|^2)N^{-1} + |c_i|^2 |\alpha_N^0|^2 \tag{6.27}$$

Proof. Let us first evaluate $\mathbb{E}\{|\alpha_N|^2\}$ using (6.5):

$$\begin{aligned}\mathbb{E}\{|\alpha_N|^2\} &= \frac{1}{N^2}\sum_{n,m}\mathbb{E}\{e^{j(\Psi_{in}-\Psi_{im})}\} \\ &= \frac{1}{N} + \frac{1}{N^2}\sum_{m\neq n}\mathbb{E}\{e^{j\Psi_{in}}\}\,\mathbb{E}\{e^{-j\Psi_{im}}\} \qquad (6.28) \\ &= \frac{1}{N} + \frac{1}{N^2}\sum_{n}\mathbb{E}\{e^{j\Psi_{in}}\}\sum_{m}\mathbb{E}\{e^{-j\Psi_{im}}\} - \frac{1}{N^2}\sum_{n}\mathbb{E}\{e^{j\Psi_{in}}\}\,\mathbb{E}\{e^{-j\Psi_{in}}\} \\ &= \frac{1}{N} + |\,\mathbb{E}\{\alpha_N\}|^2 - \frac{1}{N^2}\sum_{n}|\,\mathbb{E}\{e^{j\Psi_{in}}\}|^2 \qquad (6.29)\end{aligned}$$

where (6.28) follows from independence of Ψ_{in} and Ψ_{im} for $n \neq m$. Next, using (6.5) and (6.9), one obtains:

$$\mathbb{E}\{\alpha_N\} = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\{e^{j\Psi_{in}}\} = \frac{c_i}{N}\sum_{n=1}^{N} e^{j\Psi^0_{in}} = c_i\alpha^0_N \tag{6.30}$$

where

$$\mathbb{E}\{e^{j\Psi_{in}}\} = \mathbb{E}\{e^{j(\Psi^0_{in}+\Delta\Psi_{in})}\} = c_i e^{j\Psi^0_{in}} \tag{6.31}$$

Using (6.29)-(6.30), (6.27) follows:

$$\mathbb{E}\{|\alpha_N|^2\} = N^{-1} + |c_i|^2|\alpha^0_N|^2 - N^{-1}|c_i|^2 \tag{6.32}$$

This completes the proof of Lemma 7. □

Finally, note that (6.27) implies the following convergence as $N \to \infty$:

$$\mathbb{E}\{|\alpha_N|^2\} \to |c_i|^2|\alpha^0_N|^2 \tag{6.33}$$

This establishes (6.15)-(6.17), where perturbations do not have to be Gaussian (no such assumption was made in the above derivations). For Gaussian perturbations, from (6.51), $c_i = \mathbb{E}\{e^{j\Delta\Psi_{in}}\} = e^{-\delta^2/2} \neq 0$ and therefore, from (6.33),

$$\lim_{N\to\infty}\mathbb{E}\{|\alpha_N|^2\} = 0 \quad \text{iff} \quad \lim_{N\to\infty}|\alpha^0_N|^2 = 0 \tag{6.34}$$

and, since $|\alpha_N|^2 \xrightarrow{mse} \mathbb{E}\{|\alpha_N|^2\}$ as established in (6.25) above, (6.14) follows (in the MSE sense):

$$|\alpha_N|^2 \xrightarrow{mse} 0 \quad \text{iff} \quad \lim_{N\to\infty}|\alpha^0_N|^2 = 0 \tag{6.35}$$

For non-Gaussian perturbations, $c_i = 0$ is possible, in which case, from (6.25) and (6.27),

$$|\alpha_N|^2 \xrightarrow{mse} \mathbb{E}\{|\alpha_N|^2\} \to 0 \tag{6.36}$$

even if $\lim_{N\to\infty} |\alpha_N^0|^2 \neq 0$, i.e., the condition $\lim_{N\to\infty} |\alpha_N^0|^2 = 0$ is not necessary. This completes the proof for the MSE convergence.

6.3.2 Convergence in Probability

To establish (6.15) with convergence in probability, we use (6.11) with $z_N = |\alpha_N|^2$, $a_N = \mathbb{E}\{|\alpha_N|^2\}$ and Chebyshev inequality [106, p. 46]:

$$\Pr\{|z_N - \mathbb{E}\{z_N\}| > \epsilon\} \leq \mathrm{Var}\{z_N\}\epsilon^{-2} \tag{6.37}$$

which holds for any $\epsilon > 0$. Using (6.37) and (6.21), one obtains:

$$\begin{aligned} \lim_{N\to\infty} \Pr\{||\alpha_N|^2 - \mathbb{E}\{|\alpha_N|^2\}| > \epsilon\} &\leq \lim_{N\to\infty} \mathrm{Var}\{|\alpha_N|^2\}\epsilon^{-2} \\ &\leq \lim_{N\to\infty} 12(N\epsilon^2)^{-1} = 0 \end{aligned} \tag{6.38}$$

This establishes (6.15) with convergence in probability,

$$|\alpha_N|^2 \xrightarrow{\Pr} \mathbb{E}\{|\alpha_N|^2\} \tag{6.39}$$

Combining it with (6.16) and (6.17) established above (note that they apply to the deterministic sequences and therefore do not depend on the stochastic convergence definitions), we finally obtain (6.14) with convergence in probability.

6.3.3 Almost Sure Convergence

To establish (6.15) with almost sure convergence[1], $|\alpha_N|^2 \xrightarrow{a.s.} \mathbb{E}\{|\alpha_N|^2\}$, we need the following technical Lemma, which is an extension of [109, Theorem 1] to two-dimensional sequences with non-zero means.

Lemma 8. *Consider the empirical mean z_N of N^2 uniformly-bounded random variables z_{nm},*

$$z_N = \frac{1}{N^2} \sum_{n,m=1}^{N} z_{nm} \tag{6.40}$$

Then, $z_N \xrightarrow{a.s.} \mathbb{E}\{z_N\}$ if:

$$\text{(a) } |z_{nm}| \leq 1 \text{ and (b) } \sum_{N=1}^{\infty} \frac{1}{N} \operatorname{Var}\{z_N\} < \infty \tag{6.41}$$

Proof. See Appendix A.3. □

Now, use Lemma 8 with $z_N = |\alpha_N|^2$,

$$z_N = |\alpha_N|^2 = \frac{1}{N^2} \sum_{n,m} e^{j(\Psi_{in} - \Psi_{im})} \tag{6.42}$$

which is the empirical mean of N^2 random variables $z_{nm} = \exp\{j\Psi_{in} - j\Psi_{im}\}$ as in (6.40). Observe that $|z_{nm}| \leq 1$, which satisfies (6.41)(a). Using (6.21), $\operatorname{Var}\{z_N\} \leq 12N^{-1}$ so that

$$\sum_{N\geq 1} N^{-1} \operatorname{Var}\{z_N\} \leq 12 \sum_{N\geq 1} N^{-2} = 2\pi^2 < \infty \tag{6.43}$$

i.e., (6.41)(b) is also satisfied. Therefore, $|\alpha_N|^2 \xrightarrow{a.s.} \mathbb{E}\{|\alpha_N|^2\}$, as required. This establishes (6.15) with almost sure convergence. Combining it with (6.16) and (6.17) established above, (6.14) follows with almost sure convergence.

[1] note that this convergence mode is not implied by the two convergence modes above [106].

6.4 The Distribution and a Bound of $|\alpha_N|^2$

Eq. (6.15) implies that we can use $\mathbb{E}\{|\alpha_N|^2\}$ to estimate the actual IUI, $|\alpha_N|^2 = |\alpha_{1i,N}|^2$, for large N. However, as shown in Fig. 6.3, $|\alpha_N|^2$ exhibits fluctuations for finite N and, hence, $|\alpha_N|^2$ can have values exceeding $\mathbb{E}\{|\alpha_N|^2\}$. We should consider this fluctuation possibility for a more reliable design with finite N. A simple way to address this is to provide an upper bound for $|\alpha_N|^2$, which holds with high probability for large N and sufficiently large m,

$$|\alpha_N|^2 \lessapprox |\alpha_N^{up}|^2 = \mathbb{E}\{|\alpha_N|^2\} + m\sigma_{|\alpha_N|^2} \tag{6.44}$$

where $\sigma^2_{|\alpha_N|^2} = \mathrm{Var}\{|\alpha_N|^2\}$ is the variance, and $m = 1..3$ controls the outage probability, (i.e., the probability that $|\alpha_N|^2$ exceeds its estimated value). The estimation in (6.44) with $m = 0, 1$ and 3 is illustratred in Fig. 6.3, where one can observe that using $m = 0$ is not sufficient and $m = 1..3$ provides more reliable design. Here, larger m corresponds to smaller outage probability.

The following Proposition gives an approximations for the distribution and variance of $|\alpha_N|^2$ with finite-N, which can be used in (6.44).

Proposition 7. *For large N and if the FP holds for the nominal (no error) array at double the element spacing, i.e.,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{j2\Psi^0_{in}} = 0, \tag{6.45}$$

$|\alpha_N|^2$ is distributed as a non-central chi-squared random variable $\chi^2_2(\lambda)$ with 2 degrees of freedom,

$$|\alpha_N|^2 \sim \frac{1}{2}\sigma^2_N \chi^2_2(\lambda) \tag{6.46}$$

where $\lambda = 2\sigma_N^{-2}|\,\mathbb{E}\{\alpha_N\}|^2$ *is the noncentrality parameter, and*

$$\lambda = 2N(1-|c_i|^2)^{-1}|c_i|^2|\alpha_N^0|^2 \tag{6.47}$$

$$\sigma_N^2 = \mathrm{Var}\{\alpha_N\} = N^{-1}(1-|c_i|^2) \tag{6.48}$$

$$\sigma^2_{|\alpha_N|^2} = \mathrm{Var}\{|\alpha_N|^2\} \approx \sigma_N^2(\sigma_N^2 + 2|c_i|^2|\alpha_N^0|^2) \tag{6.49}$$

Proof. See Appendix A.4. □

The accuracy of these approximations is examined in the next section.

6.5 Examples and Discussion

In this section, we illustrate and validate the above results. Let us consider a ULA with the nominal element spacing $d = 0.5$ measured in wavelengths, see Fig. 7.1. The array has perturbations in both locations and phaseshifts, where the errors are zero-mean, i.i.d. Gaussian of variances σ_p^2 and σ_ϕ^2, respectively. Since $\Delta\Psi_{in}$ is a linear combination of these normally-distributed errors, it is also a zero-mean Gaussian random variable of variance

$$\begin{aligned}\delta^2 &= \mathrm{Var}\{\Delta\Psi_{in}\}\\ &= \mathrm{Var}\{\Delta\psi_{in} - \Delta\phi_n\}\\ &= \mathrm{Var}\{2\pi\boldsymbol{u}_i^+\Delta\boldsymbol{p}_n - \Delta\phi_n\}\\ &= 4\pi^2\sigma_p^2 + \sigma_\phi^2\end{aligned} \tag{6.50}$$

and therefore (see e.g. [61, p. 68])

$$c_i = \mathbb{E}\{e^{j\Delta\Psi_{in}}\} = e^{-\delta^2/2} \tag{6.51}$$

so that $0 < c_i \leq 1$.

First, we validate the asymptotic behavior $\mathrm{Var}\{|\alpha_N|^2\} \to 0$, which is the key part of the Theorem 3 proof, and also the approximation in (6.49). Let us consider the ULA as in Fig.

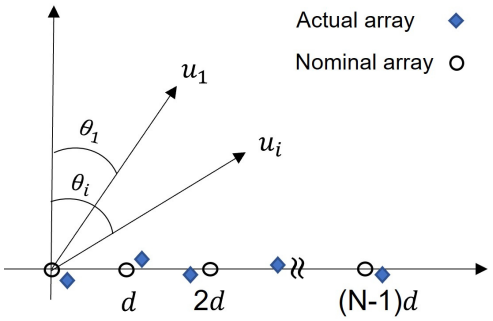


Figure 6.1: N-element ULA under perturbed element locations. θ_i is the AoA of User i, where $i = 1$ corresponds to the main user, all measured from the array broadside.

7.1. The behavior of $\text{Var}\{|\alpha_N|^2\}$ for increasing N is shown in Fig. 6.2, which confirms that $\text{Var}\{|\alpha_N|^2\} \to 0$, as expected. This also implies, from Chebyshev inequality, convergence in probability in (6.14). Note also that the empirical (Monte-Carlo (MC) simulated) variance and its estimate in (6.49) agree well with each other over the whole range of N.

$\text{Var}\{|\alpha_N|^2\}$ shows qualitatively-different behavior for smaller and larger $\sigma_{p,\phi}$, which can be explained using (6.49). In the large error regime, $\sigma_{p,\phi} = 0.1$, σ_N^2 is large and hence the first term in (6.49) dominates:

$$\sigma_N^4 \gg 2\sigma_N^2|c_i|^2|\alpha_N^0|^2 \quad \to \quad \text{Var}\{|\alpha_N|^2\} \approx \sigma_N^4 = \frac{(1-|c_i|^2)^2}{N^2} \tag{6.52}$$

which decreases monotonically with N. In the small error regime, $\sigma_{p,\phi} = 0.01$, the 2nd term in (6.49) dominates until about $N = 100$ and hence

$$\sigma_N^4 \ll 2\sigma_N^2|c_i|^2|\alpha_N^0|^2 \quad \to \quad \text{Var}\{|\alpha_N|^2\} \approx 2|c_i|^2|\alpha_N^0|^2\sigma_N^2 \tag{6.53}$$

which exhibits an oscillatory decrease with N due to $|\alpha_N^0|$.

6.5.1 Impact of Errors: Small and Large Perturbation Regimes

Large perturbation regime: The behavior of the IUI factors for increasing N in the large perturbation regime is illustrated in Fig. 6.3. We note, from the comparison of $|\alpha_N|^2$ and $|\alpha_N^0|^2$, that perturbations in this regime have a significant impact on the IUI and cannot be

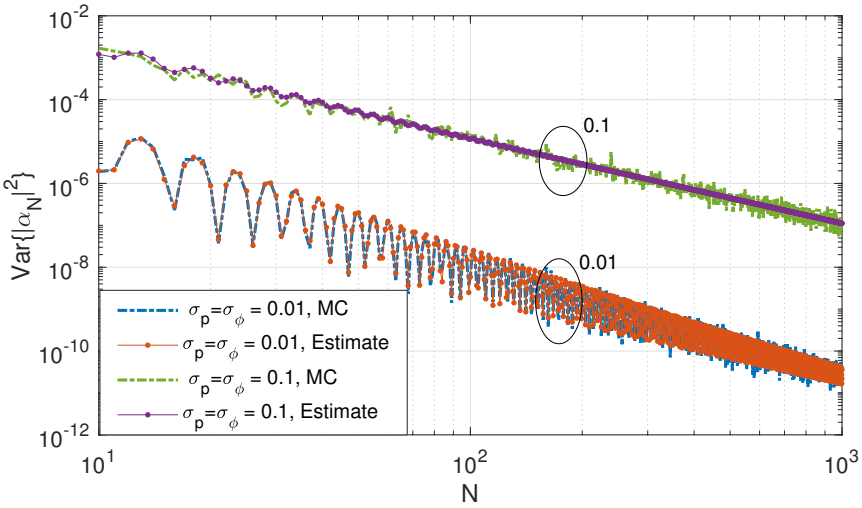


Figure 6.2: Empirical Monte-Carlo (MC) simulated $\mathrm{Var}\{|\alpha_N|^2\}$ and its estimate in (6.49) in the presence of zero-mean i.i.d. Gaussian errors of variances σ_p^2 and σ_ϕ^2, for the ULA with $d = 0.5$, the user angles-of-arrival are $\theta_1 = 0$, $\theta_i = \pi/8$ (with respect to the array broadside). The MC variance was evaluated over 100 trials generated independently for each N.

ignored for any finite N. It can be observed that all three IUI factors, namely, the random IUI $|\alpha_N|^2$, its mean $\mathbb{E}\{|\alpha_N|^2\}$, and the nominal $|\alpha_N^0|^2$ values, decrease with N, which is consistent with Theorem 3. Also, while $|\alpha_N|^2$ exhibits the slowest decrease and statistical fluctuations due to random location and phase errors, the nominal IUI $|\alpha_N^0|^2$ shows the fastest decrease. In general, there is significant difference between all three IUI factors. Given the significant statistical fluctuations around the mean in this regime, using the mean IUI as an estimate, as done in some studies, can lead to a significant underestimation of the actual IUI level. Similarly, the nominal IUI factor $|\alpha_N^0|^2$ is not a good approximation of $|\alpha_N|^2$ and significantly underestimates the actual IUI.

