

Foundations of MIMO Communication

Understand the fundamentals of wireless and MIMO communication with this accessible and comprehensive text. Viewing the subject through an information theory lens, but also drawing on other perspectives, it provides a sound treatment of the key concepts underpinning contemporary wireless communication and MIMO, all the way to massive MIMO. Authoritative and insightful, it includes over 330 worked examples and 450 homework problems, with solutions and MATLAB® code and data available online. Altogether, this is an excellent resource for instructors and graduate students, and a great reference for researchers and practicing engineers.

Robert W. Heath Jr. is a Cullen Trust for Higher Education Endowed Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin. He is a Fellow of the IEEE.

Angel Lozano is a Professor in the Department of Information and Communication Technologies at Universitat Pompeu Fabra, Barcelona. He is a Fellow of the IEEE.

“*Foundations of MIMO Communication* by Heath and Lozano is a much needed tour-de-force on a critically important topic for wireless system design. The book provides an in-depth comprehensive treatment of the theoretical fundamentals in single-user, multiuser, and massive MIMO communications. It covers information-theoretic performance limits as well as estimation, channel modeling, matrix theory, and optimization techniques that are key to many aspects of MIMO system design. Hundreds of worked examples drive home the value of the theory in solving practical problems, and over one thousand important references are provided for further reading. This masterfully written book will serve as an invaluable textbook for students as well as a definitive reference for researchers and practicing engineers.”

Andrea Goldsmith, *Stanford University*

“This masterpiece from two preeminent scholars and innovators was worth the wait: this book took over a decade to write. Heath and Lozano are renowned information and communication theorists who have made many pioneering contributions to the theory of MIMO. However, they are unusual in also having done much to bring MIMO into practice: Heath at a pioneering Silicon Valley startup and Lozano at Bell Labs. They hold dozens of fundamental patents on MIMO and their broad perspective gives this book a unique depth and relevance. The book is well-suited for teaching or self study, and stands apart from others in the beauty and completeness of the exposition.”

Jeffrey Andrews, *The University of Texas at Austin*

“Remarkably, and almost three decades after its inception, MIMO technology is holding on to its status as one of the most impactful and versatile technologies, likely to underpin the design radio access networks for years to come still. With this highly complete textbook, Heath and Lozano are sharing their deep expertise and long-term experience in an area that has seen so many successes and revolutions. This material will be invaluable to teachers, students, and practitioners alike.”

David Gesbert, *Eurecom*

“MIMO and Massive MIMO are key technologies for enhancing the capacity and coverage of 3G/4G/5G and Beyond 5G wireless systems across all frequency bands. Heath and Lozano provide a comprehensive treatment of this essential technology from both a theoretical and practical point of view. An excellent and must read for graduate students and researchers in the field of wireless communication.”

Amitava Ghosh, *Nokia Bell Labs*

Foundations of MIMO Communication

ROBERT W. HEATH JR.
University of Texas at Austin

ANGEL LOZANO
Universitat Pompeu Fabra





CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
 a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
 education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9780521762281

DOI: 10.1017/9781139049276

© Cambridge University Press & Assessment 2019

This publication is in copyright. Subject to statutory exception and to the provisions
 of relevant collective licensing agreements, no reproduction of any part may take
 place without the written permission of Cambridge University Press & Assessment.

First published 2019

Reprinted 2019

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data

Names: Heath, Robert W., Jr, 1973– author. | Lozano, Angel, 1968– author.

Title: Foundations of MIMO communication / Robert W. Heath Jr.,
 University of Texas, Austin, Angel Lozano, Universitat Pompeu Fabra.

Description: Cambridge, United Kingdom ; New York, NY, USA :
 Cambridge University Press, 2019.

Identifiers: LCCN 2018035156 | ISBN 9780521762281 (hardback : alk. paper)

Subjects: LCSH: MIMO systems.

Classification: LCC TK5103.4836 .H43 2019 | DDC 621.384–dc23

LC record available at <https://lccn.loc.gov/2018035156>

ISBN 978-0-521-76228-1 Hardback

Additional resources for this publication at www.cambridge.org/heathlozano

Cambridge University Press & Assessment has no responsibility for the persistence
 or accuracy of URLs for external or third-party internet websites referred to in this
 publication and does not guarantee that any content on such websites is, or will
 remain, accurate or appropriate.

To Pia and Rohan, for putting up with their busy Papa,
and to Nuria.

R. W. H.

For Ester, Carla and Bernat, beams of light in my life.
And for my parents and siblings.

A. L.

Contents

| | |
|---|----------------|
| <i>Preface</i> | <i>page</i> xv |
| <i>Acknowledgments</i> | xvii |
| <i>A brief historical account</i> | xix |
| <i>About this book</i> | xxii |
| Part I Wireless Communication Theory | 1 |
| 1 A primer on information theory and MMSE estimation | 3 |
| 1.1 Introduction | 3 |
| 1.2 Signal distributions | 4 |
| 1.3 Information content | 6 |
| 1.3.1 Entropy | 7 |
| 1.3.2 Differential entropy | 9 |
| 1.3.3 Entropy rate | 10 |
| 1.4 Information dependence | 11 |
| 1.4.1 Relative entropy | 11 |
| 1.4.2 Mutual information | 12 |
| 1.5 Reliable communication | 17 |
| 1.5.1 Information-theoretic abstraction | 17 |
| 1.5.2 Capacity | 20 |
| 1.5.3 Coding and decoding | 22 |
| 1.5.4 Bit-interleaved coded modulation | 30 |
| 1.5.5 Finite-length codewords | 35 |
| 1.5.6 Hybrid-ARQ | 36 |
| 1.5.7 Extension to MIMO | 37 |
| 1.6 MMSE estimation | 39 |
| 1.6.1 The conditional-mean estimator | 40 |
| 1.6.2 MMSE estimation in Gaussian noise | 41 |
| 1.6.3 The I-MMSE relationship in Gaussian noise | 44 |
| 1.7 LMMSE estimation | 47 |
| 1.7.1 Random variables | 48 |
| 1.7.2 Random processes | 50 |
| 1.8 Summary | 51 |
| Problems | 51 |

