UNIVERSITY OF CALIFORNIA, SAN DIEGO

Design and Analysis of MIMO Systems With Practical Channel State Information Assumptions

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in
Electrical and Computer Engineering
(Communication Theory and Systems)

by

Jun Zheng

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor Robert Bitmead
Professor William S. Hodgkiss
Professor Kenneth Kreutz-Delgado
Professor Paul H. Siegel

2006

The dissertation of Jun Zheng is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2006

To Ning,

and my parents.

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Table of Contents . . . v

List of Figures . . . ix

Acknowledgements . . . xiii

Vita and Publications . . . xv

Abstract . . . xvii

1 Introduction . . . 1
- 1.1 Multiple Input Multiple Output Systems . . . 1
- 1.2 System Performance versus CSI Assumptions . . . 3
  - 1.2.1 No CSIT and Perfect CSIR . . . 3
  - 1.2.2 Perfect CSIT and Perfect CSIR . . . 4
  - 1.2.3 No CSIT and No CSIR . . . 5
  - 1.2.4 Role of Channel State Information . . . 6
- 1.3 Practical CSI Considerations . . . 7
  - 1.3.1 Systems with Partial CSIT . . . 7
  - 1.3.2 Systems with unknown CSIR . . . 9

2 Generalized Vector Quantization Framework and Asymptotic Analysis . 12
- 2.1 Motivation . . . 12
- 2.2 Generalized Vector Quantization Framework . . . 17
  - 2.2.1 Motivation for Generalization . . . 17
  - 2.2.2 Problem Formulation . . . 19
  - 2.2.3 Optimal Partitioning of the Source Space . . . 21
  - 2.2.4 Normalized Inertial Profile . . . 23
  - 2.2.5 Heuristic Derivation of the Asymptotic Distortion Integral 26
- 2.3 Minimization of the Distortion Integral & Different Distortion Bounds . . . 29
  - 2.3.1 Asymptotic Distortion Lower Bound . . . 30
  - 2.3.2 An Alternative Distortion Lower Bound . . . 31
  - 2.3.3 Asymptotic Distortion Upper Bound . . . 32
  - 2.3.4 Losses Due in the Context of Side Information . . . 34
  - 2.3.5 Achievability of the Asymptotic Distortion Bounds . . . 35
- 2.4 Distortion Analysis of Mismatched Quantizers . . . 38
  - 2.4.1 Dimensionality Mismatch . . . 39

2.4.2 Distortion Function Mismatch . . . 39
2.4.3 Source Distribution Mismatch . . . 41
2.5 Distortion Analysis of Quantizers with Transformed Codebook . . 42
2.5.1 Problem Formulation . . . 43
2.5.2 Sub-optimal Point Density & Sub-optimal Voronoi Shape . 43
2.5.3 Characterizing the Inertial Profile of the Transformed Codebook . . . 44
2.5.4 Distortion Integral of the Transformed Codebook . . . 48
2.6 Asymptotic Analysis of Constrained Source . . . 49
2.7 Asymptotic Analysis of Complex Source . . . 52
2.7.1 Quantization of Unconstrained Source . . . 52
2.7.2 Quantization of Constrained Source . . . 53
2.8 Summary . . . 55

3 Capacity Analysis of MISO Systems with Finite-Rate CSI Feedback . . 57
3.1 Motivation . . . 57
3.2 System Model . . . 57
3.3 Problem Formulation . . . 59
3.4 Statistical Properties of the Channel Information . . . 61
3.5 Distortion Analysis for i.i.d. MISO Fading Channels . . . 65
3.6 Distortion Analysis for Correlated MISO Fading Channels . . . 69
3.6.1 Distortion lower bound $D_{\text{c-Low},1}$ for correlated channels . . 69
3.6.2 Distortion lower bound $D_{\text{c-Low},2}$ for correlated channels . . 70
3.6.3 Interesting Observations of the Distortion Bounds . . . 71
3.6.4 Numerical and Simulation Examples . . . 74
3.7 Distortion Analysis in High-SNR and Low-SNR Regimes . . . 77
3.7.1 High-SNR Distortion Analysis . . . 77
3.7.2 Low-SNR Distortion Analysis . . . 78
3.8 Distortion Comparisons between i.i.d. and Correlated channels . . 79
3.9 Summary . . . 81

4 Analysis of MISO CSI Quantizers with Mismatched Codebooks and Transformed Codebooks . . . 82
4.1 Motivation . . . 82
4.2 Mismatched Analysis of Quantized MISO Beamforming Systems . 83
4.2.1 Dimensionality Mismatch and Quantization Criterion Mismatch . . . 83
4.2.2 Source Distribution Mismatch (or Point Density Mismatch) 87
4.2.3 Comparisons with Other Channel Quantizers . . . 88
4.3 Analysis of MISO Channel Quantizers with Transformed Codebook 92
4.3.1 Problem Formulation . . . 93
4.3.2 Distortion Analysis of Transformed Codebooks . . . 93
4.3.3 Comparison with Optimal MISO CSI-Quantizers . . . 96
4.4 Summary . . . 100

5 Capacity Analysis of MIMO Systems with Finite-Rate CSI Feedback . . 103
5.1 Motivation . . . 103
5.2 System Model of MIMO Systems with Finite Rate CSI Feedback . 104
5.2.1 Fading Channel Model . . . 104
5.2.2 Finite-Rate Channel State Information Feedback . . . 104
5.2.3 Practical Transmit Pre-Coders with Quantized CSIT . . . 106
5.3 Analysis of MIMO Pre-coders with Quantized CSI . . . 107
5.3.1 Formulating the MIMO CSI-Quantizer as A General Vector Quantization Problem . . . 107
5.3.2 High-Resolution Distortion Analysis . . . 112
5.3.3 Interesting Observations of the Distortion Lower Bounds . 116
5.3.4 Analysis of CSI-Quantizers using Mismatched High-SNR and Low-SNR Codebooks . . . 118
5.3.5 Analysis of MIMO Pre-Coding Schemes with Multi-Mode spatial Multiplexing Strategy . . . 121
5.4 Numerical and Simulation Results . . . 123
5.4.1 High-Rate Capacity Analysis . . . 123
5.4.2 Analysis of Mismatched High-SNR and Low-SNR Codebooks 124
5.4.3 Performance of Multi-Mode spatial Multiplexing Schemes . 126
5.5 Summary . . . 128

6 Capacity Analysis of MIMO Systems with Unknown Channel State Information . . . 129
6.1 Motivation . . . 129
6.2 System Model . . . 132
6.3 Improved Capacity Lower Bound of Unknown MIMO Channels . . 135
6.3.1 Upper Bound of System Mutual Information Rate . . . 135
6.3.2 Tight Mutual Information Rate Upper Bound Leads to Improved Capacity Lower Bound . . . 139
6.4 Analysis of the Mutual Information Upper Bound . . . 148
6.4.1 Optimization of Pilot Structures . . . 148
6.4.2 Optimization of Pilot and Data Slot Allocations (under Equal Power Assumptions) . . . 154
6.4.3 Optimization of the Number of Active Transmit Antennas 158
6.4.4 Optimization of Power Allocations (ρ_τ, ρ_d) between Training and Data Symbols . . . 161
6.4.5 Low SNR Regimes . . . 166
6.5 Numerical and Simulation Results . . . 167
6.5.1 Orthogonal Pilot Structure . . . 168
6.5.2 Pilot and Data Slot Allocations (under Equal Power Assumptions) . . . 169
6.5.3 Power Allocations between Training and Data Symbols . . 172
6.5.4 Comparison with Other Capacity Analysis Results . . . 173

6.6 Summary . . . 173

7 Design of LDPC-Coded MIMO Systems With Unknown Block Fading Channels . . . 175
7.1 Motivation . . . 175
7.2 System Model . . . 178
7.2.1 MIMO transmitter structure . . . 178
7.2.2 MIMO receiver structure . . . 180
7.3 Soft-Input Soft-Output MIMO Detector . . . 181
7.3.1 Optimal soft MIMO detector . . . 183
7.3.2 Sub-optimal soft MIMO detector . . . 184
7.3.3 Sub-optimal butterfly soft MIMO detector . . . 187
7.3.4 Modified EM-based MIMO detector . . . 189
7.4 Design of LDPC-coded MIMO Systems . . . 195
7.4.1 Receiver structure of the LDPC-coded MIMO systems . . . 196
7.4.2 Analysis of extrinsic information transfer characteristics . . . 197
7.4.3 LDPC code optimization . . . 201
7.5 Numerical and Simulation Results . . . 203
7.5.1 Elimination of positive feedback in EM-based MIMO detectors . . . 203
7.5.2 EXIT function comparison between different soft MIMO detectors . . . 205
7.5.3 LDPC code degree profile optimization . . . 207
7.5.4 Overall coded MIMO system performance . . . 209
7.6 Summary . . . 211

8 Conclusions and Future Work . . . 213
8.1 Analysis of MIMO Systems with Finite-Rate Feedback . . . 213
8.2 Design and Analysis of MIMO Systems with Unknown CSI . . . 216
8.3 Future Work . . . 218
8.3.1 CSI Quantization with Practical Assumptions . . . 218
8.3.2 Practical Quantizer Design for CSI Feedback . . . 219
8.3.3 Precoder Design for Multiuser MIMO with Partial CSIT . . . 220

Bibliography . . . 221

LIST OF FIGURES

1.1 MIMO system model 2

3.1 Capacity loss of a 3×1 MISO transmit beamforming system with finite rate feedback 68
3.2 Capacity loss versus CSI feedback rate B of a 3×1 correlated MISO transmit beamforming system with normalized antenna spacing $D/\lambda = 0.5$, and signal to noise ratio $\rho = -10, 0$ and 20dB. . . . 74
3.3 Capacity loss versus CSI feedback rate B of a 3×1 correlated MISO transmit beamforming system with normalized antenna spacing $D/\lambda = 0.2, 0.3, 0.5, 2.0$, and signal to noise ratio $\rho = 20$dB. . . . 75
3.4 Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) versus transmit antenna spacing D/λ of a 3×1 MISO transmit beamforming system with signal to noise ratio $\rho =$ 5dB under CSI feedback rate $B = 10$ bits. 76
3.5 Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) versus transmit antenna spacing D/λ of a 3×1 MISO transmit beamforming system in high-SNR regime with $\rho =$ 20dB. 80

4.1 Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs MMSE quantizer). 86
4.2 Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs Mismatched codebook for i.i.d. fading channels). 89
4.3 Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) comparison of a 3×1 MISO transmit beamforming with optimal and mismatched codebooks versus antenna spacing $d = D/\lambda$, in high and low SNR regimes ($\rho = -10$ and 20 dB). . . . 92
4.4 Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $d = D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs Transformed codebook). 95
4.5 Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) comparison of a 3×1 MISO transmit beamforming with optimal and transformed codebooks versus antenna spacing $d = D/\lambda$, in low SNR regimes ($\rho = -10$ dB). 100

4.6 Demonstration of the tightness of the distortion bounds $D_{\text{c-tr-Low}}$ and $D_{\text{c-tr-Upp}}$ for a MISO system using transformed codebook over correlated fading channels with different number of transmit antennas of antenna spacing $d = D/\lambda = 0.5$. 101

5.1 Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels with finite-rate CSI feedback ($B = 8$), and using multi-mode spatial multiplexing transmission schemes. 122
5.2 Capacity loss versus CSI feedback rate B of a 4×2 MIMO system ($t = 4$, $r = 2$ and $n = 2$) over i.i.d. Rayleigh fading channels and with signal to noise ratio $\rho = -10, 0$ and 20dB. 124
5.3 Performance losses ($L_{\text{H-snr}}$ and $L_{\text{L-snr}}$) versus signal to noise ratio ρ of a 4×3 MIMO system ($t = 4$, $r = 3$, and $n = 2$ over i.i.d. Rayleigh fading channels with feedback rate $B = 8$ bits per channel update. . 125
5.4 Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels with feedback rate $B = 8$ bits per channel update, and using multi-mode spatial multiplexing (MMSM) transmission schemes. 126
5.5 Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels using MMSM transmission scheme, and with several different CSI feedback rate ($B = 1, 3, 5, 8$ bits per channel update). 127

6.1 MIMO system model composed of M transmit antennas and N receive antennas 132
6.2 Transmitted symbol structure of the MIMO system 134
6.3 Mutual information rate comparison (between actual system mutual information rate, MMSE-based capacity (or mutual information) lower bound, and the proposed mutual information rate upper bound) of a 2×2 MIMO system with channel coherent time $T = 6$ and signal to noise ratio $\rho_\tau = \rho_d = 4dB$. 138
6.4 Histogram of $\Re[v_{i,j}]$ with different number of transmit antennas $M = 1, 2, 3$. 145
6.5 Differential entropy $h(\mathbf{V})$ of Gaussian product matrix $\mathbf{V}$ versus the number of transmit antennas M. 146
6.6 Capacity and mutual information comparison between non-coherent MIMO channel capacity C, the actual system mutual information rate R, and the proposed mutual information upper bound $\overline{R}$ of an $M \times 2$ MIMO system with channel coherent time $T = 1$ and signal to noise ratio $\rho_\tau = \rho_d = 4dB$. 147
6.7 Mutual information rate upper bound of a 6×6 MIMO system under equal power allocation scheme of SNR $\rho = 4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$ 158

6.8 Mutual information upper bound of a $M \times 6$ MIMO system with SNR $\rho = 4dB$ and $T_\tau = M$, under different coherent time intervals $T = 8, 10, 12, 14, 16, 18, 20$ 162
6.9 Mutual information upper bound of a 6×6 MIMO system under optimal power allocation scheme of SNR $\rho = 4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$ 163
6.10 Optimal power allocation of a 6×6 MIMO system with SNR $\rho = 4dB$, under different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15,$ and 20 164
6.11 Mutual information rate gain of optimal power allocations over equal power allocations of a 6×6 MIMO system with SNR $\rho = 4dB$, under different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$ 165
6.12 Mutual information upper bound comparison between orthogonal pilot structures and random pilot structures under equal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$, and data interval $T_d = 1, 2, 3, 4$. 168
6.13 Mutual information rate upper bound of 6×6 MIMO system with coherent time interval $T = 10$, with different data slots allocation $T_d = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ 169
6.14 Mutual information rate upper bound of a 6×6 MIMO system under equal power allocation scheme of SNR $\rho = -4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$ 170
6.15 Mutual information upper bound comparison between equal power allocation and optimal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$ and different data slot allocation $T_d = 1, 2, 3, 4, 5, 6, 10$ 171
6.16 Comparison between the MMSE-based capacity lower bounds and the improved capacity lower bounds (mutual information upper bounds) under equal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$, and data interval $T_d = 1, 2, 3, 4$ 172

7.1 Transmitter model of LDPC-coded MIMO systems 178
7.2 Transmitted symbol structure of the coded MIMO system 179
7.3 Conventional receiver structure of LDPC-coded MIMO systems . . 180
7.4 Sub-optimal soft MIMO detector structure 184
7.5 Sub-optimal soft MIMO detector using butterfly structure 187
7.6 Conventional receiver structure of LDPC-coded MIMO systems . . 196
7.7 New receiver structure of LDPC-coded MIMO systems 196
7.8 Decoding trajectory of a regular $(3, 6)$ LDPC-coded MIMO system with optimal soft MIMO detector over a 2×2 unknown MIMO channel with coherence time $T = 6$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$. 199

7.9 Extrinsic information transfer characteristic of the EM-based MIMO detectors over a 2×2 unknown MIMO channel with training length $T_\tau = 2$ and signal to noise ratio $\rho = 4dB$. 204
7.10 Comparison of the extrinsic information transfer characteristic of different MIMO detectors in a 2×2 unknown MIMO system with coherence time $T = 6$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$. 205
7.11 Comparison of the extrinsic information transfer characteristic of different MIMO detectors in a 2×2 unknown MIMO system with coherence time $T = 18$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$. 206
7.12 EXIT-chart curve of a 2×2 regular $(3, 6)$ LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$, and training number $T_\tau = 2$ using optimal soft MIMO detectors, under different signal to noise ratios $\rho = -2, 0, 2.2, 4, 6, 8\ dB$. 207
7.13 EXIT-chart curve of a 2×2 optimized LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$, and training number $T_\tau = 2$ using optimal soft MIMO detectors, under different signal to noise ratios $\rho = -2, 0, 1.3, 4, 6, 8\ dB$. 208
7.14 Probability of bit error of a 2×2 regular $(3, 6)$ LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$ and training number $T_\tau = 2$ using several different soft MIMO detectors. 210
7.15 Probability of bit error of a 2×2 optimized LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$ and training number $T_\tau = 2$ using several different soft MIMO detectors. 211

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. Bhaskar D. Rao, for his inspirational guidance, constant support and encouragement throughout my Ph.D. work. He guided me at every important step of my Ph.D. path in UCSD, giving me not only academic advice and dedicating great amount of time and efforts on my thesis topic, but also generously understanding and encouraging me through all the time. He instilled in me not only the knowledge required for a doctoral degree, but also the understanding of research philosophy and ability critical to an independent researcher. I am also thankful to my committee members, Dr. Paul H. Siegel, Dr. Kenneth Kreutz-Delgado, Dr. Robert Bitmead, and Dr. William S. Hodgkiss for their valuable comments and time. Without the help, contribution and affection from my advisor and committee members, I would not be able to complete this work.

I owe a lot to my collaborators Chandra R. Murthy and Ethan R. Duni, and former lab-mate June C. Roh. I would like to thank them for many stimulating discussions and critical feedback, which greatly helped with the development of this work. I also would like to thank other group members: David Wipf, Yogananda Isukapalli, Aditya Jagannatham Cecile Levasseur, Shankar Shivappa, Wenyi Zhang, and Hairuo Zhuang, for the encouragement and insightful discussions.

Finally, I dedicate my dissertation to my father, mother, and my girlfriend Ning. Their love are the cornerstones of my life.

This work was supported in part by CoRe grant No. 02-10109 sponsored by Ericsson and in part by the U. S. Army Research Office under the Multi-University Research Initiative (MURI) grant No. W911NF-04-1-0224.

This dissertation is a collection of papers that were published or submitted for publication. The text of Chap. 2 is in part a reprint of the material which was coauthored with Ethan R. Duni and Bhaskar D. Rao and has been accepted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of*

multiple antenna systems with finite rate feedback using high resolution quantization theory". Chap. 3 and Chap. 4, in part and under some rearrangements, are reprints of papers which were coauthored with Bhaskar D. Rao and have been submitted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels"*, and in *IEEE Journal on Selected Areas in Communications* under the title *"Analysis of vector quantizers using transformed codebook with application to feedback-based multiple antenna systems"* respectively. Chap. 5 is in part a reprint of the paper which was coauthored with Bhaskar D. Rao and has been accepted in *Proceedings of IEEE Asilomar Conference 2006*, and will be submitted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of MIMO systems with finite-rate channel state information feedback"*. Chap. 6 is in part a reprint of the paper which was coauthored with Bhaskar D. Rao and has been submitted for publication in *IEEE Transactions on Information Theory* under the title *"A study of limits on training via capacity analysis of MIMO systems with unknown channel state information"*. The text of Chap. 7 is in part a reprint of the material which was coauthored with Bhaskar D. Rao and has been published in *IEEE Transactions on Signal Processing* under the title *"LDPC-coded MIMO systems with unknown block fading channels: soft MIMO detector design, channel estimation, and code optimization"*. The dissertation author was the primary researcher and author, and the co-authors listed in these publications contributed to or supervised the research which forms the basis for this dissertation.

VITA

1978	Born, Hangzhou, China
2001	B.S. in Electrical Engineering, Tsinghua University, Beijing, China
2001–2003	Research Assistant, Department of Electrical and Computer Engineering, Texas A& M University, College Station, TX
2003	M.S. in Electrical and Computer Engineering, Texas A& M University, College Station, TX
2003–2006	Research Assistant, Department of Electrical and Computer Engineering, University of California, San Diego
2006	Ph.D. in Electrical and Computer Engineering, University of California, San Diego, CA

PUBLICATIONS

J. Zheng and B. D. Rao, "LDPC-coded MIMO systems with unknown block fading channels: soft MIMO detector design, channel estimation, and code optimization," *IEEE Trans. on Signal Processing*, vol. 54, pp. 1504–1518, Apr. 2006.

J. Zheng, E. R. Duni, and B. D. Rao, "Analysis of multiple antenna systems with finite rate feedback using high resolution quantization theory," *IEEE Trans. on Signal Processing*, 2006 (to appear).

J. Zheng and B. D. Rao, "Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels," *submitted to IEEE Trans. on Signal Processing*, 2006.

J. Zheng and B. D. Rao, "Analysis of MIMO systems with finite-rate channel state information feedback," *IEEE Trans. on Signal Processing*, in preparation.

J. Zheng and B. D. Rao, "A study of limits on training via capacity analysis of MIMO systems with unknown channel state information," *submitted to IEEE Trans. on Information Theory*, Feb. 2005.

J. Zheng and B. D. Rao, "Analysis of vector quantizers using transformed codebook with application to feedback-based multiple antenna systems," *submitted to IEEE Journal on Selected Areas in Communications*, June 2006.

J. Zheng and B. D.Rao, "Capacity analysis of MIMO systems with unknown channel state information," in *IEEE Information Theory Workshop 2004*, San Antonio, Oct. 2004.

J. Zheng, E. R. Duni, and B. D. Rao, "Analysis of multiple antenna systems with finite-rate feedback using high resolution quantization theory," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2006, pp. 73–82.

J. Zheng and B. D. Rao, "Capacity analysis of multple antenna systems with mismatched channel quantization schemes," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

J. Zheng and B. D. Rao, "Capacity analysis of correlated multiple antenna systems with finite rate feedback," in *IEEE International Conference on Communications*, Istanbul, Turkey, June 2006.

J. Zheng and B. D. Rao, "Analysis of vector quantizers using transformed codebook with application to feedback-based multiple antenna systems," in *Proceedings of EUSIPCO*, Florence, Italy, Sept. 2006 (to appear).

C. R. Murthy, J. Zheng, and B. D. Rao, "Multiple antenna systems with finite rate feedback," in *IEEE Military Communications Conference, 2005*, Atlantic City, NJ, Oct. 2005, pp. 1–7.

J. Zheng and B. D. Rao, "LDPC-Coded MIMO receiver design over unknown fading channels," in *Proc. IEEE Globecom 2005*, St. Louis, MO, Nov. 2005, pp. 2942–2947.

J. Zheng and B. D. Rao, "EM-based receiver design of LDPC-Coded MIMO systems over unknown fading channels," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, Mar. 2005, pp. 1009–1012.

FIELDS OF STUDY

Major Field: Engineering

Studies in communication theory, digital signal processing, information theory, estimation theory, and their applications in the optimization of MIMO, OFDM and CDMA wireless communication systems.

Professor Bhaskar D. Rao, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Design and Analysis of MIMO Systems With Practical Channel State Information Assumptions

by

Jun Zheng

Doctor of Philosophy in Electrical and Computer Engineering
(Communication Theory and Systems)

University of California, San Diego, 2006

Professor Bhaskar D. Rao, Chair

Using multiple antennas at both the transmitter and the receiver is one of the most promising techniques that can offer significant increases in channel capacity of a communication system in a wireless fading environment. However, the performance of the MIMO system depends heavily upon the availability of the channel state information (CSI) at the transmitter (CSIT) and at the receiver (CSIR). In this dissertation, we focus our attention on the design and analysis of MIMO systems over wireless fading channels with practical CSI assumptions, which can broadly be divided into the following two categories.

The first part considers the development of a general framework for the analysis of multiple antenna systems with finite-rate feedback, wherein the CSI is quantized at the receiver and conveyed back to the transmitter through a rate-constrained reverse link. Inspired by the results of classical high resolution quantization theory, the problem of finite rate quantized communication system is formulated as a general fixed-rate vector quantization problem with side information available at the encoder (or the quantizer) but unavailable at the decoder. The

framework of the quantization problem is sufficiently general to include quantization schemes with general non-mean square distortion functions, and constrained source vectors. Asymptotic distortion analysis of the proposed general quantization problem is provided by extending the vector version of the Bennett's integral. Specifically, tight lower and upper bounds of the average asymptotic distortion are provided together with useful insights from a source coding perspective. The proposed general methodology provides a powerful analytical tool to study a wide range of finite-rate feedback systems which includes both MISO systems over spatially correlated fading channels and MIMO systems over i.i.d. fading channels. The established framework is also versatile enough to provide analysis of suboptimal mismatched CSI quantizers and quantizers with transformed codebooks.

The second part of this dissertation is focused the on the design and analysis of MIMO systems over fading channels with CSI unavailable both at the transmitter and at the receiver. To be specific, we first provide an improved capacity lower bound for MIMO systems with unknown CSI. By analyzing (and optimizing) the proposed capacity lower bound with respect to different system parameters, we improve our intuition and understanding of the effects of training on the overall performance of MIMO systems under unknown CSI assumptions. Moreover, based on the capacity analysis results, we also provide the design of practical LDPC-coded MIMO systems under the same unknown CSI assumption at both component level and structural level. We first propose at the component level several soft-input soft-output MIMO detectors whose performances are much better than the conventional MMSE-based detectors. At the structural level, an unconventional iterative decoding scheme is proposed whose structure leads to a simple and efficient LDPC code degree profile optimization algorithm with proven global optimality and guaranteed convergence from any initialization.

1 Introduction

In this dissertation, we consider systems with multiple antennas at the transmitter and receiver, which are commonly referred to as multiple input multiple output (MIMO) systems. Communication systems using multiple antennas at both the transmitter and the receiver have recently received increased attention due to their ability to provide great capacity increases in a wireless fading environment. The initial excitement about MIMO was leaded by the pioneering work of Winters [1], Foschini [2], Gans [3] and Telatar [4, 5], which predicts remarkable spectral efficiencies for wireless systems with multiple transmit and receive antennas. The performance of MIMO system depends on the availability of the channel state information (CSI) at the transmitter and at the receiver. Most MIMO capacity analysis and system design often assumes perfect knowledge of the CSI is available at the receiver, and sometimes at the transmitter as well. This is not a realistic assumption for most practical communication systems especially for systems using frequency division duplexing schemes or over relatively fast fading channels. In this dissertation, we examine multiple antenna systems with practical channel state information assumptions.

1.1 Multiple Input Multiple Output Systems

Consider a MIMO system with M_T transmit antennas and M_R receive antennas. One can denote the impulse response between the j^{th} transmit antenna and the i^{th} receive antenna by $h_{i,j}(\tau, t)$, the MIMO channel is given by the $M_R \times M_T$

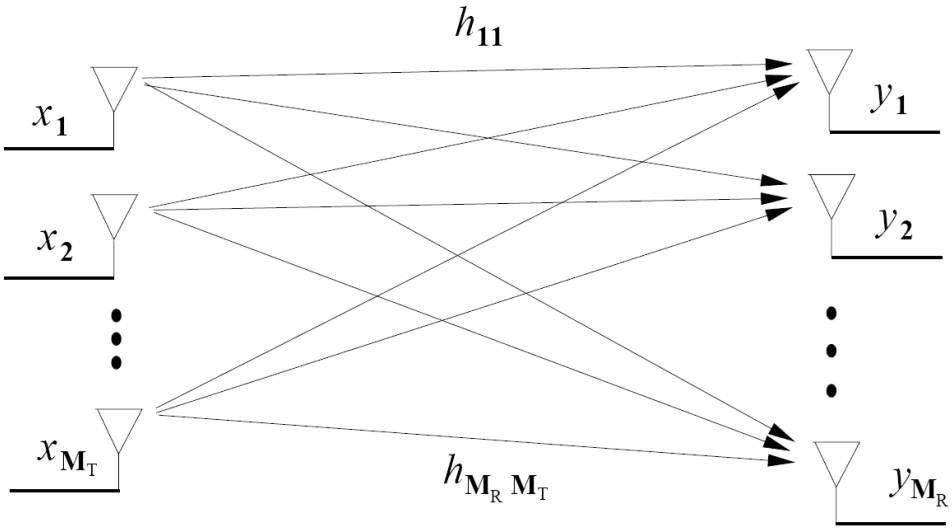

Figure 1.1: MIMO system model

matrix $\mathbf{H}(\tau, t)$ with

$$\mathbf{H}(\tau,t) = \begin{bmatrix} h_{1,1}(\tau,t) & h_{1,2}(\tau,t) & \cdots & h_{1,M_T}(\tau,t) \\ h_{2,1}(\tau,t) & h_{2,2}(\tau,t) & \cdots & h_{2,M_T}(\tau,t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{1,1}(\tau,t) & h_{1,2}(\tau,t) & \cdots & h_{1,M_T}(\tau,t) \end{bmatrix} . \tag{1.1}$$

If one denotes $s_j(t)$ as the signal transmitted from the j^{th} transmit antenna, the signal received at the i^{th} receive antenna, $y_i(t)$, can be represented as

$$y_i(t) = \sum_{j=1}^{M_T} h_{i,j}(\tau,t) \star s_j(t), \qquad i = 1, 2, \cdots, M_R \ , \tag{1.2}$$

where "$\star$" represents the time domain convolution between the transmitted signal and the channel impulse response.