Small perturbation regime: The IUI factors for the small perturbation regime with $\sigma_p = \sigma_\phi = 0.01$ are depicted in Fig. 6.4. Unlike Fig. 6.3, all three factors in this case – the random IUI, its mean, and nominal values– exhibit similar behavior up to approximately $N = 100$. Hence, the impact of random errors is almost negligible within this interval. This can be justified using (6.15) and (6.16), where in the small perturbation regime, the first term of (6.16) dominates, for which $|c_i| \approx 1$ and therefore

$$(1 - |c_i|^2)N^{-1} \ll |c_i|^2|\alpha_N^0|^2 \tag{6.54}$$

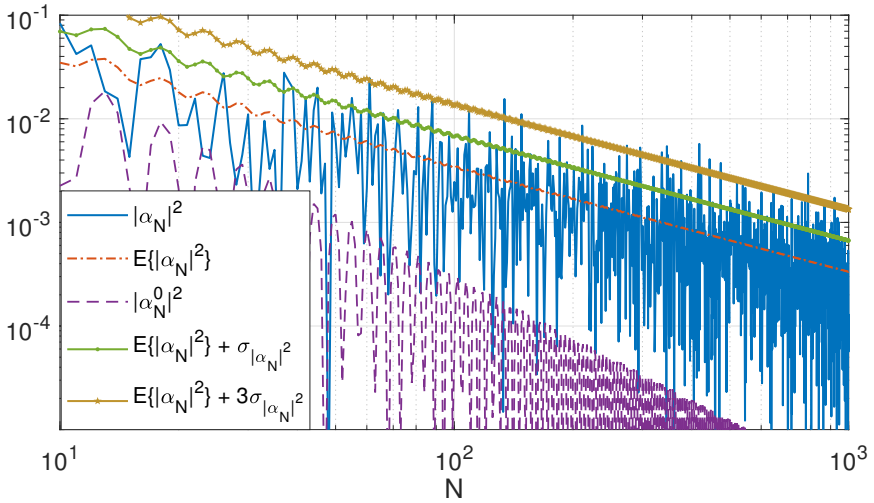


Figure 6.3: IUI factor $|\alpha_N|^2$, its mean and nominal values for the N-element ULA with nominal element spacing $d = 0.5$, under zero-mean i.i.d. Gaussian perturbations with $\sigma_p = \sigma_\phi = 0.1$ (generated independently for each N); $\theta_1 = 0$, $\theta_i = \pi/8$. While all decrease with N, $|\alpha_N|^2$ fluctuates due to random errors and $|\alpha_N^0|^2$ exhibits fastest decrease. Note that perturbations have non-negligible impact for any finite N.

so that, from (6.15),

$$|\alpha_N|^2 \approx \mathbb{E}\{|\alpha_N|^2\} \approx |\alpha_N^0|^2 \tag{6.55}$$

In this case, random perturbations have almost no impact because the nominal and random leakage factors are nearly identical. Utilizing (6.51) for Gaussian perturbations, (6.54) is equivalent to

$$\delta^2 \ll \ln(1 + N|\alpha_N^0|^2) \tag{6.56}$$

which represents the notion of the small perturbation regime, in which phase and location errors have negligible effects, and therefore, (6.55) holds. Consequently, the three IUI leakage factors are almost similar. Note that (6.56) does not imply that larger errors become tolerable as N increases or that arbitrarily large errors are permitted as $N \to \infty$. The reason is that α_N^0 also depends on N, and in many cases, $|\alpha_N^0|^2 \sim N^{-2}$. Consequently, the overall scaling of the upper bound in (6.56) is as $\ln(1 + N^{-1}) \sim N^{-1}$, which contrasts with a naive interpretation.

Note also that random errors affect both $|\alpha_N|^2$ and its mean value $\mathbb{E}\{|\alpha_N|^2\}$, as shown in (6.55) and Fig. 6.3 and 6.4. While in the small perturbation regime, $\mathbb{E}\{|\alpha_N|^2\} \approx |\alpha_N^0|^2$

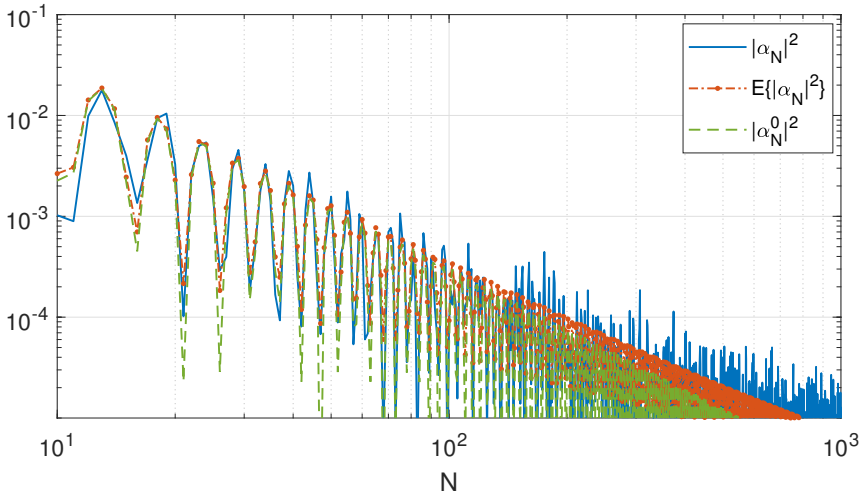


Figure 6.4: The IUI factors as in Fig. 6.3 for $\sigma_p = \sigma_\phi = 0.01$ (small perturbation regime). Note that, unlike Fig. 6.3, all 3 behave similarly until about $N = 100$ and the impact of random errors is negligible.

up to about $N = 100$ (see Fig. 6.4), this approximation no longer holds in the large perturbation regime (see Fig. 6.3) when

$$\mathbb{E}\{|\alpha_N|^2\} \approx (1 - |c_i|^2)N^{-1} \tag{6.57}$$

The observations above are not restricted to ULA geometry but also hold for other array geometries. This is also observed for uniform circular array (UCA) under i.i.d. Gaussian perturbations with $\sigma_p = \sigma_\phi = 0.1$, in Fig. 6.5. In this figure, we can see while the actual values of IUI are different from those for the ULA, the general tendencies are the same as in Fig. 6.4. The FP also holds for UCA under random perturbations because it holds for its nominal array, as established in [57]. This is consistent with Theorem 3, which holds for any array geometry. Comparing the UCA and ULA cases with the same error standard deviations in Fig. 6.5 and Fig. 6.3, we can conclude that the UCA is more robust to random errors than the ULA. This is due to the fact that $|\alpha_N^0|^2 \sim N^{-2}$ for the ULA but only $|\alpha_N^0|^2 \sim N^{-1}$ for the UCA, resulting in a smaller decrease with N. Therefore, random errors have a less significant impact on the UCA .

The impact of i.i.d. Gaussian perturbations both in the large and small perturbation

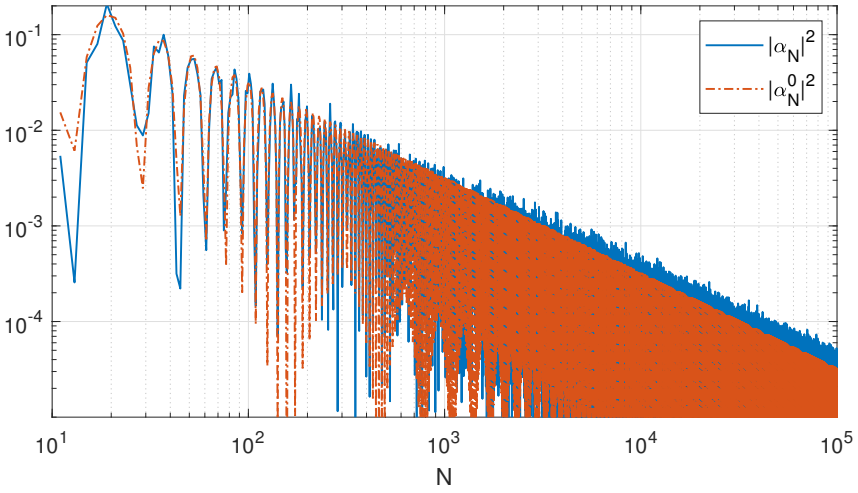


Figure 6.5: IUI factor $|\alpha_N|^2$ and its nominal values for the N-element UCA with nominal element spacing $d = 0.5$, under zero-mean i.i.d. Gaussian perturbations with $\sigma_p = \sigma_\phi = 0.1$; $\theta_1 = 0,\ \theta_i = \pi/8$. Note the same tendencies as in Fig. 6.4 and significantly smaller impact of errors compared to Fig. 6.3.

regimes for ULA with larger element spacing ($d = 1$) is illustrated in Figs. 6.6 and 6.7. Using (6.57), in the large perturbation regime ($\sigma_p = \sigma_\phi = 0.1$), we have $|\alpha_N|^2 \to \mathbb{E}\{|\alpha_N|^2\} \approx (1 - |c_i|^2)N^{-1}$. Importantly, this behavior is independent of the element spacing d. This can be observed by comparing Fig. 6.3 and Fig. 6.6, both of which depict the large perturbation regime with $d = 0.5$ for the former and $d = 1$ for the latter. While the array with a larger d has a smaller nominal IUI with an equal number of elements N, due to having a larger aperture size, the actual and expected IUI values are similar in both cases. However, in the small perturbation regime ($\sigma_p = \sigma_\phi = 0.01$), we have $|\alpha_N|^2 \to \mathbb{E}\{|\alpha_N|^2\} \approx |c_i|^2|\alpha_N^\circ|^2$, as indicated in (6.54). Hence, the convergence depends on the $|\alpha_N^\circ|^2$ and so on d. This can be observed in Fig. 6.7, where we can see that up to about $N = 100$, the condition in (6.56) is satisfied, resulting in all three graphs being similar. As N increases, the graph transitions out of the small perturbation regime because (6.56) is no longer valid. This transition occurs more rapidly with a larger d since a larger d implies a smaller nominal IUI for a fixed N. This can be observed by comparing Fig. 6.4 and Fig. 6.7.

It should be noted that the above observations also hold for non-Gaussian errors. The IUI factors under uniformly-distributed errors, with all other parameters identical to those

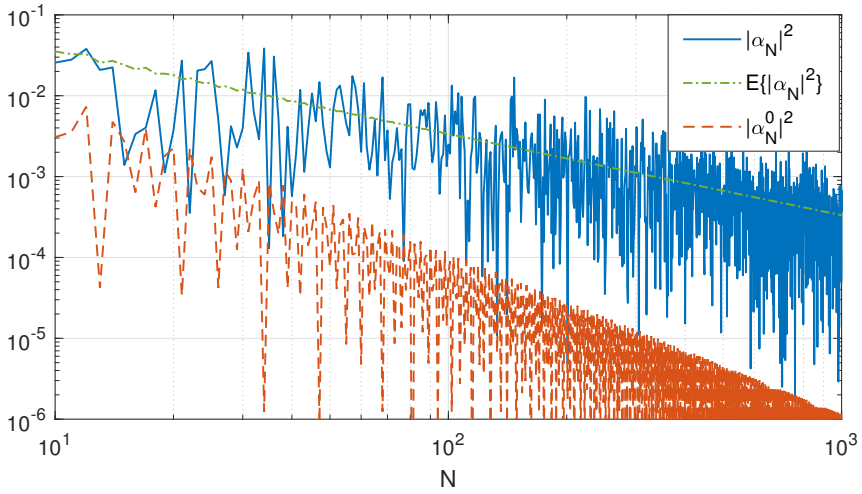


Figure 6.6: IUI factor $|\alpha_N|^2$, its mean and nominal values for the N-element ULA with nominal element spacing $d = 1$, under zero-mean i.i.d. Gaussian perturbations with $\sigma_p = \sigma_\phi = 0.1$; $\theta_1 = 0,\ \theta_i = \pi/8$.

in Fig. 6.4, are shown in Fig. 6.8. Comparison of these two figures shows that error impacts are almost the same for both. Therefore, we conclude that random errors are more influenced by their standard deviation than by their shape of distribution. This also agrees with Theorem 3 and Proposition 7, which hold for any error distribution as long as the errors are i.i.d. and of finite variance.

6.5.2 IUI Scaling With N and Design Guidelines

It follows from Theorem 3 that, if the FP holds for the nominal array, then it also holds for the perturbed one and all three IUI factors converge to zero as $N \to \infty$,

$$|\alpha_N|^2,\ \mathbb{E}\{|\alpha_N|^2\},\ |\alpha_N^0|^2 \to 0 \tag{6.58}$$

One should note that, in spite of their similar convergence points, their convergence speeds differ significantly. As an example, for a ULA that has fixed element spacing, distinct AoAs, and no grating lobes, the nominal array's convergence is scaled as

$$|\alpha_N^0|^2 \sim N^{-2} \tag{6.59}$$

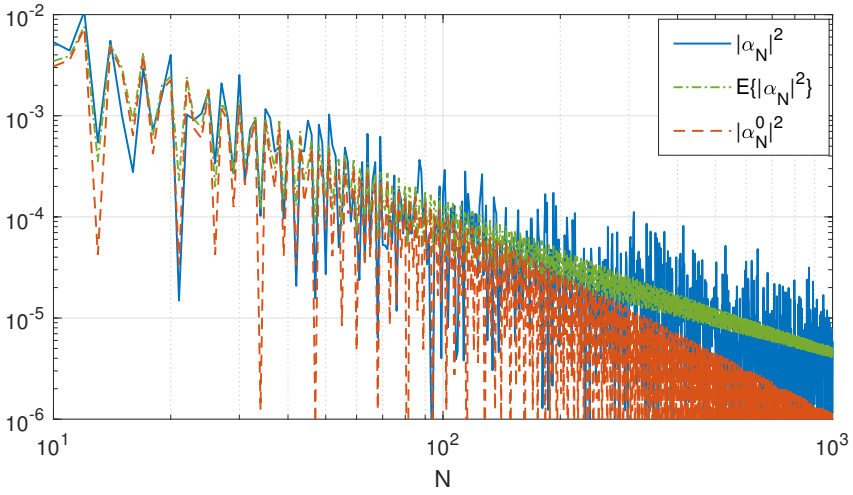


Figure 6.7: IUI factor $|\alpha_N|^2$, its mean and nominal values for the N-element ULA with nominal element spacing $d = 1$, under zero-mean i.i.d. Gaussian perturbations with $\sigma_p = \sigma_\phi = 0.01$; $\theta_1 = 0,\ \theta_i = \pi/8$.

meaning that the IUI decreases 20 dB per decade, while for the perturbed ULA, it is

$$|\alpha_N|^2,\ \mathbb{E}\{|\alpha_N|^2\} \sim N^{-1} \tag{6.60}$$

i.e., 10 dB per decade. In other words, the random errors impact is to slow down the convergence speed from N^{-2} to N^{-1}, even if the FP holds. In this case, more antennas are required under random errors to achieve the same low IUI leakage as the nominal array. Further confirmation of this result can be seen in Fig. 6.3, where at $N = 100$, the IUI is about -20 dB under random errors. This result is in agreement with the respective experimental results in [13, Fig. 11]. In addition, we compare the above scalings with those in [56]. In their study, it was observed that in a LOS channel with uniformly-random users (distributed on a sphere) and without any errors, the IUI power scales as $N^{-2}\ln^2 N \approx N^{-2}$, whereas in an i.i.d. Rayleigh fading channel, it scales only as N^{-1}. However, (6.60) indicates that in the presence of random location and phase errors, the scaling in the LOS channel slows down to N^{-1}, which is similar to a fading channel.

We note that (6.56) may be challenging to apply in practice. This is primarily because $|\alpha_N^0|^2$ depends on user AoAs, which may either be unknown or known with limited precision. From a practical perspective, a sensible design should tolerate *any* IUI leakage not

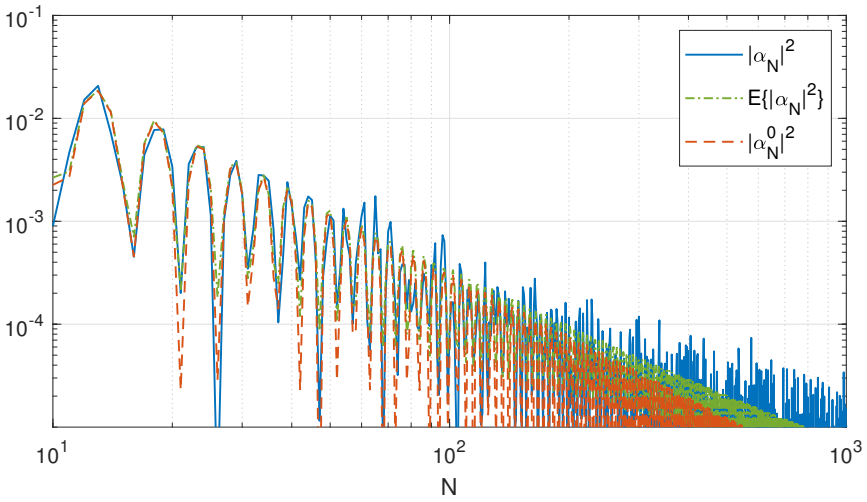


Figure 6.8: IUI factor $|\alpha_{1i,N}|^2$, its mean and nominal values for the N-element ULA as in Fig. 6.4 under uniform perturbations with $\sigma_p = \sigma_\phi = 0.01$. Note the similarity to Fig. 6.4, where perturbations are Gaussian.

exceeding a certain threshold ε. It means that any $|\alpha_N|^2 \leq \varepsilon$ is acceptable, regardless of its actual value. In this case, one can apply this target IUI threshold to (6.56) and obtain

$$\delta^2 \ll \ln(1 + N\varepsilon) \tag{6.61}$$

which would ensure that the threshold is not exceeded under random errors. To illustrate this, let $N = 100$ and $\varepsilon = 10^{-3}$ (i.e., -30 dB) for Fig. 6.4. In this scenario, using (6.61), one obtains $\delta^2 \ll 0.1$ makes the impact of random errors negligible for this design. This is consistent with $\delta^2 = 4\pi^2\sigma_p^2 + \sigma_\phi^2 \approx 0.004$ in Fig. 6.4. Using (6.59) and (6.60), one concludes that the number of required antennas to achieve a low IUI threshold $\varepsilon \ll 1$ scales as $1/\varepsilon$ with errors and only as $\sqrt{1/\varepsilon}$ without errors. Hence, for small ε, more antennas are needed in cases with errors.