| | | |
|----------|---|-----|
| 2 | A signal processing perspective | 57 |
| 2.1 | Introduction | 57 |
| 2.2 | Signal, channel, and noise representations | 58 |
| 2.2.1 | Passband signals and complex baseband equivalents | 58 |
| 2.2.2 | Complex baseband channel response | 61 |
| 2.2.3 | Time discretization | 64 |
| 2.2.4 | Pulse shaping | 66 |
| 2.2.5 | Additive noise | 71 |
| 2.2.6 | Energy and power | 73 |
| 2.2.7 | Channel normalization | 74 |
| 2.2.8 | Vector representation | 75 |
| 2.3 | Signal, channel, and noise representations: extension to MIMO | 77 |
| 2.3.1 | Vector and matrix representations | 77 |
| 2.3.2 | Channel normalization | 78 |
| 2.3.3 | Stacked vector representation | 79 |
| 2.3.4 | Precoding | 80 |
| 2.3.5 | Signal constraints | 83 |
| 2.4 | Linear channel equalization | 87 |
| 2.4.1 | Linear ZF equalization | 88 |
| 2.4.2 | LMMSE equalization | 94 |
| 2.5 | Single-carrier frequency-domain equalization | 98 |
| 2.5.1 | Basic formulation | 98 |
| 2.5.2 | Extension to MIMO | 102 |
| 2.6 | OFDM | 104 |
| 2.6.1 | Basic formulation | 104 |
| 2.6.2 | Extension to MIMO | 110 |
| 2.7 | Channel estimation | 110 |
| 2.7.1 | Single-carrier channel estimation | 111 |
| 2.7.2 | OFDM channel estimation | 121 |
| 2.8 | Summary and outlook | 122 |
| | Problems | 124 |
| 3 | Channel modeling | 131 |
| 3.1 | Introduction | 131 |
| 3.2 | Preliminaries | 132 |
| 3.2.1 | Basics of radio propagation | 132 |
| 3.2.2 | Modeling approaches | 133 |
| 3.3 | Large-scale phenomena | 135 |
| 3.3.1 | Pathloss and shadow fading | 135 |
| 3.3.2 | Free-space model | 138 |
| 3.3.3 | Macrocell models | 139 |
| 3.3.4 | Microcell models | 141 |
| 3.3.5 | Picocell and indoor models | 142 |
| 3.4 | Small-scale fading | 143 |

| | | |
|----------|--|------------|
| 3.4.1 | Multipath propagation | 143 |
| 3.4.2 | Space selectivity | 146 |
| 3.4.3 | Time selectivity | 151 |
| 3.4.4 | Frequency selectivity | 158 |
| 3.4.5 | Time–frequency double selectivity | 166 |
| 3.5 | Interlude: essential notions of antenna arrays | 170 |
| 3.5.1 | Array steering vectors | 170 |
| 3.5.2 | Array factor and beamforming | 173 |
| 3.6 | Modeling of MIMO channels | 175 |
| 3.6.1 | Analytical models | 177 |
| 3.6.2 | Parametric models | 188 |
| 3.7 | Channel estimation revisited | 192 |
| 3.7.1 | Large-scale phenomena | 193 |
| 3.7.2 | Small-scale fading | 195 |
| 3.8 | MIMO channel models in standards | 199 |
| 3.8.1 | 3GPP spatial channel model | 200 |
| 3.8.2 | SUI models for IEEE 802.16 | 201 |
| 3.8.3 | IEEE 802.11 channel model | 201 |
| 3.9 | Summary and outlook | 202 |
| | Problems | 205 |
| 4 | Single-user SISO | 209 |
| 4.1 | Introduction | 209 |
| 4.2 | Interplay of bit rate, power, and bandwidth | 209 |
| 4.2.1 | Low-SNR regime | 215 |
| 4.2.2 | High-SNR regime | 218 |
| 4.3 | AWGN channel | 220 |
| 4.3.1 | Capacity | 220 |
| 4.3.2 | Discrete constellations | 224 |
| 4.3.3 | Sneak preview of link adaptation | 225 |
| 4.4 | Frequency-selective channel | 228 |
| 4.4.1 | Partition into parallel subchannels | 231 |
| 4.4.2 | Waterfilling power allocation | 233 |
| 4.4.3 | Capacity | 235 |
| 4.4.4 | Discrete constellations | 238 |
| 4.4.5 | CESM, MIESM, and EESM mapping methods | 241 |
| 4.5 | Frequency-flat fading channel | 244 |
| 4.5.1 | CSIR and CSIT | 245 |
| 4.5.2 | No CSIT | 249 |
| 4.5.3 | No CSI | 261 |
| 4.6 | Frequency-selective fading channel | 267 |
| 4.7 | Which fading setting applies? | 268 |
| 4.8 | Pilot-assisted communication | 271 |
| 4.8.1 | Frequency-flat fading | 271 |

