UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament de Teoria del Senyal
i Comunicacions

telecos
BCN

PHYSICALLY CONSISTENT WIRELESS COMMUNICATIONS WITH STATISTICAL CHANNEL STATE INFORMATION

A PhD dissertation submitted to
the Departament de Teoria del Senyal i Comunicacions
of the Universitat Politècnica de Catalunya by

Aniol Martí Espelt

In partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

Supervised by

Jaume Riba Sagarra
Meritxell Lamarca Orozco

Barcelona, June 2026

Un fracàs no s'improvisa.
– Joan Fuster

Cadascú basteix les seves quimeres
d'acord amb les seves il·lusions més recòndites.
– Pere Calders

Abstract

The deployment of 5G and 6G wireless networks enforces a paradigm shift in communication system design, driven by the move towards sub-THz frequencies and extremely large antenna arrays. These advancements leave traditional channel modeling—which often relies on far-field assumptions and neglects electromagnetic interactions between antenna elements—physically inconsistent. Furthermore, the overhead associated with acquiring instantaneous channel state information (CSI) in massive multiple-input multiple-output (mMIMO) systems presents a critical bottleneck, particularly for low-latency or high-mobility communications.

This thesis addresses the aforementioned challenges by developing a framework for the design and analysis of noncoherent wireless communication systems that operate solely with statistical, rather than instantaneous, CSI. The core of this work is the establishment of a physically consistent channel model that accurately incorporates the effects of near-field spherical wavefronts and mutual coupling. We demonstrate that these complex physical phenomena can be effectively captured within a correlated Rayleigh fading model, providing a tractable yet realistic foundation for system analysis.

Using this framework, we investigate the performance of one-shot, energy-based communication schemes, which are particularly well suited for low-latency applications. A key result is the existence of a fundamental error floor at high signal-to-noise ratio (SNR) for constellations with more than two energy levels when no CSI is available at the transmitter. However, we also prove that this error vanishes as the number of receiver antennas grows, highlighting the channel hardening benefits of massive arrays.

A widely adopted receiver in energy-based noncoherent systems is the so-called energy detector. Although it is optimal under uncorrelated fading, its performance degrades significantly in correlated channels. To address this limitation, we introduce a novel class of quadratic detectors, including the best quadratic unbiased estimator (BQUE) as well as a practical implementation called assisted BQUE. These detectors leverage statistical CSI to achieve near-optimal performance. Furthermore, two strategies for enhancing reliability are proposed and evaluated: a constellation design methodology that minimizes the analytical symbol error rate by leveraging statistical CSI at the transmitter, and a permutational index modulation (PIM) scheme that introduces coding gain with minimal complexity.

Finally, the thesis explores the impact of model mismatch, revealing that noncoherent systems exhibit greater robustness to mutual coupling than their coherent counterparts. We also demonstrate that wavefront curvature can be exploited well beyond the classical Fraunhofer distance. Moreover, we show that large antenna arrays enable the multiplexing and low-complexity detection of multiple users, even when employing noncoherent processing.

Resum

El desplegament de les xarxes sense fils 5G i 6G imposa un canvi de paradigma en el disseny dels sistemes de comunicació, impulsat per l'avenç cap a les freqüències sub-THz i les agrupacions extremadament grans d'antenes. Aquests avenços deixen la modelització de canal tradicional —que sovint es basa en hipòtesis de camp llunyà i negligeix les interaccions electromagnètiques entre antenes— físicament inconsistent. A més a més, la càrrega associada a l'adquisició de la informació del canal (CSI) instantània en sistemes amb canal d'entrada i sortida múltiples massius (mMIMO) representa un coll d'ampolla crític, especialment per a comunicacions de baixa latència o alta mobilitat.

Aquesta tesi afronta els reptes anteriors mitjançant el desenvolupament d'un marc per al disseny i l'anàlisi de sistemes de comunicació sense fils no coherents que operen únicament amb CSI estadística, en lloc d'instantània. El nucli d'aquest treball és l'establiment d'un model de canal físicament consistent que incorpora amb precisió els efectes de les ones esfèriques en camp proper i de l'acoblament mutu. Demostrem que aquests fenòmens físics complexos poden ser representats de manera efectiva dins d'un model d'esvaïment de Rayleigh correlat, proporcionant una base tractable però realista per a l'anàlisi del sistema.

Amb aquest marc, investiguem el rendiment d'esquemes de comunicació de tir únic basats en energia, especialment adequats per a aplicacions de baixa latència. Un resultat clau és l'existència d'un terra d'error a relacions senyal-soroll (SNR) altes, per a constel·lacions amb més de dos nivells d'energia i quan no hi ha CSI al transmissor. Tanmateix, també demostrem que aquest error desapareix a mesura que creix el nombre d'antenes receptores, destacant els beneficis de l'enduriment de canal quan s'utilitzen agrupaments massius d'antenes.

Un receptor habitual en sistemes no coherents basats en energia és l'anomenat detector d'energia. Tot i que és òptim en esvaïment incorrelat, el seu rendiment es deteriora considerablement en canals correlats. Per superar aquesta limitació, introduïm una nova classe de detectors quadràtics, incloent-hi el millor detector quadràtic no esbiaixat (BQUE) i una implementació pràctica anomenada BQUE assistit. Aquests detectors aprofiten la CSI estadística per assolir un rendiment gairebé òptim. A més, es proposen i s'avaluen dues estratègies per millorar la fiabilitat del sistema: un mètode de disseny de constel·lacions que minimitza la probabilitat d'error aprofitant la CSI al transmissor, i un esquema de modulació amb permutació d'índexs (PIM) que introdueix guany de codificació amb una complexitat mínima.

Finalment, la tesi revela que els sistemes no coherents són més robustos a l'acoblament mutu que els coherents. També demostrem que la curvatura del front d'ona és aprofitable més enllà de la distància de Fraunhofer. Així mateix, les grans agrupacions d'antenes permeten multiplexar i detectar usuaris amb baixa complexitat, inclús amb processament no coherent.

Acknowledgements

Escriure la tesi no ha estat fàcil, però escriure aquestes línies ho és menys. He fet unes simulacions i resulta que el contingut tècnic de la tesi se'l llegiran unes set persones (fita superior), però en canvi els agraïments se'ls llegirà tothom que tingui la tesi als dits (bon dia tinguis, personal del servei de biblioteques que l'està catalogant). Però no és només això! També és importantíssim que no et deixis ningú, perquè si en una equació et deixes un signe ningú s'enfadarà, però si als agraïments t'oblides de mencionar a segons qui potser et passes la resta de dinars de Nadal relegat a la taula de la mainada.

Permeteu-me que faci una pausa—evidentment que m'ho permeteu, que és la meva tesi—per reconèixer que quan duia gairebé una pàgina d'agraïments em vaig plantejar si esborrar-ho tot i simplement escriure «tots els que hi heu de ser ja ho sabeu», però prefereixo no quedar com un pocavergonya (i sobretot no quedar-me sense canelons per Nadal).

Bé, ara sí, *comencem*. Primer de tot les formalitats: Jaume i Xell, gràcies per aguantar la meva tossuderia i impertinències. De fet, segurament tots dos esteu fent que no amb el cap mentre llegiu aquestes línies. En qualsevol cas, alguns resultats d'aquesta tesi són fruit de discussions que hem tingut i que molt probablement jo sol no hauria trobat. En aquest sentit també he de donar les gràcies a la resta de membres de SPCOM com també a la Nuria Duffo, l'Albert Aguasca, en Jordi Romeu, en Joan M. Rius i l'Ignasi Corbella. La meva falta de coneixement d'antenes i circuits d'alta freqüència va propiciar que els molestés més d'una vegada.

A part dels companys que han contribuït directament en la tesi, hi ha més gent de la universitat a qui vull donar les gràcies: Sisco Vallverdú, Jaume Comellas, Joan M. Gené, Carlos López, Xavier Gràcia, Sergio Zarza i Josep Pujal. Evidentment no em puc deixar l'Orestes Mas. A part d'haver adaptat la seva plantilla de LaTeX per aquesta tesi, les parets del seu despatx han estat (massa) sovint testimoni de renecs i lamentacions.

Vorrei anche ringraziare tutta la gente di UniPi. Sebbene parlavamo in inglese, credo che posso scrivere queste parole in italiano. Luca, Giacomo, Lorenzo, Irched (spero che adesso parli italiano) e specialmente Leo; grazie a tutti per avermi fatto sentire benvenuto.

Mereixen un paràgraf exclusiu els amics que han fet la tesi paral·lelament amb mi: Marc, Ferran i Mireia. M'ho he passat molt bé amb vosaltres, i de fet segueixo passant-m'ho bé cada cop que ens veiem! També han de ser en aquest paràgraf en Juan, en Rachid i en Jordi, que tot i haver-hi coincidit menys, han contribuït a fer aquesta etapa molt més agradable.

Evidentment, també he de donar les gràcies a la gent que m'ha alimentat aquests anys. I no, no parlo de l'AGAUR, em refereixo al personal de Camins (especialment en Jordi) i de la FIB (sí, en Marc i en Gerardo també).

Vull dedicar aquestes darreres línies (i pàgina, pel que sembla) a la meva família i amics (intentaré llistar-vos a tots, però si us plau disculpeu-me si em deixo algú). En primer lloc, als meus pares per tot el suport al llarg d'aquests anys, i a les meves àvies Maria i Conxita, que tot i que ja no hi eres quan vaig començar la tesi sé que n'estaries molt orgullosa. També he de donar-li les gràcies a l'Àlex per ser-hi sempre, sota qualsevol circumstància. I ara sí, per acabar: Gerard, Albert, Martí, Oriol G., Oriol C., Cris, David R., Anna, Oriol X., Vicenç, Roc, Irene, Eudald, Marc, Núria, Adrià, David P., Abel, Abde, Jan, Josep i Héctor. A tots vosaltres, agrair-vos les bones estones que em feu passar.

This thesis has been supported by:

- Grant 2023 FI "Joan Oró" 00050 by Departament de Recerca i Universitats de la Generalitat de Catalunya and the ESF+.
- Project MAYTE (PID2022-136512OB-C21) by MICIU/AEI/10.13039/501100011033 and ERDF/EU.
- Grant 2021 SGR 01033 by Departament de Recerca i Universitats de la Generalitat de Catalunya.

Acronyms

ABQUE assisted best quadratic unbiased estimator

AC alternating current

AMC adaptive modulation and coding

ASK amplitude-shift keying

AWGN additive white Gaussian noise

BLUE best linear unbiased estimator

BQUE best quadratic unbiased estimator

BS base station

CLT central limit theorem

CMS canonical minimum scattering

CRB Cramér–Rao bound

CSI channel state information

CSIR channel state information at the receiver

CSIT channel state information at the transmitter

DPSK differential phase shift keying

ED energy detector

EHF extremely high frequency

eMBB enhanced mobile broadband

FDD frequency-division duplex

HSNR high signal-to-noise ratio

IIoT industrial internet of things

IM index modulation

IoT internet of things

ISAC integrated sensing and communications

ITU International Telecommunication Union

KCL Kirchhoff's current law

KKT Karush–Kuhn–Tucker

KLD Kullback–Leibler divergence

KVL Kirchhoff's voltage law

LHS left-hand side

LLR log-likelihood ratio

LMMSE linear minimum mean squared error

LNA low-noise amplifier

LoS line-of-sight

LTE long term evolution

MI mutual information

MIMO multiple-input multiple-output

MISO multiple-input single-output

ML maximum likelihood

mMIMO massive multiple-input multiple-output

mMTC massive machine-type communication

MSE mean squared error

MVUE minimum variance unbiased estimator

NLoS non-line-of-sight

NUSW non-uniform spherical wave

OFDM orthogonal frequency-division multiplexing

OTFS orthogonal time frequency space

PAM pulse-amplitude modulation

PDF probability density function

PEP pairwise error probability

PIM permutational index modulation

PM permutation modulation

PSK phase shift keying

QMMSE quadratic minimum mean squared error

RHS right-hand side

SE spectral efficiency

SER symbol error rate

SIMO single-input multiple-output

SINR signal-to-interference-plus-noise ratio

SISO single-input single-output

SNR signal-to-noise ratio

TDD time-division duplex

UE user equipment

ULA uniform linear array

UPA uniform planar array

UPW uniform plane wave

URLLC ultra-reliable and low-latency communication

USTM unitary space-time modulation

USW uniform spherical wave

V2X vehicle to everything

Notation

α, a	Scalar
$\mathbf{a}$	Column vector
$\mathbf{A}$	Matrix
$[\mathbf{a}]_i$	i-th element of vector $\mathbf{a}$
$[\mathbf{A}]_{i,j}$	Element at the i-th row and j-th column of $\mathbf{A}$
$[\mathbf{A}]_{i,*}$	All elements in i-th row of $\mathbf{A}$
$[\mathbf{A}]_{*,j}$	All elements in j-th column of $\mathbf{A}$
$\mathfrak{R}(a)$	Real part of a complex number a
$\mathfrak{I}(a)$	Imaginary part of a complex number a
$\lvert a \rvert$	Absolute value of a
a^*	Complex conjugate of a
$\lVert\mathbf{a}\rVert_p$	p-norm of vector $\mathbf{a}$
$\mathbf{A}^{\mathrm{T}}$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^*$	Conjugate of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{H}}$	Conjugate transpose of matrix $\mathbf{A}$
$\lVert\mathbf{A}\rVert_{\mathrm{F}}$	Frobenius norm of matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\mathrm{rank}(\mathbf{A})$	Rank of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	Determinant of matrix $\mathbf{A}$
$\mathrm{col}(\mathbf{A})$	Column space of matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{A})$	Vector with the main diagonal of the matrix $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{vec}(\mathbf{A})$	Column-wise vectorization of matrix $\mathbf{A}$

$\mathbf{0}_{M\times N}$	$M \times N$ matrix of zeros
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{N\times N}$	Set of $N \times N$ matrices with real entries
$\mathbb{C}$	Set of complex numbers
$\mathbb{C}^{N\times N}$	Set of $N \times N$ matrices with complex entries
ℓ^p	Space of p-summable sequences
e	Euler's number
j	Imaginary unit
$\exp(\cdot)$	Exponential function
$\log_a(\cdot)$	Logarithm in base a
$\log(\cdot)$	Logarithm in base 2
$\ln(\cdot)$	Natural logarithm
$\operatorname{sinc}(\cdot)$	Normalized sinc function
$\Gamma(\cdot)$	Gamma function
$\mathrm{H}(\cdot)$	Shannon entropy
$\mathrm{h}(\cdot)$	Differential entropy
$\mathscr{P}(a_0, \ldots, a_N)$	Polynomial $a_N x^N + a_{N-1}x^{N-1} + \cdots + a_0$
$\max(\mathscr{X})$	Maximum value in the set $\mathscr{X}$
$\min(\mathscr{X})$	Minimum value in the set $\mathscr{X}$
$\arg\max_{x\in\mathscr{X}} f(x)$	Value of x that maximizes f in the set $\mathscr{X}$
$\arg\min_{x\in\mathscr{X}} f(x)$	Value of x that minimizes f in the set $\mathscr{X}$
a, $\mathbf{a}$, $\mathbf{A}$	Random variables
$\hat{a}$	Estimate of a
$\mathrm{P}(\cdot)$	Probability function
$\mathscr{N}(\mathbf{b}, \mathbf{C})$	Normal distribution with mean $\mathbf{b}$ and covariance matrix $\mathbf{C}$
$\mathscr{CN}(\mathbf{b}, \mathbf{C})$	Complex normal distribution with mean $\mathbf{b}$ and covariance matrix $\mathbf{C}$
$\mathrm{F}_{\chi^2}(\cdot\,; n)$	Cumulative distribution function of a chi-squared random variable with n degrees of freedom

$\mathrm{E}_{\mathrm{a}}[\cdot]$	Expectation with respect to the distribution of a
$\xrightarrow{\mathrm{d}}$	Convergence in distribution
$\mathrm{Q}(\cdot)$	Tail distribution function of the standard normal distribution
$\mathrm{var}(\cdot)$	Variance
$\mathrm{b}(\cdot)$	Bias
$\mathrm{MSE}(\cdot)$	Mean squared error
(r, θ, ϕ)	Spherical coordinates following the physics convention
λ	Wavelength in meters
$k = 2\pi/\lambda$	Angular wavenumber
$\eta \approx 376.73\,\Omega$	Impedance of free space
D	Antenna/array aperture
$d_{\mathrm{r}} = 0.62\sqrt{D^3/\lambda}$	Fresnel distance
$d_{\mathrm{F}} = 2D^2/\lambda$	Fraunhofer distance
d_{UP}	Arbitrary uniform-power distance
P_ϵ	Error probability
N	Number of antennas
M	Constellation size
$\rho_{\mathrm{N}} \in \mathbb{C}$	Noise correlation coefficient
$\rho \in \mathbb{C}$	Exponential channel correlation coefficient

Contents

List of tables

List of figures

CHAPTER 1

Introduction

Communication can be defined as the capacity to transmit information from a source to a receiver. Human beings have been able to communicate through long distances for years using methods such as smoke signals, or since the 19th century with electrical telecommunication systems such as the telegraph [1].

Nevertheless, it was not until the following century that a mathematical basis for communication was established. This milestone was achieved by Claude Shannon in 1948, with his seminal work "A mathematical theory of communication" [2]. Since then, communication systems have continuously evolved, new standards have been released and data rates have incremented drastically.

Current communication systems, however, are not only expected to offer very high data rates but also to accommodate a large number and variety of devices such as mobile phones, sensor networks or even vehicles [3]. For 5G systems, the International Telecommunication Union (ITU) already identified three use cases [4]:

- **Enhanced mobile broadband (eMBB):** this mode is an extension of the existing 4G long term evolution (LTE) broadband services. Its goal is to provide higher data rates (up to 10 Gb/s in uplink and 20 Gb/s in downlink) in very dense areas to allow applications such as augmented reality or video streaming.
- **Ultra-reliable and low-latency communication (URLLC):** critical communications such as autonomous vehicles, vehicle to everything (V2X), and remote surgery must have very low latency ($\leq$ 5 ms, end-to-end) and error rate ($\leq 10^{-5}$). In this mode, data is usually transferred in small packets.
- **Massive machine-type communication (mMTC):** to support sensor networks and industrial internet of things (IIoT), systems must be able to accommodate a large number of devices ($\sim 10^6$ devices/km^2) with low data rate and energy requirements.

In order to fulfill these requirements, the 5G standard considers operating in extremely high frequency (EHF) bands (*i.e.* mmWave), mostly underutilized due to its high atmospheric attenuation. Another key technology that has already been implemented in 5G is massive multiple-input multiple-output (mMIMO), which consists in deploying a large number of antennas (up to a few hundred) at the base stations (BSs). In mMIMO, the spectral efficiency

(SE) grows monotonically with the number of antennas and the spatial resolution increases [5], which means that sensing capabilities of the array, data rate and link reliability also increase.

Nevertheless, these technologies alone are insufficient to address the demands of next-generation wireless communications, particularly the 6G use cases announced by the ITU in 2023 [6]. Specifically, three extensions of the 5G usage scenarios are envisioned: immersive communications (eMBB+), massive communication (mMTC+), and hyper-reliable low-latency communications (URLLC+). In addition, three entirely new use cases are introduced: integrated sensing and communications (ISAC), integrated artificial intelligence and communications, and ubiquitous connectivity [7].

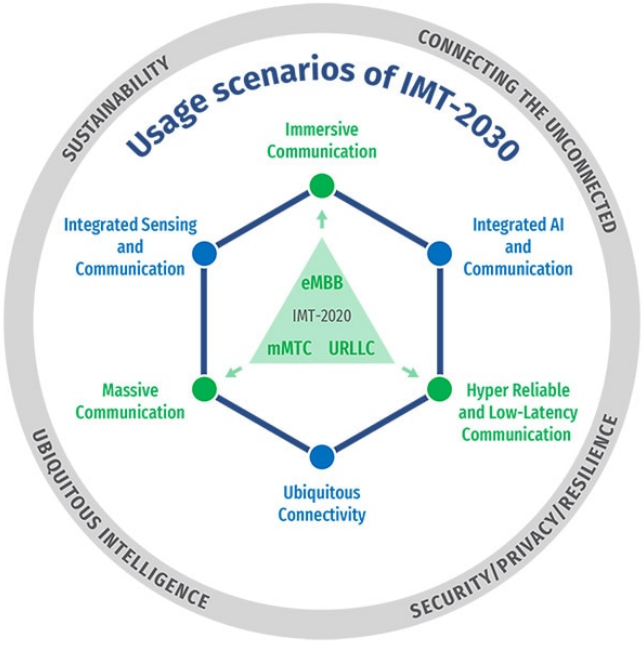


Figure 1.1: Usage scenarios of 5G (IMT-2020) and 6G (IMT-2030).

Source: ITU-R M.2160-0 [6].

Two strategies proposed to address these new scenarios are to further increase the number of antennas—either by building larger arrays or by placing them more densely—and to raise the operating frequency to the sub-THz bands [3], [7], [8]. Inevitably, such approaches result in near-field propagation conditions, and when antennas are spaced closer than half a wavelength, mutual coupling effects in the array must also be taken into account.

On the other hand, the massive number of antennas imposes significant complexity constraints on channel state information (CSI) acquisition, especially in frequency-division duplex (FDD) systems where channel reciprocity does not hold. In time-division duplex (TDD), by contrast, performance is instead limited by the number of users, which in mMTC scenarios is expected to be very large. Thus, regardless of the duplexing scheme, CSI acquisition remains a major challenge in future wireless communications, and developing efficient or complementary strategies for it is central to the design of 6G networks.

1.1 Scope

This dissertation addresses the design and performance of wireless communication systems that operate with statistical, rather than instantaneous, CSI. This paradigm, in which the receiver lacks knowledge of the instantaneous CSI, is usually referred to as *noncoherent communications*. In early analog communication systems, the term noncoherent indicated that the receiver had no knowledge of the channel phase. Indeed, the first mobile systems were noncoherent due to technical limitations; however, as soon as technology advanced sufficiently, systems began to exploit phase information, since doing so is more robust to additive noise [9, Appendix D].

Recent studies [3], [10], [11] have pointed out that noncoherent communication may play a critical role in future wireless networks. This renewed interest stems from the observation that operating without full instantaneous channel knowledge can substantially reduce the overhead traditionally required for channel estimation. Such overhead becomes particularly significant in systems with very large antenna arrays or in environments characterized by rapid channel variations [12]. Among the different noncoherent approaches—*i.e.* differential schemes, Grassmannian signaling and energy detection (see section 2.4.1)—our investigation focuses specifically on one-shot, energy-based communication schemes. These schemes rely on channel statistics rather than instantaneous channel knowledge, making them particularly well suited for short-packet, ultra-low-latency applications such as the IIoT [13].

For these reasons, recent works have begun to focus on improving the SE of noncoherent systems [10], [14]–[16]. A central challenge, however, is that many of these analyses still rely predominantly on the isotropic Rayleigh fading model. This assumption is physically inconsistent with the propagation characteristics envisioned for 6G, such as transmissions in the sub-THz bands and the use of extremely large or dense antenna arrays. In such scenarios, phenomena like sparse scattering, wavefront curvature in the near field, and mutual coupling between antenna elements become dominant, inducing significant and structured channel correlation [8], [17], [18]. Failing to account for these features can lead to design inconsistencies, such as underestimating system performance or mispredicting array behavior. Furthermore, the short channel coherence time associated with the small operating wavelengths in the sub-THz bands aggravate the challenges of channel estimation.

Motivated by these challenges, the core contribution of this thesis is twofold. First, we develop a physically consistent channel modeling framework that rigorously incorporates near-field propagation and mutual coupling effects. We demonstrate that these complex physical interactions can be effectively captured within a generalized correlated Rayleigh fading model. Second, using this robust framework, we conduct a thorough analysis of one-shot energy-based noncoherent systems. A substantial part of this thesis is consequently devoted to investigating their fundamental performance limits, proposing novel detection schemes, and developing performance enhancement techniques under these realistic, arbitrarily correlated channel conditions. Furthermore, we examine the performance of communication systems that rely on model-based receivers, and compare the behaviors predicted by physically consistent models with those obtained from classical models that ignore fundamental characteristics of the wireless propagation channel.

1.2 Organization and contributions

The present thesis is organized into seven chapters: an introduction, a state-of-the-art review (which also includes a novel contribution in chapter 3), three core chapters forming the main body of the work, and a final chapter that concludes the thesis and discusses future research lines. In what follows, we briefly summarize the contents of each chapter and outline the technical works on which they are based:

- **Chapter 2** introduces different models of wireless channels in the near field. It also discusses the main effects of mobility, such as fading, and their implications for channel estimation. Finally, it presents the noncoherent communication paradigm along with its most common approaches.

- **Chapter 3** focuses on the phenomenon of mutual coupling that arises when different antennas in an array interact. In particular, the study leverages the multiport communication theory framework, which is introduced in its most general form in this same chapter. Furthermore, a novel analysis of the coupling effects from the receiver to the transmitter is presented. The importance of this contribution is that, under the assumption of no coupling from the receiver to the transmitter, the communication channel can still be modeled as a correlated Rayleigh fading channel.

  The work presented in this chapter has resulted in the following publications:

  - A. Martí, J. Riba, M. Lamarca, and X. Gràcia, "Asymptotic analysis of near-field coupling in massive MISO and massive SIMO systems", *IEEE Communications Letters*, vol. 28, no. 8, pp. 1929–1933, 2024.

- **Chapter 4** focuses on one of the three most widely used noncoherent communication approaches: energy detection. The performance of one-shot systems over channels with correlated Rayleigh fading and colored Gaussian noise is carefully analyzed. We provide a theoretical study of the limitations of unipolar pulse-amplitude modulation (PAM), revealing the existence of a fundamental error floor at high signal-to-noise ratio (SNR) for constellations with more than two energy levels when no (statistical) CSI is available at the transmitter. Finally, we introduce a design framework for quadratic detectors, together with the analytical expression of their asymptotic symbol error probability.

  The work presented in this chapter has resulted in the following publications:

  - M. Vilà-Insa, A. Martí, J. Riba, and M. Lamarca, "Quadratic detection in noncoherent massive SIMO systems over correlated channels", *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14 259–14 272, 2024.

- **Chapter 5** introduces two techniques to improve the reliability of the one-shot schemes presented in the previous chapter. The first approach proposes an optimization method to design constellations that minimize the asymptotic symbol error probability derived in chapter 4, assuming statistical CSI is available at the transmitter (CSIT). The second introduces a coding technique based on permutational index modulation (PIM) that embeds information in the ordering by which a set of values is mapped onto different subcarriers, relying only on statistical CSI at the receiver (CSIR).

The work presented in this chapter has resulted in the following publications:

- A. Martí, M. Vilà-Insa, J. Riba, and M. Lamarca, "Constellation design for quadratic detection in noncoherent massive SIMO communications", in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lucca, Italy, 2024, pp. 566–570.
- M. Vilà-Insa, A. Martí, M. Lamarca, and J. Riba, "Low-complexity detection of permutational index modulation for noncoherent communications", *IEEE Wireless Communications Letters*, vol. 14, no. 10, pp. 3059–3063, 2025.

- **Chapter 6** analyzes the error probability of receivers employing parametric models in the presence of model mismatch, with special emphasis on energy-based schemes. Two main scenarios are considered: in the first, a base station (BS) equipped with a densely packed antenna array detects the signal of a single user while neglecting mutual coupling. In the second, we consider a multiuser environment where users may be located in either the near field or the far field. A spatial channel correlation model valid in both regions is adopted. With this model, we demonstrate that energy-based noncoherent systems can exploit the curvature of spherical wavefronts even beyond the Fraunhofer distance. Furthermore, we show that large antenna arrays enable near-optimal detection at low computational complexity.

  The work presented in this chapter has resulted in the following publications:

  - A. Martí, L. Sanguinetti, J. Riba, and M. Lamarca, "Coherent and noncoherent detection in dense arrays: Can we ignore mutual coupling?", in *2025 33rd European Signal Processing Conference (EUSIPCO)*, Palermo, Italy, 2025, pp. 2027–2031.
  - A. Martí, L. Sanguinetti, M. Lamarca, and J. Riba, "Harnessing wavefront curvature and spatial correlation in noncoherent MIMO communications", *IEEE Wireless Communications Letters*, vol. 14, no. 8, pp. 2461–2465, 2025.

- **Chapter 7** concludes the thesis and outlines potential directions for future research.

Other contributions

The author of this thesis has also contributed to the following works during the PhD:

- A. Martí, J. Portell, J. Riba, and O. Mas, "Context-aware lossless and lossy compression of radio frequency signals", *Sensors*, vol. 23, no. 7, 2023.
- A. Martí, F. de Cabrera, and J. Riba, "On the estimation of Tsallis entropy and a novel information measure based on its properties", *IEEE Signal Processing Letters*, vol. 30, pp. 818–822, 2023.

A complete outline of the thesis is depicted in figure 1.2.

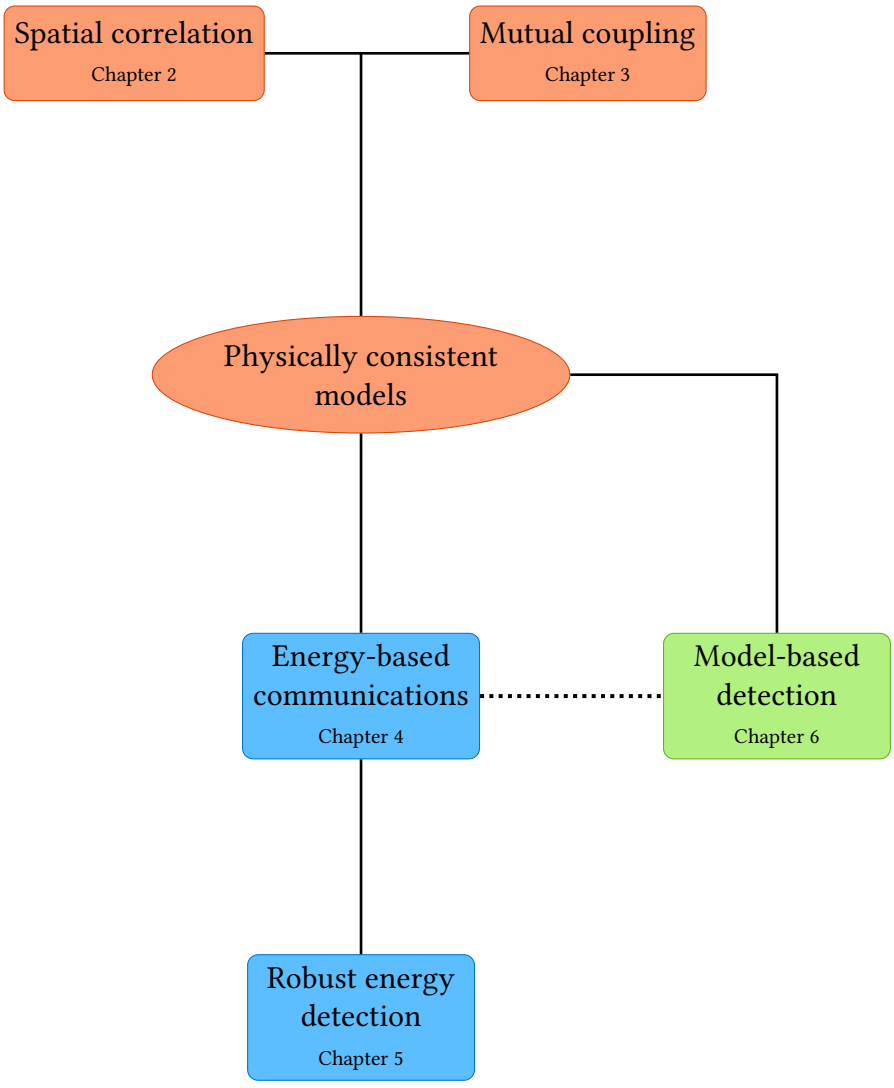


Figure 1.2: Thesis outline.

CHAPTER 2

Foundations of wireless propagation

The behavior of a communication channel—whether wireless or not—is fundamentally governed by the laws of physics, particularly Maxwell's equations [18], [27]. While solving these equations yields accurate channel models, the resulting expressions are too complex and difficult to work with directly. Fortunately, they can be simplified and adapted to specific scenarios, enabling practical application in communication system design [8].

Traditional communication systems typically assumed uniform planar waves, which significantly simplified modeling. However, in modern systems operating at high frequencies or employing large arrays—either physically or in number of antennas—, such approximations are no longer valid and more accurate models are needed to capture previously negligible effects [7], [8], [28].

At the same time, the wireless medium is not only shaped by deterministic field regions but also by random variations introduced by the environment and user mobility. Large obstacles block or attenuate signals, producing shadowing [29, Sec. 3.3]. Multipath reflections, diffractions, and scatterings combine causing rapid fluctuations in received power and phase [29, Sec. 3.4].

This chapter reviews these foundations from physical laws to mobile channel models. We begin by examining the near-field and far-field regions, focusing on wavefront curvature and its implications for antenna arrays. We then introduce channel modeling approaches that account for spherical, parabolic and planar wavefronts. Finally, we connect these physical models to mobile communication scenarios and we discuss how these effects shape the distinction between coherent and noncoherent communications.

2.1 Electromagnetic field regions

First of all, we examine the regions of the electromagnetic field around a radiating object, such as an antenna or an array. Traditionally, three regions have been defined: the reactive near field, the radiative near field and the far field [30, Sec. 2.2.4]. However, more recent literature in communication theory further subdivides the radiative near field in uniform and nonuniform regions, depending on whether power variations, in addition to phase variations, must be accounted for in the modeling [7], [28].

It is important to note that these regions are not physically real, in the sense that they arise as a convenient way to simplify the electric field expression in non-trivial antennas, as we shall see next.

2.1.1 Electrically small antennas

We begin by considering a thin antenna of length l, located along the z axis, and operating at wavelength λ. Such an antenna is said to be electrically small when $l \ll \lambda$. Assuming a uniform current density, the electric field at a distance r perpendicular to the propagation direction is

$$E(r) = \frac{\mathrm{j}\eta I_1 l}{2\lambda r}\,\mathrm{e}^{-\mathrm{j}kr}\left(1 - \frac{\mathrm{j}}{kr} - \frac{1}{(kr)^2}\right), \tag{2.1}$$

where $k = 2\pi/\lambda$ is the angular wavenumber, η is the impedance of free space and I_1 is the antenna driving current. This expression corresponds to the polar component of the electric field radiated by a Hertzian dipole, which is described in detail in section 3.2.1.

The term in parentheses represents the so-called evanescent waves, which do not contribute to radiated power but only to reactive power. In many practical scenarios, this term can be omitted, since it approaches 1 when $kr \gg 1$ [30, Sec. 4.2.2]:

$$\left|1 - \frac{\mathrm{j}}{kr} - \frac{1}{(kr)^2}\right|^2 = 1 - \frac{1}{(kr)^2} + \frac{1}{(kr)^4} \approx 1, \quad r \geq 2\lambda. \tag{2.2}$$

Therefore, the electric field at $r \geq 2\lambda$ can be approximated by

$$E(r) = \frac{\mathrm{j}\eta I_1 l}{2\lambda r}\,\mathrm{e}^{-\mathrm{j}kr} = \frac{E_0}{\sqrt{4\pi r}}\,\mathrm{e}^{-\mathrm{j}kr}, \quad E_0 \in \mathbb{C}, \tag{2.3}$$

which corresponds to the field of a spherical wave. The zone where this approximation is valid is referred to as the far field, whereas the region $r \leq \lambda$ is the near field and $\lambda < r < 2\lambda$ is the transition zone of the electrically small antenna [8], [30, Sec. 4.2].

2.1.2 Electrically large antennas

To illustrate the behavior of electrically large antennas,[1] we consider an electrically small antenna located at (x, y, z) transmitting toward an antenna of length $D > \lambda$ centered along the z axis, as sketched in figure 2.1.

For now, assume that communication takes place in the radiation zone [30, Sec. 4.2], so that the evanescent waves term can be neglected. In this case, the electric field at a point $(0, 0, z')$ of the receiving antenna corresponds to that of a spherical wave:

$$E(x, y, z) = \frac{E_0}{\sqrt{4\pi R^2}}\mathrm{e}^{-\mathrm{j}kR}, \quad R = \sqrt{x^2 + y^2 + (z - z')^2}. \tag{2.4}$$

The field regions arise from subsequent approximations of R, by limiting the maximum phase and amplitude errors across the antenna. Expanding the expression of R in (2.4), it can be written as

$$R = \sqrt{x^2 + y^2 + (z - z')^2} = \sqrt{r^2 + \left(-2rz'\cos\theta + z'^2\right)}, \tag{2.5}$$

[1] An electrically large antenna can be modeled as a combination of many electrically small antennas [30, Ch. 4].

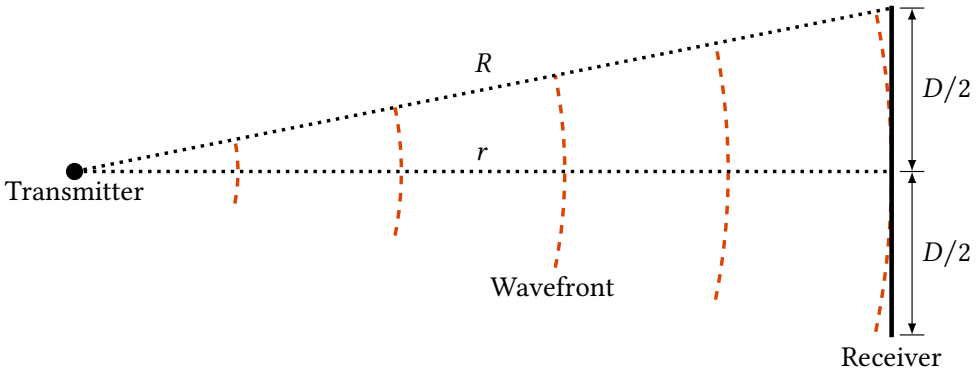


Figure 2.1: A point-source transmitter impinging on an aperture receiver located at a distance r with an angle of incidence $\theta = \pi/2$ rad. The resulting wavefront curvature at the receiver is illustrated.

where $r = \sqrt{x^2 + y^2 + z^2}$ and $z = r\cos\theta$, with θ the inclination angle of the source.

Using the binomial expansion [31, Ch. 23], (2.5) can be written as

$$R = r - z'\cos\theta + \frac{1}{r}\left(\frac{z'^2}{2}\sin(\theta)^2\right) + \frac{1}{r^2}\left(\frac{z'^3}{2}\cos\theta\sin(\theta)^2\right) + \cdots \tag{2.6}$$

It is worth mentioning that (2.3) is obtained by considering only the first term in (2.6), which is equivalent to setting $z' = 0$ due to the antenna being electrically small.

Far-field approximation

The next logic step is to consider not only the first term of the expansion but also the second one, so $R \approx r - z'\cos\theta$. The maximum error introduced with this approximation is $D^2/8r$, which corresponds to the maximum of the third term in (2.6) when $z' = D/2$ and $\theta = \pi/2$ rad [30, Sec. 4.4.1].

A known result from antenna theory is that phase errors below $\pi/8$ rad can be neglected when analyzing antenna parameters [30, Sec. 4.4.1], so it must be fulfilled that

$$\frac{kD^2}{8r} \leq \frac{\pi}{8} \implies r \geq \frac{2D^2}{\lambda} = d_\mathrm{F}. \tag{2.7}$$

This distance is the *Fraunhofer distance* and sets the boundary between the near field and the far field.

As for power–or, equivalently, amplitude—variations, the maximum variation is given by

$$\frac{r^2}{\left(r + \frac{D^2}{8r}\right)^2}, \tag{2.8}$$

which is the ratio between the approximated value of R^2 at the center of the antenna and the value at the edge.

Different uniform-power distance definitions have been proposed in literature [7]. For instance, in [32], [33], the *Björnson distance* $d_\mathrm{B} = 2D$ was proposed, which yields a maximum power variation of approximately 0.94. Since $d_\mathrm{F} = d_\mathrm{B}\frac{D}{\lambda} \geq d_\mathrm{B}$ for $D > \lambda$, power variations

become negligible in the far field of an electrically large antenna, and the wave can be approximated as planar. In what follows, when referring to a generic uniform-power distance, we denote it by d_{UP}.

In conclusion, assuming that $r \geq d_{\mathrm{F}}$, the far-field approximation can be summarized as

$$R \approx r \quad \text{for amplitude terms,} \tag{2.9}$$
$$R \approx r - z' \cos\theta \quad \text{for phase terms.} \tag{2.10}$$

Near-field approximation

The situation changes when the observation point is located at a distance $r < d_{\mathrm{F}}$, since the phase error becomes greater than $\pi/8$ rad. In this case, the third term in (2.6) must also be retained:

$$R \approx r - z' \cos\theta + \frac{1}{r}\left(\frac{z'^2}{2}\sin(\theta)^2\right). \tag{2.11}$$

The maximum approximation error then occurs when the first neglected term (*i.e.* the fourth term in the expansion) reaches its maximum value. By differentiating this term with respect to θ and setting the derivative to zero, one finds that the maximum error occurs at $z' = D/2$ and $\theta = \arctan(\pm\sqrt{2})$, as shown in [30, Sec. 4.4.2]. Thus,

$$\frac{k}{r^2}\frac{D^3}{16} \cdot \cos\Big(\arctan(\pm\sqrt{2})\Big)\sin\Big(\arctan(\pm\sqrt{2})\Big)^2 = \frac{\pi}{12\sqrt{3}} \cdot \frac{D^3}{\lambda r^2} \leq \frac{\pi}{8}, \tag{2.12}$$

which yields

$$r \geq 0.62\sqrt{\frac{D^3}{\lambda}} = d_{\mathrm{r}}. \tag{2.13}$$

This distance is known as the *Fresnel distance* and defines the boundary between the reactive and the radiative near field or Fresnel region [30, Sec. 4.4.2].

In the radiative near field, power variations cannot always be neglected. Comparing the Fresnel distance with the Björnson distance,

$$\frac{0.62\sqrt{D^3/\lambda}}{2D} = 0.31\sqrt{\frac{D}{\lambda}} \geq 1 \iff D \geq 10.41\lambda. \tag{2.14}$$

Therefore, power variations in the radiative near field can be ignored only when $D \geq 10.41\lambda$. In classical antenna theory, no distinction is made regarding whether amplitude (or power) variations across the antenna aperture should be accounted for or not. However, from a communication-oriented modeling and analysis perspective, this distinction becomes relevant. We will therefore examine it in more detail in section 2.2.

A final note regarding the region $r < d_{\mathrm{r}}$ is required. This region is usually referred to as the reactive near field, where reactive power components cannot be ignored. In this case, evanescent waves must be included in the analysis, and the wavefront is no longer spherical.

2.1.3 Arrays of electrically small antennas

Finally, instead of a receiver with a single electrically large antenna, consider an array of electrically small antennas, with maximum aperture D, centered at the origin. If the distance between transmitter and receiver is $r = \sqrt{x^2 + y^2 + z^2} \in [\max(d_{\mathrm{r}}, d_{\mathrm{UP}}), d_{\mathrm{F}})$, the electric field at the receiver can be approximated by

$$E(x, y, z) = \frac{E_0}{\sqrt{4\pi} r} \mathrm{e}^{-\mathrm{j}kR}, \quad R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}, \tag{2.15}$$

with (x', y', z') a coordinate in the array surface.

In other words, power variations are negligible over the array but phase differences remain significant. Although this expression is equivalent to that for an electrically large antenna in the Fresnel region (see (2.4) for $r \geq \max(d_{\mathrm{r}}, d_{\mathrm{UP}})$), the key difference is that, in this case, every antenna in the array is electrically small, hence they are located in the far field of the transmitter [32], [34].

The previous feature allows for array processing analysis—such as beamforming—similar to that of a plane wave impinging the array with an angle $\theta \neq \pi/2$ rad, where the power on each antenna is equal but phases differ.[2] With proper hardware processing, signals from each antenna can be combined to suppress the phase differences and increase the array gain. This is not possible when employing a single large antenna [8], [32].

Remark on power variations

It is important to note that, in certain scenarios, power variations must be taken into account. In particular, when analyzing the asymptotic behavior of large arrays (*i.e.* when $D \to \infty$), neglecting power variations can lead to results that are not physically meaningful. For instance, this consideration must be explicitly considered in the analysis in section 3.3.

2.2 Near-field channel modeling

Based on the previous observations, near-field channel models must account for wavefront curvature and, if the reactive near-field region is considered, also include inductive wave components. In this thesis, however, we are mainly interested in the radiative field region. Therefore, it is always assumed that the operating distance is greater than d_{r}, and inductive waves are neglected, with the only exception being the intra-array coupling phenomenon.

Even in the radiative near field, in some scenarios power variations across the array cannot be omitted. For this reason, the models described in the sequel take into account both situations:

1. Non-uniform spherical wave (NUSW).
2. Uniform spherical wave (USW).

In addition, the classical uniform plane wave (UPW) approximation is also presented. These electromagnetic field regions are illustrated in figure 2.2.

[2] It should be noted that phase differences in the radiative near field scenario are not linear.

It should be remarked that the model in figure 2.2 is a conservative simplification. In practice, not only the distance but also the angle have an impact on the importance of power and phase variations (see, for instance, [7, Fig. 9]). Moreover, as will be shown shortly, the plane-wave approximation can be derived from the first-order Taylor expansion of the spherical wavefront phase. Including the second-order term in the Taylor expansion yields the parabolic approximation, which preserves the curvature of the wavefront while offering a better mathematical tractability. For this reason, this approximation is also introduced in the sequel.

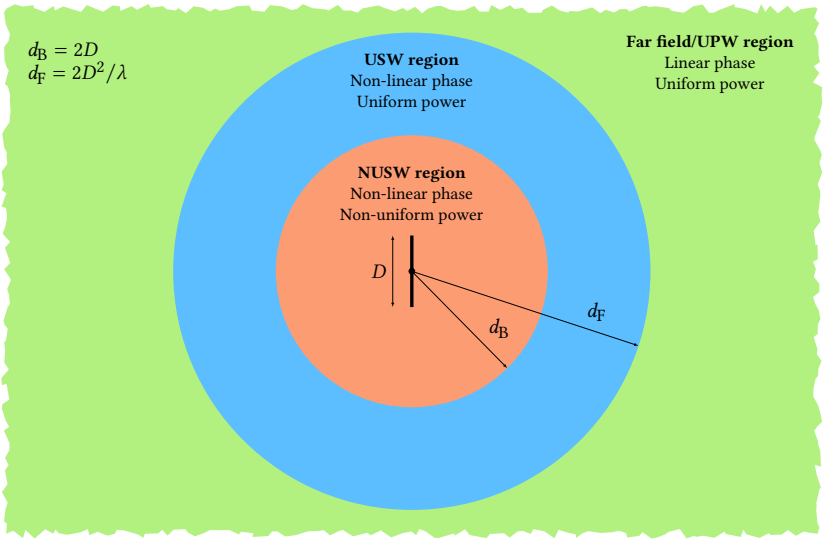


Figure 2.2: Illustration of the radiative electromagnetic field regions.

2.2.1 Near-field array response vector

An array response vector mathematically characterizes the phase and amplitude response of an antenna array to an incoming signal. To determine the array response in the near field, we consider an isotropic signal source located at $\mathbf{s}$, and an N-element antenna array of arbitrary geometry. For simplicity, the coordinate origin is placed at one of the array elements, and each array element is positioned at $\mathbf{u}_n$, for $n = 0, 1, \ldots, N-1$, as illustrated in figure 2.3. The signal source may represent either an active transmitter or a passive scatterer within the environment.

The array response for the system just described has the structure

$$\mathbf{a}(r, \theta, \phi) = \left(g_0 \mathrm{e}^{-\mathrm{j}k(r_0 - r)}, \ldots, g_{N-1} \mathrm{e}^{-\mathrm{j}k(r_{N-1} - r)}\right)^{\mathrm{T}}, \tag{2.16}$$

where g_n, for $0 \leq n \leq N-1$, depends on r_n and accounts for the individual characteristics of each element [7]. Here, $r = \|\mathbf{s}\|$ is the distance between the reference element (*i.e.* the origin) and the signal source, and

$$r_n = \|\mathbf{s} - \mathbf{u}_n\| = \sqrt{r^2 - 2\mathbf{s}^{\mathrm{T}}\mathbf{u}_n + \|\mathbf{u}_n\|^2} = r\sqrt{1 - \frac{2\mathbf{k}(\theta, \phi)^{\mathrm{T}}\mathbf{u}_n}{r} + \frac{\|\mathbf{u}_n\|^2}{r^2}}, \tag{2.17}$$

is the distance between the n-th element and the source, with

$$\mathbf{k}(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^{\mathrm{T}} \tag{2.18}$$

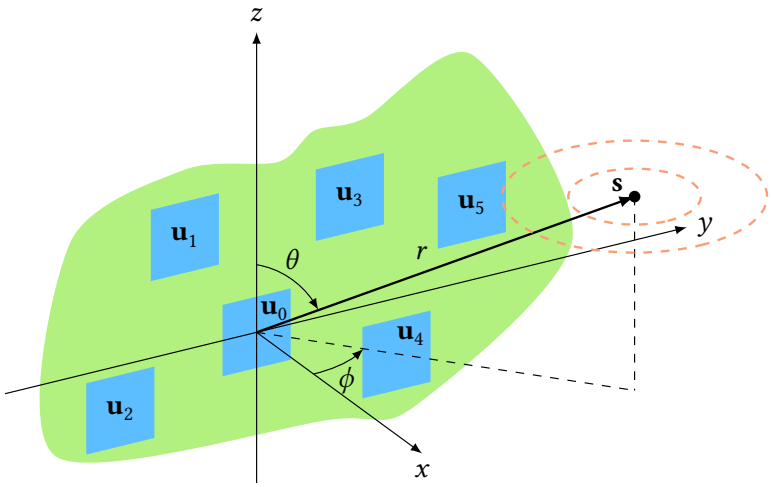


Figure 2.3: An isotropic signal source **s** and an antenna array of arbitrary architecture, with one of its elements at the coordinate origin.

denoting the propagation direction vector from the signal source to the array [35]. Substituting (2.17) in (2.16) yields the NUSW array response vector. The USW model is obtained as a special case by setting $g_n = 1$ in (2.16).

It is worth noting that the array response described here represents the spatial response relative to a reference point, namely the origin. To this end, the common phase term $\exp(-\mathrm{j}kr)$ is factored out of the response vector. Similarly, under the assumption of uniform attenuation, where $g_n = 1/r$ for all n, this scaling factor is also extracted from the vector. These shared terms are subsequently absorbed into the global channel gain, as detailed in sections 2.2.2 and 2.2.3.

Parabolic wavefront and far-field approximation

In the radiative near field, the first-order Maclaurin polynomial $\sqrt{1+x} \approx 1 + \frac{1}{2}x$ is often employed to approximate (2.17) resulting in

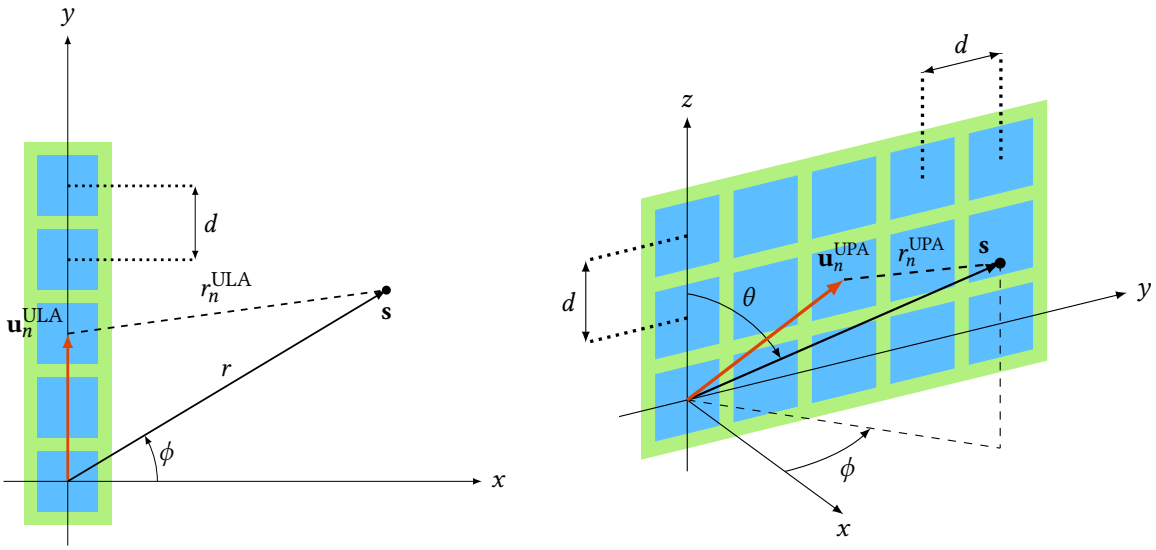
$$\Delta r_n = r_n - r \approx -\mathbf{k}(\theta,\phi)^{\mathrm{T}}\mathbf{u}_n + \frac{\|\mathbf{u}_n\|^2}{2r}, \tag{2.19}$$

which substituted in (2.16) corresponds to the so-called parabolic wavefront or Fresnel approximation of the spherical wave. This approximation is applicable both in the NUSW and USW models.

The classical far-field expression is obtained in a similar manner by only considering the first term of the Taylor expansion

$$\Delta r_n \approx -\mathbf{k}(\theta,\phi)^{\mathrm{T}}\mathbf{u}_n. \tag{2.20}$$

Unlike the near-field models presented in (2.17) and (2.19), equation (2.20) reveals that, in the far field, the phase variations across the array elements do not depend on r.



(a) Geometry of ULA response vector.

(b) Geometry of UPA response vector.

Figure 2.4: Geometry of ULA and UPA response vectors, showing element positions, spacing, angles, and distances to an isotropic signal source.

ULA response

Up to this point, we have considered an arbitrary array geometry to derive the array response vector. We now particularize this result for two important cases: the uniform linear array (ULA) and the uniform planar array (UPA).

We begin with a ULA consisting of N antennas with uniform inter-element spacing d. In this case, we can assume the coordinate system shown in figure 2.4a, where all array elements lie on the xy-plane, and thus we can set $\theta = \pi/2$ rad. By placing the coordinate origin at the first element of the ULA, the position of the n-th antenna is given by $\mathbf{u}_n^{\text{ULA}} = (0, nd)^{\text{T}}$, for $0 \leq n \leq N-1$. Similarly, the coordinates of the signal source are $\mathbf{s} = r\mathbf{k}(\phi) = r\,(\cos\phi, \sin\phi)^{\text{T}}$.

In order to obtain the array response vector we just need to particularize r_n in (2.17) to the ULA geometry:

$$\Delta r_{n,a}^{\text{ULA}} = r_n^{\text{ULA}} - r = \sqrt{r^2 - 2rnd\sin\phi + (nd)^2} - r. \tag{2.21}$$

Under the Fresnel approximation, $\Delta r_{n,a}^{\text{ULA}}$ can be simplified to

$$\Delta r_{n,b}^{\text{ULA}} = -nd\sin\phi + \frac{(nd)^2}{2r}. \tag{2.22}$$

Omitting the last term results in the well-known plane-wave approximation

$$\Delta r_{n,c}^{\text{ULA}} = -nd\sin\phi. \tag{2.23}$$

Finally, substituting the previous distances into (2.16) yields the ULA response vector for the

spherical, parabolic and planar wavefront assumptions:

$$[\mathbf{a}_{\mathrm{ULA}}(r,\phi)]_{n+1} = g_{n,a}^{\mathrm{ULA}} \exp\Big(\mathrm{j}k\Big(r - \sqrt{r^2 - 2rnd\sin\phi + (nd)^2}\Big)\Big), \tag{2.24}$$

$$[\mathbf{b}_{\mathrm{ULA}}(r,\phi)]_{n+1} = g_{n,b}^{\mathrm{ULA}} \exp\left(\mathrm{j}k\left(nd\sin\phi - \frac{(nd)^2}{2r}\right)\right), \tag{2.25}$$

$$[\mathbf{c}_{\mathrm{ULA}}(r,\phi)]_{n+1} = g_{n,c}^{\mathrm{ULA}} \exp\big(\mathrm{j}knd\sin\phi\big). \tag{2.26}$$

As previously discussed, the uniform power response is obtained by setting $g_{n,a}^{\mathrm{ULA}} = g_{n,b}^{\mathrm{ULA}} = g_{n,c}^{\mathrm{ULA}} = 1$, yielding

$$[\bar{\mathbf{a}}_{\mathrm{ULA}}(r,\phi)]_{n+1} = \exp\Big(\mathrm{j}k\Big(r - \sqrt{r^2 - 2rnd\sin\phi + (nd)^2}\Big)\Big), \tag{2.27}$$

$$\big[\bar{\mathbf{b}}_{\mathrm{ULA}}(r,\phi)\big]_{n+1} = \exp\left(\mathrm{j}k\left(nd\sin\phi - \frac{(nd)^2}{2r}\right)\right), \tag{2.28}$$

$$[\bar{\mathbf{c}}_{\mathrm{ULA}}(\phi)]_{n+1} = \exp\big(\mathrm{j}knd\sin\phi\big). \tag{2.29}$$

Observe that (2.29) corresponds to the classical steering vector commonly used in array processing [36, Ch. 8].

Throughout this thesis, the following notation is adopted: $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ denote the spherical, parabolic, and planar wavefront array responses, respectively. The corresponding overbar notation represents the respective wavefront model under the uniform power assumption.