Overall, random errors do have an impact on the IUI leakage factor, but it is still possible to achieve small values of $|\alpha_N|^2$, provided that N is sufficiently large, or, equivalently, if δ^2 is sufficiently small. This is consistent with the result in Theorem 3.

6.5.3 Impact of Errors on the Main User

Finally, we examine how random errors affect the main user's normalized power gain $|\alpha_{11,N}|^2$, which equals 1 in the absence of errors. Using (6.15)-(6.17) in Theorem 3, one obtains:

$$\begin{aligned} |\alpha_{11,N}|^2 &\to \mathbb{E}\{|\alpha_{11,N}|^2\} \\ &= |c_1|^2 |\alpha^0_{11,N}|^2 + (1-|c_1|^2)N^{-1} \\ &= |c_1|^2 + (1-|c_1|^2)N^{-1} \qquad (6.62) \\ &\to |c_1|^2 = e^{-\delta^2} \end{aligned}$$

where the last equality holds for Gaussian errors, and the others apply for both Gaussian and non-Gaussian errors; (6.62) is due to $|\alpha^0_{11,N}|^2 = 1$, which follows from (6.7) and (6.9) with $\Psi^0_{1n} = 0$. Using (6.51), one can observe that $|c_1|^2 \leq 1$, which measures the average normalized power received by the main user under random errors. It is important to note that random errors do affect the main user's power/SNR, but this impact is not significant as long as δ^2 remains small:

$$|\alpha_{11,N}|^2 \approx 1 \text{ if } \delta^2 \ll 1 \qquad (6.63)$$

The values of $|\alpha_{11,N}|^2$, its mean, and nominal values for an N-element ULA are illustrated in Fig. 6.9. Note that power loss is not considerable for $\sigma_p, \sigma_\phi \leq 0.1$ but it quickly increases when the variance of errors go above that threshold. By increasing N, the random fluctuations of $|\alpha_{11,N}|^2$ caused by random errors do decrease, but the significant power/SNR loss remains, irrespective of N, for $\sigma_p, \sigma_\phi > 0.1$. Thus, in practice, one should avoid this regime.

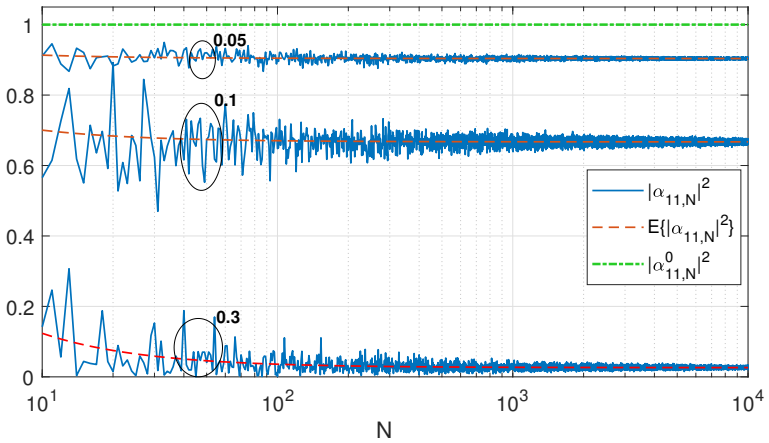


Figure 6.9: The main user normalized power gain $|\alpha_{11,N}|^2$, its mean and nominal values for the N-element ULA with $d = 0.5$ and $\theta_1 = 0$, under zero-mean i.i.d. Gaussian errors with $\sigma_p = \sigma_\phi = 0.05, 0.1, 0.3$ (generated independently for each value of N). While power loss is not large for $\sigma_p, \sigma_\phi \leq 0.1$, it quickly increases otherwise and $|\alpha^0_{11,N}|^2 = 1$ for the nominal (no errors) array.

6.6 Conclusion

While existing studies have typically considered favorable propagation in massive MIMO scenarios without considering implementation errors, this study investigates the impact of random errors on the FP. In particular, the FP property is rigorously shown to hold asymptotically for the perturbed array if it holds for the unperturbed one. Thus, random errors do not have a catastrophic impact, even as the number of antennas increases without bound. While random errors do not affect the FP asymptotically, they significantly slow down convergence to the asymptotic value. This means that more antennas are needed under random errors to achieve the same low IUI as without errors. These results are general enough to include arbitrary array geometry as well as non-Gaussian error distributions. Practical guidelines as to what accuracy is needed to make the impact of errors negligible for a finite number of antenna are given. The analytical results are validated via numerical simulations and are in agreement with measurement-based studies.

Chapter 7

Optimizing Uniform Linear Arrays for low IUI

7.1 Introduction

In this chapter, we study multi-user mMIMO systems with uniform linear arrays and a finite number of antennas. To reduce implementation complexity, the design of the ULA is optimized so that its total number of antennas is minimized under the SINR constraint. Optimization problems are solved for sum as well as per-user interference constraints. Due to the non-convexity of the problem, standard convex optimization tools are not applicable. Therefore, a novel analytical approach is proposed and some globally optimal closed-form solutions are derived. The optimal number of antennas depends on the target SINR, the total number of users, and their angular separation. The designs are robust since they do not require precise knowledge of the directions of arrival of interfering users.

7.2 Channel model and problem formulation

The system model defined in (3.2) is considered, where an N-element BS serves $M+1$ single antenna users simultaneously. In this chapter, we index the main user as 0, for technical reasons. The MRC beamformer (tuned to the main user) with the beamforming weight $\boldsymbol{w} = \boldsymbol{h}_0/|\boldsymbol{h}_0|$ is used in the receiver to detect the main user signal. As discussed previously, see Chapter 3, the SINR in (3.4) is maximized (SINR $= \gamma_0$) when the number of antennas increases without bound ($N \to \infty$) and favorable propagation holds. However, having an infinite number of antennas is not feasible in practice; instead, the number of antennas may be large but still finite and bounded. Therefore, we consider a finite number of antennas in this chapter and do not require the asymptotic favorable propagation to hold ($N \to \infty$), as this may exclude many practical scenarios. When the SNRs of all users are equal and large enough, $\gamma_i = \gamma_0 \gg 1$, which corresponds to the interference-dominated regime, the SINR can be approximated as follows:

$$\text{SINR} \approx \alpha_N^{-2}, \quad \alpha_N^2 = \sum_{i=1}^{M} |\alpha_{iN}|^2 \tag{7.1}$$

where α_N^2 represents the normalized aggregate interference. To have an acceptable system performance, the SINR should meet the requirement.

$$\text{SINR} \geq \gamma \tag{7.2}$$

where γ is a threshold SINR. It is desirable to minimize the complexity and cost of the design while meeting this constraint. To this end, it is necessary to minimize the number N of antennas. Therefore, the following question arises:

Q. *What is the minimum N required to guarantee that the SINR is not below its threshold γ?*

Without constraints on the angles-of-arrival (AoA) θ_0 and θ_i of the main and interfering

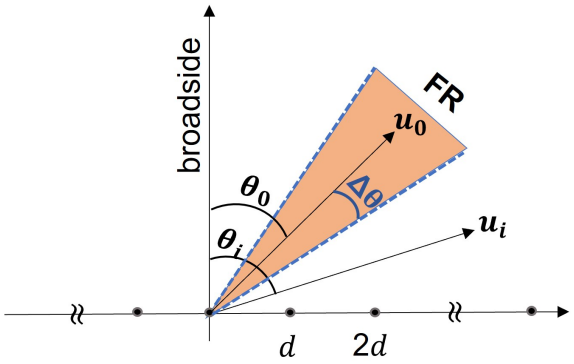


Figure 7.1: An N-element ULA of element spacing d; θ_0 and θ_i are the AoAs of the main and i-th interfering user, all measured from the broadside, $-\pi/2 \leq \theta_i \leq \pi/2$. The forbidden region FR of width $2\Delta\theta > 0$ protects the main user from excessive interference, so that $\theta_i \notin$ FR.

users, achieving a high SINR is not achievable since, if $\theta_i = \theta_0$, the main and i-th user cannot be separated, $|\alpha_{iN}| = 1$, and so the SINR is low (≤ 1). To avoid this situation, we introduce a forbidden region (FR) around the main user:

$$\text{FR} = \{\theta : |\theta - \theta_0| < \Delta\theta\} \tag{7.3}$$

where $2\Delta\theta > 0$ is angular "span". Interfering users must be outside of the FR, i.e. they are not allowed to get too close to the main user, $|\theta_i - \theta_0| \geq \Delta\theta$, so that $\Delta\theta$ is the minimum angular distance; see Fig. 7.1 for an illustration.

Using (7.1) and (7.3), the above question can be expressed as the following constrained optimization problem:

$$\text{(P1)}: \quad \min N \quad \text{s.t.} \quad \alpha_N^{-2} \geq \gamma, \quad \forall \theta_i \notin \text{FR}, \quad i = 1..M$$

where we implicitly assume that N is an integer. It is noteworthy that the FR constraint, $\theta_i \notin$ FR, does not require precise knowledge of the AoAs θ_i of the interfering users. It is similar to the power control strategy in [56], in which some particularly 'bad' users are dropped from service, improving the performance significantly. Considering a BS with a N-element ULA, as shown in Fig. 7.1, the respective IUI "leakage" factor $|\alpha_{iN}|$ of User i

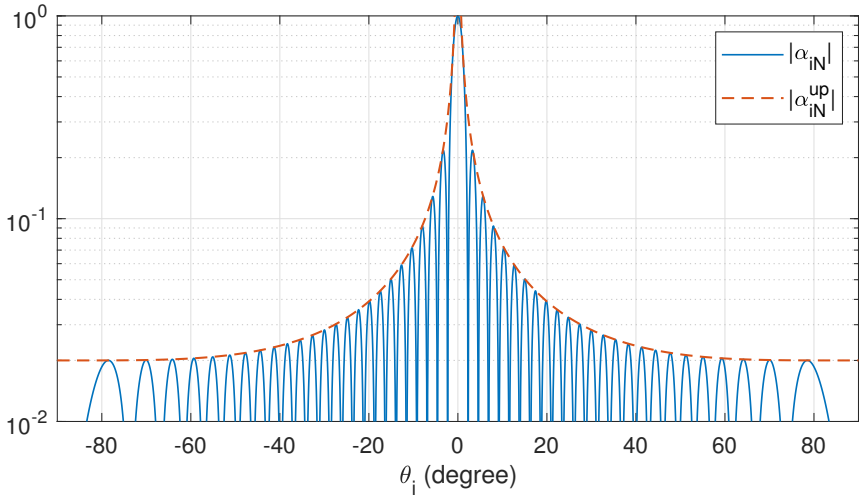


Figure 7.2: $|\alpha_{iN}|$ and α_{iN}^{up} for the ULA with $d = 1/2$, $\theta_0 = 0$ and $N = 50$. Note that the latter is a tight envelope of the former.

can be expressed as (see (5.3)):

$$|\alpha_{iN}| = \left| \frac{\sin(N\psi_i/2)}{N\sin(\psi_i/2)} \right|, \quad \psi_i = 2\pi d(\sin(\theta_i) - \sin(\theta_0)) \tag{7.4}$$

$$\leq \frac{1}{N|\sin(\psi_i/2)|} = \alpha_{iN}^{up} \tag{7.5}$$

where θ_0 and θ_i are the AoA's of the main and i-th user, respectively, measured from the array broadside. Note that this model can be extended to weakly-directional elements (e.g. short dipoles or patch antennas) in a strightforward way. However, this will have only a minor impact on our results and therefore is not pursued here. We further assume no back-plane radiation so that $|\theta_i| \leq \pi/2$. This is the case in practice due to the presence of a ground plane, element design or other shielding mechanisms. To simplify the derivations, we further assume that $|\theta_0| + \Delta\theta \leq \pi/2$, and our results can also be extended to $|\theta_0| + \Delta\theta > \pi/2$ with some more complicated manipulations.

For large N, $\sin(N\psi_i/2)$ in (7.4) oscilates very quickly and since the condition in (P1) must hold for $\forall\theta_i \notin$ FR, $|\sin(N\psi_i/2)| = 1$ is possible for many θ_i outside the forbidden region. Hence, the upper bound α_{iN}^{up} obtained in (7.5) is tight. Fig. 7.2 illustrates this point for $N = 50$ and $d = 1/2$: note that α_{iN}^{up} is a tight envelope of $|\alpha_{iN}|$ and they are equal for many AoAs θ_i. If grating lobes (GL) exist in the array pattern, $\alpha_{iN} = 1$ at GL

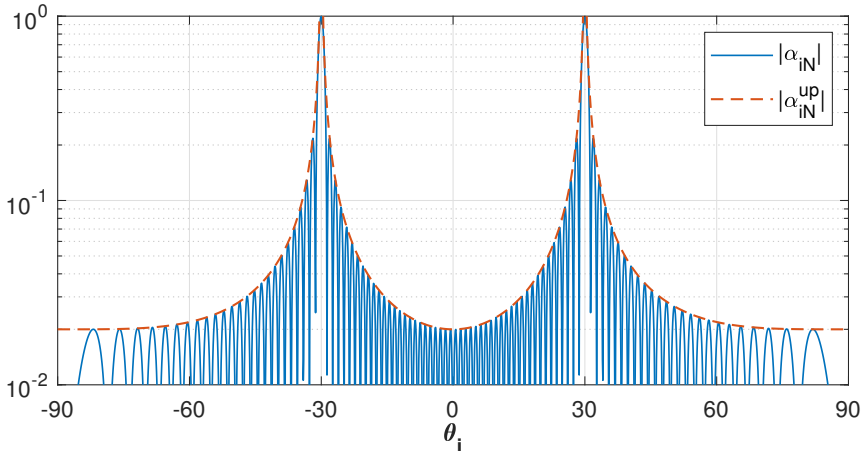


Figure 7.3: $|\alpha_{iN}|$ and α_{iN}^{up} for the ULA as in Fig. 7.2 but with larger spacing $d = 1$ and $\theta_0 = 30^0$. Note that, similarly to Fig. 7.2, α_{iN}^{up} is still a tight envelope for $|\alpha_{iN}|$ but, with larger spacing, a grating lobe appears at $\theta_i = -30^0$ and the favorable propagation does not hold anymore (large SINR is not achievable at the GL direction).

directions and therefore high SINR is not achievable if interfering users appear at those directions [89]. Hence, the SINR constraint in (P1) cannot be satisfied for $\gamma > 1$, see Fig. 7.3. To avoid GLs, the element spacing d should satisfies the following condition [61][62]:

$$d < d_{max} = \frac{1}{1 + \sin|\theta_0|} \tag{7.6}$$

Using (7.5), an upper bound on the total IUI leakage, defined in (7.1), is:

$$\alpha_N^2 = \sum_{i=1}^{M} |\alpha_{iN}|^2 \leq \sum_{i=1}^{M} |\alpha_{iN}^{up}|^2 = (\alpha_N^{up})^2 \tag{7.7}$$

and it is tight, as shown in Fig. 7.2, 7.3. Thus, by using this bound for optimization, there is no loss of accuracy, the analysis is simplified and leads us to closed-form global-optimal solutions. The problem (P1) can be re-formulated as follows:

$$\begin{aligned} (\text{P2}): \quad \min_{N,d} N \quad \text{s.t.} \quad & \alpha_N^{up} \leq 1/\sqrt{\gamma}, \quad \forall \theta_i \notin \text{FR}, \quad i = 1..M \\ & 0 < d < d_{max} \end{aligned} \tag{7.8}$$

with understanding that N is an integer. To obtain a more efficient design, we also allow optimization over d.

7.3 Optimal ULA with per-user IUI constraints

In this section, we re-formulate (P2) by replacing the sum IUI constraint with per-user ones. This is based on three reasons: (i) it offers more robust solutions, which are not sensitive to the AoAs of intervening users; their knowledge is not necessary as long as they are out of the FR; (ii) it makes the problem solvable analytically so that it is possible to obtain globally optimal closed-form solutions, and (iii) one can set per-user constraints in a way to satisfy the sum constraint in (P2), as well.