| | | |
|---------------------------------|--|------------|
| 4.8.2 | Pilot power boosting | 278 |
| 4.8.3 | Frequency-selective fading | 278 |
| 4.9 | Channels with interference | 279 |
| 4.10 | Summary and outlook | 284 |
| | Problems | 287 |
| Part II Single-user MIMO | | 295 |
| 5 | SU-MIMO with optimum receivers | 297 |
| 5.1 | Introduction | 297 |
| 5.2 | Initial considerations | 298 |
| 5.3 | CSIR and CSIT | 301 |
| 5.3.1 | Quasi-static setting | 301 |
| 5.3.2 | Ergodic setting | 307 |
| 5.4 | No CSIT | 311 |
| 5.4.1 | Quasi-static setting | 311 |
| 5.4.2 | Ergodic setting | 314 |
| 5.5 | No CSI | 338 |
| 5.6 | Pilot-assisted communication | 341 |
| 5.7 | Channels with interference | 345 |
| 5.8 | Optimum transmitter and receiver structures | 348 |
| 5.8.1 | Single codeword versus multiple codewords | 348 |
| 5.8.2 | LMMSE-SIC receiver | 349 |
| 5.8.3 | The layered architecture | 352 |
| 5.8.4 | BICM implementations | 353 |
| 5.9 | Link adaptation | 358 |
| 5.9.1 | Single codeword | 358 |
| 5.9.2 | Multiple codewords | 359 |
| 5.10 | Reciprocity and CSI feedback | 361 |
| 5.10.1 | Channel reciprocity | 362 |
| 5.10.2 | Analog feedback | 363 |
| 5.10.3 | Digital feedback | 364 |
| 5.11 | Summary and outlook | 374 |
| | Problems | 378 |
| 6 | SU-MIMO with linear receivers | 386 |
| 6.1 | Introduction | 386 |
| 6.2 | General characteristics of linear MIMO receivers | 387 |
| 6.3 | Linear ZF receiver | 388 |
| 6.3.1 | Receiver structure | 388 |
| 6.3.2 | Output SNR distribution | 390 |
| 6.3.3 | Ergodic spectral efficiency | 392 |
| 6.4 | LMMSE receiver | 396 |

| | | |
|--------------------------------|---|------------|
| 6.4.1 | Receiver structure | 396 |
| 6.4.2 | Output SINR distribution | 398 |
| 6.4.3 | Ergodic spectral efficiency | 403 |
| 6.5 | Relationship between the LMMSE and the optimum receiver | 407 |
| 6.6 | Summary and outlook | 408 |
| | Problems | 410 |
| Part III Multiuser MIMO | | 413 |
| 7 | Multiuser communication prelude | 415 |
| 7.1 | Introduction | 415 |
| 7.2 | Spectral efficiency region | 416 |
| 7.3 | Orthogonal channel sharing | 418 |
| 7.3.1 | Time-division | 418 |
| 7.3.2 | Frequency-division | 419 |
| 7.3.3 | OFDMA | 420 |
| 7.4 | Non-orthogonal channel sharing | 420 |
| 7.5 | Scalar metrics | 422 |
| 7.5.1 | Sum of the spectral efficiencies | 422 |
| 7.5.2 | Weighted sum of the spectral efficiencies | 422 |
| 7.5.3 | Equal spectral efficiencies | 424 |
| 7.5.4 | Minimum of the spectral efficiencies | 424 |
| 7.5.5 | Proportional fairness | 424 |
| 7.5.6 | Generalized proportional fairness | 425 |
| 7.6 | User selection and resource allocation | 426 |
| 7.6.1 | The proportional-fair algorithm | 428 |
| 7.7 | Low-SNR regime | 429 |
| 7.8 | Summary and outlook | 431 |
| | Problems | 433 |
| 8 | MU-MIMO with optimum transceivers | 436 |
| 8.1 | Introduction | 436 |
| 8.2 | The multiple-access channel | 437 |
| 8.3 | Multiple-access channel with CSIR and CSIT | 440 |
| 8.3.1 | Quasi-static setting | 440 |
| 8.3.2 | Optimum receiver structure | 444 |
| 8.3.3 | Precoder optimization | 447 |
| 8.3.4 | High-SNR regime | 449 |
| 8.3.5 | Ergodic setting | 454 |
| 8.4 | Multiple-access channel with no CSIT | 457 |
| 8.4.1 | Quasi-static setting | 457 |
| 8.4.2 | Ergodic setting | 458 |
| 8.5 | Multiple-access channel with no CSI | 460 |

| | | |
|----------|--|-----|
| 8.6 | Pilot-assisted multiple-access channel | 461 |
| 8.7 | Duality between the multiple access and broadcast channels | 462 |
| 8.7.1 | Description and significance | 462 |
| 8.7.2 | Dual versus actual multiple-access channels | 467 |
| 8.8 | The broadcast channel | 467 |
| 8.9 | Broadcast channel with CSIR and CSIT | 469 |
| 8.9.1 | Optimum transmitter structure | 469 |
| 8.9.2 | Quasi-static setting | 472 |
| 8.9.3 | Precoder and power allocation optimization | 478 |
| 8.9.4 | High-SNR regime | 480 |
| 8.9.5 | Ergodic setting | 484 |
| 8.10 | Broadcast channel with no CSIT | 487 |
| 8.11 | Summary and outlook | 490 |
| | Problems | 490 |
| 9 | MU-MIMO with linear transceivers | 497 |
| 9.1 | Introduction | 497 |
| 9.2 | Linear receivers for the multiple-access channel | 498 |
| 9.3 | Linear ZF receiver for the multiple-access channel | 500 |
| 9.3.1 | Receiver structure | 500 |
| 9.3.2 | Output SNR distribution | 501 |
| 9.3.3 | Ergodic spectral efficiency | 501 |
| 9.3.4 | High-SNR regime | 505 |
| 9.4 | LMMSE receiver for the multiple-access channel | 507 |
| 9.4.1 | Receiver structure | 507 |
| 9.4.2 | Output SINR distribution | 509 |
| 9.4.3 | Ergodic spectral efficiency | 509 |
| 9.4.4 | High-SNR regime | 513 |
| 9.5 | Duality with linear transceivers | 514 |
| 9.6 | Linear transmitters for the broadcast channel | 516 |
| 9.7 | Linear ZF transmitter for the MU-MISO broadcast channel | 517 |
| 9.7.1 | Transmitter structure | 517 |
| 9.7.2 | SNR distribution | 518 |
| 9.7.3 | Power allocation | 520 |
| 9.7.4 | Ergodic spectral efficiency | 520 |
| 9.7.5 | High-SNR regime | 524 |
| 9.7.6 | Pilot-assisted ZF transmission | 528 |
| 9.8 | Block-diagonalization for the broadcast channel | 554 |
| 9.8.1 | Transmitter structure | 555 |
| 9.8.2 | Power allocation | 556 |
| 9.8.3 | Ergodic spectral efficiency | 558 |
| 9.8.4 | High-SNR regime | 559 |
| 9.9 | Regularized ZF transmitter for the broadcast channel | 562 |
| 9.9.1 | Regularizing term | 563 |