UPA response

The analysis for a UPA with $N = N_{\mathrm{H}} \times N_{\mathrm{V}}$ antennas with a separation d is analogous to that of a ULA. Assuming that the array is deployed in the yz-plane, the n-th antenna is located at $\mathbf{u}_n^{\mathrm{UPA}} = (0, i_n d, j_n d)^{\mathrm{T}}$, where $i_n = \mathrm{mod}(n-1, N_{\mathrm{H}})$ and $j_n = \left\lfloor \frac{n-1}{N_{\mathrm{H}}} \right\rfloor$, for $0 \le n \le N-1$, are the horizontal and vertical indices of element n, as depicted in figure 2.4b. The distance between the signal source and the n-th array element is

$$\Delta r_{n,a}^{\mathrm{UPA}} = r_n^{\mathrm{UPA}} - r = \sqrt{r^2 - 2rd(i_n\sin\theta\sin\phi + j_n\cos\theta) + d^2(i_n^2 + j_n^2)} - r, \tag{2.30}$$

which, in the Fresnel region, can be approximated as

$$\Delta r_{n,b}^{\mathrm{UPA}} = -d(i_n\sin\theta\sin\phi + j_n\cos\theta) + d^2\frac{i_n^2 + j_n^2}{2r}, \tag{2.31}$$

and, in the far field, as

$$\Delta r_{n,c}^{\mathrm{UPA}} = -d(i_n\sin\theta\sin\phi + j_n\cos\theta). \tag{2.32}$$

Therefore, the non-uniform power response vectors are given by:

$$[\mathbf{a}_{\mathrm{UPA}}(r,\theta,\phi)]_{n+1} = g_{n,a}^{\mathrm{UPA}} \exp\Big(\mathrm{j}k\Big(r - \sqrt{r^2 - 2rd(i_n\sin\theta\sin\phi + j_n\cos\theta) + d^2(i_n^2 + j_n^2)}\Big)\Big), \tag{2.33}$$

$$[\mathbf{b}_{\mathrm{UPA}}(r,\theta,\phi)]_{n+1} = g_{n,b}^{\mathrm{UPA}} \exp\left(\mathrm{j}k\left(d(i_n\sin\theta\sin\phi - j_n\cos\theta) + d^2\frac{i_n^2 + j_n^2}{2r}\right)\right), \tag{2.34}$$

$$[\mathbf{c}_{\mathrm{UPA}}(r,\theta,\phi)]_{n+1} = g_{n,c}^{\mathrm{UPA}} \exp\big(\mathrm{j}kd(i_n\sin\theta\sin\phi + j_n\cos\theta)\big). \tag{2.35}$$

Like in the ULA case, under the uniform power assumption $g_{n,a}^{\mathrm{UPA}} = g_{n,b}^{\mathrm{UPA}} = g_{n,c}^{\mathrm{UPA}} = 1$, yielding

$$[\overline{\mathbf{a}}_{\mathrm{UPA}}(r,\theta,\phi)]_{n+1} = \exp\Bigg(\mathrm{j}k\Bigg(r - \sqrt{r^2 - 2rd(i_n \sin\theta \sin\phi + j_n \cos\theta) + d^2(i_n^2 + j_n^2)}\Bigg)\Bigg), \tag{2.36}$$

$$\left[\overline{\mathbf{b}}_{\mathrm{UPA}}(r,\theta,\phi)\right]_{n+1} = \exp\Bigg(\mathrm{j}k\Bigg(d(i_n \sin\theta \sin\phi - j_n \cos\theta) + d^2\frac{i_n^2 + j_n^2}{2r}\Bigg)\Bigg), \tag{2.37}$$

$$[\overline{\mathbf{c}}_{\mathrm{UPA}}(\theta,\phi)]_{n+1} = \exp\left(\mathrm{j}kd(i_n \sin\theta \sin\phi + j_n \cos\theta)\right). \tag{2.38}$$

It is also worth noting that, under this assumption, the array responses in the Fresnel and far-field regions become separable along the two axes. Specifically, the exponential term can be factorized into two components, one depending on i_n and the other on j_n. This property allows the response vector of a UPA to be expressed as the Kronecker product of two ULA response vectors [28], one along the y-axis and the other along the z-axis:

$$\overline{\mathbf{b}}_{\mathrm{UPA}}(r,\theta,\phi) = \overline{\mathbf{b}}_{\mathrm{ULA},y}(r,\theta,\phi) \otimes \overline{\mathbf{b}}_{\mathrm{ULA},z}(r,\theta), \tag{2.39}$$

$$\overline{\mathbf{c}}_{\mathrm{UPA}}(\theta,\phi) = \overline{\mathbf{c}}_{\mathrm{ULA},y}(\theta,\phi) \otimes \overline{\mathbf{c}}_{\mathrm{ULA},z}(\theta). \tag{2.40}$$

Note that this does not hold under a general spherical wavefront.

2.2.2 Near-field channel models for SIMO and MISO systems

In section 2.2.1, we derived the general array response vector in the near field, along with its particular forms for ULA and UPA geometries. Building on these results, we can now directly obtain the near-field channel expressions for both line-of-sight (LoS) and non-line-of-sight (NLoS) propagation scenarios.

Under the block-fading channel assumption (see section 2.3), a single LoS path channel can be modeled as

$$\mathbf{h}^{\mathrm{LoS}} = \alpha \mathbf{a}(r,\theta,\phi), \tag{2.41}$$

where α is the (complex) channel gain at the reference [7], and $\mathbf{a}(r,\theta,\phi)$ is the array response vector, which can be any of those described in section 2.2.1, depending on the wavefront model considered [7], [35].

On the other hand, when multiple transmission paths with no LoS component are present, the attenuation factors $\{\alpha_l\}_{l=1}^{L}$ are assumed to be mutually uncorrelated random variables, and the multipath channel can be expressed as

$$\mathbf{h}^{\mathrm{NLoS}} = \sum_{l=1}^{L} \alpha_l \mathbf{a}(r_l,\theta_l,\phi_l), \tag{2.42}$$

where (r_l,θ_l,ϕ_l) denotes the location of the l-th scatterer [7], [35]. Under rich scattering, the channel vector follows a spatially *correlated Rayleigh fading model* [37], [38],

$$\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_{\mathbf{h}}), \tag{2.43}$$

where the spatial correlation matrix is given by

$$\mathbf{C}_{\mathbf{h}} = \mathrm{E}\left[\mathbf{h}\mathbf{h}^{\mathrm{H}}\right] = \sum_{l=1}^{L} \beta_l \mathbf{a}(r_l,\theta_l,\phi_l)\mathbf{a}(r_l,\theta_l,\phi_l)^{\mathrm{H}}, \tag{2.44}$$

with $\beta_l = \mathrm{E}\left[|\alpha_l|^2\right]$ the power of the l-th path. Even with a moderate number of scatterers, the central limit theorem (CLT) ensures that the channel coefficients can still be modeled as Rayleigh distributed, which makes the Rayleigh fading model applicable also to sparse channels such as those in the mmWave bands.

In denser scenarios, it is possible to assume a continuum of scatterers, in which case the covariance matrix can be expressed as

$$\mathbf{C_h} \to \beta \int_{r_1}^{r_2} \int_{\theta_1}^{\theta_2} \int_{\phi_1}^{\phi_2} f(r,\theta,\phi)\mathbf{a}(r,\theta,\phi)\mathbf{a}(r,\theta,\phi)^{\mathrm{H}} \,\mathrm{d}\phi \,\mathrm{d}\theta \,\mathrm{d}r. \tag{2.45}$$

In (2.45), β denotes the average power at the reference array element [37], [39], $\theta \in (\theta_1, \theta_2)$, $\phi \in (\phi_1, \phi_2)$ and $r \in (r_1, r_2)$ define the spatial distribution range of the channel, and $f(r,\theta,\phi)$ is the normalized spatial scattering function satisfying $\int_{r_1}^{r_2} \int_{\theta_1}^{\theta_2} \int_{\phi_1}^{\phi_2} f(r,\theta,\phi) \,\mathrm{d}\phi \,\mathrm{d}\theta \,\mathrm{d}r = 1$. Under the uniform power hypothesis it is satisfied that $\beta = \mathrm{tr}(\mathbf{C_h})/N$.

Uncorrelated Rayleigh fading discussion

The well-known uncorrelated Rayleigh fading model arises when $\mathbf{C_h} = \beta \mathbf{I}_N$. It should be clear by now that the applicability of this model is very limited, even with rich scattering in the far field [40], [41]. We now examine this limitation more closely.

In the far field, scatterers can only be distinguished by their direction, not their distance, so (2.45) simplifies to

$$\mathbf{C_h} = \beta \int_{\theta_1}^{\theta_2} \int_{\phi_1}^{\phi_2} f(\theta,\phi)\overline{\mathbf{c}}(\theta,\phi)\overline{\mathbf{c}}(\theta,\phi)^{\mathrm{H}} \,\mathrm{d}\phi \,\mathrm{d}\theta, \tag{2.46}$$

which corresponds to uncorrelated Rayleigh fading only if the double integral equals the identity matrix. To show that this does not hold even for very simple scenarios, we assume a ULA along the z axis with isotropic scattering in the halfspace in front of it, so

$$[\mathbf{C_h}]_{m,n} = \beta \int_{-\pi/2}^{\pi/2} \mathrm{d}\phi \int_0^{\pi} \frac{\sin\theta}{2\pi} \exp\left(\mathrm{j}\frac{2\pi}{\lambda} d(m-n)\cos\theta\right) \mathrm{d}\theta = \beta \mathrm{sinc}\left(2\frac{d}{\lambda}(m-n)\right). \tag{2.47}$$

Fading is uncorrelated when $[\mathbf{C_h}]_{m,n} = 0$ for $m \neq n$, but this only occurs when $d = l \cdot \frac{\lambda}{2}$, $l \in \mathbb{Z}$.

In conclusion, even in the far field with a ULA, the channel is spatially uncorrelated only when antennas are separated by a multiple of $\lambda/2$ in the presence of rich scattering.[3] Indeed, this is the only situation where spatial uncorrelatedness arises [17], [37].

Furthermore, it should be noted that the considerations in this chapter only account for the array spatial correlation, but not for the mutual coupling between antennas which, in practice, also induces additional channel correlation, as we will see in chapter 3.

[3] Since (2.47) decays with d, antennas that are sufficiently far apart can be treated as uncorrelated even if their separation is not a multiple of $\lambda/2$.

Exponential correlation model

Although the spatial correlation models (2.44) and (2.45) are highly realistic, their large number of parameters can make them difficult to use in certain scenarios. Moreover, since uncorrelated Rayleigh models are not physically consistent in general, intermediate models which account for spatial correlation while relying on fewer parameters have been proposed.

A notable example is the *exponential correlation model* for ULAs [42], [43],[4] which captures the higher correlation between neighboring antennas compared to distant ones, while depending on only a single parameter. The corresponding channel covariance matrix is Hermitian and Toeplitz, with each entry given by

$$[\mathbf{C_h}]_{m,n} = c_{m,n} = \begin{cases} \rho^{n-m}, & m \leq n, \\ c_{n,m}^*, & m > n, \end{cases} \tag{2.48}$$

where $\rho \in \mathbb{C}$ is the correlation coefficient, satisfying $|\rho| < 1$.

In this model, correlation decays exponentially with antenna separation. This decay is significantly faster than the hyperbolic behavior observed in (2.47). Nevertheless, by appropriately tuning the correlation coefficient, the exponential model can serve as a coarse yet qualitatively similar approximation of more physically accurate models. This is particularly useful when assessing performance or algorithms in which the channel is not the main focus, but where incorporating correlation remains important to reflect realistic conditions.

2.2.3 Near-field channel models for MIMO systems

Having examined the single-input multiple-output (SIMO) and multiple-input single-output (MISO) channel models in the previous section, we now extend the analysis to MIMO systems. Contrary to the far field, where the LoS MIMO channel matrix can be expressed using the response vectors of the transmitter and receiver arrays, in the near field it must be determined element by element. To do so, consider a communication system with two arrays of N_R and N_T antennas, respectively. The antennas in the first array are located at positions $\mathbf{u}_n$, for $0 \leq n \leq N_\mathrm{R} - 1$, while those in the second array are located at $\mathbf{s}_m$, for $0 \leq m \leq N_\mathrm{T} - 1$, as depicted in figure 2.5.

Following an approach analogous to that of the previous section, the channel matrix can be expressed as

$$[\mathbf{H}^{\mathrm{LoS}}]_{m+1,n+1} = \alpha g_{m,n} \mathrm{e}^{-\mathrm{j}k(r_{m,n}-r)}, \tag{2.49}$$

where α denotes the channel gain at the array's reference element and $g_{m,n}$ is an element-specific factor that depends on $r_{m,n}$ [7], [28], with $r_{m,n}$ given by

$$\begin{aligned} r_{m,n} &= \|\mathbf{s}_m - \mathbf{u}_n\| = \|\mathbf{s}_0 - \mathbf{u}_n + (\mathbf{s}_m - \mathbf{s}_0)\| \\ &= \sqrt{r^2 + 2r\mathbf{k}(\theta,\phi)^{\mathrm{T}}(\mathbf{s}_m - \mathbf{s}_0 - \mathbf{u}_n) - 2(\mathbf{s}_m - \mathbf{s}_0)^{\mathrm{T}}\mathbf{u}_n + \|\mathbf{u}_n\|^2 + \|\mathbf{s}_m - \mathbf{s}_0\|^2}, \end{aligned} \tag{2.50}$$

with $r = \|\mathbf{s}_0 - \mathbf{u}_0\|$. Note that only the first two terms inside the square root depend on r, since $\mathbf{k}(\theta,\phi)$ denotes the propagation direction vector from the origin to the transmitting

[4] An extension of the exponential correlation model to UPA configurations can be obtained via a Kronecker approximation [44], [45].

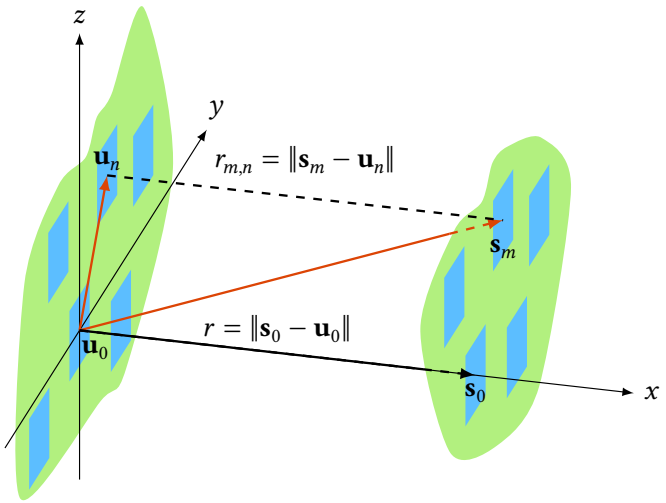


Figure 2.5: Illustration of a LoS MIMO communication system involving two arrays with arbitrary geometry.

---

array, $\mathbf{u}_n$ represents the position of the n-th receiving antenna, and $\mathbf{s}_m - \mathbf{s}_0$ is the vector from the reference element in the transmitting array to the m-th transmitting antenna.

Applying the Fresnel approximation to (2.50) yields

$$\Delta r_{m,n} = r_{m,n} - r \approx \mathbf{k}(\theta,\phi)^{\mathrm{T}}(\mathbf{s}_m - \mathbf{s}_0 - \mathbf{u}_n) - \frac{(\mathbf{s}_m - \mathbf{s}_0)^{\mathrm{T}}\mathbf{u}_n}{r} + \frac{\|\mathbf{u}_n\|^2 + \|\mathbf{s}_m - \mathbf{s}_0\|^2}{2r}. \tag{2.51}$$

The plane-wave approximation is particularly interesting, as it simplifies the MIMO channel representation to the outer product of the transmit and receive array response vectors. Dropping the terms that decrease with r, the far-field distance becomes

$$\Delta r_{m,n} \approx \mathbf{k}(\theta,\phi)^{\mathrm{T}}(\mathbf{s}_m - \mathbf{s}_0 - \mathbf{u}_n) \tag{2.52}$$

and

$$[\mathbf{H}_{\mathrm{FF}}^{\mathrm{LoS}}]_{m+1,n+1} = \alpha \exp\big(-\mathrm{j}k\mathbf{k}(\theta,\phi)^{\mathrm{T}}(\mathbf{s}_m - \mathbf{s}_0)\big)\exp\big(\mathrm{j}k\mathbf{k}(\theta,\phi)^{\mathrm{T}}\mathbf{u}_n\big). \tag{2.53}$$

The last two terms in (2.53) correspond to the transmitting and receiving array response vectors. Hence, the far-field MIMO channel can be written as

$$\mathbf{H}_{\mathrm{FF}}^{\mathrm{LoS}} = \alpha\bar{\mathbf{c}}_{\mathrm{R}}(r,\theta,\phi)\bar{\mathbf{c}}_{\mathrm{T}}(r,\theta,\phi)^{\mathrm{H}}, \tag{2.54}$$

with

$$[\bar{\mathbf{c}}_{\mathrm{R}}(r,\theta,\phi)]_{n+1} = \exp\big(\mathrm{j}k\mathbf{k}(\theta,\phi)^{\mathrm{T}}\mathbf{u}_n\big) \tag{2.55}$$

and

$$[\bar{\mathbf{c}}_{\mathrm{T}}(r,\theta,\phi)]_{m+1} = \exp\big(\mathrm{j}k\mathbf{k}(\theta,\phi)^{\mathrm{T}}(\mathbf{s}_m - \mathbf{s}_0)\big) \tag{2.56}$$

the array response vectors under the UPW approximation.

Based on (2.54), it has been proposed to model the near-field channel as the outer product of the two array response vectors [7], [35]. In this approach, the transmit array is treated as a single point to obtain the near-field receive array response $\mathbf{a}_{\mathrm{R}}(r,\theta,\phi)$, whose n-th element is given by

$$[\mathbf{a}_{\mathrm{R}}(r,\theta,\phi)]_{n+1} = g_{\mathrm{R},n}\mathrm{e}^{-\mathrm{j}k\|\mathbf{u}_n - \mathbf{s}_0\|}. \tag{2.57}$$

Similarly, the transmit array response $\mathbf{a}_{\mathrm{T}}(r,\theta,\phi)$ is determined by treating the receive array as a single point, resulting in

$$[\mathbf{a}_{\mathrm{T}}(r,\theta,\phi)]_{m+1} = g_{\mathrm{T},m}\mathrm{e}^{-\mathrm{j}k\|\mathbf{s}_m-\mathbf{u}_0\|} = g_{\mathrm{T},m}\mathrm{e}^{-\mathrm{j}k\|\mathbf{s}_m\|}. \tag{2.58}$$

Then, the near-field MIMO channel is given by

$$\mathbf{H}^{\mathrm{LoS}} = \alpha\mathbf{a}_{\mathrm{R}}(r,\theta,\phi)\mathbf{a}_{\mathrm{T}}(r,\theta,\phi)^{\mathrm{H}}. \tag{2.59}$$

The main difference between these two models is that (2.49) rank can be greater than one, while (2.59) is always rank one. The latter, however, presents better mathematical tractability thanks to its outer product structure.

This outer product formulation is specially appropriate for multipath modeling. By representing each scatterer as a point source, the overall channel can be expressed as a sum of terms analogous to (2.59):

$$\mathbf{H}^{\mathrm{NLoS}} = \sum_{l=1}^{L}\alpha_l\mathbf{a}_{\mathrm{R}}(r_l,\theta_l,\phi_l)\mathbf{a}_{\mathrm{T}}(r_l,\theta_l,\phi_l)^{\mathrm{H}}, \tag{2.60}$$

with α_l denoting the attenuation of the l-th path and (r_l,θ_l,ϕ_l) the spherical coordinates of the l-th scatterer. Of course, when $L\to\infty$, the model in (2.60) converges to the Rayleigh fading model

$$\mathrm{vec}(\mathbf{H}) \sim \mathscr{CN}(\mathbf{0}_{N_{\mathrm{T}}N_{\mathrm{R}}}, \mathbf{C}_{\mathbf{H}}), \quad \mathbf{C}_{\mathbf{H}} = \mathrm{E}\Big[\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^{\mathrm{H}}\Big]. \tag{2.61}$$

Among the several models presented in this section, (2.43) and (2.61) are the most suitable for dynamic scenarios, such as mobile communications, thanks to their inherently stochastic behavior.

2.3 Mobile communications

In mobile communications, the most common scenario involves a fixed BS equipped with an N-antenna array communicating with a single-antenna user equipment (UE). The wireless channel established between both ends is primarily affected by [46, Ch. 2]:

- **Path loss:** reduction in power density of an electromagnetic wave as it propagates through space. Path loss models are typically deterministic and depend on the transmitter-receiver distance and the system frequency.
- **Shadowing:** variation in received signal power caused by large obstacles (like buildings, hills, or trees) blocking or attenuating the signal. It occurs over medium to large distances compared to the wavelength, leading to slow changes in signal strength. Shadowing is usually modeled as a log-normal distribution.
- **Multipath:** rapid fluctuation of the received signal's amplitude, phase, or power over short time intervals or distances. It is caused by the superposition of multiple signal paths due to reflection, diffraction, and scattering. These paths can interfere constructively when phase shifts are aligned, or destructively when phase shifts are misaligned.

Variations in received power caused by path loss and shadowing occur over relatively large distances and are therefore referred to as large-scale propagation effects. In contrast, power fluctuations resulting from the constructive and destructive interference of multipath components occur over distances on the order of the wavelength. These variations are typically referred to as small-scale propagation effects. The effect of all three phenomena is illustrated in figure 2.6.

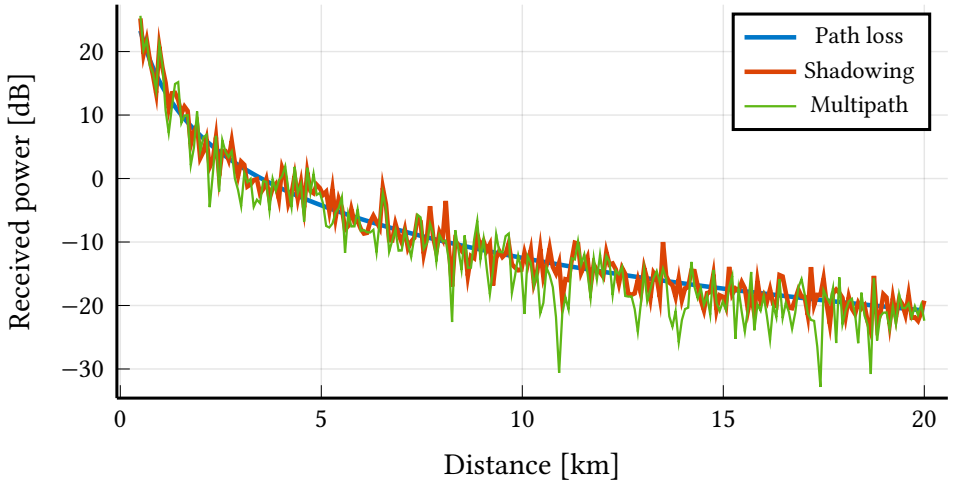


Figure 2.6: Received signal power versus distance, illustrating large-scale effects (path loss and shadowing) and small-scale effects (multipath propagation).

Of the models presented in the previous section, the one that best fits the aforementioned scenario is the correlated Rayleigh model, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{h})$, given in (2.43). In this model, the Gaussian distribution captures the small-scale fading effects, while the spatial correlation matrix describes the macroscopic propagation characteristics, including antenna gains and radiation patterns [38]. The average received power at the array reference element is often modeled in decibels as

$$\beta = \mathrm{Y} - 10\gamma \log_{10}\left(\frac{d}{1\,\mathrm{km}}\right) + \chi, \tag{2.62}$$

where d is the distance in kilometers between the transmitter and the receiver, Y and γ are determined according to an established propagation model, such as the Hata–COST231 model for urban areas [29, Sec. 3.3], and $\chi \sim \mathcal{N}\left(0, \sigma_\chi^2\right)$ models the shadow fading [38]. Note that the β in (2.62) is the same as in (2.45).

2.3.1 Block-fading model

Having described several channel models as well as the main mobile channel phenomena a natural question arises: for how long does a given channel observation remain valid?

The main limiting factor is the multipath phenomenon, which varies over distances on the order of the wavelength and therefore changes much more rapidly than shadowing or path loss. A common simplification for the small-scale fading is the block-fading model, in which the channel is assumed to remain constant over a short time interval known as the coherence time, and change to an independent realization in the next coherence block [38], [29, Sec. 3.4].

Coherence time is related to the maximum Doppler shift and is usually assumed to be

$$T_{\mathrm{c}} = \frac{\lambda}{4v}, \tag{2.63}$$

where v is the velocity of the UE. If the transmitted signal duration is much shorter than the coherence time, the channel remains essentially constant over the signal, a condition referred to as slow fading. On the other hand, if the signal duration exceeds the coherence time, the channel varies during the transmission, leading to fast fading. Expression (2.63) also highlights that high-speed mobility and high-frequency communications exhibit shorter coherence times, thereby requiring the channel to be estimated more often [38]. For example, the coherence time for a UE operating at $f = 3$ GHz in a high-speed train moving at 300 km/h is $T_{\mathrm{c}} = 5\,\mu$s.

In addition to their finite duration in the time domain, coherence blocks also span a limited range in frequency, known as the coherence bandwidth. This quantity is determined by the phase differences introduced by multipath propagation and can be approximated as

$$B_{\mathrm{c}} = \frac{1}{2T_{\mathrm{d}}}, \tag{2.64}$$

where T_{d} is the delay spread, defined as the time difference between the earliest and latest arriving multipath components [46, Sec. 2.4]. If the signal bandwidth is much smaller than B_{c}, the channel response is essentially constant across the band, a condition known as frequency-flat fading. Conversely, when the signal bandwidth exceeds B_{c}, different frequency components experience different channel gains and phases, leading to frequency-selective fading.

2.4 Coherent and noncoherent communications

In modern communication systems, a common approach to handle channels with unknown coefficients is to first estimate them using pilot sequences during a training phase. The resulting information, known as CSI, is then used for data transmission in the subsequent communication phase. When the receiver is assumed to have perfect instantaneous CSI, the scheme is referred to as coherent communication. In contrast, noncoherent communication refers to schemes in which instantaneous CSI is not available.

For channels with infinite coherence time, a training phase followed by data transmission has been shown to be optimal [47]. However, when the coherence time T_{c} is finite, the resources that can be allocated to training are limited. In particular, if T_{c} is small, the channel quickly changes into an independent realization, and the validity of the estimated channel rapidly deteriorates [48].

This effect is further noticeable in multi-antenna systems. For instance, in a channel with an N_{T}-antenna transmitter, at least N_{T} orthogonal pilot symbols must be transmitted to the receiver in order to estimate the channel [38, Sec. 1.3.5]. Consequently, for a coherence block of T samples, the fraction of resources dedicated to training is N_{T}/T, leaving only a portion of $1 - N_{\mathrm{T}}/T$ of the coherence block for the communication phase [10]. The increasing training requirements with the number of antennas can lead to significant overhead in mMIMO systems [49], [50]. For this reason, noncoherent communication schemes are regaining

attention, specially since large antenna arrays are expected to be one of the key technologies for 6G [3], [51].

An attentive reader may note that this bottleneck can, in some cases, be mitigated by exploiting channel reciprocity. A typical configuration in cellular multiuser systems involves a BS equipped with a very large number of antennas, N, and UEs with only a few antennas each. For simplicity, consider the case of K single-antenna UEs. In this setting, CSI can be acquired through uplink training and then exploited in the downlink by leveraging channel reciprocity [52]. As a result, the training length scales with K rather than N.

Even with this reduced CSI acquisition overhead, the previous approach is not without its limitations. First, reusing uplink CSI for the downlink is only possible in TDD, as channel reciprocity does not hold in FDD [29, Sec. 5.10]. Second, the problem of *pilot contamination* arises [53]. When training-based schemes are employed in the uplink, the received signal at one of the BS antennas is given by the $\tau_{\mathrm{p}} \times 1$ vector

$$\mathbf{y} = \sum_{k=1}^{K} \mathsf{h}_k \mathbf{p}_k + \mathbf{z}, \tag{2.65}$$

where $\mathbf{z}$ is additive Gaussian noise, h_k is the channel between the k-th user and the BS, and $\mathbf{p}_k \in \mathbb{C}^{\tau_{\mathrm{p}} \times 1}$ with $\|\mathbf{p}_k\| = 1$ is the pilot sequence of the k-th user [54]. Ideally, to avoid pilot contamination, the condition $\mathbf{p}_i \perp \mathbf{p}_j$ for $i \neq j$ should be satisfied. In practice, however, this is not feasible as $K > \tau_{\mathrm{p}}$, so pilot sequences must be reused leading to pilot contamination. Although promising approaches have been proposed to mitigate this issue in coherent communications [55]–[57], noncoherent communications remain a viable alternative in certain scenarios [11], [58], [59]. Therefore, further research in this direction continues to be of significant importance.

2.4.1 Noncoherent communications approaches

The term noncoherent communications has been interpreted with slight variations across the literature. In this thesis, we adopt the definition that a system operates noncoherently when the channel realization is unknown at both the receiver and the transmitter. Under this assumption, the design of noncoherent communication schemes focuses on identifying components of the transmitted signal that remain invariant under fading, and can therefore be reliably detected at a receiver without CSI.

The three main approaches in multi-antenna systems are described next [14], and their potential applications are illustrated in figure 2.7.

Differential schemes

Differential detection schemes are particularly useful in scenarios where channel fading evolves gradually over time without abrupt variations. By comparing successive received signals, the noncoherent receiver can reliably detect transitions of the transmitted signal across channel uses [60]–[63].

To illustrate these schemes, we briefly introduce differential phase shift keying (DPSK) modulation in a conventional single-input single-output (SISO) system [64, Sec. 12.1]. Let

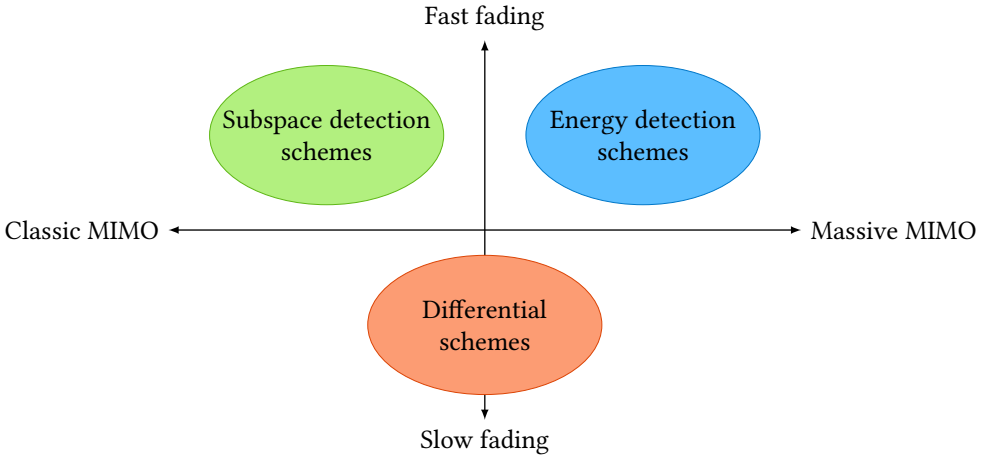


Figure 2.7: Illustration of noncoherent communications approaches with their potential use cases.

$\mathcal{S} = \{\exp\left(\mathrm{j}2\pi m/M\right) \mid m = 0, \ldots, M-1\}$ denote a M-ary phase shift keying (PSK) constellation and $s(n) \in \mathcal{S}$ represent the transmitted signal. The differential scheme starts by transmitting a reference symbol $x(0) = 1$, and it is then followed by a differential encoding process such that the signal transmitted at the n-th time slot is

$$x(n) = x(n-1)s(n). \tag{2.66}$$

Representing the received signals corresponding to the $(n-1)$-th and n-th symbols as

$$\mathsf{y}(n-1) = \mathsf{h}\mathsf{x}(n-1) + \mathsf{z}(n-1), \tag{2.67}$$
$$\mathsf{y}(n) = \mathsf{h}\mathsf{x}(n) + \mathsf{z}(n), \tag{2.68}$$

where $\mathsf{h} \sim \mathcal{CN}(0,1)$ is the channel coefficient and $\mathsf{z}(n)$ is additive noise. It follows that

$$\mathsf{y}(n) = \mathsf{h}\mathsf{x}(n-1)\mathsf{s}(n) + \mathsf{z}(n) \tag{2.69}$$
$$= \mathsf{y}(n-1)\mathsf{s}(n) + \mathsf{z}(n) - \mathsf{z}(n-1)\mathsf{s}(n) \tag{2.70}$$
$$= \mathsf{y}(n-1)\mathsf{s}(n) + \mathsf{z}'(n). \tag{2.71}$$

where $\mathsf{z} \sim \mathcal{CN}(0, P_{\mathsf{z}})$ and $\mathsf{z}' \sim \mathcal{CN}(0, 2P_{\mathsf{z}})$.

Comparing (2.68) and (2.71) it can be observed that they share the same structure. Consequently, the transmitted symbol $s(n)$ can be recovered in the same way as in coherent detection, but treating $\mathsf{y}(n-1)$ as the channel and $\mathsf{z}'(n)$ as the additive noise. It is worth noting that noise power is doubled, which results in a SNR loss of 3 dB.

Subspace detection schemes

Subspace detection methods exploit the property that, at high SNR, the wireless channel preserves the subspace spanned by the transmitted symbols. This is because the unknown channel matrix scales and rotates the basis of the transmitted signal matrix without altering its column space. Consequently, the information embedded in this subspace remains intact, motivating the design of constellations where each symbol corresponds to a different subspace. In mathematics, the set of all subspaces of a given dimension within a vector space is known

as the Grassmannian, which has the structure of a differentiable manifold (*i.e.* a smooth surface that locally resembles the Euclidean space) [65, Ch. 1].

To further illustrate subspace detection methods, we now proceed to describe the unitary space-time modulation (USTM) scheme [66]. Consider a Rayleigh block-fading channel with N_{T} transmitting antennas and N_{R} receiving antennas satisfying $N_{\mathrm{T}} \leq N_{\mathrm{R}}$. The channel matrix is $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{R}}}$, with each term distributed as $\mathcal{CN}(0,1)$, and it remains constant over a coherence block of duration $T \geq 2$ symbols. Letting $\mathbf{X} \in \mathbb{C}^{T \times N_{\mathrm{T}}}$ be the transmitted signal in a coherence block, the received signal can be represented by

$$\mathbf{Y} = \mathbf{X}\mathbf{H} + \sqrt{\frac{N_{\mathrm{T}}}{TP}}\mathbf{Z}, \tag{2.72}$$

where $\mathbf{Z}$ is additive white Gaussian noise (AWGN) and P denotes the SNR.

In coherent communications, the capacity-achieving constellation follows a Gaussian distribution [67, Sec. 6.5]. However, this is not the case in the noncoherent setting, where the channel realization is unknown [68]. Instead, when $T \geq N_{\mathrm{T}} + N_{\mathrm{R}}$ and $P \to \infty$, capacity is achieved by transmitting unitary space-time signals [47], that is,

$$\mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{I}_{N_{\mathrm{T}}}. \tag{2.73}$$

The asymptotical optimality of USTM arises from the fact that $\mathbf{X}$ and $\mathbf{XH}$ span the same subspace, *i.e.* the same element of the Grassmann manifold [47], [66]. Accordingly, Grassmannian constellations are formally defined as

$$\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_M \mid \mathbf{X}_i^{\mathrm{H}}\mathbf{X}_i = \mathbf{I}_{N_{\mathrm{T}}}\}. \tag{2.74}$$

Therefore, the design of a Grassmannian constellation can be formulated as a point-packing problem on the Grassmann manifold, where each codeword corresponds to a subspace [14], [47], [69].

Energy detection schemes

Most primitive communication systems were based on energy detection, but these schemes were soon overtaken by more sophisticated alternatives capable of exploiting phase information, which offered greater robustness against additive noise [9, Appendix D]. With the arrival of mMIMO, however, energy detection schemes are regaining interest. The reason is that the large number of antennas significantly reduces the impact of small-scale fading and the channel performs as deterministic, phenomenon known as *channel hardening* [70], [71].

Although channel hardening is typically exploited in coherent communications, it can also be leveraged in the noncoherent setting by relying on the second-order statistics of the channel [49]. To show this, consider a SIMO wireless system with N receiving antennas. The received signal in complex baseband form is

$$\mathbf{y} = \mathbf{h}\mathsf{x} + \mathbf{z}, \quad \mathbf{h} \sim \mathcal{CN}(\mathbf{0}_N, P_{\mathsf{h}}\mathbf{I}_N),\ \mathbf{z} \sim \mathcal{CN}(\mathbf{0}_N, P_{\mathsf{z}}\mathbf{I}_N), \tag{2.75}$$

and x denotes the transmitted symbol belonging to a unipolar constellation. Taking the normalized received energy gives

$$\frac{1}{N}\|\mathbf{y}\|^2 = \frac{1}{N}\|\mathbf{h}\|^2\mathsf{x}^2 + \frac{1}{N}\|\mathbf{z}\|^2 + \frac{2}{N}\Re\big(\mathbf{h}^{\mathrm{H}}\mathbf{z}\big)\mathsf{x}. \tag{2.76}$$

As $N \to \infty$, it follows from the law of large numbers that

$$\frac{1}{N}\|\mathbf{h}\|^2 \to P_{\mathsf{h}}, \tag{2.77}$$

$$\frac{1}{N}\|\mathbf{z}\|^2 \to P_{\mathsf{z}}, \tag{2.78}$$

$$\frac{2}{N}\Re\left(\mathbf{h}^{\mathrm{H}}\mathbf{z}\right) \to 0. \tag{2.79}$$

Consequently, $\|\mathbf{y}\|^2/N \to P_{\mathsf{h}}\mathsf{x}^2 + P_{\mathsf{z}}$ and, since P_{h} and P_{z} are assumed known, x can be detected reliably [49], [72].

Channel hardening can also be exploited in certain correlated channel models. Energy detection schemes under spatially correlated channels constitute the main focus of this thesis, and are studied in detail in chapter 4.

CHAPTER 3

Mutual coupling modeling in near-field communications

The increasing demand for high data rates in modern wireless communication systems has driven significant research into multi-antenna techniques. Conventional MIMO systems, and mMIMO in particular, offer substantial gains in both SE and link reliability [38, Sec.4.4]. Nevertheless, mMIMO alone cannot fully meet the performance requirements envisioned for next-generation networks, which include data rates on the order of Tbps and end-to-end latencies below 0.5 ms [73].

For this reason, new research directions are appearing, with the most prominent ones being large intelligent surfaces [74] and holographic MIMO [75], [76]. The latter technology refers to an array with a massive number of densely deployed antennas, which unavoidably results in mutual coupling [30, Sec. 8.7]. By contrast, the majority of the literature on classical and massive MIMO assumes half-wavelength element spacing, where coupling can be safely neglected [38].

In the previous chapter, the analysis focused primarily on the propagation channel, considering only the environment while neglecting the effects of the antennas themselves. As a result, the developed models are physically consistent for mMIMO and 5G systems, but they fall short when addressing the requirements of next generation communications and holographic MIMO.

In this chapter, we extend the framework to explicitly account for mutual coupling among antenna elements. It is important to highlight that coupling naturally arises in any configuration, whether in the near or far field. However, its impact becomes particularly critical in emerging technologies such as holographic MIMO, where the extremely dense deployment of antennas amplifies electromagnetic interactions and substantially alters channel behavior. The consequences of neglecting mutual coupling are examined in chapter 6.

Among the several methodologies available to study mutual coupling [27], [77], [78], this thesis adopts the perspective of *multiport communication theory* [79], [80]. This approach is particularly useful, since it models the communication system employing multiport networks and reduces the electromagnetic coupling problem into circuit theory, facilitating the analytical tractability.

Next, we present multiport communication theory in its most general form, while emphasizing the most common simplifications that are typically introduced for practical analysis, mainly the unilateral approximation [79], [81, Sec. 11.2], [82, Sec. 5.4]. As will be shown, the unilateral approximation significantly simplifies the analysis of multiports. However, the proof of its validity in the far field [79] does not extend to the near-field regime.

Since, as discussed in the previous chapter, communication systems are increasingly operating in the near field, it becomes essential to determine whether this approximation can still be reliably adopted. To this end, we provide a novel assessment of its validity, clarifying the conditions under which it holds and examining its implications for the accurate modeling of coupled multi-antenna systems.

3.1 Multiport communication theory

Multiport communication theory is a framework that involves a circuit-theoretic approach where inputs and outputs of the MIMO communication system are associated with ports of a multiport black-box, described by impedance or scattering matrices [83]–[85]. In multiport communication, the complexity of electromagnetic field theory is encapsulated within the model, thus simplifying the mathematical treatment [79]. Furthermore, noise modeling is more straightforward in circuit than in field theory, as the latter is a purely deterministic theory.

Multiport theory has already been used in MIMO systems literature to study different aspects such as the array gain [86], [87], the diversity gain [88], the multistreaming capability of compact arrays [89], the uplink/downlink reciprocity [90] and more recently the impact of mutual coupling in channel estimation [84] or to derive consistent models for holographic MIMO [83].

We now proceed to describe the multiport system model. Consider a narrowband communication system with N_T antennas at the transmitter and N_R at the receiver. The discrete baseband representation of the received signal is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad \mathbf{y}, \mathbf{z} \in \mathbb{C}^{N_\mathrm{R}},\ \mathbf{x} \in \mathbb{C}^{N_\mathrm{T}},\ \mathbf{H} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}, \tag{3.1}$$

where $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}_{N_\mathrm{R}}, \sigma^2 \mathbf{I}_{N_\mathrm{R}}\right)$ is AWGN, $\mathbf{H}$ denotes the MIMO channel matrix that encodes all the relevant physical context and $\mathbf{x}$ is the transmitted symbol vector. Multiport communication comes into play by providing physically consistent structures for $\mathbf{H}$ and $\mathbf{z}$. This modeling is shown in figure 3.1, and it consists of four basic parts: signal generation, impedance matching, antenna mutual coupling and noise.

More specifically, signal originates at the voltage sources ($\mathrm{v}_{\mathrm{G},n}$), is matched via a transmission network ($\mathbf{Z}_\mathrm{MT}$), and reaches the central coupling block ($\mathbf{Z}_\mathrm{A}$), which represents the physical core of the system where the electromagnetic interaction between antennas occurs. Finally, a receiving matching network ($\mathbf{Z}_\mathrm{MR}$) and the loads (Z_L) are included.

Noise is modeled through extrinsic components at the transmitter ($\mathrm{v}_{\mathrm{TEN},n}$) and the receiver ($\mathrm{v}_{\mathrm{REN},m}$), as well as intrinsic components at the receiver ($\mathrm{v}_{\mathrm{LNA},m}$ and $\mathrm{i}_{\mathrm{LNA},m}$).

All the previous steps are detailed in the subsequent sections.

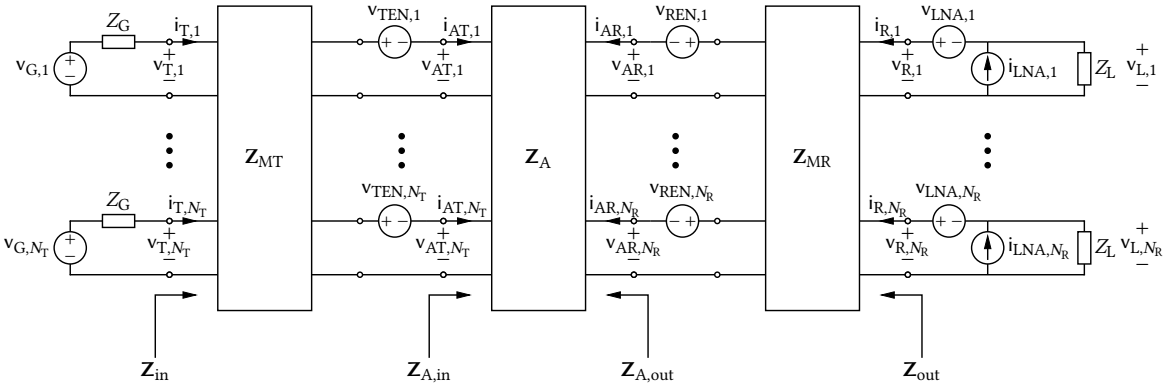


Figure 3.1: Circuit-theoretic multiport model of a multi-antenna system.

Throughout this thesis, we consider the circuit variables defined in figure 3.1, as well as their vectorized versions. For instance, the vectors $\mathbf{v}_T$ and $\mathbf{i}_T$ are defined as:

$$\mathbf{v}_T = \left(v_{T,1}, \dots, v_{T,N_T}\right)^T, \quad \mathbf{i}_T = \left(i_{T,1}, \dots, i_{T,N_T}\right)^T. \tag{3.2}$$

All voltages and currents are described by their complex envelopes. Equivalent impedances are denoted with an arrow starting at the computation point toward the corresponding circuit. For instance, $\mathbf{Z}_{in}$ represents the input impedance seen by the signal sources, while $\mathbf{Z}_{out}$ denotes the output impedance seen by the load.

Since multiport theory can be complex, the reader is encouraged to consult appendix A, where the fundamentals of circuit theory and network analysis are reviewed. Concepts such as active power, phasors, and impedance parameters are essential for following the derivations in this chapter and are detailed there for ease of reference.

3.1.1 Signal generation

The generation of the physical signal to transmit is modeled by N_T voltage sources $v_{G,n}$ with internal impedances $Z_G = R_G + jX_G$. The maximum average power that can be delivered by each generator is

$$P_{A,n} = \frac{E\left[|v_{G,n}|^2\right]}{4R_G}. \tag{3.3}$$

Therefore, the total available power is

$$P_A = \sum_n^{N_T} P_{A,n} = \frac{E\left[\|\mathbf{v}_G\|_2^2\right]}{4R_G}, \quad \mathbf{v}_G = \left(v_{G,1}, \dots, v_{G,N_T}\right)^T. \tag{3.4}$$

3.1.2 Impedance matching

Impedance matching networks can be beneficial when placed between the antenna array and the amplifiers or signal generators. These networks can be tailored to achieve specific objectives such as maximizing power transfer from the signal generators to the antennas (*i.e.* power matching), optimizing the SNR at the outputs of the receive amplifiers (*i.e.* noise

matching), or fulfilling other desired performance criteria. The transmitter matching network is described by $\mathbf{Z}_{\mathrm{MT}} \in \mathbb{C}^{2N_{\mathrm{T}} \times 2N_{\mathrm{T}}}$ such that

$$\begin{pmatrix} \mathbf{v}_{\mathrm{T}} \\ \mathbf{v}_{\mathrm{AT}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{\mathrm{MT},11} & \mathbf{Z}_{\mathrm{MT},12} \\ \mathbf{Z}_{\mathrm{MT},21} & \mathbf{Z}_{\mathrm{MT},22} \end{pmatrix} \begin{pmatrix} \mathbf{i}_{\mathrm{T}} \\ -\mathbf{i}_{\mathrm{AT}} \end{pmatrix}. \tag{3.5}$$

In a similar way, the receiver-side impedance matching network is described by $\mathbf{Z}_{\mathrm{MR}} \in \mathbb{C}^{2N_{\mathrm{R}} \times 2N_{\mathrm{R}}}$:

$$\begin{pmatrix} \mathbf{v}_{\mathrm{R}} \\ \mathbf{v}_{\mathrm{AR}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{\mathrm{MR},11} & \mathbf{Z}_{\mathrm{MR},12} \\ \mathbf{Z}_{\mathrm{MR},21} & \mathbf{Z}_{\mathrm{MR},22} \end{pmatrix} \begin{pmatrix} \mathbf{i}_{\mathrm{R}} \\ -\mathbf{i}_{\mathrm{AR}} \end{pmatrix}. \tag{3.6}$$

It is assumed that both matching networks are (see appendix A):

- **Reciprocal:** the network contains no active devices or nonreciprocal media.
- **Lossless:** the network does not absorb power.
- **Noiseless:** the network does not introduce noise into the system.

Hence,

$$\begin{aligned} \mathbf{Z}_{\mathrm{MT}} &= \mathbf{Z}_{\mathrm{MT}}^{\mathrm{T}}, \quad \mathfrak{R}(\mathbf{Z}_{\mathrm{MT}}) = \mathbf{0}_{2N_{\mathrm{T}} \times 2N_{\mathrm{T}}}, \\ \mathbf{Z}_{\mathrm{MR}} &= \mathbf{Z}_{\mathrm{MR}}^{\mathrm{T}}, \quad \mathfrak{R}(\mathbf{Z}_{\mathrm{MR}}) = \mathbf{0}_{2N_{\mathrm{R}} \times 2N_{\mathrm{R}}}. \end{aligned} \tag{3.7}$$

3.1.3 Antenna mutual coupling

The antenna multiport $\mathbf{Z}_{\mathrm{A}} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{R}}) \times (N_{\mathrm{T}}+N_{\mathrm{R}})}$ is a random matrix that models mutual coupling and can be partitioned as:

$$\begin{pmatrix} \mathbf{v}_{\mathrm{AT}} \\ \mathbf{v}_{\mathrm{AR}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{\mathrm{AT}} & \mathbf{Z}_{\mathrm{ATR}} \\ \mathbf{Z}_{\mathrm{ART}} & \mathbf{Z}_{\mathrm{AR}} \end{pmatrix} \begin{pmatrix} \mathbf{i}_{\mathrm{AT}} \\ \mathbf{i}_{\mathrm{AR}} \end{pmatrix}, \tag{3.8}$$

where $\mathbf{Z}_{\mathrm{AT}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ and $\mathbf{Z}_{\mathrm{AR}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ describe the mutual coupling between antennas in the transmit and receive arrays (intra-array coupling), and $\mathbf{Z}_{\mathrm{ATR}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{R}}}$ and $\mathbf{Z}_{\mathrm{ART}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ model the mutual coupling between the transmit and receive sides (inter-array coupling). Taking into account that antennas are reciprocal [30, Sec. 3.8], it is fulfilled that $\mathbf{Z}_{\mathrm{AT}} = \mathbf{Z}_{\mathrm{AT}}^{\mathrm{T}}$, $\mathbf{Z}_{\mathrm{AR}} = \mathbf{Z}_{\mathrm{AR}}^{\mathrm{T}}$ and $\mathbf{Z}_{\mathrm{ATR}} = \mathbf{Z}_{\mathrm{ART}}^{\mathrm{T}}$. In general, the elements of $\mathbf{Z}_{\mathrm{A}}$ depend on the type of antenna in the array as well as the communication environment. Different models for $\mathbf{Z}_{\mathrm{A}}$ are discussed in sections 3.2.1 and 3.2.2.

It is usually assumed—specially in the far field—that

$$\|\mathbf{Z}_{\mathrm{ATR}}\mathbf{i}_{\mathrm{AR}}\| \ll \|\mathbf{Z}_{\mathrm{AT}}\mathbf{i}_{\mathrm{AT}}\|, \quad \forall \mathbf{i}_{\mathrm{AR}} \in \mathbb{C}^{N_{\mathrm{R}}},\ \mathbf{i}_{\mathrm{AT}} \in \mathbb{C}^{N_{\mathrm{T}}}, \tag{3.9}$$

due to the attenuation between transmitter and receiver. This motivates the *unilateral approximation* for $\mathbf{Z}_{\mathrm{A}}$:

$$\mathbf{Z}_{\mathrm{A}} \approx \begin{pmatrix} \mathbf{Z}_{\mathrm{AT}} & \mathbf{0}_{N_{\mathrm{T}} \times N_{\mathrm{R}}} \\ \mathbf{Z}_{\mathrm{ART}} & \mathbf{Z}_{\mathrm{AR}} \end{pmatrix}. \tag{3.10}$$

It implies that $\mathbf{v}_{\mathrm{AT}} \approx \mathbf{Z}_{\mathrm{AT}}\mathbf{i}_{\mathrm{AT}}$. That is, the electrical properties at the transmit-side antenna ports are independent of the receiver, thus simplifying the analysis of the system. The validity of this approximation in near-field communications is further discussed in section 3.3.

Lossy antennas

In this thesis we only consider lossless antennas. However, with the multiport model it is possible to take into account lossy antennas [83], [87]. In order to do so, a dissipation resistance R_d must be added to each antenna port, resulting in:

$$\mathbf{Z}_\mathrm{AT} \to \mathbf{Z}_\mathrm{AT} + R_\mathrm{d}\mathbf{I}_{N_\mathrm{T}}, \quad \mathbf{Z}_\mathrm{AR} \to \mathbf{Z}_\mathrm{AR} + R_\mathrm{d}\mathbf{I}_{N_\mathrm{R}}. \tag{3.11}$$

Of course, if different antennas are used, instead of a multiple of the identity matrix an arbitrary diagonal matrix shall be used.

3.1.4 Noise

The multiport model considers both extrinsic and intrinsic noise sources. The former, originated by background radiation, is modeled by voltage sources $\{\mathrm{v}_{\mathrm{TEN},m}\}_{1\leq m\leq N_\mathrm{T}}$ and $\{\mathrm{v}_{\mathrm{REN},m}\}_{1\leq m\leq N_\mathrm{R}}$ connected to the transmitter and receiver antenna ports when no current flows (*i.e.* open-circuit) [79], [83]. Noise components are zero-mean correlated random variables $\mathbf{v}_\mathrm{TEN} \sim \mathscr{CN}(\mathbf{0}_{N_\mathrm{T}}, \mathbf{C}_\mathrm{TEN})$, $\mathbf{v}_\mathrm{REN} \sim \mathscr{CN}(\mathbf{0}_{N_\mathrm{R}}, \mathbf{C}_\mathrm{REN})$ with covariances [87]:

$$\mathbf{C}_\mathrm{TEN} = 4k_\mathrm{B}T_\mathrm{A}B_\mathrm{W}\mathfrak{R}\big(\mathbf{Z}_\mathrm{A,in}\big), \quad \mathbf{C}_\mathrm{REN} = 4k_\mathrm{B}T_\mathrm{A}B_\mathrm{W}\mathfrak{R}\big(\mathbf{Z}_\mathrm{A,out}\big), \tag{3.12}$$

where k_B is the Boltzmann constant, T_A is the noise temperature of the antennas and B_W is the equivalent noise bandwidth.

Intrinsic noise, on the other hand, is produced by the components placed after the receiver matching network, mainly from the low-noise amplifiers (LNAs) [79], [83]. It can be modeled by using N_R voltage sources $\mathrm{v}_{\mathrm{LNA},m}$ and N_R current sources $\mathrm{i}_{\mathrm{LNA},m}$ such that $\mathbf{i}_\mathrm{LNA} \sim \mathscr{CN}(\mathbf{0}_{N_\mathrm{R}}, \sigma_i^2\mathbf{I}_{N_\mathrm{R}})$ and $\mathbf{v}_\mathrm{LNA} \sim \mathscr{CN}(\mathbf{0}_{N_\mathrm{R}}, R_\mathrm{N}^2\sigma_i^2\mathbf{I}_{N_\mathrm{R}})$, with R_N the noise resistance of the LNAs given by the manufacturer. Although extrinsic and intrinsic noise are assumed to be uncorrelated, the voltage and current in the same amplifier are not:

$$\mathrm{E}\big[\mathbf{v}_\mathrm{LNA}\mathbf{i}_\mathrm{LNA}^\mathrm{H}\big] = \rho_\mathrm{N}R_\mathrm{N}\sigma_i^2\mathbf{I}_{N_\mathrm{R}}, \tag{3.13}$$

where ρ_N is the complex noise correlation coefficient of the LNAs.

3.1.5 Input-output relation

The input-output relation is an expression that relates the source voltages $\mathbf{v}_\mathrm{G}$ with the load voltages $\mathbf{v}_\mathrm{L}$. Taking into account noise and system linearity, it can be expressed as [90]:

$$\mathbf{v}_\mathrm{L} = \mathbf{D}\mathbf{v}_\mathrm{G} + \mathbf{n}, \quad \mathbf{n} \sim \mathscr{CN}(\mathbf{0}_{N_\mathrm{R}}, \mathbf{C}_\mathbf{n}). \tag{3.14}$$

Leveraging the superposition theorem, the first term corresponds to the noise-free voltage $\mathbf{v}_{\mathrm{L}|\mathrm{nf}} = \mathbf{D}\mathbf{v}_\mathrm{G}$ (*i.e.* $\mathbf{v}_\mathrm{TEN} = \mathbf{v}_\mathrm{REN} = \mathbf{v}_\mathrm{LNA} = \mathbf{i}_\mathrm{LNA} = \mathbf{0}_{N_\mathrm{R}}$) and the second to source-free voltage $\mathbf{v}_{\mathrm{L}|\mathrm{sf}} = \mathbf{n}$ (*i.e.* $\mathbf{v}_\mathrm{G} = \mathbf{0}_{N_\mathrm{T}}$).

For simplicity of notation, we omit the subscripts |nf and |sf in the following subsections, as it will be clear from the context.

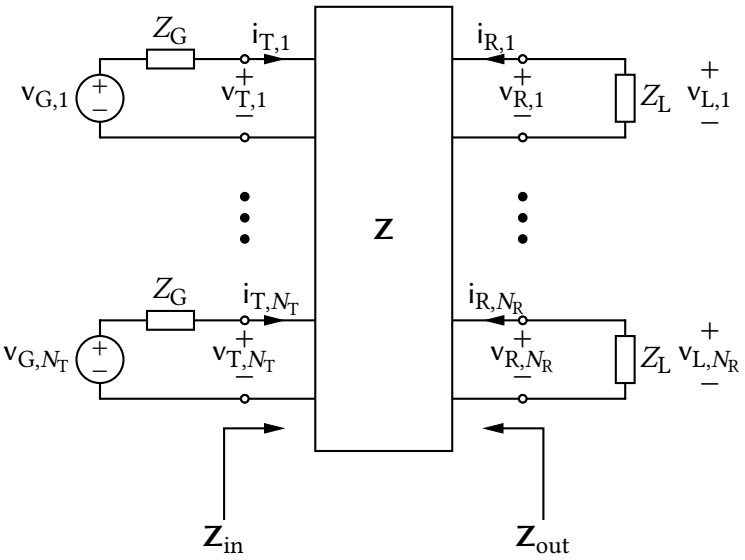


Figure 3.2: Multiport model of a noiseless MIMO system.