7.3.1 Per-user IUI constraints with optimal d

Using the per-user IUI constraints:

$$|\alpha_{iN}^{up}|^2 \leq 1/\gamma, \quad i = 1..M \tag{7.9}$$

the problem (P2) can be reformulated as follows:

$$\begin{aligned} (\text{P3}): \quad \min_{N,d} N \quad \text{s.t.} \quad & N|\sin(\psi_i(d,\theta_i)/2)| \geq \sqrt{\gamma}, \quad \forall \theta_i \notin \text{FR}, \quad i = 1..M \\ & 0 < d < d_{max} \end{aligned} \tag{7.10}$$

where we emphasize that ψ_i is a function of d and θ_i. It should be noted that (P3) is not a convex problem for two reasons: (i) its objective N is an integer, and (ii) its constraints are not convex. Standard optimization tools, such as KKT conditions [111], cannot be applied here; instead, different tools must be used, as described below.

The 1st constraint in (P3) implies that $N \geq \sqrt{\gamma}$ for any feasible N and therefore the minimum N satisfies $N_{min} \geq \sqrt{\gamma}$, i.e. it grows at least as the square root of the desired SINR ("5 dB per decade"). The exact globally-optimal solution for (P3) is presented in the following theorem, showing that this scaling is indeed tight.

Theorem 4. *The globally-optimal $N_{min}^{(3)}$ and d^* in (P3) are as follows:*

$$N_{min}^{(3)} = \left\lceil \frac{\sqrt{\gamma}}{\sin(\pi d^*(1+\sin(|\theta_0|)))} \right\rceil \tag{7.11}$$

$$d^* = (1+\sin(|\theta_0|+\Delta\theta))^{-1} \tag{7.12}$$

Proof. The following technical Lemma is instrumental to prove Theorem 4.

Lemma 9. *Let $|\theta_i| \leq \pi/2, \Delta\theta > 0$ and $|\theta_0| + \Delta\theta < \pi/2$. Then, the following inequalities hold for any d and any $\theta_i \notin FR$:*

$$0 < 2a_1 \leq |\psi_i(d,\theta_i)| \leq 2a_2 < 2\pi \tag{7.13}$$

where

$$a_1 = \pi d\left(\sin\left(|\theta_0|+\Delta\theta\right) - \sin|\theta_0|\right) \tag{7.14}$$

$$a_2 = \pi d\left(1+\sin|\theta_0|\right) \tag{7.15}$$

and the bounds are tight:

$$|\psi_i(d,\theta_i)| = 2a_2 \text{ if } \theta_i = -\operatorname{sign}(\theta_0)\,\pi/2 \tag{7.16}$$

$$|\psi_i(d,\theta_i)| = 2a_1 \text{ if } \theta_i = \theta_0 + \operatorname{sign}(\theta_0)\,\Delta\theta \tag{7.17}$$

Proof. To establish the lower bound in (7.13), we consider 2 different cases, depending on the sign of $\theta_i - \theta_0$.

1. $\theta_i - \theta_0 \geq 0$: in this case, $\theta_i \geq \theta_0 + \Delta\theta > \theta_0$ for any $\theta_i \notin$ FR and therefore

$$|\sin\theta_i - \sin\theta_0| = \sin\theta_i - \sin\theta_0 \tag{7.18}$$

$$\geq \sin(\theta_0+\Delta\theta) - \sin\theta_0 \tag{7.19}$$

$$= 2\cos(\theta_0+\Delta\theta/2)\sin(\Delta\theta/2) \tag{7.20}$$

$$= 2\cos|\theta_0+\Delta\theta/2|\sin(\Delta\theta/2) \tag{7.21}$$

$$\geq 2\cos(|\theta_0|+\Delta\theta/2)\sin(\Delta\theta/2) \tag{7.22}$$

$$= \sin(|\theta_0|+\Delta\theta) - \sin|\theta_0| \tag{7.23}$$

where (7.18) and (7.19) are due to $\sin x$ being an increasing function for $|x| \leq \pi/2$ and (7.22) is due to $\cos x$ being a decreasing function for $0 \leq x \leq \pi$.

2. $\theta_i - \theta_0 \leq 0$: this follows along the same lines as in (7.18)-(7.23) above.

Combining Cases 1 and 2 , for any $\theta_i \notin$ FR,

$$|\sin\theta_i - \sin\theta_0| \geq \sin\left(|\theta_0| + \Delta\theta\right) - \sin|\theta_0| \tag{7.24}$$

and using this inequality in (7.4), the lower bounds in (7.13) follows.

To establish the upper bounds in (7.13), note the following:

$$|\sin\theta_i - \sin\theta_0| \leq |\sin\theta_i| + |\sin\theta_0| \leq 1 + |\sin\theta_0| \tag{7.25}$$

and use this in (7.4). Using (7.16) and (7.17) in (7.4), it can be verified that the internal bounds in (7.13) are achieved. This completes the proof of Lemma 9. □

Next, note that the 1st constraint in (P3) is equivalent to:

$$\begin{aligned} & N \min_{\theta_i \notin \mathrm{FR}} \left|\sin\left(\psi_i\left(d,\theta_i\right)/2\right)\right| \\ & \quad = N \min_{\theta_i \notin \mathrm{FR}} \sin\left|\psi_i\left(d,\theta_i\right)/2\right| \geq \sqrt{\gamma} \end{aligned} \tag{7.26}$$

where the equality is due to $\left|\psi_i\left(d,\theta_i\right)/2\right| < \pi$ as in (7.13). To find the minimum in (7.26), we use (7.13) and set $z = \left|\psi_i\left(d,\theta_i\right)\right|/2$ so that (7.26) is equivalent to

$$N \min_{a_1 \leq z \leq a_2} \sin(z) \geq \sqrt{\gamma} \tag{7.27}$$

Since $\sin(z)$ is concave in the interval $0 \leq z \leq \pi$, its minimum is achieved on a boundary

point of this interval [111]. Hence,

$$\min_{a_1 \le z \le a_2} \sin(z) = \min\left\{\sin\left(a_1\right), \sin\left(a_2\right)\right\} \tag{7.28}$$

$$= \begin{cases} \sin\left(a_1\right) & \text{if } a_1 \le \pi - a_2 \\ \sin\left(a_2\right) & \text{if } a_1 \ge \pi - a_2 \end{cases} \tag{7.29}$$

and, using (7.14) and (7.15),

$$a_1 \le \pi - a_2 \quad \Leftrightarrow \quad d \le d^* = \left(1 + \sin\left(|\theta_0| + \Delta\theta\right)\right)^{-1} \tag{7.30}$$

Using (7.27) and (7.29), the optimal value $N_{min}^{(3)}$ of (P3) can be expressed as

$$N_{min}^{(3)} = \min_d \min_N N = \min\left\{N_1, N_2\right\} \tag{7.31}$$

where $\min_d \min_N$ are subject to the constraints in (P3) and, for 2nd equality, we have split the minimization over d into two sub-intervals:

$$N_1 = \left\lceil \min_d \frac{\sqrt{\gamma}}{\sin\left(a_1(d)\right)} \right\rceil \text{ s.t. } 0 < d \le d^* \tag{7.32}$$

$$N_2 = \left\lceil \min_d \frac{\sqrt{\gamma}}{\sin\left(a_2(d)\right)} \right\rceil \text{ s.t. } d^* \le d < d_{\max} \tag{7.33}$$

Next, we solve (7.32) and (7.33). To this end, observe that

$$a_1(d) = \pi d \left(\sin\left(|\theta_0| + \Delta\theta\right) - \sin|\theta_0|\right) \tag{7.34}$$

$$\le \pi d^* \left(\sin\left(|\theta_0| + \Delta\theta\right) - \sin|\theta_0|\right) = a_1\left(d^*\right) \tag{7.35}$$

$$= \frac{\pi\left(\sin\left(|\theta_0| + \Delta\theta\right) - \sin|\theta_0|\right)}{1 + \sin\left(|\theta_0| + \Delta\theta\right)} \tag{7.36}$$

$$= \pi\left(1 - \frac{1 + \sin|\theta_0|}{1 + \sin\left(|\theta_0| + \Delta\theta\right)}\right) \le \frac{\pi}{2} \tag{7.37}$$

where (7.35) is due to $d \leq d^*$ and $a_1(d) > 0$; (7.36) is due to (7.30) and (7.37) is due to

$$\frac{1+\sin|\theta_0|}{1+\sin\left(|\theta_0|+\Delta\theta\right)} \geq \frac{1+\sin|\theta_0|}{2} \geq \frac{1}{2} \tag{7.38}$$

where the last inequality is due to $0 \leq |\theta_0| < \pi/2$. Hence, using (7.13) and (7.35)-(7.37),

$$0 < a_1(d) \leq a_1\left(d^*\right) \leq \pi/2 \tag{7.39}$$

Since $\sin\left(a_1\right)$ is monotonically increasing in the the interval $0 < a_1 \leq \pi/2, d = d^*$ is a minimizer in (7.32),

$$N_1 = \left\lceil \frac{\sqrt{\gamma}}{\sin\left(a_1\left(d^*\right)\right)} \right\rceil \tag{7.40}$$

where an optimal spacing is $d_1^* = d^*$ (note that optimal spacing is not necessarily unique here). To find N_2, observe that

$$a_2(d) = \pi d\left(1+\sin|\theta_0|\right) \tag{7.41}$$

$$\geq \pi d^*\left(1+\sin|\theta_0|\right) = a_2\left(d^*\right) \tag{7.42}$$

$$= \frac{\pi\left(1+\sin\left(|\theta_0|\right)\right)}{1+\sin\left(|\theta_0|+\Delta\theta\right)} \geq \pi/2 \tag{7.43}$$

where (7.42) is due to $d \geq d^*$ in (7.33), the inequality in (7.43) is due to (7.38). Hence, using (7.42), (7.43) and (7.13), one obtains

$$\pi/2 \leq a_2\left(d^*\right) \leq a_2(d) < \pi \tag{7.44}$$

Since $\sin\left(a_2\right)$ is monotonically decreasing in the the interval $\pi/2 \leq a_2 < \pi, d = d^*$ is a minimizer in (7.33),

$$N_2 = \left\lceil \frac{\sqrt{\gamma}}{\sin\left(a_2\left(d^*\right)\right)} \right\rceil \tag{7.45}$$

and $d_2^* = d^*$ is an optimum spacing. Further note that $\sin\left(a_2\left(d^*\right)\right) = \sin\left(a_1\left(d^*\right)\right)$, due to $a_1\left(d^*\right) = \pi - a_2\left(d^*\right)$, and, therefore, $N_1 = N_2 = N_{min}^{(3)}$, along with $d^* = d_1^* = d_2^*$, as

required. This completes the proof. □

Note that the globally-optimal $N_{min}^{(3)}$ and d^* in Theorem 4 are independent of θ_i (as long as $\theta_i \notin$ FR) so that its precise knowledge is not needed. Also note that the optimal spacing d^* is independent of the threshold SINR γ but depends on θ_0 and $\Delta\theta$, while $N_{\min}^{(3)}$ does depend on γ.

To get further insight, let us consider the case of small forbidden region $\Delta\theta \ll 1$ (typical for massive MIMO applications with large N). In this case, using Taylor series [106], (7.12) simplifies to

$$d^* \approx d_{\max} - \Delta d, \quad \Delta d = \frac{\Delta\theta \cos\theta_0}{(1+\sin|\theta_0|)^2} \tag{7.46}$$

$$= 1 - \Delta\theta \tag{7.47}$$

where the last equality holds when the main user is at broadside, $\theta_0 = 0$. Note that Δd is small for small $\Delta\theta$, i.e. d^* is slightly smaller than $d_{\max}$ so that the ULA has its maximum angular resolution without a GL degrading its performance (this avoids a significant drop in the SINR). Likewise, (7.11) simplifies to

$$N_{min}^{(3)} \approx \left\lceil \frac{\sqrt{\gamma}}{\Delta\theta} \frac{1+\sin|\theta_0|}{\pi\cos\theta_0} \right\rceil \geq \left\lceil \frac{\sqrt{\gamma}}{\pi\Delta\theta} \right\rceil \tag{7.48}$$

where the inequality holds with equality when the main user is at broadside, $\theta_0 = 0$. Note from (7.48) that $N_{min}^{(3)}$ scales linearly with $\sqrt{\gamma}$ (i.e. "5 dB per decade of SINR" rule) and $1/\Delta\theta$ (i.e. "10 dB per decade of inverse FR"), i.e. it is more sensitive to the forbidden region angular measure $\Delta\theta$ than to the target SINR γ. Beam steering with larger $|\theta_0|$ requires more antennas to maintain the same SINR and the smallest number of antennas is needed for $\theta_0 = 0$.

When the beam steering is over a bounded sector $|\theta_0| \leq \theta_{max}, N_{min}^{(3)}$ should be set for the largest possible steering angle $|\theta_0| = \theta_{max}$ and this will also work for the entire region $|\theta_0| \leq \theta_{max}$. Hence, for this design, a precise knowledge of θ_0 is not needed and θ_{max} is sufficient.

7.3.2 Per-user IUI constraints with fixed d

Some antenna array designs prefer a fixed element spacing (often, $d = 1/2$) and hence optimization over d is not possible. Thus, we consider here an optimal ULA with fixed d and reformulate (P3) as follows:

$$\text{(P4):}\ \min_{N} N \text{ s.t. } N\left|\sin\left(\psi_i\left(\theta_i\right)/2\right)\right| \geq \sqrt{\gamma} \quad \forall \theta_i \notin \text{FR} \tag{7.49}$$

where $i = 1..M$ and minimization is over N alone, $\psi_i\left(\theta_i\right)$ is a function of θ_i only and the spacing d is a fixed parameter. Proposition 8 below presents the globally-optimal solution of the non-convex problem (P4).

Proposition 8. *The globally-minimal N in (P4) is:*

$$N_{min}^{(4)} = \begin{cases} \left\lceil \sqrt{\gamma}/\sin\left(a_1(d)\right)\right\rceil, & 0 < d \leq d^* \\ \left\lceil \sqrt{\gamma}/\sin\left(a_2(d)\right)\right\rceil, & d^* \leq d < d_{\max} \end{cases} \tag{7.50}$$

where d^* is as in (7.12), and

$$a_1(d) = \pi d\left(\sin\left(|\theta_0| + \Delta\theta\right) - \sin|\theta_0|\right) \tag{7.51}$$

$$a_2(d) = \pi d\left(1 + \sin|\theta_0|\right) = \pi d/d_{\max} \tag{7.52}$$

Proof. Follows from (7.31)-(7.33) in the proof of Theorem 4, with minimization over d removed. □

Note that, in general, the optimal values of (P3) and (P4), $N_{min}^{(3)}$ and $N_{min}^{(4)}$, are related as

$$N_{min}^{(3)} \leq N_{min}^{(4)} \tag{7.53}$$

and the inequality is strict unless d is sufficiently close to d^*, i.e., in general, larger N is needed for a fixed d design.

To get further insight, (7.50) can be simplified in the special case of a small forbidden region $\Delta\theta \ll \min\{1, (\pi d)^{-1}\}$ and $d \leq d^*$

$$N_{min}^{(4)} \approx \left\lceil \frac{\sqrt{\gamma}}{\pi \Delta\theta d \cos\theta_0} \right\rceil \geq \left\lceil \frac{\sqrt{\gamma}}{\pi \Delta\theta d} \right\rceil \tag{7.54}$$

with equality if $\theta_0 = 0$. Comparing (7.48) and (7.54), we conclude that the same linear scaling with $\sqrt{\gamma}$ and $1/\Delta\theta$ holds in both cases, so that the optimization over d in (P3) offers only a fixed offset (independent of γ and $\Delta\theta$) for N_{min}. A popular choice in many practical array designs is $d = 1/2$, for which (7.54) further simplifies to

$$N_{min}^{(4)} \approx \left\lceil \frac{2\sqrt{\gamma}}{\pi \Delta\theta \cos\theta_0} \right\rceil \geq \left\lceil \frac{2\sqrt{\gamma}}{\pi \Delta\theta} \right\rceil \tag{7.55}$$

Comparing this to (7.48), we conclude that the optimization over d allows one to reduce N_{min} by a factor of

$$2\left(1 + \sin|\theta_0|\right)^{-1} \leq 2 \tag{7.56}$$

with corresponding reduction in complexity and this gain is maximal ($= 2$) at $\theta_0 = 0$.