In this dissertation, we consider a narrow-band MIMO system over block-fading channels. A narrow-band (frequency-flat) point-to-point communication system of M_T transmit and M_R receive antennas is shown in Fig. 1.1. This system can be represented by the following concise discrete time model:

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n} \ , \tag{1.3}$$

where $\mathbf{x}$ represents the M_T-dimensional transmitted symbol, $\mathbf{y}$ represents the M_R-dimensional received signal, $\mathbf{n}$ is the M_R-dimensional noise vector, and $\mathbf{H}$ is the $M_R \times M_T$ matrix of channel gains with the $(i,j)^{\text{th}}$ element $h_{i,j}$ representing the gain from transmit antenna j to receive antenna i. The additive noise is assumed to have a circularly symmetric complex Gaussian (CSCG) distribution with zero mean and identity covariance matrix, i.e. $\mathbf{n} \sim \mathcal{N}_c(\mathbf{0},\, I_t)$. The transmitted signal $\mathbf{x}$ is normalized to have a power constraint

$$E\left[\|\mathbf{x}\|^2\right] = \rho \ , \tag{1.4}$$

where ρ represents the average signal to noise ratio (SNR) of the MIMO system per receive antenna under unit channel gain. Furthermore, a block fading model is adopted in this thesis in the sense that the channel impulse response $\mathbf{H}$ is assumed to be constant within a coherent fading block and vary independently from one coherent block to another.

1.2 System Performance versus CSI Assumptions

The performance of a MIMO system depends on the availability of the channel state information at the transmitter as well as at the receiver. We briefly describe in the following some capacity analysis results of MIMO systems under several different CSI assumptions.

1.2.1 No CSIT and Perfect CSIR

For a MIMO system transmitting over i.i.d. Rayleigh flat fading channels, when the channel state information is perfectly accessible at the receiver side (perfect-CSIR), and completely unknown at the transmitter side (no-CSIT), the transmitted signal vector $\mathbf{x}$ is chosen to be statistically non-preferential, i.e.

$$E\Big[\mathbf{x}\mathbf{x}^{\mathsf{H}}\Big] = \frac{\rho}{M_T}\, I_{M_T} \ . \tag{1.5}$$

This means that the signals are independent and equal-powered at each transmit antennas. The system capacity (or mutual information rate) of a MIMO system under such CSI assumptions is given by [4],

$$C = E\left[\log_2\left(\mathbf{det}\Big(I_{M_R} + \rho \cdot \mathbf{H}\,\mathbf{H}^{\mathsf{H}}\Big)\right)\right] \ . \tag{1.6}$$

It is equivalent to the following form

$$C = E\left[\sum_{i=1}^{\min(M_T,M_R)} \log_2\left(1 + \frac{\rho \cdot \lambda_i}{M_T}\right)\right] \ , \tag{1.7}$$

where λ_i are the non-zero eigen-values of $\mathbf{HH}^{\mathsf{H}}$.

1.2.2 Perfect CSIT and Perfect CSIR

When the channel state information is unknown to the transmitter, as we have seen, equal power allocation across the transmit antenna array is a logical transmission scheme. On the other hand, if perfect channel state information is also available at the transmitter, adaptive transmission strategies, for example various transmit pre-coding schemes, can be applied. In this case, the channel capacity of a MIMO system under the perfect CSIR and perfect CSIT assumption is given by [4]

$$C = E\left[\max_{\mathbf{Q}:\mathbf{tr}(\mathbf{Q})=1} \log_2\left(\mathbf{det}\Big(I_{M_R} + \rho \cdot \mathbf{HQH}^{\mathsf{H}}\Big)\right)\right] \ , \tag{1.8}$$

where $\mathbf{Q}$ represents the covariance matrix of the input signal vector $\mathbf{x}$ based on a particular channel realization.

Telatar [4] showed that the MIMO channel can be converted to parallel, non-interfering single-input single-output (SISO) channels through a singular value decomposition (SVD) of the channel matrix. The SVD yields $\min(M_T, M_R)$ parallel channels with gains corresponding to the eigen-values values of $\mathbf{HH}^{\mathsf{H}}$. Water-filling the transmit power over these parallel channels leads to the power allocation

$$P_i = \left(\mu - \frac{1}{\rho\lambda_i}\right)^+, \qquad 1 \le i \le \min(M_T, M_R) \ , \tag{1.9}$$

where P_i is the power in the i^{th} eigen-mode of the channel, x^+ is defined as $x^+ \triangleq \max(x, 0)$, and μ is the waterfill level that satisfies,

$$\sum_{i=1}^{\min(M_T, M_R)} P_i = 1 \ . \tag{1.10}$$

The channel capacity is shown to be given by the following concise form

$$C = E\left[\sum_{i=1}^{\min(M_T, M_R)} \log_2\left(1 + \rho \cdot P_i \cdot \lambda_i\right)\right] \ . \tag{1.11}$$

Note that the assumption of perfect CSIT and perfect CSIR models a fading channel that changes slow enough to be reliably measured by the receiver and fed back to the transmitter without significant delay. However, for most practical channel environments, it is a stringent condition to satisfy.

1.2.3 No CSIT and No CSIR

As described in last subsection, with perfect CSIR channel capacity grows linearly with the minimum number of transmit and receive antennas. However, reliable channel estimation may not be possible for a mobile receiver that experiences rapid fluctuations of the channel responses. Since user mobility is the principal driving force for any wireless communication systems, the capacity behavior under the assumption that both transmitter and receiver only have channel statistical distributions is of particular interest.

In this section, we summarize some MIMO capacity results under the CSI assumption that both CSIT and CSIR are absent. One of the first papers to address this issue is by Marzetta and Hochwald [6], where they modeled the components of the channel matrix as i.i.d. complex Gaussian random variables that remain constant for a coherence interval of T symbol periods after which they change to another independent realization. The authors proved that the capacity is achieved when the transmitted signal matrix is equal to the product of two statistically independent matrices, i.e.

$$\mathbf{S} = \mathbf{\Phi} \cdot \mathbf{V} \ , \tag{1.12}$$

where $\mathbf{S} \in \mathbb{C}^{M_R \times T}$ is the capacity achieving transmitted signal, matrix $\mathbf{\Phi}$ is a $T \times T$ isotropically distributed unitary matrix, and $\mathbf{V}$ is a $M_R \times T$ independent real, nonnegative, diagonal matrix. Furthermore, the joint density of the diagonal elements of $\mathbf{V}$ is unchanged by rearrangements of its arguments. Marzetta and Hochwald showed that, for a fixed number of antennas, as the length of the coherence interval T increases, the capacity approaches the capacity obtained as if the receiver knew the propagation coefficients. In contrast to the linear growth of capacity with $\min(M_R, M_T)$ under the perfect CSIR assumption, the most interesting result shown in [6] is that in the absence of CSIT and CSIR, capacity does not increase at all as the number of transmit antennas M_T is increased beyond the length of the coherence interval T.

The MIMO capacity for this model was further explored by Zheng and Tse in [7]. The authors showed that in high SNR regimes, capacity is achieved when using no more than $M^* = \min\left(M_T, M_R, T/2\right)$ transmit antennas. Particularly, having more transmit antennas than receive antennas does not provide any capacity increase in high SNR regions. Zheng and Tse also computed the asymptotic capacity of the non-coherent MIMO channel in high SNR regions in terms of M_T, M_R, and T and showed that the capacity gain is $M^*(1 - M^*/T)$ bits per second per hertz for every 3-dB increase in SNR.

1.2.4 Role of Channel State Information

It can observed from the aforementioned capacity analysis results that the performance (e.g. in terms of capacity) of a MIMO system depends heavily upon the availability of the channel state information at both the transmitter and the receiver. It is also evident that MIMO systems with ideal CSIT and CSIR outperform systems with only CSIR and no CSIT, which also provide better performance than systems with no CSIR and no CSIT. Therefore, intuitively speaking, the role CSI in a MIMO system is to provide a mechanism for the transmitter and receiver to better adjust or adapt their communication strategies and hence achieve better

performance. However, a considerable amount of further effort is required in order to design and analyze MIMO systems with different CSI assumptions especially for practical environments.

1.3 Practical CSI Considerations

1.3.1 Systems with Partial CSIT

Considerable amount of work on the design and analysis of MIMO systems adopt two extreme CSIT assumptions: complete CSIT [4] [8] where channel state information is perfectly known at the transmitter and no CSIT [4]. However, in practical situations, the CSI assumption usually lie in between these two extremes, where the transmitter only has part of the channel information. Most frequency division duplex (FDD) systems are perfect examples of this kind of CSI assumptions, where the channel reciprocity is not valid due to asymmetry between the uplink and downlink channels. Hence the CSIT in this case has to be conveyed from the receiver through a reverse link, which is always rate constrained due to practical reasons. Even for time division dulplexing (TDD) systems, due to the asymmetry of the power amplifiers, some marginal channel information is still required from the receiver in order to gain full knowledge of the fading channel status. Therefore, a practical MIMO system has to face the challenge of designing transmission strategies that are able to make efficient usage of the partial CSI available at the transmitter. It is also equally challenging to analyze multiple antenna systems with partial CSIT, as well as to understand the effects of the rate constraint of the reverse link on the overall system performance.

The first part of this dissertation, which include Chap. 2 - Chap. 5, is focused on providing capacity analysis of multiple antenna systems with finite-rate CSI feedback. In Chap. 2, a general framework for the analysis of quantized feedback multiple antenna systems is developed by leveraging of the vast body of source coding theory, particularly high resolution quantization theory. Specifically,

the channel quantization is formulated as a general finite-rate vector quantization problem with attributes tailored to meet the general issues that arise in feedback based communication systems, including encoder side information, source vectors with constrained parameterizations, and general non-mean-squared distortion functions. Asymptotic (high quantization rate) distortion analysis of the proposed general quantization problem is provided by extending the vector version of the Bennett's integral. Specifically, tight lower and upper bounds of the average asymptotic distortion are proposed. The framework is then extended to provide asymptotic distortion analysis for the important practical problem of sub-optimal quantizers resulting from mismatches in the distortion functions, source statistics, and quantization criteria. Moreover, sub-optimal quantizer with transformed codebooks is also investigated and its distortion analysis is provided. Finally, distortion analysis of source variables with constrained parametrization, and analysis of complex source variables are also provided in this chapter.

The proposed methodology in Chap. 2 is quite general and provides a powerful analytical tool to study a wide range of finite rate feedback-based multiple antenna systems. By utilizing the proposed general distortion analysis, Chap. 3 investigates the effects of finite-rate CSI quantization on MISO systems over both spatially i.i.d. and correlated fading channels. Specifically, tight lower bounds of the average asymptotic distortion, which is defined as the system capacity loss due to the finite-rate channel quantization, are provided. Closed-form analysis of the capacity loss in high-SNR and low-SNR regimes are also provided. As an extended application of the proposed general framework, two types of mismatched MISO CSI quantizers are investigated in Chap. 4. These include quantizers that are designed with minimum mean square error (MMSE) criterion but the desired measure is ergodic capacity loss (i.e. mismatched design criterion), and quantizers whose codebooks are designed with a mismatched channel covariance matrix (i.e. mismatched statistics). Moreover, a MISO system transmitting over spatially correlated fading channels but using channel quantizers whose codebook is trans-

formed from spatially i.i.d. fading channels is also considered in Chap. 4. Bounds on the system capacity loss of these sub-optimal CSI feedback schemes, i.e. MISO systems with CSI quantizers using mismatched codebooks and transformed codebooks, are provided and compared to that of the optimal quantizers. Finally, the analysis of CSI-feedback-based multiple antenna systems is further extended to MIMO fading channels. In Chap. 5, tight lower bounds on the capacity loss of MIMO systems over i.i.d. Rayleigh flat fading channels due to the finite-rate channel quantization are provided. Moreover, MIMO CSI-quantizers using mismatched codebooks that only optimized for high-SNR and low-SNR regimes, as well as MIMO systems using multi-mode spatial multiplexing transmission schemes are also investigated in the same chapter.

1.3.2 Systems with unknown CSIR

MIMO capacity analysis and system design is often based on the assumption that the fading channel coefficient between each transmit and receive antenna pair is perfectly known at the receiver. In most of these cases, a quasi-static fading channel model is adopted where and channel state changes extremely slow and sufficient pilots are used to perform a close-to-ideal channel estimate. Based on the ideal CSIR, a coherent data detection is performed at the MIMO receiver. However, ideal CSIR is not a realistic assumption for most practical communication systems especially in fast fading channels. It is because perfect channel estimation at the receiver requires a sufficient long training sequence which is not feasible for a time-varying fading channel. Moreover, channel estimation always occupies some time slots that can be used for data transmission which decreases the spectral efficiency. Hence, there exists a trade-off between the amount of time and power spent on data symbols and training symbols. It is therefore an interesting problem to study the effects of training on the performance of multiple antenna systems with unknown CSIR assumptions. On the other hand, it is also a challenging problem to design practical coded MIMO systems under the same CSI assumptions.

The second part of this dissertation, which include Chap. 6 and Chap. 7, is focused the on the design and analysis of MIMO systems over fading channels with CSI unavailable both at the transmitter and at the receiver. In Chap. 6, a detailed capacity analysis of a MIMO system composed of M transmit and N receive antennas operating in a block fading environment with unknown CSI assumption is provided. Specifically, we first propose a mutual information upper bound of the unknown MIMO channel under the assumption that the input distribution is restricted to a certain structure and form but without assuming any specific channel estimation algorithm. The proposed mutual information upper bound is shown to be tight when M is moderately large or the system SNR is small, thereby leading to a valid capacity lower bound of the unknown MIMO channel. By analyzing the proposed capacity lower bound with respect to different system parameters, the effects of training symbols on the performance of MIMO systems are investigated from several different perspectives, including the design of optimal pilot structure, time slot allocation and power allocation between training and data symbols, as well as the selection of active number of transmit antennas.

Chap. 7 of this dissertation focuses on the design of practical LDPC-coded MIMO systems employing a soft iterative receiver structure with joint channel estimation and data detection scheme. To be specific, we first propose at the component level several soft-input soft-output MIMO detectors whose performances are much better than the conventional MMSE-based detectors. In particular, one optimal soft MIMO detector and two simplified sub-optimal detectors are developed that do not require an explicit channel estimate and offer an effective tradeoff between complexity and performance. In addition, a modified EM-based MIMO detector is developed which completely removes positive feedback between input and output extrinsic information and provide much better performance compared to the direct EM-based detector that has strong correlations especially in fast fading channels. At the structural level, the LDPC-coded MIMO receiver is constructed in an unconventional manner where the soft MIMO detector and LDPC

variable node decoder form one super soft-decoding unit, and the LDPC check node decoder forms the other component of the iterative decoding scheme. By exploiting the proposed receiver structure, tractable extrinsic information transfer functions of the component soft decoders are obtained, which further lead to a simple and efficient LDPC code degree profile optimization algorithm with proven global optimality and guaranteed convergence from any initialization.

2 Generalized Vector Quantization Framework and Asymptotic Analysis

2.1 Motivation

Communication systems using multiple antennas at both the transmitter and the receiver have recently received much attention due to their promise of providing significant capacity increases in a wireless fading environment, as reported by Telatar [4] and Foschini [2]. The performance of multiple antenna systems depends on the availability of the channel state information (CSI) at the transmitter (CSIT) and at the receiver (CSIR). Often in MIMO system design and analysis, two extreme CSIT assumptions are adopted: *complete CSIT* [4] [8] where channel state information is perfectly known at the transmitter and *no CSIT* [4]. This chapter considers systems with CSI assumptions in between these extremes. Perfect CSIR is assumed to be available at the receiver, and attention is focused on MIMO systems where CSI is conveyed from the receiver to the transmitter through a finite rate feedback link. Recently, several interesting papers have appeared, proposing design algorithms as well as analytically quantifying the performance of finite rate feedback multiple antenna systems [9] - [29].

Most of the widely used models for studying the effects of partial CSI at the transmitter fall into two categories: statistical feedback and instantaneous

feedback. In the statistical feedback approach, it is assumed that the channel is rapidly changing, and the coherence time is too small to feed back every instantaneous channel realization. However, the channel statistics varies sufficiently slowly, so that the mean and variance of the channel can be fed back to the transmitter accurately. The channel is then modeled as a Gaussian distribution with the given mean and variance, and the system performance is optimized with respect to the input distribution and analytically characterized [9]- [12]. In the instantaneous feedback approach, which is the focus of this work, a block fading channel model is assumed and the receiver conveys back to the transmitter current CSI through an error-free feedback link with limited bandwidth. To focus attention on the effects of the finite rate quantization of the CSI information, this chapter assumes there is no delay in the feedback channel. More specifically, given B bits of feedback, the receiver maps the current channel instantiation into one of $N = 2^B$ integer indices, with each index representing a particular mode of the fading channel. The transmitter hence optimizes or adapts its transmission strategy based on the feedback information. This imposes great challenges in the design and analysis of the optimal quantizer that takes into account both the underlying channel distribution and the performance metric, such as received SNR, channel capacity, bit error rate, etc. The topic of instantaneous channel feedback of multi-antenna systems has received much attention in the past few years, notably in [13] - [29].

Skoglund et. al. [13] investigated the finite rate MIMO system from an information theoretic perspective. They proved that when the receiver has full knowledge of the CSI and the feedback information, the capacity-achieving encoder can be split into two parts, a fixed codebook encoder and an adaptive weighting matrix based on the feedback information. Lau et. al. [14] extended the results to a more general case where the channel state s, the CSIR v, and the CSIT u have a joint statistical relation $p\big(s, u, v\big)$. They proved that the system capacity is obtained through a joint optimization of both the quantization partitions and the conditional input distributions, which is difficult to solve and can only be

resolved through numerical approximations. Therefore, most of the attention in recent literature is focussed on the case of ideal CSIR with simple sub-optimal (i.i.d. Gaussian) input distributions.

Narula et. al. considered in [15] a multiple transmit antennas and single receive antenna (MISO) system which employs finite-rate feedback to describe the beamforming vector. The Lloyd algorithm [16] was utilized to design the optimum beamforming vector codebook, where both the channel gain and the system mutual information were used as performance metrics. By relating the problem to rate distortion theory, the authors obtained an analysis of the SNR loss due to quantized beamforming and connect it to the system feedback rate B and the number of antennas t (when t is large). Based on the geometrical properties of the channel space, Mukkavilli et. al. [17] derived a universal lower bound on the outage probability of quantized MISO beamforming systems with arbitrary number of transmit antennas t over i.i.d. Rayleigh fading channels. The authors also proposed a codebook design criterion based on minimizing the maximum inner product between any two distinct beamforming vectors in the codebook. Love and Heath [18] [19] also derived the same min-max criterion in a i.i.d. Rayleigh fading MIMO channel setting, and related the problem to that of Grassmannian line packing [20], and proposed a random computer search algorithm to generate the codebook that optimizes the Grassmannian beamforming criterion. The authors also investigated in [21] the problem of quantizing the beamforming vector under a per-antenna power constraint, also referred to as quantized equal gain transmission.

Vector quantization (VQ) techniques along with the Lloyd algorithm can be used to design codebooks that specifically optimize for both the statistical distribution of the vector (or matrix) channel as well as the specific performance metric (for example, the mutual information rate). This approach was used by Xia et. al. in [22] [23] and Roh et. al. in [24] [25], where the authors derived an (weighted) inner product criterion and used the Lloyd algorithm [16] to generate the codebook. Both of these works analyzed the performance of MISO systems with limited rate-

feedback in the case of i.i.d. Rayleigh fading channels, and obtained closed-form expressions of the capacity loss (or SNR loss) in terms of feedback rate B and antenna size t. In [26] [27], Roh et. al. extended the results from MISO channels to the case of MIMO systems with quantized feedback. They employed a transmission scheme with a fixed number of spatial channels and equal power allocation, and proposed a new criterion for designing the codebook of beamforming matrices. By utilizing the complex multivariate beta distribution and tractable approximations to the Voronoi regions, they also provided corresponding analytical results of the system capacity loss for the case of i.i.d. Rayleigh fading channel in high SNR regimes. Furthermore, a multi-mode spatial multiplexing transmission strategy was proposed to compensate for the degradation due to the equal power allocation assumption which achieved a significant amount of the system capacity. A variant of the multi-model spatial multiplexing was also presented in [28]. The problem of quantized equal gain transmission was recently revisited by Murthy et. al. wherein a VQ approach was suggested for codebook design [29] and a closed-form capacity loss analysis was conducted.

The analysis of finite rate feedback systems has proven to be difficult and results available to date are quite limited: i.i.d. channels and mainly MISO channels. This chapter attempts to provide a general framework for the analysis of quantized feedback multiple antenna systems. This is done by exploiting the similarities between classical fixed-rate source coding and the channel quantization. For example, in the fixed-rate quantization problem, the encoder attempts to describe a random source (scalar or vector) using a finite number of bits with the goal being to minimize a chosen distortion measure (for example, a power of the Euclidean norm of the quantization error). In multiple antenna feedback systems, the channel state information is described using a finite number of bits with the goal being to optimize a given performance metric (such as the received SNR, system mutual information rate or BER). These similarities would be extremely helpful in the design and analysis of finite rate feedback MIMO systems as they

would benefit from the vast body of source coding theory, particularly high resolution quantization theory and VQ-based codebook design methodology. Although several authors have remarked on this similarity (including [15], [17] [23]), the exact and deeper connection between the two fields still remains elusive. A closer examination reveals that there are enough differences between the problems that a direct use of high resolution results from source coding is not feasible. Fortunately, however, it is possible to extend some of the results to the problem at hand and provide an interesting general framework for analyzing finite rate feedback systems.

Without narrowing the scope to a specific multi-antenna channel quantization scheme, this chapter formulates the problem as a general finite rate vector quantization problem with attributes tailored to meet the general issues that arise in feedback based communication systems. These attributes include side information available at the encoder (or quantizer) but unavailable at the decoder, general non-mean square distortion functions, and source vectors with constraints. Source coding with side-information-dependent distortion measures has been considered in [30] [31]. Those works focused on the classic rate-distortion approach, which is suited to variable rate coders, whereas this work focuses on the problem of fixed-rate quantizers and their associated high-rate theory. Additionally, those works considered a more limited form of the distortion measure than this paper. Asymptotic distortion analysis of the proposed general quantization problem with side information, constrained quantization space, and general distortion functions is provided by extending Bennett's classic analysis [32] as well as its corresponding vector extensions [33] [34] [35]. To be specific, tight lower and upper bounds of the average asymptotic distortion are proposed. Sufficient conditions on the achievability of the distortion bounds are also provided and related to corresponding classical fixed-rate quantization problems. Based on the general framework, the asymptotic distortion analysis is further extended to the important practical problem of sub-optimal quantizers resulting from mismatches in the distortion functions, source statistics, and quantization criteria. As a further demonstration

of the utility of the framework, sub-optimal vector quantizers using transformed codebooks are also investigated. Moreover, distortion analysis of complex source variables are also investigated in this chapter. It is shown that under certain necessary and sufficient conditions, the distortion analysis of complex source variables can be performed in a concise manner without first transforming the problem into real domains.

The proposed methodology from the source coding perspective provides a powerful analytical tool to study a wide range of finite rate feedback systems. Specific applications of the proposed general framework as well as the high-rate distortion analysis to multiple antenna systems over fading channels with finite-rate CSI feedback are provided in Chap. 3-5.

2.2 Generalized Vector Quantization Framework

In this section, the finite rate feedback based multiple antenna system is formulated as a generalized fixed-rate vector quantization problem and analyzed by adapting tools from high resolution quantization theory. The new attributes of this generalization are additional side information available to the encoder, constrained parameterizations of the vectors to be quantized, and non-mean-squared performance metrics.

2.2.1 Motivation for Generalization

To better understand the need for this generalization, an illustrative example is useful. For this purpose, consider a MISO system with t transmit antennas and a single receive antenna where the CSI to be quantized is the vector channel realization $\mathbf{h} \in \mathbb{C}^t$, which is equivalent to a real vector of $2t$ dimensions. In classical source coding, the encoder (or the quantizer) describes the random source $\mathbf{s} \in \mathbb{R}^k$ by one of the entries of a finite alphabet codebook denoted $\left\{ \widehat{\mathbf{s}}_1, \cdots, \widehat{\mathbf{s}}_N \right\}$, where $\widehat{\mathbf{s}}_i \in \mathbb{R}^k$. The encoder and the codebook are designed to minimize the dis-

tortion between $\mathbf{s}$ and its quantized representation $\widehat{\mathbf{s}}_i$ (for example, the expected r^{th} power of the Euclidean distance). The design of finite rate MISO feedback systems is a generalized channel (vector or matrix) quantization problem because of the following key differences:

1. *Redundant Parameters:* Not all channel parameters need to be quantized. For example, consider the quantization of the maximum ratio transmission (MRT) beamforming vector in a MISO system, which is given by $\mathbf{v} = \mathbf{h}/\|\mathbf{h}\|$ [36]. The transmitter only requires the channel directional information (vector) $\mathbf{v}$. Therefore, it is redundant to directly quantize the channel instantiation $\mathbf{h}$.

2. *Constrained Vector Parameterization:* The channel instantiation and the actual variable to be quantized may lie in different spaces and may have different dimensions because the information to be quantized may have certain constraints. In the example of quantized MRT beamforming, the vector $\mathbf{v} \in \mathbb{C}^t$ to be quantized is constrained to be unit-norm and hence lies on the unit hyper-sphere or manifold, whereas the channel instantiation $\mathbf{h}$ could be anywhere in $\mathbb{C}^t$ space.

3. *Encoder Side Information:* The side information which is not the quantization objective, for example the gain $\|\mathbf{h}\|$ of the MISO channel, can be utilized as side information at the quantizer (or the encoder) to improve the quantization performance.

4. *Non mean-squared performance metric:* The distortion measure is often a more general non-mean-square error function[1]. In the example of quantized MRT beamforming, if the average received SNR loss ρ_{L} is the performance metric, then the distortion measure is given by

$$\rho_{\text{L}} = \rho_{\text{P}} - \rho_{\text{Q}} = E_{\mathbf{h}}\left[\rho \cdot \|\mathbf{h}\|^2 \cdot \left(1 - \left|\langle \mathbf{v}, \widehat{\mathbf{v}} \rangle\right|^2\right)\right],$$

[1]Such distortion measures have also been considered in the source coding literature, and it is an unavoidable new attribute in the context of analyzing the multiple antenna systems with finite rate feedback.

where ρ_{P} is the expected received SNR with perfect beamforming feedback and ρ_{Q} is the expected received SNR with quantized beamforming, and ρ is the average system SNR. Notice that the distortion function is clearly not of the form $\|\mathbf{v} - \widehat{\mathbf{v}}\|^r$.

Due to the above-mentioned differences, high resolution quantization theory results from classical source coding cannot be directly applied to the design and analysis of finite rate feedback systems. In order to take advantage of the vast body of literature on source coding, the analysis must be extended to allow for encoder side information, constrained quantization variables and non-mean-squared distortion measures. The following sections extend Bennett's asymptotic distortion analysis to this more general framework thereby providing insight (asymptotic analysis) into the finite rate feedback based communication system problem.

2.2.2 Problem Formulation

Let $\mathbf{y}$ be a $k_{\mathrm{q}} \times 1$ random vector belonging to vector space $\mathbb{Q}\left(\mathbb{R}^{k_{\mathrm{q}} \times 1}\right)$, and $\mathbf{z}$ be a $k_{\mathrm{z}} \times 1$ random vector belonging to vector space $\mathbb{Z}\left(\mathbb{R}^{k_{\mathrm{z}} \times 1}\right)$ respectively. Vector $\mathbf{y}$ and $\mathbf{z}$ can be combined into $\mathbf{x} = \big(\mathbf{y}, \mathbf{z}\big)$, which belongs to space $\mathbb{S} = \mathbb{Q} \times \mathbb{Z}$ and has joint probability density $p\big(\mathbf{x}\big) = p\big(\mathbf{y}, \mathbf{z}\big)$. A quantization scheme is to be designed to quantize the random vector $\mathbf{y}$ into one of the N vectors (or code points) $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$, each belonging to space $\mathbb{Q}$. The performance of the quantizer is measured by the distortion function D_{Q}. In the conventional source coding context, where the objective is to quantize and reproduce the source information $\mathbf{y}$, the most commonly used distortion measure is the r^{th} power of the mean-squared quantization error ($\widetilde{\mathbf{y}} = \mathbf{y} - \widehat{\mathbf{y}}$), i.e.

$$D_{\mathrm{norm}} = E\Big[\, \big\|\mathbf{y} - \widehat{\mathbf{y}}\big\|_2^r \Big] \,, \tag{2.1}$$

where $\|\cdot\|_2$ denotes the l_2 norm. The definition of the distortion measure is extended to a general non-mean-square form, which is also parameterized by the

side information $\mathbf{z}$, and given by

$$D = E_{\mathbf{x}}\Big[D_{\mathrm{Q}}\Big(\mathbf{y}, \widehat{\mathbf{y}}\,;\, \mathbf{z}\Big)\Big] \ , \tag{2.2}$$

where for a given instantiation of $\mathbf{z}$, $D_{\mathrm{Q}}\big(\,\cdot\,,\,\cdot\,;\,\mathbf{z}\big)$ is a mapping from space $\mathbb{Q}\times\mathbb{Q}$ to the real domain $\mathbb{R}^{+}$. The distortion function D_{Q} can be viewed as a generalized multi-dimensional distance function between $\mathbf{y}$ and $\widehat{\mathbf{y}}$ that is parameterized by $\mathbf{z}$, and it is assumed to have the following property

$$D_{\mathrm{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}}\,;\, \mathbf{z}\big) \geq 0 \ , \tag{2.3}$$

with equality in the neighborhood of $\widehat{\mathbf{y}}$ if and only if $\mathbf{y} = \widehat{\mathbf{y}}$ (local minima). It is also assumed that the side information $\mathbf{z}$ is available at the encoder (or the quantizer), but not available at the decoder. Therefore, the quantized output $\widehat{\mathbf{y}}$ is a function of both the quantization variable (objective) $\mathbf{y}$ and the side information $\mathbf{z}$, i.e.

$$\widehat{\mathbf{y}} = Q\big(\mathbf{x}\big) = Q\big(\mathbf{y}\,,\,\mathbf{z}\big) \ .$$

Due to the unavailability of the side information $\mathbf{z}$ at the decoder, a single codebook (*not* multiple codebooks indexed by $\mathbf{z}$) is used and known to both the encoder and the decoder.