| | | |
|-----------|---|-----|
| 9.9.2 | Power allocation and ergodic spectral efficiency | 566 |
| 9.9.3 | High-SNR regime | 571 |
| 9.10 | Summary and outlook | 571 |
| | Problems | 574 |
| 10 | Massive MIMO | 578 |
| 10.1 | Introduction | 578 |
| 10.2 | Going massive | 579 |
| 10.2.1 | The massive MIMO regime | 579 |
| 10.2.2 | Excess antennas | 581 |
| 10.3 | Reverse-link channel estimation | 582 |
| 10.3.1 | Pilot reuse | 582 |
| 10.3.2 | Pilot contamination | 583 |
| 10.4 | Reverse-link data transmission | 587 |
| 10.4.1 | Channel hardening | 590 |
| 10.4.2 | Matched-filter receiver | 592 |
| 10.4.3 | LMMSE receiver | 607 |
| 10.5 | Forward-link data transmission | 615 |
| 10.5.1 | Matched-filter transmitter | 616 |
| 10.5.2 | Regularized ZF transmitter | 620 |
| 10.6 | Mitigation of pilot contamination | 623 |
| 10.6.1 | Subspace methods | 624 |
| 10.6.2 | Coordinated pilot assignment | 624 |
| 10.6.3 | Reception and precoding with other-cell awareness | 625 |
| 10.6.4 | Large-scale multicell processing | 626 |
| 10.7 | Practical considerations | 628 |
| 10.8 | Summary and outlook | 631 |
| | Problems | 636 |
| 11 | Afterword | 643 |
| 11.1 | Beyond cellular | 643 |
| 11.2 | Beyond wireless | 644 |
| | Appendices | 647 |
| | Appendix A Transforms | 649 |
| A.1 | Fourier transforms | 649 |
| A.2 | Z-transform | 651 |
| | Appendix B Matrix algebra | 653 |
| B.1 | Column space, row space, null spaces | 653 |
| B.2 | Special matrices | 654 |
| B.3 | Matrix decompositions | 656 |

| | | |
|--|------------------------------|-----|
| B.4 | Trace and determinant | 658 |
| B.5 | Frobenius norm | 658 |
| B.6 | Moore–Penrose pseudoinverse | 659 |
| B.7 | Matrix inversion lemma | 659 |
| B.8 | Kronecker product | 659 |
| Appendix C Random variables and processes | | 661 |
| C.1 | Random variables | 661 |
| C.2 | Large random matrices | 671 |
| C.3 | Random processes | 672 |
| Appendix D Gradient operator | | 674 |
| Appendix E Special functions | | 676 |
| E.1 | Gamma function | 676 |
| E.2 | Digamma function | 677 |
| E.3 | Exponential integrals | 677 |
| E.4 | Bessel functions | 678 |
| E.5 | Q-function | 678 |
| E.6 | Hypergeometric functions | 679 |
| Appendix F Landau symbols | | 680 |
| Appendix G Convex optimization | | 681 |
| G.1 | Convex sets | 681 |
| G.2 | Convex and concave functions | 681 |
| G.3 | Convex optimization problems | 681 |
| G.4 | KKT optimality conditions | 682 |
| G.5 | Lagrange multipliers | 683 |
| G.6 | Jensen’s inequality | 684 |
| <i>References</i> | | 685 |
| <i>Index</i> | | 752 |

Preface

The art of wireless communications is arguably one of the biggest technological revolutions in history, and a crowning achievement of modern engineering. Its impact on the functioning of contemporary society cannot be overstated. It seems utterly impossible to conceive today’s world without the myriad devices that are wirelessly connected at all times.

In the decades since their inception, cellular systems have undergone five generational transitions. The first generation unfolded during the 1980s and offered only analog telephony. The second generation (2G), rolled out in the 1990s and whose dominant standard was the global system for mobile communications (GSM), saw the changeover to digital and the birth of text messaging. The third generation (3G), circa the 2000s, incorporated data and multimedia applications. Subsequently, the fourth generation (4G) brought about, besides faster bit rates and lower latencies, a complete adoption of packet switching as a platform for the mobile internet and the convergence of all standards worldwide into the long-term evolution (LTE) system. Honoring its name, LTE then evolved into the fifth generation (5G), augmented by another radio access interface termed new radio (NR) that allows operating on a much wider range of frequencies.

In the transitional period between 2G and 3G, as the soaring costs of spectrum collided with the pressures to increase bit rates so as to accommodate data and multimedia applications, the interest in radically improving the spectral and power efficiencies became acute. This propelled research initiatives that blossomed into major advances, chief among which stands multiple-input multiple-output (MIMO) communication, the subject of this book. In short, MIMO amounts to the transmission of concurrent signals from multiple antennas at one end of the link, with multiple antennas also at the receiving end.