7.4 From per-user to sum IUI constraint

The above results obtained under the per-user IUI constraints in (P3)(P4) can also be extended to the sum IUI constraint. To this end, observe that one way to satisfy the sum constraint in (P2) is to ensure that $|\alpha_{iN}^{up}| \leq 1/\sqrt{M\gamma}$ for all $i = 1 \ldots M$ and $\theta_i \notin \text{FR}$, so that

$$\alpha_N^2 = \sum_{i=1}^{M} |\alpha_{iN}|^2 \leq 1/\gamma \tag{7.57}$$

as required. Replacing the target SINR γ with $M\gamma$ in Theorem 4 , one obtains an upper bound on the optimal value $N_{min}^{(2)}$ of (P2). In particular, using (7.11)(7.48),

$$N_{min}^{(2)} \leq \left\lceil \frac{\sqrt{M\gamma}}{\sin\left(\pi d^* \left(1+\sin|\theta_0|\right)\right)} \right\rceil \approx \left\lceil \frac{\sqrt{M\gamma}}{\Delta\theta} \frac{1+\sin|\theta_0|}{\pi\cos\theta_0} \right\rceil \tag{7.58}$$

and the upper bound is tight (holds with equality) when all interfering users are $\Delta\theta$ away from the main user (at the FR boundary - the worst-case scenario). Note that $N_{min}^{(2)}$ scales with γ and $1/\Delta\theta$ at most as $N_{min}^{(3)}$, and at most as $\sqrt{M}$ with the number of interfering users ("5 dB per decade"). These scalings are independent of interfering users' AoAs, provided they are outside of the FR. In fact, a ULA design based on the upper bound in (7.58) is also independent of these AoAs, a welcome robustness property for a practical design. To see whether the above scalings are tight in other scenarios, we consider uniformly-distributed interfering users below.

7.5 Optimal ULA for uniformly-distributed users

This section aims to obtain a globally-optimal solution for (P2) without explicitly imposing a forbidden region around the main user so that interfering users can approach it more closely as their number is increasing. To make the analysis and optimization analytically tractable, we consider a setting where M interfering users are distributed uniformly around the main user so that their AoAs are as follows,

$$\theta_0 = 0, \quad \theta_i = i\pi/M, \quad i = \pm 1 \ldots \pm M/2 \tag{7.59}$$

where we assume that M is even. Clearly, as M increases, adjacent users are getting closer and closer to each other so that scenarios of various user densities can be modelled. For this setting, (P2) is modified as follows:

$$\text{(P5):} \ \min_N N \text{ s.t. } \alpha_N^{up} \leq 1/\sqrt{\gamma} \tag{7.60}$$

where there is no forbidden region and we consider fixed- d design with $d \leq 2/\pi$ to avoid GLs; α_N^{up} is as in (7.7). The following Proposition gives a closed-form globally-optimal solution of this problem.

Proposition 9. *Let the user AoAs be as in (7.59) and the element spacing $0 < d \leq 2/\pi$ be fixed. Then, the globally-optimal value $N_{min}^{(5)}$ of $(P5)$ is*

$$N_{min}^{(5)} = \frac{M}{\pi d}\sqrt{\frac{\gamma}{3}}(1+o(1)) \approx \frac{M}{\pi d}\sqrt{\frac{\gamma}{3}} \tag{7.61}$$

where $o(1) \to 0$ as $M \to \infty$ and the approximation holds for sufficiently-large M.

Proof. Two technical Lemmas are needed for the proof.

Lemma 10. *[113] The following inequality holds for $0 \leq x \leq 2$:*

$$\frac{x}{\sqrt{1+x^2}} \leq \sin(x) \leq x \tag{7.62}$$

Lemma 11. *The following relationship holds as $M \to \infty$:*

$$S_M = \sum_{i=1}^{M/2} \frac{1}{\sin^2\left(\frac{i\pi}{M}\right)} = \frac{1}{6}M^2(1+o(1)) \tag{7.63}$$

Proof. Using the lower bound in (7.62),

$$S_M = \sum_{i=1}^{M/2} \frac{1}{\sin^2\left(\frac{i\pi}{M}\right)} \leq \sum_{i=1}^{M/2}\left(1+\frac{M^2}{(\pi i)^2}\right) \tag{7.64}$$

$$= \frac{M}{2} + \frac{M^2}{\pi^2}\sum_{i=1}^{M/2}\frac{1}{i^2} = \frac{1}{6}M^2(1+o(1)) \tag{7.65}$$

where the last equality is due to $\sum_{i=1}^{\infty} i^{-2} = \pi^2/6$ so that $\sum_{i=1}^{M/2} i^{-2} = \pi^2/6(1+o(1))$ as $M \to \infty$. On the other hand, using the upper bound in (7.62),

$$S_M \geq \sum_{i=1}^{M/2} \frac{M^2}{(\pi i)^2} = \frac{1}{6}M^2(1+o(1)) \tag{7.66}$$

Combining these two bounds for S_M, (7.63) follows. □

Using (7.5) and Lemma 10,

$$(N\pi d\sin\theta_i)^{-2} \le |\alpha_{iN}^{up}|^2 \le N^{-2}\left(1+(\pi d\sin\theta_i)^{-2}\right) \tag{7.67}$$

so that

$$(\alpha_N^{up})^2 = \sum_{i\neq 0}|\alpha_{iN}^{up}|^2 = 2\sum_{i=1}^{M/2}|\alpha_{iN}^{up}|^2 \tag{7.68}$$

$$\le \frac{2}{N^2}\sum_{i=1}^{M/2}\left(1+(\pi d\sin\theta_i)^{-2}\right) \tag{7.69}$$

$$= \frac{M}{N^2} + \frac{2}{(N\pi d)^2}\sum_{i=1}^{M/2}\frac{1}{\sin^2\left(\frac{i\pi}{M}\right)} \tag{7.70}$$

$$= \frac{M}{N^2} + \frac{M^2}{3(N\pi d)^2}(1+o(1)) \tag{7.71}$$

$$= \frac{M^2}{3(N\pi d)^2}(1+o(1)) \tag{7.72}$$

where 2nd equality in (7.68) is due to symmetery of $|\alpha_{iN}|$ around $i=0$, (7.69) is due to (7.67); (7.71) is due to Lemma 11. On the other hand,

$$(\alpha_N^{up})^2 \ge \frac{2}{(N\pi d)^2}\sum_{i=1}^{M/2}\frac{1}{\sin^2\left(\frac{i\pi}{M}\right)} = \frac{M^2}{3(N\pi d)^2}(1+o(1)) \tag{7.73}$$

where the inequality is due to (7.67) and the equality is due to Lemma 11. Combining (7.72) and (7.73),

$$\alpha_N^{up} = \frac{M}{\sqrt{3}N\pi d}(1+o(1)) \tag{7.74}$$

and using the constraint in (P5),

$$N \ge \frac{M\sqrt{\gamma}}{\sqrt{3}\pi d}(1+o(1)) \tag{7.75}$$

from which (7.61) follows. This completes the proof. □

With the popular choice of $d = 1/2$, the minimum number of elements is

$$N_{min}^{(5)} = \frac{2M}{\pi}\sqrt{\frac{\gamma}{3}}(1 + o(1)) \approx \frac{2M}{\pi}\sqrt{\frac{\gamma}{3}} \tag{7.76}$$

Note from (7.61), (7.76) that the scaling of $N_{min}^{(5)}$ with γ is the same as in (7.58) (and also in (7.11), (7.50)) but the scaling with M is remarkably different from that in (7.58): $N_{min}^{(5)} \sim M$ vs. $\sqrt{M}$ in (7.58), i.e. much faster for the former. This can be explained by the absence of the FR in (P5), where the minimal user angular spacing is π/M, i.e. decreases with M, while it is fixed at $\Delta\theta$ for all M in (P2). The lack of the FR calls for a larger number of antennas to maintain the same target SINR. One can further consider an extension of (P5) whereby the optimization includes the element spacing d as well,

$$\text{(P6)}: \min_{N,d} N \text{ s.t. } \alpha_N^{up} \leq 1/\sqrt{\gamma}, \quad 0 < d \leq 2/\pi \tag{7.77}$$

Since $N_{min}^{(5)}$ is monotonically-decreasing in d, it follows that the largest feasible d is optimal, $d^* = 2/\pi$ and therefore

$$N_{min}^{(6)} = \frac{M}{2}\sqrt{\frac{\gamma}{3}}(1 + o(1)) \approx \frac{M}{2}\sqrt{\frac{\gamma}{3}} \tag{7.78}$$

i.e a reduction in the number of elements by the factor $N_{min}^{(5)}/N_{min}^{(6)} \approx 4/\pi$ compared to $d = 1/2$.

7.6 Examples

In this section, we present several examples to illustrate and validate the above analytical results. Let us begin with the fixed- d design under the per-user constraint in (P4) with $\gamma = 20$ dB, $\theta_0 = 0$ and $\Delta\theta = 5°$. The actual (achieved) SINR is $|\alpha_{iN}|^{-2}$, where α_{iN} is as in (7.4), and its lower bound (LB) is $|\alpha_{iN}^{up}|^{-2}$, where α_{iN}^{up} is as in (7.5). Using (7.54), $N_{min}^{(4)} = 74$. Fig. 7.4 illustrates the actual SINR for this design vs. AoA θ_i of the interfering user. As expected, the target SINR is met everywhere outside of the forbidden region $|\theta_i| < \Delta\theta = 5°$.

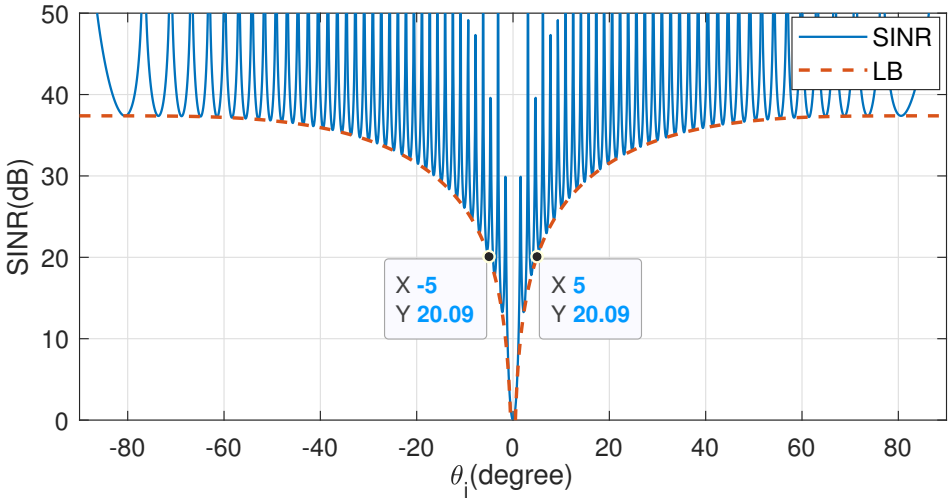


Figure 7.4: The actual SINR $|\alpha_{iN}|^{-2}$ and its lower bound (LB) $|\alpha_{iN}^{up}|^{-2}$ for the fixed-d ULA with $d = 1/2$ and $N = N_{min}^{(4)} = 74$, $\theta_0 = 0$, $\Delta\theta = 5^0$, $\gamma = 20$dB. Note that the actual SINR$\geq$ 20 dB for all $|\theta_i| \geq 5^0$ 5o, as required by (P1).

Although the actual SINR oscillates due to the nulls in the antenna array pattern, its LB (envelope) monotonically increases with $|\theta_i|$ as the interfering user moves away from the main user, which is indeed a tight envelope..

We now consider the optimal- d design in (P3) for the settings of Fig. 7.4 to see the benefit of optimizing over d. From (7.11) and (7.12), $d^* = 0.92$ and $N_{min}^{(3)} = 41$ in this case, i.e. much smaller N is needed for the optimal- d design compared to the fixed- d design with $N_{min}^{(4)} = 74$, which also translates into smaller complexity/cost. Fig. 7.5 illustrates the actual SINR for this design vs. AoA θ_i of the interfering user. As expected, the target SINR is met everywhere outside of the forbidden region $|\theta_i| < \Delta\theta = 5°$.

Note the different behavior of Fig. 7.5 and Fig. 7.4. Unlike the latter, where the SINR envelope (LB) increases monotonically with $|\theta_i - \theta_0|$, for the former, the SINR envelope can decrease despite the interfering user moving further away from the main user, e.g. around $\theta_i = \pm 90°$. This is attributed to the grating lobe, which partially appears around $\theta_i = \pm 90°$, see e.g. [89], due to larger spacing $d = d^* = 0.92$.

Consider now (P5) with $M = 50$ uniformly-distributed users as in (7.59) and $\gamma = 20$ dB, $d = 0.5$. Using (7.61), it follows that $N_{min}^{(5)} = 185$, i.e. significantly larger than $N_{min}^{(3)} = 41$ and $N_{min}^{(4)} = 74$ above, which is due to the sum rather than per-user IUI

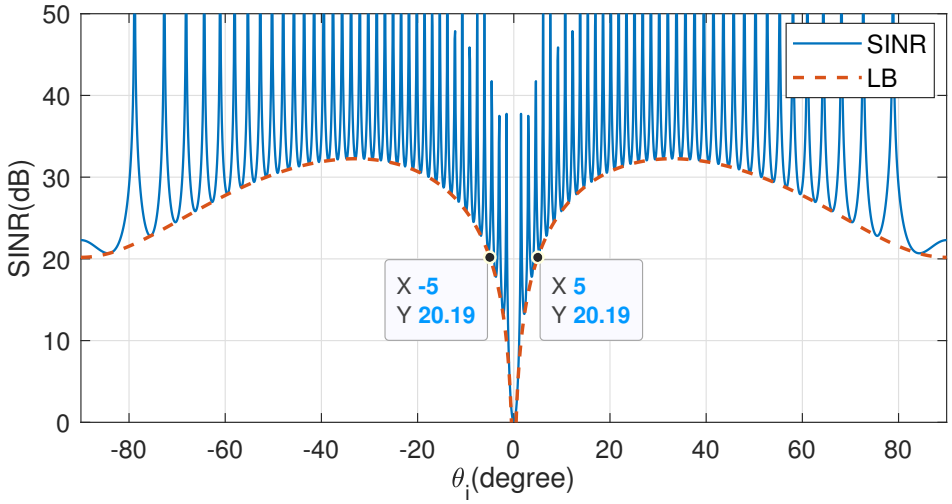


Figure 7.5: The actual SINR $|\alpha_{iN}|^{-2}$ and its lower bound (LB) $|\alpha_{iN}^{up}|^{-2}$ for the optimized-d ULA with $d = d^* = 0.92$ and $N = N_{min}^{(3)} = 41, \theta_0 = 0, \Delta\theta = 5°, \gamma = 20$ dB. Smaller N is needed here compared to Fig. 7.4.

constraint and the absence of the FR. Fig. 7.6 illustrates this case. Note that the LB $|\alpha_N^{up}|^{-2}$ is a tight envelope of the actual SINR.

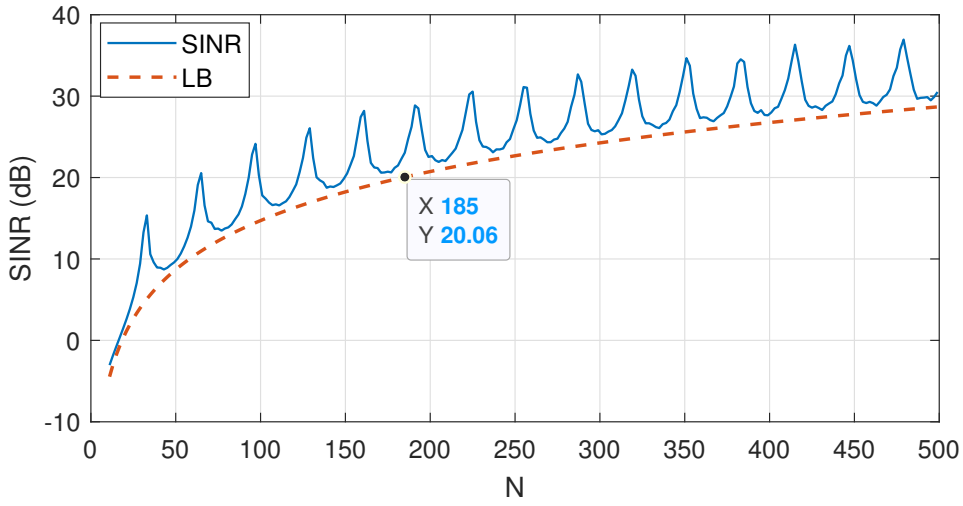


Figure 7.6: Fig. 6. SINR and its lower bound (LB) for a ULA with $d = 0.5$ and $M = 50$ uniformly-distributed users; the main user AoA is $\theta_0 = 0°$. To ensure SINR ≥ 20 dB, the number of antennas $N \geq N_{min}^{(5)} \approx 185$, i.e. significantly larger than in Fig. 7.4 and 7.5.

7.7 Conclusion

The FP for uniform linear arrays is considered in the non-asymptotic regime (finite number of antennas). The objective is to minimize the number of antennas (complexity/cost) while satisfying SINR constraints in both single and multiple interfering-user scenarios. To solve these non-convex optimization problems, novel inequalities are developed, and the overall feasible set is partitioned into several sub-sets, over which the optimization is performed. The obtained closed-form globally optimal solutions/designs reveal how the optimal number of antennas scales with the target SINR, angular user separation, and the total number of users. In particular, it is shown that the number of antennas can be reduced by almost 50% if variable antenna spacing is allowed compared to the fixed spacing of half a wavelength.