This chapter considers a fixed-rate[2] (B bits) quantizer with $N = 2^{B}$ quantization levels. For each of the N vectors (or code points) in the codebook, there is an unique region $\mathbb{S}_i \subseteq \mathbb{S}$ corresponding to the set of the source input that is quantized into $\widehat{\mathbf{y}}_i$, defined as

$$\mathbb{S}_i = \Big\{\mathbf{x}\,\Big|\, Q\big(\mathbf{x}\big) = \widehat{\mathbf{y}}_i\Big\} \ . \tag{2.4}$$

A quantizer can be specified by the output points ("codebook") and by the partition of the input source space $\mathbb{S}$, which is composed of N disjoint and exhaustive regions $\mathbb{S}_1, \mathbb{S}_2, \cdots, \mathbb{S}_N$, i.e.

$$\mathbb{S} = \bigcup_{i=1}^{N} \mathbb{S}_i, \qquad \mathbb{S}_i \bigcap \mathbb{S}_j = \phi, \ i \neq j \ .$$

[2]The asymptotic analysis provided in this chapter can be extended to variable rate quantizers with fixed output entropy (or the average message length) following the work in [33].

Finally, viewed from a conventional source coding perspective, the described general quantization problem is equivalent to the quantization of a mixed density source with each source component having probability density given by $p(\mathbf{y}|\mathbf{z})$, and parameterized distortion function given by $D_{\mathrm{Q}}(\mathbf{y}, \widehat{\mathbf{y}}\,;\, \mathbf{z})$. The side information $\mathbf{z}$ (index information of the source component) is available at the encoder (or the quantizer) but not available at the decoder. In the following subsections, a detailed asymptotic distortion analysis is provided for the proposed general vector quantization problem.

2.2.3 Optimal Partitioning of the Source Space

Similar to the analysis provided in [33], this analysis begins by exploring the geometrical properties of the partitions of the source space $\mathbb{S}$. First observe the partitions in the space $\mathbb{Q}$ by projecting the partition region $\mathbb{S}_i$ onto $\mathbb{Q}$ conditioned on $\mathbf{z}$:

$$\mathbb{Q}_{\mathbf{z},i} = \left\{ \mathbf{y} \,\middle|\, (\mathbf{y},\, \mathbf{z}) \in \mathbb{S}_i \right\} . \tag{2.5}$$

Therefore, the conditional space $\mathbb{Q}_{\mathbf{z}}$, given by

$$\mathbb{Q}_{\mathbf{z}} = \left\{ \mathbf{y} \,\middle|\, (\mathbf{y},\, \mathbf{z}) \in \mathbb{S} \right\} , \tag{2.6}$$

can be represented as a union of all the non-overlapping projected partitions $\mathbb{Q}_{\mathbf{z},i}$ conditioned on $\mathbf{z}$, i.e.

$$\mathbb{Q}_{\mathbf{z}} = \bigcup_{i=1}^{N} \mathbb{Q}_{\mathbf{z},i} \ .$$

In order to separate the effects of the original source distribution on the quantization systems, following the approach in [33], first consider the optimal quantizer for a random source which is uniformly distributed on space $\mathbb{S}$. The optimal quantizer that minimizes distortion will satisfy the following two conditions. First, for any quantization points (or codebook) $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$ in $\mathbb{Q}$, the optimal quantizer has a Dirichlet partition, given by

$$\mathbb{S}_i = \left\{ \mathbf{x} \,\middle|\, D_{\mathrm{Q}}(\mathbf{y},\, \widehat{\mathbf{y}}_i\,;\, \mathbf{z}) \le D_{\mathrm{Q}}(\mathbf{y},\, \widehat{\mathbf{y}}_j\,;\, \mathbf{z}) ,\ \forall\, j \neq i \right\} , \qquad 1 \le i \le N \ . \tag{2.7}$$

It is known that a quantizer is characterized by the its Voronoi partitions as well as the corresponding centroid (or code points). For an optimal quantizer, first it can be shown that each projected partition conditioned a particular side information is also Dirichlet partition.

Lemma 1 *With uniformly distributed input source* $\mathbf{x}$*, if partition* $\mathbb{S}_1, \mathbb{S}_2, \cdots, \mathbb{S}_N$ *is a Dirichlet partition with respect to the quantization points (or codebook)* $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$ *in* $\mathbb{Q}$*, then each projected partition* $\mathbb{Q}_{\mathbf{z},1}, \mathbb{Q}_{\mathbf{z},2}, \cdots, \mathbb{Q}_{\mathbf{z},N}$ *is also a Dirichlet partition w.r.t. points* $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$ *in space* $\mathbb{Q}_{\mathbf{z}}$*, i.e.*

$$\mathbb{Q}_{\mathbf{z},i} = \Big\{ \mathbf{y} \,\Big|\, D_Q\big(\mathbf{y}, \widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big) \le D_Q\big(\mathbf{y}, \widehat{\mathbf{y}}_j \,;\, \mathbf{z}\big)\,,\ \forall\, j \neq i \Big\}\,, \qquad 1 \le i \le N\ . \tag{2.8}$$

Proof: According to the definition of the projected partitions $\mathbb{Q}_{\mathbf{z},i}$ in (2.5), for any two elements $\mathbf{y}_i \in \mathbb{Q}_{\mathbf{z},i}$ and $\mathbf{y}_j \in \mathbb{Q}_{\mathbf{z},j}$, there exists $\mathbf{x}_i$ and $\mathbf{x}_j$ such that

$$\mathbf{x}_i = \big(\mathbf{y}_i,\, \mathbf{z}\big) \in \mathbb{S}_i, \qquad \mathbf{x}_j = \big(\mathbf{y}_j,\, \mathbf{z}\big) \in \mathbb{S}_j\ . \tag{2.9}$$

Therefore, according to the definition of the Dirichlet partition given by (2.7), the following inequality is true for any $j \neq i$,

$$D_{\mathrm{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big) \le D_{\mathrm{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}}_j \,;\, \mathbf{z}\big)\ . \tag{2.10}$$

This is exactly the condition of the Dirichlet partion on projected space $\mathbb{Q}_{\mathbf{z}}$ given by (2.8). ■

Second, each output point $\widehat{\mathbf{y}}_i$ is the centroid of its corresponding region, in the sense that

$$\widehat{\mathbf{y}}_i = \arg\min_{\widehat{\mathbf{y}}} \iint_{(\mathbf{y},\,\mathbf{z}) \in \mathbb{S}_i} D_{\mathrm{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}} \,;\, \mathbf{z}\big)\, d\mathbf{y}\, d\mathbf{z}\ . \tag{2.11}$$

Note that generally, the centroid point $\widehat{\mathbf{y}}_i$ of $\mathbb{S}_i$ might *not* be the centroid of each projected region $\mathbb{Q}_{\mathbf{z},i}$.

Due to the assumption that the side information $\mathbf{z}$ is only available at the encoder (or the quantizer), all source components ($\mathbf{y}$ conditioned on different

instantiations of $\mathbf{z}$) share the same codebook $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$. Therefore, once the codebook is fixed, the Dirichlet partitions on each projected space $\mathbb{Q}_{\mathbf{z}}$ are determined by the distortion function $D_{\mathrm{Q}}\big(\,\cdot\,,\,\cdot\,;\,\mathbf{z}\big)$ parameterized by $\mathbf{z}$. This means that the partitions of a specific quantizer on each different projected space $\mathbb{Q}_{\mathbf{z}}$ are related to each other. Therefore, the optimal codebook (or the placement of the code points) should be designed to minimize the overall distortion, as opposed to the distortion for any one conditional source component. This imposes a great challenge not only on the quantizer design but also on the distortion analysis.

2.2.4 Normalized Inertial Profile

By performing a Taylor series expansion on the distortion function $D_{\mathrm{Q}}\big(\mathbf{y},\,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)$ about $\mathbf{y} = \widehat{\mathbf{y}}$, the distortion measure can be represented in the following form:

$$\begin{aligned} D_{\mathrm{Q}}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big) &= D_{\mathrm{Q}}\big(\,\widehat{\mathbf{y}}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big) + \mathbf{d}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\,\big(\mathbf{y}-\widehat{\mathbf{y}}\big) \\ &+ \big(\mathbf{y}-\widehat{\mathbf{y}}\big)^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\,\big(\mathbf{y}-\widehat{\mathbf{y}}\big) + O(\|\mathbf{y}-\widehat{\mathbf{y}}\|^3)\ , \end{aligned} \tag{2.12}$$

where $\mathbf{d}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big) \in \mathbb{R}^{1\times k_{\mathrm{q}}}$ is the gradient of D_{Q} given by

$$\mathbf{d}\big(\widehat{\mathbf{y}}\big) = \frac{\partial}{\partial\,\mathbf{y}}\bigg|_{\mathbf{y}=\widehat{\mathbf{y}}} D_{\mathrm{Q}}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)\ ,$$

and $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big) \in \mathbb{R}^{k_{\mathrm{q}}\times k_{\mathrm{q}}}$ is the Hessian matrix of the distortion function with the $(i,j)^{\mathrm{th}}$ element given by

$$w_{i,j} = \frac{1}{2}\cdot\frac{\partial^2}{\partial\,y_i\partial\,y_j}\bigg|_{\mathbf{y}=\widehat{\mathbf{y}}} D_{\mathrm{Q}}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)\ .$$

According to the definition of the distortion function as well as the property given by (2.3) that $\mathbf{y} = \widehat{\mathbf{y}}$ is the local minimum of D_{Q}, both $D_{\mathrm{Q}}\big(\,\widehat{\mathbf{y}}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)$ and $\mathbf{d}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$ are zero, and the Hessian matrix $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$ is positive semi-definite. Therefore, as N (or B) gets large (i.e. the high resolution assumption) D_{Q} can be approximated by the following second order Taylor series expansion:

$$D_{\mathrm{Q}}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big) \approx \big(\mathbf{y}-\widehat{\mathbf{y}}\big)^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\,\big(\mathbf{y}-\widehat{\mathbf{y}}\big)\ . \tag{2.13}$$

The quadratic matrix $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$ is an extension of the "sensitivity matrix" defined in [34], which describes the scalar sensitivities of the parameters and cross-sensitivity terms related to the interaction in quantizing multiple parameters simultaneously.

The quantization region $\mathbb{Q}_{\mathbf{z},i}$ can be viewed as a shifted area $\mathbb{E}_{\mathbf{z},i}$ centered at $\widehat{\mathbf{y}}_i$, defined as

$$\mathbb{E}_{\mathbf{z},i} = \Big\{ \mathbf{e} \,\Big|\, \mathbf{e} + \widehat{\mathbf{y}}_i \in \mathbb{Q}_{\mathbf{z},i} \Big\} \ , \tag{2.14}$$

where $\mathbb{E}_{\mathbf{z},i}$ is the Voronoi region of code point $\widehat{\mathbf{y}}_i$ in space $\mathbb{Q}_{\mathbf{z}}$. Therefore, the average distortion $D_{\mathbf{z},i}$ in the quantization region $\mathbb{Q}_{\mathbf{z},i}$ depends on its location $\widehat{\mathbf{y}}_i$ and the adopted Voronoi shape $\mathbb{E}_{\mathbf{z},i}$, which is given by

$$\begin{aligned} D_{\mathbf{z},i} &= \int_{\mathbb{Q}_{\mathbf{z},i}} D_{\mathrm{Q}}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big)\ d\,\mathbf{y} = \int_{\mathbf{e}+\widehat{\mathbf{y}}_i \in \mathbb{Q}_{\mathbf{z},i}} D_{\mathrm{Q}}\big(\,\mathbf{e}+\widehat{\mathbf{y}}_i\,,\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big)\ d\,\mathbf{e} \\ &\approx \int_{\mathbf{e}\in\mathbb{E}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ \ d\,\mathbf{e} = V\big(\mathbb{Q}_{\mathbf{z},i}\big)^{1+\frac{2}{k_{\mathrm{q}}}} \cdot I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big), \end{aligned} \tag{2.15}$$

where $V\big(\mathbb{Q}_{\mathbf{z},i}\big)$ or $V\big(\mathbb{E}_{\mathbf{z},i}\big)$ is the volume of the region $\mathbb{Q}_{\mathbf{z},i}$ defined as

$$V\big(\mathbb{Q}_{\mathbf{z},i}\big) = V\big(\mathbb{E}_{\mathbf{z},i}\big) = \int_{\mathbb{E}_{\mathbf{z},i}} d\mathbf{e}\ , \tag{2.16}$$

and $I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big)$ is the normalized inertial profile defined as

$$I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) \ \triangleq\ V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}\in\mathbb{E}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ \ d\,\mathbf{e}\ . \tag{2.17}$$

The normalized inertial profile given by (2.17) depends on the shape, but not the size, of the region $\mathbb{E}_{\mathbf{z},i}$ and on the sensitivity matrix $\mathbf{W}_{\mathbf{z}}(\widehat{\mathbf{y}}_i)$. Hence, it is invariant to an arbitrary scaling, which is proved in the following lemma.

Lemma 2 *The normalized inertial profile* $I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big)$ *(including the constraint inertial profile) given by (2.17) and (2.97) depends only on the shape of the Voronoi region* $\mathbb{E}_{\mathbf{z},i}$ *and the local smoothness of the distortion function at its current location* $\widehat{\mathbf{y}}_i$*. It is invariant to the following scaling operation (within the small neighborhood of point* $\widehat{\mathbf{y}}_i$*), i.e.*

$$I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) = I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\alpha\,\mathbb{E}_{\mathbf{z},i}\Big)\ . \tag{2.18}$$

Proof: First, according to the definition of the volume of the region $\mathbb{E}_{\mathbf{z},i}$ given by (2.16), it is clear that the volume of the scaled region $\alpha\mathbb{E}_{\mathbf{z},i}$, defined as

$$\alpha\mathbb{E}_{\mathbf{z},i} = \Big\{ \alpha\,\mathbf{e} \,\Big|\, \mathbf{e} \in \mathbb{E}_{\mathbf{z},i} \Big\} \ , \tag{2.19}$$

is given by

$$V\big(\alpha\,\mathbb{E}_{\mathbf{z},i}\big) = \int_{\alpha\,\mathbb{E}_{\mathbf{z},i}} d\mathbf{e} = \int_{\mathbb{E}_{\mathbf{z},i}} d\left(\alpha\,\mathbf{e}'\right) = \int_{\mathbb{E}_{\mathbf{z},i}} \alpha^{k_{\mathrm{q}}}\ d\,\mathbf{e}' = \alpha^{k_{\mathrm{q}}} \cdot V\big(\mathbb{E}_{\mathbf{z},i}\big) \ . \tag{2.20}$$

With the same reasoning, the inertial profile of a scaled region is given by

$$\begin{aligned} I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\alpha\,\mathbb{E}_{\mathbf{z},i}\Big) &= V\big(\alpha\,\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}\in\,\alpha\mathbb{E}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ d\,\mathbf{e} \\ &= \Big(\alpha^{k_{\mathrm{q}}}\,V\big(\mathbb{E}_{\mathbf{z},i}\big)\Big)^{-\left(1+\frac{2}{k_{\mathrm{q}}}\right)} \int_{\mathbf{e}'\in\mathbb{E}_{\mathbf{z},i}} \alpha\mathbf{e}'^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\alpha\mathbf{e}'\ d\left(\alpha\mathbf{e}'\right) \\ &= \alpha^{-(k_{\mathrm{q}}+2)}\,\alpha^{k_{\mathrm{q}}+2}\ \cdot V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}'\in\mathbb{E}_{\mathbf{z},i}} \mathbf{e}'^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}'\ d\,\mathbf{e}' \\ &= I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) \ . \end{aligned} \tag{2.21}$$

■

According to Gersho's conjecture [33], for large N most cells of the optimal quantizer are congruent (shifted, rotated, scaled, and elongated versions) of the tessellating polytope $\mathcal{H}$. Therefore, the optimal Voronoi region that minimizes the inertial profile must belong to the set of admissible tessellating polytopes in space $\mathbb{Q}_{\mathbf{z}}$ (i.e. regions that can tile the space). The optimal inertial profile is defined as the minimum inertia of all admissible regions $\mathbb{E}_{\mathbf{z},i}$, i.e.

$$I_{\mathrm{opt}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\Big) = \min_{\mathbb{E}_{\mathbf{z},i}\in\mathcal{H}_{\mathrm{Q}}} I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) \ , \tag{2.22}$$

where $\mathcal{H}_{\mathrm{Q}}$ is the class of all admissible polytopes. It is proved in [34] [37] [35] [38] that the inertial profile of any Voronoi shape $\mathbb{E}_{\mathbf{z},i}$ is lower bounded by that of a "M-shaped" hyper-ellipsoid, i.e.

$$\begin{aligned} I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) &= V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}\in\mathbb{E}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ d\,\mathbf{e} \\ &\geq V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}\,\in\,\mathcal{T}\left(\mathbf{0},\,\mathbf{W}_{\mathbf{z}}(\widehat{\mathbf{y}}_i),\,V(\mathbb{E}_{\mathbf{z},i})\right)} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ d\,\mathbf{e} \ , \end{aligned} \tag{2.23}$$

where $\mathcal{T}\big(\mathbf{y},\, \mathbf{M},\, v\big)$ is the hyper-ellipsoidal set centered at $\mathbf{y}$ with volume v, defined as

$$\mathcal{T}\big(\mathbf{y},\, \mathbf{M},\, v\big) = \left\{ \mathbf{x} \,\middle|\, \left(\frac{\kappa_{k_\mathrm{q}}^2}{v^2 |\mathbf{M}|} \right)^{1/k_\mathrm{q}} \big(\mathbf{x}-\mathbf{y}\big)^{\mathsf{T}} \mathbf{M} \big(\mathbf{x}-\mathbf{y}\big) \;\leq 1 \right\} , \tag{2.24}$$

where κ_n is the volume of a n-dimensional unit sphere [39] given by

$$\kappa_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)} \;\; .$$

Carrying out the multi-dimensional integration using the same approach as in [34] [38], the lower bound of the inertial profile is given by,

$$I_{\mathrm{opt}}\big(\widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big) \gtrapprox \widetilde{I}_{\mathrm{opt}}\big(\widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big) = \frac{k_\mathrm{q}}{k_\mathrm{q}+2} \cdot \left(\frac{\big|\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)\big|}{\kappa_{k_\mathrm{q}}^2} \right)^{1/k_\mathrm{q}} \;\; . \tag{2.25}$$

As reported in [38], the lower bound is tight (denoted as "$\gtrapprox$") in most cases and the optimal Voronoi regions can be well approximated by the "M-shaped" hyper-ellipsoids. In [40], the error introduced by the hyper-ellipsoidal approximation was investigated for spaces up to dimensions 10 and was shown to be insignificant. Therefore, although the hyper-ellipsoids cannot form a lattice partition of space $\mathbb{Q}_{\mathbf{z}}$, the difference in the inertial profiles is insignificant. On the other hand, it is also evident that the inertial profile of any admissible polytope is an upper bound on $I_{\mathrm{opt}}\big(\widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big)$.

In some practical cases, the sensitivity matrix $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)$ is independent of the location $\widehat{\mathbf{y}}_i$, i.e.

$$\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big) = \mathbf{W}_{\mathbf{z}} \;\; . \tag{2.26}$$

Then, it is reasonable to assume that the cell regions $\mathbb{E}_{\mathbf{z},i}$ of an optimal quantizer are only scaled and rotated versions of the polytope shape $\mathcal{H}$. Hence, the inertial profile $I_{\mathrm{opt}}\big(\widehat{\mathbf{y}}_i \,;\, \mathbf{z}\big)$ of an optimal quantizer reduces to be the optimal inertia coefficient $I_{\mathrm{opt}}(\mathbf{z})$, which is constant for any location vector $\widehat{\mathbf{y}}_i$. This is also true for the tight lower bound $\widetilde{I}_{\mathrm{opt}}(\mathbf{z})$ given by (2.25).

2.2.5 Heuristic Derivation of the Asymptotic Distortion Integral

To generalize the concept of point density introduced by Lloyd [41] in one-dimensional quantization, and extended to vector quantization in [33], denote

by $\lambda_{\mathbf{z},N}(\widehat{\mathbf{y}}_i)$ the *specific point density function* for a given side information $\mathbf{z}$, i.e.

$$\lambda_{\mathbf{z},N}(\widehat{\mathbf{y}}_i) = \frac{1}{NV(\mathbb{Q}_{\mathbf{z},i})} \ , \tag{2.27}$$

where $V(\mathbb{Q}_{\mathbf{z},i})$ is the volume of $\mathbb{Q}_{\mathbf{z},i}$. When the number of the quantization level N is sufficiently large, $\lambda_{\mathbf{z},N}(\widehat{\mathbf{y}}_i)$ has an *asymptotic point density* given by

$$\lambda_{\mathbf{z}}(\widehat{\mathbf{y}}) = \lim_{N\to\infty} \lambda_{\mathbf{z},N}(\widehat{\mathbf{y}}_i) \ . \tag{2.28}$$

Similar to Bennett's integral provided in [32], the system distortion conditioned on a particular side information $\mathbf{z}$ can be rewritten in the following form:

$$\begin{aligned}
D(\mathbf{z}) &= E_{\mathbf{y}|\mathbf{z}}\Big[D_{\mathrm{Q}}\Big(\mathbf{y}\,, Q(\mathbf{y},\mathbf{z})\,;\,\mathbf{z}\Big)\Big] \\
&= \int_{\mathbb{Q}_{\mathbf{z}}} D_{\mathrm{Q}}\Big(\mathbf{y}\,, Q(\mathbf{y},\mathbf{z})\,;\,\mathbf{z}\Big)\cdot p(\mathbf{y}|\,\mathbf{z})\,d\,\mathbf{y} \\
&= \sum_{i=1}^{N}\int_{\mathbb{Q}_{\mathbf{z},i}} D_{\mathrm{Q}}\Big(\mathbf{y}\,, Q(\mathbf{y},\mathbf{z})\,;\,\mathbf{z}\Big)\cdot p(\mathbf{y}|\,\mathbf{z})\,d\,\mathbf{y} \\
&\approx \sum_{i=1}^{N} p(\widehat{\mathbf{y}}_i|\mathbf{z})\int_{\mathbb{Q}_{\mathbf{z},i}} D_{\mathrm{Q}}(\,\mathbf{y}\,,\widehat{\mathbf{y}}_i\,;\,\mathbf{z})\ d\,\mathbf{y} = \sum_{i=1}^{N} p(\widehat{\mathbf{y}}_i|\mathbf{z})\cdot D_{\mathbf{z},i} \\
&= \sum_{i=1}^{N} 2^{-\frac{2B}{k_{\mathrm{q}}}}\cdot I(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i})\cdot p(\widehat{\mathbf{y}}_i|\mathbf{z})\cdot\lambda_{\mathbf{z},N}(\widehat{\mathbf{y}}_i)^{-\frac{2}{k_{\mathrm{q}}}}\cdot V(\mathbb{Q}_{\mathbf{z},i}) \\
&\approx \left(\int_{\mathbb{Q}_{\mathbf{z}}} I(\mathbf{y}\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z}}(\mathbf{y}))\cdot p(\mathbf{y}|\mathbf{z})\cdot\lambda_{\mathbf{z}}(\mathbf{y})^{-\frac{2}{k_{\mathrm{q}}}}\ d\,\mathbf{y}\right)\cdot 2^{-\frac{2B}{k_{\mathrm{q}}}} \ ,
\end{aligned} \tag{2.29}$$

where $\mathbb{E}_{\mathbf{z}}(\mathbf{y})$ denotes the asymptotic Voronoi shape of the cell that contains $\mathbf{y}$ when N approaches infinity.

Since the code points of the quantizer (or the codebook) do not depend on $\mathbf{z}$, it is shown in the following lemma that the asymptotic point density function $\lambda_{\mathbf{z}}(\mathbf{y})$ given by (2.28) does not depend on the side information $\mathbf{z}$.

Lemma 3 *If the side information* $\mathbf{z}$ *is only available at the encoder (or quantizer) but not at the decoder, the asymptotic point density function* $\lambda_{\mathbf{z}}(\widehat{\mathbf{y}})$ *does not depend on* $\mathbf{z}$*, i.e.*

$$\lambda_{\mathbf{z}}(\widehat{\mathbf{y}}) = \lambda(\widehat{\mathbf{y}}) \ , \tag{2.30}$$

assuming the distortion function D_Q has a continuous sensitivity matrix $\mathbf{W}_{\mathbf{z}}(\mathbf{y})$ w.r.t. source variable $\mathbf{y}$.

Proof: The original definition of point density function $\lambda_{\mathbf{z}}(\mathbf{y})$ is given by equation (2.27) and (2.28). In order to see the independence of the point density function on side information $\mathbf{z}$, let us consider an alternative definition $\lambda_{\text{alt}}(\mathbf{y})$, which is given by the following form

$$\lambda_{\text{alt}}\big(\widehat{\mathbf{y}}\big) = \lim_{V\big(\mathcal{R}(\widehat{\mathbf{y}})\big)\to 0} \frac{n\big(\widehat{\mathbf{y}}\big)/N}{V\big(\mathcal{R}(\widehat{\mathbf{y}})\big)} \ , \tag{2.31}$$

where $\mathcal{R}(\widehat{\mathbf{y}})$ is a small neighborhood region centered at $\widehat{\mathbf{y}}$, and $V\big(\mathcal{R}(\widehat{\mathbf{y}})\big)$ is the volume of this small region. $n\big(\widehat{\mathbf{y}}\big)$ is the number of code points in $\mathcal{R}(\widehat{\mathbf{y}})$, and $n\big(\widehat{\mathbf{y}}\big)/N$ is the relative frequency of the points in $\mathcal{R}(\widehat{\mathbf{y}})$. It is evident from the definition that $\lambda_{\text{alt}}\big(\widehat{\mathbf{y}}\big)$ does not depend on the side information $\mathbf{z}$, simply because $n\big(\widehat{\mathbf{y}}\big)$ is independent of $\mathbf{z}$. If the distortion function D_{Q} has a continuous sensitivity matrix, $\mathbf{W}_{\mathbf{z}}(\mathbf{y})$ can be regarded as constant matrix equal to $\mathbf{W}_{\mathbf{z}}(\widehat{\mathbf{y}})$ for source variables in the infinitesimal regions $\mathcal{R}(\widehat{\mathbf{y}})$ centered at $\widehat{\mathbf{y}}$ such that $V\big(\mathcal{R}(\widehat{\mathbf{y}})\big) \to 0$. Then, conditioned on a particular realization of $\mathbf{z}$, any projected quantization cells $\mathbb{Q}_{\mathbf{z},i}$ within region $\mathcal{R}(\widehat{\mathbf{y}})$ should be of the same shape and size[3]. Hence, one can rewrite the alternative point density definition (2.31) by the following form,

$$\lambda_{\text{alt}}\big(\widehat{\mathbf{y}}\big) = \lim_{V\big(\mathcal{R}(\widehat{\mathbf{y}})\big)\to 0} \frac{1}{N\cdot\Big(V\big(\mathcal{R}(\widehat{\mathbf{y}})\big)/n\left(\widehat{\mathbf{y}}\right)\Big)} = \lim_{N\to\infty} \frac{1}{N\cdot V\big(\mathbb{Q}_{\mathbf{z},i}\big)} \ , \tag{2.32}$$

which is equivalent to the original definition $\lambda\big(\widehat{\mathbf{y}}\big)$ given by (2.28). Therefore, by establishing the equivalence of the two definitions of the point density function, one can show that $\lambda_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$ is independent of $\mathbf{z}$. ■

Note that the independence between the point density and the side information proved in Lemma 3 does not necessarily mean that the encoder side information does not affect the quantizer performance. In fact, it plays an important role in determining the partitions (or tessellations) of the quantization space

[3]Due to the high-rate assumption, one can always find a sufficiently small (with infinitesimal volume) neighboring region $\mathcal{R}(\widehat{\mathbf{y}})$ that contains enough code points, so that the edge effects can be ignored.

$\mathbb{Q}_{\mathbf{z}}$. Hence the shapes of quantization cells $\mathbb{Q}_{\mathbf{z},\,i}$ corresponds to the same code point $\widehat{\mathbf{y}}_i$ but conditioned on different realizations of $\mathbf{z}$ are different and closely depend on $\mathbf{z}$. This represents the finer structure of the quantizer, whereas point density function is a coarser characteristic.

Hence, by the substituting (2.30) into (2.29), the overall distortion of a finite rate quantization system can be represented as

$$\begin{aligned} D = E_{\mathbf{z}}\left[\, D(\mathbf{z})\,\right] &= \int_{\mathbb{Z}} D(\mathbf{z}) \cdot p\,(\mathbf{z})\;\, d\,\mathbf{z} \\ &= \left(\iint_{(\mathbf{y},\mathbf{z})\,\in\,\mathbb{S}} I\big(\mathbf{y}\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z}}(\mathbf{y})\big) \cdot p\,(\mathbf{y},\,\mathbf{z}) \cdot \lambda\big(\mathbf{y}\big)^{-\frac{2}{k_{\mathrm{q}}}}\; d\,\mathbf{y}\, d\,\mathbf{z} \right) \cdot 2^{-\frac{2B}{k_{\mathrm{q}}}}. \end{aligned} \tag{2.33}$$

Note that the integration given by equation (2.33) can be viewed as an extension of Bennett's integral to a generalized fixed-rate vector quantization problem with additional encoder side information and general distortion metric function.