Noise-free relation

In the noise-free scenario, the three multiports in figure 3.1 are connected in cascade. Therefore, they can be combined into a single equivalent multiport, as illustrated in figure 3.2. The relation between voltages and currents is given by

$$\begin{pmatrix}\mathbf{v}_T\\ \mathbf{v}_R\end{pmatrix} = \mathbf{Z}\cdot\begin{pmatrix}\mathbf{i}_T\\ \mathbf{i}_R\end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix}\mathbf{Z}_T & \mathbf{Z}_{TR}\\ \mathbf{Z}_{RT} & \mathbf{Z}_R\end{pmatrix}, \tag{3.15}$$

where each block of $\mathbf{Z}$ is obtained from circuit analysis as (see appendix B)

$$\begin{aligned}\mathbf{Z}_T &= \mathbf{Z}_{MT,11} - \mathbf{Z}_{MT,12}\left(\mathbf{A}_T - \mathbf{Z}_{ATR}\mathbf{A}_R^{-1}\mathbf{Z}_{ART}\right)^{-1}\mathbf{Z}_{MT,21},\\ \mathbf{Z}_R &= \mathbf{Z}_{MR,11} - \mathbf{Z}_{MR,12}\left(\mathbf{A}_R - \mathbf{Z}_{ART}\mathbf{A}_T^{-1}\mathbf{Z}_{ATR}\right)^{-1}\mathbf{Z}_{MR,21},\\ \mathbf{Z}_{TR} &= \mathbf{Z}_{MT,12}\left(\mathbf{A}_T - \mathbf{Z}_{ATR}\mathbf{A}_R^{-1}\mathbf{Z}_{ART}\right)^{-1}\mathbf{Z}_{ATR}\mathbf{A}_R^{-1}\mathbf{Z}_{MR,21},\end{aligned} \tag{3.16}$$

with the following auxiliary variables:

$$\mathbf{A}_R = \mathbf{Z}_{AR} + \mathbf{Z}_{MR,22}, \quad \mathbf{A}_T = \mathbf{Z}_{AT} + \mathbf{Z}_{MT,22}. \tag{3.17}$$

This network is also reciprocal, as it is a cascade connection of three reciprocal networks [91], hence $\mathbf{Z}_T = \mathbf{Z}_T^T$, $\mathbf{Z}_R = \mathbf{Z}_R^T$ and

$$\mathbf{Z}_{RT} = \mathbf{Z}_{TR}^T = \mathbf{Z}_{MR,12}\mathbf{A}_R^{-1}\mathbf{Z}_{ART}\left(\mathbf{A}_T - \mathbf{Z}_{ATR}\mathbf{A}_R^{-1}\mathbf{Z}_{ART}\right)^{-1}\mathbf{Z}_{MT,21}. \tag{3.18}$$

From (3.15) and Ohm's law, we can obtain the matrix of the (noiseless) input-output relationship $\mathbf{v}_{L|nf} = \mathbf{D}\mathbf{v}_G$:

$$\mathbf{D} = Z_L\left(Z_L\mathbf{I}_{N_R} + \mathbf{Z}_R\right)^{-1}\mathbf{Z}_{RT}(Z_G\mathbf{I}_{N_T} + \mathbf{Z}_T - \mathbf{Z}_{TR}(Z_L\mathbf{I}_{N_R} + \mathbf{Z}_R)^{-1}\mathbf{Z}_{RT})^{-1}. \tag{3.19}$$

Applying the Woodbury matrix identity [19], [92, Sec. 0.7], $\mathbf{D}$ can also be expressed as

$$\mathbf{D} = Z_L\left(Z_L\mathbf{I}_{N_R} + \mathbf{Z}_R - \mathbf{Z}_{RT}(Z_G\mathbf{I}_{N_T} + \mathbf{Z}_T)^{-1}\mathbf{Z}_{TR}\right)^{-1}\mathbf{Z}_{RT}(Z_G\mathbf{I}_{N_T} + \mathbf{Z}_T)^{-1}. \tag{3.20}$$

When the receiver and transmitter are sufficiently far apart, the unilateral approximation $\mathbf{Z}_{\text{ATR}} \approx \mathbf{0}_{N_\text{T}\times N_\text{R}}$ holds so $\mathbf{Z}_{\text{TR}} \approx \mathbf{0}_{N_\text{T}\times N_\text{R}}$ too, and $\mathbf{D}$ can be simplified to

$$\mathbf{D}_{\text{UA}} = Z_\text{L}\left(Z_\text{L}\mathbf{I}_{N_\text{R}} + \mathbf{Z}_\text{R}\right)^{-1}\mathbf{Z}_{\text{RT}}(Z_\text{G}\mathbf{I}_{N_\text{T}} + \mathbf{Z}_\text{T})^{-1}. \tag{3.21}$$

Observe that now $\mathbf{Z}_\text{T}$ and $\mathbf{Z}_\text{R}$ are not random, since $\mathbf{Z}_{\text{ATR}}\mathbf{A}_\text{R}^{-1}\mathbf{Z}_{\text{ART}}$ and $\mathbf{Z}_{\text{ART}}\mathbf{A}_\text{T}^{-1}\mathbf{Z}_{\text{ATR}}$ are approximately zero and all the other terms are deterministic.

These same expressions can be given in terms of $\mathbf{Z}_{\text{in}}$ and $\mathbf{Z}_{\text{out}}$. Following appendix A.3, both matrices can be computed as:

$$\begin{aligned}\mathbf{Z}_{\text{in}} &= \mathbf{Z}_\text{T} - \mathbf{Z}_{\text{TR}}\left(Z_\text{L}\mathbf{I}_{N_\text{R}} + \mathbf{Z}_\text{R}\right)^{-1}\mathbf{Z}_{\text{RT}},\\ \mathbf{Z}_{\text{out}} &= \mathbf{Z}_\text{R} - \mathbf{Z}_{\text{RT}}\left(Z_\text{G}\mathbf{I}_{N_\text{T}} + \mathbf{Z}_\text{T}\right)^{-1}\mathbf{Z}_{\text{TR}}.\end{aligned} \tag{3.22}$$

Therefore,

$$\begin{aligned}\mathbf{D} &= Z_\text{L}\left(Z_\text{L}\mathbf{I}_{N_\text{R}} + \mathbf{Z}_\text{R}\right)^{-1}\mathbf{Z}_{\text{RT}}(Z_\text{G}\mathbf{I}_{N_\text{T}} + \mathbf{Z}_{\text{in}})^{-1}\\ &= Z_\text{L}\left(Z_\text{L}\mathbf{I}_{N_\text{R}} + \mathbf{Z}_{\text{out}}\right)^{-1}\mathbf{Z}_{\text{RT}}(Z_\text{G}\mathbf{I}_{N_\text{T}} + \mathbf{Z}_\text{T})^{-1} = \mathbf{Q}\mathbf{Z}_{\text{RT}}(Z_\text{G}\mathbf{I}_{N_\text{T}} + \mathbf{Z}_\text{T})^{-1},\end{aligned} \tag{3.23}$$

where we have defined

$$\mathbf{Q} = Z_\text{L}\left(Z_\text{L}\mathbf{I}_{N_\text{R}} + \mathbf{Z}_{\text{out}}\right)^{-1}, \tag{3.24}$$

since it will be useful later on.

Under the unilateral approximation, $\mathbf{Z}_{\text{in,UA}} = \mathbf{Z}_\text{T}$ and $\mathbf{Z}_{\text{out,UA}} = \mathbf{Z}_\text{R}$, so (3.23) is equivalent to (3.21).

Signal-free relation

The noise term is obtained in a similar manner setting $\mathbf{v}_\text{G} = \mathbf{0}_{N_\text{T}}$. Due to the higher number of noise sources, we leverage the superposition principle to facilitate the computation of $\mathbf{n}$ by first considering the receiver-side noises and then the transmitter-side noise.

Letting $\mathbf{v}_{\text{TEN}} = \mathbf{0}_{N_\text{T}}$ and applying the Thévenin theorem at the output of the receiver matching network, the equivalent circuit in figure 3.3 is obtained [93], where $\mathbf{v}_{\text{RN}} \sim \mathcal{CN}(\mathbf{0}_{N_\text{R}}, \mathbf{C}_{\text{RN}})$ and $\mathbf{C}_{\text{RN}} = 4k_\text{B}T_\text{A}B_\text{W}\mathfrak{R}(\mathbf{Z}_{\text{out}})$.

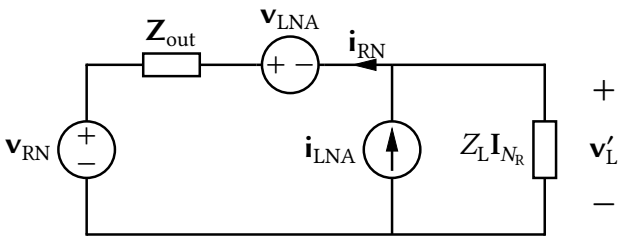


Figure 3.3: Equivalent circuit to find the receiver-side noise expression.

Through Kirchhoff's current law (KCL) and Kirchhoff's voltage law (KVL):

$$\begin{aligned}&\text{KCL: } \mathbf{i}_{\text{RN}} = \mathbf{i}_{\text{LNA}} - \frac{\mathbf{v}'_\text{L}}{Z_\text{L}},\\ &\text{KVL: } \mathbf{v}'_\text{L} + \mathbf{v}_{\text{LNA}} - \mathbf{Z}_{\text{out}}\mathbf{i}_{\text{RN}} - \mathbf{v}_{\text{RN}} = \mathbf{0}_{N_\text{R}}.\end{aligned} \tag{3.25}$$

Substituting the first equation into the second one and isolating $\mathbf{v}'_{\mathrm{L}}$:

$$\mathbf{v}'_{\mathrm{L}} = \mathbf{Q}(\mathbf{v}_{\mathrm{RN}} - \mathbf{v}_{\mathrm{LNA}} + \mathbf{Z}_{\mathrm{out}}\mathbf{i}_{\mathrm{LNA}}). \tag{3.26}$$

The effects of the transmitting antennas noise can be computed by setting all sources but $\mathbf{v}_{\mathrm{TEN}}$ to zero, which results in the circuit in figure 3.4, with $\mathbf{v}_{\mathrm{TN}} \sim \mathcal{CN}(\mathbf{0}_{N_{\mathrm{T}}}, \mathbf{C}_{\mathrm{TN}})$ and $\mathbf{C}_{\mathrm{TN}} = 4k_{\mathrm{B}}T_{\mathrm{A}}B_{\mathrm{W}}\Re(\mathbf{Z}_{\mathrm{in}})$.

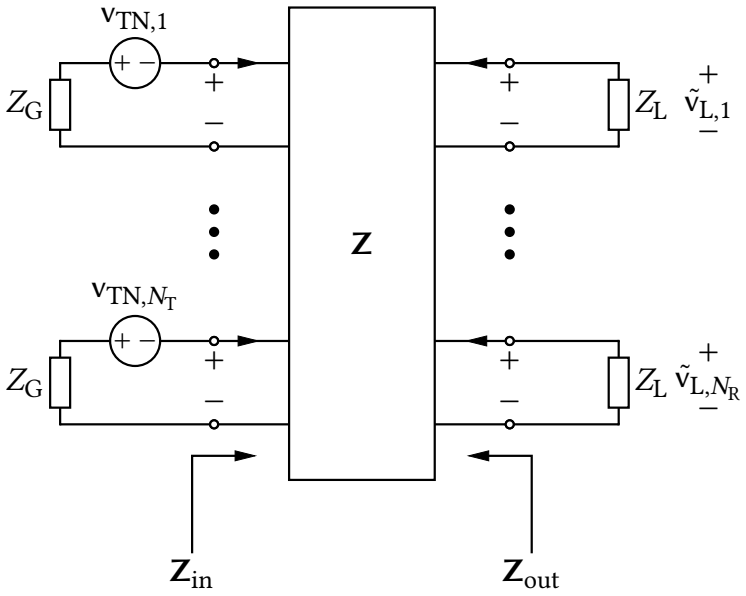


Figure 3.4: Multiport network with transmitting antenna noise before the matching network.

This circuit is similar to the noiseless one shown in figure 3.2, but with the source and its impedance swapped, and the source reversed. Since exchanging these two components does not affect $\mathbf{i}_{\mathrm{T}}$, from the noise-free analysis it follows that:

$$\tilde{\mathbf{v}}_{\mathrm{L}} = -\mathbf{D}\mathbf{v}_{\mathrm{TN}} = -\mathbf{Q}\mathbf{Z}_{\mathrm{RT}}(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}})^{-1}\mathbf{v}_{\mathrm{TN}}. \tag{3.27}$$

Finally, the noise component at the receiver is obtained by superposition:

$$\mathbf{n} = \mathbf{v}'_{\mathrm{L}} + \tilde{\mathbf{v}}_{\mathrm{L}} = \mathbf{Q}\big(\mathbf{v}_{\mathrm{RN}} - \mathbf{v}_{\mathrm{LNA}} + \mathbf{Z}_{\mathrm{out}}\mathbf{i}_{\mathrm{LNA}} - \mathbf{Z}_{\mathrm{RT}}(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}})^{-1}\mathbf{v}_{\mathrm{TN}}\big). \tag{3.28}$$

Similar to the noise-free scenario, the unilateral approximation greatly simplifies the noise expression. Besides the aforementioned simplifications of $\mathbf{Q}$ and $\mathbf{Z}_{\mathrm{out}}$, if the unilateral approximation is valid the transmitting antennas noise can be omitted [79], so $\mathbf{v}_{\mathrm{TN}} \approx \mathbf{0}_{N_{\mathrm{T}}}$ and

$$\mathbf{n}_{\mathrm{UA}} = \mathbf{Q}(\mathbf{v}_{\mathrm{RN}} - \mathbf{v}_{\mathrm{LNA}} + \mathbf{Z}_{\mathrm{R}}\mathbf{i}_{\mathrm{LNA}}). \tag{3.29}$$

A summary of the different expressions in the general case and under the unilateral approximation is given in section 3.1.8.

3.1.6 Transmit power and noise covariance matrix

Transmit power is defined as the average active power delivered to the transmitting antennas. When expressed in terms of complex envelopes, it can be written as

$$P_{\mathrm{T}} = \mathrm{E}\big[\Re\big(\mathbf{v}_{\mathrm{AT}}^{\mathrm{H}}\mathbf{i}_{\mathrm{AT}}\big)\big] = \mathrm{E}\big[\Re\big(\mathbf{v}_{\mathrm{T}}^{\mathrm{H}}\mathbf{i}_{\mathrm{T}}\big)\big], \tag{3.30}$$

where the second equality follows from assuming that the matching networks are either lossless or absent. Moreover, if antennas are also lossless, the transmit power is equal to the radiated power.

From Ohm's law and the multiport equation we know that

$$\begin{aligned} \mathbf{v}_{\mathrm{T}} &= \mathbf{v}_{\mathrm{G}} - Z_{\mathrm{G}}\mathbf{i}_{\mathrm{T}} \\ \mathbf{v}_{\mathrm{T}} &= \mathbf{Z}_{\mathrm{in}}\mathbf{i}_{\mathrm{T}}, \end{aligned} \tag{3.31}$$

and subtracting both equations results in

$$\mathbf{i}_{\mathrm{T}} = \big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-1}\mathbf{v}_{\mathrm{G}}. \tag{3.32}$$

Taking into account that $\mathbf{Z}_{\mathrm{in}}$ is symmetric, substituting $\mathbf{v}_{\mathrm{T}}$ and $\mathbf{i}_{\mathrm{T}}$ in (3.30) yields

$$\begin{aligned} P_{\mathrm{T}} &= \mathrm{E}\Big[\Re\Big(\mathbf{v}_{\mathrm{G}}^{\mathrm{H}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-\mathrm{H}}\mathbf{Z}_{\mathrm{in}}^{\mathrm{H}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-1}\mathbf{v}_{\mathrm{G}}\Big)\Big] \\ &= \mathrm{E}\Big[\mathbf{v}_{\mathrm{G}}^{\mathrm{H}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-\mathrm{H}}\Re(\mathbf{Z}_{\mathrm{in}})\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-1}\mathbf{v}_{\mathrm{G}}\Big] \\ &= \frac{\mathrm{E}\big[\mathbf{v}_{\mathrm{G}}^{\mathrm{H}}\mathbf{B}\mathbf{v}_{\mathrm{G}}\big]}{R_{\mathrm{G}}}, \end{aligned} \tag{3.33}$$

where $\mathbf{B}$ denotes the power coupling matrix, defined as

$$\mathbf{B} = R_{\mathrm{G}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-\mathrm{H}}\Re(\mathbf{Z}_{\mathrm{in}})\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\big)^{-1}. \tag{3.34}$$

Under the unilateral approximation, the power coupling matrix is obtained replacing $\mathbf{Z}_{\mathrm{in}}$ with $\mathbf{Z}_{\mathrm{T}}$ in the expression above, resulting in a deterministic power coupling matrix.

In the previous section, we derived an expression for noise following the multiport model. Since noise is a purely random phenomenon modeled as a zero-mean Gaussian process, it is fully characterized by its second-order moment. In particular, the noise covariance matrix is given by

$$\begin{aligned} \mathbf{C}_{\mathbf{n}} = \mathrm{E}\big[\mathbf{n}\mathbf{n}^{\mathrm{H}}\big] &= \mathbf{Q}\mathrm{E}\Big[\mathbf{v}_{\mathrm{RN}}\mathbf{v}_{\mathrm{RN}}^{\mathrm{H}} - \mathbf{v}_{\mathrm{LNA}}\mathbf{v}_{\mathrm{LNA}}^{\mathrm{H}} - \mathbf{v}_{\mathrm{LNA}}\mathbf{i}_{\mathrm{LNA}}^{\mathrm{H}}\mathbf{Z}_{\mathrm{out}}^{\mathrm{H}} - \mathbf{Z}_{\mathrm{out}}\mathbf{i}_{\mathrm{LNA}}\mathbf{v}_{\mathrm{LNA}}^{\mathrm{H}} \\ &\qquad + \mathbf{Z}_{\mathrm{out}}\mathbf{i}_{\mathrm{LNA}}\mathbf{i}_{\mathrm{LNA}}^{\mathrm{H}}\mathbf{Z}_{\mathrm{out}}^{\mathrm{H}} + \mathbf{Z}_{\mathrm{RT}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\big)^{-1}\mathbf{v}_{\mathrm{TN}}\mathbf{v}_{\mathrm{TN}}^{\mathrm{H}}\big(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\big)^{-\mathrm{H}}\Big]\mathbf{Q}^{\mathrm{H}} \\ &= \mathbf{Q}\Big(\mathbf{C}_{\mathrm{RN}} + \sigma_i^2 R_{\mathrm{N}}^2\mathbf{I}_{N_{\mathrm{R}}} + \sigma_i^2\mathbf{Z}_{\mathrm{out}}\mathbf{Z}_{\mathrm{out}}^{\mathrm{H}} - 2\sigma_i^2 R_{\mathrm{N}}\Re(\rho_{\mathrm{N}}^{*}\mathbf{Z}_{\mathrm{out}})\Big)\mathbf{Q}^{\mathrm{H}} + \mathbf{D}\mathbf{C}_{\mathrm{TN}}\mathbf{D}^{\mathrm{H}}. \end{aligned} \tag{3.35}$$

If the unilateral approximation holds, the last term can be neglected setting $\mathbf{C}_{\mathrm{TN}} = \mathbf{0}_{N_{\mathrm{T}}}$, and $\mathbf{Z}_{\mathrm{out}}$ should be replaced with $\mathbf{Z}_{\mathrm{R}}$.

3.1.7 Matching networks optimization

From a practical point of view, it is convenient to design the transmitter matching network, $\mathbf{Z}_{\mathrm{MT}}$, such that maximum power transfer is achieved. Ideally, this configuration ensures that all the power available at the sources is effectively transmitted to the antenna array, a condition known as power matching. Formally, $P_{\mathrm{T}} = P_{\mathrm{A}}$. This is achieved with a network such that the input impedance seen from the sources is the conjugate match of the source impedance, $\mathbf{Z}_{\mathrm{in}} = Z_{\mathrm{G}}^{*}\mathbf{I}_{N_{\mathrm{T}}}$, yielding $\mathbf{B} = \mathbf{I}_{N_{\mathrm{T}}}/4$. Similarly, on the receiving end, the network $\mathbf{Z}_{\mathrm{MR}}$ should be designed with the primary goal of maximizing the SNR at the output, a process referred to as noise matching.

However, achieving these conditions is not straightforward, especially when the unilateral approximation does not hold. In this situation, (3.16) and (3.22) evidence that $\mathbf{Z}_{\mathrm{MT}}$ and $\mathbf{Z}_{\mathrm{MR}}$ are a coupled transcendental equation with no closed-form solution and only numerical approximations are possible [94], [95].

The complexity of the problem simplifies considerably under the unilateral approximation. By setting $\mathbf{Z}_{\mathrm{ATR}} = \mathbf{0}_{N_{\mathrm{T}}\times N_{\mathrm{R}}}$, it follows that $\mathbf{Z}_{\mathrm{in}} = \mathbf{Z}_{\mathrm{T}}$, so the transmitter power matching network can be computed as

$$\mathbf{Z}_{\mathrm{MT}}^{\star} = \begin{pmatrix} -\mathrm{j}X_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} & -\mathrm{j}\sqrt{R_{\mathrm{G}}}\mathfrak{R}(\mathbf{Z}_{\mathrm{AT}})^{1/2} \\ -\mathrm{j}\sqrt{R_{\mathrm{G}}}\mathfrak{R}(\mathbf{Z}_{\mathrm{AT}})^{1/2} & -\mathrm{j}\mathfrak{F}(\mathbf{Z}_{\mathrm{AT}}) \end{pmatrix}. \tag{3.36}$$

On the other hand, the receiver noise matching network is given by

$$\mathbf{Z}_{\mathrm{MR}}^{\star} = \begin{pmatrix} \mathrm{j}Z_{\mathrm{opt}}\mathbf{I}_{N_{\mathrm{R}}} & \mathrm{j}\sqrt{\mathfrak{R}\left(Z_{\mathrm{opt}}\right)}\mathfrak{R}(\mathbf{Z}_{\mathrm{AR}})^{1/2} \\ \mathrm{j}\sqrt{\mathfrak{R}\left(Z_{\mathrm{opt}}\right)}\mathfrak{R}(\mathbf{Z}_{\mathrm{AR}})^{1/2} & -\mathrm{j}\mathfrak{F}(\mathbf{Z}_{\mathrm{AR}}) \end{pmatrix}, \tag{3.37}$$

which yields $\mathbf{Z}_{\mathrm{R}} = Z_{\mathrm{opt}}\mathbf{I}_{N_{\mathrm{R}}}$, with $Z_{\mathrm{opt}} = R_{\mathrm{N}}\left(\sqrt{1-\mathfrak{F}(\rho_{\mathrm{N}})^{2}} + \mathrm{j}\mathfrak{F}(\rho_{\mathrm{N}})\right)$. Note that by employing the matching network in (3.37), the receiving antennas are decoupled, since $\mathbf{Z}_{\mathrm{R}}$ is a diagonal matrix. However, this is not equivalent to an uncoupled array, given that in a decoupled array, the receiving array mutual coupling appears in $\mathbf{D}$ through $\mathbf{Z}_{\mathrm{AR}}$.

Even if the unilateral approximation is valid, the design of optimal matching networks is very challenging for arrays with a large number of antennas [79], [80], [83]. For this reason, in mMIMO they are usually omitted [87], [90] or replaced by self-impedance matching networks [96]. This latter approach neglects the mutual coupling between antennas by replacing the impedance matrices $\mathbf{Z}_{\mathrm{AT}}$ and $\mathbf{Z}_{\mathrm{AR}}$ with their diagonal counterparts, $\mathrm{diag}(\mathbf{Z}_{\mathrm{AT}})$ and $\mathrm{diag}(\mathbf{Z}_{\mathrm{AR}})$.

3.1.8 Summary

Given the large number of parameters involved in the multiport communication model, as well as their difference in the general case and under the unilateral approximation, we find it helpful to provide an overview in table 3.1.

Recall that $\mathbf{A}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{AT}} + \mathbf{Z}_{\mathrm{MT},22}$ and $\mathbf{A}_{\mathrm{R}} = \mathbf{Z}_{\mathrm{AR}} + \mathbf{Z}_{\mathrm{MR},22}$.

Table 3.1: Summary of multiport communication theory variables.

General expression	Unilateral approximation
$\mathbf{Z}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{MT},11} - \mathbf{Z}_{\mathrm{MT},12}\left(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}\right)^{-1}\mathbf{Z}_{\mathrm{MT},21}$	$\mathbf{Z}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{MT},11} - \mathbf{Z}_{\mathrm{MT},12}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{MT},21}$
$\mathbf{Z}_{\mathrm{R}} = \mathbf{Z}_{\mathrm{MR},11} - \mathbf{Z}_{\mathrm{MR},12}\left(\mathbf{A}_{\mathrm{R}} - \mathbf{Z}_{\mathrm{ART}}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{ATR}}\right)^{-1}\mathbf{Z}_{\mathrm{MR},21}$	$\mathbf{Z}_{\mathrm{R}} = \mathbf{Z}_{\mathrm{MR},11} - \mathbf{Z}_{\mathrm{MR},12}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{MR},21}$
$\mathbf{Z}_{\mathrm{TR}} = \mathbf{Z}_{\mathrm{MT},12}\left(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}\right)^{-1}\mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{MR},21}$	$\mathbf{Z}_{\mathrm{TR}} = \mathbf{0}_{N_{\mathrm{T}}\times N_{\mathrm{R}}}$
$\mathbf{Z}_{\mathrm{RT}} = \mathbf{Z}_{\mathrm{MR},12}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}\left(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}\right)^{-1}\mathbf{Z}_{\mathrm{MT},21}$	$\mathbf{Z}_{\mathrm{RT}} = \mathbf{Z}_{\mathrm{MR},12}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{MT},21}$
$\mathbf{Z}_{\mathrm{in}} = \mathbf{Z}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{TR}}\left(Z_{\mathrm{L}}\mathbf{I}_{N_{\mathrm{R}}} + \mathbf{Z}_{\mathrm{R}}\right)^{-1}\mathbf{Z}_{\mathrm{RT}}$	$\mathbf{Z}_{\mathrm{in}} = \mathbf{Z}_{\mathrm{T}}$
$\mathbf{Z}_{\mathrm{out}} = \mathbf{Z}_{\mathrm{R}} - \mathbf{Z}_{\mathrm{RT}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-1}\mathbf{Z}_{\mathrm{TR}}$	$\mathbf{Z}_{\mathrm{out}} = \mathbf{Z}_{\mathrm{R}}$
$\mathbf{Q} = Z_{\mathrm{L}}\left(Z_{\mathrm{L}}\mathbf{I}_{N_{\mathrm{R}}} + \mathbf{Z}_{\mathrm{out}}\right)^{-1}$	$\mathbf{Q} = Z_{\mathrm{L}}\left(Z_{\mathrm{L}}\mathbf{I}_{N_{\mathrm{R}}} + \mathbf{Z}_{\mathrm{R}}\right)^{-1}$
$\mathbf{D} = \mathbf{Q}\mathbf{Z}_{\mathrm{RT}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-1}$	$\mathbf{D} = \mathbf{Q}\mathbf{Z}_{\mathrm{RT}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-1}$
$\mathbf{B} = R_{\mathrm{G}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\right)^{-\mathrm{H}}\Re(\mathbf{Z}_{\mathrm{in}})\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{in}}\right)^{-1}$	$\mathbf{B} = R_{\mathrm{G}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-\mathrm{H}}\Re(\mathbf{Z}_{\mathrm{T}})\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-1}$
$\mathbf{n} = \mathbf{Q}\left(\mathbf{v}_{\mathrm{RN}} - \mathbf{v}_{\mathrm{LNA}} + \mathbf{Z}_{\mathrm{out}}\mathbf{i}_{\mathrm{LNA}} - \mathbf{Z}_{\mathrm{RT}}\left(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}\right)^{-1}\mathbf{v}_{\mathrm{TN}}\right)$	$\mathbf{n} = \mathbf{Q}\left(\mathbf{v}_{\mathrm{RN}} - \mathbf{v}_{\mathrm{LNA}} + \mathbf{Z}_{\mathrm{R}}\mathbf{i}_{\mathrm{LNA}}\right)$

3.2 Mutual coupling impedance matrices

The multiport model discussed in the previous section introduces the mutual coupling matrix $\mathbf{Z}_{\mathrm{A}}$, which can be decomposed into four submatrices representing different coupling interactions within the system:

- Mutual coupling among the transmit antennas: $\mathbf{Z}_{\mathrm{AT}}$.
- Mutual coupling among the receive antennas: $\mathbf{Z}_{\mathrm{AR}}$.
- Mutual coupling from the transmit antennas to the receive antennas: $\mathbf{Z}_{\mathrm{ART}}$.
- Mutual coupling from the receive antennas to the transmit antennas: $\mathbf{Z}_{\mathrm{ATR}}$.

Before computing the elements of these matrices for different types of antenna, we first define the concepts of *self-impedance* and *mutual impedance* and clarify their physical meaning.

First, consider a single current element (*i.e.* an antenna) radiating into an unbounded medium. Its self-impedance is given by the ratio of the induced voltage to the driving current [30, Sec. 8.5]:

$$Z_{\mathrm{in}} = \frac{V_{\mathrm{in}}}{I_{\mathrm{in}}}. \tag{3.38}$$

Its real part denotes the radiation resistance of the antenna, whereas the imaginary part represents the self-reactance.

The self-impedance of an antenna coincides with its input impedance only in the absence of any other current elements. The presence of an obstacle—such as another nearby antenna element—can alter the current distribution on the antenna, modify the radiated electromagnetic field, and consequently change its input impedance. As a result, the input impedance of an antenna in the presence of surrounding elements depends not only on its self-impedance but also on the mutual impedance between the driven element and the neighboring elements [97, Ch. 13].

For the analysis, we consider a pair of current elements and represent them as a two-port network such that

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}, \tag{3.39}$$

where I_1 and I_2 are the driving currents and V_1 and V_2 the induced voltages on each element [98]. When the second port is open-circuited, the input impedance of the first antenna is given by

$$Z_{11} = \frac{V_1}{I_1}|_{I_2=0}. \tag{3.40}$$

Similarly, when the first port is open-circuited,

$$Z_{22} = \frac{V_2}{I_2}|_{I_1=0} \tag{3.41}$$

is the input impedance of the second antenna. These values coincide with the self-impedances of the first and second antennas if they are canonical minimum scattering (CMS) antennas, which means that they are "invisible" when the accessible waveguide terminals are open-circuited [83], [99], [100].

It is important to note that the pairwise treatment performed in sections 3.2.1 and 3.2.2 relies on the antennas being CMS. If this condition is not met, the interactions with neighboring antennas must be explicitly considered. Fortunately, thin antennas generally satisfy this property [79], [80], [83].

On the other hand, the mutual impedance at the first port due to a current in the second port is given by

$$Z_{12} = \frac{V_1}{I_2}|_{I_1=0}. \tag{3.42}$$

Since antennas are reciprocal devices, it is fulfilled that

$$Z_{21} = Z_{12} = \frac{V_2}{I_1}|_{I_2=0}, \tag{3.43}$$

which is the mutual impedance at the second port due to a current in the first port.

Therefore, the presence of another element modifies the input impedance of an antenna [97, Ch. 13]. The extent and nature of this effect depend on three factors: the type of antenna, the relative positioning of the elements, and the type of feed used to excite them [30, Sec. 8.6]. In the following, we describe the construction of the mutual coupling matrix $\mathbf{Z}_{\mathrm{A}}$ for different antenna types. We begin by examining the intra-array coupling matrices, followed by the inter-array coupling matrices.

3.2.1 Intra-array coupling impedance matrix

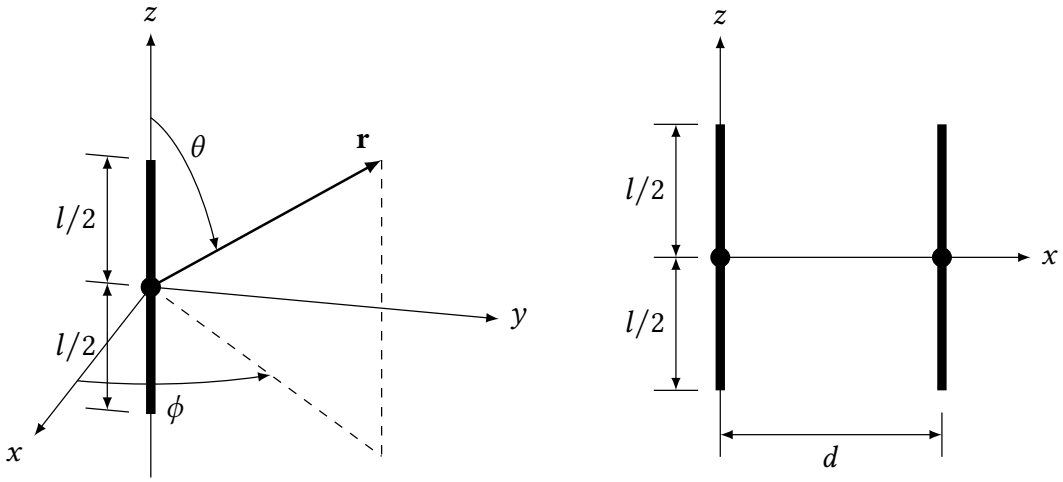
Matrices $\mathbf{Z}_{\mathrm{AT}}$ and $\mathbf{Z}_{\mathrm{AR}}$ previously defined in section 3.1.3 represent the intra-array coupling at the transmitter and receiver, respectively. These matrices contain both the self-impedances and mutual impedances of the antenna elements. Specifically, the entry at position (m, n) denotes the mutual impedance observed at the m-th port due to a current at the n-th port, with all other ports open-circuited. Under this definition, the diagonal elements correspond to the self-impedances of each antenna [36, Sec. 6.1].

As previously discussed, mutual impedances depend on both the spatial configuration of the array and the type of antenna employed. While many array configurations and antenna types exist, we restrict our analysis to ULAs with identical elements which can be either Hertzian dipoles or half-wavelength dipoles.

In the sequel, we construct the intra-array coupling matrices for each of these antenna types and compare their corresponding mutual coupling functions, including that of the isotropic radiator. For simplicity, we use the notation $\mathbf{Z}_{\mathrm{A}\bullet}$ to denote the intra-array coupling matrix, where $\bullet$ stands for either T or R.

Hertzian dipoles

The Hertzian dipole, also known as infinitesimal dipole, is the most elementary type of antenna that fulfills Maxwell equations. It consists of an electrically short and infinitesimally thin wire with an almost uniform current density. Although Hertzian dipoles cannot be physically constructed, they serve as building blocks for more complex antenna geometries [30, Sec. 4.2].



(a) A Hertzian dipole in spherical coordinates. **(b)** Two side-by-side Hertzian dipoles.

Figure 3.5: Representation of a single and two side-by-side Hertzian dipoles.

In order to find the mutual impedance matrix for an array of side-by-side Hertzian dipoles, we first consider a single Hertzian dipole of length $l \ll \lambda$ located along the z axis, with a uniform current I_1, as depicted in figure 3.5a [30, Sec. 4.2]. Its electric field at a point (in spherical coordinates) $\mathbf{r} = (r, \theta, \phi)$ is given by

$$E_r = \eta \frac{I_1 \, l \cos\theta}{2\pi k} \left(\frac{k}{r^2} - \frac{\mathrm{j}}{r^3} \right) \mathrm{e}^{-\mathrm{j}kr}, \tag{3.44}$$

$$E_\theta = \mathrm{j}\eta \frac{I_1 \, l \sin\theta}{4\pi k} \left(\frac{k^2}{r} - \frac{\mathrm{j}k}{r^2} - \frac{1}{r^3} \right) \mathrm{e}^{-\mathrm{j}kr}, \tag{3.45}$$

$$E_\phi = 0. \tag{3.46}$$

If a second (open-circuited) dipole is placed at coordinates $(d, 0, 0)$, oriented as the first one (see figure 3.5b), the electric field produced by the first dipole induces an open-circuit voltage

on the second [101, Sec. 25.3]. This voltage is given by

$$V_2 = -\int_{-l/2}^{l/2} E_z(z)\,\mathrm{d}z. \tag{3.47}$$

Since a Hertzian dipole is electrically small, the z component of the electric field can be well approximated by the θ component [98]:

$$E_z(r) \approx -E_\theta(r = d, \theta = \pi/2) = \mathrm{j}\eta\frac{I_1\, l}{4\pi k}\left(-\frac{k^2}{r} + \frac{\mathrm{j}k}{r^2} + \frac{1}{r^3}\right)\mathrm{e}^{-\mathrm{j}kr}. \tag{3.48}$$

Substituting into (3.47),

$$V_2 = \int_{-l/2}^{l/2} E_\theta(r = d, \theta = \pi/2)\,\mathrm{d}z = \mathrm{j}\eta\frac{I_1\, l^2}{4\pi k}\left(\frac{k^2}{d} - \frac{\mathrm{j}k}{d^2} - \frac{1}{d^3}\right)\mathrm{e}^{-\mathrm{j}kd}, \tag{3.49}$$

that yields

$$Z_{21} = \frac{V_2}{I_1}|_{I_2=0} = \mathrm{j}\eta\frac{l^2}{4\pi k}\left(\frac{k^2}{d} - \frac{\mathrm{j}k}{d^2} - \frac{1}{d^3}\right)\mathrm{e}^{-\mathrm{j}kd}. \tag{3.50}$$

Taking into account that the radiation resistance of a Hertzian dipole is given by [30, Sec. 4.2]

$$R_\mathrm{r} = \frac{2\pi}{3}\eta\left(\frac{l}{\lambda}\right)^2, \tag{3.51}$$

we can also express Z_{21} in terms of R_r:

$$Z_{21} = R_\mathrm{r}\,\psi(kd), \tag{3.52}$$

where ψ is defined as

$$\psi(x) = \frac{3}{2}\mathrm{j}\mathrm{e}^{-\mathrm{j}x}\left(\frac{1}{x} - \frac{\mathrm{j}}{x^2} - \frac{1}{x^3}\right). \tag{3.53}$$

Since a Hertzian dipole is infinitesimally thin, its self-impedance can be obtained by taking the limit of the mutual impedance between two identical dipoles as their separation $d \to 0$:

$$Z_{11} = \lim_{d\to 0} Z_{21} \implies \begin{cases}\mathfrak{R}(Z_{11}) = R_\mathrm{r},\\ \mathfrak{I}(Z_{11}) \to -\infty.\end{cases} \tag{3.54}$$

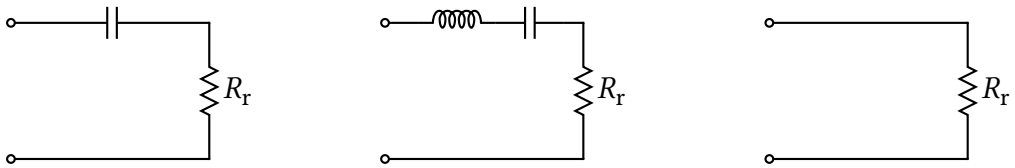
That is, the real part corresponds to the radiation resistance, while the imaginary part tends toward negative infinity, indicating an infinite capacitive reactance. Conceptually, this capacitive behavior can be compensated by connecting an inductor in series with the dipole, canceling out the reactance [81, Ch. 2]. The circuit models for the raw, matched and equivalent Hertzian dipole are shown in figure 3.6.

Assuming matched dipoles, the extension of the previous results to a N-antenna ULA is straightforward:

$$\mathbf{Z}_{\mathrm{A}\bullet} = R_\mathrm{r}\begin{pmatrix} 1 & \psi(kd) & \cdots & \psi(Nkd)\\ \psi(kd) & 1 & \cdots & \psi((N-1)kd)\\ \vdots & \vdots & \ddots & \vdots\\ \psi(Nkd) & \psi((N-1)kd) & \cdots & 1\end{pmatrix}, \tag{3.55}$$

or, compactly,

$$\begin{cases}[\mathbf{Z}_{\mathrm{A}\bullet}]_{m,n} = R_\mathrm{r},\ m = n,\ 1 \le m, n \le N,\\ [\mathbf{Z}_{\mathrm{A}\bullet}]_{m,n} = R_\mathrm{r}\,\psi(kd|m-n|),\ m \ne n.\end{cases} \tag{3.56}$$



(a) Raw Hertzian dipole. **(b)** Matched Hertzian dipole. **(c)** Equivalent circuit.

Figure 3.6: Circuit models of a Hertzian dipole.

Half-wavelength dipoles

The half-wavelength dipole, as its name suggests, is a linear antenna with a total length of $\lambda/2$, consisting of two conducting elements—typically wires—each $\lambda/4$ long, symmetrically arranged around a central feed point. In contrast to the Hertzian dipoles, half-wavelength dipoles not only can be manufactured but they are commonly used in practice [30, Sec. 4.6]. For this reason, their coupling characteristics have been extensively studied in the literature.

The self-impedance of a half-wavelength dipole can be computed numerically and its value is $73 + 42.5\mathrm{j}\,\Omega$. Similarly, the mutual impedance between two dipoles cannot be determined analytically and must also be evaluated numerically [30, Sec. 8.6]. Following the notation previously introduced, the mutual impedance between two side-by-side dipoles (*i.e.* figure 3.5b) is given by

$$\begin{aligned} \mathfrak{R}(Z_{21}) &= \frac{\eta}{4\pi}[2\mathrm{Ci}(u) - \mathrm{Ci}(v) - \mathrm{Ci}(w)], \\ \mathfrak{I}(Z_{21}) &= -\frac{\eta}{4\pi}[2\mathrm{Si}(u) - \mathrm{Si}(v) - \mathrm{Si}(w)], \end{aligned} \tag{3.57}$$

where Ci and Si are the cosine and sine integrals [30, Appendix III],

$$\begin{aligned} \mathrm{Ci}(x) &= -\int_x^{+\infty} \frac{\cos\tau}{\tau}\,\mathrm{d}\tau, \\ \mathrm{Si}(x) &= \int_0^x \frac{\sin\tau}{\tau}\,\mathrm{d}\tau, \end{aligned} \tag{3.58}$$

and u, v and w are auxiliary variables defined as

$$\begin{aligned} u &= kd, \\ v &= k\left(\sqrt{d^2 + l^2} + l\right), \\ w &= k\left(\sqrt{d^2 + l^2} - l\right), \end{aligned} \tag{3.59}$$

with $l = \lambda/2$.

As in the case of Hertzian dipoles, obtaining the impedance matrix for a ULA of N half-wavelength dipoles is straightforward. Letting

$$\begin{aligned} u_{m,n} &= kd|m-n|, \\ v_{m,n} &= k\left(\sqrt{d^2(m-n)^2 + l^2} + l\right), \\ w_{m,n} &= k\left(\sqrt{d^2(m-n)^2 + l^2} - l\right), \end{aligned} \tag{3.60}$$

the mutual impedance matrix of the array is given by

$$\begin{aligned}[\Re(\mathbf{Z}_{\mathrm{A}\bullet})]_{m,n} &= \frac{\eta}{4\pi}\big[2\mathrm{Ci}(u_{m,n}) - \mathrm{Ci}(v_{m,n}) - \mathrm{Ci}(w_{m,n})\big],\\ [\Im(\mathbf{Z}_{\mathrm{A}\bullet})]_{m,n} &= -\frac{\eta}{4\pi}\big[2\mathrm{Si}(u_{m,n}) - \mathrm{Si}(v_{m,n}) - \mathrm{Si}(w_{m,n})\big],\end{aligned} \tag{3.61}$$

for $m \neq n$, and

$$[\mathbf{Z}_{\mathrm{A}\bullet}]_{m,n} = 73.1 + 42.5\mathrm{j},\ m = n,\ 1 \leq m, n \leq N. \tag{3.62}$$

For instance, the mutual impedance matrix for three half-wavelength dipoles separated $d = \lambda$ can be computed numerically and is given by [36, Sec. 6.4]

$$\mathbf{Z}_{\mathrm{A}\bullet} = \begin{pmatrix} 73.1 + 42.5\mathrm{j} & 4.0 + 17.7\mathrm{j} & 1.1 + 9.4\mathrm{j} \\ 4.0 + 17.7\mathrm{j} & 73.1 + 42.5\mathrm{j} & 4.0 + 17.7\mathrm{j} \\ 1.1 + 9.4\mathrm{j} & 4.0 + 17.7\mathrm{j} & 73.1 + 42.5\mathrm{j} \end{pmatrix}. \tag{3.63}$$

To conclude this section, in figure 3.7 we depict the real part of the coupling function between two half-wavelength dipoles, together with the coupling functions for two Hertzian dipoles and two isotropic radiators (*i.e.* theoretical point sources radiating a spherical electric field). It can be observed that both Hertzian and half-wavelength dipoles exhibit practically the same mutual coupling effect, yet a scaling factor due to their different radiation resistance should be taken into account. On its turn, the coupling of two isotropic radiators follows the pattern of a cardinal sine [98, Eq. (7)], which is qualitatively very similar to that of the dipoles, as it was previously noted in [98].

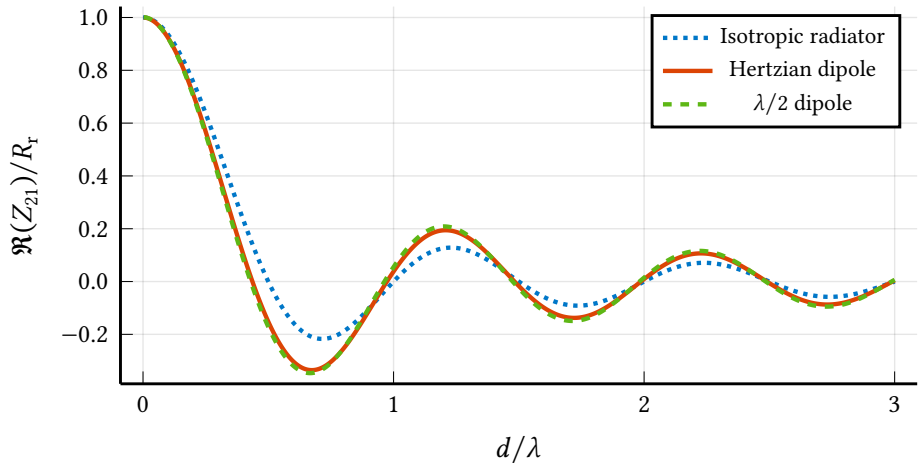


Figure 3.7: Mutual coupling between two isotropic radiators, Hertzian dipoles or half-wavelength dipoles at different distances.

3.2.2 Inter-array coupling impedance matrix

The inter-array coupling naturally depends on the characteristics of the propagation environment between the transmitter and the receiver [80]. Consequently, the channel models presented in chapter 2—or any alternative channel model in the literature—can be employed to determine the inter-array coupling matrix, with an appropriate scaling factor in ohms.

In order to model $\mathbf{Z}_{\text{ART}}$—and consequently $\mathbf{Z}_{\text{ATR}}$ due to reciprocity—we consider the most general case of a transmitting array with N_{T} antennas and a receiving array with N_{R} antennas. Following the geometry in section 2.2, transmitting array elements are located at positions $\mathbf{s}_m$, for $0 \leq m \leq N_{\text{T}} - 1$, and receiving array elements at positions $\mathbf{u}_n$, for $0 \leq n \leq N_{\text{R}} - 1$. The coordinate origin is chosen at $\mathbf{u}_0$. Naturally, the SIMO and MISO matrices can be obtained by setting $N_{\text{T}} = 1$ or $N_{\text{R}} = 1$, respectively.

Next, we review the main inter-array impedance matrices used throughout this thesis, considering both LoS and NLoS propagation conditions. Alternative models can be found, for example, in [79], [83], [89].

Free-space propagation

Under free-space propagation in the radiative field, $\mathbf{Z}_{\text{ART}}$ follows the model in (2.49) with its entries given by

$$[\mathbf{Z}_{\text{ART}}]_{m+1,n+1} = \Gamma g_{m,n} \mathrm{e}^{-\mathrm{j}k(r_{m,n}-r)}, \tag{3.64}$$

where Γ is an impedance scaling factor. Moreover, $\mathbf{Z}_{\text{ART}}$ must coincide with (3.56) for large d, up to an angle-dependent scaling factor. For simplicity, we set this factor equal to one.[1] Recall that for the specific geometry of side-by-side dipoles, this factor evaluates to 1.5.

Therefore, we have that

$$[\mathbf{Z}_{\text{ART}}]_{m+1,n+1} = \mathrm{j}R_{\text{r}} \frac{\mathrm{e}^{-\mathrm{j}kr_{m,n}}}{kr_{m,n}}, \tag{3.65}$$

or, in terms of the array response vectors,

$$\mathbf{Z}_{\text{ART}} = \mathrm{j}R_{\text{r}} \mathbf{a}_{\text{R}}(r,\theta,\phi)\mathbf{a}_{\text{T}}(r,\theta,\phi)^{\text{H}}, \tag{3.66}$$

which corresponds to the model in (2.59), with $g_{\text{R},n} = \frac{1}{k\|\mathbf{u}_n - \mathbf{s}_0\|}$ and $g_{\text{T},m} = \frac{1}{k\|\mathbf{s}_m\|}$.

Multipath propagation

The inter-array impedance matrix for a MIMO multipath propagation environment can be obtained from (2.60) as

$$\mathbf{Z}_{\text{ART}} = \sum_{l=1}^{L} \Gamma_l \mathbf{a}_{\text{R}}(r_l,\theta_l,\phi_l)\mathbf{a}_{\text{T}}(r_l,\theta_l,\phi_l)^{\text{H}}, \tag{3.67}$$

and, since scatterers are modeled as point sources, it follows that

$$\mathbf{Z}_{\text{ART}} = \mathrm{j}R_{\text{r}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\text{R}}(r_l,\theta_l,\phi_l)\mathbf{a}_{\text{T}}(r_l,\theta_l,\phi_l)^{\text{H}}. \tag{3.68}$$

Under rich scattering the CLT holds, and the correlated Rayleigh model

$$\text{vec}(\mathbf{Z}_{\text{ART}}) \sim \mathscr{CN}\left(\mathbf{0}_{N_{\text{R}}N_{\text{T}}}, \mathbf{C}_{\mathbf{Z}_{\text{ART}}}\right), \quad \mathbf{C}_{\mathbf{Z}_{\text{ART}}} = \text{E}\left[\text{vec}(\mathbf{Z}_{\text{ART}})\text{vec}(\mathbf{Z}_{\text{ART}})^{\text{H}}\right] \tag{3.69}$$

can be employed. This formulation also allows to use the classical uncorrelated Rayleigh fading model by setting $\mathbf{C}_{\mathbf{Z}_{\text{ART}}} = |\Gamma|^2 \mathbf{I}_{N_{\text{R}}N_{\text{T}}}$.

[1] The angle-dependent scaling factor can be absorbed into the SNR.

3.3 Analysis of receiver to transmitter coupling

In section 3.1.3 we introduced the unilateral approximation $\mathbf{Z}_{\mathrm{ATR}} \approx \mathbf{0}_{N_{\mathrm{T}} \times N_{\mathrm{R}}}$, which leads to significantly simpler expressions for the input-output relations of a coupled system (see table 3.1). However, this is a coarse approximation [19]. Actually, for the unilateral approximation to be valid—and therefore for the system to behave as if the network was unilateral—we do not need $\mathbf{Z}_{\mathrm{TR}} \approx \mathbf{0}_{N_{\mathrm{T}} \times N_{\mathrm{R}}}$ but only inter-array coupling to be much smaller than intra-array coupling.

Using the input impedance of the multiport network in (3.22) as a reference, the unilateral approximation is valid whenever $\mathbf{Z}_{\mathrm{in}} \approx \mathbf{Z}_{\mathrm{T}}$. This condition is satisfied if

$$\left\| \mathbf{Z}_{\mathrm{TR}} \left(Z_{\mathrm{L}} \mathbf{I}_{N_{\mathrm{R}}} + \mathbf{Z}_{\mathrm{R}} \right)^{-1} \mathbf{Z}_{\mathrm{RT}} \right\|_{\mathrm{F}} \ll \| \mathbf{Z}_{\mathrm{T}} \|_{\mathrm{F}}. \tag{3.70}$$

Similarly, by considering the output impedance, the condition may be written as

$$\left\| \mathbf{Z}_{\mathrm{RT}} \left(Z_{\mathrm{G}} \mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}} \right)^{-1} \mathbf{Z}_{\mathrm{TR}} \right\|_{\mathrm{F}} \ll \| \mathbf{Z}_{\mathrm{R}} \|_{\mathrm{F}}, \tag{3.71}$$

since in both cases $\mathbf{D} \approx \mathbf{D}_{\mathrm{UA}}$, as follows from (3.23).

3.3.1 System model

In this section of the thesis, we analyze the validity of the previous conditions under LoS propagation, considering two complementary cases:

1. Downlink: massive MISO with $N_{\mathrm{T}} = N$ and $N_{\mathrm{R}} = 1$.
2. Uplink: massive SIMO with $N_{\mathrm{T}} = 1$ and $N_{\mathrm{R}} = N$.

The array is assumed to be a ULA with aperture size D and radiating elements are Hertzian dipoles of length l with a separation d between them. Hertzian dipoles are used instead of half-wavelength dipoles in order to enable analytical tractability. To compensate for the capacitive reactance of the dipoles, loading inductors are assumed to be inserted between each antenna and its feedline, as previously illustrated in figure 3.6. Since N is assumed to be large and near-field operation is expected, matching networks are omitted at both the transmitter and the receiver.

In the scenario just described, the intra-array mutual impedance matrix is given by (3.56), whereas the inter-array coupling impedance matrix in (3.65) degenerates into a vector with components

$$[\mathbf{z}_{\mathrm{TR}}]_n = \mathrm{j} R_{\mathrm{r}} \frac{\mathrm{e}^{-\mathrm{j}kr_n}}{kr_n}, \tag{3.72}$$

where r_n is the distance between the single-antenna device and the n-th element in the array.

Next, we proceed to analyze the inter-array coupling in the MISO and SIMO scenarios just described. For ease of reading, we summarize the system parameters:

- Number of antennas: N.
- Array size: D.

- Inter-element spacing: d.
- Dipole length: l.
- Wavelength: λ.
- Distance between transmitter and receiver: r.
- Incidence angle: θ.
- Source and load impedances: Z_{G}, Z_{L}.

3.3.2 Massive MISO

In a MISO system, the unilateral approximation will be valid when condition (3.70) is fulfilled:

$$\left\| \frac{\mathbf{z}_{\mathrm{TR}} \mathbf{z}_{\mathrm{TR}}^{\mathrm{T}}}{Z_{\mathrm{L}} + R_{\mathrm{r}}} \right\|_{\mathrm{F}} \ll \|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}. \tag{3.73}$$

Since $\mathbf{z}_{\mathrm{TR}} \mathbf{z}_{\mathrm{TR}}^{\mathrm{T}}$ is rank-one, it is equivalent to

$$\frac{\|\mathbf{z}_{\mathrm{TR}}\|_2^2}{|Z_{\mathrm{L}} + R_{\mathrm{r}}|} \ll \|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}. \tag{3.74}$$

Assuming the intra-array coupling model in (3.56), we can further manipulate the right-hand side (RHS) in (3.74). In particular, we lower bound $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ as follows:

$$\begin{aligned} \|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}} = \sqrt{\sum_{m=1}^{N} \sum_{n=1}^{N} |[\mathbf{Z}_{\mathrm{T}}]_{m,n}|^2} &= \sqrt{\sum_{m=1}^{N} \left\| [\mathbf{Z}_{\mathrm{T}}]_{m,*} \right\|_2^2} \\ &\geq \sqrt{N} \cdot \min_{m} \left\| [\mathbf{Z}_{\mathrm{T}}]_{m,*} \right\|_2 = \sqrt{N} \cdot \left\| [\mathbf{Z}_{\mathrm{T}}]_{1,*} \right\|_2. \end{aligned} \tag{3.75}$$

In the last step, we used the fact that antennas at the corners of the ULA experience the smallest coupling. In fact, the last row could be used equivalently instead of the first one. It is also worth to remark that $\sqrt{N} \left\| [\mathbf{Z}_{\mathrm{T}}]_{1,*} \right\|_2 \to \|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ when $N \to \infty$, because $\mathbf{Z}_{\mathrm{T}}$ is Toeplitz [102]. Therefore, the bound in (3.75) is asymptotically tight.

We now proceed to compute the Euclidean norm of the first row:

$$\begin{aligned} \left\| [\mathbf{Z}_{\mathrm{T}}]_{1,*} \right\|_2^2 &= R_{\mathrm{r}}^2 + \sum_{n=1}^{N-1} R_{\mathrm{r}}^2 |\psi(kdn)|^2 = R_{\mathrm{r}}^2 + \frac{9}{4} R_{\mathrm{r}}^2 \sum_{n=1}^{N-1} \frac{1}{(kdn)^2} - \frac{1}{(kdn)^4} + \frac{1}{(kdn)^6} \\ &= R_{\mathrm{r}}^2 + \frac{9}{4} R_{\mathrm{r}}^2 \left(\frac{H_{N-1}^{(2)}}{(kd)^2} - \frac{H_{N-1}^{(4)}}{(kd)^4} + \frac{H_{N-1}^{(6)}}{(kd)^6} \right), \end{aligned} \tag{3.76}$$

where $H_n^{(p)}$ is the generalized harmonic number of order p [103, Sec. 1.2.7],

$$H_n^{(p)} = \sum_{m=1}^{n} \frac{1}{m^p}. \tag{3.77}$$

Multiplying by N we obtain the bound in (3.75):

$$\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}^2 \geq NR_{\mathrm{r}}^2 + \frac{9}{4}NR_{\mathrm{r}}^2\left(\frac{H_{N-1}^{(2)}}{(kd)^2} - \frac{H_{N-1}^{(4)}}{(kd)^4} + \frac{H_{N-1}^{(6)}}{(kd)^6}\right). \tag{3.78}$$

Substituting the square root of the RHS in (3.78) into the RHS of (3.74) results in a sufficient condition for the unilateral approximation to hold in MISO.

Now we analyze this condition in two limiting cases: fixed inter-element spacing and fixed array size.

Fixed inter-element spacing

If d is fixed, then $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ diverges as $O(\sqrt{N})$. Furthermore, it can be assumed that the receiver is located at a distance r meters and incidence angle $\theta = \pi/2$ to the first antenna in the transmitting array, as the asymptotic behavior does not depend on the location of the receiving antenna. Under these assumptions, the left-hand side (LHS) in (3.74) simplifies to

$$\|\mathbf{z}_{\mathrm{TR}}\|_2^2 = \frac{R_{\mathrm{r}}^2}{k^2}\sum_{n=0}^{N-1}\frac{1}{r_n^2} = \frac{R_{\mathrm{r}}^2}{k^2}\sum_{n=0}^{N-1}\frac{1}{r^2+(nd)^2}. \tag{3.79}$$

The limit when $N \to \infty$ of the sum above is convergent and can be computed using the Poisson summation formula [104, Sec. 7.3]:

$$\lim_{N\to+\infty}\frac{\|\mathbf{z}_{\mathrm{TR}}\|_2^2}{|Z_{\mathrm{L}}+R_{\mathrm{r}}|} = \frac{R_{\mathrm{r}}^2}{|Z_{\mathrm{L}}+R_{\mathrm{r}}|}\left(\frac{d + r\pi\coth\left(\pi\frac{r}{d}\right)}{2d(rk)^2}\right) < +\infty, \tag{3.80}$$

which means that (3.74) is true asymptotically. Therefore, in a massive MISO environment, the unilateral approximation is asymptotically valid even in the (radiative) near field when the inter-element spacing is kept constant.