Chapter 8

Conclusion

8.1 Summary

Mobile communication technologies are evolving rapidly to meet the surging demand for data, particularly video content. While 5G has marked significant progress, emerging applications such as the Internet of Everything and Holographic Telepresence demand even more advanced networks. This growing need has paved the way for 6G mobile networks, which prioritize innovative spectrum utilization and energy-efficient transmission. Among the key technologies in 5G and beyond wireless networks, massive MIMO stands out for its remarkable enhancement of spectral and energy efficiency. By employing a large array of antennas, mMIMO significantly improves spatial resolution, facilitating orthogonal (non-interfering) channel vectors for different users. This phenomenon, known as favorable propagation, enables the use of simple linear processing, like matched filtering, as the optimal approach.

In this thesis, after emphasizing the importance of massive MIMO and its key property, favorable propagation, in Chapter 1, we conducted a comprehensive study of existing literature on favorable propagation in massive MIMO systems, encompassing various channel conditions and antenna geometries, in Chapter 2.

In Chapter 3, we introduced a fixed MIMO channel model with M independent single-antenna users transmitting simultaneously to an N-element access point, which forms the basis for the analyses in this thesis. To detect the main user signal, we employed matched filter beamforming, which allowed us to derive the SINR as a function of the main user SNR and the aggregated IUI of the interfering users. This aggregated IUI served as our primary metric for evaluating the FP. We also presented a comparison between various linear processing techniques, including matched filtering, zero-forcing, and minimum-mean-square error, in terms of computational complexity and their performance under the FP, as well as their robustness. While matched filtering boasts lower computational complexity and greater robustness—each antenna operates independently of others in the array—it exhibits inferior performance in mitigating IUI. Nevertheless, this performance gap diminishes with an increasing number of antennas.

In Chapter 4, we provided rigorous proof of favorable propagation in LOS scenarios for both uniform circular and cylindrical arrays with fixed antenna spacing and a finite number M of users, each located at distinct positions with unique angles of arrival. Previous methods used to establish the FP for linear and planar arrays are not applicable to circular arrays. Therefore, we employed the Bessel series expansion to obtain novel upper bounds for the IUI leakage factor, enabling us to analyze its asymptotic behavior. As the FP holds asymptotically, it becomes possible to reduce inter-user interference to negligible levels by employing a sufficiently large number of antennas. To approach the FP closely, we established a condition for the number of antennas, N, through an asymptotic analysis. The FP does not hold in general if either M grows unbounded with N or if the AoAs are allowed to approach each other. However, it might hold in some special cases.

In Chapter 5, we identified a gap in existing studies on favorable propagation for uniform linear arrays. Our analysis revealed that the presence of grating lobes in a ULA can lead to violations of the conditions for favorable propagation, even when users have distinct angles of arrival. To address this issue, we introduced a novel design—a block-partitioned nonuniform linear array—where all grating lobes are effectively canceled, restoring favor-

able propagation for all distinct angles of arrival. In the NULA design, each block (subarray) is a ULA, but the overall array is non-uniform. We carefully selected the number and spacing of subarrays (blocks) to ensure that, grating lobes are canceled by nulls in the block array factor. Notably, our analysis and design exhibit frequency domain robustness, making them suitable for wide-band systems.

Existing studies have typically investigated favorable propagation in massive MIMO systems without considering implementation errors. However, real-world systems inevitably involve implementation errors. Chapter 6 explores the impact of random errors in antenna locations and beamforming phases on the FP, and rigorously demonstrates that the FP holds asymptotically for the perturbed array if it holds for the unperturbed one. This condition becomes both necessary and sufficient when random errors have a Gaussian distribution. Although random errors do not impact the asymptotic FP, they notably decelerate the convergence of the IUI leakage factor toward its asymptotic value from $1/N^2$ (no errors) to $1/N$ (with errors). This means that more antennas are needed under random errors to achieve the same low IUI as without errors. These results are general enough to include arbitrary array geometry. Practical guidelines on the level of accuracy required to make the impact of errors negligible for a finite number of antennas was provided.

In Chapter 7, we investigated the FP for uniform linear arrays with a finite number of antennas (non-asymptotic regime). When the asymptotic FP holds, it means that the IUI is zero when we have an infinite number of antennas. However, depending on the requirements of a system, a minimum SINR may be tolerable, and therefore, zero IUI is not always necessary. Our objective was to determine the minimum number of antennas (complexity/cost) required to satisfy SINR constraints in both single and multiple interfering-user scenarios. To solve these non-convex optimization problems, we developed novel inequalities and partition the overall feasible set into several subsets for optimization. The closed-form globally optimal solutions/designs obtained reveal how the optimal number of antennas scales with the target SINR, angular user separation, and the total number of users. Specifically, the study demonstrates that allowing variable antenna spacing can

reduce the number of antennas by almost 50% compared to using a fixed spacing of half a wavelength.

8.2 Future Work

Despite extensive studies on favorable propagation in massive MIMO, there are still unresolved issues. In this section, we explore several possibilities for extending the current work.

Multi-user massive MIMO in high-altitude platform stations (HAPS): Ultra-wide coverage is a crucial concept in 6G mobile communications, and HAPS have received significant attention as a leading solution for extending coverage [91][114]. Massive MIMO plays a key role in HAPS systems due to its ability to deliver high spectral and energy efficiency, which are essential in HAPS systems. Cylindrical arrays are considered strong candidates for these systems because they can provide extensive coverage [91][92]. However, a question arises: how many antennas are needed in the cylindrical array to serve a given number M of users while ensuring a minimum SINR? In other words, what is the maximum number of users that can be served using an N-element antenna array? This problem has already been addressed for uniform linear arrays in Chapter 7, but it would be more practical to extend this solution to cylindrical and planar arrays.

FP analysis in the non-asymptotic regime with non-uniform weighting: In Chapter 7, we optimize uniform linear arrays in the non-asymptotic regime (finite N) using uniform weighting in the antenna array. Future work may involve optimizing the performance of finite N systems by applying non-uniform weights, such as Chebyshev weighting[1], to expedite convergence to asymptotic orthogonality. This, in turn, would result in a lower number of antennas, leading to reduced complexity and cost. Preliminary studies on the impact of nonuniform weighting (e.g., Chebyshev weighting) on the FP show a substantial

[1]The array employing Chebyshev weighting achieves maximum directivity for a given sidelobe ratio when all sidelobes are of equal height [62].

decrease in the number of required antennas compared to uniform weighting arrays to achieve the same level of SINR. These results with non-uniform weighting can also be extended to planar and circular arrays.

FP improvement using intelligent reflecting surface (IRS): IRS is a promising technology for enhancing the spectral and energy efficiency of 5G and 6G wireless networks [115]. IRS utilizes large, reconfigurable metasurfaces or arrays of reflective elements to manipulate electromagnetic waves in real-time, making it especially valuable in challenging propagation environments like indoor spaces and urban areas [115]. Exploring how IRS can further enhance favorable propagation in massive MIMO is an intriguing direction for future research.

Analysis and improvement of the FP in Holographic massive MIMO: In holographic MIMO, an array of antennas is used to synthesize and manipulate the phase and amplitude of electromagnetic waves in a highly controlled manner [116]. This enables the creation of precise and adaptive radiation patterns that can be dynamically adjusted to focus energy where it's needed and nullify interference. Some of the key characteristics and advantages of holographic MIMO include spatial precision, adaptivity, improved coverage and capacity, and low latency [116]. An analysis of favorable propagation and its potential to improve complexity and cost in holographic MIMO systems remains an area for future exploration.

Investigating broader geometric configurations for the favorable propagation property: Exploring a more general result for arbitrary geometric configurations (which increases with the number of antennas) and investigating the possibility of establishing the general condition under which the FP property would hold are key objectives for future research. For instance, exploring the possibility of having antennas randomly placed within an arbitrary area while maintaining a fixed density of the antennas could offer valuable insights.

Exploring the favorable propagation property in realistic scenarios: In realistic scenarios, mobile user devices may have more than one antennas, and the propagation environment tends to be more complex, characterized by multipath components and frequency selec-

tivity. Investigating the existence or applicability of the favorable propagation property within these realistic conditions stands as a promising avenue for future research.

Bibliography

[1] M. Shafi et al., 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice, IEEE Journal Sel. Areas Comm., vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[2] Ericsson Mobility Report, June 2023. http://www.ericsson.com/en/mobility-report.

[3] 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, pp. 154–655, Nov. 2017.

[4] C. de Alwis et al., Survey on 6G Frontiers: Trends, Applications, Requirements, Technologies and Future Research, IEEE Open Journal Comm. Soc., vol. 2, pp. 836–886, Apr. 2021.

[5] P. Yang et al., 6G Wireless Communications: Vision and Potential Techniques, IEEE Netw., vol. 33, no. 4, pp. 70–75, Jul. 2019.

[6] E. Biglieri et al., MIMO Wireless Communications. Cambridge University Press, 2010.

[7] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, Fundamentals of Massive MIMO. Cambridge University Press, 2016.

[8] T. L. Marzetta, Massive MIMO: An Introduction, Bell Labs Tech. Journal., vol. 20, pp. 11–22, Mar. 2015.

[9] T. S. Rappaport et al., Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond, IEEE Access., vol. 7, pp. 78729–78757, Jun 2019.

[10] E. G. Larsson, O. Edfors, F. Tufvesson and T. L. Marzetta, Massive MIMO for Next Generation Wireless Systems, IEEE Comm. Magazine., vol. 52, no. 2, pp. 186-195, Feb. 2014.

[11] T. Marzetta, Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas, IEEE Trans. Wireless Comm., vol. 9, no. 11, pp. 3590 –3600, Nov. 2010.

[12] H. Q. Ngo and E. Larsson, No Downlink Pilots Are Needed in TDD Massive MIMO, IEEE Trans. Wireless Comm., vol. 16, no. 5, pp. 2921–2935, May 2017.

[13] P. Harris et al., Performance Characterization of a Real-Time Massive MIMO System With LOS Mobile Channels, IEEE Journal Sel. Areas Comm., vol. 35, no. 6, pp. 1244-1253, Jun. 2017.

[14] F. Bohagen, P. Orten, G. E. Oien, Design of Optimal High-Rank Line-of-Sight MIMO Channels, IEEE Trans. Wireless Comm., vol. 6, no. 4, pp. 1420-1425, Apr. 2007.

[15] P. Wang et al, Tens of Gigabits Wireless Communications Over E-Band LoS MIMO Channels With Uniform Linear Antenna Arrays, IEEE Trans. Wireless Comm., vol. 13, no. 7, pp. 3791-3805, Jul. 2014.

[16] H. Do, N. Lee, A. Lozano, Reconfigurable ULAs for Line-of-Sight MIMO Transmission, IEEE Trans. Wireless Comm., vol. 20, no. 5, pp. 2933-2947, May 2021.

[17] E. G. Larsson, T. L. Marzetta, H. Q. Ngo, and H. Yang, Antenna Count for Massive MIMO: 1.9 GHz vs. 60 GHz, IEEE Comm. Mag., vol. 56, no. 9, pp. 132–137, Sep. 2018.

[18] H. Tataria et al, 6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities, Proc. IEEE., vol. 109, no. 7, pp. 1166-1199, Jul. 2021.

[19] S. Rangan, T. S. Rappaport, and E. Erkip, Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges, Proc. IEEE., vol. 102, no. 3, pp. 366-385, Mar. 2014.

[20] E. F.W. Alexanderson, Transoceanic Radio Communication, Trans. Amer. Inst. Elect. Eng., vol. 38, no. 2, pp. 1269–1285, Jul. 1919.

[21] B. S. Tsybakov, The Capacity of a Memoryless Gaussian Vector Channel, Problems of Inform. Transmission., vol. 1, no. 1, pp. 18–29, 1965.

[22] I. E. Telatar, Capacity of Multi-Antenna Gaussian Channels, AT&T Bell Labs, Internal Tech. Memo, Jun. 1995, (European Trans. Telecom., vol.10, no. 6, Dec. 1995).

[23] G. J. Foschini and M. J. Gans, On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas, Wireless Pers. Commun., vol. 6, pp. 311–335, Mar. 1998.

[24] G. G. Raleigh and J. M. Cioffi, Spatio-Temporal Coding for Wireless Communication, IEEE Trans. Comm., vol. 46, no. 3, pp. 357–366, Mar. 1998.

[25] R. W. Heath Jr, A. Lozano, Foundations of MIMO Communication. Cambridge University Press, 2018.

[26] Special Issue on MIMO Systems, IEEE Trans. Sig. Proc., vol. 50, no. 10, Oct. 2002.

[27] Special Issue on MIMO Systems and Applications: Parts I and II, IEEE Journal Sel. Areas Comm., vol. 21, no. 3 and 5, Apr. and Jun. 2003.

[28] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[29] T. L. Marzetta et al., Special Issue on Massive MIMO, Journal of Communications and Networks., vol. 15, no. 4, pp. 333-337, Aug. 2013.

[30] Special Issue on Large-Scale Multiple Antenna Wireless Systems, IEEE Journal Sel. Areas Comm., vol. 31, no. 2, Feb 2013.

[31] D. Gesbert et al., From Single User to Multiuser Communications: Shifting the MIMO Paradigm, IEEE Signal Process. Mag., vol. 24, no. 5, pp. 36-36, Sep. 2007.

[32] P. Viswanath and D. N. C. Tse, Sum Capacity of a Vector Gaussian Broadcast Channel and Uplink-Downlink Duality, IEEE Trans. Inf. Theory., vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[33] H. Weingarten, Y. Steinberg, and S. Shamai, The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel, IEEE Trans. Inf. Theory., vol. 52, no. 9, pp. 3936–3964, Sep. 2006

[34] T. L. Marzetta, How Much Training is Required for Multiuser MIMO?, 40th Asilomar Conference on Signals, Systems and Computers (ACSSC), Pacific Grove, CA, USA, pp. 359-363, Nov. 2006.

[35] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, Pilot Contamination Problem in Multicell TDD Systems, IEEE Int. Symp. Information Theory, Seoul, Korea (South), pp. 2184-2188, Feb. 2009.

[36] E. Bjornson, J. Hoydis, L. Sanguinetti, Massive MIMO Has Unlimited Capacity, IEEE Trans. Wireless Comm., vol. 17, no. 1, pp. 574-590, Jan. 2018.

[37] R. W. Heath et al., An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems, IEEE Journal Sel. Topics Signal Process., vol. 10, no. 3, pp. 436–453, Apr. 2016.

[38] S. A. Busari et al., Millimeter-Wave Massive MIMO Communication for Future Wireless Systems, IEEE Comm. Surveys Tuts., vol. 20, no. 2, pp. 836–869, Dec. 2017.

[39] M. Xiao et al., Millimeter Wave Communications for Future Mobile Networks, IEEE Journal Sel. Areas Comm., vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[40] A. L. Swindlehurst et al., Millimeter-Wave Massive MIMO: The Next Wireless Revolution? IEEE Comm. Mag., vol. 52, no. 9, pp. 56–62, Sep. 2014.

[41] L. Lu et al., An Overview of Massive MIMO: Benefits and Challenges, IEEE Journal Sel. Topics in Sig. Proc., vol. 8, no. 5, pp.742-758, Oct. 2014.

[42] F. Rusek et al., Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays, Signal Processing Magazine., vol. 30, no. 1, pp. 40-60, Jan. 2013.

[43] I. Ahmed, H. Khammari, and A. Shahid, Resource Allocation for Transmit Hybrid Beamforming in Decoupled Millimeter Wave Multiuser-MIMO Downlink, IEEE Access., vol. 5, pp. 170–182, Dec. 2017.

[44] I. Ahmed et al., A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives, IEEE Comm. Surveys Tuts., vol. 20, no. 4, pp. 3060–3097, Jun. 2018.

[45] 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 Comm. Mag., vol. 53, no. 1, pp. 186–194, Jan. 2015.

[46] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, Aspects of Favorable Propagation in Massive MIMO, in European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, pp. 76–80, Sep. 2014.

[47] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems, IEEE Trans. Comm., vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[48] J. H. Chen, When Does Asymptotic Orthogonality Exist for Very Large Arrays?, IEEE GlobeCom, Atlanta, GA, USA, pp. 4146-4150, Nov. 2013.

[49] C. Masouros and M. Matthaiou, Physically Constrained Massive MIMO: Hitting the Wall of Favorable Propagation, IEEE Comm. Lett., vol. 19, no. 5, pp. 771–774, May 2015.

[50] J. Hoydis et al., Channel Measurements for Large Antenna Arrays, Int. Symp. Wireless Comm Systems (ISWCS), Paris, France, pp. 811-815, Aug. 2012.

[51] X. Gao et al., Massive MIMO Performance Evaluation Based on Measured Propagation Data, IEEE Trans. Wireless Comm., vol. 14, no. 7, pp. 3899-3911, Jul. 2015.

[52] M. Gauger et al., Channel Measurements with Different Antenna Array Geometries for Massive MIMO Systems, 10th Int. ITG Conf. Systems Comm. Coding (SCC), Hamburg, Germany, pp. 1-6, Feb. 2015.