2.3 Minimization of the Distortion Integral & Different Distortion Bounds

Due to the new attribute of the encoder side information, the generalized vector quantization problem can be viewed as quantizing a multi-component source with different distortion functions. Therefore, the codebook should be designed to match the overall distortion averaged over all source components. A tight distortion lower bound is derived in this section to characterize the minimum system distortion achieved by the optimal quantizer. Furthermore, due to the unavailability of encoder side information at the decoder, a quantization scheme with multiple codebooks, designed to match each of source component, is not feasible. Hence, an alternative distortion lower bound is derived based on the distortion of this virtual multi-codebook quantization scheme. On the other hand, the generalized vector quantizer also benefits from the availability of the side information at the encoder. Therefore, the distortion of a side-information-aided quantizer is less than that of quantizing a mixed source with an average distortion function over the

components, which leads to a distortion upper bound. In the rest of this section, derivations of these distortion bounds are provided and related to corresponding classical fixed-rate quantization problems.

2.3.1 Asymptotic Distortion Lower Bound

The distortion integral (2.33) allows the minimization of the system distortion by optimizing the choice of the Voronoi shape $\mathbb{E}_{\mathbf{z}}(\mathbf{y})$ and the point density function $\lambda(\mathbf{y})$. By substituting the lower bound of the inertial profile (2.22) into equation (2.29), the following conditional distortion lower bound of the optimal quantizer can be obtained:

$$D(\mathbf{z}) \geq D_{\text{opt}}(\mathbf{z}) \overset{a}{\geq} \left(\int_{\mathbb{Q}_{\mathbf{z}}} I_{\text{opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big) \cdot p\big(\mathbf{y}\big|\mathbf{z}\big) \cdot \lambda_{\mathbf{z}}\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}}\, d\,\mathbf{y} \right) \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ . \tag{2.34}$$

Detailed discussion on the achievability of the above inequality is provided in Section 2.3.5. After some manipulations, the overall distortion of the finite rate quantization system can be represented by the following form

$$D_{\text{Opt}} = E_{\mathbf{z}}\Big[\, D_{\text{opt}}(\mathbf{z}) \,\Big] \geq \left(\int_{\mathbb{Q}} I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big) \cdot p\big(\mathbf{y}\big) \cdot \lambda\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}}\, d\,\mathbf{y} \right) \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.35}$$

where D_{Opt} represents the distortion of the optimal quantizer, and $I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big)$ is the average optimal inertial profile defined as

$$I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big) = \int_{\mathbb{Z}} I_{\text{opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big) \cdot p\big(\mathbf{z}\big|\mathbf{y}\big)\ d\,\mathbf{z}\ . \tag{2.36}$$

By utilizing the Holder's inequality, the optimal point density that minimizes the asymptotic distortion (2.35) is given by

$$\lambda^*\big(\mathbf{y}\big) = \Big(I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big) \cdot p\big(\mathbf{y}\big) \Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} \cdot \left(\int_{\mathbb{Q}} \Big(I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big) \cdot p\big(\mathbf{y}\big) \Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}}\, d\,\mathbf{y} \right)^{-1} \ . \tag{2.37}$$

By substituting (2.37) into (2.35), the asymptotic distortion of the optimal quantizer is lower bounded by the following form,

$$D_{\text{Opt}} \geq D_{\text{Low,1}} = \left(\int_{\mathbb{Q}} \Big(I^{\text{w}}_{\text{opt}}\big(\mathbf{y}\big) \cdot p\big(\mathbf{y}\big) \Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}}\, d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ . \tag{2.38}$$

Furthermore, by substituting the inertial profile lower bound (2.25) into the above equation, a tight lower bound on $D_{\text{Low},1}$ can be obtained:

$$D_{\text{Low},1} \gtrapprox \widetilde{D}_{\text{Low},1} = \left(\int_{\mathbb{Q}} \left(\widetilde{I}^{\text{w}}_{\text{opt}}(\mathbf{y}) \cdot p(\mathbf{y}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.39}$$

where $\gtrapprox$ means the lower bound is tight and $\widetilde{D}_{\text{Low},1}$ well approximates $D_{\text{Low},1}$, and $\widetilde{I}^{\text{w}}_{\text{opt}}(\mathbf{y})$ is given by

$$\widetilde{I}^{\text{w}}_{\text{opt}}(\mathbf{y}) = \int_{\mathbb{Z}} \widetilde{I}_{\text{opt}}(\mathbf{y}\,;\,\mathbf{z}) \cdot p(\mathbf{z}|\,\mathbf{y})\ d\,\mathbf{z}\ . \tag{2.40}$$

The above analysis results (equations (2.38) and (2.39)) can be viewed as an extension of the asymptotic distortion analysis provided in [34] to the generalized quantization problem with side information.

2.3.2 An Alternative Distortion Lower Bound

In some cases, the weighted inertial profile (2.36) is hard to obtain due to the intractability of the conditional probability $p(\mathbf{z}|\mathbf{y})$. In these situations, a second lower bound $D_{\text{Low},2}$ is proposed which is itself a lower bound on $D_{\text{Low},1}$. To be specific, the following lower bound on the conditional distortion $D_{\text{opt}}(\mathbf{z})$ is obtained from equation (2.34) without restricting the point density $\lambda_{\mathbf{z}}(\mathbf{y})$ to be independent of $\mathbf{z}$:

$$D_{\text{opt}}(\mathbf{z}) \geq D_{\text{Low},2}(\mathbf{z}) = \left(\int_{\mathbb{Q}} \left(I_{\text{opt}}(\mathbf{y}\,;\,\mathbf{z}) \cdot p(\mathbf{y}|\,\mathbf{z}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.41}$$

with equality if and only if $\lambda_{\mathbf{z}}(\mathbf{y})$ satisfies

$$\begin{aligned} &\lambda_{\mathbf{z}}(\mathbf{y}) = \lambda^{*}_{\mathbf{z}}(\mathbf{y}) \\ &= \left(I_{\text{opt}}(\mathbf{y}\,;\,\mathbf{z}) \cdot p(\mathbf{y}|\,\mathbf{z}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} \cdot \left(\int_{\mathbb{Q}} \left(I_{\text{opt}}(\mathbf{y}\,;\,\mathbf{z}) \cdot p(\mathbf{y}|\,\mathbf{z}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{-1} . \end{aligned} \tag{2.42}$$

Therefore, the overall asymptotic distortion is lower bounded as follows,

$$D_{\text{Opt}} \geq D_{\text{Low},1} \overset{a}{\geq} D_{\text{Low},2} = E_{\mathbf{z}}\Big[D_{\text{Low},2}(\mathbf{z}) \Big] = \int_{\mathbb{Z}} D_{\text{Low},2}(\mathbf{z}) \cdot p(\mathbf{z})\ d\,\mathbf{z}\ . \tag{2.43}$$

where (a) is due to the fact that inequality (2.41) is valid for any point density including the optimized point density (2.37). Again, by applying the hyper-ellipsoidal approximation on the Voronoi shapes, similar tight distortion lower bounds (or approximations) of $D_{\text{Low},2}\big(\mathbf{z}\big)$ and $D_{\text{Low},2}$ can be obtained, i.e.

$$D_{\text{Low},2}\big(\mathbf{z}\big) \gtrapprox \widetilde{D}_{\text{Low},2}\big(\mathbf{z}\big) = \left(\int_{\mathbb{Q}} \Big(\widetilde{I}_{\text{opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big) \cdot p\big(\mathbf{y}\big|\,\mathbf{z}\big) \Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} \, d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \quad (2.44)$$

and

$$D_{\text{Low},2} \gtrapprox \widetilde{D}_{\text{Low},2} = E_{\mathbf{z}}\Big[\widetilde{D}_{\text{Low},2}\Big] = \int_{\mathbb{Z}} \widetilde{D}_{\text{Low},2}\big(\mathbf{z}\big) \cdot p\big(\mathbf{z}\big) \ d\,\mathbf{z} \ . \quad (2.45)$$

An intuitive explanation of the above asymptotic distortion lower bound can be provided as below. First, each conditional distortion $D_{\text{Low},2}\big(\mathbf{z}\big)$ can be viewed as the minimal (or optimum) asymptotic distortion by quantizing a source $\mathbf{y}$ with distribution $p\big(\mathbf{y}\big|\,\mathbf{z}\big)$, and distortion function $D_{\text{Q}}\big(\mathbf{y},\,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)$. Each component quantizer (parameterized by $\mathbf{z}$) is assumed to have independent optimal Voronoi shapes and hence independent optimal inertial profile $I_{\text{opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big)$. In order to achieve the lower bound $D_{\text{Low},2}\big(\mathbf{z}\big)$, each component quantizer also uses different optimized point density $\lambda^*_{\mathbf{z}}(\mathbf{y})$. This means that multiple codebooks $\widehat{\mathbf{y}}_{\mathbf{z},1}, \widehat{\mathbf{y}}_{\mathbf{z},2}, \cdots, \widehat{\mathbf{y}}_{\mathbf{z},N}$ are utilized for each source component. This is equivalent to saying that the system distortion is lower bounded by quantizing a multi-component mixed source with the component index (or the side information) $\mathbf{z}$ available at both the encoder (or the quantizer) and the decoder. Therefore, $D_{\text{Low},2}$ can be viewed as the average minimum distortion of a class of finite rate quantizers with different source distributions and distortion functions.

2.3.3 Asymptotic Distortion Upper Bound

This subsection provides a distortion upper bound that can be used as an alternative distortion measurement of a finite rate quantization system. Suppose a sub-optimal quantizer is constructed by applying the same conditional partitions $\mathbb{Q}_{\mathbf{z},\,i} = \mathbb{Q}_i$ (or $\mathbb{E}_{\mathbf{z},\,i} = \mathbb{E}_i$) for different instantiations of $\mathbf{z}$, the asymptotic distortion

of the optimal quantizer is upper bounded by the following form

$$D_{\text{Opt}} \leq D_{\text{Upp}} = \left(\int_{\mathbb{Q}} I^{\text{w}}_{\text{Upp}}(\mathbf{y}) \cdot p(\mathbf{y}) \cdot \lambda(\mathbf{y})^{-\frac{2}{k_{\text{q}}}} \, d\,\mathbf{y} \right) \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.46}$$

where the sub-optimal average inertial profile $I^{\text{w}}_{\text{Upp}}(\mathbf{y})$ is given by

$$\begin{aligned} I^{\text{w}}_{\text{Upp}}\left(\widehat{\mathbf{y}}_i\right) &= \min_{\mathbb{E}_i \in \mathcal{H}_{\text{Q}}} \int_{\mathbb{Z}} I\left(\widehat{\mathbf{y}}_i \, ; \, \mathbf{z} \, ; \, \mathbb{E}_i\right) \cdot p\left(\mathbf{z} \middle| \, \widehat{\mathbf{y}}_i\right) \, d\,\mathbf{z} \\ &= \min_{\mathbb{E}_i \in \mathcal{H}_{\text{Q}}} \left(V\left(\mathbb{E}_i\right)^{-(1+\frac{2}{k_{\text{q}}})} \int_{\mathbf{e} \in \mathbb{E}_i} \mathbf{e}^{\mathsf{T}} \, \mathbf{W}^{\text{w}}\left(\widehat{\mathbf{y}}_i\right) \mathbf{e} \ d\,\mathbf{e} \right) \ , \end{aligned} \tag{2.47}$$

with the average sensitivity matrix $\mathbf{W}^{\text{w}}\left(\widehat{\mathbf{y}}_i\right)$ given by

$$\mathbf{W}^{\text{w}}\left(\widehat{\mathbf{y}}_i\right) = \int_{\mathbb{Z}} \mathbf{W}_{\mathbf{z}}\left(\widehat{\mathbf{y}}_i\right) \cdot p\left(\mathbf{z} \middle| \, \widehat{\mathbf{y}}_i\right) \, d\,\mathbf{z} \ . \tag{2.48}$$

By applying the same hyper-ellipsoidal approximation on the cell shape $\mathbb{E}_i$, similar tight inertia lower bound (or approximation) can be obtained

$$I^{\text{w}}_{\text{Upp}}\left(\widehat{\mathbf{y}}_i\right) \gtrapprox \widetilde{I}^{\text{w}}_{\text{Upp}}\left(\widehat{\mathbf{y}}_i\right) = \frac{k_{\text{q}}}{k_{\text{q}}+2} \cdot \left(\frac{\left| \mathbf{W}^{\text{w}}\left(\widehat{\mathbf{y}}_i\right) \right|}{\kappa^2_{k_{\text{q}}}} \right)^{1/k_{\text{q}}} \ . \tag{2.49}$$

Therefore, by utilizing Holder's inequality, the asymptotic distortion of an optimal quantizer can be upper bounded by the following form,

$$D_{\text{Opt}} \leq D_{\text{Upp}} = \left(\int_{\mathbb{Q}} \left(I^{\text{w}}_{\text{Upp}}(\mathbf{y}) \cdot p(\mathbf{y}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.50}$$

with the optimal point density that minimizes (2.50) given by

$$\lambda^*(\mathbf{y}) = \left(I^{\text{w}}_{\text{Upp}}(\mathbf{y}) \cdot p(\mathbf{y}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} \cdot \left(\int_{\mathbb{Q}} \left(I^{\text{w}}_{\text{Upp}}(\mathbf{y}) \cdot p(\mathbf{y}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{-1} \ . \tag{2.51}$$

Similarly, by substituting (2.49) into the distortion analysis (2.50), one can also obtain the tight lower bound[4] (or good approximation) of the asymptotic upper bound D_{Upp}, which is given by

$$D_{\text{Upp}} \gtrapprox \widetilde{D}_{\text{Upp}} = 2^{-\frac{2B}{k_{\text{q}}}} \cdot \left(\int_{\mathbb{Q}} \left(\widetilde{I}^{\text{w}}_{\text{Upp}}(\mathbf{y}) \cdot p(\mathbf{y}) \right)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} d\,\mathbf{y} \right)^{\frac{2+k_{\text{q}}}{k_{\text{q}}}} \ . \tag{2.52}$$

[4]Note that $\widetilde{D}_{\text{Upp}}$ is a lower bound on the distortion upper bound D_{Upp}. Hence, it may not even be a valid bound. However, since the lower bound $\widetilde{D}_{\text{Upp}}$ is very tight and well approximates D_{Upp}, it remains to be a valid upper bound of the optimal distortion D_{Opt} in most of the cases.

The asymptotic distortion upper bound D_{Upp} also has an intuitive connection to the traditional source coding problem. By using the same quantization region $\mathbb{E}_{\mathbf{z},i} = \mathbb{E}_i$ for different instantiations of $\mathbf{z}$, the side information $\mathbf{z}$ is completely ignored at the encoder (or the quantizer) and the quantization variable (or objective) $\mathbf{y}$ is quantized directly. Therefore, it can viewed as an equivalent problem of quantizing the source vector $\mathbf{y}$ with marginal distribution $p\big(\mathbf{y}\big)$, and a weighted distortion function D_{Q}^{w} given by

$$D_{\text{Q}}^{\text{w}}\big(\mathbf{y},\, \widehat{\mathbf{y}}\big) = \int_{\mathbb{Z}} D_{\text{Q}}\big(\mathbf{y},\, \widehat{\mathbf{y}}\,;\, \mathbf{z}\big) \cdot p\big(\mathbf{z}\big|\, \mathbf{y}\big)\; d\,\mathbf{z}\;. \tag{2.53}$$

By performing a second order Taylor series expansion on the weighted distortion function (2.53), it results in the weighted sensitivity matrix $\mathbf{W}^{\text{w}}\big(\mathbf{y}\big)$ given by (2.48). A classical asymptotic distortion analysis [34] of a source vector $\mathbf{y}$ (without side information) with pdf $p\big(\mathbf{y}\big)$ and distortion function $D_{\text{Q}}^{\text{w}}(\mathbf{y},\, \widehat{\mathbf{y}})$ will lead to the same distortion upper bound given by (2.50) and (2.52).

2.3.4 Losses Due in the Context of Side Information

Armed with the above-derived bounds and their corresponding interpretations, the performance loss for quantization with side information can be quantified. First, consider the loss due to ignorance of the side information at the decoder. As discussed above, the point density is constrained to be independent of the side information in this case, giving rise to a performance loss:

$$L_{\text{dec}} = \frac{\widetilde{D}_{\text{low},1}}{\widetilde{D}_{\text{low},2}}\;. \tag{2.54}$$

Next, consider the loss due to ignorance of the side information at both the encoder and decoder. In this case, the cell-shapes are constrained to be constant, and should be designed under an "averaged" distortion measure. The performance loss in this case is given by:

$$L_{\text{enc}} = \frac{\widetilde{D}_{\text{Upp}}}{\widetilde{D}_{\text{low},1}}\;. \tag{2.55}$$

This term represents the additional loss due *solely* to encoder ignorance, and so the total loss of a system with no access to the side information, relative to a system in which both the encoder and receiver have the side information, is given by $L_{\text{tot}} = L_{\text{enc}} \cdot L_{\text{dec}}$. Note that these loss functions specify the penalty in terms of excess distortion, and so the minimum loss is 1. The units can easily be converted into bits per dimension as $\frac{1}{2}\log_2(L)$.

2.3.5 Achievability of the Asymptotic Distortion Bounds

According to the connections of the distortion bounds provided in Section 2.3 and their related conventional fixed-rate quantization problems, the encoder side information plays an important role in determining the achievability of these bounds. In this part, strict achievability is first provided that corresponds to the case where the encoder side information is irrelevant to the encoding process. Hence there is no penalty of only knowing the side information at the encoder, and no advantages of knowing it at the decoder. Moreover, for large-dimensional source having factorable determinant of the sensitivity matrix, by utilizing a properly designed random codebook, sources can be quantized as if the side information is also available at the decoder. Detailed discussions of the achievability of the distortion bounds are provided in the rest of this section.

Strict Achievability of $D_{\mathbf{Low,1}}$ and $D_{\mathbf{Upp}}$

Due to the unavailability of the side information $\mathbf{z}$ at the decoder, a single codebook is used in a sense that the codebook, $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$, is fixed for any realizations of $\mathbf{z}$. Consequently, the projected quantization regions $\mathbb{Q}_{\mathbf{z},i}$ are different for different instantiations of $\mathbf{z}$ but depend on each other. To be specific, once the codebook is fixed, the Voronoi regions $\mathbb{Q}_{\mathbf{z},i}$ are determined by the distortion function $D_{\text{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big)$ or the sensitivity matrix $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)$. Therefore, the code points in general *cannot* be located in a way that the Voronoi region $\mathbb{Q}_{\mathbf{z},i}$ (or $\mathbb{E}_{\mathbf{z},i}$) is optimal, in a sense minimizing the inertial profile $I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big)$,

for all different realizations of $\mathbf{z}$. Hence, inequality (a) in (2.34) is strict and both equation (2.35) and (2.46) are strict asymptotic distortion lower and upper bounds respectively, if the optimal Voronoi regions $\mathbb{Q}_{\mathbf{z},i}$ for different $\mathbf{z}$ are not the same. A sufficient condition which guarantees the achievability of the distortion lower and upper bounds is given by

$$D_{\mathrm{Q}}\big(\mathbf{y},\, \widehat{\mathbf{y}}\,;\, \mathbf{z}\big) = f\big(\mathbf{z}\big) \cdot D_{\mathrm{Q}}\big(\mathbf{y},\, \widehat{\mathbf{y}}\big) \ . \tag{2.56}$$

For the asymptotic analysis provided in this chapter, where the distortion function is approximated by its second order Taylor series expansion, the sufficient condition can be reduced to the following form

$$\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big) = f\big(\mathbf{z}\big) \cdot \mathbf{W}\big(\widehat{\mathbf{y}}\big) \ . \tag{2.57}$$

In such cases, it is easy to prove that the Voronoi regions $\mathbb{Q}_{\mathbf{z},i}$ for different realizations of $\mathbf{z}$ are the same, and hence the optimal inertial profile $I_{\mathrm{opt}}\big(\mathbf{y}\,;\mathbf{z}\big)$ is achievable for every instantiations of $\mathbf{z}$ and also proportional to $f\big(\mathbf{z}\big)$, i.e.

$$I_{\mathrm{opt}}\big(\mathbf{y}\,;\mathbf{z}\big) = f\big(\mathbf{z}\big) \cdot I_{\mathrm{opt}}\big(\mathbf{y}\big) \ . \tag{2.58}$$

From another point of view, the side information $\mathbf{z}$ at the quantizer is irrelevant to the quantization process if the distortion function or the sensitivity matrix satisfies the product structure given by (2.56) or (2.57). It is therefore equivalent to quantizing vector $\mathbf{y}$ directly. The overall distortion in this situation is the average distortion of a mixed source with each component having the same optimal Voronoi shape and point density $\lambda(\mathbf{y})$, but with different conditional source distributions $p\big(\mathbf{y}\big|\,\mathbf{z}\big)$ and weighted distortion function $f(\mathbf{z}) \cdot D_{\mathrm{Q}}\big(\mathbf{y},\, \widehat{\mathbf{y}}\big)$.

Strict Achievability of $D_{\mathbf{Low,2}}$

If the distortion function satisfies factorable condition (2.57) and the source vector $\mathbf{y}$ and encoder side information $\mathbf{z}$ are further statistically independent from each other, it can be easily proved that the optimal point density $\lambda^*_{\mathbf{z}}(\mathbf{y})$

given by (2.42) does not depend on side information $\mathbf{z}$. In this case, all proposed distortion bounds are achievable, i.e.

$$D_{\text{Opt}} = D_{\text{Upp}} = D_{\text{Low},1} = D_{\text{Low},2} \ , \tag{2.59}$$

and the system asymptotic distortion can be viewed as the distortion of a mixed source with marginal distribution $p\big(\mathbf{y}\big)$ and weighted distortion function $D_{\text{Q}}^{\text{w}}\big(\mathbf{y}, \widehat{\mathbf{y}}\big)$ given by

$$D_{\text{Q}}^{\text{w}}\big(\mathbf{y}, \widehat{\mathbf{y}}\big) = \bigg(\int_{\mathbb{Z}} f\big(\mathbf{z}\big) \ d\,\mathbf{z} \bigg) \cdot D_{\text{Q}}\big(\mathbf{y}, \widehat{\mathbf{y}}\big) \ .$$

Asymptotic Achievabilities of Sources With Large Dimensions

Interestingly, for cases where the distortion function or sensitivity matrix does not satisfy the factorable condition (2.56) or (2.57), the distortion bounds $D_{\text{Low},1}$ and D_{Upp} provided in previous sections are also tight for sources with large dimensions and under high resolutions. To see this, consider a sub-optimal encoder that employs a random codebook [42] [43], code vectors $\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \cdots, \widehat{\mathbf{y}}_N$ generated independently from a known pdf, denoted by $p_{\text{c}}(\mathbf{y})$. Following the same derivations provided in [42] [43], and the average value (expectation over the random codebook) of the asymptotic distortion of the generalized vector quantizer with random codebook can be represented as the following form

$$D_{\text{Rand}} = \Big(1 + \frac{2}{k_{\text{q}}}\Big)\, \Gamma\Big(1 + \frac{2}{k_{\text{q}}}\Big) \cdot \bigg(\int_{\mathbb{Q}} I_{\text{opt}}^{\text{w}}\big(\mathbf{y}\big) \cdot p\big(\mathbf{y}\big) \cdot p_{\text{c}}\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}} \ d\,\mathbf{y} \bigg) \cdot 2^{-\frac{2B}{k_{\text{q}}}} \ , \tag{2.60}$$

where $\Gamma(\cdot)$ is the Gamma function. Optimizing the right-hand side of equation (2.60) with respect to $p_{\text{c}}(\mathbf{y})$ will lead to the same optimal point density given by (2.37), i.e.

$$p_{\text{c}}(\mathbf{y}) = \lambda^*(\mathbf{y}) \ , \tag{2.61}$$

and the optimal performance of a random codebook quantizer is given by

$$D_{\text{opt-Rand}} = \eta(k_{\text{q}}) \cdot D_{\text{opt}} \, , \qquad\qquad \eta(k_{\text{q}}) = \Big(1 + \frac{2}{k_{\text{q}}}\Big)\, \Gamma\Big(1 + \frac{2}{k_{\text{q}}}\Big) \ . \tag{2.62}$$

It can be observed from equation (2.62) that the ratio $\eta(k_{\text{q}})$ between the distortions of the random codebook and the minimal system distortion approaches 1 as k_{q}

increases. This means that the distortion lower bound $D_{\text{Low},1}$ can be achieved by a quantizer with random codebook for source vectors with large dimensions even though the sensitivity matrix $\mathbf{W}_{\mathbf{z}}(\mathbf{y})$ is not factorable.

Furthermore, if the following product is factorable

$$I_{\text{opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big)\cdot p\big(\mathbf{y}\big|\,\mathbf{z}\big) = g_1(\mathbf{y})\cdot g_2(\mathbf{z})\ \ , \tag{2.63}$$

it can be shown after straightforward manipulations that optimal point density $\lambda^*_{\mathbf{z}}(\mathbf{y})$ given by equation (2.42) does not depend on the side information $\mathbf{z}$. Hence, the distortion lower bound $D_{\text{Low},2}$ is achievable. Moreover, as a direct result of the above condition, if the distortion lower bound $\widetilde{D}_{\text{Low},2}$ is considered and vector $\mathbf{y}$ and $\mathbf{z}$ are independent, the achievability condition further reduces to the factorablity of the determinant of the sensitivity matrix $\mathbf{W}_{\mathbf{z}}(\mathbf{y})$, given by

$$\big|\,\mathbf{W}_{\mathbf{z}}(\mathbf{y})\,\big| = g_1(\mathbf{y})\cdot g_2(\mathbf{z})\ \ , \tag{2.64}$$

which is weaker than the matrix factorable condition given by (2.57).

2.4 Distortion Analysis of Mismatched Quantizers

In the previous section and in past work, the analytical results were derived under the assumption that both the encoder and the decoder have perfect knowledge of the source distribution, distortion function, and are using the most efficient quantization algorithm. This is clearly not always true as practical constraints often result in approximations and various types of suboptimal choices in the design of feedback-based wireless communication systems. These suboptimal choices often result in various types of mismatches. In this subsection, asymptotic analysis of mismatched quantizers is provided for the following three different categories: dimensionality mismatch, distortion function mismatch and source distribution mismatch.

2.4.1 Dimensionality Mismatch

The optimal quantizer is designed to quantize the source variable $\mathbf{y}$ (or objective) with the minimal free dimensions k_{q} or $k'_{\mathrm{q}} = k_{\mathrm{q}} - k_{\mathrm{c}}$ in the case of constraint source. Dimensionality mismatch occurs when the sub-optimal quantizer is designed to quantize a redundant source variable $\mathbf{y}_{\mathrm{R}}$. As an example, for the MISO problem one may quantize directly the channel vector $\mathbf{h}$ instead of the directional vector $\mathbf{v} = \mathbf{h}/\|\mathbf{h}\|$. Hence, vector quantization is carried out in a space with dimension $k_{\text{R-q}}$ ($k_{\text{R-q}} > k_{\mathrm{q}}$), and the distortion function D_{Q} is represented in its redundant form $D_{\text{Q-R}}(\mathbf{y}_{\mathrm{R}}, \widehat{\mathbf{y}}_{\mathrm{R}}\,;\, \mathbf{z})$. In this case, by following the methodology provided in [44] [45], the final asymptotic analysis of the mismatched quantizer can be obtained which has a form similar to the lower bounds give by (2.39) and (2.45), i.e.

$$D_{\text{mis-R-Low}} = c \cdot 2^{-\frac{2B}{k_{\text{R-q}}}}\ , \tag{2.65}$$

where c is a constant coefficient that depends on $k_{\text{R-q}}$, distortion function $D_{\text{Q-R}}$, and the source distribution $p(\mathbf{y}_{\mathrm{R}}, \mathbf{z})$. An important and general observation from equation (2.65) is that by quantizing the redundant source variable $\mathbf{y}_{\mathrm{R}}$, the system asymptotic distortion will have a smaller exponential slope ($-2/k_{\text{R-q}}$) when compared to that of quantizing the minimal free-dimensional vector $\mathbf{y}$ with exponential distortion slope ($-2/k_{\mathrm{q}}$).

2.4.2 Distortion Function Mismatch

In some cases, the quantizer (or the codebook) is designed or trained by using a distortion measure $D_{\text{mis-Q}}$ that is different from the actual system distortion function D_{Q}. An example of such a situation in practice is when the approximated distortion function $D_{\text{mis-Q}}$ leads to simple and efficient quantization schemes and codebook design algorithms [38]. More specifically for the MISO problem, one can envision designing a quantizer based on an SNR maximization criteria for simplicity and evaluating it using the capacity loss criteria. We provide in this subsection

an asymptotic analysis of the general vector quantizer with mismatched distortion function.