Although MIMO is sometimes defined as the incorporation of the space domain to the communication process, this is not quite precise. The space domain has been at the crux of wireless systems since their onset, and in fact it is inherent to the concept of a cellular network: a region is tessellated into cells and the time/frequency signaling resources are reused repeatedly over such cells, i.e., they are spatially reused.

A more satisfying definition of MIMO might be the reusing of signaling resources, not over faraway cells, but rather across antennas—typically collocated—belonging to a common transmitter/receiver, such that new signaling dimensions are unlocked as antennas are added. The many advantages that this brings about are spelled out in detail throughout the text. By arranging these spatial dimensions, each seen as a scalar quantity, into a vector, it can be said that MIMO amounts to a vectorization of the transmission. This subsumes specific instances, such as *phased arrays* and *antenna diversity*, which predate the general formulation of MIMO.

Being in space rather than in time/frequency, the signaling dimensions harnessed with

MIMO behave differently, and this book is about how to communicate over them. These new signaling dimensions multiply the signaling opportunities in time/frequency, hence it can be said that MIMO provides additional “bandwidth” without an increase in the actual electromagnetic bandwidth [1].

Acknowledgments

There are a number of people without whose helping hand this project would not have come to fruition.

First of all, we are indebted to Ramya Bhavagatula (Google), Chanbyoung Chae (Yonsei University), Robert Daniels (PHAZR), Antonio Forenza (Artemis Networks), Kaibin Huang (University of Hong Kong), Insoo Hwang (Google), Takao Inoue (National Instruments), and Rahul Vaze (Tata Institute of Fundamental Research), all of whom contributed to preliminary chapter drafts of an earlier book that never made it to press. That work served as an inspiration for this book.

Second, we are grateful to research collaborators who have influenced the contents and their presentation. Chief among these stand Jeffrey Andrews (The University of Texas at Austin), Federico Boccardi (Ofcom), Helmut Bölcskei (ETH Zürich), Iain Collings (Macquarie University), Gerard Foschini (Bell Labs), David Gesbert (Eurecom), Martin Haenggi (University of Notre Dame), Andrea Goldsmith (Stanford University), Nuria González-Prelcic (Universidade de Vigo), Howard Huang (Bell Labs), Nihar Jindal (Google), Geert Leus (TU Delft), David Love (Purdue University), Thomas Marzetta (New York University), Matt McKay (HKUST), Constantinos Papadias (AIT), Steven Peters (Sterne Kessler Goldstein Fox), Farrokh Rashid-Farrokhi, Shilpa Talwar (Intel), Jose Tellado (HP Enterprise), Ted Rappaport (New York University), Dragan Samardzija (Bell Labs), Anthony Soong (Huawei), Antonia Tulino (Bell Labs), Reinaldo Valenzuela (Bell Labs), Sivarama Venkatesan (Bell Labs), Sergio Verdú (Princeton University), and Harish Viswanathan (Bell Labs).

We also want to thank all the students and colleagues who revised portions of the text. At The University of Texas at Austin, this includes Rebal AlJurdi, Ahmed Alkhateeb, Raquel Flores Buckley, Caleb Lo, Preeti Kumari, Khurram U. Mazher, Amine Mezghani, Nitin Jonathan Myers, Sungwoo Park, Andrew Thornburg, Yuyang Wang, Hongxiang Xie, Yi Zhang, Dalin Zhu, as well as the graduate students of EE 381V and EE 381S. At Universitat Pompeu Fabra, in turn, it includes Masoud Attarifar, Geordie George, David Morales-Jimenez, Rasoul Nikbakht, and Ratheesh Mungara, as well as Albert Guillen and Alfonso Martinez, who clarified aspects of BICM at a moment's notice. Further mention must be made of Dmitry Chizhik (Bell Labs), who kindly examined the channel modeling descriptions, of Giovanni Geraci (Nokia), who reviewed the massive MIMO material, of Luca Sanguinetti (Università di Pisa), who helped polish the description of pilot decontamination methods, and of Aldebaro Klautau (Universidade Federal do Pará), who scrutinized the signal processing coverage. Ezio Biglieri (Universitat Pompeu Fabra) provided invaluable assistance with coding aspects, and his endless wisdom on all the facets of technical writing and of the presentation of ideas in general. Markus Rupp (Technische Universität

Wien) provided references and shared his know-how on performance mapping methods. Very special thanks go to Matthew Valenti (West Virginia University), who supplied the turbo coding data from his coded modulation library, and to Jack Winters, who helped us keep the historical account honest.

Last, but not least, we appreciate the support received from the editorial staff at Cambridge University Press over the years. Phil Meyler provided the initial impetus, Elizabeth Horne carried the torch, and Julie Lancashire brought it home. Additional gratitude goes to Annie Toynbee for overseeing the production, to Gary Smith and Susan Parkinson for thoroughly proofreading the manuscript, and to the texline team for their assistance with all TeX-related matters. And, since first impressions matter, we are thankful to Carla Gironella for the picture that graces the cover.

This book would also not have been possible without support from a number of sponsors, including the National Science Foundation, the European Research Council, the Office of Naval Research, the Army Research Labs, AGAUR, DARPA, AT&T Labs, ICREA, Intel, Huawei, Freescale, MINECO, Motorola, Nokia, CommScope, Crown Castle, Mitsubishi Electronic Research Laboratories, and Samsung.

Robert W. Heath Jr.
The University of Texas at Austin
Austin, Texas

Angel Lozano
Universitat Pompeu Fabra
Barcelona

A brief historical account

We usually imagine that invention occurs in a flash, with a eureka moment that leads a lone inventor towards a startling epiphany. In truth, large leaps forward in technology rarely have a precise point of origin. At the start, forces that precede an invention merely begin to align, often imperceptibly, as a group of people and ideas converge, until over the course of months or years they gain clarity and momentum and the help of additional ideas and actors. Luck seems to matter, and so does timing, for it tends to be the case that the right answers, the right people, the right place—perhaps all three—require a serendipitous encounter with the right problem. And then, sometimes, a leap. Only in retrospect do such leaps look obvious.