[53] A. O. Martínez et al., An Experimental Study of Massive MIMO Properties in 5G Scenarios, IEEE Trans. Antennas Propag., vol. 66, no. 12, pp. 7206–7215, Dec. 2018.

[54] J. Li and Y. Zhao, Measurement-Based Asymptotic User Orthogonality Analysis and Modelling for Massive MIMO, IEEE Comm. Lett., vol. 21, no. 12, pp. 2762–2765, Dec. 2017.

[55] J. Flordelis et al., Spatial Separation of Closely-Located Users in Measured Massive MIMO Channels, IEEE Access., vol. 6, pp. 40253–40266, Jul. 2018.

[56] H. Yang, T. L. Marzetta, Massive MIMO With Max-Min Power Control in Line-of-Sight Propagation Environment, IEEE Trans. Comm., vol. 65, no. 11, pp. 4685-4693, Nov. 2017.

[57] E. Anarakifirooz, S. Loyka, Favorable Propagation for Massive MIMO with Circular and Cylindrical Antenna Arrays, IEEE Wireless Comm. Letters., vol. 11, no. 3, pp. 458-462, Mar. 2022.

[58] X. Gao et al., Linear Pre-Coding Performance in Measured Very-Large MIMO Channels, IEEE Vehicular Technology Conference (VTC Fall), San Francisco, CA, USA, pp. 1-5, Sep. 2011.

[59] M. Matthaiou et al., Does Massive MIMO Fail in Ricean Channels?, IEEE Wireless Comm Letters., vol. 8, no. 1, pp. 61-64, Feb. 2019.

[60] J. Zhang et al., 3D MIMO for 5G NR: Several Observations from 32 to Massive 256 Antennas Based on Channel Measurement, IEEE Comm. Mag., vol. 56, no. 3, pp. 62–70, Mar. 2018.

[61] L. H. Van Trees, Optimum Array Processing. Hoboken, NJ, USA: Wiley, 2002.

[62] R.C. Hansen, Phased Array Antennas, New York, USA: Wiley, 1998.

[63] Y. V. Krivosheev, A. V. Shishlov, and V. V. Denisenko, Grating Lobe Suppression in Aperiodic Phased Array Antennas Composed of Periodic Subarrays With Large Element Spacing, IEEE Antennas Propag. Mag., vol. 57, no. 1, pp. 76-85, Feb. 2015.

[64] R.L. Haupt, Antenna Arrays: A Computational Approach, Wiley-IEEE Press, 2010.

[65] Z. Iqbal, M. Pour, Grating Lobe Reduction in Scanning Phased Array Antennas With Large Element Spacing, IEEE Trans. Antennas Prop., vol. 66, no. 12, pp. 6965-6974, Oct. 2018.

[66] A. Puglielli et al., Design of Energy and Cost-Efficient Massive MIMO Arrays, Proc. IEEE, vol. 104, no. 3, pp. 586–606, Mar. 2016.

[67] N. Mukit et al., Designing Large-Scale Antenna Array Using Sub-Array, Bull. Electr. Eng. Inform., vol. 8, no. 3, pp. 906-915, Sep. 2019.

[68] W. Hong et al., Study and Prototyping of Practically Large-Scale mmWave Antenna Systems for 5G Cellular Devices, IEEE Comm. Mag., vol. 52, no. 9, pp. 63–69, Sep. 2014.

[69] R. F. Harrington, Sidelobe Reduction by Nonuniform Element Spacing , IEEE Trans. Antennas Propagat., vol. 9, no. 2, pp. 187-192, Mar. 1961.

[70] F. Hodjat and S. A. Hovanessian, Nonuniformly Spaced Linear and Planar Array Antennas for Sidelobe Reduction, IEEE Trans. Antennas Propag., vol. 26, no. 2, pp. 198–204, Mar. 1978.

[71] X. F. Ren, J. A. Azevedo, and A. M. Casimiro, Synthesis of Non-Uniformly Spaced Arrays Using the Fourier Transform and Window Techniques, IET Microwaves, Antennas Propag., vol. 3, no. 8, pp. 1245–1253, Dec. 2009.

[72] À. O. Martínez, E. de Carvalho, and J. Ø. Nielsen, Towards Very Large Aperture Massive MIMO: A measurement Based Study, IEEE Globecom Workshops (GC Wkshps), Austin, TX, USA, pp. 281–286, Dec. 2014.

[73] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally? IEEE Trans. Commun., vol. 63, no. 11, pp. 3917–3928, Nov. 2015.

[74] X. Gao et al., Measured Propagation Characteristics for Very-Large MIMO at 2.6 GHz, 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR), Pacific Grove, CA, USA, pp. 295-299, Nov. 2012.

[75] A. O. Martínez, E. De Carvalho, and J. Ø. Nielsen, Massive MIMO Properties based on Measured Channels: Channel Hardening, User Decorrelation and Channel Sparsity, 50th Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, USA, pp. 1804–1808, Nov. 2016.

[76] J. Flordelis et al., Spatial Separation of Closely-Spaced Users in Measured Massive Multi-User MIMO Channels, IEEE Int. Conf. Commun. (ICC), London, U.K., pp. 1441–1446, Jun. 2015.

[77] E. Bjornson, M. Matthaiou and M. Debbah, Massive MIMO with Non-Ideal Arbitrary Arrays: Hardware Scaling Laws and Circuit-Aware Design, IEEE Trans. Wireless Comm., vol. 14, no. 8, pp. 4353-4368, Aug. 2015.

[78] E. Bjornson, P. Zetterberg, M. Bengtsson and B. Ottersten, Capacity Limits and Multiplexing Gains of MIMO Channels with Transceiver Impairments, IEEE Comm. Letters., vol. 17, no. 1, pp. 91-94, Jan. 2013.

[79] S. A. Vorobyov, A. B. Gershman, Z.-Q. Luo, Robust Adaptive Beamforming Using Worst-Case Performance Optimization: A Solution to The Signal Mismatch Problem, IEEE Trans. Signal Process., vol. 51, no. 2, pp. 313-324, Feb. 2003.

[80] R. G. Lorenz, S. P. Boyd, Robust Minimum Variance Beamforming, IEEE Trans. Signal Process., vol. 53, pp. 1684-1696, May 2005.

[81] J. Li, P. Stoica, Robust Adaptive Beamforming. New York: Wiley, 2006.

[82] S. Loyka, C. D. Charalambous, On the Compound Capacity of a Class of MIMO Channels Subject to Normed Uncertainty, IEEE Trans. Info. Theory., vol. 58, no. 4, pp. 2048-2063, Apr. 2012.

[83] S. Loyka, C. D. Charalambous, Novel Matrix Singular Value Inequalities and Their Applications to Uncertain MIMO Channels, IEEE Trans. Info. Theory., vol. 61, no. 12, pp. 6623–6634, Dec. 2015.

[84] C. Studer, M. Wenk, and A. Burg, MIMO Transmission with Residual Transmit-RF Impairments, International ITG Workshop on Smart Antennas (WSA), Bremen, Germany, pp. 189–196, Feb. 2010.

[85] B. E. Priyanto et al., Assessing and Modeling the Effect of RF Impairments on UTRA LTE Uplink Performance, IEEE VTC-Fall, Baltimore, MD, USA, pp. 1213–1217, Sep. 2007.

[86] E. Bjornson et al, Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits, IEEE Trans. Info. Theory., vol. 60, no. 11, pp. 7112-7139, Nov. 2014.

[87] M. Wenk, MIMO-OFDM Testbed: Challenges, Implementations, and Measurement Results (Microelectronics). Konstanz, Germany: Hartung-Gorre, 2010.

[88] P. Zetterberg, Experimental Investigation of TDD Reciprocity-Based Zero-Forcing Transmit Precoding, EURASIP J. Adv. Signal Process., vol. 2011, pp.1-10, Dec. 2011..

[89] E. Anarakifirooz, S. Loyka, Structural Design of Non-Uniform Linear Arrays for Favorable Propagation in Massive MIMO, IEEE Comm. Letters., vol. 27, no. 1, pp. 367-371, Jan. 2023.

[90] S. Willhammer et al., Channel Hardening in Massive MIMO: Model Parameters and Experimental Assessment, IEEE Open Journal of the Comm. Society., vol. 1, pp. 501-512, Apr. 2020.

[91] K. Tashiro, K. Hoshino and A. Nagate, Cylindrical Massive MIMO System for HAPS: Capacity Enhancement and Coverage Extension, IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, pp. 1-6, Apr. 2021.

[92] https://www.softbank.jp/en/corp/news/info/2021/20210902_01.

[93] S. Loyka, and M. Khojastehnia, Comments on "On Favorable Propagation in Massive MIMO Systems and Different Antenna Configurations", IEEE Access., vol. 7, pp. 185369-185372, Dec. 2019.

[94] E. Anarakifirooz and S. Loyka, Favorable Propagation for Massive MIMO With Circular and Cylindrical Antenna Arrays, IEEE Wireless Comm. Letters, vol. 11, no. 3, pp. 458-462, Mar. 2022.

[95] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications to Statistics and Econometrics, Wiley, 1999.

[96] A. Puglielli et al., Design of Energy- and Cost-Efficient Massive MIMO Arrays, Proc. IEEE., vol. 104, no. 3, pp. 586-606, Mar. 2016.

[97] K. H. Rosen, Handbook of Discrete and Combinatorial Mathematics. Boca Raton, FL, USA: CRC Press, 1999.

[98] C.N. Chen et al, 38-GHz Phased Array Transmitter and Receiver Based on Scalable Phased Array Modules with Endfire Antenna Arrays for 5G MMW Data Links, IEEE Trans. Microwave Theory Tech., vol. 69, no. 1, pp. 980-999, Jan. 2021.

[99] C. A. Balanis, Antenna theory analysis and design-4th edition, John Wiley & Sons, New York, 1998.

[100] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, Digital Communications, 3rd ed. Norwell, MA: Kluwer, 2004

[101] G. N. Watson. A Treatise on the Theory of Bessel Functions. 2d ed. Cambridge University Press, 1966.

[102] L. Landau, Bessel Functions: Monotonicity and Bounds, Journal of the London Mathematical Society, v. 61, no. 1, pp. 197–215, Feb. 2000.

[103] I. Krasikov, Uniform Bounds for Bessel Functions, Journal of Applied Analysis., vol. 12, no. 1, pp. 83–91, Jun. 2006

[104] E.G. Larsson, T.L. Marzetta, H.Q. Ngo, and H. Yang, Antenna Count for Massive MIMO: 1.9 GHz vs. 60 GHz, IEEE Comm. Mag., vol. 56, no. 9, pp. 132-137, Sep. 2018.

[105] S. Loyka, G. Levin, Finite-SNR Diversity-Multiplexing Tradeoff via Asymptotic Analysis of Large MIMO Systems, IEEE Trans. Info. Theory., vol. 56, no. 10, pp. 4781-4792, Oct. 2010.

[106] A. N. Shiryaev, Probability (3rd Ed.), New York, NY, USA: Springer-Verlag, 2015.

[107] A. A. Borovkov, Probability Theory (5th Ed.), London: Springer, 2013.

[108] W. Feller, An Introduction to Probability Theory and Its Applications., vol. 2, John Wiley & Sons, 1971.

[109] R. Lyons, Strong Laws of Large Numbers for Weakly Correlated Random Variables, Michigan Math. J., vol. 35, no. 3, pp. 353-359, Jan. 1988.

[110] I. N. Bronshtein, K. A. Semendyayev et al, Handbook of Mathematics, 5th ed. New York: Springer, 2007.

[111] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[112] N. L. Johnson et al, Continuous Univariate Distributions (2nd Ed.), New York: John Wiley & Sons, 1995.

[113] A. J. Kroopnick, A Lower Bound for sin x. The Mathematical Gazette., vol. 81, No. 490, pp. 88-89, Mar. 1997.

[114] W. Jaafar and H. Yanikomeroglu, HAPS-ITS: enabling future ITS services in trans-continental highways, IEEE Commun. Mag., vol. 60, no. 10, pp. 80–86, Oct. 2022.

[115] M. Di Renzo et al, Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How it Works, State of Research, and The Road Ahead, IEEE JSAC, v. 38, no. 11, pp. 2450–2525, Nov. 2020.

[116] C. Huang et al., Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends, IEEE Wireless Commun., vol. 27, no. 5, pp. 118–125, Oct. 2020.

APPENDICES

A.1 Interchanging Limit Orders

By using induction, one can prove the order of limit and summation over a finite number of sequences can be changed:

$$\lim_{n\to\infty}\sum_{k=1}^{k'} a_{kn} = \sum_{k=1}^{k'}\lim_{n\to\infty} a_{kn} \tag{1}$$

Using this theorem, exchanging the summation orders of a convergent sequence $f(n,m)$ can be easily proved:

$$\begin{aligned}\sum_{m=1}^{M}\sum_{n=-\infty}^{+\infty} f(n,m) &= \sum_{m=1}^{M}\lim_{N\to\infty}\sum_{n=-N}^{N} f(n,m) = \lim_{N\to\infty}\sum_{m=1}^{M}\sum_{n=-N}^{N} f(n,m)\\ &= \lim_{N\to\infty}\sum_{n=-N}^{N}\sum_{m=1}^{N} f(n,m) = \sum_{n=-\infty}^{+\infty}\sum_{m=1}^{M} f(n,m)\end{aligned} \tag{2}$$

the third equality in (1) derived from the fact that, two finite sum are interchangeable.

A.2 Proof of Lemma 5

First, we prove the upper bound. Using (6.5), one obtains

$$\begin{aligned}\mathbb{E}\{|\alpha_N|^4\} &= \mathbb{E}\{\alpha_N\alpha_N^*\alpha_N\alpha_N^*\} \\ &= \frac{1}{N^4}\,\mathbb{E}\{\sum_{n_1=1}^{N} e^{j\Psi_{in_1}} \sum_{n_2=1}^{N} e^{-j\Psi_{in_2}} \sum_{n_3=1}^{N} e^{j\Psi_{in_3}} \sum_{n_4=1}^{N} e^{-j\Psi_{in_4}}\} \\ &= \frac{1}{N^4} \sum_{n_1,n_2,n_3,n_4} \mathbb{E}\{e^{j(\Psi_{in_1}-\Psi_{in_2}+\Psi_{in_3}-\Psi_{in_4})}\} \end{aligned} \tag{3}$$

where $1 \leq n_k \leq N$. Since $e^{j\Psi_{in_k}}$ are independent for different n_k, we divide the total set S_t of $\{n_1, ..., n_4\}$ into the set of distinct indices S_d and its complementary set $S_d^c = S_t - S_d$,

$$S_t = \{\{n_1, n_2, n_3, n_4\}, 1 \leq n_k \leq N\} \tag{4}$$

$$S_d = \{\{n_1, n_2, n_3, n_4\}, n_i \neq n_j, \forall i \neq j\} \tag{5}$$

Next, we determine the cardinalities of these sets. First, note that the cardinality of S_t is $|S_t| = N^4$ and, likewise,

$$|S_d| = N(N-1)(N-2)(N-3) \tag{6}$$

(the latter follows from the number $(N)_k$ of all possible ordered selections of k distinct items out of a set of N distinct items, $(N)_k = \frac{N!}{(N-k)!}$ [110, p. 745]). Hence, the cardinality of the complementary set S_d^c is

$$|S_d^c| = N^4 - |S_d| = 6N^3 - 11N^2 + 6N \tag{7}$$

To simplify the derivations, define

$$\beta_{i,n_1..n_4} = \prod_{k=1}^{4} \mathbb{E}\{e^{j(-1)^{k+1}\Psi_{in_k}}\} \tag{8}$$

$$\beta'_{i,n_1..n_4} = \mathbb{E}\{\prod_{k=1}^{4} e^{j(-1)^{k+1}\Psi_{in_k}}\} \tag{9}$$

Since $e^{j\Psi_{in_k}}$ are independent for different n_k,

$$\beta_{i,n_1..n_4} = \beta'_{i,n_1..n_4} \text{ if } \{n_1..n_4\} \in S_d \tag{10}$$

Using (3) and (10),

$$\mathbb{E}\{|\alpha_N|^4\} = \frac{1}{N^4} \sum_{n_1..n_4} \beta'_{i,n_1..n_4} \tag{11}$$

$$= \frac{1}{N^4}\{\sum_{S_d} \beta_{i,n_1..n_4} + \sum_{S_d^c} \beta'_{i,n_1..n_4}\} \tag{12}$$

$$= \frac{1}{N^4}\{\sum_{S_d} \beta_{i,n_1..n_4} + \sum_{S_d^c} \beta_{i,n_1..n_4}\}$$

$$+ \frac{1}{N^4}\{\sum_{S_d^c} \beta'_{i,n_1..n_4} - \sum_{S_d^c} \beta_{i,n_1..n_4}\} \tag{13}$$

$$= \frac{1}{N^4}\{\sum_{n_1..n_4} \beta_{i,n_1..n_4} + \sum_{S_d^c} (\beta'_{i,n_1..n_4} - \beta_{i,n_1..n_4})\} \tag{14}$$

$$\leq |\mathbb{E}\{\alpha_N\}|^4 + N^{-4} \sum_{S_d^c} |\beta'_{i,n_1..n_4} - \beta_{i,n_1..n_4}| \tag{15}$$

$$\leq |\mathbb{E}\{\alpha_N\}|^4 + 12N^{-1} - 22N^{-2} + 12N^{-3} \tag{16}$$

$$\leq |\mathbb{E}\{\alpha_N\}|^4 + 12N^{-1} \tag{17}$$

where (12) is due to (10), (15) is due to the triangle inequality and

$$|\mathbb{E}\{\alpha_N\}|^4 = \frac{1}{N^4} \sum_{n_1..n_4} \beta_{i,n_1..n_4} \tag{18}$$

and (16) is due to (7) and

$$|\beta'_{i,n_1..n_4} - \beta_{i,n_1..n_4}| \leq 2 \tag{19}$$

which follows from $|\beta'_{i,n_1..n_4}|, |\beta_{i,n_1..n_4}| \leq 1$; (17) is due to

$$-22N^{-2} + 12N^{-3} < 0, \quad N \geq 1 \tag{20}$$

This establishes the upper bound. The lower bound in (6.19) follows from Jensen's inequality, since $|\cdot|^4$ is a convex function, see e.g. [111].