The distortion of interest is denoted by D_{Q} and the distortion function used for designing the quantizer is denoted by $D_{\text{mis-Q}}$. Since $D_{\text{mis-Q}}$ is the basis of the quantizer, it determines the Voronoi region and the point density function. A parameter of interest in this context is the sensitivity matrix of the mismatched distortion function $D_{\text{mis-Q}}$ which is denoted by $\mathbf{W}_{\text{mis},\,\mathbf{z}}(\mathbf{y})$ and is the Hessian matrix of $D_{\text{mis-Q}}$ w.r.t. vector $\mathbf{y}$. Codebook generated or trained by the mismatched sensitivity matrix leads to a mismatched Voronoi region $\mathbb{E}_{\text{mis},\mathbf{z}}(\mathbf{y})$, which can be approximated by a hyper ellipsoid $\mathcal{T}\big(\mathbf{0},\, \mathbf{W}_{\text{mis},\,\mathbf{z}}(\widehat{\mathbf{y}}_i),\, V(\mathbb{E}_{\text{mis},\mathbf{z}})(\mathbf{y})\big)$ with its definition given by equation (2.24), where $V\big(\mathbb{E}_{\text{mis},\mathbf{z}}(\mathbf{y})\big)$ is the volume of the mismatched Voronoi region. Since the quantizer is evaluated using the true distortion function D_{Q}, by substituting the approximated $\mathbb{E}_{\text{mis},\mathbf{z}}(\mathbf{y})$ into (2.17), the mismatched inertial profile utilizing the sub-optimal codebook can be closely approximated by

$$I_{\text{mis-D}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \approx \widetilde{I}_{\text{mis-D}}\left(\mathbf{y}\,;\,\mathbf{z}\right) = V\big(\mathbb{E}_{\text{mis},\mathbf{z}}(\mathbf{y})\big)^{-\frac{2+k_{\text{q}}}{k_{\text{q}}}} \cdot \int_{\mathbf{y}' \in \mathcal{T}\left(\mathbf{0},\, \mathbf{W}_{\text{mis},\,\mathbf{z}}(\mathbf{y}),\, V(\mathbb{E}_{\text{mis},\mathbf{z}}(\mathbf{y}))\right)} (\mathbf{y}'-\mathbf{y})^{\mathsf{T}}\, \mathbf{W}_{\mathbf{z}}\big(\,\mathbf{y}\,\big)\, (\mathbf{y}'-\mathbf{y})\;\, d\,\mathbf{y}'\;\;, \tag{2.66}$$

Following the multi-dimensional integration approach provided in [34] [38], the mismatched inertial profile $\widetilde{I}_{\text{mis-D}}\left(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\right)$ can be shown to be given by the following closed form expression

$$\widetilde{I}_{\text{mis-D}}\left(\mathbf{y}\,;\,\mathbf{z}\right) = \frac{1}{k_{\text{q}}+2}\left(\frac{\big|\mathbf{W}_{\text{mis},\,\mathbf{z}}\big(\mathbf{y}\big)\big|}{\kappa_{k_{\text{q}}}^{2}}\right)^{\frac{1}{k_{\text{q}}}} \mathbf{tr}\Big(\mathbf{W}_{\text{mis},\,\mathbf{z}}^{-1}\big(\mathbf{y}\big)\, \mathbf{W}_{\mathbf{z}}\big(\mathbf{y}\big)\Big) \geq \widetilde{I}_{\text{opt}}\left(\mathbf{y}\,;\,\mathbf{z}\right). \tag{2.67}$$

Consequently, the average mismatched inertial profile $\widetilde{I}^{\text{w}}_{\text{mis-D}}\big(\mathbf{y}\big)$ can be represented as

$$\widetilde{I}^{\text{w}}_{\text{mis-D}}\big(\mathbf{y}\big) = \int_{\mathbb{Z}} \widetilde{I}_{\text{mis-D}}\big(\mathbf{y}\,;\,\mathbf{z}\big) \cdot p\big(\mathbf{z}\big|\,\mathbf{y}\big)\;\, d\,\mathbf{z}\;. \tag{2.68}$$

In addition, the mismatched sensitivity matrix also leads to a mismatched point density function having the following form, from (2.37)

$$\lambda_{\text{mis-D}}\big(\mathbf{y}\big) = \Big(\widetilde{I}^{\text{w}}_{\text{opt-mis}}\big(\mathbf{y}\big) \cdot p\,\big(\mathbf{y}\big)\Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}} \cdot \bigg(\int_{\mathbb{Q}} \Big(\widetilde{I}^{\text{w}}_{\text{opt-mis}}\big(\mathbf{y}\big) \cdot p\,\big(\mathbf{y}\big)\Big)^{\frac{k_{\text{q}}}{2+k_{\text{q}}}}\, d\,\mathbf{y}\bigg)^{-1}\;\;, \tag{2.69}$$

where $\widetilde{I}^{\mathrm{w}}_{\text{opt-mis}}(\mathbf{y})$ is the optimal average inertia profile of a system with actual distortion function equal to $D_{\text{mis-Q}}$. Finally, by substituting the above mismatched average inertial profile (2.68) and mismatched point density (2.69) into the distortion lower bound $D_{\text{Low},1}$ ($\widetilde{D}_{\text{Low},1}$) given by (2.39), the average distortion of a quantizer with mismatched distortion function can be obtained as:

$$\widetilde{D}_{\text{mis-D-Low},1} = 2^{-\frac{2B}{k_{\mathrm{q}}}} \cdot \int_{\mathbb{Q}} \widetilde{I}^{\mathrm{w}}_{\text{mis-D}}(\mathbf{y}) \cdot p(\mathbf{y}) \cdot \widetilde{\lambda}_{\text{mis-D}}(\mathbf{y})^{-\frac{2}{k_{\mathrm{q}}}} \, d\,\mathbf{y} \ , \tag{2.70}$$

Utilizing a similar approach, other mismatched distortion analysis, such as distortion lower bound $\widetilde{D}_{\text{mis-D-Low},2}$, can also be obtained.

2.4.3 Source Distribution Mismatch

It is evident that the optimal quantizer (or the optimal codebook) is designed to match not only the distortion function D_{Q} but also the underlying source distribution $p(\mathbf{y}, \mathbf{z})$. In situations where the source distribution is hard to obtain or is subject to errors, the performance of the quantized system will degrade with the use of the sub-optimal codebook generated using the mismatched source distribution, which is denoted as $p_{\text{mis}}(\mathbf{y}, \mathbf{z})$. As an example, for the MISO problem one may use a codebook designed assuming i.i.d. channels for correlated channels. The mismatched source distribution results in a mismatched average inertial profile, which is given by

$$\widetilde{I}^{\mathrm{w}}_{\text{mis-P}}(\mathbf{y}) = \int_{\mathbb{Z}} \widetilde{I}_{\text{opt}}(\mathbf{y}\,;\,\mathbf{z}) \cdot p_{\text{mis}}(\mathbf{z}|\,\mathbf{y}) \ d\,\mathbf{z} \ . \tag{2.71}$$

The mismatched average inertial profile $\widetilde{I}^{\mathrm{w}}_{\text{mis-P}}(\mathbf{y})$ together with $p_{\text{mis}}(\mathbf{y})$ further lead to a mismatched point density function $\lambda_{\text{mis-P}}(\mathbf{y})$,

$$\lambda_{\text{mis-P}}(\mathbf{y}) = \left(\widetilde{I}^{\mathrm{w}}_{\text{mis-P}}(\mathbf{y}) \cdot p_{\text{mis}}(\mathbf{y})\right)^{\frac{k_{\mathrm{q}}}{2+k_{\mathrm{q}}}} \cdot \left(\int_{\mathbb{Q}} \left(\widetilde{I}^{\mathrm{w}}_{\text{mis-P}}(\mathbf{y}) \cdot p_{\text{mis}}(\mathbf{y})\right)^{\frac{k_{\mathrm{q}}}{2+k_{\mathrm{q}}}} d\,\mathbf{y}\right)^{-1}, \tag{2.72}$$

which is not optimized to match the actual source distribution as compared to the optimal point density function given by (2.37). Therefore, the asymptotic

distortion of a sub-optimal quantizer with mismatched source distribution is given by

$$\widetilde{D}_{\text{mis-P-Low,1}} = 2^{-\frac{2B}{k_{\text{q}}}} \cdot \int_{\mathbb{Q}} \widetilde{I}_{\text{opt}}^{\text{w}}(\mathbf{y}) \cdot p(\mathbf{y}) \cdot \widetilde{\lambda}_{\text{mis-P}}(\mathbf{y})^{-\frac{2}{k_{\text{q}}}} \, d\,\mathbf{y} \ . \tag{2.73}$$

Again, other asymptotic distortion bounds due to the source distribution mismatch, such as $\widetilde{D}_{\text{mis-P-Low,2}}$ can also be obtained.

In summary, the mismatched analysis provided in this section shows that the system performance degradation (or the distortion increment) due to the mismatch in the distortion function as well the source distribution only impacts the coefficient in front of the exponential term $2^{-2B/k_{\text{q}}}$. However, the dimensionality mismatch caused by quantizing a redundant source vector $\mathbf{y}_{\text{R}}$ has a more significant effect on the system performance. It reduces the slope of the exponential components in equations (2.39) and (2.45), and hence leads to a larger distortion $2^{-2B/k_{\text{R-q}}}$ $\left(2^{-2B/k_{\text{R-q}}} \gg 2^{-2B/k_{\text{q}}}\right)$ than that of an optimal quantizer especially in the high resolution regimes.

2.5 Distortion Analysis of Quantizers with Transformed Codebook

In certain situations, the underlying source distribution $p(\mathbf{y}, \mathbf{z})$ or the distortion function D_{Q} of the source variable varies during the quantization process. It is practically infeasible to design separate codebooks optimized for every different source distribution and distortion function, or the encoder and the decoder may not have the ability to store a large number of codebooks. In these situations, it is convenient to use a quantizer whose codebook is constructed by a transformation of a given codebook, potentially optimum for a particular set of statistical conditions, to best match the statistical environment at hand. This type of quantizers are generally called transformed quantizers [35] [46], and have been used in conventional source coding area with a linear orthogonal transformation followed by a product quantizer. We provide in this section an analysis of the gen-

eralized vector quantizer, which is described in Section 2.2, when a transformed codebook is used. Detailed applications to finite-rate feedback MISO systems with transformed codebook over correlated fading channels are provided in Chap. 4.

2.5.1 Problem Formulation

It is first assumed that all the codebooks are generated from one fixed codebook $\mathcal{C}_0$ which is designed to match source distribution $p_0(\mathbf{y},\, \mathbf{z})$, and distortion function $D_{0,\,\mathrm{Q}}$ with sensitivity matrix $\mathbf{W}_{0,\,\mathbf{z}}(\mathbf{y})$. Codebook $\mathcal{C}_0$ has a point density given by $\lambda_0(\mathbf{y})$, and a normalized inertial profile $I_0\big(\mathbf{y}; \mathbf{z}; \mathbb{E}_{0,\mathbf{z}}(\mathbf{y})\big)$ that is optimized to matches the distortion function $D_{0,\,\mathrm{Q}}$, with $\mathbb{E}_{0,\mathbf{z}}(\mathbf{y})$ representing the asymptotic Voronoi cell that contains $\mathbf{y}$ with side information $\mathbf{z}$. Let the source distribution change from $p_0(\mathbf{y},\, \mathbf{z})$ to $p(\mathbf{y},\, \mathbf{z})$ and the distortion function become D_{Q} instead of $D_{0,\,\mathrm{Q}}$ with sensitivity matrix $\mathbf{W}_{\mathbf{z}}(\mathbf{y})$ instead of $\mathbf{W}_{0,\,\mathbf{z}}(\mathbf{y})$. Then the encoder and decoder will correspondingly adopt a transformed codebook $\mathcal{C}$ obtained from $\mathcal{C}_0$ by a general one-to-one mapping $\mathbf{F}(\cdot)$ with both of its domain and codomain in space $\mathbb{Q}$, i.e.

$$\mathcal{C} = \Big\{ \mathbf{F}(\widehat{\mathbf{y}}) \;\Big|\; \widehat{\mathbf{y}} \in \mathcal{C}_0 \Big\} \,. \tag{2.74}$$

2.5.2 Sub-optimal Point Density & Sub-optimal Voronoi Shape

Assuming the codebook transformation function $\mathbf{F}(\cdot)$ has continuous first order derivative, two types of sub-optimality arise when the transformed quantizer is used. One comes from the sub-optimal point density $\lambda_{\mathrm{tr}}(\mathbf{y})$, which can be derived from $\lambda_0(\mathbf{y})$ by the following transformation

$$\lambda_{\mathrm{tr}}(\mathbf{y}) = \frac{\lambda_0\left(\mathbf{F}^{-1}(\mathbf{y})\right)}{\left|\mathbf{F}_{\mathrm{d}}\big(\mathbf{F}^{-1}(\mathbf{y})\big)\right|}\,, \qquad \mathbf{F}_{\mathrm{d}}(\mathbf{y}) = \frac{\partial\, \mathbf{F}(\mathbf{y})}{\partial\, \mathbf{y}} \;. \tag{2.75}$$

If the source variable is subject to k_{c} constraints given by the vector equation $\mathbf{g}(\mathbf{y}) = \mathbf{0}$, the transformed point density is given by

$$\lambda_{\text{c-tr}}(\mathbf{y}) = \frac{\lambda_0\left(\mathbf{F}^{-1}(\mathbf{y})\right)}{\Big|\mathbf{V}_2(\mathbf{y})^{\mathsf{T}} \cdot \mathbf{F}_{\mathrm{d}}\big(\mathbf{F}^{-1}(\mathbf{y})\big) \cdot \mathbf{V}_2\big(\mathbf{F}^{-1}(\mathbf{y})\big)\Big|} \;, \tag{2.76}$$

where $\mathbf{V}_2(\mathbf{y})$ is an orthonormal matrix with its columns constituting an orthonormal basis of the null space $\mathcal{N}\left(\frac{\partial}{\partial \mathbf{y}}\,\mathbf{g}(\mathbf{y})\right)$. Compared to the optimal point density $\lambda^*(\mathbf{y})$ given by equation (2.37) that matches to $p(\mathbf{y},\,\mathbf{z})$, $\lambda_{\rm tr}(\mathbf{y})$ given by equation (2.75) is always sub-optimal and hence results in performance degradation. The other sub-optimality arises from the fixed location of the code points in the transformed codebook $\mathcal{C}$, in the sense that the Voronoi shape of the transformed code cannot match to the distortion function $D_{\rm Q}$ and hence is not optimized to minimize the inertial profile. Note that these two sub-optimalities, named as point density loss and cell shape loss, were also discussed in [35] in the setting of the conventional product quantizers and further applied to study the distortion performance of quantizers with transformed codebooks.

2.5.3 Characterizing the Inertial Profile of the Transformed Codebook

Unfortunately, the Voronoi region $\mathbb{E}_{{\rm tr},\mathbf{z}}\big(\widehat{\mathbf{y}}'_i\big)$ of code points $\widehat{\mathbf{y}}'_i$ in the transformed codebook $\mathcal{C}$, which is defined to be

$$\mathbb{E}_{{\rm tr},\mathbf{z}}\big(\widehat{\mathbf{y}}'_i\big) \stackrel{\triangle}{=} \left\{ \mathbf{y} \;\middle|\; D_{\rm Q}\big(\mathbf{y},\,\widehat{\mathbf{y}}'_i\,;\,\mathbf{z}\big) \le D_{\rm Q}\big(\mathbf{y},\,\widehat{\mathbf{y}}'_j\,;\,\mathbf{z}\big)\,, \quad \forall\ \widehat{\mathbf{y}}_j \in \mathcal{C}\ \ \&\ \ \widehat{\mathbf{y}}'_j \neq \widehat{\mathbf{y}}'_i \right\}, \tag{2.77}$$

is hard to characterize and depends both on the transformation $\mathbf{F}$ as well as the distortion function $D_{\rm Q}$. In order to characterize the effects of the transformed Voronoi shape on the system distortion, lower and upper bounds of the normalized inertial profile of the transformed code are provided. First, let us consider a sub-optimal quantizer $\mathcal{Q}_{\rm sub}(\cdot)$ with transformed codebook $\mathcal{C}$ but using a sub-optimal encoding process, given by

$$\widehat{\mathbf{y}} = \mathcal{Q}_{\rm sub}\big(\mathbf{y},\mathbf{z}\big) = \mathbf{F}\Big(\,\mathcal{Q}\Big(\mathbf{F}^{-1}\big(\mathbf{y},\mathbf{z}\big)\Big)\Big)\ , \tag{2.78}$$

where $\mathcal{Q}(\cdot)$ is the optimal encoder that matches to distortion function $D_{0,\,\rm Q}$. This sub-optimal encoder can be viewed as an extension of the "companding" model introduced by Bennett in [32] to the general vector quantization problem. It was

originally used in conventional scalar quantizers, where the encoder is a combination of a monotonically increasing nonlinear mapping $E(x)$, the compressor, followed by a uniform quantizer; and the corresponding decoder is composed of a uniform decoder followed by an inverse mapping E^{-1}, the expander. In the case of the generalized vector quantizer discussed here, the Voronoi shape of the sub-optimal transformed encoder $\mathcal{Q}_{\text{sub}}$ can be analytically characterized as

$$\mathbb{E}_{\text{sub},\mathbf{z}}\big(\mathbf{F}(\mathbf{y})\big) = \Big\{ \mathbf{F}(\mathbf{y}') \;\Big|\; \mathbf{y}' \in \mathbb{E}_{0,\mathbf{z}}(\mathbf{y}) \Big\} \;, \tag{2.79}$$

where $\mathbb{E}_{0,\mathbf{z}}(\mathbf{y})$ is the optimal Voronoi shape of the original codebook $\mathcal{C}_0$ corresponding to distortion function $D_{0,\,\text{Q}}$. Due to the sub-optimality of encoder $\mathcal{Q}_{\text{sub}}$, the normalized inertial profile of the transformed Voronoi shape $\mathbb{E}_{\text{tr},\mathbf{z}}(\mathbf{y})$ is upper bounded by the inertial profile of $\mathbb{E}_{\text{sub},\mathbf{z}}(\mathbf{y})$ given by (2.79), but lower bounded the inertial profile of the optimal Voronoi shape $\mathbf{E}_{\mathbf{z}}(\mathbf{y})$ corresponding to distortion function D_{Q}.

Proposition 1 *Under high resolution assumptions, the approximated inertial profile* $\widetilde{I}_{tr}\big(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\big)$ *of a quantizer with transformed codebook can be upper and lower bounded by the following form,*

$$\begin{aligned}
&\frac{k_q}{k_q+2}\cdot\left(\frac{\big|\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\mathbf{y})\big)\big|}{\kappa_{k_q}^2}\right)^{\frac{1}{k_q}} = \widetilde{I}_{opt}\big(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\big) \overset{a}{\leq} \widetilde{I}_{tr}\big(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\big) \overset{b}{\leq} \widetilde{I}_{sub}\big(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\big) \\
&= \frac{\big|\mathbf{F}_d(\mathbf{y})\big|^{-\frac{2}{k_q}}}{k_q+2}\left(\frac{\big|\mathbf{W}_{0,\,\mathbf{z}}\big(\mathbf{y}\big)\big|}{\kappa_{k_q}^2}\right)^{\frac{1}{k_q}} \mathbf{tr}\Big(\mathbf{W}_{0,\,\mathbf{z}}(\mathbf{y})^{-1}\cdot\mathbf{F}_d(\mathbf{y})^{\mathsf{T}}\cdot\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\mathbf{y})\big)\cdot\mathbf{F}_d(\mathbf{y})\Big).
\end{aligned} \tag{2.80}$$

Furthermore, if the source variable is subject to k_c constraints given by the vector equation $\mathbf{g}(\mathbf{y}) = \mathbf{0}$, the constrained inertial profile $\widetilde{I}_{c\text{-}tr}\big(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\big)$ *can be similarly*

bounded by

$$\begin{aligned}
&\frac{k'_q}{k'_q+2}\cdot\left(\frac{\left|\mathbf{V}_2\big(\mathbf{F}(\mathbf{y})\big)^{\mathsf{T}}\cdot\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\mathbf{y})\big)\cdot\mathbf{V}_2\big(\mathbf{F}(\mathbf{y})\big)\right|}{\kappa^2_{k_{q'}}}\right)^{\frac{1}{k_{q'}}}\\
&= \widetilde{I}_{c\text{-}opt}\left(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\right)\overset{a}{\leq}\ \widetilde{I}_{c\text{-}tr}\left(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\right)\ \overset{b}{\leq}\ \widetilde{I}_{c\text{-}sub}\left(\mathbf{F}(\mathbf{y})\,;\,\mathbf{z}\right)\\
&= \frac{\left|\mathbf{V}_2\big(\mathbf{F}(\mathbf{y})\big)^{\mathsf{T}}\cdot\mathbf{F}_d(\mathbf{y})\cdot\mathbf{V}_2(\mathbf{y})\right|^{-\frac{2}{k'_q}}}{k'_q+2}\left(\frac{\left|\mathbf{V}_2(\mathbf{y})^{\mathsf{T}}\cdot\mathbf{W}_{0,\mathbf{z}}(\mathbf{y})\cdot\mathbf{V}_2(\mathbf{y})\right|}{\kappa^2_{k'_q}}\right)^{\frac{1}{k'_q}}\\
&\cdot\mathbf{tr}\left(\Big(\mathbf{V}_2(\mathbf{y})^{\mathsf{T}}\mathbf{W}_{0,\mathbf{z}}(\mathbf{y})\mathbf{V}_2(\mathbf{y})\Big)^{-1}\cdot\mathbf{V}_2(\mathbf{y})^{\mathsf{T}}\mathbf{F}_d(\mathbf{y})^{\mathsf{T}}\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\mathbf{y})\big)\,\mathbf{F}_d(\mathbf{y})\mathbf{V}_2(\mathbf{y})\right),
\end{aligned}\tag{2.81}$$

where $\mathbf{V}_2(\mathbf{y})$ *is an orthonormal matrix with its columns constituting an orthonormal basis of the null space* $\mathcal{N}\left(\frac{\partial}{\partial\mathbf{y}}\,\mathbf{g}(\mathbf{y})\right)$.

Proof: Due to the fixed location of the code points in the transformed codebook $\mathcal{C}$, which can not be optimized to minimize the normalized inertial profile, it is evident that the transformed inertial profile $\widetilde{I}_{\rm tr}$ is lower bounded by the optimal inertial profile $\widetilde{I}_{\rm opt}$ given by equation (2.25). Hence, inequality (a) in (2.80) can be obtained after some manipulations. The same reasonings are valid for inequality (a) in (2.81) for the constraint source.

As for inequality (b) in (2.80), since function $\mathbf{F}(\cdot)$ is first order continuous, any points in the vicinity of the transformed code point $\mathbf{F}(\widehat{\mathbf{y}})$ can be first order Taylor series expanded as

$$\mathbf{F}(\mathbf{y})\approx\mathbf{F}(\widehat{\mathbf{y}})+\mathbf{F}_{\rm d}(\widehat{\mathbf{y}})\cdot(\mathbf{y}-\widehat{\mathbf{y}})\,,\qquad \mathbf{F}_{\rm d}(\widehat{\mathbf{y}})=\left.\frac{\partial}{\partial\,\mathbf{y}}\right|_{\mathbf{y}=\widehat{\mathbf{y}}}\mathbf{F}(\mathbf{y})\ .\tag{2.82}$$

Moreover, due to the fact the $\mathbf{F}(\cdot)$ is a one-to-one mapping, for any point $\mathbf{y}'$ in the vicinity of $\mathbf{F}(\widehat{\mathbf{y}})$, there exists a unique point $\mathbf{y}$ in the neighborhood of $\widehat{\mathbf{y}}$ such that $\mathbf{y}'=\mathbf{F}(\mathbf{y})$. Therefore, under high resolutions, the distortion function $D_{\rm Q}$ can be expanded around point $\mathbf{F}(\widehat{\mathbf{y}})$ by the following form

$$\begin{aligned}
D_{\rm Q}\big(\mathbf{y}',\,\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\big)&\approx\big(\mathbf{y}'-\mathbf{F}(\widehat{\mathbf{y}})\big)^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\,\big(\mathbf{y}'-\mathbf{F}(\widehat{\mathbf{y}})\big)\\
&\approx(\mathbf{y}-\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\Big(\mathbf{F}_{\rm d}(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\cdot\mathbf{F}_{\rm d}(\widehat{\mathbf{y}})\Big)\cdot(\mathbf{y}-\widehat{\mathbf{y}})\ ,
\end{aligned}\tag{2.83}$$

which has quadratic format but with transformed sensitivity matrix. By substituting equation (2.83) as well as the Voronoi shape of the sub-optimal encoder given by equation (2.79) into the definition of the inertial profile given by (2.17), we can obtain the following normalized inertial profile of the transformed code with sub-optimal encoder,

$$\begin{aligned}\widetilde{I}_{\mathrm{tr}}\left(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\right) &\leq \widetilde{I}_{\mathrm{sub}}\left(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\right)\\ &= \frac{\left|\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\right|^{-2/k_{\mathrm{q}}}}{k_{\mathrm{q}}+2}\left(\frac{\left|\mathbf{W}_{0,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\right|}{\kappa_{k_{\mathrm{q}}}^{2}}\right)^{\frac{1}{k_{\mathrm{q}}}}\cdot\mathbf{tr}\Big(\mathbf{W}_{0,\mathbf{z}}(\widehat{\mathbf{y}})^{-1}\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})^{\mathsf{T}}\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\Big),\end{aligned} \tag{2.84}$$

which corresponds to inequality (b) in (2.80).

If the source variable (vector) $\mathbf{y}$ is further subject to k_{c} *constraints* given by the vector equation $\mathbf{g}(\mathbf{y}) = \mathbf{0}$, the distortion function D_{Q} can be similarly expanded around point $\mathbf{F}(\widehat{\mathbf{y}})$ as

$$\begin{aligned}D_{\mathrm{Q}}\big(\mathbf{y}',\,\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\big) \approx (\mathbf{y}-\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\Big(\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\cdot\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\Big)\cdot(\mathbf{y}-\widehat{\mathbf{y}})\\ \mathbf{e}^{\mathsf{T}}\cdot\Big(\mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\cdot\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\cdot\mathbf{V}_2(\widehat{\mathbf{y}})\Big)\cdot\mathbf{e}\ ,\end{aligned} \tag{2.85}$$

where $\mathbf{e}$ is the projected error vector with respect to point $\widehat{\mathbf{y}}$ given by

$$\mathbf{e} = \mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot(\mathbf{y}-\widehat{\mathbf{y}})\ . \tag{2.86}$$

Similarly, by substituting (2.85) and the sub-optimal Voronoi shape (2.79) into the inertial profile definition (2.17), we can obtain the sub-optimal inertial profile of the transformed code with constraint source

$$\begin{aligned}&\widetilde{I}_{\text{tr-c}}\left(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\right) \leq \widetilde{I}_{\text{sub-c}}\left(\mathbf{F}(\widehat{\mathbf{y}})\,;\,\mathbf{z}\right)\\ &= \frac{\Big|\mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\cdot\mathbf{V}_2(\widehat{\mathbf{y}})\Big|^{-\frac{2}{k'_{\mathrm{q}}}}}{k'_{\mathrm{q}}+2}\left(\frac{\Big|\mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\cdot\mathbf{W}_{0,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\cdot\mathbf{V}_2(\widehat{\mathbf{y}})\Big|}{\kappa_{k_{\mathrm{q}}}^{2}}\right)^{\frac{1}{k_{\mathrm{q}}}}\\ &\cdot\mathbf{tr}\left(\Big(\mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\mathbf{W}_{0,\,\mathbf{z}}(\widehat{\mathbf{y}})\,\mathbf{V}_2(\widehat{\mathbf{y}})\Big)^{-1}\cdot\mathbf{V}_2(\widehat{\mathbf{y}})^{\mathsf{T}}\,\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})^{\mathsf{T}}\,\mathbf{W}_{\mathbf{z}}\big(\mathbf{F}(\widehat{\mathbf{y}})\big)\,\mathbf{F}_{\mathrm{d}}(\widehat{\mathbf{y}})\,\mathbf{V}_2(\widehat{\mathbf{y}})\right),\end{aligned} \tag{2.87}$$

which corresponds to inequality (b) in (2.81). ■

2.5.4 Distortion Integral of the Transformed Codebook

By substituting the transformed point density (2.75) and the bounds of the transformed inertial profile given by (2.80) into the distortion integration (2.33), we can upper and lower bound the asymptotic system distortion of a transformed quantizer by the following form

$$\begin{aligned}
&\widetilde{D}_{\text{tr-Low}} = 2^{-\frac{2B}{k_{\text{q}}}} \cdot \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{opt}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{tr}}\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right) \\
&\leq \widetilde{D}_{\text{tr}} = 2^{-\frac{2B}{k_{\text{q}}}} \cdot \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{tr}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{tr}}\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right) \\
&\leq \widetilde{D}_{\text{tr-Upp}} = 2^{-\frac{2B}{k_{\text{q}}}} \cdot \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{sub}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{tr}}\big(\mathbf{y}\big)^{-\frac{2}{k_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right). \qquad (2.88)
\end{aligned}$$

Similarly, by substituting (2.76) and (2.81) into (2.33), the asymptotic distortion of a constrained quantizer with transformed codebook is bounded by

$$\begin{aligned}
&\widetilde{D}_{\text{c-tr-Low}} = 2^{-\frac{2B}{k_{\text{q}'}}} \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{c-opt}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{c-tr}}\big(\mathbf{y}\big)^{-\frac{2}{k'_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right) \\
&\leq \widetilde{D}_{\text{c-tr}} = 2^{-\frac{2B}{k_{\text{q}'}}} \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{c-tr}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{c-tr}}\big(\mathbf{y}\big)^{-\frac{2}{k'_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right) \\
&\leq \widetilde{D}_{\text{c-tr-Upp}} = 2^{-\frac{2B}{k'_{\text{q}}}} \left(\int_{\mathbb{Z}} \int_{\mathbb{Q}} \widetilde{I}_{\text{c-sub}}\left(\mathbf{y}\,;\,\mathbf{z}\right) \cdot p\big(\mathbf{y},\,\mathbf{z}\big) \cdot \lambda_{\text{c-tr}}\big(\mathbf{y}\big)^{-\frac{2}{k'_{\text{q}}}}\, d\,\mathbf{y}\; d\,\mathbf{z} \right). \qquad (2.89)
\end{aligned}$$

Similar to conventional product transformed code [35], there exist tradeoffs between the two sub-optimalities: point density loss and Voronoi shape loss. To be specific, it is always possible to find a transformation $\mathbf{F}(\cdot)$ such that the transformed point density $\lambda_{\text{tr}}(\mathbf{y})$ matches exactly the optimal point density $\lambda^*(\mathbf{y})$. However, by doing so, the transformation may cause severe "oblongitis" of the Voronoi shape in some cases, which will lead to significant increment of the normalized inertial profile. Therefore, a compromised transformation that optimally trades off the two losses should be employed. This tradeoff is directly reflected in the distortion bound $\widetilde{D}_{\text{tr, Upp}}$ where both $\widetilde{I}_{\text{sub}}\left(\mathbf{y}\,;\,\mathbf{z}\right)$ and $\lambda_{\text{tr}}(\mathbf{y})$ in (2.88) depend on the transformation $\mathbf{F}(\cdot)$. So is distortion bound $\widetilde{D}_{\text{c-tr, Upp}}$ given by (2.89).