Fixed array size

In order to keep the array size constant, we must let the inter-element distance decrease with the number of antennas: $d = \frac{D}{N-1}$. Substituting it in (3.78), we observe that $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ grows as $O(N^{3.5})$, compared with $O(\sqrt{N})$ from the previous scenario. Note that this is consistent with the fact that placing antennas closer together increases mutual coupling.

Besides, substituting d in (3.79), we obtain that each term of the series is given by

$$a_n = \frac{R_{\mathrm{r}}^2/k^2}{r^2+(\frac{nD}{N-1})^2} \geq \frac{R_{\mathrm{r}}^2/k^2}{r^2+(\frac{nD}{n})^2} = \frac{R_{\mathrm{r}}^2/k^2}{r^2+D^2}. \tag{3.81}$$

Since a_n is lower-bounded by a constant, it follows from the vanishing condition and the comparison test [31, Ch. 23] that $\|\mathbf{z}_{\mathrm{TR}}\|_2^2$ diverges as $O(N)$. As the RHS in (3.74) grows much faster than its LHS, we conclude that the unilateral approximation is also asymptotically valid when the array size is fixed.

3.3.3 Massive SIMO

From (3.71), the unilateral approximation condition in a SIMO system becomes

$$\frac{\|\mathbf{z}_{\mathrm{RT}}\|_2^2}{|Z_{\mathrm{G}} + R_{\mathrm{r}}|} \ll \|\mathbf{Z}_{\mathrm{R}}\|_{\mathrm{F}}, \tag{3.82}$$

which is equivalent to (3.74) derived for MISO.

If the system considered in section 3.3.2 was now employed as a SIMO system in the uplink, then the expressions derived for MISO (*i.e.* (3.74) to (3.80)) would be valid replacing $\mathbf{Z}_{\mathrm{T}}$ and Z_{L} by $\mathbf{Z}_{\mathrm{R}}$ and Z_{G}. This allows to extend the downlink results to the uplink, which means that the unilateral approximation is also asymptotically valid for a SIMO system operating in the radiative near field.

3.3.4 Numerical results

In this subsection, the theoretical results developed throughout section 3.3 are numerically illustrated in two representative scenarios with realistic parameters.

In the first scenario, we consider a ULA with a fixed inter-element spacing $d = \lambda/2$ placed at a distance $r = 55\,\mathrm{m}$ of the UE. This geometry constrains the maximum array aperture to $D_{\max} = 2\,\mathrm{m}$ in order to ensure operation within the Fresnel region.

In the second scenario, a ULA with a fixed aperture $D = 1\,\mathrm{m}$ is analyzed, with the separation between transmitter and receiver equal to the Fresnel distance $r = 0.62\sqrt{D^3/\lambda} \approx 20\,\mathrm{m}$. Inter-element spacing is equal to $d = D/(N-1)$, so the array aperture is kept constant.

In both cases, the system operates at $\lambda = 1\,\mathrm{mm}$ and the dipole length is $l = \lambda/20$. The source and load impedances $Z_{\mathrm{G}} = Z_{\mathrm{L}} = 186 - 31.6\mathrm{j}\,\Omega$ are taken from experimental measurements in [105] and have been recently adopted in [83]. A summary of this setup is given in table 3.2. For the sake of clarity, in this section we employ the notation for MISO systems; however, as established in section 3.3.3, the results remain valid for SIMO systems thanks to the equivalence between both link directions.

Table 3.2: Simulation parameters for testing the unilateral approximation.

Parameter	Scenario 1	Scenario 2
N	$[10, 2000]$	$[10, 10^6]$
D	$[4.5\,\mathrm{mm}, 2\,\mathrm{m}]$	$1\,\mathrm{m}$
d	$\lambda/2$	$D/(N-1)$
l	$\lambda/20$	$\lambda/20$
λ	$1\,\mathrm{mm}$	$1\,\mathrm{mm}$
r	$55\,\mathrm{m}$	$20\,\mathrm{m}$
θ	$\pi/2\,\mathrm{rad}$	$\pi/2\,\mathrm{rad}$
Z_{G}, Z_{L}	$186 - 31.6\mathrm{j}\,\Omega$	$186 - 31.6\mathrm{j}\,\Omega$

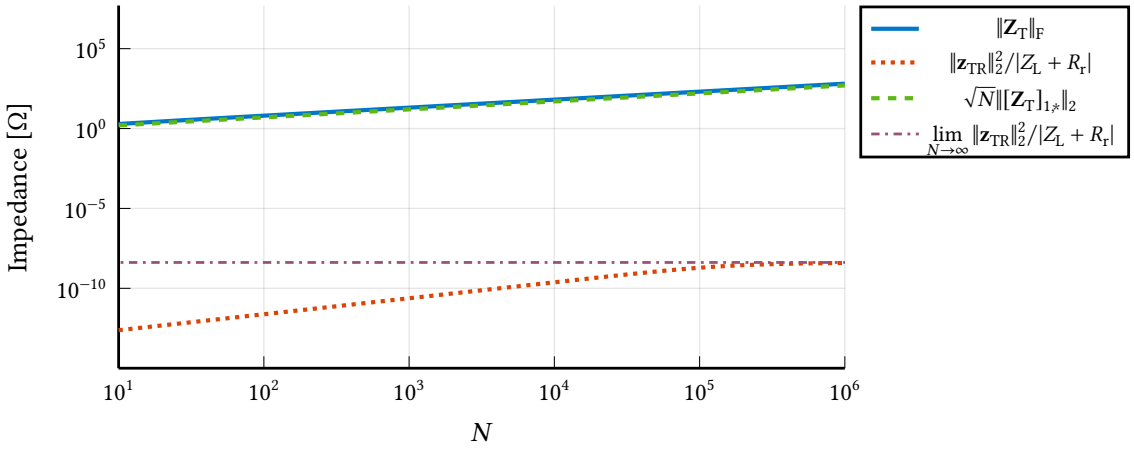


Figure 3.8: Fixed inter-element spacing $d = \lambda/2$ coupling condition as a function of N with $r = 55$ m.

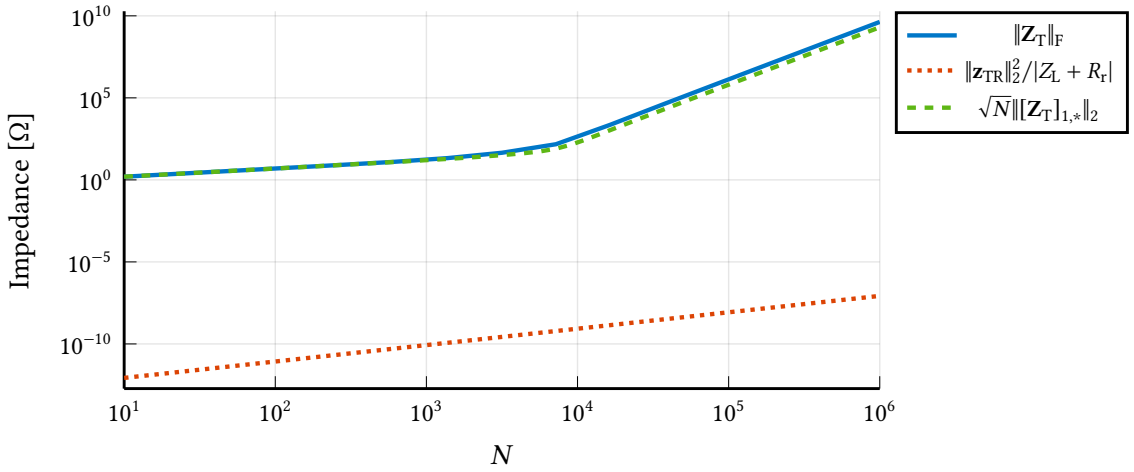


Figure 3.9: Fixed array size $D = 1$ m coupling condition as a function of N with $r \approx 20$ m (*i.e.* the Fresnel distance).

In figure 3.8 the results for the first scenario are shown. We depict $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ and its lower bound from (3.75), as well as (3.79) and (3.80). It can be observed that, for this particular setup, the unilateral approximation holds for any practical number of antennas, as $\|\mathbf{z}_{\mathrm{TR}}\|_2^2/|Z_{\mathrm{L}}+R_{\mathrm{r}}|$ is always much smaller (at least 10 times) than $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$. It should be noted that the results for $N > 2000$ are not valid for our model, as the system would be operating in the inductive near field, but they are drawn to illustrate the convergence of (3.79) into (3.80).

In figure 3.9, we plot the results for the second simulation scenario. As expected, $\|\mathbf{z}_{\mathrm{TR}}\|_2^2/|Z_{\mathrm{L}}+R_{\mathrm{r}}|$ grows without bounds, but $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ grows much faster, thus ensuring the asymptotic validity of the unilateral approximation. Furthermore, for all N it is fulfilled $\|\mathbf{z}_{\mathrm{TR}}\|_2^2/|Z_{\mathrm{L}}+R_{\mathrm{r}}| \ll \|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$. Therefore, inter-array coupling can also be neglected in this scenario.

Finally, we shall remark that the lower bound for $\|\mathbf{Z}_{\mathrm{T}}\|_{\mathrm{F}}$ derived in (3.75) is tight in both scenarios.

3.4 Relation between information-theoretic and physical models

In order to encode the physical information into the channel matrix we need to associate physical quantities to information-theoretic variables $\mathbf{x}$ and $\mathbf{y}$ from (3.1). Following the circuital approach, it is natural to relate the channel input with $\mathbf{v}_\mathrm{G}$ and the channel output with $\mathbf{v}_\mathrm{L}$ [80], [83]. Letting c be an arbitrary constant measured in V^2,

$$\frac{\mathbf{v}_\mathrm{L}}{\sqrt{c}} = \mathbf{D}\frac{\mathbf{v}_\mathrm{G}}{\sqrt{c}} + \frac{\mathbf{n}}{\sqrt{c}}. \tag{3.83}$$

Since each term is now dimensionless, they can be identified with the corresponding information-theoretic variables.

To ensure consistency between the two models, both the transmitted power and the noise terms must coincide. Consequently, we impose

$$P_\mathrm{T} = \frac{\mathrm{E}\left[\mathbf{v}_\mathrm{G}^\mathrm{H}\mathbf{B}\mathbf{v}_\mathrm{G}\right]}{R_\mathrm{G}} = c \cdot \mathrm{E}\left[\mathbf{x}^\mathrm{H}\mathbf{x}\right], \tag{3.84}$$

$$\mathbf{z} = \frac{\mathbf{n}}{\sqrt{c}}, \tag{3.85}$$

where $\mathbf{B}$ is the power coupling matrix defined in (3.34).

From the power constraint in (3.84), we choose the input vector as

$$\mathbf{x} = \frac{\mathbf{B}^{1/2}\mathbf{v}_\mathrm{G}}{\sqrt{c \cdot R_\mathrm{G}}}. \tag{3.86}$$

Substituting the signal and noise vectors (3.85) and (3.86) into (3.1) yields

$$\mathbf{y} = \mathbf{H}\frac{\mathbf{B}^{1/2}\mathbf{v}_\mathrm{G}}{\sqrt{c \cdot R_\mathrm{G}}} + \frac{\mathbf{n}}{\sqrt{c}} = \frac{\mathbf{v}_\mathrm{L}}{\sqrt{c}}. \tag{3.87}$$

Finally, comparing this expression with (3.83) implies that the channel matrix must satisfy

$$\mathbf{H} = \sqrt{R_\mathrm{G}}\mathbf{D}\mathbf{B}^{-1/2}, \tag{3.88}$$

so that (3.1) and (3.83) are equivalent.

In particular, under optimal power matching at the transmitter, $\mathbf{B}$ becomes proportional to the identity matrix, so $\mathbf{x}$ is proportional to $\mathbf{v}_\mathrm{G}$ and $\mathbf{H}$ is proportional to $\mathbf{D}$.

Recalling the expressions in table 3.1, in the general case $\mathbf{B}$ is a random matrix, and $\mathbf{D}$ is given by the product of several random matrices, which leads to a non-Gaussian distribution of the channel even if inter-array coupling is assumed to be Gaussian. Furthermore, the distribution of $\mathbf{x}$ is also affected by $\mathbf{B}^{1/2}$, which can result in signaling strategies being physically inconsistent.[2]

[2] It should be noted that $\mathbf{B}^{-1/2}$ from the channel matrix and $\mathbf{B}^{1/2}$ from the symbol vector cancel each other, so such strategies can still be interpreted at the level of the physical channel.

Fortunately, under the unilateral approximation, $\mathbf{Q}$, $\mathbf{B}$ and $\mathbf{Z}_{\mathrm{T}}$ are deterministic, so

$$\mathbf{H} = \sqrt{R_{\mathrm{G}}}\mathbf{Q}\mathbf{Z}_{\mathrm{RT}}(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}})^{-1}\mathbf{B}^{-1/2} \tag{3.89}$$

and $\mathbf{H}$ follows the same distribution as $\mathbf{Z}_{\mathrm{RT}}$, up to a linear transformation. Notice that the presence or absence of matching networks does not affect the distribution of either $\mathbf{Z}_{\mathrm{RT}}$ or $\mathbf{H}$, since the matching networks are deterministic.

For simplicity, we will now assume that no matching networks are present at either the transmitter or the receiver, without detriment to the generality of the result. In this case, $\mathbf{Z}_{\mathrm{R}}$, $\mathbf{Z}_{\mathrm{T}}$ and $\mathbf{Z}_{\mathrm{RT}}$ can be directly obtained following the approaches in sections 3.2.1 and 3.2.2.

Assuming a multipath propagation scenario, it is fair to consider a Rayleigh fading model for $\mathbf{Z}_{\mathrm{RT}}$. Consequently, $\mathbf{H}$ is also a normally distributed random variable with covariance

$$\mathbf{C}_{\mathbf{H}} = \Big(\mathbf{B}^{-\mathrm{T}/2}(Z_{\mathrm{G}}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}})^{-1} \otimes \mathbf{Q}\Big)\mathbf{C}_{\mathbf{Z}_{\mathrm{RT}}}\Big(\mathbf{B}^{-\mathrm{T}/2}(Z_{\mathrm{G}}^{*}\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{Z}_{\mathrm{T}}^{*})^{-1} \otimes \mathbf{Q}^{*}\Big), \tag{3.90}$$

where we have used that $\mathbf{B}$ is Hermitian, $\mathbf{Z}_{\mathrm{T}}$ and $\mathbf{Q}$ are symmetric, and where $\mathbf{C}_{\mathbf{Z}_{\mathrm{RT}}}$ denotes the covariance of the inter-array coupling matrix defined in (3.69).

Equation (3.90) clearly illustrates the combined effect of mutual coupling and spatial correlation on the communication channel. The terms in parentheses, namely $\mathbf{Z}_{\mathrm{T}}$, $\mathbf{Q}$ and $\mathbf{B}$, account for mutual coupling and reduce to a multiple of the identity matrix in its abscence. Conversely, $\mathbf{C}_{\mathbf{Z}_{\mathrm{RT}}}$ encapsulates spatial correlation and becomes a multiple of the identity matrix under rich scattering in the far field, as discussed in section 2.2.

The conclusion is that $\mathbf{H}$ follows a correlated Rayleigh fading model. Even if the propagation environment itself is assumed to produce spatially uncorrelated fading, the presence of mutual coupling among antennas induces correlation, thereby modifying the statistical structure of the channel. Note, however, that this result relies on the validity of the unilateral approximation. Fortunately, it holds even in the near field of a SIMO and MISO systems, ensuring that the model retains practical relevance in such scenarios.

CHAPTER 4

Energy-based noncoherent communications

With the advent of massive antenna arrays and the use of ever-higher frequency bands, the acquisition of accurate CSI has become increasingly challenging due to factors such as pilot overhead, hardware impairments, and rapid channel variations in high-mobility scenarios. As a result, there has been a renewed interest in noncoherent communication schemes, which do not require instantaneous CSI at the transmitter or receiver and can thus offer a more scalable and robust alternative in scenarios where conventional channel estimation becomes impractical [3].

As discussed in section 2.4, there exist several approaches that enable reliable communication without explicit channel knowledge, most notably subspace detection, differential encoding, and energy detection methods. Among these, energy-based schemes are particularly appealing in scenarios with a very large number of antennas or under stringent low-latency requirements. A representative use case is IIoT, envisioned as an evolution of the well-established IoT, but tailored for mission-critical applications in industrial environments [106], [107]. Wireless networks in IIoT scenarios are typically characterized by a high density of low-power/low-complexity terminals [113], the majority of which remain inactive most of the time. The information being transmitted through these networks mainly consists in control and telemetry data collected by machine sensors. Therefore, while the volume of traffic is low, its transmission must be highly reliable and meet stringent latency requirements. To accomplish these challenging requirements, IIoT leverages two operation modes of 5G: mMTC and URLLC [13].

The design of noncoherent systems for URLLC applications has therefore attracted considerable attention. Many works focus on single-antenna transmitters [108]–[111], and several propose schemes based on energy detection [12], [42], [49], [72], [112], given their simplicity and suitability for low-complexity devices. Nevertheless, the vast majority of studies on energy-based noncoherent detection adopt the isotropic Rayleigh fading model, an assumption that—as shown in chapters 2 and 3—is not physically consistent when the electromagnetic properties of antenna arrays and mutual coupling are properly taken into account. A notable exception is [42], where the maximum likelihood (ML) detector is analyzed under correlated fading and an optimized constellation design is developed.

The extremely low-latency constraints typically imposed in IIoT scenarios have two major consequences. First, only statistical CSIR is assumed to be available; that is, the receiver has full knowledge of the channel statistics but is unaware of the specific channel realizations. Second, information is decoded on a symbol-by-symbol basis, implying that communication is *one-shot* [121].

Based on these considerations, in this chapter we analyze the fundamental limitations of one-shot SIMO noncoherent communications under correlated Rayleigh fading. This assumption is sustained by section 3.3, where we have shown that even in the near field it is possible to consider such model. Furthermore, we propose a quadratic framework that generalizes the energy detector (ED) commonly used in the literature and we derive an analytic expression for its symbol error rate (SER) [20].

4.1 Signal model

Consider a narrowband massive SIMO architecture with a single-antenna transmitter and a receiver BS equipped with N antennas. The communication is one-shot and performed through a fast fading channel, modeled as a random variable $\mathbf{h}$ that remains constant for a single channel use and changes into an independent realization in the next one. The transmitter sends an equiprobable symbol $x \in \mathbb{C}$ selected from a M-ary constellation $\mathcal{X} = \{x_1, \dots, x_M\}$, for $M \geq 2$. An average transmitted power constraint is assumed (*i.e.* $\mathrm{E}_{\mathsf{x}}\big[|\mathsf{x}|^2\big] = 1$).

The signal at the receiver is expressed using a complex baseband representation:

$$\mathbf{y} = \mathbf{h}\mathsf{x} + \mathbf{z}, \quad \mathbf{y}, \mathbf{h}, \mathbf{z} \in \mathbb{C}^N, \tag{4.1}$$

being $\mathbf{z}$ an additive Gaussian noise component. The receiver has full statistical CSI,[1] *i.e.* it is aware of the distributions of both $\mathbf{h}$ and $\mathbf{z}$, but not their realizations. The transmitter is completely unaware of the channel state, *i.e.* there is no CSIT. The fading is assumed correlated Rayleigh: $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{h})$. This is consistent with general, well-established channel models [29, Ch. 3], as well as state-of-the-art ones, such as those described in chapters 2 and 3. Similarly, the noise is distributed as $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{z})$ with arbitrary correlation, which accounts for both colored noise and multiuser interference. The average SNR at the receiver is defined as follows:

$$\alpha = \frac{\mathrm{E}_{\mathsf{x},\mathbf{h}}\big[|\mathsf{x}|^2\|\mathbf{h}\|^2\big]}{\mathrm{E}_\mathbf{z}\big[\|\mathbf{z}\|^2\big]} = \frac{\mathrm{E}_{\mathsf{x}}\big[|\mathsf{x}|^2\big]\mathrm{tr}\big(\mathrm{E}_\mathbf{h}\big[\mathbf{h}\mathbf{h}^{\mathrm{H}}\big]\big)}{\mathrm{tr}\big(\mathrm{E}_\mathbf{z}\big[\mathbf{z}\mathbf{z}^{\mathrm{H}}\big]\big)} = \frac{\mathrm{tr}(\mathbf{C}_\mathbf{h})}{\mathrm{tr}(\mathbf{C}_\mathbf{z})}, \tag{4.2}$$

which has been derived using independence between $\mathbf{z}$, $\mathbf{h}$ and x, the cyclic property of the trace and linearity of the expectation.

4.2 Maximum likelihood detector

For a constellation $\mathcal{X}$ with equiprobable symbols, the error probability is defined as

$$P_\epsilon = \frac{1}{M}\sum_{x\in\mathcal{X}} \mathrm{P}(\hat{x}(\mathbf{y}) \neq x \,|\, \mathsf{x} = x), \tag{4.3}$$

[1] Acquisition of second-order statistical properties of both channel and noise is a prominent research topic in communications. Refer, for example, to [114]–[117] and references therein for a variety of methods and techniques.

where $\hat{x}(\mathbf{y})$ is the output of a symbol detector applied to the received signal $\mathbf{y}$. The receiver that minimizes P_ϵ is the ML detector [67, Sec. 4.1-1].

The likelihood function of the received signal (4.1) for a transmitted symbol and given a channel realization is

$$f_{\mathbf{y}|x,\mathbf{h}}(\mathrm{y}) = \frac{\exp(-(\mathrm{y}-\mathbf{h}x)^{\mathrm{H}}\mathbf{C}_{\mathbf{z}}^{-1}(\mathrm{y}-\mathbf{h}x))}{\pi^N|\mathbf{C}_{\mathbf{z}}|}. \tag{4.4}$$

The channel realization is unknown at the receiver and is removed by marginalizing the previous function (*i.e.* unconditional model [118]). This results in the following likelihood function:

$$f_{\mathbf{y}|x}(\mathrm{y}) = \mathrm{E}_{\mathbf{h}}\Big[f_{\mathbf{y}|x,\mathbf{h}}(\mathrm{y})\Big] = \frac{\exp(-\mathrm{y}^{\mathrm{H}}\mathbf{C}_{\mathbf{y}|x}^{-1}\mathrm{y})}{\pi^N|\mathbf{C}_{\mathbf{y}|x}|}, \tag{4.5}$$

where $\mathbf{C}_{\mathbf{y}|x} = |x|^2\mathbf{C}_{\mathbf{h}} + \mathbf{C}_{\mathbf{z}}$ is the covariance matrix of the received signal for a given x.

The ML detector is obtained by maximizing the likelihood function over all possible symbols in $\mathscr{X}$:

$$\hat{x}_{\mathrm{ML}} = \arg\max_{x\in\mathscr{X}} f_{\mathbf{y}|x}(\mathrm{y}) = \arg\max_{x\in\mathscr{X}} \ln f_{\mathbf{y}|x}(\mathrm{y}) = \arg\min_{x\in\mathscr{X}} \mathrm{y}^{\mathrm{H}}\mathbf{C}_{\mathbf{y}|x}^{-1}\mathrm{y} + \ln|\mathbf{C}_{\mathbf{y}|x}|. \tag{4.6}$$

It is clear from (4.5) and (4.6) that the phase information of x cannot be retrieved from $\mathbf{y}$, since the noncoherent ML detector only perceives its energy $\varepsilon_i = |x_i|^2$. For this reason, $\mathscr{X}$ will be constructed from a unipolar PAM:

$$\mathscr{X} = \{\sqrt{\varepsilon_1} = 0 < \sqrt{\varepsilon_2} < \cdots < \sqrt{\varepsilon_M}\}. \tag{4.7}$$

In various related scenarios, it has been proven that the capacity-achieving input distribution is composed of a finite number of discrete mass points, one of which is found at the origin [119], [120]. Hence, we set the first symbol in our constellation at 0, in the same manner as in most other works on the topic [42], [49], [72], [121].

We might express

$$\mathbf{C}_{\mathbf{y}|x} = \mathbf{C}_{\mathbf{z}}^{\frac{1}{2}}(|x|^2\mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}}\mathbf{C}_{\mathbf{h}}\mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}} + \mathbf{I}_N)\mathbf{C}_{\mathbf{z}}^{\frac{1}{2}}, \tag{4.8}$$

and then eigendecompose $\mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}}\mathbf{C}_{\mathbf{h}}\mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}}$ as $\mathbf{U}\boldsymbol{\Gamma}\mathbf{U}^{\mathrm{H}}$. The diagonal matrix $\boldsymbol{\Gamma}$ contains the spectrum $\{\gamma_n\}_{1\le n\le N}$ and $\mathbf{U}$ is its eigenbasis. Without loss of generality, both $\mathbf{C}_{\mathbf{h}}$ and $\mathbf{C}_{\mathbf{z}}$ are assumed full-rank, since otherwise the analysis could be restricted to a lower-dimensional subspace. Moreover, if $\mathbf{C}_{\mathbf{z}}$ were rank-deficient, information could be transmitted noise-free through its null space. With these transformations, we can define

$$\mathbf{r} = \mathbf{U}^{\mathrm{H}}\mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}}\mathbf{y} \Longrightarrow \mathbf{r}|x \sim \mathscr{CN}\big(\mathbf{0}_N, \mathbf{C}_{\mathbf{r}|x}\big), \tag{4.9}$$

with $\mathbf{C}_{\mathbf{r}|x} = |x|^2\boldsymbol{\Gamma} + \mathbf{I}_N$. It is obtained by noise whitening and subsequent decorrelation of the received signal $\mathbf{y}$. Notice how there is a one-to-one correspondence between $\mathbf{y}$ and $\mathbf{r}$, thus the ML detector (4.6) can be equivalently expressed as

$$\hat{x}_{\mathrm{ML}} = \arg\min_{x\in\mathscr{X}} \mathbf{r}^{\mathrm{H}}\mathbf{C}_{\mathbf{r}|x}^{-1}\mathbf{r} + \ln\big|\mathbf{C}_{\mathbf{r}|x}\big| = \arg\min_{x\in\mathscr{X}} \sum_{n=1}^{N} \frac{|r_n|^2}{|x|^2\gamma_n + 1} + \ln\big(|x|^2\gamma_n + 1\big), \tag{4.10}$$

with $r_n = [\mathbf{r}]_n$. This detector coincides with the one presented in [42]. From an algebraic perspective, it is more convenient to treat (4.10) rather than (4.6) in subsequent derivations, since $\mathbf{C}_{\mathbf{r}|x}$ is diagonal. Therefore, we will refer to $\mathbf{r}|x$ instead of $\mathbf{y}|x$ without loss of generality, even though the analysis is valid for a receiver operating directly on $\mathbf{y}|x$.

Complexity of detection using (4.10) is $O(MN)$, with an additional preprocessing cost of $O(N^2)$. Although the quadratic term constitutes the main bottleneck, the linear term can also be significant in moderate antenna regimes when large constellations are employed.

4.3 Asymptotic regimes

It is of theoretical interest to analyze the performance of the presented communication system in various asymptotic regimes, namely as the number of receiving antennas or SNR grow without bound. Let

$$P_{a\to b} = \mathrm{P}(\hat{x} = x_b \,|\, \mathsf{x} = x_a) = \mathrm{P}\Big(f_{\mathbf{r}|x_a}(\mathbf{r}) \le f_{\mathbf{r}|x_b}(\mathbf{r}) \,|\, \mathsf{x} = x_a\Big) = \mathrm{P}\big(\mathrm{L}_{a,b}(\mathbf{r}) \le 0 \,|\, \mathsf{x} = x_a\big) \tag{4.11}$$

be the pairwise error probability (PEP) associated with transmitting x_a and detecting x_b at the receiver, in which

$$\mathrm{L}_{a,b}(\mathbf{r}) = \ln \frac{f_{\mathbf{r}|x_a}(\mathbf{r})}{f_{\mathbf{r}|x_b}(\mathbf{r})} = \mathbf{r}^{\mathrm{H}}\Big(\mathbf{C}^{-1}_{\mathbf{r}|x_b} - \mathbf{C}^{-1}_{\mathbf{r}|x_a}\Big)\mathbf{r} + \ln \frac{|\mathbf{C}_{\mathbf{r}|x_b}|}{|\mathbf{C}_{\mathbf{r}|x_a}|} \tag{4.12}$$

is the log-likelihood ratio (LLR) [122, Ch. 3] between hypotheses a and b. With the maximum PEP of the constellation, we can bound the error probability presented in (4.3), as stated in [67, Sec. 4.2-3]:

$$\max_{x_a \neq x_b \in \mathcal{X}} \frac{1}{M} P_{a\to b} \le P_\epsilon \le \max_{x_a \neq x_b \in \mathcal{X}} (M-1) P_{a\to b}. \tag{4.13}$$

This implies that the error probability will vanish if and only if the maximum PEP does as well:

$$\begin{aligned} \lim_{N\to\infty} \max_{x_a \neq x_b \in \mathcal{X}} P_{a\to b} = 0 \iff \lim_{N\to\infty} P_\epsilon = 0 \\ \lim_{\alpha\to\infty} \max_{x_a \neq x_b \in \mathcal{X}} P_{a\to b} = 0 \iff \lim_{\alpha\to\infty} P_\epsilon = 0. \end{aligned} \tag{4.14}$$

We now state two fundamental results regarding the performance of the presented system under asymptotic regimes.

Theorem 4.1. Let L be a positive constant. For a spectral matrix $\mathbf{\Gamma}$ with monotonically decreasing entries satisfying $0 < \gamma_n \le L$, the error probability of a constellation $\mathcal{X}$ vanishes as $N \to \infty$ if and only if

$$|x_a|^2 \neq |x_b|^2 \iff x_a \neq x_b,\ \forall x_a, x_b \in \mathcal{X} \qquad \text{and} \qquad \lim_{N\to\infty} \mathrm{tr}\big(\mathbf{\Gamma}^2\big) = \infty. \tag{4.15}$$

Proof. See appendix C.1. □

Theorem 4.2. For a finite number of receiving antennas, the error probability of a constellation of type (4.7) vanishes for increasing SNR if and only if $M = 2$.

Proof. See appendix C.2. □

What these results imply is that communication with the system considered herein is asymptotically error-free for an increasing number of receiving antennas but not for increasing SNR. This behavior for an amplitude-shift keying (ASK) constellation is illustrated in figure 4.1.

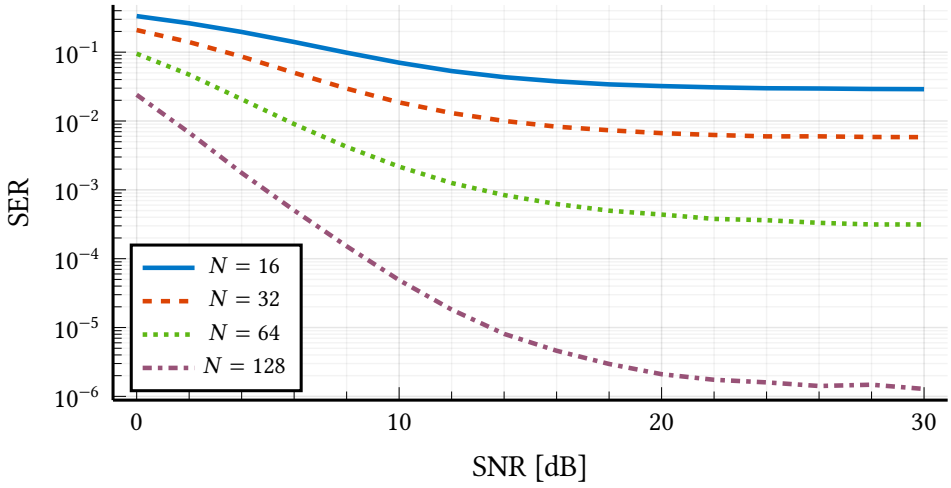


Figure 4.1: Monte Carlo results of the SER for the ML detector in terms of SNR. Various numbers of antennas have been considered under an exponentially correlated Rayleigh channel with $\rho = 0.8$ and a uniform unipolar 4-ASK modulation (see section 2.2.2 for a detailed channel model description).

4.4 Isotropic channel and high SNR approximation

There is a special case of the ML problem presented in (4.6) which has been widely studied in the literature [12], [121]: the isotropic channel with white noise. Under this model, the spectral matrix $\mathbf{\Gamma}$ is proportional to the identity (*i.e.* $\mathbf{\Gamma} = \alpha \mathbf{I}_N$). This assumption greatly simplifies (4.10), since $\mathbf{r}|x \sim \mathcal{CN}(\mathbf{0}_N, (|x|^2\alpha + 1)\mathbf{I}_N)$:

$$\hat{x}_{\mathrm{ML}}^{(\mathrm{iso})} = \arg\min_{x \in \mathcal{X}} \frac{\|\mathbf{r}\|^2}{|x|^2\alpha + 1} + N \ln\big(|x|^2\alpha + 1\big). \tag{4.16}$$

In this scenario, $\|\mathbf{r}\|^2$ is a sufficient statistic [121]. Furthermore, the computational complexity reduces from $O(N^2) + O(MN)$ to $O(N) + O(M)$. Our objective is therefore to implement a near-ML detector, in the sense of error probability, with reduced computational complexity.

Another relevant simplification of the ML detector that results in a similar expression emerges at high SNR (HSNR). That is, when $\alpha \to \infty$. Given a constellation $\mathcal{X}^*$ that does not include the null symbol $x = 0$, the covariance matrix of the received signal reduces to

$$\lim_{\alpha \to \infty} \mathbf{C}_{\mathbf{r}|x} = \lim_{\alpha \to \infty} |x|^2 \mathbf{\Gamma}, \tag{4.17}$$

with which detector (4.10) simplifies to

$$\lim_{\alpha \to \infty} \hat{x}_{\mathrm{ML}} = \lim_{\alpha \to \infty} \arg\min_{x \in \mathcal{X}^*} \frac{\mathbf{r}^{\mathrm{H}} \mathbf{\Gamma}^{-1} \mathbf{r}}{|x|^2} + N \ln\big(|x|^2\big) \tag{4.18}$$

and $\mathbf{r}^{\mathrm{H}} \mathbf{\Gamma}^{-1} \mathbf{r}$ becomes a sufficient statistic.

The previous result is no longer valid for the constellation considered in (4.7), as $\mathbf{C}_{\mathbf{r}|x=0} = \mathbf{I}_N$ even when $\alpha \to \infty$, so this symbol has to be treated separately to perform ML detection. Specifically, (4.10) should be evaluated for $x = 0$, while (4.18) applies to the remaining $M - 1$ symbols. The symbol yielding the minimum value among all candidates corresponds to the ML decision of the transmitted symbol.

In both special cases (4.16) and (4.18), the ML decoder can be decoupled into a two-stage process:

1. The computation of a continuous quadratic statistic.
2. A one-dimensional decision problem that detects the transmitted symbol solely using the proposed statistic.

Motivated by this observation, a generalization of this two-step approach for arbitrary channel and noise spectra is proposed in the sequel. Sections 4.5 and 4.6 address the design of the first and second steps of this procedure.

4.5 Energy statistic

The first simplified receiver described in section 4.4, known as ED, is widely used in the literature due to its low complexity and optimality within the isotropic channel. Nevertheless, when the channel is arbitrarily correlated, the ED does not fully exploit the statistical CSI and is no longer optimal. Therefore, in this chapter, we present a family of one-dimensional detectors that incorporate channel correlation. We have chosen a quadratic structure (in data) that generalizes the one for ED and is adequate when dealing with variables related to second-order moments (such as the transmitted energy $|x|^2$, in which we convey information):

$$\hat{\varepsilon}(\mathbf{r}) = \mathbf{r}^{\mathrm{H}}\mathbf{A}\mathbf{r} + c, \tag{4.19}$$

where $\mathbf{A} \in \mathbb{C}^{N\times N}$ denotes the estimator matrix to be designed. Since $\hat{\varepsilon}(\mathbf{r}) \geq 0$, matrix $\mathbf{A}$ must be Hermitian and positive semi-definite. Furthermore, the decorrelation step in (4.9) restricts $\mathbf{A}$ to be diagonal with real and nonnegative entries. The linear term is omitted, as it is not relevant in this setting due to the fact that $\mathrm{E}[\mathbf{r}|x] = \mathbf{0}_N$, whereas the affine term c is included to ensure an unbiased estimate of ε. It should be noted, however, that from a symbol-detection perspective, the constant term c does not affect the decision process.

The detection procedure is then split into two phases:

1. Perform the energy estimation with (4.19). In this step, ε is assumed to be continuous.
2. Detect the transmitted symbol by classifying the estimate according to some detection regions.

The corresponding block diagram is shown in figure 4.2.

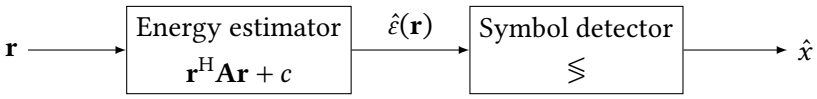


Figure 4.2: Block diagram of the energy detection procedure.

Throughout this section, we consider the received signal from (4.9), $\mathbf{r}|\varepsilon \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_{\mathbf{r}|\varepsilon})$, with transmitted energy $\varepsilon = |x|^2$ and $\mathbf{C}_{\mathbf{r}|\varepsilon} = \varepsilon\boldsymbol{\Gamma} + \mathbf{I}_N$. In addition, ε is treated as continuous variable during the energy estimation stage. Although the optimal scheme in related scenarios [119], [120] as well as practical implementations rely on discrete constellations, decoupling the detector into estimation and detection stages simplifies receiver design. The system is initially designed ignoring the discrete nature of ε, which is later introduced in section 4.6 via threshold-based classification.

4.5.1 Information-theoretic design criteria

The detector based on (4.19) is suboptimal for most scenarios because $\hat{\varepsilon}$ is not a sufficient statistic in the general case. Nevertheless, we wish to design the coefficients of the quadratic estimator so that it preserves as much information as possible on the transmitted symbol, thereby aiming to mitigate the error rate performance penalty. Hence, the mutual information arises as a natural choice for the design criterion for the coefficients of the quadratic estimator.

We propose to design coefficients $\mathbf{A}$ and c to maximize the mutual information (MI) between the transmitted symbol and the estimator output, *i.e.*

$$\mathrm{I}(\varepsilon;\hat{\varepsilon}) = \mathrm{h}(\varepsilon) - \mathrm{h}(\varepsilon|\hat{\varepsilon}) \implies \hat{\varepsilon}_{\mathrm{OPT}} = \arg\max_{\hat{\varepsilon}\in\mathcal{Q}} \mathrm{I}(\varepsilon;\hat{\varepsilon}), \tag{4.20}$$

where $\mathcal{Q}$ is the set of all estimators of the form (4.19). This problem is, in general, very hard to solve analytically. Motivated by [123], we obtain a lower bound on the MI:

$$\mathrm{I}(\varepsilon;\hat{\varepsilon}) = \mathrm{h}(\varepsilon) - \mathrm{h}(\varepsilon - \hat{\varepsilon}|\hat{\varepsilon}) \tag{4.21}$$

$$\geq \mathrm{h}(\varepsilon) - \mathrm{h}(\varepsilon - \hat{\varepsilon}). \tag{4.22}$$

Equality (4.21) is due to the translation invariance of differential entropy [124, Th. 8.6.3], while inequality (4.22) follows from the fact that conditioning does not increase entropy [124, Th. 8.6.1]. If $\xi = \varepsilon - \hat{\varepsilon}$ is interpreted as the estimation error, this lower bound becomes tighter the less ξ depends on $\hat{\varepsilon}$.

By using the maximum value of $\mathrm{h}(\xi)$, which is achieved when ξ is Gaussian [124, Th. 8.6.5], an even simpler lower bound is derived[2]:

$$\mathrm{I}(\varepsilon;\hat{\varepsilon}) \geq \mathrm{h}(\varepsilon) - \frac{1}{2}(1 + \ln(2\pi\mathrm{var}(\xi))) = \mathrm{I}_{\mathrm{LOW}}. \tag{4.23}$$

Therefore, the proposed design criterion simplifies to

$$\hat{\varepsilon} = \arg\max_{\hat{\varepsilon}\in\mathcal{Q}} \mathrm{I}_{\mathrm{LOW}} \equiv \arg\min_{\hat{\varepsilon}\in\mathcal{Q}} \mathrm{var}(\xi). \tag{4.24}$$

Remarkably, this information-theoretic approach reduces to a variance minimization problem. Moreover, the variance minimization criterion is meaningful on its own, since a smaller estimation error variance implies a tighter concentration of $\hat{\varepsilon}$, which is expected to improve symbol separation and reduce detection error probability.

The variance in (4.24) can be obtained from the average mean squared error (MSE) and bias of $\hat{\varepsilon}$, by applying the law of total expectation [125, Th. 9.1.5]:

$$\mathrm{var}(\xi) = \mathrm{MSE}(\hat{\varepsilon}) - \mathrm{b}(\hat{\varepsilon})^2 = \mathrm{E}_{\varepsilon}\big[\mathrm{E}_{\mathbf{r}|\varepsilon}\big[(\varepsilon - \hat{\varepsilon})^2\big]\big] - \mathrm{E}_{\varepsilon}\big[\varepsilon - \mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}]\big]^2. \tag{4.25}$$

[2] In section 4.6.1, we prove the test statistic $\hat{\varepsilon}|x$ is increasingly Gaussian for large N, thus the bound is expected to be asymptotically tight.

4.5.2 Best quadratic unbiased estimator

Before dealing with the full optimization problem, it is of conceptual interest to study a simpler version. Consider a *genie-aided* estimator with quadratic structure (4.19) that aims at minimizing $\mathrm{var}(\xi)$ but is aware of the exact transmitted energy ε. In such a scenario, (4.25) reduces to:

$$\mathrm{var}(\xi) = \mathrm{E}_{\mathbf{r}|\varepsilon}\big[(\varepsilon - \hat{\varepsilon})^2\big] - \big(\varepsilon - \mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}]\big)^2 = \mathrm{E}_{\mathbf{r}|\varepsilon}\big[\hat{\varepsilon}^2\big] - \mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}]^2 \equiv \mathrm{var}(\hat{\varepsilon}|\varepsilon). \tag{4.26}$$

It depends on the conditional first and second-order moments of $\hat{\varepsilon}$, which have been derived in appendix D:

$$\begin{aligned} \mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}] &= \mathrm{tr}\big(\mathbf{A}\mathbf{C}_{\mathbf{r}|\varepsilon}\big) + c \\ \mathrm{E}_{\mathbf{r}|\varepsilon}\big[\hat{\varepsilon}^2\big] &= \mathrm{tr}\Big(\mathbf{A}^2\mathbf{C}_{\mathbf{r}|\varepsilon}^2\Big) + \big(\mathrm{tr}\big(\mathbf{A}\mathbf{C}_{\mathbf{r}|\varepsilon}\big) + c\big)^2. \end{aligned} \tag{4.27}$$

Using them in (4.26), the conditional variance of $\hat{\varepsilon}$ becomes

$$\mathrm{var}(\hat{\varepsilon}|\varepsilon) = \big\|\mathbf{A}\mathbf{C}_{\mathbf{r}|\varepsilon}\big\|_{\mathrm{F}}^2. \tag{4.28}$$

As the affine term c does not affect the variance we want to optimize, we can freely set it. For convenience, we propose to choose it to obtain a conditionally unbiased estimator, which translates into

$$\mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}] = \mathrm{tr}(\mathbf{A}(\varepsilon\boldsymbol{\Gamma} + \mathbf{I}_N)) + c = \varepsilon \implies \begin{cases} c = -\mathrm{tr}(\mathbf{A}), \\ \mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) = 1. \end{cases} \tag{4.29}$$

Under conditional unbiasedness, $\mathrm{var}(\hat{\varepsilon}|\varepsilon)$ is equivalent to the conditional MSE, *i.e.* $\mathrm{MSE}(\hat{\varepsilon}|\varepsilon)$. Following a similar reasoning to the one behind the well-known best linear unbiased estimator (BLUE) [126, Ch. 6], with the quadratic structure presented, we will term the estimator that minimizes $\mathrm{MSE}(\hat{\varepsilon}|\varepsilon)$ *best quadratic unbiased estimator (BQUE)*.[3]

Accounting for the constraints in (4.29), we can solve the optimization problem using the method of Lagrange multipliers:

$$\mathscr{L}(\mathbf{A}, \lambda) = \big\|\mathbf{A}\mathbf{C}_{\mathbf{r}|\varepsilon}\big\|_{\mathrm{F}}^2 + \lambda(\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) - 1). \tag{4.30}$$

We differentiate $\mathscr{L}$ with respect to $\mathbf{A}$ and equate it to $\mathbf{0}$:

$$\frac{\partial \mathscr{L}(\mathbf{A}, \lambda)}{\partial \mathbf{A}} = 2\mathbf{C}_{\mathbf{r}|\varepsilon}\mathbf{A}\mathbf{C}_{\mathbf{r}|\varepsilon} + \lambda\boldsymbol{\Gamma} = \mathbf{0}_{N\times N} \implies \mathbf{A} = -\frac{\lambda}{2}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}. \tag{4.31}$$

Taking into account that $\boldsymbol{\Gamma}$ and $\mathbf{C}_{\mathbf{r}|\varepsilon}$ are diagonal matrices, $\mathbf{A}$ must also be diagonal. Enforcing the unbiasedness constraint leaves us with:

$$\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) = -\frac{\lambda}{2}\Big\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\Big\|_{\mathrm{F}}^2 = 1 \implies -\frac{\lambda}{2} = \frac{1}{\Big\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\Big\|_{\mathrm{F}}^2}. \tag{4.32}$$

Substituting in (4.31) we obtain the matrix of the quadratic form:

$$\mathbf{A}_{\mathrm{BQUE}} = \frac{\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-2}}{\Big\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\Big\|_{\mathrm{F}}^2}. \tag{4.33}$$

[3] In [127], [128], the authors propose the same name for an estimator with similar structural constraints.

Finally, the BQUE is given by

$$\hat{\varepsilon}_{\mathrm{BQUE}}(\mathbf{r}) = \frac{\mathbf{r}^{\mathrm{H}}\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-2}\mathbf{r} - \mathrm{tr}\left(\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-2}\right)}{\left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\right\|_{\mathrm{F}}^{2}}. \tag{4.34}$$

If we analyze the conditional MSE of this estimator, *i.e.*

$$\mathrm{MSE}\left(\hat{\varepsilon}_{\mathrm{BQUE}}|\varepsilon\right) = \left\|\mathbf{A}_{\mathrm{BQUE}}\mathbf{C}_{\mathbf{r}|\varepsilon}\right\|_{\mathrm{F}}^{2} = \frac{\left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\right\|_{\mathrm{F}}^{2}}{\left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\right\|_{\mathrm{F}}^{4}} = \left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\right\|_{\mathrm{F}}^{-2}, \tag{4.35}$$

we observe it coincides with the Cramér–Rao bound (CRB) associated with the estimation of ε, which has been derived in appendix E. Although this condition is sufficient to state if an estimator is efficient, the BQUE depends on the parameter to be estimated, making it unrealizable. Therefore, it is not the minimum variance unbiased estimator (MVUE) [126, Ch. 2].

4.5.3 Quadratic minimum mean squared error estimator

Developing the BQUE has provided us with valuable insights on problem (4.24) from the perspective of classical estimation theory. We now derive the structure of the quadratic estimator that maximizes $\mathrm{I}_{\mathrm{LOW}}$. Recall (4.25), which depends on the average MSE and squared mean bias of $\hat{\varepsilon}$:

$$\begin{aligned} \mathrm{MSE}(\hat{\varepsilon}) &= \mathrm{E}_{\varepsilon}\left[\varepsilon^2 - 2\varepsilon\mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}] + \mathrm{E}_{\mathbf{r}|\varepsilon}\left[\hat{\varepsilon}^2\right]\right] \\ \mathrm{b}(\hat{\varepsilon})^2 &= 1 - 2\mathrm{E}_{\varepsilon}\left[\varepsilon\mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}]\right] + \mathrm{E}_{\varepsilon}\left[\mathrm{E}_{\mathbf{r}|\varepsilon}[\hat{\varepsilon}]\right]^2. \end{aligned} \tag{4.36}$$

Plugging the first and second-order moments from (4.27) in the previous expressions yields

$$\begin{aligned} \mathrm{MSE}(\hat{\varepsilon}) &= \left(c + \mathrm{tr}\left(\mathbf{A}\overline{\mathbf{C}}\right) - 1\right)^2 + \sigma_{\varepsilon}^2\left(1 - \mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma})\right)^2 + \mathrm{tr}\left(\mathbf{A}^2\mathring{\mathbf{C}}^2\right) \\ \mathrm{b}(\hat{\varepsilon})^2 &= \left(c + \mathrm{tr}\left(\mathbf{A}\overline{\mathbf{C}}\right) - 1\right)^2, \end{aligned} \tag{4.37}$$

with $\overline{\mathbf{C}} = \boldsymbol{\Gamma} + \mathbf{I}_N$, $\mathring{\mathbf{C}}^2 = (\sigma_{\varepsilon}^2 + 1)\boldsymbol{\Gamma}^2 + 2\boldsymbol{\Gamma} + \mathbf{I}_N$ and $\sigma_{\varepsilon}^2 = \mathrm{var}(\varepsilon)$. Therefore, the variance to be minimized results in

$$\mathrm{var}(\xi) = \sigma_{\varepsilon}^2\left(1 - \mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma})\right)^2 + \left\|\mathbf{A}\mathring{\mathbf{C}}\right\|_{\mathrm{F}}^2, \tag{4.38}$$

which is not affected by the value of c, as expected. If we set it to $c = 1 - \mathrm{tr}\left(\mathbf{A}\overline{\mathbf{C}}\right)$, the squared mean bias term cancels (*i.e.* $\mathrm{b}(\hat{\varepsilon})^2 = 0$). In this situation, the criterion of maximum $\mathrm{I}_{\mathrm{LOW}}$ is equivalent to that of minimum MSE on average. This results in the *(Bayesian) quadratic minimum mean squared error (QMMSE)* estimator, which is an extension of the well-known linear minimum mean squared error (LMMSE) estimator [126, Ch. 12].

To obtain its expression, we shall proceed as in section 4.5.2, by differentiating (4.38) and equating it to $\mathbf{0}$:

$$\frac{\partial \mathrm{var}(\xi)}{\partial \mathbf{A}} = -2\sigma_{\varepsilon}^2\left(1 - \mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma})\right)\boldsymbol{\Gamma} + 2\mathbf{A}\mathring{\mathbf{C}}^2 = \mathbf{0}_{N\times N}. \tag{4.39}$$

By isolating $\mathbf{A}$, multiplying it by $\boldsymbol{\Gamma}$ and computing the trace, we obtain the following term:

$$\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) = \frac{\sigma_{\varepsilon}^2\left\|\boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-1}\right\|_{\mathrm{F}}^2}{1 + \sigma_{\varepsilon}^2\left\|\boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-1}\right\|_{\mathrm{F}}^2}, \tag{4.40}$$

which yields the matrix

$$\mathbf{A}_{\text{QMMSE}} = \frac{\sigma_\varepsilon^2 \boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-2}}{1 + \sigma_\varepsilon^2 \left\| \boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-1} \right\|_{\text{F}}^2} \tag{4.41}$$

and the estimator

$$\hat{\varepsilon}_{\text{QMMSE}}(\mathbf{r}) = \frac{\sigma_\varepsilon^2 \mathbf{r}^{\text{H}} \boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-2}\mathbf{r} + 1 - \sigma_\varepsilon^2 \text{tr}\left(\boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-2}\right)}{1 + \sigma_\varepsilon^2 \left\| \boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-1} \right\|_{\text{F}}^2}. \tag{4.42}$$

The mean MSE value it reaches is

$$\text{MSE}(\hat{\varepsilon}_{\text{QMMSE}}) = \frac{\sigma_\varepsilon^2}{1 + \sigma_\varepsilon^2 \left\| \boldsymbol{\Gamma}\mathring{\mathbf{C}}^{-1} \right\|_{\text{F}}^2}. \tag{4.43}$$

4.5.4 Unified framework for quadratic detectors

As stated in section 4.4, the framework outlined in (4.19) is general enough to encapsulate a variety of energy estimators as a first step in symbol detection. For instance, an estimator of the form

$$\mathbf{A}_{\text{ED}} = \frac{\mathbf{I}_N}{\text{tr}(\boldsymbol{\Gamma})} \implies \hat{\varepsilon}_{\text{ED}}(\mathbf{r}) = \frac{\|\mathbf{r}\|^2 - N}{\text{tr}(\boldsymbol{\Gamma})} \tag{4.44}$$

with a posterior classification (with suitable decision regions) is equivalent to the energy detector from (4.16). Similarly, for the high SNR detector in (4.18), its corresponding quadratic statistic is

$$\mathbf{A}_{\text{HSNR}} = \frac{\boldsymbol{\Gamma}^{-1}}{N} \implies \hat{\varepsilon}_{\text{HSNR}}(\mathbf{r}) = \frac{\mathbf{r}^{\text{H}}\boldsymbol{\Gamma}^{-1}\mathbf{r} - \text{tr}\left(\boldsymbol{\Gamma}^{-1}\right)}{N}. \tag{4.45}$$

Notice that the affine term in both cases has been set to $c = 1 - \text{tr}\left(\mathbf{A}\overline{\mathbf{C}}\right)$ to more easily compare them with the QMMSE estimator. As expected, this makes them conditionally unbiased.

An interesting aspect regarding $\hat{\varepsilon}_{\text{QMMSE}}$ is that, in the high SNR regime, its quadratic term matrix becomes

$$\lim_{\alpha\to\infty} \mathbf{A}_{\text{QMMSE}} = \lim_{\alpha\to\infty} \frac{1}{1 + \frac{1}{\sigma_\varepsilon^2} + N} \boldsymbol{\Gamma}^{-1}. \tag{4.46}$$

This is a scaled version of $\mathbf{A}_{\text{HSNR}}$ and, in fact, both matrices coincide for asymptotically large N. Therefore, at high SNR, both $\hat{\varepsilon}_{\text{QMMSE}}$ and $\hat{\varepsilon}_{\text{HSNR}}$ will present the same performance and error floor.

Another important property of $\hat{\varepsilon}_{\text{ED}}$, $\hat{\varepsilon}_{\text{HSNR}}$ and $\hat{\varepsilon}_{\text{BQUE}}$ is that they all converge to the same estimator when the channel is assumed to be uncorrelated, so they are all optimal in that case:

$$\hat{\varepsilon}(\mathbf{r}) = \frac{1}{\alpha}\left(\frac{1}{N}\|\mathbf{r}\|^2 - 1\right). \tag{4.47}$$

Similarly, $\hat{\varepsilon}_{\text{QMMSE}}$ becomes

$$\hat{\varepsilon}'(\mathbf{r}) = \frac{\|\mathbf{r}\|^2 + \frac{(\alpha+1)^2}{\sigma_\varepsilon^2\alpha} + \alpha - N}{\frac{(\alpha+1)^2}{\sigma_\varepsilon^2\alpha} + \alpha + N\alpha}, \tag{4.48}$$

which is a scaled and translated version of (4.47), *i.e.* they define equivalent detectors with appropriate decision regions, thus it is also optimal under uncorrelated Rayleigh fading.

4.6 Symbol detection

Once the energy estimation phase has been studied, we now move on to the second step in the architecture proposed in section 4.5: symbol detection. For a given constellation, we describe the detection regions considering that the energy statistic asymptotically follows a Gaussian distribution. Afterwards, we analyze the probability of error of quadratic detectors and present a decision-directed design that leverages the BQUE to improve performance.

4.6.1 Detection regions

Quadratic energy estimators of the form (4.19) can be expressed as a summation:

$$\hat{\varepsilon}(\mathbf{r}) = \mathbf{r}^{\mathrm{H}}\mathbf{A}\mathbf{r} + c = \sum_{n=1}^{N} a_n |r_n|^2 + c_n, \tag{4.49}$$

where $a_n = [\mathbf{A}]_{n,n}$ and $\sum_n c_n = c$. Particularizing for the estimators presented in the previous section (*i.e.* (4.38)), we can set $c = 1 - \mathrm{tr}(\mathbf{A}(\boldsymbol{\Gamma} + \mathbf{I}_N))$, since it simplifies the analysis. Notice how (4.49) produces a real output because $\mathbf{A}$ is Hermitian. Therefore, defining a decision region $\mathscr{R}_x$ for each $x \in \mathscr{X}$ consists in determining detection thresholds on the real line:

$$\hat{x} = \begin{cases} x_1, & \hat{\varepsilon} \leq \tau_1, \\ x_i, & \tau_{i-1} < \hat{\varepsilon} < \tau_i, \\ x_M, & \tau_{M-1} \leq \hat{\varepsilon}. \end{cases} \tag{4.50}$$

The thresholds that minimize the error probability, $\mathscr{T} = \{\tau_i\}_{1 \leq i \leq M-1}$, are found at the intersection between the densities of the estimator output conditioned to every transmitted symbol of $\mathscr{X}$.

Since $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$ is a quadratic form of a complex normal vector, it follows a generalized chi-squared distribution [129]. Its probability density function (PDF) can be obtained analytically in specific cases [130], [131]. Nevertheless, the resulting expressions for general quadratic detectors and arbitrary correlation models are usually very involved [132, Ch. 1], [133] and do not allow for a simple derivation of detection thresholds.

In the context of massive SIMO, we can exploit the asymptotic properties of $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$ for large N. By computing its mean and variance (see appendix D) and invoking the CLT, we obtain

$$\hat{\varepsilon}(\mathbf{r}|\varepsilon) \xrightarrow{\mathrm{d}} \mathscr{N}\Big(1 - (1-\varepsilon)\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}),\, \|\mathbf{A}(\varepsilon\boldsymbol{\Gamma} + \mathbf{I}_N)\|_{\mathrm{F}}^2\Big), \tag{4.51}$$

as $N \to \infty$. For the unbiased estimators, $\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) = 1$, thus the Gaussian density function is centered at the symbol energy level. Observe that the terms in sum (4.49) are independent but not identically distributed. To ensure the CLT can be applied in this scenario, it is sufficient to prove that Lyapunov's condition is satisfied. In appendix C.3 we show that it does indeed hold, thus validating the use of the CLT. This result motivates employing a Gaussian approximation for the densities of $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$.