A.3 Proof of Lemma 8

We will need the following technical Lemmas.

Lemma 12 ([109]). *Let $\{a_N\}_{N=1}^{\infty}$ be a sequence of real numbers such that*

$$a_N \geq 0, \ \sum_{N=1}^{\infty} \frac{a_N}{N} < \infty \tag{21}$$

Then, there exists an increasing sequence of integers $\{N_k\}_{k=1}^{\infty}$ such that $N_{k+1}/N_k \to 1$ as $k \to \infty$ and

$$\sum_{k=1}^{\infty} a_{N_k} < \infty \tag{22}$$

Lemma 13 ([109]). *If y_n are random variables such that $\sum_{n=1}^{\infty} \mathbb{E}\{|y_n|^2\} < \infty$, then $y_n \xrightarrow{a.s.} 0$ as $n \to \infty$.*

Now, use Lemma 12 with

$$a_N = \mathrm{Var}\{z_N\} = \mathbb{E}\{|z_N - \mathbb{E}\{z_N\}|^2\} \tag{23}$$

which satisfy (21), to conclude that there exits an increasing sequence $\{N_k\}$ such that

$N_{k+1}/N_k \to 1$ and

$$\sum_{k=1}^{\infty} \mathbb{E}\{|z_{N_k} - \mathbb{E}\{z_{N_k}\}|^2\} < \infty \tag{24}$$

Now, using Lemma 13 with $y_k = z_{N_k} - \mathbb{E}\{z_{N_k}\}$,

$$z_{N_k} \xrightarrow{a.s.} \mathbb{E}\{z_{N_k}\} \text{ as } k \to \infty \tag{25}$$

i.e. almost sure convergence holds for the subsequence $\{N_k\}$. It remains to prove that it also holds for any N in-between, $N_k < N < N_{k+1}$. To this end, let $y_{nm} = z_{nm} - \mathbb{E}\{z_{nm}\}$, define the following sets $I_1,\ I_2$ of indexes (n, m):

$$\begin{aligned} I_1 &= \{(n,m) : 1 \le n \le N_k,\ N_k < m \le N\} \\ I_2 &= \{(n,m) : N_k < n \le N,\ 1 \le m \le N\} \end{aligned} \tag{26}$$

and note that

$$|y_N| = |z_N - \mathbb{E}\{z_N\}| \tag{27}$$

$$= \frac{1}{N^2}\left|\sum_{n,m\le N} y_{nm}\right| \tag{28}$$

$$\le \frac{1}{N_k^2}\left|\sum_{n,m\le N} y_{nm}\right| \tag{29}$$

$$= \frac{1}{N_k^2}\left|S_k + \sum_{(n,m)\in I_1} y_{nm} + \sum_{(n,m)\in I_2} y_{nm}\right| \tag{30}$$

$$\le \frac{|S_k|}{N_k^2} + \sum_{(n,m)\in I_1} \frac{|y_{nm}|}{N_k^2} + \sum_{(n,m)\in I_2} \frac{|y_{nm}|}{N_k^2} \tag{31}$$

$$\le \frac{|S_k|}{N_k^2} + \sum_{\substack{n\le N_k \\ N_k < m \le N_{k+1}}} \frac{|y_{nm}|}{N_k^2} + \sum_{\substack{m\le N_{k+1} \\ N_k < n \le N_{k+1}}} \frac{|y_{nm}|}{N_k^2} \tag{32}$$

$$\le \frac{|S_k|}{N_k^2} + \frac{2(N_{k+1} - N_k)}{N_k} + \frac{2N_{k+1}(N_{k+1} - N_k)}{N_k^2} \tag{33}$$

$$\xrightarrow{a.s.} 0 \tag{34}$$

where $S_k = \sum_{n,m \leq N_k} y_{nm}$; (29) is due to $N > N_k$, (32) is due to $N_{k+1} > N_k$ and all summation terms being positive, (33) is due to

$$|y_{nm}| = |z_{nm} - \mathbb{E}\{z_{nm}\}| \leq |z_{nm}| + \mathbb{E}\{|z_{nm}|\} \leq 2 \tag{35}$$

since $|z_{nm}| \leq 1$; (34) follows since, from (25),

$$\frac{S_k}{N_k^2} = z_{N_k} - \mathbb{E}\{z_{N_k}\} \xrightarrow{a.s.} 0 \tag{36}$$

and $N_{k+1}/N_k \to 1$. Therefore, from (34), $z_N \xrightarrow{a.s.} \mathbb{E}\{z_N\}$, as required.

A.4 Proof of Proposition 7

Using (6.26) and (6.27):

$$\sigma_N^2 = \mathbb{E}\{|\alpha_N|^2\} - |\,\mathbb{E}\{\alpha_N\}|^2 = N^{-1}(1 - |c_i|^2) \tag{37}$$

To estimate $|\alpha_N|^2$ for large N, we first obtain the asymptotic distribution of $\alpha_N = \alpha_{N1} + j\alpha_{N2}$. To this end, observe from (6.5) that it is an empirical average of i.i.d. random variables with finite variance. Therefore, from the central limit theorem [106, p. 406], its real α_{N1} and imaginary α_{N2} parts are asymptotically Gaussian,

$$\alpha_{Nk} \sim \mathcal{N}(\mathbb{E}\{\alpha_{Nk}\}, \sigma_{Nk}^2), \;\; k = 1, 2 \tag{38}$$

where $\sigma^2_{Nk} = \text{Var}\{\alpha_{Nk}\}$ and, using (6.30), $\mathbb{E}\{\alpha_N\} = c_i\alpha^0_N$ from which $\mathbb{E}\{\alpha_{Nk}\}$ can be found. Next, σ^2_{N1} can be evaluated as follows:

$$\sigma^2_{N1} = \text{Var}\left\{\frac{1}{N}\sum_{n=1}^{N}\cos(\Psi_{in})\right\} \tag{39}$$

$$= \frac{1}{N^2}\sum_{n=1}^{N}\text{Var}\{\cos(\Psi_{in})\} \tag{40}$$

$$= \frac{1}{N^2}\sum_{n=1}^{N}(\mathbb{E}\{\cos^2(\Psi_{in})\} - (\mathbb{E}\{\cos(\Psi_{in})\})^2) \tag{41}$$

$$= \frac{1}{N^2}\sum_{n=1}^{N}\left(\frac{1}{2} + \frac{1}{4}\mathbb{E}\{e^{j2\Psi_{in}} + e^{-j2\Psi_{in}}\} - \frac{1}{4}(\mathbb{E}\{e^{j\Psi_{in}} + e^{-j\Psi_{in}}\})^2\right) \tag{42}$$

$$= \frac{1}{N^2}\sum_{n=1}^{N}\left(\frac{1}{2} + \frac{1}{4}e^{j2\Psi^0_{in}}(c'_i - c_i^2) + \frac{1}{4}e^{-j2\Psi^0_{in}}(c'_i - c_i^2)^* - \frac{1}{2}|c_i|^2\right) \tag{43}$$

$$= \frac{1}{2N}(1 - |c_i|^2) + \sum_{n=1}^{N}\left(\frac{1}{4}e^{j2\Psi^0_{in}}(c'_i - c_i^2) + \frac{1}{4}e^{-j2\Psi^0_{in}}(c'_i - c_i^2)^*\right) \tag{44}$$

$$= \frac{1}{2N}(1 - |c_i|^2) + \frac{1}{2N}\text{Re}\{(c'_i - c_i^2)\beta^0_N\} \tag{45}$$

$$= \frac{1}{2N}(1 - |c_i|^2)(1 + o(1)) \tag{46}$$

where (40) is due Ψ_{in} being independent for different n; (43) is obtained using (6.6) and substituting the below expectation terms:

$$\mathbb{E}\{e^{j2\Psi_{in}}\} = e^{j2\Psi^0_{in}}c'_i, \quad c'_i = \mathbb{E}\{e^{j2\Delta\Psi_{in}}\} \tag{47}$$

$$\mathbb{E}\{e^{j\Psi_{in}}\} = e^{j\Psi^0_{in}}c_i, \quad c_i = \mathbb{E}\{e^{j\Delta\Psi_{in}}\} \tag{48}$$

(44) follows by removing the fixed terms out of the summation and (45) is obtained using (6.45) and the fact that $X + X^* = 2\text{Re}\{X\}$. Note that if FP holds at double the frequency, $\beta^0_N \sim N^{-1}$, see (6.59). Hence, the second term in (45) scales as N^{-2} and (46) follows, where $o(1) \to 0$ as $N \to \infty$. Using (37) and following the steps in (39)-(46) for α_{N2}, one

obtains:

$$\sigma_{Nk}^2 = \frac{1}{2}\sigma_N^2 - \frac{(-1)^k}{2N}\mathrm{Re}\{(c_i' - c_i^2)\beta_N^o\} \tag{49}$$

$$= \frac{1}{2}\sigma_N^2(1 + o(1)) \approx \frac{1}{2}\sigma_N^2 \tag{50}$$

Next, to obtain the asymptotic distribution of $|\alpha_N|^2$, we shown that α_{N1} and α_{N2} are asymptotically uncorrelated and hence independent. To this end, we use the normalized random variables x and y:

$$x = \sigma_{N1}^{-1}\alpha_{N1} \sim \mathcal{N}(\mathbb{E}\{\alpha_{N1}\}, 1) \tag{51}$$

$$y = \sigma_{N2}^{-1}\alpha_{N2} \sim \mathcal{N}(\mathbb{E}\{\alpha_{N2}\}, 1) \tag{52}$$

and evaluate their covariance $r_N = \mathrm{Cov}(x, y)$:

$$r_N = \mathbb{E}\{xy\} - \mathbb{E}\{x\}\,\mathbb{E}\{y\} \tag{53}$$

$$= (\sigma_{N1}\sigma_{N2})^{-1}(\mathbb{E}\{\alpha_{N1}\alpha_{N2}\} - \mathbb{E}\{\alpha_{N1}\}\,\mathbb{E}\{\alpha_{N2}\}) \tag{54}$$

$$= \frac{1}{\sigma_{N1}\sigma_{N2}N^2}\sum_n\sum_m(\mathbb{E}\{\cos(\Psi_{in})\sin(\Psi_{im})\} - \mathbb{E}\{\cos(\Psi_{in})\}\,\mathbb{E}\{\sin(\Psi_{im})\}) \tag{55}$$

$$= \frac{1}{\sigma_{N1}\sigma_{N2}N^2}\sum_n(\frac{1}{2}\,\mathbb{E}\{\sin(2\Psi_{in})\} - E\{\cos(\Psi_{in})\}\,\mathbb{E}\{\sin(\Psi_{in})\}) \tag{56}$$

$$= \frac{1}{\sigma_{N1}\sigma_{N2}N^2}\sum_n(\frac{1}{2}\,\mathbb{E}\{\frac{e^{j2\Psi_{in}} - e^{-j2\Psi_{in}}}{2j}\} - \mathbb{E}\{\frac{e^{j\Psi_{in}} + e^{-j\Psi_{in}}}{2}\}\,\mathbb{E}\{\frac{e^{j\Psi_{in}} - e^{-j\Psi_{in}}}{2j}\}) \tag{57}$$

$$= \frac{1}{\sigma_{N1}\sigma_{N2}N^2}\sum_n\frac{1}{4j}((c_i' - c_i^2)e^{j2\Psi_{in}^0} - (c_i' - c_i^2)^* e^{-j2\Psi_{in}^0}) \tag{58}$$

$$= \frac{1}{2\sigma_{N1}\sigma_{N2}N}\mathrm{Im}\{(c_i' - c_i^2)\beta_N^0\} \tag{59}$$

$$= \mathrm{Im}\{(c_i' - c_i^2)\beta_N^0\}\left((1 - |c_i|^2)^2 - (\mathrm{Re}\{(c_i' - c_i^2)\beta_N^0\})^2\right)^{-1/2} \tag{60}$$

where (55) is obtained by substituting the real and imaginary parts of α_N, defined in (6.5), and the double sum in (55) turns into a single sum in (56) since Ψ_n is independent of Ψ_m for $n \neq m$ and so $\mathbb{E}\{\cos(\Psi_{in})\sin(\Psi_{im})\} = \mathbb{E}\{\cos(\Psi_{in})\}\,\mathbb{E}\{\sin(\Psi_{im}\}$. This results in the terms inside the summations with $n \neq m$ being zero. (58) is obtained by substituting

the expectation terms with (47) and (48). (59) is obtained using (6.45) and the fact that $-j(X - X^*) = 2\text{Im}\{X\}$ and finally, (60) follows by substituting σ_{N1} and σ_{N2} by (49). Therefore, using (60):

$$\lim_{N\to\infty} r_N = \lim_{n\to\infty} \frac{\text{Im}\{(c_i' - c_i^2)\beta_N^0\}}{\sqrt{(1-|c_i|^2)^2 - (\text{Re}\{(c_i' - c_i^2)\beta_N^0\})^2}} = 0 \tag{61}$$

since $\lim_{N\to\infty} \beta_N^0 = 0$ due to the FP holding at double the frequency. This shows that x and y are asymptotically uncorrelated and, since they are Gaussian, also independent of each other. Next, we use the following technical lemma to find the distribution of $|\alpha_N^2|$ [112, p. 447].

Lemma 14. *Let $(X_1, X_2, \ldots, X_k)$ be k independent, normally distributed random variables with means μ_i and unit variances. Then the random variable $X = \sum_{i=1}^k X_i^2$ is distributed according to the noncentral chi-squared distribution with parameters k as the degrees of freedom and λ as the noncentrality parameter.*

$$X = \sum_{i=1}^{k} X_i^2 \sim \chi_k^2(\lambda), \quad \lambda = \sum_{i=1}^{k} \mu_i^2 \tag{62}$$

$$\text{Var}\{X\} = 2(k + 2\lambda) \tag{63}$$

Since x and y are two unit variance and asymptotically independent random variables with means $\mathbb{E}\{\alpha_{N1}\}$ and $\mathbb{E}\{\alpha_{N2}\}$, one can use Lemma 14 and (49) and obtains:

$$\frac{\alpha_{N1}^2}{0.5\sigma_N^2} + \frac{\alpha_{N2}^2}{0.5\sigma_N^2} \sim \chi_2^2(\lambda) \tag{64}$$

and using (6.27) and (62) ,λ is

$$\lambda = \frac{\mathbb{E}\{\alpha_{N1}\}^2}{0.5\sigma_N^2} + \frac{\mathbb{E}\{\alpha_{N2}\}^2}{0.5\sigma_N^2} \tag{65}$$

$$= \frac{2|\,\mathbb{E}\{\alpha_N\}|^2}{\sigma_N^2} = 2\sigma_N^{-2}|c_i|^2|\alpha_N^0|^2 \tag{66}$$

and using (63) with $k = 2$ degrees of freedom and the λ defined in (66), the variance is:

$$\mathrm{Var}\{\frac{\alpha_{N1}^2}{0.5\sigma_N^2} + \frac{\alpha_{N2}^2}{0.5\sigma_N^2}\} = \frac{4}{\sigma_N^4}\,\mathrm{Var}\{|\alpha_N|^2\} \tag{67}$$

$$\approx 2(2 + 4\sigma_N^{-2}|c_i|^2|\alpha_N^0|^2) \tag{68}$$

Hence, using (64), the IUI is distributed as

$$|\alpha_N|^2 = \alpha_{N1}^2 + \alpha_{N2}^2 \sim 0.5\sigma_N^2\chi_2^2(\lambda) \tag{69}$$

and using (68), $\mathrm{Var}\{|\alpha_N|^2\}$ can be approximated as

$$\sigma_{|\alpha_N|^2}^2 = \mathrm{Var}\{|\alpha_N|^2\} \approx \sigma_N^4 + 2|c_i|^2|\alpha_N^0|^2\sigma_N^2 \tag{70}$$