2.6 Asymptotic Analysis of Constrained Source

The analysis provided above is for the case that the input source $\mathbf{y}$ is a free random vector of dimension $k_{\rm q}$. In some situations, it is required to quantize the $k_{\rm q}$ dimensional source vector $\mathbf{y} \in Q$ subject to constraints,

$$\mathbf{g}\big(\mathbf{y}\big) = \mathbf{0} \ , \tag{2.90}$$

where $\mathbf{g}(\cdot)$ is a multi-dimensional function of size $k_{\rm c} \times 1$. In this case, the degrees of freedom in $\mathbf{y}$ reduce from $k_{\rm q}$ to $k'_{\rm q} = k_{\rm q} - k_{\rm c}$. This subsection provides an asymptotic analysis for such a constrained source $\mathbf{y}$, which is an extension of the problem addressed in Section 2.2.2

First perform a singular value decomposition (SVD) on the derivative[5] of the constraint function $\mathbf{g}(\mathbf{y})$ at point $\mathbf{y} = \widehat{\mathbf{y}}$, which is given by

$$\mathbf{G}\big(\widehat{\mathbf{y}}\big) = \frac{\partial}{\partial \mathbf{y}} \mathbf{g}\big(\,\mathbf{y}\big)\bigg|_{\mathbf{y}=\widehat{\mathbf{y}}} = \mathbf{U}_{\rm G} \left[\mathbf{\Sigma}_{\rm G} \ \ \mathbf{0}\right] \mathbf{V}_{\rm G}^{\mathsf{T}} \ \ , \tag{2.91}$$

where $\mathbf{\Sigma}_{\rm G}$ is the $k_{\rm c} \times k_{\rm c}$ diagonal matrix, and $\mathbf{U}_{\rm G}$ and $\mathbf{V}_{\rm G}$ are unitary matrices of sizes $k_{\rm c} \times k_{\rm c}$ and $k_{\rm q} \times k_{\rm q}$ with matrix $\mathbf{V}_{\rm G}$ further decomposed into $\mathbf{V}_{\rm G} = \big[\mathbf{V}_1 \, \mathbf{V}_2\big]$ with $\mathbf{V}_1$ of size $k_{\rm q} \times k_{\rm c}$ and $\mathbf{V}_2$ of size $k_{\rm q} \times k'_{\rm q}$. Hence, in the neighborhood of point $\widehat{\mathbf{y}}$, the distance vector $(\mathbf{y} - \widehat{\mathbf{y}})$ is approximately constrained to lie in the tangential space of $\mathbf{g}\,(\mathbf{v})$ at $\widehat{\mathbf{v}}$, which is invariant to the following orthogonal projection

$$\mathbf{y} - \widehat{\mathbf{y}} = \big(I - \mathbf{V}_1 \mathbf{V}_1^{\mathsf{T}}\big) \cdot (\mathbf{y} - \widehat{\mathbf{y}}) = \mathbf{V}_2 \mathbf{V}_2^{\mathsf{T}} \cdot (\mathbf{y} - \widehat{\mathbf{y}}) \ \ . \tag{2.92}$$

By substituting the above equality into the conventional Taylor series expansion given by (2.12), one can obtain the constraint Taylor expansion (second order) of the distortion function $D_{\rm Q}\big(\mathbf{y}, \widehat{\mathbf{y}}\,; \mathbf{z}\big)$ about $\mathbf{y} = \widehat{\mathbf{y}}$

$$D_{\rm Q}\big(\,\mathbf{y}\,, \widehat{\mathbf{y}}\,; \mathbf{z}\big) = D_{\rm Q}\big(\,\widehat{\mathbf{y}}\,, \widehat{\mathbf{y}}\,; \mathbf{z}\big) + \mathbf{d}_{\rm c,\,\mathbf{z}}\,\big(\widehat{\mathbf{y}}\big) \cdot \mathbf{e} + \mathbf{e}^{\mathsf{T}} \cdot \mathbf{W}_{\rm c,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big) \cdot \mathbf{e} + O(\|\mathbf{y} - \widehat{\mathbf{y}}\|^3). \tag{2.93}$$

where $\mathbf{e}$ of size $k'_{\rm q} \times 1$ is the transformed distance vector given by

$$\mathbf{e} = \mathbf{V}_2^{\mathsf{T}} \cdot \big(\mathbf{y} - \widehat{\mathbf{y}}\big) \ \ , \tag{2.94}$$

[5] It is assumed that for points $\mathbf{y}$ satisfying constraint (2.90), the derivative $\mathbf{G}\big(\mathbf{y}\big)$ exists and is continuous and full rank.

and the corresponding transformed gradient vector $\mathbf{d}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}}) \in \mathbb{R}^{1\times k'_{\mathrm{q}}}$ and Hessian matrix $\mathbf{W}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}}) \in \mathbb{R}^{k'_{\mathrm{q}}\times k'_{\mathrm{q}}}$ (for the case of constraint source) are given by,

$$\mathbf{d}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}}) = \mathbf{d}_{\mathbf{z}}(\widehat{\mathbf{y}}) \cdot \mathbf{V}_2\,, \qquad \mathbf{W}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}}) = \mathbf{V}_2^{\mathsf{T}} \cdot \mathbf{W}_{\mathbf{z}}(\widehat{\mathbf{y}}) \cdot \mathbf{V}_2\ . \tag{2.95}$$

Employing the same reasoning that $\mathbf{y} = \widehat{\mathbf{y}}$ is the local minimum of the distortion function, one obtains the following similar second order distortion approximation:

$$D_{\mathrm{Q}}(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}) \approx\ \mathbf{e}^{\mathsf{T}} \cdot \mathbf{W}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}}) \cdot \mathbf{e}\ , \tag{2.96}$$

where $\mathbf{W}_{\mathrm{c},\mathbf{z}}(\widehat{\mathbf{y}})$ (of size $k'_{\mathrm{q}} \times k'_{\mathrm{q}}$) is the constraint sensitivity matrix with respect to the transformed distance vector $\mathbf{e}$.

Under high resolution assumptions, the Jacobian matrix $\mathbf{J}(\mathbf{e})$ between the original constraint vector $\mathbf{y}$ and the re-parameterized free vector $\mathbf{e}$ can be approximated as,

$$\mathbf{J}(\mathbf{e}) = \frac{\partial}{\partial\,\mathbf{e}}\left(\mathbf{y}\big|_{\mathbf{g}(\mathbf{y})=0}\right) \approx \mathbf{J}(\mathbf{0}) \overset{a}{=} \left(\mathbf{V}_2^{\mathsf{T}}\mathbf{V}_2\right)^{-1} = I_{k'_{\mathrm{q}}}\ ,$$

with determinant $|\mathbf{J}(\mathbf{e})| \approx 1$, where (a) follows the derivative chain-rule given by

$$\frac{\partial \mathbf{e}}{\partial \mathbf{y}|_{\mathbf{g}(\mathbf{y})=0}} = \left(\frac{\partial \mathbf{e}}{\partial \mathbf{y}}\right)\left(\frac{\partial \mathbf{y}}{\partial \mathbf{y}|_{\mathbf{g}(\mathbf{y})=0}}\right)\ .$$

Therefore, similar to the definition given by (2.17), the normalized inertial profile for the constraint source $\mathbf{y}$ can be approximated as

$$I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) \approx I_{\mathrm{c}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\Big) = V\big(\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}'}})} \int_{\mathbf{e}\in\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{T}}\,\mathbf{W}_{\mathrm{c},\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ \ d\,\mathbf{e}, \tag{2.97}$$

where $\mathbb{E}^{\mathrm{c}}_{\mathbf{z},\,i}$ is the projected Voronoi region defined as $\mathbb{E}^{\mathrm{c}}_{\mathbf{z},\,i} = \Big\{\mathbf{V}_2\cdot\mathbf{e}\,\Big|\,\mathbf{e}\in\mathbb{E}_{\mathbf{z},\,i}\Big\}$ with $\mathbb{E}_{\mathbf{z},\,i}$ defined in (2.14), and $V\big(\mathbb{E}^{\mathrm{c}}_{\mathbf{z},\,i}\big)$ is the corresponding volume of the projected cell, given by

$$V\big(\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\big) = \int_{\mathbf{e}\in\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}} d\,\mathbf{e}\ . \tag{2.98}$$

Similarly to the case of the un-constrained source case, the normalized inertial profile of the constrained source given by (2.97) depends only on the shape

$\mathbb{E}_{\mathbf{z},i}$ (or $\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}$), and the constrained sensitivity matrix $\mathbf{W}^{\mathrm{c}}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)$ of the distortion function at its current location. Moreover, derivations similar to the ones given in Lemma 2 can be carried out to prove that the inertial profile has the same invariant scaling property, given by the following form

$$I_{\mathrm{c}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\Big) = I_{\mathrm{c}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\alpha\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\Big)\ , \tag{2.99}$$

where the definition of the scaling transformation $\alpha\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}$ is given by (2.19). Furthermore, it is also true that the normalized inertial profile of any Voronoi shape $\mathbb{E}_{\mathbf{z},i}$ for a constrained source $\mathbf{y}$ is also lower bounded by that of a "M-shaped" hyper-ellipsoid, i.e.

$$\begin{aligned} I_{\mathrm{c}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i}\Big) &\geq I_{\mathrm{c,\,opt}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\Big) \geq \widetilde{I}_{\mathrm{c,\,opt}}\big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big) \\ &= I_{\mathrm{c}}\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathcal{T}\Big) = \frac{k'_{\mathrm{q}}}{k'_{\mathrm{q}}+2}\cdot\left(\frac{\big|\mathbf{W}_{\mathrm{c},\,\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)\big|}{\kappa^2_{k'_{\mathrm{q}}}}\right)^{1/k'_{\mathrm{q}}}, \end{aligned} \tag{2.100}$$

where $\mathcal{T}\left(\widehat{\mathbf{y}}_i,\,\mathbf{W}_{\mathrm{c},\,\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big),\,V(\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i})\right)$ is the hyper-ellipsoid centered at $\widehat{\mathbf{y}}_i$ with volume $V(\mathbb{E}^{\mathrm{c}}_{\mathbf{z},i})$, which is defined in (2.24).

All the asymptotic analysis provided in Section 2.2.3–Section 2.3.5 are still valid as long as the dimension of the quantized space is replaced by k'_{q}, the sensitivity matrix is replaced by the constraint one $\mathbf{W}_{\mathrm{c},\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$, while the asymptotic distortion integrations are over the constrained space (2.90). For example, the asymptotic distortion lower bound $D_{\mathrm{Low},1}$ for the constrained source input, denoted as $D_{\text{c-Low},1}$, can be obtained as

$$D_{\text{c-Opt}} \geq D_{\text{c-Low},1} = \left(\int_{\mathbf{g}(\mathbf{y})=0}\ \Big(I^{\mathrm{w}}_{\mathrm{c,opt}}\big(\mathbf{y}\big)\cdot p\big(\mathbf{y}\big)\Big)^{1\Big/\left(1+\frac{2}{k'_{\mathrm{q}}}\right)}\,d\,\mathbf{y}\right)^{1+\frac{2}{k'_{\mathrm{q}}}}\cdot 2^{-2B/k'_{\mathrm{q}}}\ , \tag{2.101}$$

where the constrained average inertial profile $I^{\mathrm{w}}_{\mathrm{c,opt}}\big(\mathbf{y}\big)$ is given by

$$I^{\mathrm{w}}_{\mathrm{c,opt}}\big(\mathbf{y}\big) = \int_{\mathbb{Z}}\ I_{\mathrm{c,opt}}\big(\mathbf{y}\,;\,\mathbf{z}\big)\cdot p\big(\mathbf{z}\big|\,\mathbf{y}\big)\ d\,\mathbf{z}\ . \tag{2.102}$$

Following similar derivations, other asymptotic analysis bounds, such as $D_{\text{c-Low},2}$, $D_{\text{c-Upp}}$, $\widetilde{D}_{\text{c-Low},1}$, $\widetilde{D}_{\text{c-Low},2}$, and $\widetilde{D}_{\text{c-Upp}}$ can also be readily obtained.

2.7 Asymptotic Analysis of Complex Source

In some cases, the source variable to be quantized is a complex vector. In order to apply the asymptotic distortion analysis provided in [45], one can always transform the source vector from the complex domain to the real domain. However, under certain conditions, the proposed distortion analysis can also be extended to deal with complex source variables directly without increasing the vector size (due to the transformation) and hence save derivation efforts.

2.7.1 Quantization of Unconstrained Source

By utilizing the Wirtinger Calculus [47], the distortion function D_{Q} can be Taylor series expanded in the complex domain (without first transforming into the real domain) since D_{Q} is a real function of complex vector $\mathbf{y} \in \mathbb{C}^{k_{\mathrm{q}}}$. With the local minimal assumption of D_{Q} w.r.t. $\mathbf{y}$ at point $\mathbf{y} = \widehat{\mathbf{y}}$, the second order approximation of the distortion function is given by

$$D_{\mathrm{Q}}\big(\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big) \approx \big(\mathbf{y}-\widehat{\mathbf{y}}\big)^{\mathsf{H}}\,\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\,\big(\mathbf{y}-\widehat{\mathbf{y}}\big)+\Re\left[\big(\mathbf{y}-\widehat{\mathbf{y}}\big)^{\mathsf{T}}\,\mathbf{W}'_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\,\big(\mathbf{y}-\widehat{\mathbf{y}}\big)\right]\quad, \tag{2.103}$$

where $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big), \mathbf{W}'_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big) \in \mathbb{C}^{k_{\mathrm{q}}\times k_{\mathrm{q}}}$ are complex Hessian matrices with $(i,j)^{\mathrm{th}}$ element given by

$$w_{i,j} = \frac{\partial^2}{\partial\, y_i^*\,\partial\, y_j} D_{\mathrm{Q}}\big(\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)\,\Big|_{\mathbf{y}=\widehat{\mathbf{y}}}\,, \qquad w'_{i,j} = \frac{\partial^2}{\partial\, y_i\,\partial\, y_j} D_{\mathrm{Q}}\big(\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big)\,\Big|_{\mathbf{y}=\widehat{\mathbf{y}}} \tag{2.104}$$

with y_i^* representing the complex conjugate of y_i. If the distortion function D_{Q} satisfies the condition that matrix $\mathbf{W}'_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big) = \mathbf{0}$, by extending the same reasonings used in the analysis of real vectors, the corresponding normalized inertial profile of complex source is given by

$$I\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) = V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k_{\mathrm{q}}})} \int_{\mathbf{e}\in\mathbb{E}_{\mathbf{z},i}} \mathbf{e}^{\mathsf{H}}\,\mathbf{W}_{\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\;\,d\,\mathbf{e}\;, \tag{2.105}$$

whose tight[6] lower bound (having a "M-shaped" Hyper-ellipsoidal Voronoi region) can be represented as

$$I_{\text{opt}}\big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big) \gtrapprox \widetilde{I}_{\text{opt}}\big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big) = \frac{k_{\text{q}}}{k_{\text{q}}+1}\cdot\left(\frac{\big|\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)\big|^2}{\kappa^2_{2k_{\text{q}}}}\right)^{1/2k_{\text{q}}} \quad . \tag{2.106}$$

2.7.2 Quantization of Constrained Source

Now let us consider the case where the complex source variable $\mathbf{y}$ is further restricted under some real constraints denoted as $\mathbf{g}(\mathbf{y})$. Suppose function $\mathbf{g}(\mathbf{y})$ is of size $2k_{\text{c}}\times 1$, and can be partitioned into the following form under certain orderings

$$\mathbf{g}(\mathbf{y}) = \big[\ \mathbf{g}_1^{\mathsf{T}}(\mathbf{y})\ \ \mathbf{g}_2^{\mathsf{T}}(\mathbf{y})\ \big]^{\mathsf{T}}\ , \qquad \mathbf{g}_1(\mathbf{y}),\,\mathbf{g}_2(\mathbf{y}) \in \mathbb{R}^{k_{\text{c}}\times 1}\ .$$

The following proposition states the necessary and sufficient condition for the inertial profile of the complex constrained source to have a concise format similar to the real case.

Proposition 2 *The normalized inertial profile of the constrained complex source* $\mathbf{y}$ *can be represented as the following form*

$$I_c\Big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\,;\,\mathbb{E}_{\mathbf{z},i}\Big) = V\big(\mathbb{E}_{\mathbf{z},i}\big)^{-(1+\frac{2}{k'_q})}\int_{\mathbf{e}\in\mathbb{E}^c_{\mathbf{z},i}} \mathbf{e}^H\,\mathbf{W}_{c,\,\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)\,\mathbf{e}\ \,d\,\mathbf{e}\ , \tag{2.107}$$

where $k'_q = k_q - k_c$ *and the constrained sensitivity matrix* $\mathbf{W}_{c,\,\mathbf{z}}\big(\,\widehat{\mathbf{y}}_i\,\big)$ *is given by*

$$\mathbf{W}_{c,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big) = \mathbf{V}_2^H\cdot\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\cdot\mathbf{V}_2\ . \tag{2.108}$$

with the unconstrained sensitivity matrix $\mathbf{W}_{\mathbf{z}}\big(\widehat{\mathbf{y}}\big)$ *given by equation (2.104) and matrix* $\mathbf{V}_2$ *being an orthonormal matrix with its columns constituting an orthonormal basis of the null space* $\mathcal{N}\big(\frac{\partial}{\partial\mathbf{y}}\,\mathbf{g}_1(\mathbf{y})\big)$*, if and only if there exists a non-singular matrix* $\mathbf{\Phi}$ *satisfies the following equation,*

$$\frac{\partial\,\mathbf{g}_1(\widetilde{\mathbf{y}}')}{\partial\,\widetilde{\mathbf{y}}'} = \mathbf{\Phi}\cdot\frac{\partial\,\mathbf{g}_2(\widetilde{\mathbf{y}})}{\partial\,\widetilde{\mathbf{y}}}\,, \qquad \widetilde{\mathbf{y}} = \big[\,\mathbf{y}^T,\,\mathbf{y}^H\big]^T, \qquad \widetilde{\mathbf{y}}' = \big[\,-\mathbf{y}^T,\,\mathbf{y}^H\big]^T\ . \tag{2.109}$$

[6]Symbol "$\gtrapprox$" represents that the low bound is tight, and can be used as a good approximation in most cases.

In this case, the constrained inertial profile can also be tightly lower bounded by the following form

$$I_{c,opt}\big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big) \gtrapprox \widetilde{I}_{c,opt}\big(\widehat{\mathbf{y}}_i\,;\,\mathbf{z}\big) = \frac{k'_q}{k'_q+1}\cdot\left(\frac{\big|\mathbf{W}_{c,\mathbf{z}}\big(\widehat{\mathbf{y}}_i\big)\big|^2}{\kappa^2_{2k'_q}}\right)^{\frac{1}{2k'_q}} \quad . \tag{2.110}$$

Proof: (Sufficient Condition:) According to the property of the complex derivative provided in [48], the following equality is valid

$$\frac{\partial\,\mathbf{g}(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}} = \frac{\partial\,\mathbf{g}(\widetilde{\mathbf{y}})}{\partial\,\widetilde{\mathbf{y}}}\cdot\begin{bmatrix} I_{k_{\rm q}} & j\,I_{k_{\rm q}} \\ I_{k_{\rm q}} & -j\,I_{k_{\rm q}}\end{bmatrix},\quad \overline{\mathbf{y}} = \big[\,\mathbf{y}_{\rm R}^{\mathsf{T}},\,\mathbf{y}_{\rm I}^{\mathsf{T}}\big]^{\mathsf{T}},\quad \overline{\mathbf{y}}' = \big[\,-\mathbf{y}_{\rm I}^{\mathsf{T}},\,\mathbf{y}_{\rm R}^{\mathsf{T}}\big]^{\mathsf{T}}, \tag{2.111}$$

where $\mathbf{y}_{\rm R}$ and $\mathbf{y}_{\rm I}$ are the real and imaginary part of $\mathbf{y}$. By substituting (2.109) into (2.111) and after some manipulations, we can obtain the following relation[7],

$$\overline{\frac{\partial\,\mathbf{g}_1(\mathbf{y})}{\partial\,\mathbf{y}}} \;=\; \begin{bmatrix} I_{k_{\rm c}} & \mathbf{0} \\ \mathbf{0} & j\cdot\mathbf{\Phi}\end{bmatrix}\frac{\partial\,\mathbf{g}(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}\;, \tag{2.112}$$

Since the column vectors of matrix $\mathbf{V}_2 \in \mathbb{C}^{k_{\rm q}\times k'_{\rm q}}$ span the null space given by $\mathcal{N}\big(\frac{\partial}{\partial\,\mathbf{y}}\,\mathbf{g}_1(\mathbf{y})\big)$, it is evident that column vectors of $\overline{\mathbf{V}}_2$ should span the null space of $\mathcal{N}\left(\overline{\frac{\partial}{\partial\,\mathbf{y}}\,\mathbf{g}_1(\mathbf{y})}\right)$. Moveover, according to equation (2.112), columns of matrix $\overline{\mathbf{V}}_2$ also span the null space $\mathcal{N}\big(\frac{\partial}{\partial\,\overline{\mathbf{y}}}\,\mathbf{g}(\overline{\mathbf{y}})\big)$. By employing the same reasoning used in [45], one can obtain the following second order Taylor series expansion of the distortion function after some manipulations,

$$D_{\rm Q}\big(\,\mathbf{y}\,,\widehat{\mathbf{y}}\,;\,\mathbf{z}\big) \approx \overline{(\mathbf{y}-\widehat{\mathbf{y}})}^{\mathsf{T}}\cdot\overline{\mathbf{V}}_2\overline{\mathbf{V}}_2^{\mathsf{T}}\cdot\overline{\mathbf{W}_{\rm c,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big)}\cdot\overline{\mathbf{V}}_2\overline{\mathbf{V}}_2^{\mathsf{T}}\cdot\overline{(\mathbf{y}-\widehat{\mathbf{y}})} = \mathbf{e}^{\mathsf{H}}\,\mathbf{W}_{\rm c,\,\mathbf{z}}\big(\widehat{\mathbf{y}}\big)\mathbf{e}, \tag{2.113}$$

where vector $\mathbf{e} \in \mathbb{C}^{k'_{\rm q}}\times 1$ is given by $\mathbf{e} = \mathbf{V}_2^{\mathsf{H}}\cdot(\mathbf{y}-\widehat{\mathbf{y}})$. Again, by extending the same reasonings used in the analysis of real vectors, the constrained complex inertial profile (2.107) as well as its tight lower bound (2.110) can be obtained.

(Necessary Condition:) On the other hand, if the distortion function $D_{\rm Q}$ has a concise second order approximation given by (2.113), which can further lead

[7]Operation $\overline{\mathbf{A}}$ of a matrix $\mathbf{A}$ is defined to be $\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_{\rm R} & \mathbf{A}_{\rm I} \\ -\mathbf{A}_{\rm I} & \mathbf{A}_{\rm R}\end{bmatrix}$, where $\mathbf{A}_{\rm R}$ and $\mathbf{A}_{\rm I}$ represents the real and imaginary part of matrix $\mathbf{A}$.

to a concise inertial profile expression, the column vectors of matrix $\overline{\mathbf{V}}_2$ span the null space $\mathcal{N}\big(\frac{\partial}{\partial \overline{\mathbf{y}}}\,\mathbf{g}(\overline{\mathbf{y}})\big)$. This means that the two range spaces $\mathcal{R}\Big(\big(\frac{\partial}{\partial \overline{\mathbf{y}}}\,\mathbf{g}(\overline{\mathbf{y}})\big)^{\mathsf{T}}\Big)$ and $\mathcal{R}\Big(\big(\overline{\frac{\partial}{\partial \mathbf{y}}\,\mathbf{g}_1(\mathbf{y})}\big)^{\mathsf{T}}\Big)$ are equivalent. Moveover, it can be shown that the following equality is valid

$$\begin{aligned}\mathcal{R}\left(\left(\overline{\frac{\partial\,\mathbf{g}_1(\mathbf{y})}{\partial\,\mathbf{y}}}\right)^{\mathsf{T}}\right) &= \mathcal{R}\left(\left[\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}, \frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}}')}{\partial\,\overline{\mathbf{y}}'}^{\mathsf{T}}\right]\right) \\ &= \mathcal{R}\left(\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}\right) \oplus \mathcal{R}\left(\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}}')}{\partial\,\overline{\mathbf{y}}'}^{\mathsf{T}}\right), \qquad (2.114)\end{aligned}$$

where "$\oplus$" represents the subspace summation. Similarly, one can also obtain the following equality

$$\begin{aligned}\mathcal{R}\left(\left(\frac{\partial\,\mathbf{g}(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}\right)^{\mathsf{T}}\right) &= \mathcal{R}\left(\left[\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}, \frac{\partial\,\mathbf{g}_2(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}\right]\right) \\ &= \mathcal{R}\left(\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}\right) \oplus \mathcal{R}\left(\frac{\partial\,\mathbf{g}_2(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}^{\mathsf{T}}\right). \qquad (2.115)\end{aligned}$$

It is evident from (2.114) and (2.115) that the two range spaces represented by $\mathcal{R}\Big(\big(\frac{\partial}{\partial \overline{\mathbf{y}}'}\,\mathbf{g}_1(\overline{\mathbf{y}}')\big)^{\mathsf{T}}\Big)$ and $\mathcal{R}\Big(\big(\frac{\partial}{\partial \overline{\mathbf{y}}}\,\mathbf{g}_2(\overline{\mathbf{y}})\big)^{\mathsf{T}}\Big)$ are equivalent and there exists a non-singular matrix $\mathbf{\Psi}$ such that

$$\frac{\partial\,\mathbf{g}_1(\overline{\mathbf{y}}')}{\partial\,\overline{\mathbf{y}}'} = \mathbf{\Psi}\cdot\frac{\partial\,\mathbf{g}_2(\overline{\mathbf{y}})}{\partial\,\overline{\mathbf{y}}}\,. \qquad (2.116)$$

From equation (2.116), one can further obtain the following equality after some manipulations

$$\frac{\partial\,\mathbf{g}_1(\widetilde{\mathbf{y}}')}{\partial\,\widetilde{\mathbf{y}}'} = \left(\frac{-j\mathbf{\Psi}}{2}\right)\cdot\frac{\partial\,\mathbf{g}_2(\widetilde{\mathbf{y}})}{\partial\,\widetilde{\mathbf{y}}}\,. \qquad (2.117)$$

Therefore, as long as the rows of two derivatives matrices $\frac{\partial\,\mathbf{g}_1(\widetilde{\mathbf{y}}')}{\partial\,\widetilde{\mathbf{y}}'}$ and $\frac{\partial\,\mathbf{g}_2(\widetilde{\mathbf{y}})}{\partial\,\widetilde{\mathbf{y}}}$ span the same subspace, the complex constrained inertial profile can be expressed in a concise form. ■

2.8 Summary

This chapter has developed a general framework for the analysis of multiple antenna systems with finite rate feedback from a source coding perspective.

Without narrowing the scope to a specific channel quantization scheme, the problem was formulated as a general fixed-rate vector quantization problem with side information available at the encoder but unavailable at the decoder. The proposed framework is sufficiently general to include quantization schemes with non-mean square distortion functions, and cases where the source vector is constrained. Asymptotic distortion analysis of the proposed general quantization problem was provided by extending Bennett's classic analysis. More specifically, tight lower and upper bounds of the average asymptotic distortion and sufficient conditions for the achievability of the distortion bounds were provided. Based on the general framework, the asymptotic distortion analysis was further extended to the important practical problem of sub-optimal quantizers resulting from mismatches in the distortion functions, source statistics, and quantization criteria. As a further demonstration of the utility of the framework, sub-optimal vector quantizers using transformed codebooks were investigated. Moreover, the problem of quantizing complex source variables was also investigated in this chapter. It was shown that under certain necessary and sufficient conditions, the distortion analysis of complex source variables can be performed in a concise manner in the complex domain without first transforming the problem into real domains. The text of this chapter is in part a reprint of the material which was coauthored with Ethan R. Duni and Bhaskar D. Rao and has been accepted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of multiple antenna systems with finite rate feedback using high resolution quantization theory"*.