A representation of the detection regions, under the Gaussian hypothesis, of the BQUE when transmitting a 4-ASK constellation with $N = 128$ is depicted in figure 4.3. The first symbol,

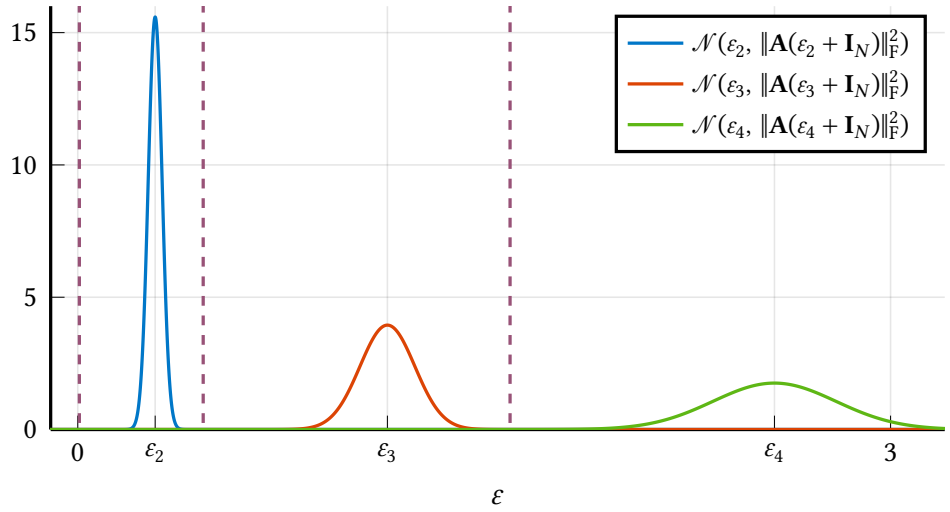


Figure 4.3: Detection regions of the BQUE for a 4-ASK constellation when $N = 128$. The symbol $\varepsilon = 0$ is omitted for visualization purposes.

$\varepsilon_1 = 0$, is omitted for visualization purposes, since its variance is notably small compared with that of the other symbols.

The vertical lines in figure 4.3 represent the detection thresholds, $\mathcal{T}$. In the Gaussian case, obtaining them reduces to finding the roots of a second-degree polynomial for each pair of adjacent densities. Since the variance increases with the symbol energy, each polynomial admits two real roots; however, only the root lying between the two likelihood functions is relevant.[4] Moreover, due to the monotonic increase in variance, the relevant intersection corresponds to the largest root for each adjacent pair. An illustration of this behavior is shown in figure 4.4, where we depict the intersection points between the natural logarithms of three different Gaussian densities. As expected, the first and third densities have a single intersection, whereas the second and third intersect twice, both to the left of the former.

The procedure of finding the thresholds is summarized in algorithm 1.

Finally, recall that channel hardening manifests as a reduction in the variance of each symbol estimate when N increases. Interestingly, this change does not significantly alter the decision regions. In other words, provided that the number of antennas is sufficiently large, the positions of the detection thresholds remain almost unaffected, since all symbols are scaled proportionally and the intersection points shift only marginally. To illustrate this phenomenon, in table 4.1 we list the detection thresholds of the BQUE when a 4-ASK constellation is transmitted, for $N \in 16, 64, 256, 1024$.

This behavior will reappear in chapter 5 when we consider the design of optimized constellations.

[4] Gaussian densities with different means but equal variance have a single intersection point.

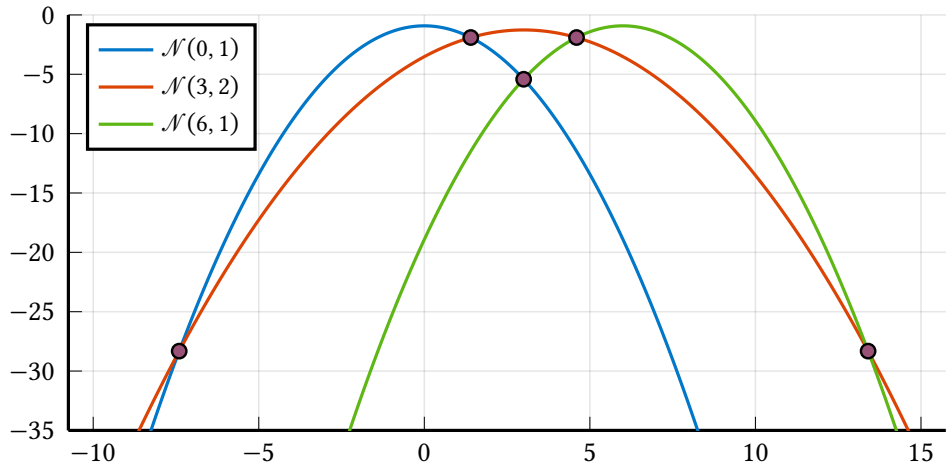


Figure 4.4: Intersection between the natural logarithms of Gaussian densities $\mathcal{N}(0,1)$, $\mathcal{N}(3,2)$ and $\mathcal{N}(6,1)$.

Algorithm 1 Thresholds between Gaussian likelihoods

Require: constellation $\mathcal{X}$, matrix $\mathbf{A}$, spectrum $\mathbf{\Gamma}$
Ensure: thresholds $\mathcal{T} = \{\tau_1, \dots, \tau_{M-1}\}$

procedure NORMALINTERSECTION($\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$)
  $a = 1/\sigma_2^2 - 1/\sigma_1^2$
  $b = 2 \cdot (\mu_1/\sigma_1^2 - \mu_2/\sigma_2^2)$
  $c = \mu_2^2/\sigma_2^2 - \mu_1^2/\sigma_1^2 + \ln(\sigma_2^2/\sigma_1^2)$
  return roots($\mathcal{P}(c, b, a)$)
end procedure
for $i = 1 : M-1$ **do**
  $\mu_1 = 1 - (1 - \varepsilon_i)\text{tr}(\mathbf{A\Gamma})$
  $\mu_2 = 1 - (1 - \varepsilon_{i+1})\text{tr}(\mathbf{A\Gamma})$
  $\sigma_1^2 = \|\mathbf{A}(\varepsilon_i\mathbf{\Gamma} + \mathbf{I}_N)\|_{\text{F}}^2$
  $\sigma_2^2 = \|\mathbf{A}(\varepsilon_{i+1}\mathbf{\Gamma} + \mathbf{I}_N)\|_{\text{F}}^2$
  $\tau_i = \max(\text{NORMALINTERSECTION}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2))$
end for
return $\mathcal{T}$

4.6.2 Probability of detection error

From (4.50) we observe that the error probability can be computed as:

$$P_\epsilon = \frac{1}{M}\Big(\text{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_1) > \tau_1) + \text{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_M) < \tau_{M-1}) + \sum_{i=2}^{M-1} \text{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_i) < \tau_{i-1}) + \text{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_i) > \tau_i)\Big). \tag{4.52}$$

Table 4.1: Detection thresholds, τ_1, τ_2 and τ_3, of the BQUE detector for a 4-ASK constellation with $N \in \{16, 64, 256, 1024\}$.

	$N = 16$	$N = 64$	$N = 256$	$N = 1024$
τ_1	0.0072	0.0061	0.0058	0.0057
τ_2	0.4899	0.4670	0.4609	0.4594
τ_3	1.6808	1.6083	1.5891	1.5843

Under the Gaussian assumption, each tail probability is approximated with the Q-function:

$$\begin{aligned} \mathrm{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_i) < \tau_{i-1}) &\approx \mathrm{Q}\left(\frac{1-(1-\varepsilon_i)\mathrm{tr}(\mathbf{A\Gamma})-\tau_{i-1}}{\|\mathbf{A}(\varepsilon_i\mathbf{\Gamma}+\mathbf{I}_N)\|_{\mathrm{F}}}\right), \\ \mathrm{P}(\hat{\varepsilon}(\mathbf{r}|\varepsilon_i) > \tau_i) &\approx \mathrm{Q}\left(\frac{\tau_i-1+(1-\varepsilon_i)\mathrm{tr}(\mathbf{A\Gamma})}{\|\mathbf{A}(\varepsilon_i\mathbf{\Gamma}+\mathbf{I}_N)\|_{\mathrm{F}}}\right), \end{aligned} \tag{4.53}$$

with ε_i defined as in (4.7). Thus, the error probability is approximately given by

$$P_\epsilon \approx \frac{1}{M}\left(\sum_{i=2}^{M}\mathrm{Q}\left(\frac{1-(1-\varepsilon_i)\mathrm{tr}(\mathbf{A\Gamma})-\tau_{i-1}}{\|\mathbf{A}(\varepsilon_i\mathbf{\Gamma}+\mathbf{I}_N)\|_{\mathrm{F}}}\right)+\sum_{j=1}^{M-1}\mathrm{Q}\left(\frac{\tau_j-1+(1-\varepsilon_j)\mathrm{tr}(\mathbf{A\Gamma})}{\|\mathbf{A}(\varepsilon_j\mathbf{\Gamma}+\mathbf{I}_N)\|_{\mathrm{F}}}\right)\right). \tag{4.54}$$

In expression (4.54) we observe that the error probability increases with the norms

$$\|\mathbf{A}(\varepsilon_j\mathbf{\Gamma}+\mathbf{I}_N)\|_{\mathrm{F}}, \tag{4.55}$$

which correspond to the square root of the MSE of the BQUE for each symbol given in (4.35). This dependence shows that the information-theoretic criteria proposed in section 4.5.1 is indeed justified, although it is not optimal, since the detection thresholds in the numerator also depend on $\mathbf{A}$. The optimization of (4.54) is further discussed in chapter 5.

Moreover, optimizing the MSE on average (*i.e.* as in QMMSE) is not consistent with the fact that the error probability of each symbol has a different impact at each SNR level. Therefore, we are interested in implementing a detector capable of leveraging the near-optimality of the BQUE.

4.6.3 Assisted best quadratic unbiased estimator

In order to implement a near-optimal detector that exploits the dependence of the SER on the MSE, we now propose to assist the BQUE by replacing the true transmitted symbol, assumed known by the genie-aided decoder, with the final decision delivered by the ED. By doing so, we obtain a hard-decision detector that offers some computational advantages over its soft counterpart, in which the energy estimate is directly exploited without an explicit decision. In the latter case, it is not possible to precompute either the BQUE detection regions or the associated matrix $\mathbf{A}_{\mathrm{BQUE}}$.

On the one hand, assisting with hard decisions is prone to error boosting at low-SNR. However, the simulations in section 4.7 will show that this effect is negligible except for extremely high antenna correlation (*i.e.* when the ED decision is not reliable at all). On the other hand, in the hard-decision scheme, the symbol plugged into the BQUE belongs to a

known constellation, thus matrices $\mathbf{A}_{\mathrm{BQUE},i}$ for each symbol only need to be computed once and can be stored for later uses:

$$\mathbf{A}_{\mathrm{BQUE},i} = \frac{\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon_i}^{-2}}{\left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon_i}^{-1}\right\|_{\mathrm{F}}^2}, \quad 1 \leq i \leq M. \tag{4.56}$$

This result has further implications. Given that the thresholds of the proposed quadratic detectors only depend on the constellation and $\mathbf{A}$, if $\mathbf{A}$ can be computed only once, thresholds can also be computed only once.

The decision-directed nature of the *assisted BQUE (ABQUE)* makes its theoretical analysis much more involved than that of the quadratic detectors considered in previous sections. Nonetheless, the BQUE can serve as a reliable benchmark to assess it, by providing an analytic bound on its error probability. Simulations in the next section show that both detectors perform very similarly (in terms of SER) in a variety of scenarios.

4.7 Numerical results

In this section, we provide some simulations with two different goals. First, to illustrate performance aspects that are not apparent from analytic results presented in the preceding sections; in particular, we display how the outage probability performance of statistical CSI-aware detectors benefits from channel hardening. Second, to numerically validate the theoretical results presented in the previous sections and compare the various detectors in terms of average SER performance.

In all simulations we have employed signal (4.1) assuming the exponential correlation channel model described in section 2.2.2, with correlation coefficient $\rho = 0.7$ (unless stated otherwise). Symbols from a uniform unipolar 8-ASK constellation have been transmitted with average power equal to one. We have chosen a standard constellation instead of one optimized to the channel statistics (*e.g.* see [42]) to better portray a realistic scenario, with a low complexity transmitter that is unaware of CSI. It has the added benefit of being robust to SNR estimation errors in transmission.

4.7.1 Outage probability

An outage event occurs when the instantaneous SER of a system is above a certain threshold ζ_{out}, given a channel realization [134]. In the context of URLLC, the outage probability is a relevant quality-of-service metric because it is intimately related to the reliability and latency of a communication setup [135], [136]. Within the one-shot scheme considered in this work, we define it as follows:

$$P_{\mathrm{out}} = \int_{\mathcal{O}} f_{\mathbf{h}}(\mathbf{h})\,\mathrm{d}\mathbf{h}, \tag{4.57}$$

where $\mathcal{O} = \{\mathbf{h} \in \mathbb{C}^N \mid P_\epsilon(\mathbf{h}) > \zeta_{\mathrm{out}}\}$, with $P_\epsilon(\mathbf{h})$ the average error probability conditioned on a given channel realization $\mathbf{h}$. In coherent communications, it is equivalent to the probability that the instantaneous SNR drops below a certain value [9, Eq. (6.46)]. On the contrary, this correspondence does not hold in noncoherent systems of the kind considered herein, as it is clear from figure 4.5, where the normalized channel norms corresponding to specific

SERs are depicted. We can observe that the highest SER values correspond to channel realizations that deviate the most from the expected value, from both above and below. Channel hardening counteracts this deviation from the mean: the higher the number of antennas at the receiver, the more concentrated the channel realizations are (in the norm), and the lower their associated SER values are.

Figure 4.6 illustrates how this hardening affects outage probability. In general terms, employing more antennas stabilizes the channel statistics and reduces the chances of dealing with an outage event. Although ED benefits from this property, its hardening gains are less pronounced than ML and BQUE ones. These latter two share similar performance unless the quality requirement becomes very stringent, in which case ML outperforms BQUE. It is worth to remark that the minimum SER in figure 4.5 coincides with the threshold ζ_{out} for which the BQUE outage probability reaches its maximum in figure 4.6.

Another relevant phenomenon that can be observed in figure 4.6 is the fact there exist channel realizations for which the ED outperforms the ML detector. Indeed, (unconditional) ML detection is optimal in terms of average error probability, but not necessarily when conditioned to a specific channel realization.

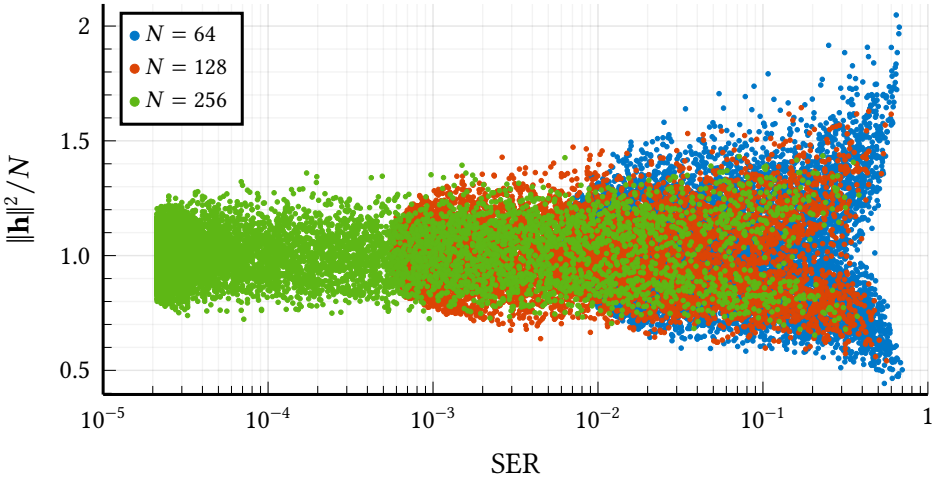


Figure 4.5: Scatter plot to assess the relevance of the channel norm as an indicator of SER performance. The BQUE error probability for 10^4 different channels at SNR = 10 dB is depicted.

4.7.2 Symbol error rate analysis

To evaluate the validity of the error probability expression from (4.54), it is compared against Monte Carlo results of the high SNR, ED, QMMSE and BQUE detectors. Analytical expressions are drawn as continuous lines, whereas simulation results are painted as dotted lines with markers sharing the same color. ML and ABQUE detectors are also assessed numerically.

In figure 4.7, the performance of all detectors discussed previously is depicted as a function of the SNR for $N = 512$ antennas. We have computed the thresholds for each quadratic detector and then classified the estimated powers, as described in section 4.6.1. A first observation of these results reveals that the ED error floor is much higher than that of the other quadratic detectors, which share it with the genie-aided detector. This error floor is very close to the

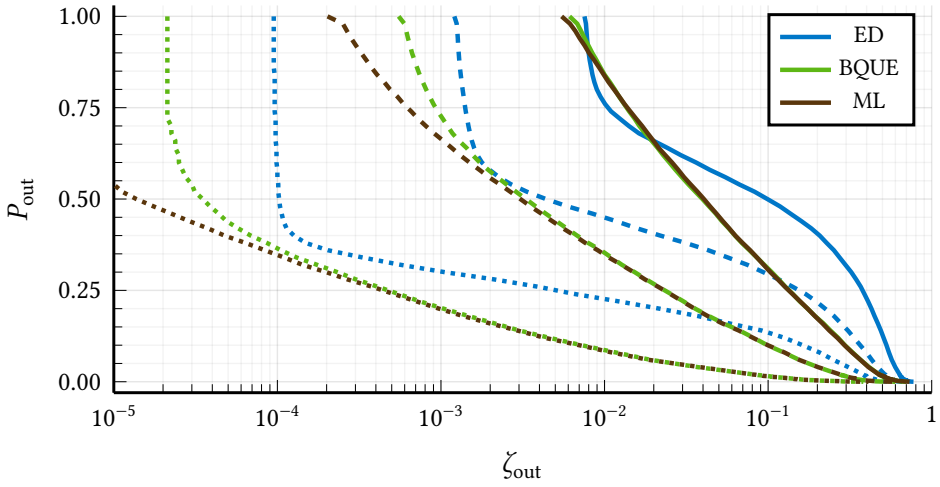


Figure 4.6: Outage probability for a SER threshold, ζ_{out}, at SNR $= 10$ dB for $N = 64$ (solid lines), $N = 128$ (dashed lined) and $N = 256$ (dotted lines).

one obtained with ML detection. The reason behind the slight discrepancy is that, although the CLT holds, the Gaussian likelihoods assumed in algorithm 1 present inaccuracies that affect the threshold positioning, thus resulting in an increased error floor. Another notable outcome is that the ABQUE exhibits a performance very close to that of the oracle (*i.e.* BQUE) with just a narrow increase in complexity.

Figures 4.8 and 4.9 illustrate the error floor level by depicting the SER of the different detectors at SNR $= 30$ dB in terms of the number of antennas and the correlation coefficient, respectively. In the former, we can observe the penalization suffered by ED. By being agnostic to channel statistics, it is not able to fully exploit the increasing number of antennas. As a result, its error floor is higher than the other methods' and it decreases at a slower rate, as well. The other quadratic detectors and ABQUE all display error levels similar to ML. Regarding the SER behavior in terms of channel correlation, the statistical CSI-aware detectors display remarkable robustness, since their error floors do not change significantly for $0 \leq \rho \leq 0.9$. Moreover, this SER is shared with the ML detector for $\rho < 0.99$ (up to the gap mentioned above). On the other hand, the error probability of the ED is acutely sensitive to channel correlation. For instance, at $\rho = 0.7$, its performance has degraded by almost a factor of 100, compared with the uncorrelated case.

A final note on these results is the remarkable accuracy of the approximated error probability expressions derived from (4.54). Indeed, in figure 4.8 we can clearly observe that the difference between the theoretical and the numerical SER is only barely noticeable for small N, thereby illustrating how the Gaussian approximation becomes increasingly accurate when the number of antennas increases.

4.8 Comparison of detection schemes

In this section, we undertake a comprehensive comparison between the established ML and energy detectors and the novel BQUE, QMMSE and ABQUE detectors introduced in this

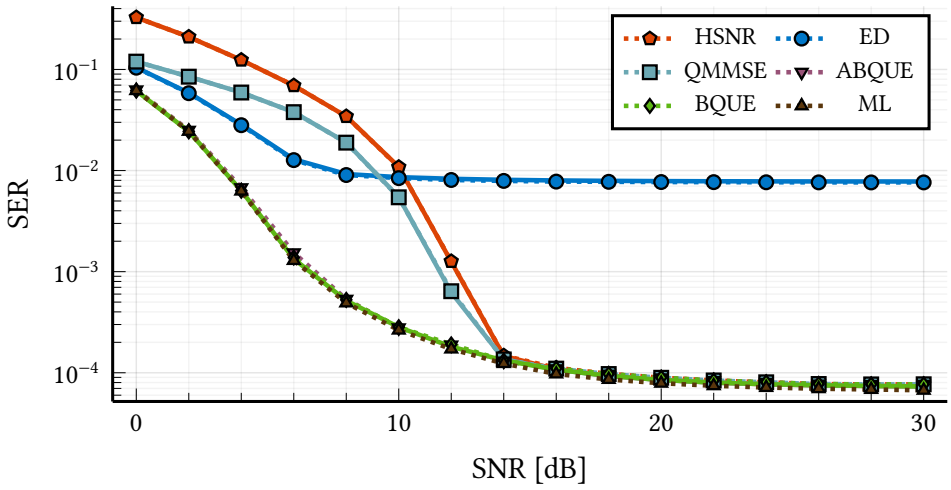


Figure 4.7: SER of the presented detectors in terms of SNR for $N = 512$.

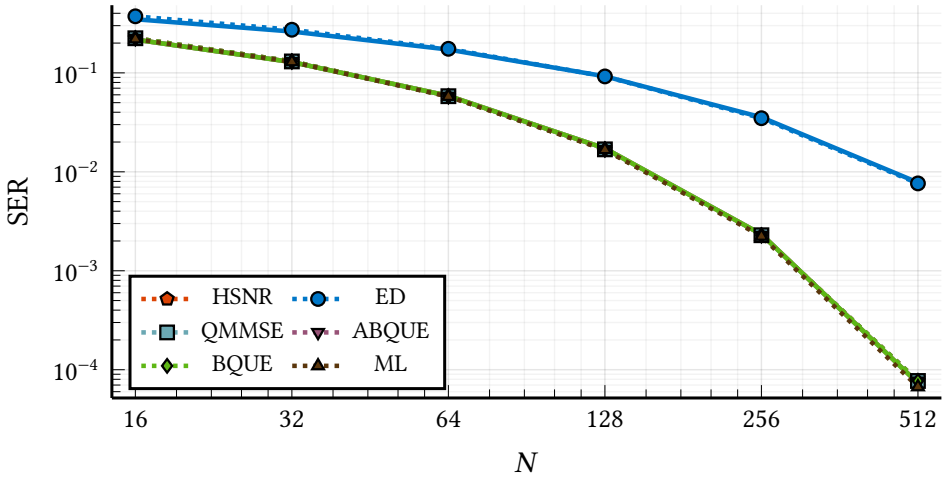


Figure 4.8: Floor level (*i.e.* SER at SNR = 30 dB) of the presented detectors in terms of N.

study. To facilitate this comparison, table 4.2 outlines six distinct properties for each detector:

1. **Complexity:** refers to the computational complexity of the entire detection process.
2. **Performance:** denotes the efficiency of the detector in terms of error probability.
3. **Graceful degradation:** reflects the resilience of the detector in front of small changes in channel correlation.
4. **Tractable:** indicates the possibility of mathematically analyzing the detector and its error probability.
5. **Implementable:** describes whether the detector depends on the transmitted symbol or not.
6. **Agnostic to CSIR:** specifies whether the detector needs to know the channel covariance.

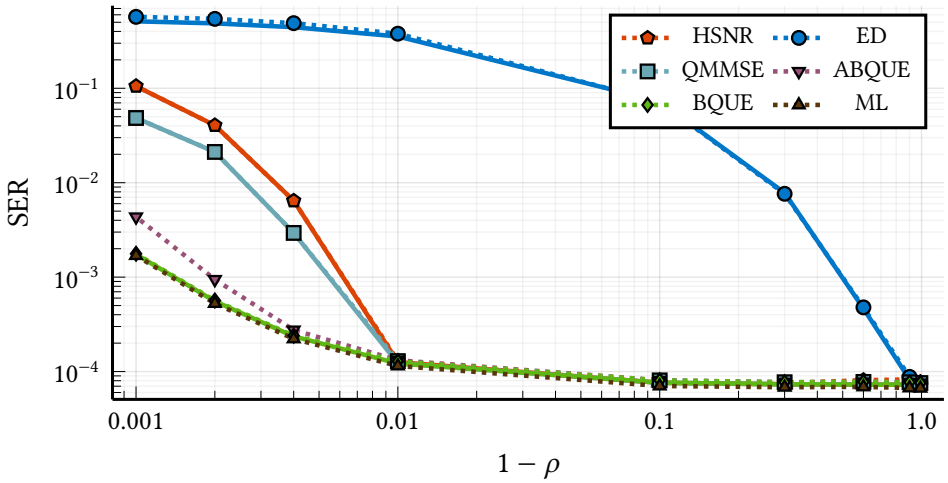


Figure 4.9: Floor level (*i.e.* SER at SNR = 30 dB) of the presented detectors in terms of channel correlation, for $N = 512$. The plot is with respect to $1-\rho$ for better visualization.

It is worth to mention that no single detector emerges as superior across all areas. Instead, there exists a trade-off among the aforementioned properties. For instance, while the ED has a very low complexity compared to other detectors, its performance is severely degraded for even slightly correlated channels (see figure 4.9).

Conversely, both the ML and BQUE detectors exhibit comparable performance and complexity. However, only the latter has an analytic expression for the error probability (see section 4.6.2). Furthermore, the tractability of BQUE can be extended to ABQUE when the channel correlation is not extremely high (see figure 4.9).

Finally, the QMMSE detector behaves like BQUE at moderate and high SNR, but its performance is inferior to that of the ED at low SNR. Even so, its resilience in the face of high channel correlation must not be overstated. A viable implementation that exhibits a high performance and is tractable is to leverage the ED at low SNR and switch to the QMMSE at moderate and high regimes.

Table 4.2: Summary of detection schemes properties.

Detector	Complexity	Performance	Graceful degradation	Tractable	Implementable	Agnostic to CSIR
ML	$O(N^2)+O(MN)$	Optimal	Yes	No	Yes	No
ED	$O(N)+O(M)$	High loss ($\rho \neq 0$)	No	Yes	Yes	Yes
BQUE	$O(N^2)+O(N)+O(M)$	Negligible loss	Yes	Yes	No	No
QMMSE	$O(N^2)+O(N)+O(M)$	Negligible loss (moderate & high SNR)	Yes ($\rho < 0.99$)	Yes	Yes	No
ABQUE	$O(N^2)+O(N)+O(M)$	Negligible loss	Yes	No	Yes	No

4.9 Conclusions

In this chapter, we have analyzed the fundamental limitations of one-shot SIMO noncoherent communication when an arbitrary PAM constellation is considered. We have also introduced a quadratic framework that generalizes the ED commonly used in the literature. We have derived an analytic approximation for the error probability of any detector exhibiting the quadratic structure proposed, as a sum of Q-functions. An improved scheme based on the combination of quadratic detectors has also been presented. Their performance in terms of average SER and outage probability has been tested through several Monte Carlo experiments, as well as the validity of the SER approximations.

We can outline some future research lines that arise from this work. The most straightforward one is to design PAM constellations optimized for a particular quadratic detector, by means of minimizing the error probability approximation (4.54). Similar approaches are studied in [12], [42], [72]. Other potential extensions of the presented analysis include the design of codes across multiple channel uses and suitable detection schemes [121], [137]. Finally, a possible application of the proposed framework is in noncoherent energy detection of index modulations [138], [139]. For instance, considering a frequency selective channel and a multicarrier modulation such as orthogonal frequency-division multiplexing (OFDM), information can be conveyed by assigning different transmission power levels to different sets of subcarriers.

Optimized constellations and index modulation schemes serve as a mean to reduce the SER of one-shot communication systems. These aspects are explored in the following chapter.

CHAPTER 5

Performance enhancement methods for one-shot communications

The error performance of the one-shot noncoherent receiver analyzed in chapter 4 is insufficient when no additional coding or redundancy is employed. For example, a massive SIMO system with $N = 512$ receive antennas, a correlation factor of $\rho = 0.7$, and an 8-ASK constellation exhibits an error floor on the order of 10^{-4}, as shown in figure 4.7.

To address this limitation, this chapter introduces two different strategies aimed at improving system performance, in terms of error probability. The first strategy focuses on the design of tailored constellations that minimize the SER given in (4.54). In the second strategy, information is embedded in the ordering by which a fixed set of values is mapped onto OFDM subcarriers, resulting in a spherical code that provides improved robustness against channel impairments.

5.1 Constellation design

The first strategy we propose to mitigate the SER in one-shot communication systems is the design of tailored constellations [21]. As observed in figures 4.5 and 4.7, the presence of an error floor can be partly attributed to a "favorable" channel realization combined with a low-power symbol being mistaken for a higher power symbol. From a detection theory point of view, there is no decision rule that effectively discriminates between both hypothesis, even for unbounded SNR [140]. This observation leads to the conclusion that fixed ASK constellations are not suitable for energy detection in noncoherent systems. Instead, one must consider alternative constellation designs that adapt to the SNR regime, thereby reducing ambiguity and improving reliability. Nevertheless, such adaptation comes at the cost of requiring statistical CSIT as well as the detection scheme employed by the receiver.

In the literature, there are several works that focus on the design of constellations for noncoherent energy-based detection. However, most of them [12], [72], [112], [121], [141], [142] rely on the isotropic channel assumption, which is not physically consistent in sparse scattering or near-field communication environments (see chapter 2). A notable exception is [42], which follows the design criteria developed in [143]. Due to the complexity of the SER, pairwise error probability and even the Chernoff bound expressions, Borran et al.

propose to employ the Kullback–Leibler divergence (KLD) as an indicator of the SER with ML detection. The authors of [42] apply this approach to design a constellation optimized for an asymptotically large number of antennas under the exponential correlation channel model (see section 2.2.2).

5.1.1 Error probability optimization

In contrast, we propose to exploit the fact that when a quadratic detector is employed the true SER can be accurately predicted with an analytic expression. Hence, the constellation design can be approached from SER minimization directly, rather than from the use of bounds or indicators. Formally, we can state the optimization problem as:

$$\begin{aligned} \min_{\boldsymbol{\varepsilon},\boldsymbol{\tau}} \quad & P_\epsilon(\mathbf{A},\boldsymbol{\varepsilon},\boldsymbol{\tau}) \\ \text{s.t.} \quad & \boldsymbol{\varepsilon},\boldsymbol{\tau} \geq \mathbf{0}, \\ & \varepsilon_1 = 0, \\ & \textstyle\sum_{i=1}^{M} \varepsilon_i = M, \\ & \varepsilon_i \leq \tau_i,\ i = 1,\dots,M-1, \\ & \tau_i \leq \varepsilon_{i+1},\ i = 1,\dots,M-1, \end{aligned} \tag{5.1}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1,\dots,\varepsilon_M)^{\mathrm{T}}$, $\boldsymbol{\tau} = (\tau_1,\dots,\tau_{M-1})^{\mathrm{T}}$, for a fixed M, and $P_\epsilon(\mathbf{A},\boldsymbol{\varepsilon},\boldsymbol{\tau})$ is the error probability given in (4.54), writing explicitly the dependence on the detector ($\mathbf{A}$), the constellation ($\boldsymbol{\varepsilon}$) and the thresholds ($\boldsymbol{\tau}$).

The solution of the nonlinear program (5.1) must satisfy the Karush–Kuhn–Tucker (KKT) conditions [144, Ch. 11], which can be solved employing interior-point methods [144, Ch. 5]. Specifically, we use JuMP [145], a modeling language for mathematical optimization embedded in Julia [146]. This approach is suitable for the energy, BQUE and HSNR detectors, but excludes the QMMSE since it depends on the fourth-order moment of the constellation (*i.e.* $\mathrm{var}(\varepsilon)$, see (4.37)). However, we can now set $\mathrm{tr}(\mathbf{A}\boldsymbol{\Gamma}) = 1$ and $c = -\mathrm{tr}(\mathbf{A})$, which greatly simplifies the problem.

Letting $\boldsymbol{\theta} = \left(\boldsymbol{\varepsilon}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}\right)^{\mathrm{T}}$ be the stacking of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\tau}$, and $\mathbf{e}_i$ the i-th vector of the canonical basis, the constraints from the original optimization problem can be rewritten as:

1. $\boldsymbol{\theta} \geq \mathbf{0}$, an intersection of $2M-1$ halfspaces.
2. $\mathbf{e}_1^{\mathrm{T}}\boldsymbol{\theta} = 0$, a hyperplane.
3. $(\mathbf{e}_1 + \cdots + \mathbf{e}_M)^{\mathrm{T}}\boldsymbol{\theta} = M$, a hyperplane.
4. $(\mathbf{e}_i - \mathbf{e}_{M+i})^{\mathrm{T}}\boldsymbol{\theta} \leq 0$, a halfspace.
5. $(\mathbf{e}_{M+i} - \mathbf{e}_{i+1})^{\mathrm{T}}\boldsymbol{\theta} \leq 0$, a halfspace.

Given that all constraints form convex sets [147, Sec. 2.2.1] and intersection preserves convexity [147, Sec. 2.3.1], we conclude that the feasible set is convex. Regarding the objective function, it is twice differentiable but not convex, as it is a combination of Q-functions.[1]

[1] Even with the Chernoff bound (convex for $x \geq 1$), the objective function would not be convex in general because the arguments of the Q-functions are not concave in general [147, Eq. (3.10)].

Hence, KKT conditions are necessary for global optimality, but their sufficiency cannot be guaranteed. Even so, numerical results in section 5.1.2 show a significant improvement in performance, particularly for the absence of error floor at high SNR.

In figure 5.1, we display various constellation designs under the exponential correlation model. Our solution at $\rho = 0$, where all quadratic detectors converge, coincides with the ML constellation from [42, Fig. 4] (up to a power scaling factor), further supporting the optimality of the solution. In the figure, it can also be observed that low-power symbols are closer than those of the 8-ASK, whereas high-power symbols are further apart. This is due to the fact that the variance of each $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$ in (4.19) increases with the symbol amplitude [148]. Therefore, low power symbols can be correctly detected even when they are very close to each other, but more powerful symbols must be more separated. Finally, we shall remark that intersymbol distances for statistical CSI-aware detectors follow the same tendency with ρ as that in [42]. The reason is that, when correlation increases, channel hardening requires more antennas to manifest and the uncertainty is higher. In this situation, the variances of each $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$ are more similar and symbols tend to be equispaced [49].

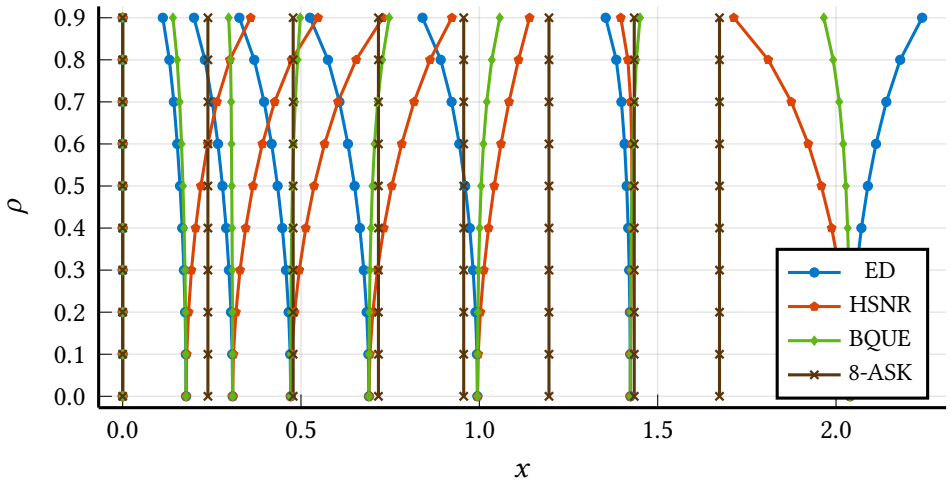


Figure 5.1: Comparison of 8-ASK and ED, HSNR and BQUE optimized constellations at a SNR equal to 15 dB and $N = 256$ for different levels of channel correlation.

On the other hand, in figure 5.2, we show that the optimized constellation is almost invariant to the number of antennas when this number is sufficiently large. This result follows directly from the discussion in section 4.6.1, where we noted that the detection thresholds are hardly affected by N.

5.1.2 Numerical results

In this section, we evaluate the performance of the proposed constellation design method in terms of error probability. First, we compare our approach with the KLD-based optimization proposed in [42], analyzing the impact of the number of antennas and channel correlation on the system performance. Subsequently, we examine the practical viability of the model by assessing its robustness against SNR estimation mismatches, as well as its capability to support the use of variable constellation sizes depending on the SNR level and target SER.

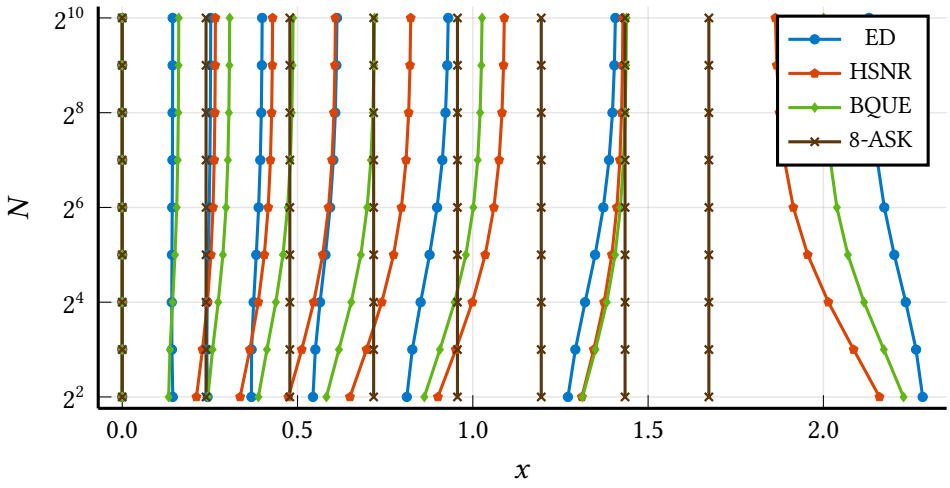


Figure 5.2: Comparison of 8-ASK and ED, HSNR and BQUE optimized constellations at a SNR equal to 15 dB and $\rho = 0.7$ for different numbers of antennas.

Comparison with KLD-based constellation designs

We devote this section to compare the detection performance of constellations designed by minimizing the asymptotic average SER, as proposed in section 5.1.1, with similar constellations available in the literature. Particularly, we consider the design proposed in [42], as the authors also rely on a large number of antennas and it explicitly accounts for correlated fading.

We compare the energy, HSNR, BQUE and ML detectors in two different scenarios, summarized in table 5.1. First, the ED, HSNR and BQUE are employed to detect their respectively optimized constellations. The ML detector is not of the form (4.19), so it does not admit a constellation tailored with our presented method. Instead, we propose to use it with the BQUE-optimized constellation, which serves as the baseline for quadratic detectors. Solid lines are employed to illustrate this configuration. Second, we use the four receivers with the KLD-optimized constellation from [42], which is depicted with dashed lines.

Table 5.1: Optimized constellations simulation setup and results.

Detector	Solid line constellation	Dashed line constellation	Best constellation
ED	ED-SER	KLD	ED-SER
HSNR	HSNR-SER	KLD	HSNR-SER
BQUE	BQUE-SER	KLD	BQUE-SER/KLD
ML	BQUE-SER	KLD	BQUE-SER

In figure 5.3, we depict the average SER of the previous setup as a function of SNR under model (2.48). It can be observed that the HSNR and ED performance is always better when the transmitted symbol belongs to a constellation optimized as proposed herein (solid lines). Regarding BQUE, its error probability coincides with both constellation designs. Notably, the ML detector performs better with the BQUE constellation than with the KLD-optimized

constellation from [42], for any SNR level. It is also worth pointing out the absence of error floor at high SNR due to the statistical CSIT, contrary to the results in the previous chapter, where the transmitter had no knowledge of the channel.

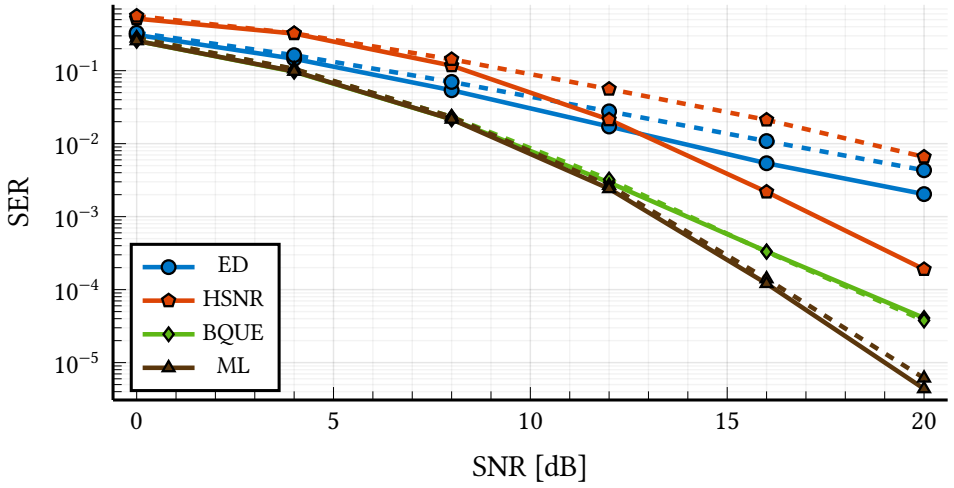


Figure 5.3: Error probability of the ED, HSNR, BQUE, and ML detectors using KLD-based (dashed lines) and SER-optimized (solid lines) constellations at different SNR levels, for $N = 128$ and $\rho = 0.7$.

Regarding the behavior of the systems for different number of antennas, observing figure 5.4 we can claim the superiority of constellations designed optimizing (4.54) instead of the KLD. This result is consistent with the fact that the design method proposed in [42] relies on the optimization of the KLD, the best achievable error exponent [124, Sec. 3.8], instead of the SER. The effect of channel hardening is also observed in the figure, as the error probability decreases when the number of antennas increases. Note that since the SNR is invariant to N, the observed gain is caused by channel hardening.

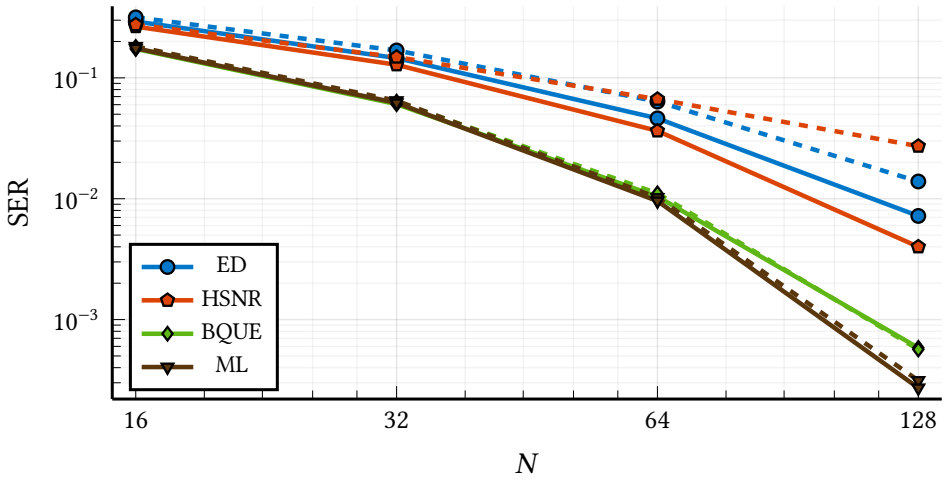


Figure 5.4: Error probability of the ED, HSNR, BQUE, and ML detectors using KLD-based (dashed lines) and SER-optimized (solid lines) constellations for different number of antennas with $\rho = 0.7$ at a SNR of 15 dB.

Finally, in figure 5.5, we show how channel correlation influences the system performance. In general, the method proposed in this work outperforms KLD optimization for each detector considered at every correlation regime. This effect is more pronounced for moderate channel correlation. Concerning the gap present at $\rho = 0$, it was also observed in the previous chapter and it is caused by the Gaussian approximation of $\hat{\varepsilon}(\mathbf{r}|\varepsilon)$. In particular, quadratic detectors classify the transmitted symbol by an estimate-and-detect procedure, which implies using thresholds. As stated before, the optimal thresholds are located at the intersection of two chi-squared densities (in the isotropic scenario), but we are computing them at the intersection of two normal densities instead. On the other hand, the ML detector performs an exhaustive search over all possible symbols, so it does not need any thresholds.

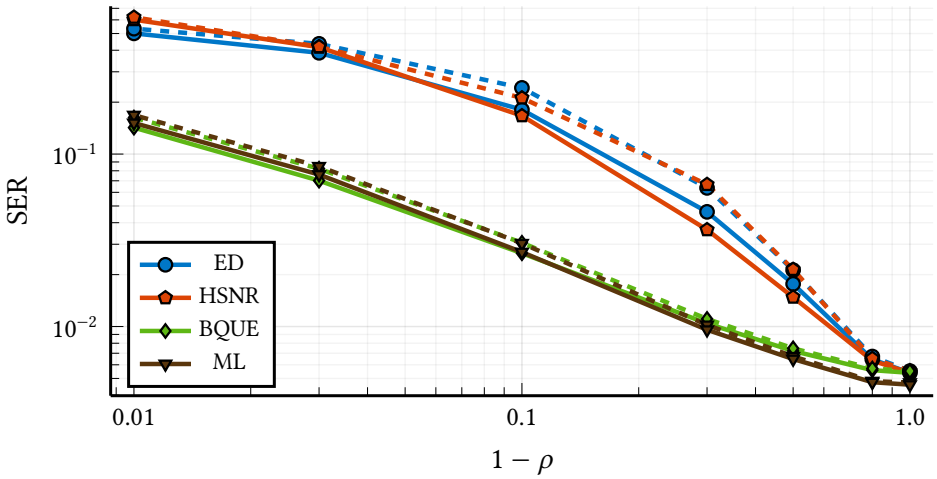


Figure 5.5: Error probability of the ED, HSNR, BQUE, and ML detectors using KLD-based (dashed lines) and SER-optimized (solid lines) constellations for different ρ with $N = 64$ at a SNR of 15 dB.

Constellation size and SNR mismatch analysis

The primary result from the previous section is that optimizing the constellation for each SNR level eliminates the error floor. This enables the implementation of schemes similar to *adaptive modulation and coding (AMC)*, where the constellation size is increased at high SNR to maximize the data rate while adhering to a target error probability [149], [150].

To illustrate this capability, figure 5.6 shows the SER as a function of the SNR for four constellation sizes: $M \in \{4, 8, 12, 16\}$. Because the error floor is absent, a specific SER target can be established, *i.e.* 10^{-4}. As shown in the figure, the system could dynamically switch constellations to maximize throughput: employing $M = 4$ when $7\,\text{dB} \leq \text{SNR} < 18\,\text{dB}$, $M = 8$ when $18\,\text{dB} \leq \text{SNR} < 29\,\text{dB}$, $M = 12$ for $\text{SNR} \geq 29\,\text{dB}$, etc.

In practical scenarios, precise SNR information may not be available at the transmitter [151], [152]. Consequently, it is essential to analyze the robustness of the designed constellations under conditions of imperfect SNR knowledge. To this end, figure 5.7 illustrates the SER performance when the constellation is optimized for a fixed value, denoted as $\alpha_{\text{opt}} \in \{5\,\text{dB}, 10\,\text{dB}, 15\,\text{dB}\}$. As shown, each curve tracks the performance of the constellation optimized to each SNR level (represented by the dashed line) near its design point before encountering an error floor at higher SNR values. Despite the emergence of these floors, the system exhibits a graceful degradation, proving the robustness of the proposed method.

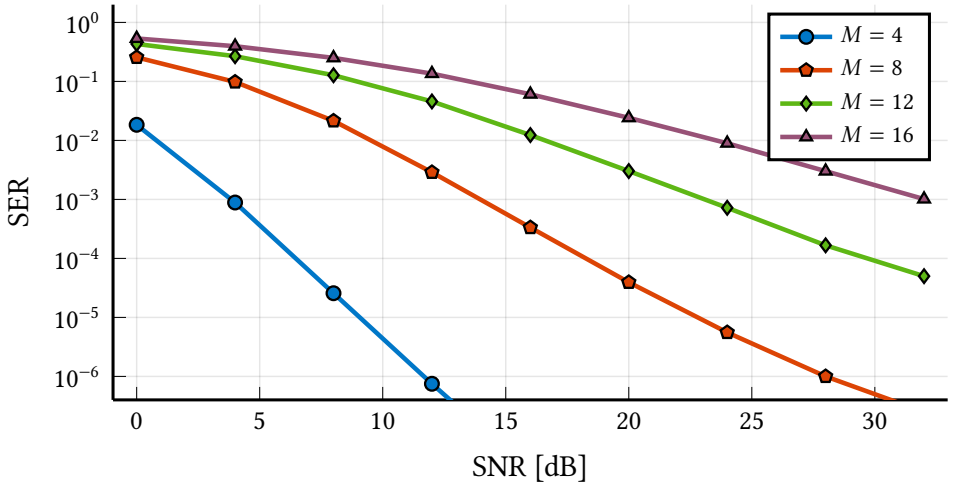


Figure 5.6: Error probability of the BQUE detector using its optimized constellation with $M \in \{4, 8, 12, 16\}$ symbols as a function of the SNR.

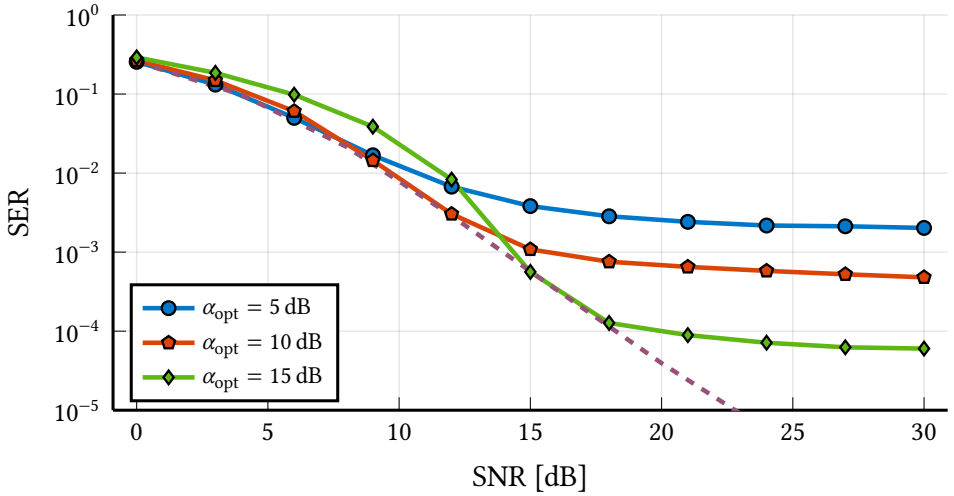


Figure 5.7: Error probability of the BQUE detector when the constellation is optimized for a particular SNR, denoted by α_{opt}. The SER achieved by the constellation optimized at each SNR level is also shown with a dashed line.

5.2 Permutational index modulation

The results obtained so far for one-shot[2] noncoherent communications—primarily those in sections 4.5 and 4.6—can be exploited to perform detection in more complex multiband systems. Since transmitted energy can be reliably detected, it becomes possible to convey information in the energy of each subcarrier of an OFDM system. Transmitting information in this manner is known as *index modulation (IM)* [153], [154]. Other forms of IM include spatial modulation, where information is encoded in the set of active or inactive antennas [155].

[2] In the context of this section, the term one-shot refers to a channel that remains constant over the duration of a single OFDM symbol.

The scheme proposed in this section can, on the one hand, be regarded as a generalization of OFDM-IM [156]. In that framework, the index corresponds to the subcarrier, which is traditionally either active or inactive. However, the SE achieved by relying on only two energy levels is limited, which has motivated significant research in recent years aimed at improving it through the use of multi-level constellations [154]. Most of these contributions have been developed in the context of coherent communications, with a few notable exceptions [139], [157].

An appealing alternative is to design the set of energy levels on each subcarrier based on a permutation of a predefined alphabet. The resulting constellation exhibits the structure of a *Variant-I permutation modulation (PM)* [158], [159], originally introduced for the Gaussian channel. Although several works have extended this concept to noncoherent fading scenarios [139], [157], to the best of our knowledge, none accounts for antenna correlation, and only [139] considers multi-level modulation.

Building upon the framework established in chapter 4, we propose a low-complexity detection scheme for multi-level OFDM-IM based on the ordering of quadratic statistics of the received signal. The proposed method leverages statistical CSI to reliably detect transmitted codewords, even in the presence of correlated fading.

5.2.1 Signal model

The system model considered here extends that of chapter 4 to a multicarrier setting. Specifically, we study a massive SIMO point-to-point system with N antennas at the receiver which uses K OFDM subcarriers. Communication is one-shot and through a fading channel that remains constant for a single OFDM block. At every band k, fading follows a correlated Rayleigh model $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_{\mathbf{h}_k})$, with $\mathrm{E}\big[\mathbf{h}_k \mathbf{h}_{k'}^{\mathrm{H}}\big] = \mathbf{0}_{N\times N}$, $\forall k \neq k'$. That is, fading is assumed to be independent among different frequency bands, which is usually achieved with techniques such as frequency interleaving [160]. The transmitter sends a codeword $\mathbf{x} = (x_1, \dots, x_K)^{\mathrm{T}} \in \mathbb{C}^K$ spanning the full K frequency bands: x_k is sent through $\mathbf{h}_k$.

The signal at the receiver in band k is modeled as[3]

$$\mathbf{y}_k = \mathbf{h}_k x_k + \mathbf{z}_k, \quad k = 1, \dots, K, \tag{5.2}$$

being $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}_N, P_z \mathbf{I}_N)$ AWGN. We consider a receiver with statistical CSI, so it is aware of the distributions of $\mathbf{h}_k$ and $\mathbf{z}_k$ but not of their realizations.

In PM [158], each codeword $\mathbf{x}$ in an alphabet $\mathscr{A}$ is constructed by permuting the elements of a reference vector $\underline{\mathbf{x}} = (\underline{x}_1, \dots, \underline{x}_K)^{\mathrm{T}}$. The set of permutations of $\underline{\mathbf{x}}$ (*i.e.* our codebook) is denoted as $\mathscr{A} = \Sigma_K(\underline{\mathbf{x}})$. The successive components of $\underline{\mathbf{x}}$ are the amplitudes $\sqrt{\varepsilon_m}$ from $\mathscr{X}$ in (4.7), in ascending order, each repeated K_m times:

$$\underline{\mathbf{x}} = \Big(\underbrace{0, \dots, 0}_{K_1}, \underbrace{\sqrt{\varepsilon_2}, \dots, \sqrt{\varepsilon_2}}_{K_2}, \dots, \underbrace{\sqrt{\varepsilon_M}, \dots, \sqrt{\varepsilon_M}}_{K_M} \Big)^{\mathrm{T}}. \tag{5.3}$$

Note that $\sum_{m=1}^{M} K_m = K$. From the point of view of IM, K_1 corresponds to the portion of inactive subcarriers in the OFDM symbol [154, Sec. 1.2.3].

[3] Note that the same model would be obtained for a system with K independent temporal channel uses. All the derivations considered herein can be applied in such scenario straightforwardly.

This scheme yields a codebook with cardinality [158]

$$|\mathscr{A}| = \frac{K!}{\prod_{m=1}^{M} K_m!}. \tag{5.4}$$

The redundancy and error correcting capabilities introduced by PIM (for finite K and M) are evidenced by comparing $|\mathscr{A}|$ with M^K, which is the total amount of possible unstructured codewords. From (5.4), the SE of the alphabet can be computed in bits per channel use [bpcu]:

$$\mathrm{R}_K(\{K_m\}) = \frac{1}{K}\log|\mathscr{A}| = \frac{1}{K}\big(\log(K!) - \sum_{m=1}^{M}\log(K_m!)\big). \tag{5.5}$$

The code rate is defined as $\mathrm{R}_K(\{K_m\})/\log M$.

It was proved in [161] that $\lim_{K\to\infty}\mathrm{R}_K(\{K_m\}) = \mathrm{H}(\{p_m\})$ for PM, where $\mathrm{H}(\cdot)$ is the Shannon entropy and $p_m = K_m/K$, for $m = 1,\dots,M$. It then follows that $\mathrm{R}_K(\{K_m\})$ is asymptotically maximized under a uniform policy of $p_m = 1/M$. An alternative proof for the asymptotic performance is given in appendix F, in which it is shown that SE approaches $\mathrm{H}(\{p_m\})$ from below. Additionally, it is also proved that uniform policy is optimal (in terms of SE) in the non-asymptotic case as well. This yields an asymptotic code rate of $\mathrm{H}(\{p_m\})/\log M \leq 1$, which means the uniform policy does not provide any redundancy for large K.

5.2.2 General maximum likelihood detection

Given an equiprobable alphabet, it is known that the probability of incorrectly detecting a transmitted codeword is minimized by the ML detector, *i.e.* the one that maximizes the likelihood function over all possible transmitted codewords. In our setting, it results in

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathbf{x}\in\mathscr{A}} \sum_{k=1}^{K} \mathbf{y}_k^{\mathrm{H}}\mathbf{C}_{\mathbf{y}|x,k}^{-1}\mathbf{y}_k + \ln\big|\mathbf{C}_{\mathbf{y}|x,k}\big|. \tag{5.6}$$

In general, detecting a single codeword involves $O(|\mathscr{A}|KN^2)$ operations to compute the statistics for every hypothesis $\mathbf{x}$ and $O(|\mathscr{A}|)$ additional ones to find the minimum one. This complexity may be reduced by applying the same pre-processing layer as in (4.9) to the received signals $\{\mathbf{y}_k\}$. Let $\mathbf{C}_{\mathbf{h}_k} = \mathbf{U}_k\boldsymbol{\Gamma}_k\mathbf{U}_k^{\mathrm{H}}$ be the eigendecomposition of the channel covariance matrix associated with band k. By working with $\{\mathbf{r}_k = \mathbf{U}_k^{\mathrm{H}}\mathbf{y}_k\}$, which are now distributed as $\mathbf{r}_k|x_k \sim \mathscr{CN}(\mathbf{0}_N, |x_k|^2\boldsymbol{\Gamma}_k + P_{\mathrm{z}}\mathbf{I}_N)$, the ML detection problem reduces to

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathbf{x}\in\mathscr{A}} \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{|[\mathbf{r}_k]_n|^2}{|x_k|^2\gamma_{n,k} + P_{\mathrm{z}}} + \ln\big(|x_k|^2\gamma_{n,k} + P_{\mathrm{z}}\big), \tag{5.7}$$

where $\gamma_{n,k} = [\boldsymbol{\Gamma}_k]_{n,n}$. This expression generalizes the one used in (4.10). Detection with (5.7) is equivalent to (5.6), since there is a one-to-one correspondence between $\mathbf{r}_k$ and $\mathbf{y}_k$.