Jon Gertner
The idea factory: Bell Labs and the great age of American innovation
Penguin Press, 2012.

MIMO builds on a long pedigree of the application of antenna arrays to communication, starting with Marconi’s transatlantic experiment in 1901 [2]. While a comprehensive formulation and a broad understanding did arguably not materialize until the 1990s and 2000s, many of the constituent pieces were present in seemingly unconnected developments well before that. Let us briefly recount these various strands and how they came together.

A first thread relates to phased arrays, a subject of interest since the 1950s [3]. By adjusting the signal’s phase at each antenna, either before transmission or upon reception, the overall radiation pattern of an array can be shaped into a beam and pointed in a desired direction, say toward a dominant propagation path. If, besides the phase, the amplitude can also be adjusted, then we have an *adaptive array*. In the 1980s, adaptive arrays evolved into smart antennas, whose array patterns can further null-out interference [4]. All of these are effectively spatial filters that rely on signal coherence across the arrays, an aspect discussed at length throughout the book and that has implications for the physical structure of those arrays and for the radio propagation. Precisely, this coherence is associated with specular propagation and with cleanly defined directions of arrival and departure.

In a second thread, arrays had long been a source of diversity. As detailed later, when multiple propagation paths exist, a signal conveyed over a wireless channel exhibits severe fluctuations, a deleterious phenomenon that can be mitigated by procuring independently fluctuating copies of that signal. An array can serve this purpose, as in such propagation conditions the antennas become uncorrelated—an opposite situation to that of adaptive arrays. Receive diversity was studied as early as 1954 [5] and implemented in early cellular generations already. Being more involved, transmit diversity is more recent [6–10].

Interestingly, diversity was shown to be compatible with the nulling-out of interference, provided the channel response was known [11].

A third thread is constituted by a variety of works that considered channels coupling multiple inputs into multiple outputs. These ranged from abstract formulations [12–14] to analyses motivated by the specific problem of communicating over coupled telephone lines [15–18]. These important precursors were the seed of MIMO, and also of multiuser detection for code-division multiple access (CDMA) [19, 20], two developments with substantial common ground. In drastic contrast with adaptive arrays and diversity, which in essence involve a single signal, these developments entail the concurrent transmission of multiple signals.

Particularly prescient was the contribution of Noach Amitay and Jack Salz, who, as early as 1984, considered a link with two orthogonally polarized antennas at each end [21]. This was extended by Jack Winters in a piece that featured multiantenna transmitters and receivers with many of the ingredients of contemporary MIMO communication: concurrent signals were transmitted from antennas collocated on a device—in time this would be known as *spatial multiplexing*—and a receiver equipped with multiple antennas recovered each of those signals by nulling-out the interference from the rest [22]. In hindsight, both of these milestone papers deserved more credit than they were given, but at the time there was no demand for enhanced wireless performance; in fact, funding for such research was being curtailed.

It was not until the 1990s that the atmosphere was primed for these ideas to coalesce. Arguably, the main catalyst was the work of Gerhard Foschini and Michael Gans, who set out to design the perfect antenna from an information-theoretic standpoint. Starting with an array and no preset conditions on how to use it, they found that, if the antennas were uncorrelated, the optimum strategy was to have each one radiate an independent signal, i.e., spatial multiplexing. This strategy was radically novel in that it sought to exploit, rather than counter or avoid, the fluctuations caused by multipath propagation. Foschini went further and proposed an architecture to effect spatial multiplexing, the so-called *layered architecture*, which was remarkable in that it could be built with off-the-shelf encoders and decoders and did not require the transmitter to know the channel's response [23]. Under the leadership of Reinaldo Valenzuela, a prototype with 12 transmit and 16 receive antennas confirmed the practicality of this proposition [24] and dispelled concerns about the feasibility of nulling-out interference from collocated antennas, i.e., concerns that in practice these signals would drown each other and be unrecoverable. Additional results by Emre Telatar consolidated the initial theoretical underpinnings of MIMO [25]. It behoves us to mention that Amitay, Salz, Winters, Foschini, Gans, Valenzuela, and Telatar were all associated with Bell Laboratories, placing this organization in a distinguished place in the history of MIMO. Angel Lozano's views on MIMO were strongly influenced by his interaction with these pioneers.

Another noteworthy advance in the 1990s was from Paulraj and Kailath, at Stanford University, who applied spatial multiplexing to TV distribution from multiple towers [26]. That idea became a core part of Iospan Wireless, an early developer of MIMO technology for fixed wireless access. Robert Heath's perspectives on MIMO were shaped by his role as an early employee of Iospan Wireless.

Also at Stanford, in yet another development in the 1990s, Raleigh and Cioffi provided a generalization to multiantenna channels of a technique (expounded in Section 4.4) that does require the transmitter to know the channel response [27, 28].

Altogether, after a slow ripening of ideas over an extensive period, a critical mass was reached. Perhaps because of the impetus provided by the soaring costs of spectrum or perhaps because, simply, the time was right, a chain reaction was sparked and spread rapidly through academia and industry. From an academic thought experiment, MIMO grew into a foundation of wireless communication standards. In the span of a few years, it was adopted by the IEEE 802.16 fixed wireless access system and by cellular standardization bodies. In particular, the 3G Partnership Project (3GPP) embraced it in a limited fashion for 3G and then as an integral part of the designs beginning with 4G LTE. In subsequent revisions of LTE, the numbers of supported antennas progressively increased and, for 5G, so-called *massive MIMO* deployments are intrinsic. However, the first commercial application of MIMO was not in the cellular arena.