3 Capacity Analysis of MISO Systems with Finite-Rate CSI Feedback

3.1 Motivation

Due to the complexity of the analysis, most of the past works on multiple antenna systems with finite-rate feedback are case specific, limited to spatially i.i.d. fading channels and mainly MISO channels, and are difficult to extend to more general scenarios. By utilizing the high-rate distortion analysis described in Chap. 2, we investigate in this chapter the performance of a MISO beamforming system with finite-rate CSI feedback over spatially correlated fading channels. The analysis of system capacity loss due to the finite-rate quantization of the channel state information is provided, and is further compared with that of MISO systems over spatially i.i.d. fading channels. The obtained analytical results provide interesting insights and demonstrate the general nature as well as the utility of the high-resolution analytical framework.

3.2 System Model

This section considers a MISO system, with t transmit antennas and one receive antenna, signaling through a frequency flat block fading channel. For the

sake of simplicity, the time index is omitted, and hence the channel model can be represented as the following form

$$y = \mathbf{h}^{\mathsf{H}} \cdot \mathbf{x} + n \ , \tag{3.1}$$

where y is the received signal (scalar), n is the additive complex Gaussian noise with zero mean and unit variance, and $\mathbf{h}^{\mathsf{H}} \in \mathbb{C}^{1\times t}$ is the correlated[1] MISO channel response with distribution given by $\mathbf{h} \sim \mathcal{N}_{\mathrm{c}}\big(\mathbf{\Sigma}_{\mathrm{h}}\big)$. The transmitted signal vector $\mathbf{x}$ is normalized to have a power constraint given by $E\big[\|\mathbf{x}\|^2\big] = \rho$, with ρ representing the average receiver signal to noise ratio.

In this chapter, the channel state information $\mathbf{h}$ is assumed to be perfectly known at the receiver but only partially available at the transmitter through a finite-rate feedback link of B bits per channel update between the transmitter and receiver. More specifically, a quantization codebook $\mathcal{C} = \big\{\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_N\big\}$, which is composed of unit-norm transmit beamforming vectors, is assumed known to both the receiver and the transmitter. Based on the channel realization $\mathbf{h}$, the receiver selects the best code point $\widehat{\mathbf{v}}$ from the codebook and sends the corresponding index back to the transmitter. At the transmitter, the unit-norm vector $\widehat{\mathbf{v}}$ is employed as the beamforming vector, i.e.

$$y = \langle \mathbf{h}, \widehat{\mathbf{v}} \rangle \cdot s + n = \|\mathbf{h}\| \cdot \langle \mathbf{v}, \widehat{\mathbf{v}} \rangle \cdot s + n \, , \qquad E\big[\,|s|^2\big] = \rho \ . \tag{3.2}$$

where $\mathbf{v}$ is the channel directional vector given by $\mathbf{v} = \mathbf{h}/\|\mathbf{h}\|$. The corresponding ergodic capacity or the maximum system mutual information rate of the quantized MISO beamforming system is given by

$$C_{\mathrm{Q}} = E\Big[\, \log_2 \Big(1 + \rho \cdot \|\mathbf{h}\|^2 \cdot |\langle \mathbf{v}, \widehat{\mathbf{v}} |^2\Big)\Big] \ . \tag{3.3}$$

With perfect channel state information available at the transmitter, which corresponds to the case of infinite rate feedback $B = \infty$, it is optimal to choose

[1] For the sake of fair comparisons, we normalize the channel covariance matrix such that the mean of the eigen values equals to one (equal to the i.i.d. channel case $\mathbf{\Sigma}_{\mathrm{h}} = I_t$).

$\mathbf{v} = \mathbf{h}/\|\mathbf{h}\|$ as the transmit beamforming vector, and the corresponding system ergodic capacity is given by

$$C_{\mathrm{p}} = E\Big[\log_2\Big(1+\rho\cdot\|\mathbf{h}\|^2\Big)\Big]\ . \tag{3.4}$$

Therefore, the performance of a CSI-feedback-based MISO system can be characterized by the capacity loss C_{Loss} due to the finite-rate quantization of the transmit beamforming vectors, which is defined as the expectation of the instantaneous mutual information rate loss,

$$C_{\mathrm{Loss}} = C_{\mathrm{p}} - C_{\mathrm{Q}} = E\big[\,C_{\mathrm{L}}(\mathbf{h},\widehat{\mathbf{v}})\,\big]\ , \tag{3.5}$$

where $C_{\mathrm{L}}(\mathbf{h},\widehat{\mathbf{v}})$ is given by the following form

$$C_{\mathrm{L}}(\mathbf{h},\widehat{\mathbf{v}}) = -\log_2\left(1-\frac{\rho\cdot\|\mathbf{h}\|^2}{1+\rho\cdot\|\mathbf{h}\|^2}\cdot\Big(1-|\langle\mathbf{v},\widehat{\mathbf{v}}\rangle|^2\Big)\right) \tag{3.6}$$

This performance metric was also used in [27] and [45].

3.3 Problem Formulation

This section provides insight into MISO beamforming systems with finite rate feedback by utilizing the proposed generalized asymptotic analysis obtained in Section 2.2. Derivations of the bounds for MISO systems are carried out step by step and with references back to the corresponding equations of the general theory in Section 2.2 for a better understanding. According to the capacity loss formula given by equation (3.5), the directional vector $\mathbf{v}$ of the MISO channel response becomes the actual variable to be quantized (or the quantization objective), denoted as

$$\mathbf{v} = \Big[r_1\,e^{j\theta_1}, r_2\,e^{j\theta_2}, \cdots, r_t\,e^{j\theta_t}\Big]^{\mathsf{T}}\ , \tag{3.7}$$

where the magnitudes r_i have constraint $\sum_{i=1}^{t} r_i^2 = 1$. Furthermore, the capacity loss is also invariant to an arbitrary phase rotation $e^{j\phi}$ on vector $\mathbf{v}$. Therefore, only the relative phase $\theta_i - \theta_1$, with $2 \le i \le t$, is of interest and hence a phase constraint

$\theta_1 = const$ can be set on vector $\mathbf{v}$. Due to the invariant transformation of the arbitrary phase rotation and also for simplifying the derivations of the distortion analysis, an equivalent phase constraint on $\mathbf{v}$ is imposed for the points in the neighborhood of $\widehat{\mathbf{v}}$,

$$\measuredangle\langle \mathbf{v},\, \widehat{\mathbf{v}} \rangle = 0 \ . \tag{3.8}$$

To make use of the framework, represent the vectors $\mathbf{v}$ and $\widehat{\mathbf{v}}$ as having the following real and imaginary components,

$$\mathbf{v} = \mathbf{v}_{\mathrm{R}} + j\,\mathbf{v}_{\mathrm{I}}\,, \qquad \widehat{\mathbf{v}} = \widehat{\mathbf{v}}_{\mathrm{R}} + j\,\widehat{\mathbf{v}}_{\mathrm{I}} \ \ , \tag{3.9}$$

where $\mathbf{v}_{\mathrm{R}}$, $\mathbf{v}_{\mathrm{I}}$, $\widehat{\mathbf{v}}_{\mathrm{R}}$, and $\widehat{\mathbf{v}}_{\mathrm{I}}$ are all real vectors of sizes $t \times 1$. Further stack the real and imaginary part of $\mathbf{v}$ together into a $2t \times 1$ real vector denoted as $\overline{\mathbf{v}}$, i.e. $\overline{\mathbf{v}} = \left[\mathbf{v}_{\mathrm{R}}^{\mathsf{T}},\, \mathbf{v}_{\mathrm{I}}^{\mathsf{T}}\right]^{\mathsf{T}}$. The finite rate feedback MISO transmit beamforming problem is now described as below. The source input $\mathbf{x}$ is equal to $\mathbf{h}$, with the quantization variable (or objective) $\mathbf{y} = \mathbf{v}$ of dimension $k_{\mathrm{q}} = 2t$, and the side information $\mathbf{z} = \alpha$ ($\alpha = \|\mathbf{h}\|^2$) of dimension $k_{\mathrm{z}} = 1$. The constraint conditions on the quantization vector $\mathbf{v}$, denoted as $\mathbf{g}(\mathbf{v})$, can be represented as the following multi-dimensional real function, i.e.

$$\mathbf{g}\big(\mathbf{v}\big) = \begin{bmatrix} \mathbf{v}_{\mathrm{R}}^{\mathsf{T}}\mathbf{v}_{\mathrm{R}} + \mathbf{v}_{\mathrm{I}}^{\mathsf{T}}\mathbf{v}_{\mathrm{I}} - 1 \\ \mathbf{v}_{\mathrm{R}}^{\mathsf{T}}\widehat{\mathbf{v}}_{\mathrm{I}} - \mathbf{v}_{\mathrm{I}}^{\mathsf{T}}\widehat{\mathbf{v}}_{\mathrm{R}} \end{bmatrix} \ , \tag{3.10}$$

with the first element representing the unit norm constraint on $\mathbf{v}$, and the second element the phase constraint. Function $\mathbf{g}(\mathbf{v})$ has size $k_{\mathrm{c}} = 2$, which leads to the actual degrees of freedom of the quantization variable $\mathbf{v}$ to be $k'_{\mathrm{q}} = 2t - 2$. The distortion function D_{Q} is given by

$$\begin{aligned} D_{\mathrm{Q}}\Big(\mathbf{y},\, \widehat{\mathbf{y}}\,;\, \mathbf{z}\Big) &= D_{\mathrm{Q}}\Big(\mathbf{v},\, \widehat{\mathbf{v}}\,;\, \alpha\Big) = -\log_2\left(1 - \frac{\rho\alpha}{1+\rho\alpha}\cdot\Big(1 - \big|\langle \mathbf{v},\, \widehat{\mathbf{v}}\rangle\big|^2\Big)\right) \\ &= -\log_2\left(1 - \frac{\rho\alpha}{1+\rho\alpha}\cdot \overline{\mathbf{v}}^{\mathsf{T}}\big(I_{2t} - \boldsymbol{\Omega}\big)\,\overline{\mathbf{v}}\right) , \end{aligned} \tag{3.11}$$

where matrix $\boldsymbol{\Omega} \in \mathbb{R}^{2k_{\mathrm{q}} \times 2k_{\mathrm{q}}}$ is given by

$$\boldsymbol{\Omega} = \left[\begin{array}{c|c} \widehat{\mathbf{v}}_{\mathrm{R}}\widehat{\mathbf{v}}_{\mathrm{R}}^{\mathsf{T}} + \widehat{\mathbf{v}}_{\mathrm{I}}\widehat{\mathbf{v}}_{\mathrm{I}}^{\mathsf{T}} & \widehat{\mathbf{v}}_{\mathrm{R}}\widehat{\mathbf{v}}_{\mathrm{I}}^{\mathsf{T}} - \widehat{\mathbf{v}}_{\mathrm{I}}\widehat{\mathbf{v}}_{\mathrm{R}}^{\mathsf{T}} \\ \hline \widehat{\mathbf{v}}_{\mathrm{I}}\widehat{\mathbf{v}}_{\mathrm{R}}^{\mathsf{T}} - \widehat{\mathbf{v}}_{\mathrm{R}}\widehat{\mathbf{v}}_{\mathrm{I}}^{\mathsf{T}} & \widehat{\mathbf{v}}_{\mathrm{R}}\widehat{\mathbf{v}}_{\mathrm{R}}^{\mathsf{T}} + \widehat{\mathbf{v}}_{\mathrm{I}}\widehat{\mathbf{v}}_{\mathrm{I}}^{\mathsf{T}} \end{array}\right] \ . \tag{3.12}$$

This corresponds back to the definition of the distortion function and satisfies the local minimal property given by (2.3).

3.4 Statistical Properties of the Channel Information

It is know that both the design as well as the analysis of a vector quantizer depends heavily on the source statistical distributions. Therefore, before dipping into the details of the analysis of MISO beamforming systems with finite-rate feedback, let us first look at some of statistical properties of the channel state information to be quantized. To be specific, we are interested in the joint and marginal probability density functions of the constrained source vectors $\mathbf{v}$ and the encoder side information α. After some manipulations, the results of the distribution functions can be described into the following lemma.

Lemma 4 *Suppose the MISO channel* $\mathbf{h} \in \mathbb{C}^{t\times 1}$ *has a complex Gaussian distribution given by* $\mathbf{h} \sim \mathcal{N}_c\big(\mathbf{0},\, \mathbf{\Sigma}_h\big)$*, and its constrained directional vector is defined to be,*

$$\mathbf{v} = \left(\sqrt{\alpha}\; e^{j\theta}\right)^{-1} \cdot \mathbf{h}\,, \qquad \alpha = \|\mathbf{h}\|^2, \qquad \theta = \measuredangle\,\langle \mathbf{v}_0,\, \mathbf{h}\rangle - \phi\ , \tag{3.13}$$

where $\mathbf{v}_0$ *is a fixed unit norm vector, and* $\phi \in [0,\, 2\pi]$ *is a fixed phase . It is equivalent as saying that vector* $\mathbf{v}$ *satisfies the following constraints*

$$\|\mathbf{v}\| = 1, \qquad \measuredangle\,\langle \mathbf{v}_0,\, \mathbf{v}\rangle = \phi\ . \tag{3.14}$$

Then the following statements are true,

1. *Random phase variable* θ *is independent of the channel power gain* α *as well as the directional vector* $\mathbf{v}$*, and* θ *is uniformly distributed between* 0 *and* 2π*, i.e.* $\theta \sim \mathcal{U}\,(0,\, 2\pi)$.

2. *If the elements of vector* $\mathbf{h}$ *are i.i.d., i.e.* $\mathbf{\Sigma}_h = I_t$ *, the channel power gain* α *is statistically independent of the directional vector* $\mathbf{v}$*. The probability density*

functions of random variables (or vector) α *and* $\mathbf{v}$ *are given by*

$$p_{\alpha}(x) = \frac{x^{t-1} \cdot e^{-x}}{(t-1)!}, \tag{3.15}$$

$$p_{\mathbf{v}}(\mathbf{x}) = 1/\gamma_t \ , \tag{3.16}$$

where $\gamma_t = \pi^{t-1}/(t-1)!$.

3. *If the singular values of* $\mathbf{\Sigma}_h$ *are positive and distinct, i.e.* $\lambda_{h,1} > \cdots > \lambda_{h,t} > 0$, *the marginal probability density functions of* α *and* $\mathbf{v}$ *are given by*

$$p_{\mathbf{v}}(\mathbf{x}) = \gamma_t^{-1} \cdot |\mathbf{\Sigma}_h|^{-1} \cdot \left(\mathbf{x}^{\mathsf{H}} \mathbf{\Sigma}_h^{-1} \mathbf{x}\right)^{-t} \ , \tag{3.17}$$

$$p_{\alpha}(x) = \sum_{i=1}^{t} \prod_{j \neq i} \left(1 - \frac{\lambda_{h,j}}{\lambda_{h,i}}\right)^{-1} \cdot \frac{1}{\lambda_{h,i}} \exp\left(-\frac{x}{\lambda_{h,i}}\right) \ . \tag{3.18}$$

And the corresponding conditional distributions are given by

$$\begin{aligned} p_{\mathbf{v}|\alpha}(\mathbf{x}) &= \left(\sum_{i=1}^{t} \prod_{j \neq i} \left(1 - \frac{\lambda_{h,j}}{\lambda_{h,i}}\right)^{-1} \cdot \frac{1}{\lambda_{h,i}} \exp\left(-\frac{\alpha}{\lambda_{h,i}}\right)\right)^{-1} \\ &\times \frac{\alpha^{t-1} \cdot \exp\left(-\alpha \cdot \mathbf{x}^{\mathsf{H}} \mathbf{\Sigma}_h^{-1} \mathbf{x}\right)}{\pi^{t-1} \cdot |\mathbf{\Sigma}_h|} \ , \end{aligned} \tag{3.19}$$

$$p_{\alpha|\mathbf{v}}(x) = \frac{x^{t-1} \cdot \left(\mathbf{v}^{\mathsf{H}} \mathbf{\Sigma}_h^{-1} \mathbf{v}\right)^t \cdot \exp\left(-x \cdot \mathbf{v}^{\mathsf{H}} \mathbf{\Sigma}_h^{-1} \mathbf{v}\right)}{(t-1)!} \ . \tag{3.20}$$

Proof: Let us first denote unitary matrix $\mathbf{P} \in \mathbb{C}^{t \times t}$ as the following form

$$\mathbf{P} = \Big[\mathbf{v}_0, \mathbf{P}_1\Big] = \Big[\mathbf{v}_0, \mathbf{v}_2, \mathbf{v}_3, \cdots, \mathbf{v}_t\Big] \ , \tag{3.21}$$

where the columns of $\mathbf{P}$ (and $\mathbf{P}_1$) are orthonormal vectors. Left multiplying both sides of equation (3.13) by matrix $\mathbf{P}^{\mathsf{H}}$, we can obtain

$$\mathbf{h}' = \mathbf{P}^{\mathsf{H}} \mathbf{h} = \sqrt{\alpha}\, e^{j\theta} \cdot \mathbf{P}^{\mathsf{H}} \mathbf{v} = \sqrt{\alpha}\, e^{j\theta} \cdot \mathbf{v}' \ , \tag{3.22}$$

where $\mathbf{v}'$ is given by

$$\mathbf{v}' = \begin{bmatrix} \langle \mathbf{v}_0, \mathbf{v} \rangle \\ \mathbf{P}_1^{\mathsf{H}} \mathbf{v} \end{bmatrix} = \Big[r_1' e^{j\phi}, r_2' e^{\theta_2'}, \cdots, r_t' e^{\theta_t'}\Big]^T \ . \tag{3.23}$$

It is evident that vector $\mathbf{h}'$ has the same distribution as $\mathbf{h}$ under the unitary transformation $\mathbf{P}$, i.e. $\mathbf{h}' \sim \mathcal{N}_{\mathrm{c}}\big(\mathbf{0},\, \mathbf{\Sigma}_{\mathrm{h}}\big)$. By taking the following variable transformation

$$\Big(h'_{\mathrm{R},1}, h'_{\mathrm{I},1}, \cdots, h'_{\mathrm{R},t}, h'_{\mathrm{I},t}\Big) \longrightarrow \Big(\alpha\,,\, \theta,\, r'_2,\, \theta'_2,\, \cdots, r'_t,\, \theta'_t\Big)\,, \tag{3.24}$$

such that

$$\begin{aligned}
h'_{\mathrm{R},1} &= \sqrt{\alpha}\cdot\big(1-\textstyle\sum_{i=2}^{t} r_i'^{\,2}\big)^{1/2}\cdot\cos\big(\theta+\phi\big)\,,\\
h'_{\mathrm{I},1} &= \sqrt{\alpha}\cdot\big(1-\textstyle\sum_{i=2}^{t} r_i'^{\,2}\big)^{1/2}\cdot\sin\big(\theta+\phi\big)\,,\\
h'_{\mathrm{R},2} &= \sqrt{\alpha}\cdot r'_2\cdot\cos\big(\theta+\theta'_2\big)\,,\\
h'_{\mathrm{I},2} &= \sqrt{\alpha}\cdot r'_2\cdot\sin\big(\theta+\theta'_2\big)\,,\\
&\vdots\\
h'_{\mathrm{R},t} &= \sqrt{\alpha}\cdot r'_t\cdot\cos\big(\theta+\theta'_t\big)\,,\\
h'_{\mathrm{I},t} &= \sqrt{\alpha}\cdot r'_t\cdot\sin\big(\theta+\theta'_t\big)\ \ ,
\end{aligned} \tag{3.25}$$

the channel vector $\mathbf{h}'$ can be equivalently transformed into $(\alpha,\, \theta,\, \mathbf{v}')$. After some manipulations, the determinant of the Jacobian matrix of the above transformation (3.25) is given by

$$\mathbf{det}\left[J\left(\frac{\partial\,\Big(h'_{\mathrm{R},1}, h'_{\mathrm{I},1}, \cdots, h'_{\mathrm{R},t}, h'_{\mathrm{I},t}\Big)}{\partial\,\Big(\alpha\,,\, \theta,\, r_2'^{\,2},\, \theta'_2,\, \cdots, r_t'^{\,2},\, \theta'_t\Big)}\right)\right] = \frac{1}{2}\;\alpha^{t-1}\;\prod_{i=2}^{t} r'_i\ . \tag{3.26}$$

Therefore, the probability density function of $\mathbf{h}'$ under the variable transformation (3.25) is given by

$$p\,(\alpha,\, \theta,\, \mathbf{v}') = \frac{\alpha^{t-1}\cdot\prod_{i=2}^{t} r'_i\cdot\exp\Big(-\alpha\cdot\mathbf{v}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathrm{h}'}^{-1}\,\mathbf{v}'\Big)}{2\,\pi^t\cdot\big|\mathbf{\Sigma}_{\mathrm{h}'}\big|}\ , \tag{3.27}$$

where $\mathbf{\Sigma}_{\mathrm{h}'} = \mathbf{P}^{\mathsf{H}}\mathbf{\Sigma}_{\mathrm{h}}\mathbf{P}$. It can be observed from the PDF expression (3.27) that θ is independent of α and $\mathbf{v}'$, and has uniform distribution $\theta \sim \mathcal{U}\,(0,\, 2\pi)$. Furthermore, due to the one-to-one mapping between $\mathbf{v}'$ and $\mathbf{v}$ under the constraint (3.14), θ is also independent of the constrained vector $\mathbf{v}$.

According to the above derivations, the joint density function of $(\alpha,\, \mathbf{v}')$ is therefore given by

$$p\,(\alpha,\, \mathbf{v}') = \frac{\alpha^{t-1}\cdot\prod_{i=2}^{t} r'_i\cdot\exp\Big(-\alpha\cdot\mathbf{v}'^{\mathsf{H}}\mathbf{\Sigma}_{\mathrm{h}'}^{-1}\,\mathbf{v}'\Big)}{\pi^{t-1}\cdot\big|\mathbf{\Sigma}_{\mathrm{h}'}\big|}\ . \tag{3.28}$$

For the sake of simplicity, the above PDF can be represented in a more efficient way by replacing vector $\mathbf{v}'$ by a redundant representation $\mathbf{v}$ subject to norm and phase constraint given by (3.14). After some manipulations, the determinant of the Jacobian matrix of the transformation $\mathbf{v} = \mathbf{P}\,\mathbf{v}'$, i.e.

$$(\, r'_2,\, \theta'_2,\, \cdots, r'_t,\, \theta'_t\,) \longrightarrow \left(\, v_{\mathrm{R},\,2},\, v_{\mathrm{I},\,2},\, \cdots,\, v_{\mathrm{R},\,t},\, v_{\mathrm{I},\,t}\,\right)\ , \tag{3.29}$$

can be obtained as

$$\mathbf{det}\left[J\left(\frac{\partial\, \left(\, v_{\mathrm{R},\,2},\, v_{\mathrm{I},\,2},\, \cdots,\, v_{\mathrm{R},\,t},\, v_{\mathrm{I},\,t}\right)}{\partial\ (\, r'_2,\, \theta'_2,\, \cdots, r'_t,\, \theta'_t\,)} \right)\right] = \left(\prod_{i=2}^{t} r'_i \right)^{-1}\ . \tag{3.30}$$

Therefore, the joint PDF of $(\alpha,\, \mathbf{v})$ is reduced to be the following form

$$p\,(\alpha,\, \mathbf{v}) = \ \frac{\alpha^{t-1} \cdot \exp\left(-\alpha \cdot \mathbf{v}^{\mathsf{H}} \mathbf{\Sigma}_{\mathrm{h}}^{-1}\, \mathbf{v}\right)}{\pi^{t-1} \cdot |\mathbf{\Sigma}_{\mathrm{h}}|}\ . \tag{3.31}$$

If the elements of vector $\mathbf{h}$ are i.i.d. Gaussian distributed, i.e. $\mathbf{\Sigma}_{\mathrm{h}} = I_t$, the joint PDF (3.31) is further reduced to be

$$p\,(\alpha,\, \mathbf{v}) = \left(\frac{\alpha^{t-1} \cdot \exp\,(-\alpha)}{(t-1)!} \right) \left(\frac{(t-1)!}{\pi^{t-1}} \right)\ , \tag{3.32}$$

where the first term corresponds to the marginal PDF of α, which is given by equation (3.15), and second term is the marginal PDF of $\mathbf{v}$, given by

$$p\,(\mathbf{v}) = \frac{(t-1)!}{\pi^{t-1}}\ . \tag{3.33}$$

Therefore, the directional vector $\mathbf{v}$ is uniformly distributed over the unit norm constrained space (3.14) and is independent of the power gain variable α of the MISO channel.

If the channel vector $\mathbf{h}$ is correlated, whose covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}$ has distinct positive eigen values, i.e. $\lambda_{\mathrm{h},1} > \cdots > \lambda_{\mathrm{h},t} > 0$, it is shown in [49] that the marginal distribution of random variable α is given by equation (3.18). By substituting the marginal pdf $p_\alpha(x)$ given by equation (3.18) into the joint pdf $p\,(\alpha,\, \mathbf{v})$ given by equation (3.31), the conditional distribution of $\mathbf{v}$ conditioned on $\mathbf{z}$ can therefore be obtained, and represented as equation (3.19). Moveover, by

integrating the joint pdf $p(\alpha, \mathbf{v})$ given by equation (3.31) w.r.t. to variable α over the region from zero to infinite, one can obtain the marginal pdf of random variable $\mathbf{v}$, which is given by equation (3.17). Similarly, by substituting the marginal pdf $p_{\mathbf{v}}(\mathbf{x})$ given by equation (3.17) into the joint pdf $p(\alpha, \mathbf{v})$ given by equation (3.31), the conditional distribution of α conditioned on $\mathbf{v}$ can therefore be obtained, which can be represented by equation (3.20).

■

3.5 Distortion Analysis for i.i.d. MISO Fading Channels

Under high resolution assumptions, a second order Taylor series expansion[2] is performed on the system distortion function given by (3.11), and the un-constrained sensitivity matrix $\mathbf{W}_{\alpha}(\widehat{\mathbf{v}})$ can be obtained as

$$\mathbf{W}_{\alpha}(\widehat{\mathbf{v}}) = \frac{\rho\alpha}{\ln 2 \cdot (1+\rho\alpha)} \cdot \Big(I - \mathbf{\Omega}\Big) \ . \tag{3.34}$$

According to the constraint condition given by (3.10), the derivative of function $\mathbf{g}(\mathbf{v})$ (Jacobian matrix) at point $\mathbf{v} = \widehat{\mathbf{v}}$ after singular value decomposition (SVD) has the following form

$$\frac{\partial}{\partial \mathbf{v}}\bigg|_{\mathbf{v}=\widehat{\mathbf{v}}} \mathbf{g}(\mathbf{v}) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \cdot \left[\begin{array}{c|c} \widehat{\mathbf{v}}_{\mathrm{R}} & \widehat{\mathbf{v}}_{\mathrm{I}} \\ \hline \widehat{\mathbf{v}}_{\mathrm{I}} & -\widehat{\mathbf{v}}_{\mathrm{R}} \end{array}\right]^{\mathsf{T}} = \mathbf{\Sigma}_{\mathrm{G}} \cdot \mathbf{V}_1^{\mathsf{T}} \ , \tag{3.35}$$

and satisfies the definition given by (2.91). Therefore, the constrained sensitivity matrix is given by

$$\mathbf{W}_{\mathrm{c},\,\alpha}(\widehat{\mathbf{v}}) = \frac{\rho\alpha}{\ln 2 \cdot (1+\rho\alpha)} \cdot \mathbf{V}_2^{\mathsf{T}}\Big(I - \mathbf{\Omega}\Big)\,\mathbf{V}_2 = \frac{\rho\alpha}{\ln 2 \cdot (1+\rho\alpha)} \cdot I_{2t-2} \ , \tag{3.36}$$

where $\mathbf{V}_2$ is an orthonormal column matrix such that $\mathbf{V}_{\mathrm{G}} = \big[\mathbf{V}_1\ \mathbf{V}_2\big]$ is a unitary matrix. This corresponds to the definition given by equation (2.95).

[2] In most communication systems, the quantization variables or objectives are complex vectors. With some modifications by utilizing the results of Wirtinger Calculus [47], the asymptotic distortion analysis can also be extended to complex distributed vectors without first transforming to the real domain.