Applying this pre-processing has an added complexity of $O(KN^2)$, but now computing the statistics requires $O(|\mathscr{A}|KN)$ operations.[4] Clearly, the complexity of directly evaluating (5.7)

[4] Diagonalizing $\{\mathbf{C}_{\mathbf{h}_k}\}$ does not incur significant computational load since fading statistics remain stable over many channel uses [38, Sec. 2.2].

scales proportionally to $|\mathscr{A}|$. Nonetheless, this task can be simplified further by applying the Viterbi algorithm in a procedure similar to the one described in [162]. Unfortunately, the total number of states in the trellis representation is $\prod_{m=1}^{M}(K_m+1)$, which greatly penalizes schemes constructed over large PAM constellations. Therefore, true ML detection of PIM might become infeasible for large K and M.

Some special cases in which ML detection admits low-complexity implementations are now considered.

5.2.3 Isotropic channel detection

Under the simplest channel correlation model (*i.e.* the isotropic channel), each frequency band is distributed as $\mathbf{h}_k \sim \mathscr{CN}(\mathbf{0}_N, \mathbf{I}_N)$. In this case, the resulting ML detector simplifies to

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\sigma\in\Sigma_K}{\arg\min} \sum_{k=1}^{K} \frac{\|\mathbf{r}_k\|^2}{|\underline{x}_{\sigma(k)}|^2 + P_{\mathrm{z}}} + N\ln(|\underline{x}_{\sigma(k)}|^2 + P_{\mathrm{z}}), \tag{5.8}$$

where σ is an index permutation. The logarithmic term is common for all codewords in the alphabet, since they are permutations of the same elements. Therefore, it can be removed from the minimization problem, leaving

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\sigma\in\Sigma_K}{\arg\min} \sum_{k=1}^{K} u_k v_{\sigma(k)} \equiv \underset{\sigma\in\Sigma_K}{\arg\min} \sum_{k=1}^{K} u_{\sigma(k)} v_k, \tag{5.9}$$

where $u_k = \|\mathbf{r}_k\|^2 \equiv \|\mathbf{y}_k\|^2$ and $v_k = (|\underline{x}_k|^2 + P_{\mathrm{z}})^{-1}$. This means that ML detection simplifies to finding the permutation $\sigma \in \Sigma_K$ that minimizes the sum $\sum_{k=1}^{K} u_{\sigma(k)} v_k$.

This problem can be solved via the rearrangement inequality [163, Proposition 6.A.3]. Consider $\{v_k\}$ are arranged in decreasing order, *i.e.* $v_1 > v_2 > \cdots > v_K$. Then, the minimum sum is obtained with the permutation that yields the terms $\{u_k\}$ in increasing order. This implies ML detection is equivalent to pairing the most energetic symbols with the most energetic received signal per band. More formally, it is expressed as

$$\varsigma = \underset{\sigma\in\Sigma_K}{\arg\operatorname{sort}} \{\|\mathbf{y}_k\|^2\}, \tag{5.10}$$

where arg sort is a function that sorts the values of $\{\|\mathbf{y}_k\|^2\}$ and returns ς, the permutation that achieves such order. The ML decision is then obtained as

$$\begin{array}{c} \{\, \|\mathbf{y}_{\varsigma(1)}\|^2 < \|\mathbf{y}_{\varsigma(2)}\|^2 < \ldots < \|\mathbf{y}_{\varsigma(K)}\|^2 \,\} \\ \Updownarrow \qquad\quad \Updownarrow \qquad\quad \Updownarrow \\ \hat{\mathbf{x}}_{\mathrm{ML}}^{\mathrm{T}} = (\ \underline{x}_{\varsigma^{-1}(1)}\ ,\ \underline{x}_{\varsigma^{-1}(2)}\ ,\ \ldots\ ,\ \underline{x}_{\varsigma^{-1}(K)}\ ). \end{array} \tag{5.11}$$

An analogous outcome was observed in [158] for the coherent detection of PM over the AWGN channel. The results in this section (namely, (5.10) and (5.11)) extend it to the noncoherent setting under isotropic fading. Using standard sorting algorithms, this procedure involves $O(KN) + O(K\ln K)$ operations, which is a significant reduction compared to the complexity of the general ML implementation.

It is worth noting that ML detection with (5.10) and (5.11) yields an additional benefit: it can assess decision reliability at almost no extra cost. Such information can be quantified by comparing the likelihood of the most likely decision, $\hat{\mathbf{x}}_{\mathrm{ML}}$, and the second most probable one, $\hat{\mathbf{x}}_2$, whose permutation is ς_2:

$$\mathrm{LLR} = \ln \frac{\mathrm{P}(\mathbf{x} = \hat{\mathbf{x}}_{\mathrm{ML}}|\mathbf{Y})}{\mathrm{P}(\mathbf{x} = \hat{\mathbf{x}}_2|\mathbf{Y})} = \sum_{k=1}^{K} (u_{\varsigma_2(k)} - u_{\varsigma(k)}) v_k. \tag{5.12}$$

These two codewords only differ in a single permutation of two symbols of consecutive power. This allows a very simple computation of (5.12):

$$\mathrm{LLR} = \min_{m \in \mathscr{M}} (u_{\varsigma(m+1)} - u_{\varsigma(m)})(v_m - v_{m+1}), \tag{5.13}$$

where $\mathscr{M} = \{\sum_{i=1}^{j} K_i,\ j = 1, \dots, M-1\}$.

5.2.4 Detection at high SNR

The previous low-complexity ML detector is only applicable under isotropic fading, which is an unlikely scenario when using large arrays [52]. In the general case (5.7), ML detection does not admit a simplified implementation based on sorting. A notable exception arises when the SNR grows without bounds, defining

$$\mathrm{SNR} = \frac{\sum_k \mathrm{E}\left[|\mathrm{x}_k|^2\right] \mathrm{E}\left[\|\mathbf{h}_k\|^2\right]}{\sum_k \mathrm{E}\left[\|\mathbf{z}_k\|^2\right]}, \tag{5.14}$$

which is a generalization of (4.2) to the multicarrier scenario.

Assuming the receiver is aware of the indices of inactive subcarriers (*i.e.* the ones with transmit power $\varepsilon_1 = 0$), which form set $\mathscr{K}_1$, we can study the limit of (5.7) when noise power vanishes:

$$\lim_{P_{\mathrm{z}} \to 0} \hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathbf{x} \in \mathscr{A}} \sum_{k \notin \mathscr{K}_1} \frac{1}{|x_k|^2} \mathbf{r}_k^{\mathrm{H}} \boldsymbol{\Gamma}_k^{-1} \mathbf{r}_k + \ln \left| |x_k|^2 \boldsymbol{\Gamma}_k \right|. \tag{5.15}$$

Once again, the logarithmic terms are removed from the minimization since they are common to all codewords. We are left with a problem similar to the one encountered in section 5.2.3. Following the same rationale, ML detection now consists in pairing the most energetic symbols with the highest values of metrics $\{\mathbf{r}_k^{\mathrm{H}} \boldsymbol{\Gamma}_k^{-1} \mathbf{r}_k\}$, which are quadratic statistics of data.

This simplification accounts for channel correlation but is only optimal in the high SNR regime and assuming inactive subcarriers have been reliably detected. Nevertheless, such a detector can be used suboptimally under any SNR and to detect every subcarrier, regardless of whether they are active or not. This sets the basis for the family of detectors introduced next.

5.2.5 General low-complexity detection

The previous analyses provide valuable insights about the ML detection problem (5.7). They motivate the derivation of a low-complexity detector based on the ordering of some relevant metrics $\{\hat{\beta}_k(\mathbf{r}_k)\}$ obtained from the received signal at each subcarrier k:

$$\tilde{\varsigma} = \arg\operatorname{sort}_{\sigma \in \Sigma_K} \{\hat{\beta}_k(\mathbf{r}_k)\}. \tag{5.16}$$

More precisely, we propose to use the framework developed in chapter 4, in which we devised a family of energy estimators based on quadratic statistics of data. While suboptimal in general, when properly designed, these methods can effectively leverage statistical CSI to provide accurate estimations of the transmitted energy in each subcarrier. Particularized to our setting, they take the form $\hat{\beta}_k(\mathbf{r}_k) = \mathbf{r}_k^{\mathrm{H}}\mathbf{A}_k\mathbf{r}_k - \mathrm{tr}(\mathbf{A}_k)$, which is the estimated energy level from subcarrier k and $\{\mathbf{A}_k\}$ are diagonal matrices that satisfy $\mathrm{tr}(\mathbf{A}_k\boldsymbol{\Gamma}_k) = 1$, so that estimation is conditionally unbiased.

As expected, the choice of weighting matrices $\{\mathbf{A}_k\}$ greatly impacts detection performance. Within the context of this work, we consider three designs from chapter 4:

1. **Energy detector (ED):** widely used in the literature for its simplicity and because it is optimal under the isotropic channel model from section 5.2.3 [121]. It is constructed with $\mathbf{A}_{\mathrm{ED},k} = \mathbf{I}_N/\mathrm{tr}(\boldsymbol{\Gamma}_k)$.
2. **High SNR (HSNR):** high SNR approximation of the ML detector for general fading corresponding to the case studied in section 5.2.4. Its associated weighting matrix is $\mathbf{A}_{\mathrm{HSNR},k} = \boldsymbol{\Gamma}_k^{-1}/N$.
3. **Assisted best quadratic unbiased estimator (ABQUE):** decision-directed implementation of the BQUE. The output $\{\hat{x}_k\}$ of a first hard decision from ED is used to compute the weighting matrices of ABQUE, defined as $\mathbf{A}_{\mathrm{ABQUE},k} = \boldsymbol{\Gamma}_k\mathbf{C}^{-2}_{\mathbf{r}|\hat{x},k} \Big/ \left\|\boldsymbol{\Gamma}_k\mathbf{C}^{-1}_{\mathbf{r}|\hat{x},k}\right\|_{\mathrm{F}}^2$. Then, a second detection round is performed on the received data with them.

The complexity of detecting a codeword with any of these schemes (without the preprocessing) is $O(KN) + O(K\ln K)$, which does not scale with $|\mathscr{A}|$, contrary to ML detection. This is especially relevant in the context of PM, whose alphabet size can grow as fast as factorially. A summary of detection complexity is given in table 5.2. Although the ED is less complex, its performance is several orders of magnitude worse than the other detectors, up to the point of being unusable in practice. Moreover, the HSNR and ABQUE detectors present the same complexity, but the latter's performance is superior, as it is shown in the following section.

Table 5.2: Summary of detection schemes complexity.

Detector	Complexity
ML	$O(KN^2) + O(\lvert\mathscr{A}\rvert KN)$
ED	$O(KN) + O(K\ln K)$
HSNR & ABQUE	$O(KN^2) + O(KN) + O(K\ln K)$

5.2.6 Numerical results

In this section the performance of the proposed low-complexity schemes against ML detection is assessed. A PIM with $K = 32$ and $M = 4$ under uniform policy is employed: $K_m = K/M = 8$, unless stated otherwise.[5] This configuration yields a SE of $\mathrm{R}_{32}(\{8\}) \approx 1.765$ bpcu, which

[5]These values are limited by the maximum ML complexity that can be simulated in a reasonably sized computation server. For instance, the number of trellis states required for the Viterbi algorithm for $K = 64$ surpasses $8 \cdot 10^5$.

amounts to a code rate of 0.8825 (the uncoded SE is 2 bpcu). Symbols are drawn from a uniform unipolar 4-ASK constellation with unit average power. While a tailored constellation and nonuniform partition are expected to yield better performance, using a standard constellation and uniform partition allows to focus the discussion on the detector design. Fading is assumed identically distributed across frequency bands (*i.e.* $\mathbf{C}_{\mathbf{h}_k} \equiv \mathbf{C}_{\mathbf{h}}$) under exponential correlation: $[\mathbf{C}_{\mathbf{h}}]_{m,n} = \rho^{|n-m|}$, being $0 \leq \rho < 1$ the correlation factor. The ML detector used as a benchmark is implemented with the Viterbi algorithm on a trellis with a total of 6561 states. Since bit coding is outside the scope, the metric considered herein is the average SER, *i.e.* the probability of incorrectly detecting each x_k.

In figure 5.8, the performance of all detectors is displayed in terms of SNR. A first examination reveals the fundamental error floor at high SNR discussed in chapter 4, for all four detectors. ED shows a higher error floor than the rest due to it being used under non-isotropic fading and not fully exploiting the channel correlation structures. HSNR suffers a severe performance degradation at low and moderate SNR (<12 dB), but reaches ML levels of SER otherwise, when the approximation upon which it is built becomes valid. The most noteworthy aspect of figure 5.8 is that ABQUE displays near-ML performance across all SNR levels in spite of its reduced complexity. It only suffers a slight penalty at very low SNR due to the poor performance of its first ED round.

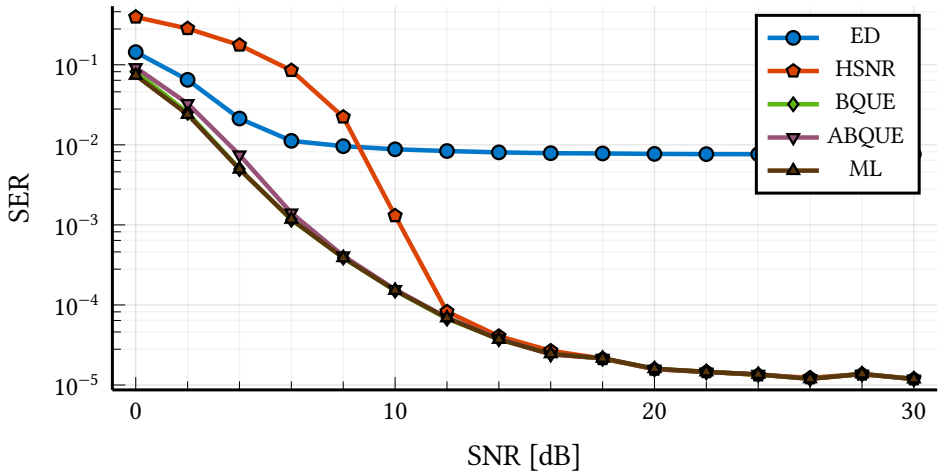


Figure 5.8: SER in terms of SNR for $N = 64$ and $\rho = 0.7$.

In figure 5.9, the minimum SNR required by each detector to achieve SER $= 10^{-3}$ is represented, for various numbers of antennas. The existence of the error floor previously discussed (which lowers for higher N values) prevents detectors from reaching the objective SER when N is below a certain threshold. Unsurprisingly, ED demands the largest N of the four. ABQUE detection reaches ML performance for $N \leq 64$, but requires a slightly higher SNR when employing more antennas. Similar performance gaps have been observed and discussed in chapter 4. Finally, the HSNR detector involves the highest SNR levels for large values of N.

Figure 5.10 depicts the impact of antenna correlation onto each detector. As expected, ED suffers a severe performance penalty with a small increase in ρ, becoming inoperative as soon as $\rho > 0.7$. The other three detectors perform similarly for $\rho \leq 0.99$, with graceful

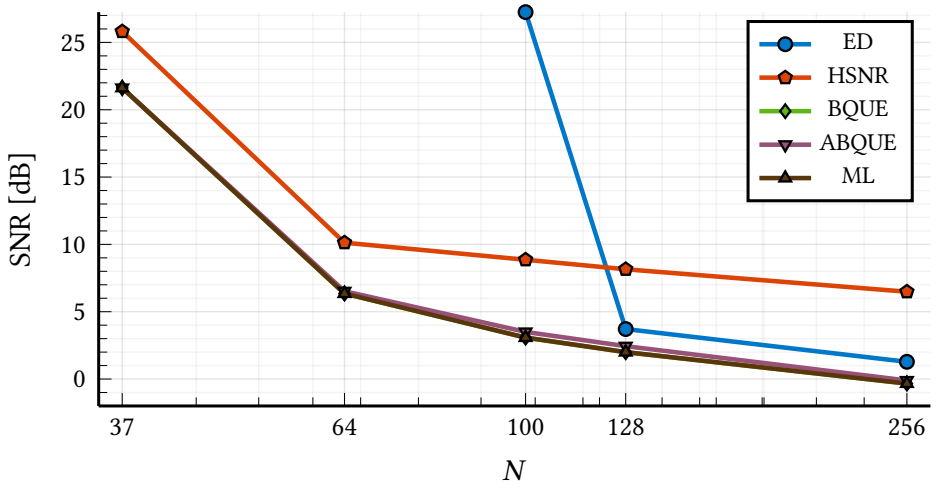


Figure 5.9: Required SNR to achieve SER $= 10^{-3}$ in terms of N, for $\rho = 0.7$.

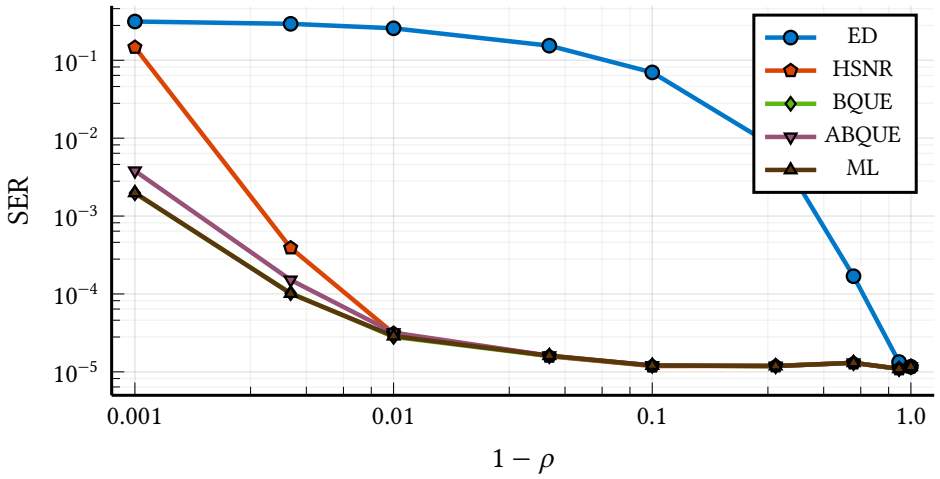


Figure 5.10: SER in terms of fading correlation for $N = 64$ and SNR $= 30$ dB.

degradation through this wide range of correlation regimes. When $\rho > 0.99$, the SER of HSNR starts to rapidly deteriorate. The performance loss for ML and ABQUE is slower and their associated SER values only diverge noticeably at extreme correlation values. Once again, a poor first ED round negatively impacts ABQUE.

Finally, in figure 5.11, we illustrate the trade-off between SE and robustness of the proposed PIM scheme in terms of K and partition policy. The SER corresponding to the ED is not provided since it is consistently outperformed by the other detectors. The one corresponding to the ML detector is omitted as well due to computational limitations, and instead the performance of the genie-aided BQUE detector is used as its benchmark (see section 4.5.2). As stated in section 5.2.1, SE increases with the number of subcarriers, at the cost of increased SER. Such trend is displayed with a dashed teal line for $M = 4$ and uniform policy. Ultimately, for $K \to \infty$, this modulation approaches the performance of the uncoded scheme.

In order to illustrate the potential advantages of using a nonuniform policy, in figure 5.11 we provide the SER/SE of an *ad hoc* code with $M = 4$ and $K = 30$, constructed from partition

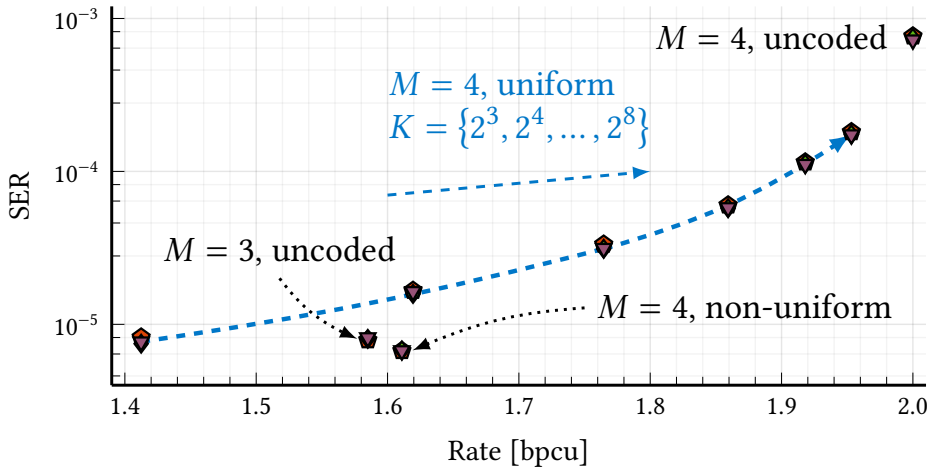


Figure 5.11: SER for various K and partitioning policies with $N = 64$, SNR $= 15$ dB and $\rho = 0.7$.

sizes $\{K_m\} = \{12, 9, 6, 3\}$. This simple code outperforms the uncoded scheme for $M = 3$ (both in terms of SER and of SE), and also the uniform partition schemes for $M = 4$, confirming the benefits of using the PIM scheme as a way of obtaining a coding gain [139]. Furthermore, if CSI is available at the transmitter, significant gains can be achieved by taking it into account in the constellation and partition design processes.

5.3 Conclusions

This chapter has investigated two strategies to enhance the robustness of one-shot communication systems based on energy detection.

The first strategy addressed the design of tailored constellations for low-complexity detection in massive SIMO systems. Assuming full statistical CSIT, we proposed minimizing the asymptotic average SER derived in the previous chapter for three quadratic receivers: the ED, the HSNR, and the BQUE. This approach successfully eliminates the fundamental error floor at high SNR. However, it requires statistical CSIT and different constellation sets optimized for each SNR regime, which increases feedback and computational overhead.

The second strategy exploited PM to introduce coding gain and robustness in multicarrier systems. Specifically, we proposed a near-ML low-complexity detector for noncoherent OFDM-PIM, built upon the ABQUE framework presented in section 4.6.3. This approach does not incur significant additional complexity, but the reduction in SER is modest.

In summary, constellation design offers superior performance in terms of SER, at the expense of higher feedback and computational demands, whereas permutation-based schemes provide a more lightweight alternative with moderate robustness gains. Importantly, both strategies are not mutually exclusive. For instance, in section 5.2 we considered standard ASK energy levels for each subcarrier, but replacing them with optimized constellations could further reduce the overall error probability.

CHAPTER 6

Symbol detection analysis with model-based receivers

As discussed in chapter 2, coherent communication systems require channel estimation through pilot training before data transmission. Similarly, noncoherent systems leveraging statistical CSI also rely on channel estimation, but with the advantage that the acquired information remains valid for a longer period of time.

In state-of-the-art high-frequency MIMO systems, a common approach is to employ *model-based strategies* for acquiring CSI, both in coherent and noncoherent settings [23], [24], [117], [164], [165]. For example, consider a BS in a high-frequency mMIMO system communicating with a single-antenna UE. Since the number of antennas is large compared to the number of paths and physical parameters [8], [52], [166], the BS can adopt the parametric model in (2.42) and estimate only the parameters α_l, r_l, θ_l, and ϕ_l, instead of the full channel vector. The overall system performance then depends on both the accuracy of the parameter estimation and the suitability of the chosen propagation model.

Model-based methods involving the estimation of environment parameters with physical sense usually lead to better estimation quality, specially in the low SNR regime. Nonetheless, an important question remains as to whether such precise modeling is always necessary, or if relaxing the model assumptions could yield simpler yet sufficiently effective receiver designs. For instance, adopting a far-field array response for a near-field system or ignoring the effects of mutual coupling in a dense array is expected to result in performance degradation, so the model cannot be overly simplified.

In this chapter, we analyze the symbol detection performance of receivers that employ parametric models under mismatched assumptions. We consider two scenarios: first, a BS with densely deployed antennas affected by mutual coupling but unaware of it; and second, an energy-based multiuser noncoherent system operating in the near field. In the latter, by adopting a near-field propagation model, we show that such systems can in fact exploit the curvature of spherical wavefronts, even at distances beyond the classical Fraunhofer limit, revealing that the classical far-field approximation may significantly underestimate system performance. Furthermore, we show that large antenna arrays facilitate the multiplexing of multiple users even under noncoherent processing, while enabling near-optimal detection with low computational complexity.

6.1 Impact of mutual coupling on detection performance

In this first section, we leverage multiport communication theory to explore the implications of mutual coupling in MIMO systems with closely spaced antennas. Unlike previous works [77], [78], [83], we do not focus on the impact of coupling on SNR or SE but rather on symbol detection performance.

Furthermore, we consider not only coherent detection but also a noncoherent approach, as, to the best of the authors' knowledge, the impact of coupling on the latter has not been investigated. In both cases, we assess the robustness of detection against mismatched channel models. Specifically, we focus on a single-user uplink scenario where the receiver is unaware of mutual coupling.

Through numerical simulations, we evidence that while the noncoherent detector generally exhibits a higher error probability, it remains robust to coupling model mismatches due to its inability to exploit phase information. Furthermore, this robustness is analytically proven in the particular case of high SNR. In contrast, the coherent detector, though more accurate under ideal conditions, becomes unreliable when antenna spacing is too small, highlighting a fundamental trade-off in system design [167, Sec. 5.2.10].

6.1.1 System model

We consider the uplink of a communication system in NLoS propagation conditions, with a single-antenna transmitter and a receiver with N half-wavelength dipole antennas[1] arranged in a side-by-side configuration as a ULA deployed in the x axis. The size of the array is D, whereas the separation between antennas is $d = D/(N-1)$ and the wavelength is λ. Hence, the position in polar coordinates of the n-th antenna with respect to the origin is $\mathbf{u}_n = (nd, 0)^{\mathrm{T}}$, with $0 \leq n \leq N-1$. The discrete baseband representation of the received signal is:

$$\mathbf{y} = \mathbf{h}\mathsf{x} + \mathbf{z}, \quad \mathbf{y}, \mathbf{h}, \mathbf{z} \in \mathbb{C}^N, \tag{6.1}$$

where $\mathbf{h} \sim \mathscr{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{h})$ denotes the channel, $\mathbf{z} \sim \mathscr{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{z})$ is additive Gaussian noise and x is the transmitted symbol. Since the noncoherent detector cannot decode phase information, symbols belong to a unipolar PAM constellation $\mathscr{X} = \{0, x_2, \dots, x_M\}$.

The inter-element separation $d = D/(N-1)$ inevitably leads to mutual coupling when the number of antennas is large. Given that the system under study is SIMO we can assume that the unilateral approximation holds (see section 3.3). Accordingly, the system can be represented by the circuit in figure 6.1. The corresponding voltage input–output relation is given by

$$\mathbf{v}_\mathrm{L} = \mathbf{d}\mathsf{v}_\mathrm{G} + \mathbf{n}, \tag{6.2}$$

where $\mathbf{d}$ denotes the channel and $\mathbf{n} \sim \mathscr{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{n})$ is additive Gaussian noise with its covariance given by (3.35) as

$$\mathbf{C}_\mathbf{n} = \mathbf{Q}\Big(\mathbf{C}_\mathrm{REN} + \sigma_i^2 R_\mathrm{N}^2 \mathbf{I}_{N_\mathrm{R}} + \sigma_i^2 \mathbf{Z}_\mathrm{R}\mathbf{Z}_\mathrm{R}^\mathrm{H} - 2\sigma_i^2 R_\mathrm{N} \mathfrak{R}(\rho_\mathrm{N}^* \mathbf{Z}_\mathrm{R})\Big)\mathbf{Q}^\mathrm{H}, \tag{6.3}$$

[1] Most communication theory works on mutual coupling consider isotropic radiators [79], [80], [98] or Hertzian dipoles [19], but a more accurate model can be obtained with half-wavelength dipoles [83].

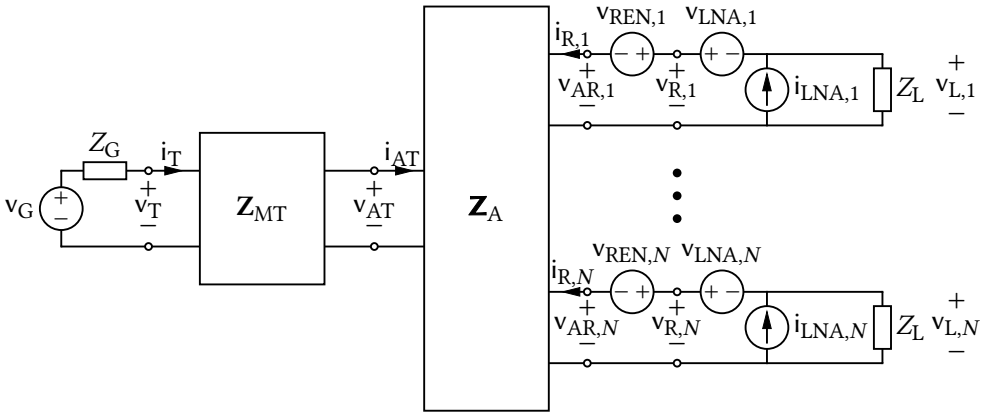


Figure 6.1: Circuit model of a SIMO communication system with power matching at the transmitter.

with $\mathbf{Q} = Z_L(Z_L\mathbf{I}_N + \mathbf{Z}_R)^{-1}$ and $\mathbf{C}_{REN} = 4k_B T_A B_W \Re(\mathbf{Z}_R)$, as defined in (3.12) and (3.24). Since the array elements are half-wavelength dipoles, $\mathbf{Z}_R$ is given by (3.61) and (3.62).

On the other hand, the channel vector is a particular case of (3.23) with $N_T = 1$. Taking into account that the transmitter is a single-antenna device, its matching network can be the optimal one given in (3.36), yielding

$$Z_T = [\mathbf{Z}_{MT}]_{1,1} - \frac{[\mathbf{Z}_{MT}]_{1,2}^2}{Z_{AT} + [\mathbf{Z}_{MT}]_{2,2}} = R_G - jX_G, \tag{6.4}$$

$$\mathbf{z}_{RT} = \frac{[\mathbf{Z}_{MT}]_{2,1}}{Z_{AT} + [\mathbf{Z}_{MT}]_{2,2}}\mathbf{z}_{ART} = -j\left(\frac{R_G}{R_r}\right)^{\frac{1}{2}}\mathbf{z}_{ART}. \tag{6.5}$$

Therefore,

$$\mathbf{d} = \frac{1}{Z_G + Z_T}\mathbf{Q}\mathbf{z}_{RT} = \frac{-j}{2\sqrt{R_G R_r}}\mathbf{Q}\mathbf{z}_{ART}, \tag{6.6}$$

with $Z_{AT} = 73.1 + 42.5j\,\Omega$ and $R_r = \Re(Z_{AT}) = 73.1\,\Omega$. Assuming that the number of scatterers is sufficiently large, it is fair to assume that $\mathbf{z}_{ART} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_{ART})$, with

$$\mathbf{C}_{ART} = R_r^2 \sum_{l=1}^{L} \beta_l \mathbf{a}_l \mathbf{a}_l^H, \tag{6.7}$$

which follows from (3.69).

Finally, recall from section 3.4 that, by setting $c = 1\,V^2$, we can identify

$$\mathbf{y} = \mathbf{v}_L, \qquad \mathbf{h} = \sqrt{\frac{R_G}{b}}\,\mathbf{d}, \qquad x = \sqrt{\frac{b}{R_G}}\,v_G, \qquad \mathbf{z} = \mathbf{n}. \tag{6.8}$$

Moreover, since the transmitter implements an optimal power matching network, $b = 1/4$. Assuming that the available transmit power is P_A, it follows that

$$E\left[|x|^2\right] = P_A, \tag{6.9}$$

$$\mathbf{C}_\mathbf{z} = \mathbf{C}_\mathbf{n}, \tag{6.10}$$

$$\mathbf{C}_\mathbf{h} = \frac{1}{R_r}\mathbf{Q}\mathbf{C}_{ART}\mathbf{Q}^H. \tag{6.11}$$

Recall that, in the absence of mutual coupling, $\mathbf{Z}_{\mathrm{R}} = (R_{\mathrm{r}} + \mathrm{j}X_{\mathrm{A}})\mathbf{I}_N$. Consequently, both $\mathbf{Q}$ and $\mathbf{C_z}$ are proportional to the identity matrix, and $\mathbf{C_h}$ is proportional to the spatial correlation matrix $\mathbf{C}_{\mathrm{ART}}$.

6.1.2 Receiver structure

The main goal of this section is to characterize the impact of ignoring antenna coupling in the detection stage. Both coherent and noncoherent approaches are considered herein. As stated in the chapter introduction, the BS implements a model-based acquisition of the channel realization and correlation. This means that, instead of estimating them directly, the receiver estimates their parameters and constructs $\mathbf{h}$ and $\mathbf{C_h}$ from them, employing the models (6.6) and (6.11).

We define the SER consistently with the definition employed in the previous chapters:

$$P_\epsilon = \frac{1}{M} \sum_{x \in \mathcal{X}} \mathrm{P}(\hat{x}(\mathbf{y}) \neq x \,|\, \mathrm{x} = x), \tag{6.12}$$

with $\hat{x}(\mathbf{y})$ the output of a symbol detector applied to $\mathbf{y}$.

In section 6.1.3, we evaluate the performance of the optimal coherent and optimal one-shot noncoherent detectors in three different scenarios: a matched detector that employs the correct signal model, a mismatched detector that accounts for the scatterers model and location but ignores mutual coupling (*i.e.* $\mathbf{Z}_{\mathrm{R}} = (R_{\mathrm{r}} + \mathrm{j}X_{\mathrm{A}})\mathbf{I}_N$), and a detector designed for an artificial uncoupled scenario where the signal is generated and detected with $\mathbf{Z}_{\mathrm{R}} = (R_{\mathrm{r}} + \mathrm{j}X_{\mathrm{A}})\mathbf{I}_N$.

Noncoherent detector

For a one-shot receiver unknowing of instantaneous channel realizations but only its statistics, the optimal symbol decision is given by the unconditional ML detector [20]:

$$\hat{x}_{\mathrm{NC}} = \arg\min_{x \in \mathcal{X}} \mathbf{y}^{\mathrm{H}} \mathbf{C}_{\mathbf{y}|x}^{-1} \mathbf{y} + \ln |\mathbf{C}_{\mathbf{y}|x}|, \tag{6.13}$$

where

$$\mathbf{C}_{\mathbf{y}|x} = |x|^2 \mathbf{C_h} + \mathbf{C_z}, \tag{6.14}$$

with $\mathbf{C_h}$ and $\mathbf{C_z}$ given by (6.10) and (6.11). The average SNR at the receiver is:

$$\mathrm{SNR} = \frac{\mathrm{E}\left[|\mathrm{x}|^2 \mathbf{h}^{\mathrm{H}} \mathbf{h}\right]}{\mathrm{E}\left[\mathbf{z}^{\mathrm{H}} \mathbf{z}\right]} = \frac{P_{\mathrm{A}} \mathrm{tr}(\mathbf{C_h})}{\mathrm{tr}(\mathbf{C_z})}. \tag{6.15}$$

If the detector is designed disregarding mutual coupling, the correlation matrix employed by the detector is not (6.14) but:

$$\widehat{\mathbf{C}}_{\mathbf{y}|x} = \gamma_1 |x|^2 \mathbf{C}_{\mathrm{ART}} + \gamma_2 \mathbf{I}_N, \tag{6.16}$$

with γ_1 and γ_2 power scaling factors that can be obtained from the multiport model setting $\mathbf{Z}_{\mathrm{R}} = (R_{\mathrm{r}} + \mathrm{j}X_{\mathrm{A}})\mathbf{I}_N$. Since γ_1 and γ_2 can be absorbed into the SNR, their explicit values are omitted.

At high SNR, and assuming $x \neq 0$, it follows from (6.1), (6.6) and (6.8) that the received signal can be written as

$$\mathbf{y}|x = \mathbf{Q} \underbrace{\frac{-\mathrm{j}}{\sqrt{\Re(Z_{\mathrm{AT}})}}}_{\alpha} x\mathbf{z}_{\mathrm{ART}} = \mathbf{Q} \underbrace{\alpha x \mathbf{z}_{\mathrm{ART}}}_{\mathbf{r}|x} = \mathbf{Q}\mathbf{r}|x, \tag{6.17}$$

where $\mathbf{r}|x$ denotes the "uncoupled" signal.

From the HSNR detector derived in section 4.4 we know that when SNR $\to \infty$, the logarithmic term in (6.13) can be neglected, so the Hermitian form is the only relevant statistic for ML detection. In terms of $\mathbf{r}|x$ it may be written as

$$\mathbf{y}^{\mathrm{H}}\mathbf{C}_{\mathbf{y}|x}^{-1}\mathbf{y} = \frac{\mathbf{r}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\left(\mathbf{Q}\mathbf{C}_{\mathrm{ART}}\mathbf{Q}^{\mathrm{H}}\right)^{-1}\mathbf{Q}\mathbf{r}}{|x|^2\beta} = \frac{\mathbf{r}^{\mathrm{H}}\mathbf{C}_{\mathrm{ART}}^{-1}\mathbf{r}}{|x|^2\beta} = \mathbf{r}^{\mathrm{H}}\mathbf{C}_{\mathbf{r}|x}^{-1}\mathbf{r}, \quad \beta = |\alpha|^2. \tag{6.18}$$

In other words, the optimal detectors of $\mathbf{y}|x$ and $\mathbf{r}|x$ are equivalent.

The conclusion is that, at the high SNR regime, mutual coupling does not have any impact on the performance of the energy detector. This result is illustrated through Monte Carlo experiments in section 6.1.3, together with the behavior at lower SNR levels.

Coherent detector

If full knowledge of every single channel realization is available at the receiver, the optimal decision is obtained with the maximal ratio combiner [46, Ch. 3]:

$$\hat{x}_{\mathrm{C}} = \arg\min_{x\in\mathcal{X}} \left| \frac{\Re[\mathbf{h}^{\mathrm{H}}\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{y}]}{\mathbf{h}^{\mathrm{H}}\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{h}} - x \right|. \tag{6.19}$$

Like in the noncoherent case, if it is assumed that the BS ignores mutual coupling, the detector operates in a mismatched mode with the channel and noise correlation given by:

$$\widehat{\mathbf{h}} = \sqrt{\gamma_1}\mathbf{z}_{\mathrm{ART}}, \tag{6.20}$$

$$\widehat{\mathbf{C}}_{\mathbf{z}} = \gamma_2 \mathbf{I}_N, \tag{6.21}$$

where γ_1 and γ_2 are the same scaling factors as before.

6.1.3 Numerical results

In order to assess the performance degradation of a receiver unaware of antenna mutual coupling, we consider a scenario with a single user located at $\mathbf{r} = (r, \phi)^{\mathrm{T}} = (25\,\mathrm{m}, -30^{\circ})^{\mathrm{T}}$, surrounded by $L = 20$ scatterers uniformly sampled from a circle centered at the user, with radius $r_{\mathrm{c}} = 3\,\mathrm{m}$, unless stated otherwise. Symbols are transmitted from a 4-PAM unipolar[2] constellation and the average SNR is 5 dB. The remaining parameters are summarized in table 6.1, with values from [83], [105] and references therein.

In general, the noncoherent detector exhibits a higher error probability but greater robustness to coupling mismatches, illustrating the classic trade-off between robustness and optimality.

Table 6.1: Mismatched coupling detection simulation parameters.

Parameter	Value	Parameter	Value
Carrier frequency	$f = 30\,\text{GHz}$	Variance of the current noise source	$\sigma_i^2 = 2k_\text{B}B_\text{W}T_\text{A}/R_\text{N}$
Bandwidth	$B_\text{W} = 20\,\text{MHz}$	Antenna impedance	$Z_\text{A} = 73 + \text{j}42.5\,\Omega$
Amplifier and load impedance	$Z_\text{G} = Z_\text{L} = 186 - \text{j}31.6\,\Omega$	Complex correlation coefficient	$\rho_\text{N} = 0.2730+\text{j}0.1793$
Noise temperature of antennas	$T_\text{A} = 290\,\text{K}$	LNA resistance	$R_\text{N} = 5\,\Omega$

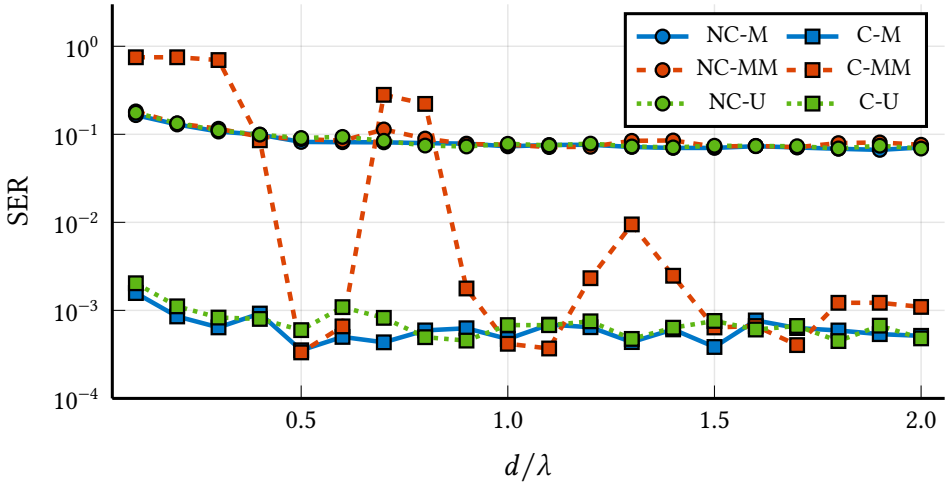


Figure 6.2: Error probability of the matched (M) and mismatched (MM), coherent (C) and noncoherent (NC), and uncoupled (U) detectors as a function of element separation for $N = 128$.

As shown in figure 6.2, the coherent mismatched detector SER mirrors the coupling function from section 3.2.1, whereas the energy detector remains nearly unaffected by the mismatch.

A similar behavior is illustrated in figure 6.3. When the array aperture is fixed but the number of antennas increases, the inter-element distance decreases and the coupling also increases. In this situation, the coherent mismatched receiver performance rapidly decays and becomes unusable. On the contrary, the noncoherent detector error probability is robust to the mismatch, and it even improves with N. In the simulated example, the noncoherent detector outperforms the coherent for $N \geq 128$.

It should also be remarked that the SER of a receiver that perfectly knows the effect of mutual coupling is practically equal to that of an uncoupled scenario, as can be observed by comparing the solid blue and dotted green lines in figure 6.2 and figure 6.3.

Although the previous results are valid in general, the exact performance is not independent of the user location [83]. In figure 6.4 it can be seen that, at end-fire, the coherent detector SER is lower than the noncoherent one, even for $N = 128$. Of course, if the number of antennas keeps increasing, the performance of the former will reduce whereas the latter's will improve, as discussed previously. Furthermore, the figure suggests that coupling primarily affects the response phase while leaving the power almost unchanged. Since the noncoherent detector relies solely on power, its performance remains unaffected by the mismatch.

[2] Although a coherent detector could employ a bipolar constellation with equal performance and lower power consumption, the system proposed herein allows for simultaneous coherent and noncoherent transmission.

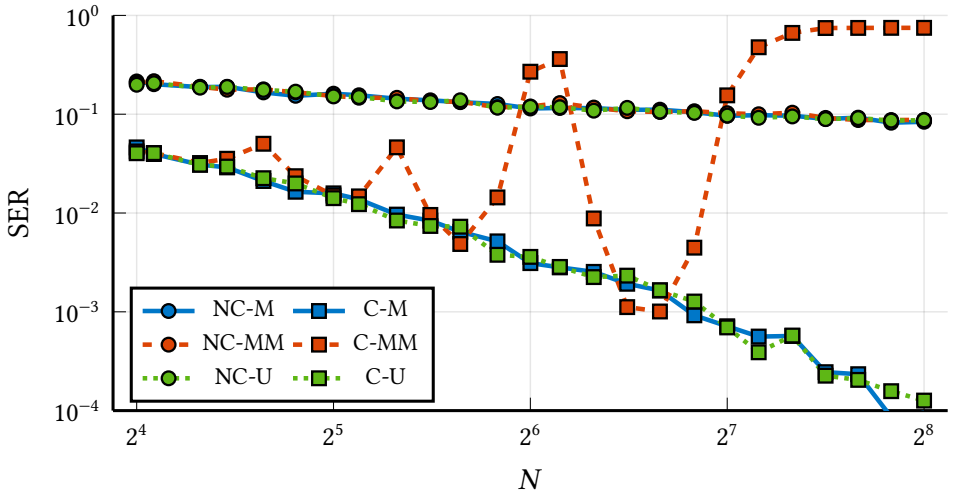


Figure 6.3: Error probability of the matched (M) and mismatched (MM), coherent (C) and noncoherent (NC), and uncoupled (U) detectors in terms of N for a fixed aperture $D = 0.5$ m.

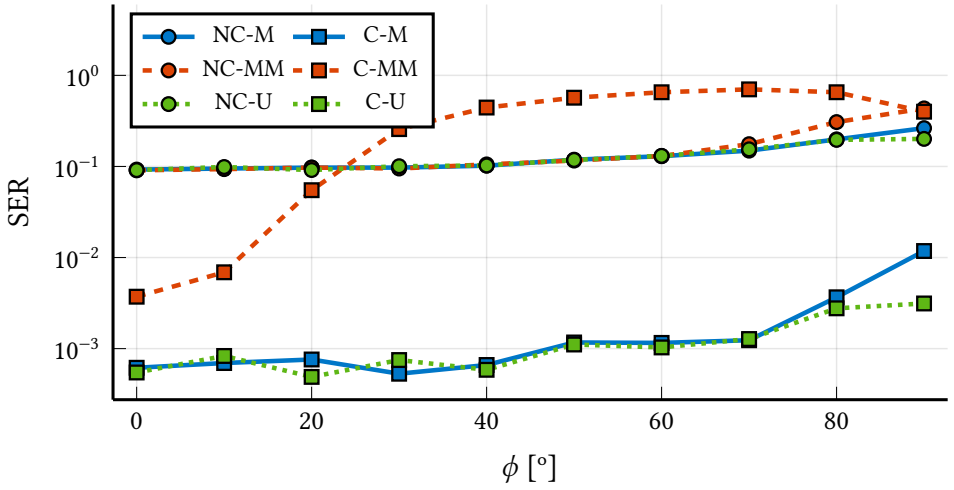


Figure 6.4: Error probability of the matched (M) and mismatched (MM), coherent (C) and noncoherent (NC), and uncoupled (U) detectors at azimuth angles from 0° to 90°. The array is $D = 0.5$ m and $N = 128$.

6.2 Leveraging wavefront curvature with noncoherent detection

In section 2.1 we showed that, at the Fraunhofer distance, the total phase error at an antenna due to the wavefront curvature is $\pi/8$ rad. From an antenna theory perspective, angles smaller than this value are usually neglected [30, Sec. 4.4.1], which is why the Fraunhofer distance has traditionally been considered the practical boundary separating the near and far field. This convention has been widely adopted in wireless system analysis, as it provides a convenient rule of thumb. But, what if the spherical wavefront is assumed even beyond the Fraunhofer distance?

Indeed, the authors in [164] showed that the classical far-field approximation underestimates SE in coherent multiuser MIMO, indicating that wavefront curvature can be exploited beyond the Fraunhofer distance. However, does this also hold for noncoherent communication systems?

The main goal of this section is demonstrating that the Fraunhofer distance neither is the appropriate boundary for the near and far fields in noncoherent communications. Furthermore, we show that multiuser noncoherent systems can leverage large arrays to multiplex different users and achieve near-optimal detection with lower complexity using a single-user detector. This suggests that simple constellations as those explored in [21] can be used in multiuser scenarios with minimal penalty if the number of antennas is sufficiently large [10].

6.2.1 System model

We consider a MIMO system with K active single-antenna UEs operating in NLoS propagation conditions. The BS is equipped with a UPA located in the yz plane, and consisting of N independent radio frequency chains, arranged into N_{V} rows, each with N_{H} antennas: $N = N_{\mathrm{H}}N_{\mathrm{V}}$, as depicted in figure 2.4b. Following the notation in section 2.2.1, the inter-element spacing across horizontal and vertical directions is d. The location in Cartesian coordinates of the n-th antenna, with $0 \leq n \leq N-1$, with respect to the origin is $\mathbf{u}_n = (0, i_n d, j_n d)^{\mathrm{T}}$. We call $\mathbf{h}_k \in \mathbb{C}^N$ the channel response between the k-th UE and the BS [37].

Since we are interested in the impact of wavefront curvature on noncoherent communications, we adopt the channel model in (2.61) with $N_{\mathrm{T}} = 1$ and $N_{\mathrm{R}} = N$. The array response vector is that given in (2.36). That is, $\mathbf{h}_k \sim \mathscr{CN}\left(\mathbf{0}_N, \mathbf{C}_{\mathbf{h}_k}\right)$, with covariance matrix

$$\mathbf{C}_{\mathbf{h}_k} = \mathrm{E}\left[\mathbf{h}_k \mathbf{h}_k^{\mathrm{H}}\right] = \sum_{l=1}^{L} \beta_{kl} \bar{\mathbf{a}}_{kl} \bar{\mathbf{a}}_{kl}^{\mathrm{H}}. \tag{6.22}$$

If the k-th UE is located in the far field of the array, then the channel can be approximated as $\mathbf{h}_k^{\mathrm{FF}} \sim \mathscr{CN}\left(\mathbf{0}_N, \mathbf{C}_{\mathbf{h}_k}^{\mathrm{FF}}\right)$, with $\mathbf{C}_{\mathbf{h}_k}^{\mathrm{FF}}$ given by

$$\mathbf{C}_{\mathbf{h}_k}^{\mathrm{FF}} = \sum_{l=1}^{L} \beta_{kl} \bar{\mathbf{c}}_{kl} \bar{\mathbf{c}}_{kl}^{\mathrm{H}}, \tag{6.23}$$

where $\bar{\mathbf{c}}_{kl}$ is the plane-wave approximation array response defined in (2.38).

Signal model

In this section, we extend the analysis from chapter 4 to a multiuser scenario. Specifically, we focus on the uplink of a one-shot communication system in which each user transmits symbols drawn from an equiprobable M-PAM constellation $\mathscr{X} = \{x_1, \ldots, x_M\}$ with $M \geq 2$, subject to an average power constraint $\mathrm{E}\left[|\mathsf{x}|^2\right] = 1$. The complex baseband signal $\mathbf{y} \in \mathbb{C}^N$ received at the BS is then given by:

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{h}_k \sqrt{p_k} \mathsf{x}_k + \mathbf{z}, \tag{6.24}$$

where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{z})$ is additive Gaussian noise, x_k is the PAM symbol transmitted by the k-th UE and p_k is a power control factor such that at the BS all signals have the same signal-to-interference-plus-noise ratio (SINR),

$$\mathrm{SINR} = \frac{p_k \mathrm{tr}\left(\mathbf{C}_{\mathbf{h}_k}\right)}{\mathrm{tr}(\mathbf{C}_\mathbf{z}) + \sum_{j \neq k} p_j \mathrm{tr}\left(\mathbf{C}_{\mathbf{h}_j}\right)}. \tag{6.25}$$

As with any noncoherent communication system, the BS is unaware of the realizations of $\{\mathbf{h}_k\}_{k=1}^K$. On the other hand, it may rely on their statistical properties.

Letting $\mathcal{X}^K$ be the K-ary Cartesian product of $\mathcal{X}$, the joint SER is:

$$P_\epsilon = \frac{1}{M^K} \sum_{\mathbf{x} \in \mathcal{X}^K} \Pr\big(\widehat{\mathbf{x}}(\mathbf{y}) \neq \mathbf{x} | \mathbf{x} = \mathbf{x}\big), \tag{6.26}$$

where $\widehat{\mathbf{x}}(\mathbf{y})$ is the output of a multiuser symbol detector applied to the received signal $\mathbf{y}$. In particular, the optimal $\widehat{\mathbf{x}}(\mathbf{y})$ (*i.e.* the one that minimizes P_ϵ) is the ML detector [67, Sec. 4.1-1]. This yields

$$\widehat{\mathbf{x}}(\mathbf{y}) = \widehat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathbf{x} \in \mathcal{X}^K} \mathbf{y}^{\mathrm{H}} \mathbf{C}_{\mathbf{y}|\mathbf{x}}^{-1} \mathbf{y} + \ln |\mathbf{C}_{\mathbf{y}|\mathbf{x}}|, \tag{6.27}$$

where

$$\mathbf{C}_{\mathbf{y}|\mathbf{x}} = (\mathbf{x}^{\mathrm{T}} \otimes \mathbf{I}_N) \mathbf{C}_{\mathrm{vec}(\mathbf{HP})} (\mathbf{x}^* \otimes \mathbf{I}_N) + \mathbf{C}_\mathbf{z}, \tag{6.28}$$

with $\mathbf{x} = (x_1, \ldots, x_K)^{\mathrm{T}}$, $\mathbf{H} \in \mathbb{C}^{N \times K} = (\mathbf{h}_1, \ldots, \mathbf{h}_K)$ and $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_K)$ the diagonal power control matrix.

The optimal decision rule (6.27) has a major drawback which is its complexity. Since it increases exponentially with K, its practical application is restricted to systems with a small number of UEs [67, Sec. 16.3–2]. A suboptimal receiver is the single-user detector that ignores the presence of other users. For the k-th UE, it takes the form:

$$\hat{x}_k = \arg\min_{x \in \mathcal{X}} \mathbf{y}^{\mathrm{H}} \mathbf{C}_{\mathbf{y}|x,k}^{-1} \mathbf{y} + \ln |\mathbf{C}_{\mathbf{y}|x,k}|, \tag{6.29}$$

where

$$\mathbf{C}_{\mathbf{y}|x,k} = |x|^2 p_k \mathbf{C}_{\mathbf{h}_k} + \mathbf{C}_\mathbf{z}. \tag{6.30}$$

Observe that this detector does not model the remaining users as noise, but rather ignores them completely.

In general, the performance gap between (6.27) and (6.29) is large. This is not, however, the case for systems with a large number of antennas (compared to the number of users) where the single-user detector achieves near-optimal performance with linear complexity, as we shall demonstrate in the sequel.

As outlined at the beginning of this chapter, our primary objective is to quantify the impact of mismatched receivers. In particular, we are interested in assessing whether the Fraunhofer distance provides a valid criterion for distinguishing between the near and far field. To this end, we evaluate the detection performance when the following covariance matrix is employed instead of (6.30):

$$\mathbf{C}_{\mathbf{y}|x,k}^{\mathrm{FF}} = |x|^2 p_k \mathbf{C}_{\mathbf{h}_k}^{\mathrm{FF}} + \mathbf{C}_\mathbf{z}. \tag{6.31}$$

Single-user asymptotic regime

Independently of the receiver considered, a necessary condition for an arbitrarily low P_ϵ in a noncoherent system is to have a uniquely identifiable constellation (see theorems 4.1 and 4.2). In the single-user scenario[3] it is defined by:

$$|x_a|^2 \neq |x_b|^2 \iff x_a \neq x_b, \quad \forall x_a, x_b \in \mathcal{X}.$$

Recall that, under this hypothesis, the error probability exhibits the following properties:

$$\lim_{N\to\infty} P_\epsilon = 0 \iff \lim_{N\to\infty} \operatorname{tr}\left(\mathbf{C}_\mathbf{h}^2\right) = \infty, \tag{6.32}$$

$$\lim_{p\to\infty} P_\epsilon = 0 \iff M = 2. \tag{6.33}$$

Although employing large arrays enables lower-complexity detection, the channel rank is still determined by the number of scatterers: $\operatorname{rank}(\mathbf{C}_\mathbf{h}) = L$, and it does not increase with N. Therefore, (6.32) is not fulfilled and the system must exhibit an error floor. Regarding (6.33), the interpretation is that, for a PAM of order $M = 2$, the zero symbol and the non-zero symbol span different subspaces and can be detected with arbitrarily low error probability. On the other hand, for $M > 2$ all the non-zero symbols share the same subspace and the only vanishing pairwise error probabilities are those between $x_1 = 0$ and any other symbol.

6.2.2 Orthogonality of channel correlation column spaces

A known result in the large-array literature is that near-field beamfocusing vectors are orthogonal when the number of antennas $N \to \infty$ and consequently the aperture $D \to \infty$ (see, for instance, [168], [169]). To the best of the authors' knowledge, such orthogonality has only been exploited in coherent communication systems. However, it can also be leveraged in noncoherent systems to multiplex different UEs, as we show in section 6.2.4.

In the following, we assume that the UE and the scatterers are located in a region delimited by $r \in (r_1, r_2)$, $\theta \in (\theta_1, \theta_2)$ and $\phi \in (\phi_1, \phi_2)$, where $\max(d_\mathrm{r}, d_\mathrm{UP}) < r_1 < r_2$, $0 < \theta_1 < \theta_2 < \pi$ and $0 < \phi_1 < \phi_2 < \pi$.

Theorem 6.1. Let $\overline{\mathbf{A}}_{\mathrm{ULA},1}, \overline{\mathbf{A}}_{\mathrm{ULA},2} \in \mathbb{C}^{N\times N}$ be the rank-L spatial correlation matrices given by (2.44) with the USW array response vector in (2.27), for two different scattering environments. Then, $\operatorname{col}\left(\overline{\mathbf{A}}_{\mathrm{ULA},1}\right) \perp \operatorname{col}\left(\overline{\mathbf{A}}_{\mathrm{ULA},2}\right)$ when $N \to \infty$.