In 1997, the IEEE 802.11 working group completed its original wireless local-area network (WLAN) standard. Successive revisions incorporated new features, with the earliest version to include MIMO being IEEE 802.11n, certified in 2007 and supporting four antennas (although two or three were more common at that time). IEEE 802.11n was succeeded by IEEE 802.11ac, supporting up to eight antennas, and further by IEEE 802.11ad, 802.11ax, and 802.11ay.

Cohorts of researchers have played a part in the advancement of MIMO, and it is nontrivial to apportion credit among individuals and institutions. A reasonable criterion to assign credit for any invention is the influence that each contribution ends up having [29], and this is the criterion applied in our exposition.

About this book

A wealth of texts on MIMO are available, including books with varying flavors and perspectives [30–40], edited books [41–43], monographs [44–46], and tutorial papers [47–49]. In addition, treatments of MIMO are included in general wireless communication books [50, 51]. The present volume, which builds on this diverse literature, is the result of 20 years of research and teaching. It is intended as a full-dress textbook for instructors and students at the graduate level, and a reference tool for researchers and practicing engineers. With this audience in mind, its aim is to be both accessible and comprehensive, and the conjunction of these objectives is the reason for the considerable length. We hope that readers will appreciate the organization of the contents, as well as the complementary features:

- 160 illustrations.
- 19 topical discussions.
- 339 examples. Some of these are titled examples, which encapsulate results of particular interest, making it easier for them to be identified and located. The rest of the examples are in the format of a problem accompanied by the corresponding solution.
- 463 homework problems proposed at the ends of the chapters.
- A companion website with the solutions to all the problems, and additional material.

There are various theoretical lenses under which MIMO, and in fact digital communication at large, can be seen. Here, the problem of MIMO communication is viewed through the lens of information theory, appropriately complemented with signal processing, estimation theory, channel modeling, optimization, linear algebra, and random matrix theory, and with a touch of stochastic geometry. The choice of information theory is not capricious, but rather the appreciation that, besides being digital, modern communications—certainly the forms that MIMO is relevant to—are built from the ground up with coding, operating very close to the fundamental limits that information theory delineates. Moreover, information theory yields surprisingly many design insights and the opportunity for extensive and informative analysis, greatly facilitating the exposition. As eloquently argued by James Massey [52], “information theory is the proper scientific basis for communication.”

In opening his treatise on quantum theory, the Nobel laureate Steven Weinberg declared that “there are parts of this book that will bring tears to the eyes of the mathematically inclined reader.” While not expecting to bring tears to the eyes of our readers, we also do not intend to sacrifice clarity or understanding at the altar of rigor. Rather, we abide by the principle that rigor should be at the service of the problem at hand, rather than the resolution of the problem being slaved to the formalisms of rigor.

Organization

The book is organized into three parts, plus a set of appendices.

- Part I, labeled *wireless communication theory*, exposes the pillars on which the edifice rests, with Chapters 1–3 successively introducing the perspectives of information theory and estimation, signal processing, and channel modeling. Along the way, the radio channel is in turn interpreted as a random transformation, as a linear impulse response, and as a stochastic process; three interpretations, each one fitting the perspective of its chapter. These perspectives converge in Chapter 4, which deals with non-MIMO communication, crisply presenting all the concepts that are then to be generalized.
- Part II is devoted to single-user MIMO communication, meaning a single transmitter and a single receiver. In this clean setting, we elaborate on how to transmit and receive with multiple antennas, first without receiver restrictions in Chapter 5, and then with a linearity restriction in Chapter 6.
- Although, with orthogonal multiplexing, single-user conditions can be created within a network, the signaling dimensions created by MIMO are best exploited when the multiuser aspects are brought into the picture. Part III deals with these aspects, first by introducing them broadly in Chapter 7, and then by delving into how to transmit to and receive from a plurality of multiantenna users. Again, this is first covered without restrictions, in Chapter 8, and then with a linearity restriction in Chapter 9. Finally, Chapter 10 broadens the scope in two respects: an entire cellular network is considered, and MIMO becomes massive MIMO.
- The appendices provide a compact tour of various mathematical results that are invoked throughout the book, conveniently couched in our notation, with the objective of rendering the text as self-contained as possible.

Based on the foregoing structure, a variety of itineraries can be defined in support of graduate courses. Some potential ones are as follows.

- A course on the information-theoretic principles of wireless communication, leading up to MIMO, can rest on Chapter 1, Chapter 2 (Sections 2.1, 2.2, and 2.6), and Chapter 4, with the necessary channel modeling taken from Sections 3.1–3.4.
- A basic course on MIMO communication can rely on Chapters 1–6 (with the possible exclusion of Sections 2.4, 2.5, and 2.7).
- An advanced course on multiuser MIMO, which could be concatenated with the one above, can rest on Chapters 7–9, with extension to massive MIMO via Chapter 10.
- A course specifically geared toward massive MIMO can be designed with Chapters 6, 9, and 10, plus whatever material is needed from Chapters 1–4 and 7.

Requisites

The background assumed on the part of the readers corresponds to senior-level or first-year graduate-level courses on signals and systems, digital communication, linear algebra, and probability. A firm grasp of random variables is particularly desirable.

No expertise on cellular networks is presumed, beyond the basic notion that such networks are organized in cells, each featuring a base station that wirelessly communicates with the population of users within. The *forward link* or *downlink* embodies the communication from the base to its users, while the *reverse link* or *uplink* embodies the return communication from the users to the base. This naturally leads to the notion of *duplexing*, which refers to how these two directions of communication are arranged. If they are simultaneous over disjoint frequency bands, we speak of frequency-division duplexing (FDD), while, if they take place on alternating time intervals over a common frequency band, we speak of time-division duplexing (TDD). Full-duplexing, deemed unfeasible in the past, is now becoming possible thanks to advanced self-interference cancelation techniques [53–55], and thus it is also considered.