Proof. Let $\overline{\mathbf{A}}_1 = \sum_{l=1}^{L} \beta_{1,l} \bar{\mathbf{a}}_{1,l} \bar{\mathbf{a}}_{1,l}^\mathrm{H}$ and $\overline{\mathbf{A}}_2 = \sum_{l=1}^{L} \beta_{2,l} \bar{\mathbf{a}}_{2,l} \bar{\mathbf{a}}_{2,l}^\mathrm{H}$ be the ULA near-field channel correlation matrices of two different scattering environments. The "ULA" subscript has been omitted for simplicity of notation.

To find a basis for $\operatorname{col}\left(\overline{\mathbf{A}}_1\right)$, observe that any $\mathbf{y} \in \operatorname{col}\left(\overline{\mathbf{A}}_1\right)$ can be written as:

$$\mathbf{y} = \overline{\mathbf{A}}_1 \mathbf{x} = \sum_{l=1}^{L} \beta_{1,l} \bar{\mathbf{a}}_{1,l} \bar{\mathbf{a}}_{1,l}^\mathrm{H} \mathbf{x} = \sum_{l=1}^{L} \gamma_l \bar{\mathbf{a}}_{1,l}, \quad \mathbf{x} \in \mathbb{C}^N, \tag{6.34}$$

[3] Note that if the condition is necessary for a single user it must also be necessary in the multiuser scenario.

and as by hypothesis $\operatorname{rank}(\overline{\mathbf{A}}_1) = L$, a basis for $\operatorname{col}(\overline{\mathbf{A}}_1)$ is $\{\overline{\mathbf{a}}_{1,l}\}_{l=1}^{L}$. Similarly, a basis for $\operatorname{col}(\overline{\mathbf{A}}_2)$ is $\{\overline{\mathbf{a}}_{2,l'}\}_{l'=1}^{L}$. Then,

$$\operatorname{col}(\overline{\mathbf{A}}_1) \perp \operatorname{col}(\overline{\mathbf{A}}_2) \iff \frac{\overline{\mathbf{a}}_{1,l}^{\mathrm{H}}\overline{\mathbf{a}}_{2,l'}}{\|\overline{\mathbf{a}}_{1,l}\|\|\overline{\mathbf{a}}_{2,l'}\|} = 0, \quad \forall l, l'. \tag{6.35}$$

Observe that $\|\overline{\mathbf{a}}_{1,l}\| = \|\overline{\mathbf{a}}_{2,l'}\| = \sqrt{N}$. Hence,

$$\frac{\overline{\mathbf{a}}_{1,l}^{\mathrm{H}}\overline{\mathbf{a}}_{2,l'}}{N} = \frac{1}{N}\sum_{n=0}^{N-1} \exp\left(\mathrm{j}\frac{2\pi}{\lambda}\left(\frac{n^2d^2(r_{2,l'} - r_{1,l})}{2r_{2,l'}r_{1,l}} - nd(\sin\phi_{1,l} - \sin\phi_{2,l'})\right)\right). \tag{6.36}$$

Taking into account that d is a constant, $r_{2,l'} - r_{1,l} \neq 0$ and $\sin\phi_{1,l} - \sin\phi_{2,l'} \neq 0$, then the exponential is different than one and $\overline{\mathbf{a}}_{1,l}^{\mathrm{H}}\overline{\mathbf{a}}_{2,l'}/N \to 0$. □

The previous theorem requires the scattering environments to be different for orthogonality to arise. Indeed, having different distances is sufficient, even if the angles are equal. This results in the following corollary for the near-field and far-field correlation matrices.

Corollary 6.1. Let $\overline{\mathbf{A}}_{\mathrm{ULA}}, \overline{\mathbf{C}}_{\mathrm{ULA}} \in \mathbb{C}^{N\times N}$ be the rank-L spatial correlation matrices given by (2.44) with the USW and UPW array response vectors in (2.27) and (2.29), respectively. Then, $\operatorname{col}(\overline{\mathbf{A}}_{\mathrm{ULA}}) \perp \operatorname{col}(\overline{\mathbf{C}}_{\mathrm{ULA}})$ when $N \to \infty$.

Proof. Without loss of generality we can assume $\overline{\mathbf{A}}_1 = \overline{\mathbf{A}}_{\mathrm{ULA}}$ and $\overline{\mathbf{A}}_2 = \overline{\mathbf{C}}_{\mathrm{ULA}}$. Then, letting $r_{2,l'} \to \infty$ in (6.36) and constraining $r_{1,l} < d_{\mathrm{F}}$ ensures that the argument of the exponential is different than zero even if $\sin\phi_{1,l} - \sin\phi_{2,l'} = 0$. Therefore, $\operatorname{col}(\overline{\mathbf{A}}_{\mathrm{ULA}}) \perp \operatorname{col}(\overline{\mathbf{C}}_{\mathrm{ULA}})$ when $N \to \infty$. □

Under the parabolic wavefront approximation, the ULA and UPA response vectors are related through the Kronecker product (see equations (2.39) and (2.40)). This allows to generalize theorem 6.1 and corollary 6.1 to UPAs, assuming that the parabolic approximation holds.

Theorem 6.2. Let $\overline{\mathbf{B}}_{\mathrm{UPA},1}, \overline{\mathbf{B}}_{\mathrm{UPA},2} \in \mathbb{C}^{N\times N}$ be the rank-L channel correlation matrices given by (2.44) with the uniform parabolic wave array response vector in (2.37), for two different scattering environments. Then, $\operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},1}) \perp \operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},2})$ when $N \to \infty$.

Proof. The proof for a UPA follows from the Kronecker product property of the channel response. Note that the subscript "ULA" is dropped throughout the proof for simplicity. Proceeding as before, $\{\overline{\mathbf{b}}_{1,y,l} \otimes \overline{\mathbf{b}}_{1,z,l}\}_{l=1}^{L}$ and $\{\overline{\mathbf{b}}_{2,y,l'} \otimes \overline{\mathbf{b}}_{2,z,l'}\}_{l'=1}^{L}$ are basis of $\operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},1})$ and $\operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},2})$, with $\overline{\mathbf{b}}_{1,y,l}, \overline{\mathbf{b}}_{2,y,l'} \in \mathbb{C}^{N_{\mathrm{H}}}$ and $\overline{\mathbf{b}}_{1,z,l}, \overline{\mathbf{b}}_{2,z,l'} \in \mathbb{C}^{N_{\mathrm{V}}}$. Therefore, to show that $\operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},1}) \perp \operatorname{col}(\overline{\mathbf{B}}_{\mathrm{UPA},2})$ asymptotically, we shall prove:

$$\frac{(\overline{\mathbf{b}}_{1,y,l} \otimes \overline{\mathbf{b}}_{1,z,l})^{\mathrm{H}}(\overline{\mathbf{b}}_{2,y,l'} \otimes \overline{\mathbf{b}}_{2,z,l'})}{N} \to 0 \tag{6.37}$$

when $N \to \infty$. Writing the limit explicitly and rearraging terms we can express it as the sum of two ULAs:

$$\begin{aligned}\lim_{N\to\infty} \frac{(\overline{\mathbf{b}}_{1,y,l} \otimes \overline{\mathbf{b}}_{1,z,l})^{\mathrm{H}}(\overline{\mathbf{b}}_{2,y,l'} \otimes \overline{\mathbf{b}}_{2,z,l'})}{N} &= \lim_{N\to\infty} \frac{(\overline{\mathbf{b}}_{1,y,l}^{\mathrm{H}}\overline{\mathbf{b}}_{2,y,l'}) \otimes (\overline{\mathbf{b}}_{1,z,l}^{\mathrm{H}}\overline{\mathbf{b}}_{2,z,l'})}{N} \\ = \lim_{N_{\mathrm{H}}\to\infty} \frac{\overline{\mathbf{b}}_{1,y,l}^{\mathrm{H}}\overline{\mathbf{b}}_{2,y,l'}}{N_{\mathrm{H}}} \cdot \lim_{N_{\mathrm{V}}\to\infty} \frac{\overline{\mathbf{b}}_{1,z,l}^{\mathrm{H}}\overline{\mathbf{b}}_{2,z,l'}}{N_{\mathrm{V}}}&.\end{aligned} \tag{6.38}$$

Applying theorem 6.1, the last two limits in (6.38) are zero. Therefore, (6.37) is satisfied and $\mathrm{col}\big(\overline{\mathbf{B}}_{\mathrm{UPA},1}\big) \perp \mathrm{col}\big(\overline{\mathbf{B}}_{\mathrm{UPA},2}\big)$. □

Corollary 6.2. Let $\overline{\mathbf{B}}_{\mathrm{UPA}}, \overline{\mathbf{C}}_{\mathrm{UPA}} \in \mathbb{C}^{N\times N}$ be the rank-L spatial correlation matrices given by (2.44) with the uniform parabolic wave and the UPW array response vectors in (2.37) and (2.38), respectively. Then, $\mathrm{col}\big(\overline{\mathbf{B}}_{\mathrm{UPA}}\big) \perp \mathrm{col}\big(\overline{\mathbf{C}}_{\mathrm{UPA}}\big)$ when $N \to \infty$.

Proof. It follows immediately from corollary 6.1 and theorem 6.2. □

In the previous propositions we have assumed that the array aperture goes to infinity when the number of antennas does as well. It should be remarked, however, that for an array with fixed aperture and variable inter-element spacing, the column spaces associated to the near-field and far-field correlation matrices are not orthogonal. For instance, in a ULA with fixed aperture D, inter-element spacing is given by $d = D/(N-1)$. Substituting in (6.36) yields

$$\frac{1}{N}\sum_{n=0}^{N-1} \exp\left(\mathrm{j}\frac{2\pi}{\lambda}\left(\frac{D^2n^2}{2(N-1)^2}\frac{r_{2,l'} - r_{1,l}}{r_{2,l'}r_{1,l}} - \frac{Dn}{N-1}(\sin\phi_{1,l} - \sin\phi_{2,l'})\right)\right), \tag{6.39}$$

which can be interpreted as a Riemann sum for the integral of a continuous function (see, for instance, [170, Sec. 7.5]):

$$\int_0^1 \exp\left(\mathrm{j}\frac{2\pi}{\lambda}\left(\frac{D^2(r_{2,l'} - r_{1,l})}{2r_{2,l'}r_{1,l}}x^2 - D(\sin\phi_{1,l} - \sin\phi_{2,l'})x\right)\right)\mathrm{d}x. \tag{6.40}$$

Therefore, the limit of $\overline{\mathbf{a}}_{1,l}^{\mathrm{H}}\overline{\mathbf{a}}_{2,l'}/N$ when $N \to \infty$ coincides with this integral. In general, (6.40) is not zero, thus the column spaces are not orthogonal when the array size is fixed.

6.2.3 Channel statistics discussion

Performance of energy-based noncoherent systems strongly depends on the channel correlation matrix. For this reason, we analyze the main properties of the near-field and far-field models (6.22) and (6.23). In doing so, we assume that scattering is local to each UE, with the k-th UE located at $\mathbf{r}_k = (r_k, \theta_k, \varphi_k)^{\mathrm{T}}$, and is uniformly sampled from a sphere centered at the user. Throughout the simulations presented in this section, we assume the radius of the sphere is $r_{\mathrm{s}} = 3\,\mathrm{m}$ and the number of scatterers is $L = 10$. Antennas at the BS are arranged as a UPA with, except where noted, $N = 16 \times 16$ and $d = \lambda$. The system operates at $f = 30\,\mathrm{GHz}$ such that $d_{\mathrm{r}} = 0.67\,\mathrm{m}$ and $d_{\mathrm{F}} = 10\,\mathrm{m}$.

Spatial correlation spectrum

We consider two UEs in the near field: $\mathbf{r}_1 = (5, 120°, -10°)$ and $\mathbf{r}_2 = (5, 110°, 0°)$, and a UE in the far field, $\mathbf{r}_3 = (25, 100°, 10°)$. Figure 6.5 reports the eigenvalues of $\mathbf{C}_{\mathbf{h}_1}$, $\mathbf{C}_{\mathbf{h}_2}$ and $\mathbf{C}_{\mathbf{h}_3}$, as obtained with (6.22), and those of their respective far-field matrices, obtained through (6.23). For all UEs, the eigenvalues of the near-field correlation matrix (rectangular markers) coincide with those of the far-field matrix (circular markers). This shows that both models exhibit the same correlation spectrum. Similar results have been previously observed in [37] and [39].

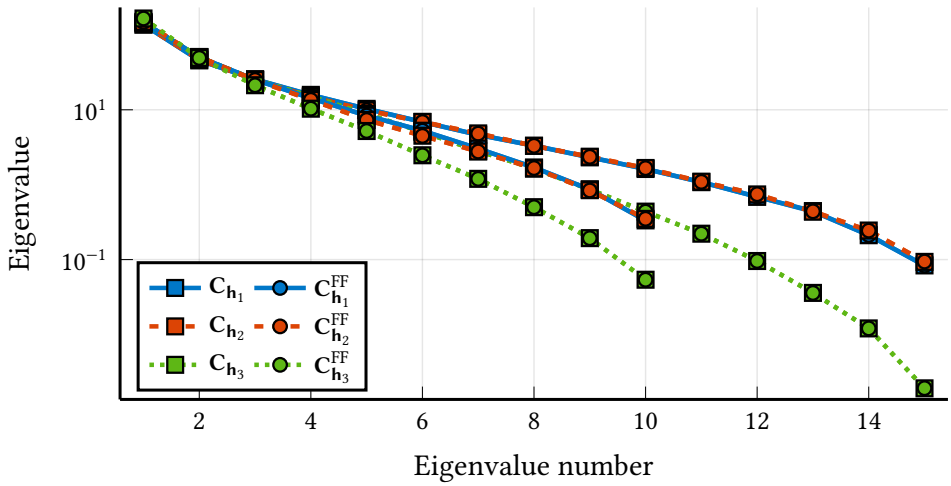


Figure 6.5: Eigenvalues of the near- and far-field channel correlation matrices for $L = 10$ and $L = 15$.

Regarding the eigenvalues of different users, we should recall that (6.22) and (6.23) only depend on the position of the scatterers, but not the user. Thus, if two users are in the same scattering cluster the eigenvalues of their channel correlation matrices coincide, even if scatterers realizations differ. Indeed, for $r_s = 3\,\mathrm{m}$ and two users located at $\mathbf{r}_1$ and $\mathbf{r}_2$, their scatterers overlap and the correlation matrices spectra are very close (see the dashed orange and solid blue lines in figure 6.5).

Distance between column spaces

The results in figure 6.5 evidence that the near-field and far-field correlation matrices from (6.22) and (6.23) exhibit identical spectra, even at short communication distances. Therefore, the primary difference between the two models must lie in their eigenvectors, *i.e.* $\mathbf{C}_{\mathbf{h}_k}$ and $\mathbf{C}^{\mathrm{FF}}_{\mathbf{h}_k}$ span different subspaces. However, as discussed in section 6.2.1, a noncoherent system such as the one here considered spans only two subspaces: the zero-symbol subspace and the non-zero-symbol subspace. Thus, measuring the distance between $\mathrm{col}\left(\mathbf{C}_{\mathbf{h}_k}\right)$ and $\mathrm{col}\left(\mathbf{C}^{\mathrm{FF}}_{\mathbf{h}_k}\right)$—which asymptotically must be zero—provides valuable insights into communication performance.

There are various methods to measure the distance between subspaces, but the most commonly used in the design of Grassmannian constellations is the *chordal distance*,[4] as it has a

[4]This is also referred to as *projection F-distance*, *e.g.* [171].

direct impact on the error probability [10], [172]. The chordal distance is defined by:

$$d_{\text{ch}}(\mathbf{X},\mathbf{Y}) = \left(\sum_{l=1}^{L} \sin^2 \alpha_l\right)^{\frac{1}{2}} = \text{tr}\left(\mathbf{I} - \mathbf{Y}^{\text{H}}\mathbf{X}\mathbf{X}^{\text{H}}\mathbf{Y}\right)^{\frac{1}{2}}, \tag{6.41}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are two L-rank matrices and α_l are the principal angles [173], [174] between the subspaces spanned by $\mathbf{X}$ and $\mathbf{Y}$. Observe that $d_{\text{ch}}(\mathbf{X},\mathbf{Y}) = 0$ if and only if $\text{col}(\mathbf{X}) = \text{col}(\mathbf{Y})$ and $d_{\text{ch}}(\mathbf{X},\mathbf{Y}) = \sqrt{L}$ if and only if $\text{col}(\mathbf{X}) \perp \text{col}(\mathbf{Y})$.

Now consider a single UE moving in a straight line from $(1, 120^\circ, -10^\circ)^{\text{T}}$ to $(200, 90.15^\circ, -10^\circ)^{\text{T}}$, so that it remains at a constant height $z = -0.5$ m with respect to the array, under three array configurations: $N = 16 \times 16$, $N = 24 \times 24$ and $N = 32 \times 32$. In figure 6.6, we represent $d_{\text{ch}}(\mathbf{U_h}, \mathbf{U_h^{\text{FF}}})/\sqrt{L}$, where $\mathbf{U_h}, \mathbf{U_h^{\text{FF}}} \in \mathbb{C}^{N \times L}$ are the matrices containing the L eigenvectors associated to the non-zero eigenvalues of $\mathbf{C_h}$ and $\mathbf{C_h^{\text{FF}}}$ (the UE subindex has been omitted for simplicity). Differently from the eigenvalues, which were already equal at $r < d_{\text{F}}$, the normalized chordal distance starts being one, that is, the subspaces spanned by $\mathbf{C_h}$ and $\mathbf{C_h^{\text{FF}}}$ are orthogonal. Although $d_{\text{ch}}(\mathbf{U_h}, \mathbf{U_h^{\text{FF}}})/\sqrt{L}$ decreases with the distance between transmitter and receiver, at the Fraunhofer distance it still is a 30% of its maximum value, and keeps decreasing at the same rate, suggesting that it is far from converging. In practice, this means that employing the far-field model makes the receiver operate in an appreciably mismatched mode, degrading the performance as it will be shown later.

On the other hand, incrementing the number of antennas (and hence the array aperture) also increases the chordal distance. Indeed, as $N \to \infty$, we have that $d_{\text{ch}}(\mathbf{U_h}, \mathbf{U_h^{\text{FF}}}) \to \sqrt{L}$. Consequently, $\text{col}(\mathbf{C_h}) \perp \text{col}\left(\mathbf{C_h^{\text{FF}}}\right)$ asymptotically, which is consistent with the theoretical subspaces orthogonality announced at the beginning of this section.

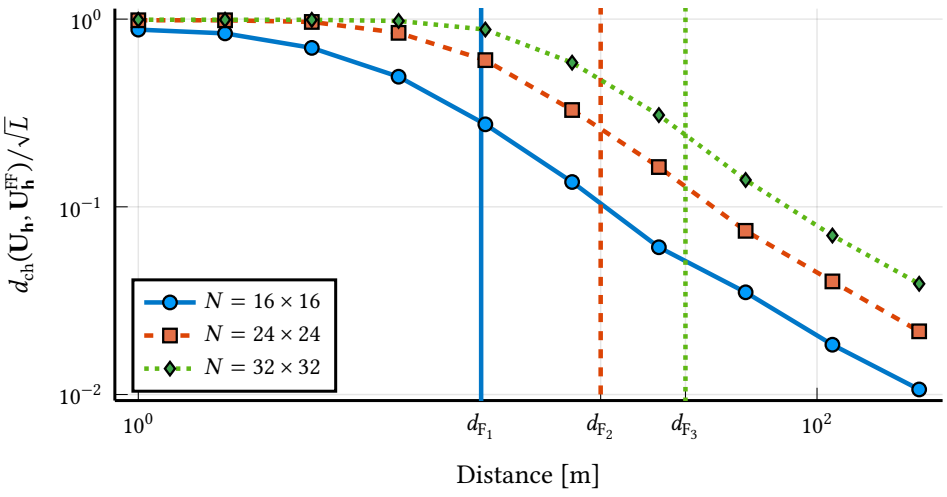


Figure 6.6: Chordal distance between $\text{col}(\mathbf{C_h})$ and $\text{col}\left(\mathbf{C_h^{\text{FF}}}\right)$ for $N = 16 \times 16$, $N = 24 \times 24$ and $N = 32 \times 32$, in terms of the distance between transmitter and receiver. The Fraunhofer distances for each case are indicated as d_{F_1}, d_{F_2} and d_{F_3}.

6.2.4 Numerical results

In order to verify the impact down to detection of a receiver employing the mismatched model (6.31) as well as the performance of the suboptimal single-user detector (6.29), we evaluate the error probability of a system employing the 4-PAM constellation that minimizes the SER of the optimal quadratic detector (see section 5.1 for further details).[5] Specifically, to assess the mismatched mode, we consider a single UE moving away from the BS. On the other hand, the performance of the suboptimal detector is analyzed in a scenario involving five static UEs simultaneously communicating with the same BS.

Single-user mismatched mode

Let us start by considering the same scenario as before: a single UE that moves along a straight line trajectory, beginning at $(1, 120^\circ, -10^\circ)^{\mathrm{T}}$ and ending at $(200, 90.15^\circ, -10^\circ)^{\mathrm{T}}$. In this case, the BS implements the receiver from (6.29), which for $K = 1$ corresponds to the ML detector that is then used to decide the transmitted symbol.

To show the impact of operating in a mismatched mode, in figure 6.7 we illustrate the SER of the proposed setup both for a receiver employing the exact correlation model and the mismatched one, for $L = 10$ and $L = 15$ scatterers. Besides the severe difference between error probabilities until $10d_{\mathrm{F}}$, using the far-field model may result in unexpected performance behaviors due to the mismatch. Regarding the increase of error probability with distance, even when employing the exact model, it is caused by the loss of resolution between scatterers in the local scattering model. That is, when the distance increases, all signal paths tend to come from the same angle and the rank of the channel correlation matrix goes to one.

The conclusions of this analysis are twofold: first, a receiver should always utilize the exact model, even beyond the Fraunhofer distance; second, noncoherent communication systems can take advantage from near field propagation conditions.

Multiuser detection

Consider five UEs arbitrarily located at positions: $\mathbf{r}_1 = (5, 120^\circ, -20^\circ)^{\mathrm{T}}$, $\mathbf{r}_2 = (10, 115^\circ, -10^\circ)^{\mathrm{T}}$, $\mathbf{r}_3 = (15, 110^\circ, 0^\circ)^{\mathrm{T}}$, $\mathbf{r}_4 = (20, 105^\circ, 10^\circ)^{\mathrm{T}}$ and $\mathbf{r}_5 = (25, 100^\circ, 20^\circ)^{\mathrm{T}}$. To multiplex them, scattering clusters must be separated, so throughout this section we consider $r_{\mathrm{s}} = 1\,\mathrm{m}$ to avoid overlapping. The receiver implements the detector in (6.29), and its performance with different number of antennas is compared with the scenario where only the UE of interest is present. For completeness, we also consider the mismatched case, where the wavefront is assumed to be planar rather than spherical.

The results of the aforementioned experiment for the fourth UE are shown in figure 6.8 (other UEs are omitted due to the symmetry of the problem). As expected, the SER of the single-user detector tends to that of the optimal detector when the number of antennas increases. In particular, for $N \geq 1500$ both receivers exhibit almost the same performance. On the other hand, the SER of the mismatched receiver is always high, independently of the number of antennas.

[5] To focus on receiver design, the constellation has been optimized for each N and each realization of scatterers location. The design is the same for near-field and far-field models since they share the same eigenvalues.

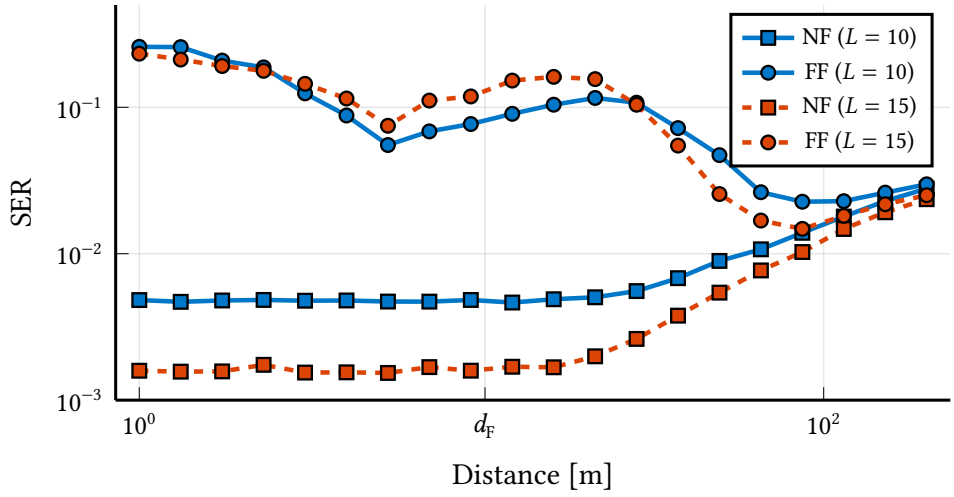


Figure 6.7: Error probability of a ML detector employing the exact or mismatched model at different distances. The simulation is performed for SNR = 20 dB.

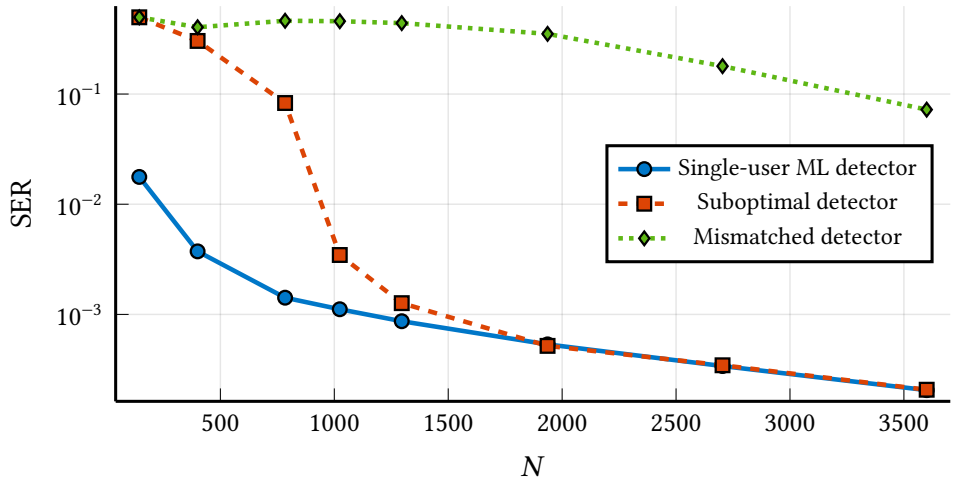


Figure 6.8: Error probability of the single-user detector in a multiuser environment for increasing number of antennas. The single-user ML and the mismatched receivers are depicted for comparison and completeness. The SINR is 20 dB.

6.3 Conclusions

In this chapter, we studied the problem of symbol detection employing model-based receivers. First, we examined the problem of inaccurate modeling of mutual coupling in a receiver with closely spaced antennas; and second, we analyzed the uplink of a NLoS one-shot noncoherent communication system, with arbitrarily located single-antenna UEs, and a BS equipped with a large antenna array operating at the mmWave band.

More specifically, in the former, we analyzed the uplink of a communication system where the BS ignores mutual coupling, and compared the resulting error probability with that obtained using the correct model. Our analysis employed both coherent and noncoherent decoding schemes, revealing that noncoherent decoding is more robust to coupling mismatch,

since it relies only on power and not on phase. This suggests that, in situations where mutual coupling cannot be accurately modeled, it may be beneficial to ignore phase information and instead rely on instantaneous channel energy detection [49]. In general, this approach provides a performance between that of the coherent detector and the average channel energy detector considered in this thesis.

In the latter, instead, we examined both exact and far-field models of the channel correlation matrix, as the performance of energy-based systems is tightly linked to them. Concretely, we showed that the subspaces spanned by these models differ significantly, even when the distance between the considered UE and the BS is greater than the Fraunhofer distance. More importantly, we illustrated that the aforementioned discrepancy clearly translates as well to a significant SER degradation at the receiver side. Furthermore, we proved that by exploiting large arrays it is possible to multiplex different UEs, even without instantaneous CSI, and achieve near-optimal performance with linear complexity (in the number of users) using a single-user detector. This result suggests that single-user constellations can be effectively used in a multiuser setting, provided that the BS has a sufficiently large number of antennas.

CHAPTER 7

Conclusions

This thesis has addressed the design and performance analysis of one-shot communication systems under physically consistent channel models, with a particular focus on error probability. By *physically consistent*, we refer to models that faithfully capture the physics of the underlying propagation mechanisms, as opposed to purely mathematical abstractions with no clear physical interpretation.

After establishing that near-field channels can still be modeled according to Rayleigh fading—even in the presence of intra-array coupling—we investigated one-shot energy detection systems under arbitrarily correlated Rayleigh fading and proposed near-optimal, low-complexity decision rules together with a constellation design method that minimizes their asymptotic error probability.

Furthermore, we examined the impact of employing mismatched models, both in terms of spatial correlation and mutual coupling. Remarkably, the results show that noncoherent systems are more robust to coupling mismatch compared to their coherent counterparts, and that they can also exploit wavefront curvature even beyond the classical Fraunhofer boundary separating the near and far field.

In particular, chapter 2 reviewed the basic concepts of wireless propagation, with special emphasis on large array systems and near-field channel modeling, their implications, and the resulting spatial correlations.

In chapter 3 we focused on the problem of mutual coupling in communication systems. Among the several alternatives to analyze it, we chose multiport communication theory, so the communication system can be represented through a circuit model. Although a large part of the chapter provides a detailed review of the state of the art and the multiport framework—with some extensions—two novel contributions were presented:

1. We first proved that there is no coupling from the receiver to the transmitter in massive SIMO and massive MISO systems, even when the number of antennas is asymptotically large and the operation takes place in the near field.

2. Building on this result, we proved that a NLoS communication channel in the near field and under the effects of mutual coupling can be modeled as Rayleigh fading, provided that the number of scatterers is sufficiently large.

Next, in chapters 4 and 5, we investigated one-shot noncoherent communication systems operating under correlated Rayleigh fading. We began by analyzing single-carrier systems in the absence of CSIT, and subsequently extended the study to account for scenarios with statistical CSIT and to multicarrier transmission. The key contributions of these two chapters can be summarized as follows:

3. Two fundamental results regarding the asymptotic performance of energy detection schemes under arbitrarily correlated Rayleigh fading:
    - When the transmitted symbols are uniquely identifiable, the error probability vanishes asymptotically with N, provided that the channel spectrum does not decay too fast when $N \to \infty$.
    - For a finite number of receiving antennas, the error probability of a fixed PAM constellation vanishes if and only if the constellation contains two symbols, one of which is zero.
4. The proposal of a quadratic detection framework that generalizes the conventional ED schemes in the literature. Within this quadratic structure, we derived an analytic approximation of the error probability—expressed as a sum of Q-functions—that applies to any detector in the framework.
5. The design of a scheme, which we called ABQUE, based on the combination of quadratic detectors, achieving near-optimal performance with significantly reduced complexity.
6. The introduction of two approaches to further improve the performance of energy detection systems:
    - A constellation design method that minimizes the asymptotic SER expression and outperforms all previously available methods in the literature, assuming statistical CSIT.
    - A near-ML, low-complexity detector for noncoherent OFDM-PIM, built upon the ABQUE detector introduced in chapter 4.

Finally, in chapter 6, we considered state-of-the-art receivers that, instead of directly estimating the channel or its covariance, construct them based on accurate physical models, where parameters such as users or scatterers distances and angles are estimated. Under these assumptions, we analyzed the importance of such physically consistent models for system performance. The main contributions of chapter 6 can be summarized as follows:

7. One-shot noncoherent communication systems are shown to be robust against mismatched modeling of mutual coupling, whereas coherent systems are not. Remarkably, in the latter case, the error probability follows the same trend as the antenna coupling function.
8. The Fraunhofer distance is not an adequate metric for delimiting the boundary between the near and far field in communication systems. In particular, noncoherent systems can exploit wavefront curvature effects well beyond the Fraunhofer distance.
9. The beamvector orthogonality property of large arrays extends to noncoherent systems. This property can be exploited to multiplex multiple users efficiently, achieving near-optimal performance with complexity that scales linearly in the number of users.

7.1 Future work

Following the conclusion of this thesis, several potential directions for future research and possible extensions of the studied problems have been identified. We begin with chapter 3:

- In section 3.3, we analyzed the coupling effects produced by the receiver on the transmitter within the multiport communication theory framework, evaluating the validity of the unilateral approximation for an asymptotic number of antennas in both uplink and downlink. This analysis, however, was limited to LoS propagation and to a single antenna at one of the ends. Therefore, two immediate extensions are:
  - Generalize the analysis to mMIMO, with arbitrary number of antennas at both transmitter and receiver.
  - Extend the study to NLoS propagation, for instance considering Rayleigh fading.

Now, we focus on research lines that follow from chapters 4 and 5:

- The analysis of OFDM-IM in section 5.2 manifests the importance of the chosen alphabet. Jointly optimizing partition sizes and energy levels and exploring alternative design criteria to trim the alphabet size could lead to improved SE/diversity tradeoffs [139].
- The PIM structure introduced is specially suitable for hierarchical and multi-resolution systems [175]. Combined coherent-noncoherent schemes based on PIM should be explored further.
- Investigate the performance of noncoherent systems within Delay-Doppler communications. While [176] proposed a noncoherent orthogonal time frequency space (OTFS) modulation based on differential detection, to the best of our knowledge energy-based OTFS remains an open topic. In addition, the extension to noncoherent OTFS-IM [177]–[179] also appears to be a promising direction.

Finally, the potential extensions that arise from chapter 6 are:

- Explore partially coherent systems with a known LoS component and multiple NLoS paths (*i.e.* Rician fading). In practice, high-frequency channels contain only a limited number of multipath components, which constrains the rank of the correlation matrix and thus bounds the error probability. Introducing a known LoS path is expected to mitigate this effect, making the performance closer to that of a coherent receiver.
- Analyze the beamfocusing capabilities of modular arrays that exploit both long-term and short-term CSI. Instead of generating the focusing pattern solely from the estimated user location (long-term CSI), the carrier phase per ULA module (short-term CSI) at the receiver can also be incorporated. This joint use may lead to an enhanced robustness by suppressing undesired grating lobes in the focusing region induced by the large inter-module separation [180], [181].

APPENDIX A

Elements of network analysis

A significant part of this thesis is devoted to analyzing the impact of mutual coupling on communication systems. Since the adopted models are grounded in circuit and network theory, in this appendix we review the fundamental concepts of circuit and network theory that will be used throughout the thesis.

The first concept required to develop the subsequent theory is that of a port. In circuit theory, a port is defined as a pair of terminals through which electrical energy can be delivered to or extracted from a circuit. For a pair of terminals to qualify as a port, they must satisfy the *port condition*, namely that the currents entering the two terminals are equal in magnitude and opposite in direction [182, Ch. 3].

A.1 Complex envelope of a bandpass signal

The complex envelope of a signal refers to the baseband complex-valued signal whose real, modulated version corresponds to the actual physical signal at the ports of an antenna or circuit.

For a real-valued signal $s(t)$, the complex envelope is derived from its analytic representation $a_s(t) = s(t) + \mathrm{j}\hat{s}(t)$, where $\hat{s}(t)$ denotes the Hilbert transform of $s(t)$ [183, Ch. 3]. The complex envelope associated with a carrier frequency ω is then defined as

$$b_s(t) = a_s(t)\mathrm{e}^{-\mathrm{j}\omega t}. \tag{A.1}$$

The original real-valued signal can be recovered by modulating the complex envelope back to passband and taking the real part:

$$s(t) = \Re\big(b_s(t)\mathrm{e}^{\mathrm{j}\omega t}\big). \tag{A.2}$$

If $s(t) = A \cdot \cos(\omega t + \theta)$, as usually assumed in alternating current (AC) circuits,[1] then its complex envelope is $b_s(t) = A\mathrm{e}^{\mathrm{j}\theta}$, which is constant in time and corresponds to the phasor associated with the real signal $s(t)$.

[1] Multiple-frequency linear AC circuits, as well as circuits driven by arbitrary periodic waveforms, can be analyzed by decomposing all signals into their sinusoidal Fourier components, each characterized by a specific magnitude and phase. Because of the superposition theorem, each frequency component may then be analyzed independently.

A.2 Active power

Assume that $v(t)$ and $i(t)$ are the baseband complex envelopes of two narrowband real-valued bandpass signals measuring the voltage and current at a given port. The average active power that flows into the port is

$$P = \frac{1}{T}\int_0^T \Re(v(t)i(t)^*)\,\mathrm{d}t. \tag{A.3}$$

If both signals are random and ergodic [184, Ch. 20], the active power can equivalently be expressed as

$$P = \mathrm{E}[\Re(\mathrm{v}\cdot\mathrm{i}^*)], \tag{A.4}$$

where the temporal average has been replaced by the statistical expectation.

Since communication signals are inherently stochastic, the definition given in (A.4) is the one adopted throughout this thesis.

A.2.1 Available power

In order to define the available power we first need to understand how much power can a load extract from a source. To do so, consider a source-load interface as shown in figure A.1. The source circuit is represented by its Thévenin equivalent circuit, with complex envelope voltage $\mathrm{v_G}$ and impedance $Z_\mathrm{G} = R_\mathrm{G}+\mathrm{j}X_\mathrm{G}$. The load is represented by an equivalent impedance $Z_\mathrm{L} = R_\mathrm{L} + \mathrm{j}X_\mathrm{L}$.

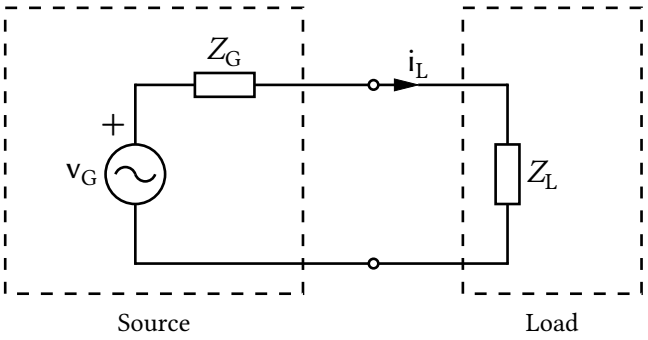


Figure A.1: Equivalent circuit with a source and a load.

Maximum power transfer from the source to the load is obtained when $Z_\mathrm{L} = Z_\mathrm{G}^*$ [182, Sec. 8.6]. In this situation, the power dissipated at the load is

$$P = \mathrm{E}\left[\Re\left(\mathrm{v_L}\cdot\mathrm{i_L^*}\right)\right] = \mathrm{E}\left[\Re\left(\frac{Z_\mathrm{L}}{2R_\mathrm{G}}\mathrm{v_G}\cdot\left(\frac{\mathrm{v_G}}{2R_\mathrm{G}}\right)^*\right)\right] = \frac{\mathrm{E}\left[|\mathrm{v_G}|^2\right]}{4R_\mathrm{G}}. \tag{A.5}$$

Since there is maximum power transfer, we say that (A.5) is the source *available power*.

A.3 Multiport networks

From the port definition provided at the beginning of this appendix, it follows naturally that a multiport is a circuit possessing an arbitrary number of ports [185, Sec. 10.1].

A.3.1 Impedance parameters

The impedance parameters of a K-port are obtained by expressing the port voltages $v_1, \dots, v_K$ in terms of port currents $i_1, \dots, i_K$:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_K \end{pmatrix} = \begin{pmatrix} Z_{1,1} & \dots & Z_{1,K} \\ \vdots & \ddots & \vdots \\ Z_{K,1} & \dots & Z_{K,K} \end{pmatrix} \begin{pmatrix} i_1 \\ \vdots \\ i_K \end{pmatrix} = \mathbf{Z}\mathbf{i}, \tag{A.6}$$

where $\mathbf{Z}$ is the impedance matrix and its elements are the impedance parameters. By definition, each parameter is obtained under open circuit conditions:

$$Z_{j,k} = \frac{v_j}{i_k}\Big|_{i_n=0,\, n\neq k}. \tag{A.7}$$

Naturally, they are measured in ohms.

A network is said to be *reciprocal* if $\mathbf{Z} = \mathbf{Z}^{\mathrm{T}}$. In other words, the voltage appearing at port j due to a current applied at port k is the same as the voltage appearing at port k when the same current is applied to port j. Networks that consist of linear passive elements are reciprocal.

On the other hand, we say that a network is *lossless* if it does not dissipate power. Formally, this condition can be stated as

$$\Re(\mathbf{Z}) = \mathbf{0}, \tag{A.8}$$

which implies that the network contains no resistive or dissipative elements [81, Sec. 4.2].

Impedance relations

Consider now the circuit in figure A.2. In this scenario, the impedance matrix can be partitioned into four blocks, analogous to the structure of a two-port network:

$$\begin{pmatrix} \mathbf{v}_{\mathrm{T}} \\ \mathbf{v}_{\mathrm{L}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{i}_{\mathrm{T}} \\ \mathbf{i}_{\mathrm{L}} \end{pmatrix}, \tag{A.9}$$

where $\mathbf{v}_{\mathrm{T}}, \mathbf{i}_{\mathrm{T}} \in \mathbb{C}^M$, $\mathbf{v}_{\mathrm{L}}, \mathbf{i}_{\mathrm{L}} \in \mathbb{C}^N$, $\mathbf{Z}_{11} \in \mathbb{C}^{M\times M}$, $\mathbf{Z}_{22} \in \mathbb{C}^{N\times N}$, $\mathbf{Z}_{12} \in \mathbb{C}^{M\times N}$, $\mathbf{Z}_{21} \in \mathbb{C}^{N\times M}$ and $M + N = K$.

The input impedance of a multiport network must satisfy

$$\mathbf{v}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{in}}\mathbf{i}_{\mathrm{T}}. \tag{A.10}$$

On the other hand, Ohm's law tells us that

$$\mathbf{v}_{\mathrm{L}} = -\mathbf{Z}_{\mathrm{L}}\mathbf{i}_{\mathrm{L}}, \tag{A.11}$$

where $\mathbf{Z}_{\mathrm{L}}$ is a diagonal matrix with elements $Z_{\mathrm{L,n}}$, for $1 \leq n \leq N$. Substituted into (A.9) yields

$$\begin{aligned} \mathbf{Z}_{\mathrm{in}}\mathbf{i}_{\mathrm{T}} &= \mathbf{Z}_{11}\mathbf{i}_{\mathrm{T}} + \mathbf{Z}_{12}\mathbf{i}_{\mathrm{L}}, \\ -\mathbf{Z}_{\mathrm{L}}\mathbf{i}_{\mathrm{L}} &= \mathbf{Z}_{21}\mathbf{i}_{\mathrm{T}} + \mathbf{Z}_{22}\mathbf{i}_{\mathrm{L}}. \end{aligned} \tag{A.12}$$

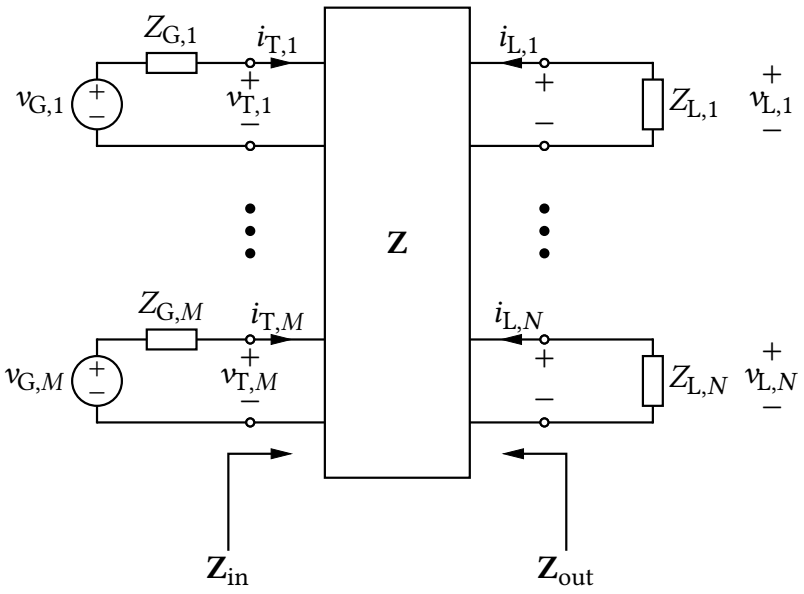


Figure A.2: Multiport network with M source ports and N load ports.

After some manipulation, the input impedance can be obtained as

$$\mathbf{Z}_{in} = \mathbf{Z}_{11} - \mathbf{Z}_{12}(\mathbf{Z}_L + \mathbf{Z}_{22})^{-1}\mathbf{Z}_{21}. \tag{A.13}$$

Following a similar approach, the output impedance is given by

$$\mathbf{Z}_{out} = \mathbf{Z}_{22} - \mathbf{Z}_{21}(\mathbf{Z}_G + \mathbf{Z}_{11})^{-1}\mathbf{Z}_{12}, \tag{A.14}$$

with $\mathbf{Z}_G$ a diagonal matrix with elements $Z_{G,n}$, for $1 \leq n \leq M$.

If $\mathbf{Z}_{in} = \mathbf{Z}_{out}$, we say that the network is *symmetrical*. Any symmetrical network is also reciprocal, with the added constraint that the main diagonal is constant [82, Sec. 5.4].

A.3.2 Scattering parameters

At high frequencies it becomes difficult to realize the short- and open-circuit terminations needed for the impedance representation of the multiport. Instead, multiports are more commonly described by their scattering parameters [81, Ch. 4], [82, Ch. 5].

At the n-th port, we define a_n and b_n representing the forward-propagating (incident) normalized power wave and the backward-propagating (reflected) normalized power wave, respectively. Letting v_n and i_n be the voltage and current at port n, the normalized power waves are given by

$$a_n = \frac{v_n + Z_n i_n}{2\sqrt{\Re(Z_n)}}, \quad b_n = \frac{v_n - Z_n^* i_n}{2\sqrt{\Re(Z_n)}}, \tag{A.15}$$

where Z_n is the port reference impedance [83]. Since a_n is related to the incident power and b_n to the reflected power, we can compute

$$|a_n|^2 - |b_n|^2 = \frac{|v_n + Z_n i_n|^2 - |v_n - Z_n^* i_n|^2}{4\Re(Z_n)} = \Re(v_n i_n^*), \tag{A.16}$$

which corresponds to the total power accepted at the n-th port [82, Sec. 5.3], [83].

Analogously to voltages and currents, power waves are related by a linear system:

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_K \end{pmatrix} = \begin{pmatrix} S_{1,1} & \cdots & S_{1,K} \\ \vdots & \ddots & \vdots \\ S_{K,1} & \cdots & S_{K,K} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_K \end{pmatrix} = \mathbf{S}\mathbf{a}, \tag{A.17}$$

where $\mathbf{S}$ is the scattering matrix. The diagonal entries $S_{i,i}$ are known as reflection coefficients, while the off-diagonal terms $S_{i,j},\ i \neq j$, are called transmission coefficients.

The scattering matrix can be obtained from the impedance matrix. First, observe that the power waves vectors can be written as

$$\begin{aligned} \mathbf{a} &= \mathbf{F}(\mathbf{v} + \mathbf{G}\mathbf{i}), \\ \mathbf{b} &= \mathbf{F}(\mathbf{v} - \mathbf{G}^*\mathbf{i}), \end{aligned} \tag{A.18}$$

where $\mathbf{F}$ is a diagonal matrix with elements $\frac{1}{2\sqrt{\Re(Z_n)}}$ and $\mathbf{G}$ is also a diagonal matrix with elements Z_n, for $1 \leq n \leq K$. By the impedance matrix relation in (A.6), we can write

$$\begin{aligned} \mathbf{a} &= \mathbf{F}(\mathbf{Z} + \mathbf{G})\mathbf{i}, \\ \mathbf{b} &= \mathbf{F}(\mathbf{Z} - \mathbf{G}^*)\mathbf{i} = \mathbf{F}(\mathbf{Z} - \mathbf{G}^*)(\mathbf{Z} + \mathbf{G})^{-1}\mathbf{F}^{-1}\mathbf{a}. \end{aligned} \tag{A.19}$$

Recall from (A.17) that the scattering matrix relates $\mathbf{a}$ and $\mathbf{b}$, so:

$$\mathbf{S} = \mathbf{F}(\mathbf{Z} - \mathbf{G}^*)(\mathbf{Z} + \mathbf{G})^{-1}\mathbf{F}^{-1}. \tag{A.20}$$

From this relation, it follows that the scattering matrix must also exhibit special properties when the network is reciprocal, symmetric or lossless. For reciprocal networks, $\mathbf{S}$ is symmetric (*i.e.* $\mathbf{S}^{\mathrm{T}} = \mathbf{S}$). If the network is additionally symmetrical, then all diagonal elements of $\mathbf{S}$ are equal. Finally, if the network is lossless, the scattering matrix satisfies $\mathbf{S}^{\mathrm{H}}\mathbf{S} = \mathbf{I}$, that is, $\mathbf{S}$ is a unitary matrix [81, Sec. 4.3].

As a final remark, for CMS antennas, the impedance description is often more convenient, as it can be obtained directly from the isolated radiation pattern with a pairwise treatment [83].

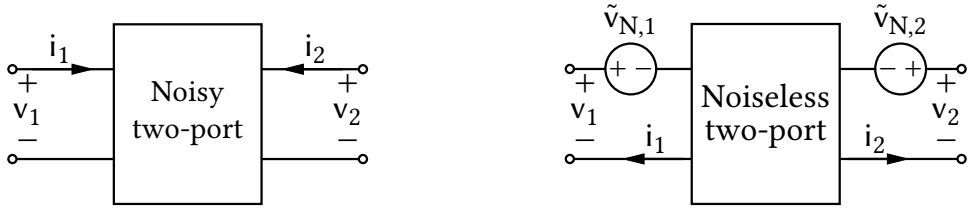
A.3.3 Noisy multiports

In some situations, such as the analysis of amplifiers, it is necessary to account for multiports that introduce noise into the system. The noise generated by a multiport is originated internally, preventing the use of a purely black-box approach. Nevertheless, by exploiting linearity, the system can be equivalently represented as a noiseless multiport with additive noise sources at each port. Mathematically, both representations are equivalent and can be expressed as

$$\mathbf{v} = \mathbf{Z}\mathbf{i} + \tilde{\mathbf{v}}_{\mathrm{N}}, \tag{A.21}$$

where $\tilde{\mathbf{v}}_{\mathrm{N}}$ denotes the vector of noise sources [80], [186]. The equivalent circuit representations with two-port networks are shown in figures A.3a and A.3b.

Focusing on a single input-output pair, an alternative formulation can be adopted. Specifically, the model in figure A.3b can be reformulated into the equivalent representation



(a) A noisy two-port. **(b)** A noiseless two-port with a noise source at each port.

Figure A.3: Equivalent circuits for a noisy two-port network.

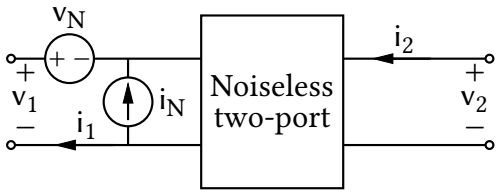


Figure A.4: A noiseless two-port with two noise sources at the same port.

in figure A.4. Although both models involve two noise sources, in the latter they are placed at the same port [80], [186], [187]. This latter representation is the one adopted in the multiport communication framework presented in section 3.1.

APPENDIX B

Analysis of multiport cascade connection

In the absence of noise, the system in figure 3.1 can be simplified to that in figure 3.2, where the multiport is characterized by equation (3.15):

$$\begin{pmatrix} \mathbf{v}_T \\ \mathbf{v}_R \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_T & \mathbf{Z}_{TR} \\ \mathbf{Z}_{RT} & \mathbf{Z}_R \end{pmatrix} \begin{pmatrix} \mathbf{i}_T \\ \mathbf{i}_R \end{pmatrix}. \tag{B.1}$$

Since $\mathbf{v}_T = \mathbf{Z}_T\mathbf{i}_T + \mathbf{Z}_{TR}\mathbf{i}_R$, both blocks may be obtained by finding the relation between $\mathbf{v}_T$ and the currents $\mathbf{i}_T$ and $\mathbf{i}_R$, as a function of $\mathbf{Z}_{MT}$, $\mathbf{Z}_A$ and $\mathbf{Z}_{MR}$.

Based on equation (3.5), the transmitter matching network impedance relations are:

$$\mathbf{v}_T = \mathbf{Z}_{MT,11}\mathbf{i}_T - \mathbf{Z}_{MT,12}\mathbf{i}_{AT}, \tag{B.2}$$
$$\mathbf{v}_{AT} = \mathbf{Z}_{MT,21}\mathbf{i}_T - \mathbf{Z}_{MT,22}\mathbf{i}_{AT}. \tag{B.3}$$

Similarly, the receiver matching network equations are obtained from (3.6) as:

$$\mathbf{v}_R = -\mathbf{Z}_{MR,12}\mathbf{i}_{AR} + \mathbf{Z}_{MR,11}\mathbf{i}_R, \tag{B.4}$$
$$\mathbf{v}_{AR} = -\mathbf{Z}_{MR,22}\mathbf{i}_{AR} + \mathbf{Z}_{MR,21}\mathbf{i}_R. \tag{B.5}$$

Finally, the antenna multiport voltages are given by:

$$\mathbf{v}_{AT} = \mathbf{Z}_{AT}\mathbf{i}_{AT} + \mathbf{Z}_{ATR}\mathbf{i}_{AR}, \tag{B.6}$$
$$\mathbf{v}_{AR} = \mathbf{Z}_{ART}\mathbf{i}_{AT} - \mathbf{Z}_{AR}\mathbf{i}_{AR}. \tag{B.7}$$

Substituting (B.5) into (B.7) and isolating $\mathbf{i}_{AR}$ yields

$$\mathbf{i}_{AR} = -(\mathbf{Z}_{MR,22} + \mathbf{Z}_{AR})^{-1}\mathbf{Z}_{ART}\mathbf{i}_{AT} + (\mathbf{Z}_{MR,22} + \mathbf{Z}_{AR})^{-1}\mathbf{Z}_{MR,21}\mathbf{i}_R, \tag{B.8}$$

which substituted in (B.6) results in

$$\mathbf{v}_{AT} = (\mathbf{Z}_{AT} - \mathbf{Z}_{ATR}(\mathbf{Z}_{MR,22} + \mathbf{Z}_{AR})^{-1}\mathbf{Z}_{ART})\mathbf{i}_{AT} + \mathbf{Z}_{ATR}(\mathbf{Z}_{MR,22} + \mathbf{Z}_{AR})^{-1}\mathbf{Z}_{MR,21}\mathbf{i}_R. \tag{B.9}$$

Equaling (B.9) with (B.3) allows to write $\mathbf{i}_{\mathrm{AT}}$ as

$$\begin{aligned}\mathbf{i}_{\mathrm{AT}} = &(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}})^{-1}\mathbf{Z}_{\mathrm{MT},21}\mathbf{i}_{\mathrm{T}} \\ &- (\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}})^{-1}\mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{MR},21}\mathbf{i}_{\mathrm{R}},\end{aligned} \tag{B.10}$$

where $\mathbf{A}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{AT}} + \mathbf{Z}_{\mathrm{MT},22}+$ and $\mathbf{A}_{\mathrm{R}} = \mathbf{Z}_{\mathrm{AR}} + \mathbf{Z}_{\mathrm{MR},22}$. Finally, $\mathbf{Z}_{\mathrm{T}}$ is obtained by substituting (B.10) in (B.2) with $\mathbf{i}_{\mathrm{R}} = \mathbf{0}$:

$$\mathbf{Z}_{\mathrm{T}} = \mathbf{Z}_{\mathrm{MT},11} - \mathbf{Z}_{\mathrm{MT},12}(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}})^{-1}\mathbf{Z}_{\mathrm{MT},21}. \tag{B.11}$$

The inter-array matrix, on the other hand, follows from setting $\mathbf{i}_{\mathrm{T}} = \mathbf{0}$ instead of $\mathbf{i}_{\mathrm{R}}$:

$$\mathbf{Z}_{\mathrm{TR}} = \mathbf{Z}_{\mathrm{MT},12}(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}})^{-1}\mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{MR},21}. \tag{B.12}$$

The remaining blocks, $\mathbf{Z}_{\mathrm{R}}$ and $\mathbf{Z}_{\mathrm{RT}}$, can be obtained by following an analogous procedure at the receiver. However, a more straightforward approach is to exploit the symmetry of the problem, simply replacing all subscripts "T" with "R" and vice versa in (B.11) and (B.12):

$$\mathbf{Z}_{\mathrm{R}} = \mathbf{Z}_{\mathrm{MR},11} - \mathbf{Z}_{\mathrm{MR},12}(\mathbf{A}_{\mathrm{R}} - \mathbf{Z}_{\mathrm{ART}}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{ATR}})^{-1}\mathbf{Z}_{\mathrm{MR},21}, \tag{B.13}$$

$$\mathbf{Z}_{\mathrm{RT}} = \mathbf{Z}_{\mathrm{MR},12}(\mathbf{A}_{\mathrm{R}} - \mathbf{Z}_{\mathrm{ART}}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{ATR}})^{-1}\mathbf{Z}_{\mathrm{ART}}\mathbf{A}_{\mathrm{T}}^{-1}\mathbf{Z}_{\mathrm{MT},21}. \tag{B.14}$$

Furthermore, since reciprocity enforces $\mathbf{Z}_{\mathrm{RT}} = \mathbf{Z}_{\mathrm{TR}}^{\mathrm{T}}$, (B.14) can also be expressed as

$$\mathbf{Z}_{\mathrm{RT}} = \mathbf{Z}_{\mathrm{MR},12}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}}(\mathbf{A}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{ATR}}\mathbf{A}_{\mathrm{R}}^{-1}\mathbf{Z}_{\mathrm{ART}})^{-1}\mathbf{Z}_{\mathrm{MT},21}. \tag{B.15}$$

The equivalence between both expressions follows directly from the Woodbury matrix identity [19], [92, Sec. 0.7].

APPENDIX C

Proofs of chapter 4

C.1 Proof of theorem 4.1

The PEP can be upper bounded making use of Cantelli's inequality [188]:

$$P_{a\to b} \leq \left(1+\Delta_{a,b}\right)^{-1}, \tag{C.1}$$

where

$$\Delta_{a,b} = \frac{\mathrm{E}_{\mathbf{r}|x_a}\left[\mathrm{L}_{a,b}(\mathbf{r}|x_a)\right]^2}{\mathrm{var}_{\mathbf{r}|x_a}\left(\mathrm{L}_{a,b}(\mathbf{r}|x_a)\right)}. \tag{C.2}$$

We must check whether this term grows without bounds, so that the PEP vanishes. This is intimately related to the channel hardening phenomenon exhibited by large arrays; notice how (C.2) is analogous to [70, Eq. (10)].

The elements involved in the computation of (C.2) are:

$$\begin{aligned}\mathrm{E}_{\mathbf{r}|x_a}\left[\mathrm{L}_{a,b}(\mathbf{r}|x_a)\right] &= \sum_{n=1}^{N}(\lambda_n - 1 - \ln\lambda_n),\\ \mathrm{var}_{\mathbf{r}|x_a}\left(\mathrm{L}_{a,b}(\mathbf{r}|x_a)\right) &= \sum_{n=1}^{N}(\lambda_n - 1)^2,\end{aligned} \tag{C.3}$$

where

$$\lambda_n = \frac{|x_a|^2\gamma_n + 1}{|x_b|^2\gamma_n + 1}. \tag{C.4}$$

Thus, we are left with

$$\Delta_{a,b} = \frac{\left(\sum_{n=1}^{N}(\lambda_n - 1 - \ln\lambda_n)\right)^2}{\sum_{n=1}^{N}(\lambda_n - 1)^2} = \frac{u_N}{d_N}. \tag{C.5}$$

Next, we distinguish two cases: $\lambda_n \nrightarrow 1$ and $\lambda_n \to 1$. Note that the first case implies that $|x_a|^2 \neq |x_b|^2$ and $\gamma_n \nrightarrow 0$, so $\mathrm{tr}\left(\mathbf{\Gamma}^2\right) \to \infty$ is automatically satisfied.

C.1.1 Non-convergence to unity

By hypothesis, $\lambda_n \nrightarrow 1$ and $\lambda_n \nrightarrow \infty$. Hence, the denominator in (C.5) is a sum of positive terms satisfying

$$c' \leq (\lambda_n - 1)^2 \leq C' \implies c'N \leq d_N \leq C'N, \tag{C.6}$$

for some positive constants c' and C'.

Therefore, the denominator grows linearly with N.

On the other hand, the terms $\lambda_n - 1 - \ln \lambda_n$ appearing in the numerator are strictly positive for $\lambda_n \neq 1$, which holds by hypothesis. Thus, each term can be bounded by positive constants c and C as

$$c \leq \lambda_n - 1 - \ln \lambda_n \leq C. \tag{C.7}$$

Since the numerator is the square of the sum of these terms, it follows that

$$c^2 N^2 \leq u_N \leq C^2 N^2, \tag{C.8}$$

so the numerator grows quadratically with N.

Therefore, since the numerator scales as N^2 while the denominator scales as N, we conclude that $\Delta_{a,b}$ grows linearly with N.

C.1.2 Convergence to unity

As previously stated, the condition $\lambda_n \to 1$ may arise either because $\gamma_n \to 0$ or because $|x_a|^2 = |x_b|^2$. Indeed, the latter case yields $\lambda_n = 1$ for all n, which has no physical meaning, since the two hypotheses in the test become identical and therefore cannot be distinguished.

It is thus clear that, under the premise of the spectrum $\{\gamma_n\}_{n\geq 1}$ being nonnegative and bounded,

$$|x_a|^2 \neq |x_b|^2 \iff x_a \neq x_b, \ \forall x_a, x_b \in \mathscr{X}, \tag{C.9}$$

which is consistent with the loss of phase information observed in the ML detector (4.6).

Property (C.9) is a specific instance of a broader concept known as *unique identification*. It has been recognized throughout the literature under various forms [109], [189], [190],[1] but we abide by the one presented in [194, Proposition 1]. In general terms, a constellation is uniquely identifiable if there is a one-to-one correspondence between each of its symbols and a distinct second-order statistical structure at the receiver. Indeed, particularized for our model (4.1) amounts to

$$x_a \mathbf{C_h} x_a^* \neq x_b \mathbf{C_h} x_b^* \iff x_a \neq x_b, \quad x_a, x_b \in \mathscr{X}, \tag{C.10}$$

which is equivalent to (C.9). Not fulfilling it would collapse the LLR in (4.12) and thus positively lower bound the PEP: it is a necessary condition for (asymptotically) error-free ML detection.

[1] These works use terms like *unique factorization*, *unique identification* or *unique determination* to define similar ideas, while others employ them to refer to different concepts [191]–[193].

Focusing on the case $\gamma_n \to 0$, instead of directly studying the divergence of $\Delta_{a,b}$, we analyze whether $\sqrt{\Delta_{a,b}}$, given by

$$\sqrt{\Delta_{a,b}} = \frac{\sum_{n=1}^{N}(\delta_n - \ln(1+\delta_n))}{\sqrt{\sum_{n=1}^{N}\delta_n^2}}, \tag{C.11}$$

diverges, since $\Delta_{a,b} \to \infty \iff \sqrt{\Delta_{a,b}} \to \infty$. In (C.11), we have defined

$$\delta_n = \lambda_n - 1, \tag{C.12}$$

so $\delta_n \to 0$ when $\lambda_n \to 1$. Note that δ_n is bounded but can be negative.

The numerator in (C.11) has a well-known Taylor expansion [31, Ch. 20]:

$$\delta_n - \ln(1+\delta_n) = \delta_n - (\delta_n - R_2(\delta_n)) = R_2(\delta_n) = \frac{\delta_n^2}{2(1+\xi_n)^2}, \tag{C.13}$$

where $R_2(\delta_n)$ is the Lagrange remainder and $|\xi_n| < |\delta_n|$, for all n.

Therefore, we can rewrite (C.11) as

$$\sqrt{\Delta_{a,b}} = \frac{\sum_{n=1}^{N}\frac{\delta_n^2}{(1+\xi_n)^2}}{2\sqrt{\sum_{n=1}^{N}\delta_n^2}}. \tag{C.14}$$

Since ξ_n is bounded, it follows that $k \le (1+\xi_n)^2 \le K$ for all n, given some positive constants k and K. Thus, (C.14) admits a global two-sided bound:

$$\frac{1}{2K}\sqrt{\sum_{n=1}^{N}\delta_n^2} \le \sqrt{\Delta_{a,b}} \le \frac{1}{2k}\sqrt{\sum_{n=1}^{N}\delta_n^2} \tag{C.15}$$

From the definition of ℓ^2-summable sequences [195, Sec. 12.1], the sum converges as $N \to \infty$ if and only if $\{\delta_n\}_{n\ge1} \in \ell^2$. Furthermore, the divergence rate of $\Delta_{a,b}$ is asymptotically equal to that of d_N.

We are left to determine the implications of $\{\delta_n\}_{n\ge1} \in \ell^2$ on the spectrum sequence $\{\gamma_n\}_{n\ge1}$. In terms of γ_n, we have

$$|\delta_n| = \left|\frac{|x_a|^2\gamma_n + 1}{|x_b|^2\gamma_n + 1} - 1\right| = \frac{\big||x_a|^2 - |x_b|^2\big|\gamma_n}{|x_b|^2\gamma_n + 1} \le \big||x_a|^2 - |x_b|^2\big|\gamma_n. \tag{C.16}$$

Hence, $\{\gamma_n\}_{n\ge1} \in \ell^2 \implies \{\delta_n\}_{n\ge1} \in \ell^2$.

To establish the converse implication, we leverage that $\gamma_n \le L$ for all n, which yields

$$|\delta_n| \ge \frac{\big||x_a|^2 - |x_b|^2\big|\gamma_n}{|x_b|^2 L + 1} \implies \gamma_n \le \frac{|\delta_n|\big(|x_b|^2 L + 1\big)}{\big||x_a|^2 - |x_b|^2\big|}. \tag{C.17}$$

It immediately follows that $\{\delta_n\}_{n\ge1} \in \ell^2 \implies \{\gamma_n\}_{n\ge1} \in \ell^2$.

Consequently, when $\gamma_n \to 0$ and $|x_a|^2 \ne |x_b|^2$, $\Delta_{a,b}$ diverges if and only if $\{\gamma_n\}_{n\ge1} \notin \ell^2$. Written in terms of the spectral matrix, $\Delta_{a,b}$ diverges if and only if $\mathrm{tr}\big(\mathbf{\Gamma}^2\big) \to \infty$. □

C.2 Proof of theorem 4.2

Referring to (4.11), the PEP can be expressed as [122, Sec. 2.4]

$$P_{a\to b} = \int_{\mathscr{D}} f_{\mathbf{r}|x_a}(\mathbf{r})\,\mathrm{d}\mathbf{r}, \tag{C.18}$$

where $\mathscr{D} = \left\{\mathbf{r}\in\mathbb{C}^N \,:\, f_{\mathbf{r}|x_b}(\mathbf{r}) \geq f_{\mathbf{r}|x_a}(\mathbf{r})\right\}$. With some simple manipulations, we can see that the boundary of this region defines an $(N-1)$-dimensional complex ellipsoid:

$$\partial\mathscr{D} = \left\{\mathbf{r}\in\mathbb{C}^N \,:\, \mathbf{r}^{\mathrm{H}}\mathbf{K}\mathbf{r} = 1\right\},\ \mathbf{K} = \frac{\mathbf{C}_{\mathbf{r}|x_b}^{-1} - \mathbf{C}_{\mathbf{r}|x_a}^{-1}}{\ln\left|\mathbf{C}_{\mathbf{r}|x_a}\mathbf{C}_{\mathbf{r}|x_b}^{-1}\right|}. \tag{C.19}$$

Without loss of generality,[2] we assume the channel is full-rank (*i.e.* $\gamma_n > 0$ for $n = 1,\dots,N$). If condition (4.15) is fulfilled, $\mathbf{K}$ is always positive definite, ensuring the ellipsoid exists.

When $|x_a|^2 > |x_b|^2$, $\mathscr{D}$ is the closure of the set of points inside ellipsoid (C.19). With a change of variable $\mathbf{r} = |\mathbf{K}|^{\frac{1}{2N}}\mathbf{K}^{-\frac{1}{2}}\mathbf{s}$, we can map it to a N-dimensional closed ball with the same volume:

$$\mathscr{D} \mapsto \mathscr{U} = \left\{\mathbf{s}\in\mathbb{C}^N \,:\, \|\mathbf{s}\|^2 \leq |\mathbf{K}|^{-\frac{1}{N}}\right\}. \tag{C.20}$$

Then, the PEP becomes:

$$P_{a\to b} = \int_{\mathscr{U}} \frac{\exp\left(-\mathbf{s}^{\mathrm{H}}\Omega^{-1}\mathbf{s}\right)}{\pi^N|\mathbf{C}_{\mathbf{r}|x_a}|}\,\mathrm{d}\mathbf{s}, \quad \Omega = \frac{\mathbf{C}_{\mathbf{r}|x_a}\mathbf{K}}{|\mathbf{K}|^{\frac{1}{N}}}, \tag{C.21}$$

which is the integral of a multivariate complex Gaussian function. We can lower bound it with the integral in $\mathscr{U}$ of a narrower Gaussian function that is tangent to it along the direction with the lowest eigenvalue of Ω, named $\underline{\omega}$. It is proportional to the PDF of $\mathbf{t}\sim\mathscr{CN}(\mathbf{0}_N, \underline{\omega}\mathbf{I}_N)$:

$$P_{a\to b} \geq \int_{\mathscr{U}} \frac{\exp\left(-\|\mathbf{t}\|^2/\underline{\omega}\right)}{\pi^N|\mathbf{C}_{\mathbf{r}|x_a}|}\,\mathrm{d}\mathbf{t}. \tag{C.22}$$

This lower bound can be expressed in terms of the cumulative distribution function of a chi-squared random variable:

$$P_{a\to b} \geq |\mathbf{C}_{\mathbf{r}|x_a}|^{-1}\underline{\omega}^N \cdot \mathrm{P}\left(\|\mathbf{t}\|^2 \leq |\mathbf{K}|^{-\frac{1}{N}}\right) = |\mathbf{C}_{\mathbf{r}|x_a}|^{-1}\underline{\omega}^N \cdot \mathrm{F}_{\chi^2}\left(2\underline{\omega}^{-1}|\mathbf{K}|^{-\frac{1}{N}}; 2N\right). \tag{C.23}$$

If neither x_a nor x_b correspond to the null symbol, the previous bound in the limit $\alpha\to\infty$ is:

$$\lim_{\alpha\to\infty} P_{a\to b} \geq \mathrm{F}_{\chi^2}\left(\frac{2N\ln\frac{|x_a|^2}{|x_b|^2}}{\frac{|x_a|^2}{|x_b|^2}-1}; 2N\right), \tag{C.24}$$

which does not vanish for finite N or $\frac{|x_a|^2}{|x_b|^2}$. The existence of this lower bound on any PEP is sufficient to prove systems with $M\geq 3$ will display a fundamental error floor at high SNR.

[2] If the channel is rank deficient, an analogous analysis can be performed on a lower dimensional subspace.

The case for $M = 2$ bears some comment. In such scenario, the two possible PEPs depend on the null symbol. Setting $x_b = 0$, we can upper bound $P_{a\to b}$ in an analogous manner as (C.22). In this case, we use a wider Gaussian function proportional to the PDF of $\mathbf{t}' \sim \mathscr{CN}(\mathbf{0}_N, \overline{\omega}\mathbf{I}_N)$, with $\overline{\omega}$ being the maximum eigenvalue of $\mathbf{\Omega}$:

$$P_{a\to b} \leq |\mathbf{C}_{\mathbf{r}|x_a}|^{-1}\overline{\omega}^N \cdot \mathrm{F}_{\chi^2}\Big(2\overline{\omega}^{-1}|\mathbf{K}|^{-\frac{1}{N}}; 2N\Big). \tag{C.25}$$

Notice how the upper bound does now vanish as $\alpha \to \infty$:

$$\begin{aligned}\lim_{\alpha\to\infty} P_{a\to b} &\leq \lim_{\alpha\to\infty} \gamma_{\mathrm{MAX}}^N|\mathbf{\Gamma}|^{-1} \cdot \mathrm{F}_{\chi^2}\left(\frac{2\ln|\mathbf{C}_{\mathbf{r}|x_a}|}{|x_a|^2\gamma_{\mathrm{MAX}}}; 2N\right) \\ &= \lim_{\alpha\to\infty} \gamma_{\mathrm{MAX}}^N|\mathbf{\Gamma}|^{-1} \cdot \mathrm{F}_{\chi^2}(0; 2N) = 0.\end{aligned} \tag{C.26}$$

For the opposite case, in which $x_a = 0$, the PEP is obtained as (C.18) but now the integration is performed through the region outside of the ellipsoid $\partial\mathscr{D}$. With the same change of variable as in (C.20), we map $\mathscr{D}$ to the outside of a ball:

$$\mathscr{D} \mapsto \mathscr{V} = \mathbb{C}^N \setminus \mathscr{U} = \Big\{\mathbf{s} \in \mathbb{C}^N \,:\, \|\mathbf{s}\|^2 > |\mathbf{K}|^{-\frac{1}{N}}\Big\}. \tag{C.27}$$

The procedure to upper bound this PEP is the same as the one used previously, with the highest eigenvalue of $\mathbf{\Omega}$:

$$\begin{aligned}P_{a\to b} &\leq \int_{\mathscr{V}} \frac{\exp\big(-\|\mathbf{t}\|^2/\overline{\omega}\big)}{\pi^N}\,\mathrm{d}\mathbf{t} = \overline{\omega}^N \cdot \mathrm{P}\Big(\|\mathbf{t}\|^2 > |\mathbf{K}|^{-\frac{1}{N}}\Big) \\ &= \overline{\omega}^N \cdot \Big(1 - \mathrm{F}_{\chi^2}\Big(2\overline{\omega}^{-1}|\mathbf{K}|^{-\frac{1}{N}}; 2N\Big)\Big).\end{aligned} \tag{C.28}$$

Once again, this upper bound does vanish for increasing SNR:

$$\lim_{\alpha\to\infty} P_{a\to b} \leq 1 - \lim_{\alpha\to\infty} \mathrm{F}_{\chi^2}\Big(2\ln\Big|\mathbf{C}_{\mathbf{r}|x_b}\Big|; 2N\Big) = 0. \tag{C.29}$$

Therefore, we conclude that systems with $M = 2$ will asymptotically be error-free for increasing SNR. □

C.3 Lyapunov central limit theorem

Lyapunov CLT is a generalization of Lindeberg–Lévy CLT. It states that a sum of a sequence of independent random variables $\{\mathsf{v}_1, \ldots, \mathsf{v}_N\}$ with mean μ_n and variance σ_n^2 converges in distribution to a normal random variable if the following condition is fulfilled (*i.e.* Lyapunov's condition [196, Ch. 5]):

$$\lim_{N\to\infty} \frac{1}{s_N^\delta} \sum_{n=1}^{N} \mathrm{E}_{\mathsf{v}_n}\Big[|\mathsf{v}_n - \mu_n|^\delta\Big] = 0, \text{ for some } \delta > 2, \tag{C.30}$$

where $s_N^2 = \sum_{n=1}^N \sigma_n^2$.

In our case, $\mathsf{v}_n = a_n|\mathsf{r}_n|^2 + c_n$ are the summands in (4.49) and N is the number of antennas at the receiver. Since $\mathsf{r}_n \sim \mathscr{CN}(0, \varepsilon\gamma_n + 1)$, each v_n follows a shifted exponential distribution with density

$$f_{\mathsf{v}_n}(v) = \frac{1}{a_n(\varepsilon\gamma_n + 1)} \exp\left(-\frac{v - c_n}{a_n(\varepsilon\gamma_n + 1)}\right), \tag{C.31}$$

defined for $v \in [c_n, \infty)$. Its mean is $\mu_n = a_n(\varepsilon\gamma_n + 1) + c_n$ and its variance is $\sigma_n^2 = a_n^2(\varepsilon\gamma_n + 1)^2$.

We proceed to verify Lyapunov's condition for $\delta = 4$. Letting $\mathsf{w}_n = \mathsf{v}_n - \mu_n$, the fourth moment can be easily computed:

$$\mathrm{E}_{\mathsf{w}_n}\left[\mathsf{w}_n^4\right] = \int_{-\sigma_n}^{+\infty} \frac{w^4 \mathrm{e}^{-1}}{\sigma_n} \exp\left(\frac{-w}{\sigma_n}\right) \mathrm{d}w = 9\sigma_n^4. \tag{C.32}$$

Substituting into (C.30) and taking into account that $\gamma_n \geq 0$:

$$\lim_{N\to\infty} \frac{1}{s_N^4} \sum_{n=1}^{N} 9a_n^4(\varepsilon\gamma_n + 1)^4 \leq \lim_{N\to\infty} \frac{9}{N}\left(\frac{a_{\mathrm{M}}(\varepsilon\gamma_{\mathrm{M}} + 1)}{a_{\mathrm{m}}(\varepsilon\gamma_{\mathrm{m}} + 1)}\right)^4 = 0, \tag{C.33}$$

where $a_{\mathrm{M}} = \max_{n=1,\ldots,N} a_n$, $a_{\mathrm{m}} = \min_{n=1,\ldots,N} a_n$, and analogously for γ_{M} and γ_{m}. Therefore, condition (C.30) is satisfied and the CLT can be applied. □

APPENDIX D

Mean and variance of random complex normal quadratic forms

Let $\mathrm{q} = \mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z} + c$, with $\mathbf{A} = \mathbf{A}^{\mathrm{H}}$ and circularly-symmetric $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{C}_\mathbf{z})$. Note that by considering an arbitrary $\mathbf{C}_\mathbf{z}$, $\mathbf{A}$ can be assumed to be diagonal and real without loss of generality. The mean of q is given by:

$$\mu_\mathrm{q} = \mathrm{E}_\mathbf{z}\left[\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z} + c\right] = \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z}) + c, \tag{D.1}$$

which follows from the cyclic property of the trace.

Using that variance is shift invariant, we obtain:

$$\mathrm{var}_\mathbf{z}(\mathrm{q}) = \mathrm{var}_\mathbf{z}\left(\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z}\right). \tag{D.2}$$

Defining $\mathbf{z} = \mathbf{z}_\mathrm{R} + \mathrm{j}\mathbf{z}_\mathrm{I}$, where $\mathbf{z}_\mathrm{R} = \Re(\mathbf{z})$ and $\mathbf{z}_\mathrm{I} = \Im(\mathbf{z})$, the quadratic form $\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z}$ can be expressed as:

$$\begin{aligned}\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z} &= \left(\mathbf{z}_\mathrm{R}^{\mathrm{T}} - \mathrm{j}\mathbf{z}_\mathrm{I}^{\mathrm{T}}\right)\mathbf{A}(\mathbf{z}_\mathrm{R} + \mathrm{j}\mathbf{z}_\mathrm{I}) = \mathbf{z}_\mathrm{R}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{R} + \mathbf{z}_\mathrm{I}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{I} + \mathrm{j}\left(\mathbf{z}_\mathrm{R}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{I} - \mathbf{z}_\mathrm{I}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{R}\right) \\ &= \mathbf{z}_\mathrm{R}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{R} + \mathbf{z}_\mathrm{I}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{I}.\end{aligned} \tag{D.3}$$

Due to $\mathbf{z}$ being circularly-symmetric, $\mathbf{z}_\mathrm{R}$, $\mathbf{z}_\mathrm{I} \sim \mathcal{N}(\mathbf{0}_N, \mathrm{C}_\mathrm{z}/2)$ are independent, and we can compute $\mathrm{var}_\mathbf{z}\left(\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z}\right)$ as twice the variance of a quadratic form in real random variables [129, Ch. 3]:

$$\mathrm{var}_\mathbf{z}\left(\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z}\right) = 2\,\mathrm{var}_{\mathbf{z}_\mathrm{R}}\left(\mathbf{z}_\mathrm{R}^{\mathrm{T}}\mathbf{A}\mathbf{z}_\mathrm{R}\right) = 4\,\mathrm{tr}\left(\mathbf{A}\frac{\mathbf{C}_\mathbf{z}}{2}\mathbf{A}\frac{\mathbf{C}_\mathbf{z}}{2}\right) = \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z}\mathbf{A}\mathbf{C}_\mathbf{z}). \tag{D.4}$$

Substituting in (D.2):

$$\mathrm{var}_\mathbf{z}(\mathrm{q}) = \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z}\mathbf{A}\mathbf{C}_\mathbf{z}). \tag{D.5}$$

Finally, the second-order moment can be easily computed:

$$\mathrm{E}_\mathbf{z}\left[\mathrm{q}^2\right] = \mathrm{var}_\mathbf{z}(\mathrm{q}) + \mu_\mathrm{q}^2 = \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z}\mathbf{A}\mathbf{C}_\mathbf{z}) + \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z})^2 + c^2 + 2c \cdot \mathrm{tr}(\mathbf{A}\mathbf{C}_\mathbf{z}). \tag{D.6}$$

APPENDIX E

Energy statistic Cramér–Rao bound

Given any unbiased estimator $\hat{\varepsilon}$ of ε, the CRB is given by the reciprocal of the Fisher information [126, Sec. 3.4]:

$$\operatorname{var}_{\varepsilon}(\hat{\varepsilon}) \geq \frac{1}{\mathrm{J}(\varepsilon)}. \tag{E.1}$$

Under mild conditions, it can be obtained from the log-likelihood function as

$$\mathrm{J}(\varepsilon) = -\mathrm{E}_{\mathbf{r}|\varepsilon}\left[\frac{\partial^2 l(\varepsilon)}{\partial \varepsilon^2}\right]. \tag{E.2}$$

In our case, the log-likelihood function is

$$l(\varepsilon) = -\ln\left(\pi^N \left|\mathbf{C}_{\mathbf{r}|\varepsilon}\right|\right) - \mathbf{r}^{\mathrm{H}}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\mathbf{r}, \quad \mathbf{C}_{\mathbf{r}|\varepsilon} = \varepsilon\boldsymbol{\Gamma} + \mathbf{I}_N, \tag{E.3}$$

with second derivative

$$\frac{\partial^2 l(\varepsilon)}{\partial \varepsilon^2} = \sum_{n=1}^{N} \gamma_n^2 \frac{\varepsilon\gamma_n + 1 - 2|r_n|^2}{(\varepsilon\gamma_n + 1)^3}. \tag{E.4}$$

The Fisher information is then

$$\mathrm{J}(\varepsilon) = -\sum_{n=1}^{N} \gamma_i^2 \frac{\varepsilon\gamma_n + 1 - 2\mathrm{E}_{\mathbf{r}|\varepsilon}\left[|\mathrm{r}_n|^2\right]}{(\varepsilon\gamma_n + 1)^3} = \sum_{n=1}^{N} \frac{\gamma_n^2}{(\varepsilon\gamma_n + 1)^2}. \tag{E.5}$$

Finally, we can express the CRB in matrix form:

$$\operatorname{var}_{\varepsilon}(\hat{\varepsilon}) \geq \left\|\boldsymbol{\Gamma}(\varepsilon\boldsymbol{\Gamma} + \mathbf{I}_N)^{-1}\right\|_{\mathrm{F}}^{-2} = \left\|\boldsymbol{\Gamma}\mathbf{C}_{\mathbf{r}|\varepsilon}^{-1}\right\|_{\mathrm{F}}^{-2}. \tag{E.6}$$

APPENDIX F

Permutation modulation spectral efficiency bounds

In order to prove $\lim_{K\to\infty} \mathrm{R}_K(\{K_m\}) = \mathrm{H}(\{p_m\})^-$ (*i.e.* from below) we leverage the following bounds on the log-factorial function derived from the Stirling's series [31, Ch. 27]:

$$\overbrace{\log\frac{\sqrt{2\pi\alpha}\alpha^\alpha}{\mathrm{e}^\alpha}+\frac{\log\mathrm{e}}{12\alpha+1}}^{g_{\mathrm{low}}(\alpha)} < \log(\alpha!) < \overbrace{\log\frac{\sqrt{2\pi\alpha}\alpha^\alpha}{\mathrm{e}^\alpha}+\frac{\log\mathrm{e}}{12\alpha}}^{g_{\mathrm{up}}(\alpha)} \tag{F.1}$$

which are asymptotically tight as $\alpha\to\infty$. Moreover, the terms $\log\mathrm{e}/12\alpha+1$ and $\log\mathrm{e}/12\alpha$ vanish asymptotically and the bounds coincide. Thus, an upper bound for the SE in (5.5) is

$$\mathrm{R}_K(\{K_m\}) < \frac{1}{K}\big(g_{\mathrm{up}}(K)-\sum_{m=1}^{M} g_{\mathrm{low}}(K_m)\big). \tag{F.2}$$

Substituting (F.1) onto it yields

$$\mathrm{R}_K(\{K_m\}) < \mathrm{H}(\{p_m\}) + \frac{1}{K}\left(\frac{\log(2\pi K)}{2}+\frac{\log\mathrm{e}}{12K}-\sum_{m=1}^{M}\frac{\log(2\pi K_m)}{2}+\frac{\log\mathrm{e}}{12K_m+1}\right) \tag{F.3}$$

which tends to $\mathrm{H}(\{p_m\})$ for $K\to\infty$. In addition, $\exists L>0$ such that $\forall K>L$,

$$\frac{\log(2\pi K)}{2}+\frac{\log\mathrm{e}}{12K}-\sum_{m=1}^{M}\frac{\log(2\pi K_m)}{2}+\frac{\log\mathrm{e}}{12K_m+1}\le 0, \tag{F.4}$$

since $\log(x)$ is subadditive for $x\ge 1$. Hence, $\mathrm{R}_K(\{K_m\})\le\mathrm{H}(\{p_m\})$ and $\lim_{K\to\infty}\mathrm{R}_K(\{K_m\}) = \mathrm{H}(\{p_m\})^-$.

To maximize the SE (defined in (5.5)) in terms of $\{K_m\}$ in the non-asymptotic case, the following optimization problem must be solved with Lagrange multipliers:

$$\arg\min_{\{K_m\}} \sum_{m=1}^{M}\log\big(\Gamma(K_m+1)\big) \text{ s.t. } \sum_{m=1}^{M} K_m = K, \tag{F.5}$$

where $\Gamma(\cdot)$ is the gamma function [31, Ch. 19]. By the Bohr–Mollerup theorem [31, Ch. 19], it is known that $\Gamma(x)$ is log-convex for $x>0$. Then, by [163, Proposition 3.C.1], the sum of convex functions (F.5) is minimized by the uniform policy: $K_m = K/M$ for all m. The resulting SE of such alphabet is $\mathrm{R}_K(M) = \frac{1}{K}(\log(K!) - M\log(K/M!))$. □

Bibliography

[1] J. Romeu and A. Elias, "Early proposals of wireless telegraphy in Spain: Francisco Salva Campillo (1751-1828)", in *IEEE Antennas and Propagation Society International Symposium. 2001 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting (Cat. No.01CH37229)*, vol. 1, 2001, pp. 10–13.

[2] C. E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[3] M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, "Twelve scientific challenges for 6G: Rethinking the foundations of communications theory", *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 868–904, 2023.

[4] M. Shafi et al., "5G: A tutorial overview of standards, trials, challenges, deployment, and practice", *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[5] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity", *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, 2018.

[6] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond", International Telecommunication Union, M.2160-0, 2023.

[7] H. Lu et al., "A tutorial on near-field XL-MIMO communications toward 6G", *IEEE Communications Surveys & Tutorials*, vol. 26, no. 4, pp. 2213–2257, 2024.

[8] E. Björnson et al., *Towards 6G MIMO: Massive spatial multiplexing, dense arrays, and interplay between electromagnetics and processing*, 2024. arXiv: 2401.02844.

[9] A. Goldsmith, *Wireless communications*. Cambridge, UK: Cambridge University Press, 2005, ISBN: 978-1-60119-764-1.

[10] D. Cuevas, "Advanced Grassmannian constellation designs for noncoherent MIMO communications", Ph.D. dissertation, Universidad de Cantabria, 2024.

[11] R. H. Gohary and H. Yanikomeroglu, "Noncoherent MIMO signaling for block-fading channels: Approaches and challenges", *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 80–88, 2019.

[12] X. Gao, J. Zhang, H. Chen, Z. Dong, and B. Vucetic, "Energy-efficient and low-latency massive SIMO using noncoherent ML detection for industrial IoT communications", *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6247–6261, 2019.

[13] M. Vaezi et al., "Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G", *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, 2022.

[14] K. Ngo et al., "Noncoherent MIMO communications: Theoretical foundation, design approaches, and future challenges", *IEEE Open Journal of Vehicular Technology*, vol. 7, pp. 381–401, 2026.

[15] X. Cai, C. Huang, P. Chen, E. Basar, and C. Yuen, "Design of non-coherent RIS-empowered DCSK with two-level nested index modulation", *IEEE Transactions on Wireless Communications*, vol. 24, no. 4, pp. 3044–3058, 2025.

[16] A. A. Siddiqui, E. Bedeer, H. H. Nguyen, and R. Barton, "Spectrally-efficient modulation on conjugate-reciprocal zeros (SE-MOCZ) for non-coherent short packet communications", *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 2226–2240, 2024.

[17] A. Pizzo, L. Sanguinetti, and T. L. Marzetta, "Fourier plane-wave series expansion for holographic MIMO communications", *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 6890–6905, 2022.

[18] A. Pizzo, L. Sanguinetti, and T. L. Marzetta, "Spatial characterization of electromagnetic random channels", *IEEE Open Journal of the Communications Society*, vol. 3, pp. 847–866, 2022.

[19] A. Martí, J. Riba, M. Lamarca, and X. Gràcia, "Asymptotic analysis of near-field coupling in massive MISO and massive SIMO systems", *IEEE Communications Letters*, vol. 28, no. 8, pp. 1929–1933, 2024.

[20] M. Vilà-Insa, A. Martí, J. Riba, and M. Lamarca, "Quadratic detection in noncoherent massive SIMO systems over correlated channels", *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14 259–14 272, 2024.

[21] A. Martí, M. Vilà-Insa, J. Riba, and M. Lamarca, "Constellation design for quadratic detection in noncoherent massive SIMO communications", in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lucca, Italy, 2024, pp. 566–570.

[22] M. Vilà-Insa, A. Martí, M. Lamarca, and J. Riba, "Low-complexity detection of permutational index modulation for noncoherent communications", *IEEE Wireless Communications Letters*, vol. 14, no. 10, pp. 3059–3063, 2025.

[23] A. Martí, L. Sanguinetti, J. Riba, and M. Lamarca, "Coherent and noncoherent detection in dense arrays: Can we ignore mutual coupling?", in *2025 33rd European Signal Processing Conference (EUSIPCO)*, Palermo, Italy, 2025, pp. 2027–2031.

[24] A. Martí, L. Sanguinetti, M. Lamarca, and J. Riba, "Harnessing wavefront curvature and spatial correlation in noncoherent MIMO communications", *IEEE Wireless Communications Letters*, vol. 14, no. 8, pp. 2461–2465, 2025.

[25] A. Martí, J. Portell, J. Riba, and O. Mas, "Context-aware lossless and lossy compression of radio frequency signals", *Sensors*, vol. 23, no. 7, 2023.

[26] A. Martí, F. de Cabrera, and J. Riba, "On the estimation of Tsallis entropy and a novel information measure based on its properties", *IEEE Signal Processing Letters*, vol. 30, pp. 818–822, 2023.

[27] M. D. Migliore, "On electromagnetics and information theory", *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 10, pp. 3188–3200, 2008.

[28] Y. Liu et al., "Near-field communications: A tutorial review", *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1999–2049, 2023.

[29] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO communication*, 1st ed. Cambridge, UK: Cambridge University Press, 2018, ISBN: 978-1-139-04927-6.

[30] C. A. Balanis, *Antenna theory: Analysis and design*, 4th ed. Hoboken, USA: Wiley, 2016, ISBN: 978-1-118-64206-1.

[31] M. Spivak, *Calculus*, 3rd ed. Houston, USA: Publish or Perish, 1994, ISBN: 978-0-914098-89-8.

[32] E. Björnson, Ö. T. Demir, and L. Sanguinetti, "A primer on near-field beamforming for arrays and reconfigurable intelligent surfaces", in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 105–112.

[33] A. Kosasih and E. Björnson, "Finite beam depth analysis for large arrays", *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 10 015–10 029, 2024.

[34] B. Friedlander, "Localization of signals in the near-field of an antenna array", *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 3885–3893, 2019.

[35] W.-X. Long et al., "Channel estimation for 6G near-field wireless communications: A comprehensive survey", *IEEE Communications Surveys & Tutorials*, vol. 28, pp. 5496–5531, 2026.

[36] R. L. Haupt, *Antenna arrays: A computational approach*. Hoboken, USA: Wiley, 2010, ISBN: 978-0-470-40775-2.

[37] Ö. T. Demir, A. Kosasih, and E. Björnson, "Spatial correlation modeling and RS-LS estimation of near-field channels with uniform planar arrays", in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2024, pp. 236–240.

[38] 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, 2017.

[39] Z. Dong and Y. Zeng, "Near-field spatial correlation for extremely large-scale array communications", *IEEE Communications Letters*, vol. 26, no. 7, pp. 1534–1538, 2022.

[40] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces", *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, 2021.

[41] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Channel modeling and channel estimation for holographic massive MIMO with planar arrays", *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 997–1001, 2022.

[42] W. Han, Z. Dong, H. Chen, and X. Gao, "Constellation design for energy-based noncoherent massive SIMO systems over correlated channels", *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2165–2169, 2022.

[43] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix", *IEEE Communications Letters*, vol. 5, no. 9, pp. 369–371, 2001.

[44] V. Croisfelt Rodrigues, J. C. Marinello, and T. Abrão, "Exponential spatial correlation with large-scale fading variations in massive MIMO channel estimation", *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 5, e3563, 2019.

[45] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model", in *2014 IEEE International Conference on Communications (ICC)*, Sydney, Australia, 2014, pp. 5865–5870.

[46] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, 1st ed. New York, USA: Cambridge University Press, 2005, ISBN: 978-0-521-84527-4.

[47] L. Zheng and D. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel", *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[48] W. Hedhly, O. Amin, B. Shihada, and M.-S. Alouini, "Hyperloop communications: Challenges, advances, and approaches", *IEEE Open Journal of the Communications Society*, vol. 2, pp. 2413–2435, 2021.

[49] L. Jing, E. De Carvalho, P. Popovski, and À. O. Martínez, "Design and performance analysis of noncoherent detection systems with massive receiver arrays", *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5000–5010, 2016.

[50] K. T. Truong and R. W. Heath Jr., "Effects of channel aging in massive MIMO systems", *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, 2013.

[51] C. Xu et al., "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures", *IEEE Access*, vol. 7, pp. 178 246–178 299, 2019.

[52] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question", *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[53] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems", in *2009 IEEE International Symposium on Information Theory*, 2009, pp. 2184–2188.

[54] F. Riera-Palou, G. Femenias, A. García Armada, and A. Pérez-Neira, "Clustered cell-free massive MIMO", in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018.

[55] A. Mishra, Y. Mao, C. K. Thomas, L. Sanguinetti, and B. Clerckx, "Mitigating intra-cell pilot contamination in massive MIMO: A rate splitting approach", *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3472–3487, 2023.

[56] M. K. Saeed, A. E. Kamal, and A. Khokhar, "Mitigating pilot contamination and enabling IoT scalability in massive MIMO systems", in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, 2023, pp. 1620–1625.

[57] M. Duran, F. Riera-Palou, G. Femenias, H. Q. Ngo, and M. Fernández-Getino García, "Superimposed training in cell-free massive MIMO: Is it really worth it?", in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 2024.

[58] M. Chowdhury, A. Manolakos, and A. J. Goldsmith, "Coherent versus noncoherent massive SIMO systems: Which has better performance?", in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 1691–1696.

[59] Y. Han and J. Lee, "Grassmannian training for massive MIMO cellular networks", in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 193–197.

[60] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation", *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 2041–2052, 2000.

[61] B. Hughes, "Differential space-time modulation", *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.

[62] V. Monzón Baeza, A. García Armada, W. Zhang, M. El-Hajjar, and L. Hanzo, "A noncoherent multiuser large-scale SIMO system relying on M-ary DPSK and BICM-ID", *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1809–1814, 2018.

[63] V. Monzón Baeza and A. García Armada, "Non-coherent massive SIMO system based on m-DPSK for Rician channels", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2413–2426, 2019.

[64] L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, wi-fi and WiMAX: Coherent versus non-coherent and cooperative turbo-transceivers*. Chichester, UK: Wiley, 2010, ISBN: 978-0-470-71175-0.

[65] J. M. Lee, *Introduction to smooth manifolds*, 2nd ed. New York, USA: Springer, 2013, ISBN: 978-1-4419-9982-5.

[66] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading", *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[67] J. G. Proakis and M. Salehi, *Digital communications*, 5th ed. Boston, USA: McGraw-Hill, 2008, ISBN: 978-0-07-295716-7.

[68] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels", *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 331–340, 2012.

[69] O. Henkel, "Sphere-packing bounds in the grassmann and stiefel manifolds", *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3445–3456, 2005.

[70] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry", *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5205–5219, 2018.

[71] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling", *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[72] A. Manolakos, M. Chowdhury, and A. Goldsmith, "Energy-based modulation for noncoherent massive SIMO systems", *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7831–7846, 2016.

[73] Z. Wang et al., "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications", *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1560–1605, 2024.

[74] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces", *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2746–2758,

[75] C. Huang et al., "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends", *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, 2020.

[76] A. Pizzo and A. Lozano, "Mutual coupling in holographic MIMO: Physical modeling and information-theoretic analysis", *IEEE Journal on Selected Areas in Information Theory*, vol. 6, pp. 111–126, 2025.

[77] R. Janaswamy, "Effect of element mutual coupling on the capacity of fixed length linear arrays", *IEEE Antennas and Wireless Propagation Letters*, vol. 1, pp. 157–160, 2002.

[78] J. W. Wallace and M. A. Jensen, "Mutual coupling in MIMO wireless systems: A rigorous network theory analysis", *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1317–1325, 2004.

[79] M. T. Ivrlač and J. A. Nossek, "Toward a circuit theory of communication", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1663–1683, 2010.

[80] M. T. Ivrlač and J. A. Nossek, "The multiport communication theory", *IEEE Circuits and Systems Magazine*, vol. 14, no. 3, pp. 27–44, 2014.

[81] D. M. Pozar, *Microwave engineering*, 4th ed. Hoboken, USA: Wiley, 2012, ISBN: 978-0-470-63155-3.

[82] F. Gustrau, *RF and microwave engineering: Fundamentals of wireless communications*. Weinheim, Germany: Wiley, 2025, ISBN: 978-1-394-28346-0.

[83] A. A. D'Amico and L. Sanguinetti, "Holographic MIMO communications: What is the benefit of closely spaced antennas?", *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 13 826–13 840, 2024.

[84] B. Tadele, V. Shyianov, F. Bellili, and A. Mezghani, "Channel estimation with tightly-coupled antenna arrays", in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[85] A. Abrardo, A. Toccafondi, and M. Di Renzo, "Analysis and optimization of reconfigurable intelligent surfaces based on s-parameters multiport network theory", in *2024 18th European Conference on Antennas and Propagation (EuCAP)*, 2024.

[86] M. T. Ivrlač and J. A. Nossek, "Receive antenna gain of uniform linear arrays of isotrops", in *2009 IEEE International Conference on Communications*, 2009.

[87] T. Laas, J. A. Nossek, and W. Xu, "Limits of transmit and receive array gain in massive MIMO", in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[88] M. T. Ivrlač and J. A. Nossek, "On the diversity performance of compact antenna arrays", in *Proc. 30th Gen. Assembly Int. Union Radio Sci.(URSI)*, 2011.

[89] M. T. Ivrlač and J. A. Nossek, "On multistreaming with compact antenna arrays", in *2011 International ITG Workshop on Smart Antennas*, 2011.

[90] T. Laas, J. A. Nossek, S. Bazzi, and W. Xu, "On reciprocity in physically consistent TDD systems with coupled antennas", *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6440–6453, 2020.

[91] B. D. O. Anderson and R. W. Newcomb, "On reciprocity in linear time-invariant networks", *Stanford Electronics Laboratories*, no. 6558-2, 1965.

[92] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed. New York, USA: Cambridge University Press, 2017, ISBN: 978-0-521-54823-6.

[93] M. T. Ivrlač and J. A. Nossek, "On physical limits of massive MISO systems", in *WSA 2016; 20th International ITG Workshop on Smart Antennas*, 2016.

[94] A. Mezghani et al., "Reincorporating circuit theory into information theory", *IEEE BITS the Information Theory Magazine*, vol. 4, no. 1, pp. 40–58, 2024.

[95] S. Phang et al., "Near-field MIMO communication links", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 9, pp. 3027–3036, 2018.

[96] K. F. Warnick, B. Woestenburg, L. Belostotski, and P. Russer, "Minimizing the noise penalty due to mutual coupling for a receiving array", *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 6, pp. 1634–1644, 2009.

[97] S. A. Schelkunoff and H. T. Friis, *Antennas: Theory and practice*. New York, USA: Wiley, 1952, ISBN: 978-0-471-75900-3.

[98] H. Yordanov, M. T. Ivrlač, P. Russer, and J. A. Nossek, "Arrays of isotropic radiators – a field-theoretic justification", in *2009 International ITG Workshop on Smart Antenna*, 2009.

[99] W. K. Kahn and H. Kurss, "Minimum-scattering antennas", *IEEE Transactions on Antennas and Propagation*, vol. 13, no. 5, pp. 671–675, 1965.

[100] W. Wasylkiwskyj and W. K. Kahn, "Theory of mutual coupling among minimum-scattering antennas", *IEEE Transactions on Antennas and Propagation*, vol. 18, no. 2, pp. 204–216, 1970.

[101] S. J. Orfanidis, *Electromagnetic waves and antennas*. 2016.

[102] R. M. Gray, "Toeplitz and circulant matrices: A review", *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2005.

[103] D. Knuth, *The art of computer programming, vol. 1: Fundamental algorithms*, 3rd ed. Boston, USA: Addison–Wesley, 1997, ISBN: 978-0-201-89683-1.

[104] R. J. Beerends, H. G. ter Morsche, J. C. van den Berg, and E. M. van de Vrie, *Fourier and laplace transforms*. Cambridge, UK: Cambridge University Press, 2003, ISBN: 978-0-511-67510-2.

[105] B. Lehmeyer, M. T. Ivrlač, and J. A. Nossek, "LNA noise parameter measurement", in *2015 European Conference on Circuit Theory and Design (ECCTD)*, 2015.

[106] A. Mahmood, S. F. Abedin, T. Sauter, M. Gidlund, and K. Landernas, "Factory 5G: A review of industry-centric features and deployment options", *IEEE Industrial Electronics Magazine*, vol. 16, no. 2, pp. 24–34, 2022.

[107] S. Mumtaz et al., "Massive internet of things for industrial applications: Addressing wireless IIoT connectivity challenges and ecosystem fragmentation", *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 28–33, 2017.

[108] A. Bana, M. Angjelichinoski, E. De Carvalho, and P. Popovski, "Massive MIMO for ultra-reliable communications with constellations for dual coherent-noncoherent detection", in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018.

[109] G. Han, Z. Dong, J. Zhang, and X. Mu, "8-QAM division for uplink massive SIMO systems", *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 48–52, 2021.

[110] S. Li, Z. Dong, H. Chen, and X. Guo, "Constellation design for noncoherent massive SIMO systems in URLLC applications", *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4387–4401, 2021.

[111] S. T. Duong, H. H. Nguyen, and E. Bedeer, "Multi-level design for multiple-symbol non-coherent unitary constellations for massive SIMO systems", *IEEE Wireless Communications Letters*, vol. 12, no. 8, pp. 1349–1353, 2023.

[112] H. Xie, W. Xu, H. Q. Ngo, and B. Li, "Non-coherent massive MIMO systems: A constellation design approach", *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3812–3825, 2020.

[113] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets", *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[114] S. Park and R. W. Heath Jr., "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach", *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8047–8062, 2018.

[115] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Super-resolution channel estimation for arbitrary arrays in hybrid millimeter-wave massive MIMO systems", *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 947–960, 2019.

[116] A. Lu, Y. Chen, and X. Gao, "2D beam domain statistical CSI estimation for massive MIMO uplink", *IEEE Transactions on Wireless Communications*, vol. 23, no. 1, pp. 749–761, 2024.

[117] W. Long, M. Moretti, M. Morelli, L. Sanguinetti, and R. Chen, "Parametric near-field MMSE channel estimation for sub-THz XL-MIMO systems", in *2025 33rd European Signal Processing Conference (EUSIPCO)*, Palermo, Italy, 2025, pp. 1198–1202.

[118] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.

[119] I. Abou-Faycal, M. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels", *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.

[120] M. C. Gursoy, H. V. Poor, and S. Verdu, "Noncoherent Rician fading channel-part II: Spectral efficiency in the low-power regime", *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2207–2221, 2005.

[121] M. Chowdhury, A. Manolakos, and A. Goldsmith, "Scaling laws for noncoherent energy-based communications in the SIMO MAC", *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1980–1992, 2016.

[122] B. C. Levy, *Principles of signal detection and parameter estimation*. Boston, USA: Springer US, 2008, ISBN: 978-0-387-76542-6.

[123] A. Pastore, "A simple capacity lower bound for communication with superimposed pilots", in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018.

[124] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Hoboken, USA: Wiley, 2006, ISBN: 978-0-471-24195-9.

[125] J. K. Blitzstein and J. Hwang, *Introduction to probability*. Boca Raton, USA: CRC Press, 2015, ISBN: 978-1-4987-0478-6.

[126] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Englewood Cliffs, USA: Prentice-Hall, 1993, ISBN: 978-0-13-345711-7.

[127] J. Villares and G. Vazquez, "Best quadratic unbiased estimator (BQUE) for timing and frequency synchronization", in *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (SSP 2001)*, 2001, pp. 413–416.

[128] J. Villares, "Sample covariance based parameter estimation for digital communications", Ph.D. dissertation, Universitat Politècnica de Catalunya, 2005.

[129] A. M. Mathai and S. B. Provost, *Quadratic forms in random variables: Theory and applications*. New York, USA: Marcel Dekker, 1992, ISBN: 978-0-8247-8691-5.

[130] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback", *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1188–1204, 2008.

[131] E. Björnson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems", *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 4027–4041, 2009.

[132] D. R. Cox, *Renewal theory*. London, UK: Methuen, 1967, ISBN: 978-0-412-20570-5.

[133] M. Brehler and M. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers", *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2383–2399, 2001.

[134] P. Lombardo, G. Fedele, and M. M. Rao, "MRC performance for binary signals in Nakagami fading with general branch correlation", *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 44–52, 1999.

[135] A. Bana, L. Sanguinetti, E. De Carvalho, and P. Popovski, "Outage analysis of downlink URLLC in massive MIMO systems with power allocation", in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 1394–1398.

[136] P. Popovski et al., "Wireless access in ultra-reliable low-latency communication (URLLC)", *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, 2019.

[137] B. Knott, M. Chowdhury, A. Manolakos, and A. Goldsmith, "Benefits of coding in a noncoherent massive SIMO system", in *2015 IEEE International Conference on Communications (ICC)*, IEEE, 2015.

[138] A. Fazeli, H. H. Nguyen, and M. Hanif, "Generalized OFDM-IM with noncoherent detection", *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4464–4479, 2020.

[139] A. Fazeli, H. H. Nguyen, H. D. Tuan, and H. V. Poor, "Non-coherent multi-level index modulation", *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2240–2255, 2022.

[140] M. Vilà-Insa and J. Riba, "Singular detection in noncoherent communications", *IEEE Wireless Communications Letters*, vol. 14, no. 4, pp. 1164–1168, 2025, ISSN: 2162-2345.

[141] A. Manolakos, M. Chowdhury, and A. Goldsmith, "Constellation design in noncoherent massive SIMO systems", in *2014 IEEE Global Communications Conference*, 2014, pp. 3690–3695.

[142] H. Xie, W. Xu, W. Xiang, K. Shao, and S. Xu, "Non-coherent massive SIMO systems in ISI channels: Constellation design and performance analysis", *IEEE Systems Journal*, vol. 13, no. 3, pp. 2252–2263, 2019.

[143] M. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of noncoherent space-time constellations", *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2332–2351, 2003.

[144] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*, 5th ed. Cham, Switzerland: Springer, 2021, ISBN: 978-3-030-85450-8.

[145] M. Lubin et al., "JuMP 1.0: Recent improvements to a modeling language for mathematical optimization", *Mathematical Programming Computation*, 2023.

[146] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing", *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.

[147] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK: Cambridge University Press, 2004, ISBN: 978-0-521-83378-3.

[148] C. Yuan, W. Xu, and X. Hu, "Adaptive modulation for ED-based non-coherent massive SIMO systems", *IEEE Access*, vol. 9, pp. 132 311–132 318, 2021.

[149] 3GPP TSG-RAN WG1, "TSGR1#17(00)1395: Adaptive modulation and coding (AMC)", 2000.

[150] K. Song, A. Ekbal, S. T. Chung, and J. M. Cioffi, "Adaptive modulation and coding (AMC) for bit-interleaved coded OFDM (BIC-OFDM)", *IEEE Transactions on Wireless Communications*, vol. 5, no. 7, pp. 1685–1694, 2006.

[151] J. Riba, J. Villares, and G. Vazquez, "A nondata-aided SNR estimation technique for multilevel modulations exploiting signal cyclostationarity", *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5767–5778, 2010.

[152] F. de Cabrera and J. Riba, "Entropy-based non-data-aided SNR estimation", in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 731–735.

[153] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: A survey", *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 315–348, 2019.

[154] M. Wen, Q. Li, and X. Cheng, *Index modulation for OFDM communications systems*. Singapore, Singapore: Springer Singapore, 2021, ISBN: 978-981-15-9406-9.

[155] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation", *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, 2014.

[156] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation", *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[157] J. Choi, "Noncoherent OFDM-IM and its performance analysis", *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 352–360, 2018.

[158] D. Slepian, "Permutation modulation", *Proceedings of the IEEE*, vol. 53, no. 3, pp. 228–236, 1965.

[159] A. Nordio and E. Viterbo, "Permutation modulation for fading channels", in *10th International Conference on Telecommunications, 2003. ICT 2003.*, vol. 2, 2003, pp. 1177–1183.

[160] S. Muller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and OFDM", *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2442–2450, 2002.

[161] E. Biglieri and M. Elia, "Optimum permutation modulation codes and their asymptotic performance (Corresp.)", *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 751–753, 1976.

[162] E. Viterbo, "Trellis decoding of permutation modulations", in *IEEE International Symposium on Information Theory, 2003. Proceedings.*, 2003, pp. 393–393.

[163] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of majorization and its applications*. New York, USA: Springer, 2011, ISBN: 978-0-387-68276-1.

[164] G. Bacci, L. Sanguinetti, and E. Björnson, "Spherical wavefronts improve MU-MIMO spectral efficiency when using electrically large arrays", *IEEE Wireless Communications Letters*, vol. 12, no. 7, pp. 1219–1223, 2023.

[165] L. Le Magoarou and S. Paquelet, "Parametric channel estimation for massive MIMO", in *2018 IEEE Statistical Signal Processing Workshop (SSP)*, 2018, pp. 30–34.

[166] A. Adhikary et al., "Joint spatial division and multiplexing for mm-wave channels", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1239–1255, 2014.

[167] H. L. Van Trees, K. L. Bell, and Z. Tian, *Detection, estimation, and modulation theory*, 2nd ed. Hoboken, USA: Wiley, 2013, ISBN: 978-0-470-54296-5.

[168] Z. Wu and L. Dai, "Multiple access for near-field communications: SDMA or LDMA?", *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 6, pp. 1918–1935, 2023.

[169] P. Ramezani and E. Björnson, "Near-field beamforming and multiplexing using extremely large aperture arrays", in *Fundamentals of 6G Communications and Networking*, X. Lin, J. Zhang, Y. Liu, and J. Kim, Eds., Cham, Switzerland: Springer, 2024, pp. 317–349, ISBN: 978-3-031-37920-8.

[170] R. G. Bartle and D. R. Sherbert, *Introduction to real analysis*, 4th ed. Hoboken, USA: Wiley, 2011, ISBN: 978-0-471-43331-6.

[171] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints", *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[172] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations", *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962–1973, 2000.

[173] J. Miao and A. Ben-Israel, "On principal angles between subspaces in $\mathbb{R}^n$", *Linear Algebra and its Applications*, vol. 171, pp. 81–98, 1992.

[174] A. Galántai and C. J. Hegedűs, "Jordan's principal angles in complex vector spaces", *Numerical Linear Algebra with Applications*, vol. 13, no. 7, pp. 589–598, 2006.

[175] K. G. Seddik et al., "Multi-resolution multicasting over the Grassmann and Stiefel manifolds", *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5296–5310, 2017.

[176] C. Xu et al., "Noncoherent orthogonal time frequency space modulation", *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 10 072–10 090, 2024.

[177] H. Zhao, D. He, Z. Kang, and H. Wang, "Orthogonal time frequency space (OTFS) with dual-mode index modulation", *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 991–995, 2021.

[178] D. Feng, J. Zheng, B. Bai, J. Jiang, and H. Chu, "Generalized index modulation for MIMO-OTFS transmission", *IEEE Wireless Communications Letters*, vol. 12, no. 5, pp. 907–911, 2023.

[179] B. A. Ozden et al., *A survey of OTFS-based index modulation techniques: Challenges, benefits, and future directions for 6G and beyond*, 2025. arXiv: 2510.20265.

[180] X. Li, Z. Dong, Y. Zeng, S. Jin, and R. Zhang, "Multi-user modular XL-MIMO communications: Near-field beam focusing pattern and user grouping", *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 13 766–13 781, 2024.

[181] J. P. González-Coma, S. Fernández, and F. Javier López-Martínez, "User selection in near-field gigantic MIMO systems with modular arrays", *IEEE Transactions on Communications*, vol. 73, no. 12, pp. 13 357–13 368, 2025.

[182] R. E. Thomas, A. J. Rosa, and G. J. Toussaint, *The analysis and design of linear circuits*, 8th ed. Hoboken, USA: Wiley, 2016, ISBN: 978-1-119-23538-5.

[183] U. Madhow, *Introduction to communication systems*. Cambridge, UK: Cambridge University Press, 2014, ISBN: 978-1-107-02277-5.

[184] A. Klenke, *Probability theory: A comprehensive course*, 2nd ed. 2014. London, UK: Springer, 2014, ISBN: 978-1-4471-5361-0.

[185] P. Russer, *Electromagnetics, microwave circuit and antenna design for communications engineering*, 2nd ed. Boston, USA: Artech House, 2006, ISBN: 978-1-58053-907-4.

[186] H. A. Haus et al., "Representation of noise in linear twoports", *Proceedings of the IRE*, vol. 48, no. 1, pp. 69–74, 1960.

[187] M. Mokari and W. Patience, "A new method of noise parameter calculation using direct matrix analysis", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 39, no. 9, pp. 767–771, 1992.

[188] K. Ngo, S. Yang, M. Guillaud, and A. Decurninge, "Joint constellation design for noncoherent MIMO multiple-access channels", *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7281–7305, 2022.

[189] Z. Dong, J. Zhang, and L. Huang, "Multi-users space-time modulation with QAM division for massive uplink communications", in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017.

[190] S. Li, J. Zhang, and X. Mu, "Design of optimal noncoherent constellations for SIMO systems", *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5706–5720, 2019.

[191] J. Zhang, F. Huang, and S. Ma, "Full diversity blind space-time block codes", *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6109–6133, 2011.

[192] L. Xiong and J. Zhang, "Energy-efficient uniquely factorable constellation designs for noncoherent SIMO channels", *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5, pp. 2130–2144, 2012.

[193] D. Xia, J. Zhang, and S. Dumitrescu, "Energy-efficient full diversity collaborative unitary space-time block code designs via unique factorization of signals", *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1678–1703, 2013.

[194] Z. Dong, H. Chen, and J. Zhang, "Design of multi-user noncoherent massive SIMO systems for scalable URLLC", *Entropy*, vol. 25, no. 9, p. 1325, 2023, ISSN: 1099-4300.

[195] A. D. R. Choudary and C. P. Niculescu, *Real analysis on intervals*. New Delhi, India: Springer India, 2014, ISBN: 9788132221470.

[196] P. Billingsley, *Probability and measure*, 3rd ed. Hoboken, USA: Wiley, 1995, ISBN: 978-8126517718.