Among the numerous acronyms that—MIMO aside—sprinkle the text, two stand out and deserve introductory remarks.

- CSI, which stands for channel-state information, alludes to complete knowledge of the channel’s response. Since a signature attribute of wireless channels is their variability, acquiring and employing CSI is instrumental when communicating over such channels. In fact, the availability of CSI is one of the main axes of the exposition in the book.
- OFDM, which stands for orthogonal frequency-division multiplexing, is the signaling technique that has come to dominate communication, wireline and wireless [56, 57]. LTE, NR, and the mentioned WLAN standards all feature OFDM. Although the mainstream alternatives are entertained too, we acknowledge the dominance of OFDM and rely on it extensively for our formulations.

Notation

In a long text such as this one, notation can be a minefield. While striving for consistency and intuition, a modicum of flexibility becomes necessary at points. As part of the effort to convey meaning consistently, some points are worth noting.

- Whenever possible, variables are named to directly reminisce of the quantities they represent, e.g., SNR for the signal-to-noise ratio, MMSE for the minimum mean-square error, or N_t and N_r for the numbers of transmit and receive antennas.
- Bold symbols denote vectors and matrices, while nonbold symbols correspond to scalars.
- Capitalization distinguishes matrices from vectors, and large-scale from small-scale quantities. (The meaning of these scales is to become clear throughout the text.) Variables that are in general matrices retain their capitalization even in the special cases in which they may adopt a vector form.
- Frequency-domain quantities are represented with sans serif fonts, in contrast with time-domain quantities, which are denoted with serif fonts.
- Dummy variables and counters match, whenever possible, their respective quantities, e.g., we use \mathbf{A} to denote a realization of the random matrix \mathbf{A} , and we use n to run a counter over N positions.

The common notational schemes and all the relevant symbols, excluding only variables appearing in intermediate derivations, are listed in the pages that follow. The notational

schemes are, as much as possible, inspired by MATLAB®. For instance, $[\mathbf{A}]_{:,j}$ denotes the j th column of matrix \mathbf{A} whereas $[\mathbf{A}]_{:,j:j'}$ denotes the submatrix containing columns j through j' .

A number of the variables are further indexed by a user indicator u whenever they are applied in multiuser contexts. For instance, in single-user settings, \mathbf{H} is the time-domain normalized channel matrix; then, in multiuser settings, \mathbf{H}_u indicates the normalized channel matrix for user u . If, besides multiple users, multiple cells are present, then the indexing is further augmented to identify the cells involved: $\mathbf{H}_{l;\ell,u}$ denotes the normalized channel matrix linking the base station at cell l with the u th user at cell ℓ th.

Further scripting, not explicitly distinguished in the listings that follow, is applied to discriminate variations in the quantities. For instance, \mathbf{W} is a generic linear receiver while \mathbf{W}^{MF} , \mathbf{W}^{ZF} , \mathbf{W}^{MMSE} are specific types thereof, namely matched-filter, zero-forcing, and minimum mean-square error linear receivers.

Common notation

| | |
|---|---|
| \propto | proportionality |
| \approx | approximation |
| \simeq | asymptotic equality |
| \sim | distribution |
| \subseteq | subset |
| $((\cdot))_K$ | modulo- K |
| \otimes | Kronecker product |
| \odot | Hadamard (entry-wise) product |
| $*$ | convolution |
| a, A | nonbold letters denote scalars |
| \mathbf{a} | bold lowercase letters denote column vectors |
| \mathbf{A} | bold uppercase letters denote matrices |
| (\cdot) | indexing for continuous signals |
| $a(t), \mathbf{a}(t), \mathbf{A}(t)$ | time-domain continuous signals at time t |
| $\mathbf{a}(f), \mathbf{a}(f), \mathbf{A}(f)$ | frequency-domain continuous signals at frequency f |
| $[\cdot]$ | indexing for discrete signals |
| $a[n], \mathbf{a}[n], \mathbf{A}[n]$ | time-domain discrete signals at time n |
| $\mathbf{a}[k], \mathbf{a}[k], \mathbf{A}[k]$ | frequency-domain discrete signals at frequency k |
| $\{\cdot\}$ | sequence |
| $[a]^+$ | $= \max(0, a)$ |
| $a _{\text{dB}}$ | $= 10 \log_{10} a$ |
| $a _{3 \text{ dB}}$ | $= \log_2 a$ |
| $ a $ | magnitude of a |
| $\ \mathbf{a}\ $ | Euclidean norm of \mathbf{a} |
| $\ \mathbf{A}\ _{\text{F}}$ | Frobenius norm of \mathbf{A} |
| $\nabla_{\mathbf{x}}$ | gradient with respect to \mathbf{x} |
| \mathbf{A}^{T} | matrix transpose |
| \mathbf{A}^* | matrix conjugate transpose |
| \mathbf{A}^{c} | matrix conjugate |
| \mathbf{A}^{-1} | matrix inverse |
| \mathbf{A}^{\dagger} | Moore–Penrose matrix pseudoinverse |
| \mathbf{A}^{\star} | value of \mathbf{A} that solves an optimization problem |
| $\hat{\mathbf{A}}$ | estimate of \mathbf{A} |
| $\tilde{\mathbf{A}}$ | error in the estimation of \mathbf{A} |
| $[\mathbf{a}]_j$ | j th entry of \mathbf{a} |
| $[\mathbf{A}]_{i,j}$ | (i, j) th entry of \mathbf{A} |
| $[\mathbf{A}]_{:,j}$ | j th column of \mathbf{A} |
| $[\mathbf{A}]_{:,j:j'}$ | submatrix containing columns j through j' of \mathbf{A} |