It is evident from (3.36) that the sensitivity matrix $\mathbf{W}_{\mathrm{c},\,\alpha}\big(\widehat{\mathbf{v}}\big)$ is factorable, i.e.

$$\mathbf{W}_{\mathrm{c},\,\alpha}\big(\widehat{\mathbf{v}}\big) = f(\alpha)\cdot I_{2t-2}\,, \qquad f(\alpha) = \frac{\rho\alpha}{\ln 2\cdot(1+\rho\alpha)}\ , \tag{3.37}$$

which satisfies the sufficient condition of the achievability of the lower bound given by (2.57). Furthermore, the sensitivity matrix is also independent of its location $\widehat{\mathbf{v}}$, which corresponds to equation (2.26). Therefore, the normalized inertial profile is reduced to be the moment of inertia coefficient $I_{\mathrm{opt}}(\alpha)$. By substituting equation (3.36) into the definition of the normalized inertial profile given by (2.97), one obtains the following inertial profile

$$I_{\mathrm{c}}\Big(\widehat{\mathbf{v}}_i\,;\,\alpha\,;\,\mathbb{E}^{\mathrm{c}}_{\alpha,i}\Big) = V\big(\mathbb{E}^{\mathrm{c}}_{\alpha,i}\big)^{-t/(t-1)}\int_{\mathbf{e}\in\mathbb{E}^{\mathrm{c}}_{\alpha,i}}\frac{\rho\alpha}{\ln 2\cdot(1+\rho\alpha)}\cdot\|\mathbf{e}\|^2\ d\,\mathbf{e}\ . \tag{3.38}$$

As discussed in Section 2.6, the normalized inertial profile of any Voronoi shape $\mathbb{E}^{\mathrm{c}}_{\alpha,i}$ for a constrained source $\mathbf{v}$ is lower bounded by that of a "M-shaped" hyper-ellipsoid, which is a hyper-sphere in this case. Therefore, by substituting (3.37) into the lower bound given by (2.100), the optimal moment of inertia coefficient is tightly lower bounded (or approximated) by the following form

$$I_{\mathrm{c,\,opt}}\Big(\widehat{\mathbf{v}}_i\,;\,\alpha\Big) = I_{\mathrm{c,\,opt}}(\alpha)\ \gtrapprox\ \widetilde{I}_{\mathrm{c,\,opt}}(\alpha) = \frac{(t-1)\cdot\gamma_t^{-1/(t-1)}}{t}\cdot f\big(\alpha\big)\ , \tag{3.39}$$

where parameter γ_t is given by

$$\gamma_t = \frac{\pi^{t-1}}{(t-1)!}\ . \tag{3.40}$$

When the elements of the channel response $\mathbf{h}$ are i.i.d. Gaussian distributed, one can observe that α and $\mathbf{v}$ are statistically independent. This is shown in Section 3.4. Hence, the weighted constrained moment of inertia coefficient is given by

$$I^{\mathrm{w}}_{\mathrm{c,opt}} \gtrapprox \widetilde{I}^{\mathrm{w}}_{\mathrm{c,\,opt}} = \frac{(t-1)\cdot\gamma_t^{-1/(t-1)}}{t}\cdot E\Big[f(\alpha)\Big]\ , \tag{3.41}$$

which corresponds to equation (2.102). By substituting (3.41) into the distortion lower bound (2.101), the asymptotic capacity loss of a finite rate feedback MISO

beamforming system can be tightly lower bounded (or approximated) by the following form

$$D_{\text{c-Low,1}} \gtrapprox \widetilde{D}_{\text{c-Low,1}} = \frac{(t-1)\cdot 2^{-B/(t-1)}}{t}\cdot E\Big[f(\alpha)\Big] \cdot \int_{\mathbf{v}:\|\mathbf{v}\|=1,\,\theta_1=0} p(\mathbf{v})\cdot\Big(\gamma_t\cdot\lambda(\mathbf{v})\Big)^{-1/(t-1)}\, d\mathbf{v}\ , \tag{3.42}$$

where the integration of $\mathbf{v}$ is over the unit hyper-sphere with phase constraint $\theta_1 = 0$. By utilizing the obtained probability density function $p_{\mathbf{v}}(\mathbf{x})$ given in Section 3.4, one can derive the following results, (also shown in [25])

$$E\Big[f\big(\alpha\big)^k\Big] = \frac{\Gamma(k+t)\cdot\rho^k}{\ln 2^k\cdot\Gamma(t)}\cdot {}_2F_0\big(t+1,\,1;\,;\,-\rho\big)\ , \tag{3.43}$$

where ${}_2F_0$ is the generalized hypergeometric function. Therefore, by substituting (3.17) and (3.43) into equation (3.42) and after some manipulations, the final asymptotic capacity loss C_{L} of MISO system with finite rate feedback is given by

$$\widetilde{D}_{\text{c-Low,1}} = \frac{(t-1)\,2^{-B/(t-1)}}{\ln 2}\cdot\Big({}_2F_0\big(t+1,\,1;\,;\,-\rho\big)\cdot\rho\Big)\ , \tag{3.44}$$

with the optimal point density $\lambda^*(\mathbf{v})$ being a uniform distribution given by

$$\lambda^*(\mathbf{v}) = \gamma_t^{-1}\,, \qquad \mathbf{v}\in\Big\{\mathbf{v}\,\Big|\,\mathbf{g}(\mathbf{v})=0\Big\}\ . \tag{3.45}$$

Finally, note that for MISO channels with finite rate feedback, the factorable condition given by equation (3.37) is satisfied and the distortion lower bound $D_{\text{c-Low,1}}$ is hence achievable. Further due to the statistical independence between α and $\mathbf{v}$ of uncorrelated MISO channels, the asymptotic distortion lower bounds $D_{\text{Low,1}}$ and $D_{\text{Low,2}}$, upper bound D_{Upp}, as well as the distortion of the optimal quantizers D_{Opt} are all the same, which is described in (2.59). Hence, due to the fact that $\widetilde{D}_{\text{c-Low,1}}$ is tight, the obtained lower bound $\widetilde{D}_{\text{c-Low,1}}$ given by (3.44) is a good approximation of the asymptotic distortion of a system employing an optimal quantizer. Moreover, the obtained distortion lower bound $\widetilde{D}_{\text{c-Low,1}}$ is consistent with the capacity loss analysis provided in [25], which was obtained from a statistical approach. Distortion bounds by adopting the system SNR loss as the

performance metric, which is not shown here due the space limitation, can also be obtained by utilizing the proposed framework and can be further shown to be consistent with the SNR loss analysis provided in [23].

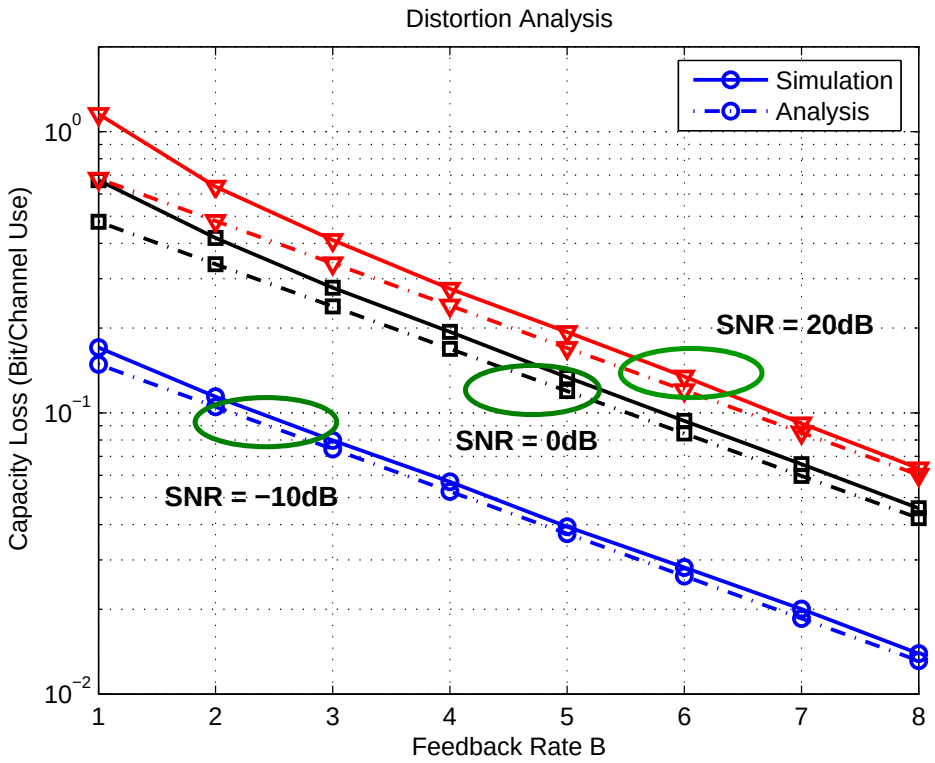

Figure 3.1: Capacity loss of a 3×1 MISO transmit beamforming system with finite rate feedback

Some numerical experiments were conducted to get a better feel for the utility of the bounds. Fig. 3.1 shows the capacity loss due to the finite rate quantization of the CSI versus feedback rate B for a 3×1 MISO system over i.i.d. Rayleigh fading channels under different system SNRs at $\rho = -10, 0$ and 20 dB, respectively. The simulation results are obtained from a MISO system using optimal CSI quantizers whose codebooks are generated by the mean squared weighted inner-product (MSwIP) criterion proposed in [25]. The analytical evaluations of the distortion lower bound $D_{\text{c-Low},1}$ provide by equation (3.44) are also included

in the plot for comparisons. It can be observed from the plot that the proposed distortion (or the capacity loss) lower bound is tight and predicts very well the actual system capacity loss obtained from Monte Carlo Simulations.

3.6 Distortion Analysis for Correlated MISO Fading Channels

In this section, the distortion (or capacity loss) analysis of MISO system with finite-rate feedback is further extended to correlated fading channels. The results provide interesting insights and demonstrate the effects of finite-rate CSI quantization and the channel correlation on system performance.

For correlated MISO fading channels $\mathbf{h} \sim \mathcal{N}_{\mathrm{c}}\big(\mathbf{0},\, \mathbf{\Sigma}_{\mathrm{h}}\big)$ with channel correlation matrix $\mathbf{\Sigma}_{\mathrm{h}}$ having distinct eigen-values, i.e. $\lambda_{\mathrm{h},1} > \cdots > \lambda_{\mathrm{h},\,t} > 0$, the high-resolution analysis for i.i.d. fading channels up to equation (3.39) is still valid. To restate the results, first of all, the constrained sensitivity matrix of the finite-rate quantized MISO beamforming system is given by

$$\mathbf{W}_{\mathrm{c},\,\alpha}\big(\widehat{\mathbf{v}}\big) = \frac{\rho\alpha}{\ln 2 \cdot (1+\rho\alpha)} \cdot I_{2t-2} \ . \tag{3.46}$$

Moreover, the optimal inertial profile is tightly lower bounded (or approximated) by the following form

$$\widetilde{I}_{\mathrm{c},\,\mathrm{opt}}\Big(\widehat{\mathbf{v}}\,;\,\alpha\Big) = \frac{(t-1)\cdot \gamma_t^{\,-\frac{1}{t-1}} \cdot \rho\,\alpha}{\ln 2 \cdot t \cdot (1+\rho\,\alpha)} \ . \tag{3.47}$$

Having obtained the inertial profile (3.47), one can then derive the following two distortion lower bounds.

3.6.1 Distortion lower bound $D_{\mathbf{c\text{-}Low},\,\mathbf{1}}$ for correlated channels

By substituting the conditional PDF $p_{\alpha|\mathbf{v}}(x)$ given by (3.20), the marginal PDF $p_{\mathbf{v}}(\mathbf{x})$ given by (3.17) and the inertial profile $\widetilde{I}_{\mathrm{c},\,\mathrm{opt}}\big(\mathbf{v}\,;\,\alpha\big)$ given by (3.47) into the distortion lower bound (2.39), the system asymptotic distortion lower bound

$\widetilde{D}_{\text{c-Low},1}$ can be expressed in the following form,

$$\widetilde{D}_{\text{c-Low},1}\left(\mathbf{\Sigma}_{\text{h}}\right) = \frac{\left(t-1\right)\gamma_t^{-\frac{t}{t-1}} \cdot \rho \cdot \beta_1\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)}{\ln 2 \cdot \left|\mathbf{\Sigma}_{\text{h}}\right|} \cdot 2^{-\frac{B}{t-1}} \quad , \tag{3.48}$$

where $\beta_1\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)$ is a constant coefficient that only depends on the number of antennas t, channel correlation matrix $\mathbf{\Sigma}_{\text{h}}$ and system SNR ρ. It is given by

$$\begin{aligned}&\beta_1\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)\\ &= \left(\int_{\mathbf{v}:\,\mathbf{g}(\mathbf{v})=0} \left(\left(\mathbf{v}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\,\mathbf{v}\right)^{-(t+1)} \cdot {}_2F_0\left(t+1,\, 1;\, ;\, -\frac{\rho}{\mathbf{v}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\,\mathbf{v}}\right)\right)^{\frac{t-1}{t}} d\,\mathbf{v}\right)^{\frac{t}{t-1}},\end{aligned} \tag{3.49}$$

with ${}_2F_0(\,;\,;\,)$ representing the generalized hypergeometric function. The optimal point density $\lambda^*\left(\mathbf{v}\right)$ that achieves the minimal distortion is given by

$$\begin{aligned}&\lambda^*\left(\mathbf{v}\right)\\ &= \beta_1\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)^{-\frac{t-1}{t}} \cdot \left(\left(\mathbf{v}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\,\mathbf{v}\right)^{-(t+1)} \cdot {}_2F_0\left(t+1,\, 1;\, ;\, -\frac{\rho}{\mathbf{v}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\,\mathbf{v}}\right)\right)^{\frac{t-1}{t}}.\end{aligned} \tag{3.50}$$

The evaluation of the coefficient β_1 is provided in the following subsections.

3.6.2 Distortion lower bound $D_{\text{c-Low},\,2}$ for correlated channels

Similarly, by substituting the conditional PDF $p_{\mathbf{v}|\alpha}(\mathbf{x})$ given by equation (3.19), the marginal PDF $p_\alpha(x)$ given by (3.17) and the inertial profile $\widetilde{I}_{\text{c},\,\text{opt}}\big(\mathbf{v}\,;\,\alpha\big)$ given by (3.47) into the distortion lower bound (2.45), the asymptotic distortion lower bound $\widetilde{D}_{\text{c-Low},2}$ can be expressed in the following form,

$$\widetilde{D}_{\text{c-Low},2}\left(\mathbf{\Sigma}_{\text{h}}\right) = \frac{\left((t-1)!\cdot\left|\mathbf{\Sigma}_{\text{h}}\right|\right)^{\frac{1}{t-1}} \cdot t^{\,t-1} \cdot \beta_2\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)}{\ln 2 \cdot (t-1)^{t-1}} \cdot 2^{-\frac{B}{t-1}} \quad , \tag{3.51}$$

where $\beta_2\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right)$ is a constant coefficient depends on system SNR ρ, number of antennas t and channel covariance matrix $\mathbf{\Sigma}_{\text{h}}$. It is given by

$$\beta_2\left(\rho,\, t,\, \mathbf{\Sigma}_{\text{h}}\right) = \int_0^\infty \frac{\rho}{1+\rho\,x} \cdot p_{\,\alpha}\left(\frac{x(t-1)}{t}\right)^{\frac{t}{t-1}} d\,x \ . \tag{3.52}$$

The evaluation of the coefficient β_2 is provided in the next subsection.

3.6.3 Interesting Observations of the Distortion Bounds

Based on the expressions for the average distortion lower bounds $\widetilde{D}_{\text{c-Low},1}\left(\mathbf{\Sigma}_{\text{h}}\right)$ and $\widetilde{D}_{\text{c-Low},2}\left(\mathbf{\Sigma}_{\text{h}}\right)$, the following observations can be made:

1. The asymptotic distortion lower bounds provided by equation (3.48) and (3.51) are described in a general format and are suitable for arbitrary channel correlations with covariance matrix $\mathbf{\Sigma}_{\text{h}}$. The average distortion of i.i.d. MISO channels is a special case where the covariance matrix $\mathbf{\Sigma}_{\text{h}}$ equals to the identity matrix. By substituting $\mathbf{\Sigma}_{\text{h}} = I_t$ into equation (3.49), the coefficient β_1 reduces to be the following form

$$\beta_1\left(\rho,\, t,\, I_t\right) = {}_2F_0\big(t+1,\, 1;\, ;\, -\rho\big) \cdot \gamma_t^{\frac{t}{t-1}} \ . \tag{3.53}$$

   Moreover, by substituting β_1 given by (3.53) into equation (3.48), the average system distortion lower bound $\widetilde{D}_{\text{c-Low},1}$ for i.i.d. MISO systems can be obtained as

$$\widetilde{D}_{\text{c-Low},1} = \left(\frac{t-1}{\ln 2} \cdot {}_2F_0\big(t+1,\, 1;\, ;\, -\rho\big) \cdot \rho\right) \cdot 2^{-\frac{B}{t-1}} \ . \tag{3.54}$$

   Similar derivations can be carried out by substituting $\mathbf{\Sigma}_{\text{h}} = I_t$ into equations (3.51) and (3.52). It can be shown that for i.i.d. fading channels, $D_{\text{c-Low},2}$ equals to $D_{\text{c-Low},1}$, which is given by equation (3.44). This result is consistent to the capacity loss analysis obtained in Section 3.5 as well as the results provided in [26].

2. Since the sensitivity matrix $\mathbf{W}_{\text{c},\,\alpha}\big(\widehat{\mathbf{v}}\big)$ given by (3.46) satisfies the factorable condition given by equation (2.57), the distortion lower bound $D_{\text{c-Low},1}$ is hence achievable and equal to the asymptotic distortions of the optimal quanizer, i.e.

$$D_{\text{c-Low},1} = D_{\text{Opt}} \gtrapprox \widetilde{D}_{\text{c-Low},1} = \widetilde{D}_{\text{Q-opt}} \ . \tag{3.55}$$

3. Both the distortion lower bounds $\widetilde{D}_{\text{c-Low},1}$ and $\widetilde{D}_{\text{c-Low},2}$ of correlated MISO channels, as well as the distortion of i.i.d. MISO channels, can be expressed

as a weighted exponential function given by

$$D = c \cdot 2^{-\frac{B}{t-1}} \ ,$$

where c is a constant coefficient that is independent of the quantization (feedback) rate B.

4. Due to the multi-dimensional integration required to evaluate the coefficient $\beta_1(\rho, t, \mathbf{\Sigma}_{\mathrm{h}})$ given by equation (3.49), the distortion lower bound $\widetilde{D}_{\text{c-Low,1}}$ lacks a closed-form expression and can only be evaluated through a Monte-Carlo simulation or a $(2t-2)$-dimensional numerical integration. Compared to the distortion lower bound $\widetilde{D}_{\text{c-Low,1}}$, $\widetilde{D}_{\text{c-Low,2}}$ (or the coefficient β_2) can be evaluated through a one-dimensional integration.

5. The distortion bounds of correlated MISO channels are smaller than that of the i.i.d. MISO channels, and satisfy the following the inequality

$$0 < \widetilde{D}_{\text{c-Low,2}}\big(\mathbf{\Sigma}_{\mathrm{h}}\big) \overset{a}{\leq} \widetilde{D}_{\text{c-Low,1}}\big(\mathbf{\Sigma}_{\mathrm{h}}\big) \overset{b}{\leq} \widetilde{D}_{\text{c-Low,1}}\big(I_t\big) \ . \tag{3.56}$$

with equality of (a) and (b) if and only if $\mathbf{\Sigma}_{\mathrm{h}} = I_t$. This means that i.i.d. channels are the worst channel to quantize in a sense of having the largest distortion (or capacity loss)[3]. This result is proved in the following proposition. Detailed comparisons of the above distortion bounds and the distortions of mismatched quantizers are provided in Section 3.7 and Section 4.2.

Proposition 3 *For a MISO system with finite-rate CSI feedback, the following orderings of the system distortions are valid for any correlated fading channels with covariance matrix* $\mathbf{\Sigma}_h$ *satisfying* $\mathbf{tr}\big(\mathbf{\Sigma}_h\big) = t$,

$$0 < \widetilde{D}_{\textit{c-Low,2}}\big(\mathbf{\Sigma}_{h}\big) \overset{a}{\leq} \widetilde{D}_{\textit{c-Low,1}}\big(\mathbf{\Sigma}_h\big) \overset{b}{\leq} \widetilde{D}_{\textit{c-Low,1}}\big(I_t\big) \ . \tag{3.57}$$

[3]This does not necessarily mean that correlated MISO channels have larger system capacities than i.i.d. channels. Since the capacity of i.i.d. MISO channels are better than that of correlated MISO channels with ideal CSI at the transmitter, the overall capacity of the finite-rate feedback-based MISO system still favors i.i.d. fading channels in the capacity sense.

Proof: First of all, inequality (a) can be proved easily according to the definition. In order to prove inequality (b), first notice that the uniform distribution is a sub-optimal solution of the point density function. Hence the resulting the average distortion is an upper bound of the optimal system distortion. By substituting the uniform point density function $\lambda(\mathbf{y})$ into the distortion integral (2.33), the averge system distortion can be upper bounded by

$$\widetilde{D}_{\text{c-Low,1}}\left(\mathbf{\Sigma}_{\text{h}}\right) \leq \widetilde{D}_{\text{upp}}\left(\mathbf{\Sigma}_{\text{h}}\right) = \frac{(t-1)\cdot\beta_5\left(\rho,\,\mathbf{\Sigma}_{\text{h}}\right)}{\ln 2\cdot t}\cdot 2^{-\frac{B}{t-1}} \quad , \tag{3.58}$$

where the constant coefficient $\beta_5\left(\rho,\,\mathbf{\Sigma}_{\text{h}}\right)$ is given by the following form

$$\beta_5\left(\rho,\,\mathbf{\Sigma}_{\text{h}}\right) = E\left[\frac{\rho\cdot\|\mathbf{h}\|^2}{1+\rho\cdot\|\mathbf{h}\|^2}\right] = E\left[\frac{\rho\cdot\mathbf{h}_0^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{h}_0}{1+\rho\cdot\mathbf{h}_0^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{h}_0}\right] \quad , \tag{3.59}$$

where vectors $\mathbf{h}$ and $\mathbf{h}_0$ have the following distribution

$$\mathbf{h}\sim\mathcal{N}_{\text{c}}\left(\mathbf{0},\mathbf{\Sigma}_{\text{h}}\right) \;, \qquad \mathbf{h}_0\sim\mathcal{N}_{\text{c}}\left(\mathbf{0},I_t\right) \;. \tag{3.60}$$

It is evident from equation (3.59) that β_5 is invariant under the following transformation,

$$\beta_5\left(\rho,\,\mathbf{\Sigma}_{\text{h}}\right) = \beta_5\left(\rho,\,\mathbf{U}\,\mathbf{\Sigma}_{\text{h}}\mathbf{U}^{\mathsf{H}}\right) \quad , \tag{3.61}$$

where $\mathbf{U}$ is any unitary matrix. Hence according to equation (3.61), if the unitary matrix $\mathbf{U}$ is set to be the eigenvectors of $\mathbf{\Sigma}_{\text{h}}$, we only need to focus our attention on the case where $\mathbf{\Sigma}_{\text{h}}$ is a diagonal matrix.

Furthermore, it is also true that β_5 is invariant to any permutations on the diagonal elements of $\mathbf{\Sigma}_{\text{h}}$,

$$\beta_5\left(\rho,\,\mathbf{\Sigma}_{\text{h}}\right) \overset{a}{=} \frac{1}{t!}\sum_{\mathbf{P}}\beta_5\left(\rho,\,\mathbf{P}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{P}\right) \overset{b}{\leq} \beta_5\left(\rho,\,\left(\frac{1}{t!}\sum_{\mathbf{P}}\mathbf{P}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{P}\right)\right) = \beta_5\left(\rho,\,I_t\right) \;, \tag{3.62}$$

where $\mathbf{P}$ is any permutation matrix, equality (a) follows the same reasoning as the invariant transformation (3.61), and (b) follows from the concavity property of function $f(x) = x/(1+x)$. At this point, by substituting the inequality (3.62) into the system distortion expression given by (3.58), inequality (b) of the distortion ordering given by (3.57) can be obtained.

■

3.6.4 Numerical and Simulation Examples

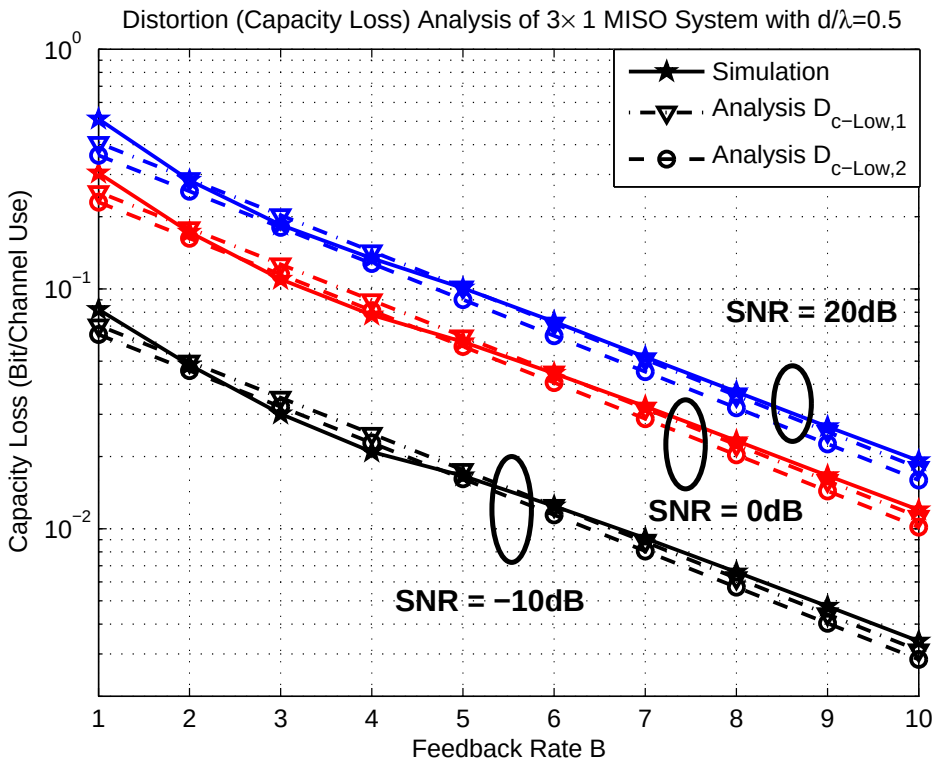

Figure 3.2: Capacity loss versus CSI feedback rate B of a 3×1 correlated MISO transmit beamforming system with normalized antenna spacing $D/\lambda = 0.5$, and signal to noise ratio $\rho = -10, 0$ and 20dB.

Some numerical experiments are now presented to provide a better feel for the utility of the bounds. Fig. 3.2 shows the capacity loss due to the finite-rate quantization of the CSI versus feedback rate B for a 3×1 MISO system over correlated Rayleigh fading channels under different system SNRs at $\rho = -10, 0$ and 20 dB, respectively. The spatially correlated channel is simulated by the correlation model in [50]: A linear antenna array with antenna spacing of half wavelength, i.e. $D/\lambda = 0.5$, uniform angular-spread in $[-30°, 30°]$ and angle of arrival $\phi = 0°$. The simulation results are obtained from a MISO system using optimal CSI quantizers whose codebooks are generated by the mean-squared weighted inner-

product (MSwIP) criterion proposed in [25]. The distortion lower bounds $\widetilde{D}_{\text{c-Low},1}$ and $\widetilde{D}_{\text{c-Low},2}$ given by equations (3.48) and (3.51) are also included in the plot for comparisons. It can be observed from the plot that the proposed distortion (or the capacity loss) lower bounds are tight and predict very well the actual system capacity loss obtained from Monte Carlo simulations.

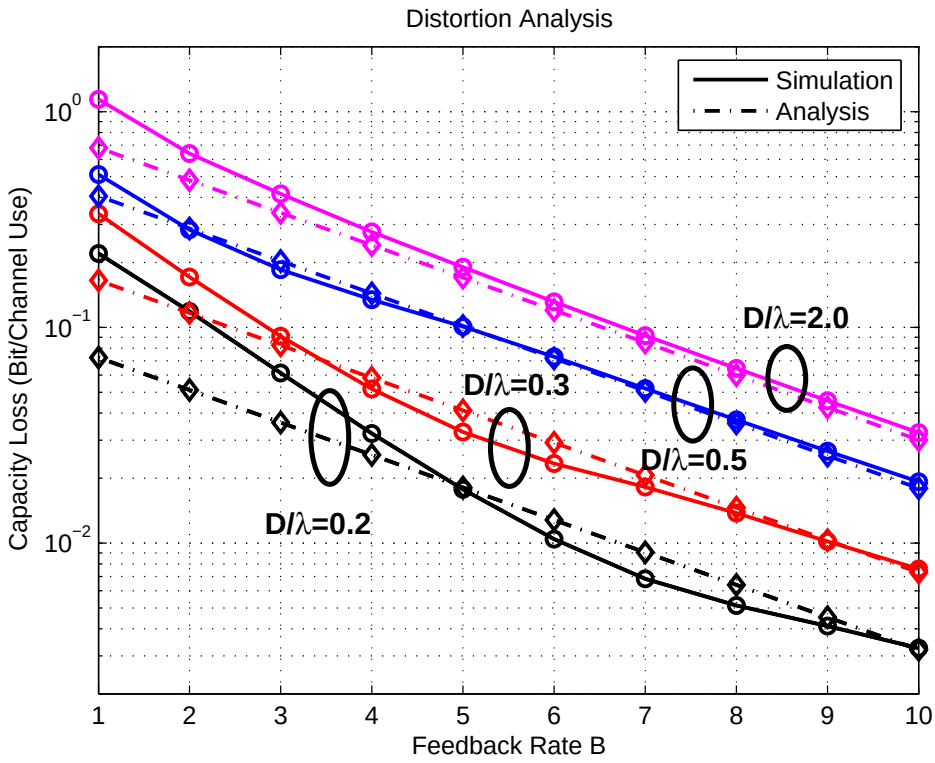

Figure 3.3: Capacity loss versus CSI feedback rate B of a 3×1 correlated MISO transmit beamforming system with normalized antenna spacing $D/\lambda = 0.2, 0.3, 0.5, 2.0$, and signal to noise ratio $\rho = 20$dB.

In order to see the effects of channel correlation on CSI quantization in a MISO system, we show in Fig. 3.3 the curves of capacity loss versus quantization rate (both simulation and analytical lower bound $\widetilde{D}_{\text{c-Low},1}$) of the same MISO system under different channel correlations obtained with adjacent antenna spacing $D/\lambda = 0.2, 0.3, 0.5, 2.0$ at SNR $\rho = 20$dB. As a comparison to uncorrelated MISO

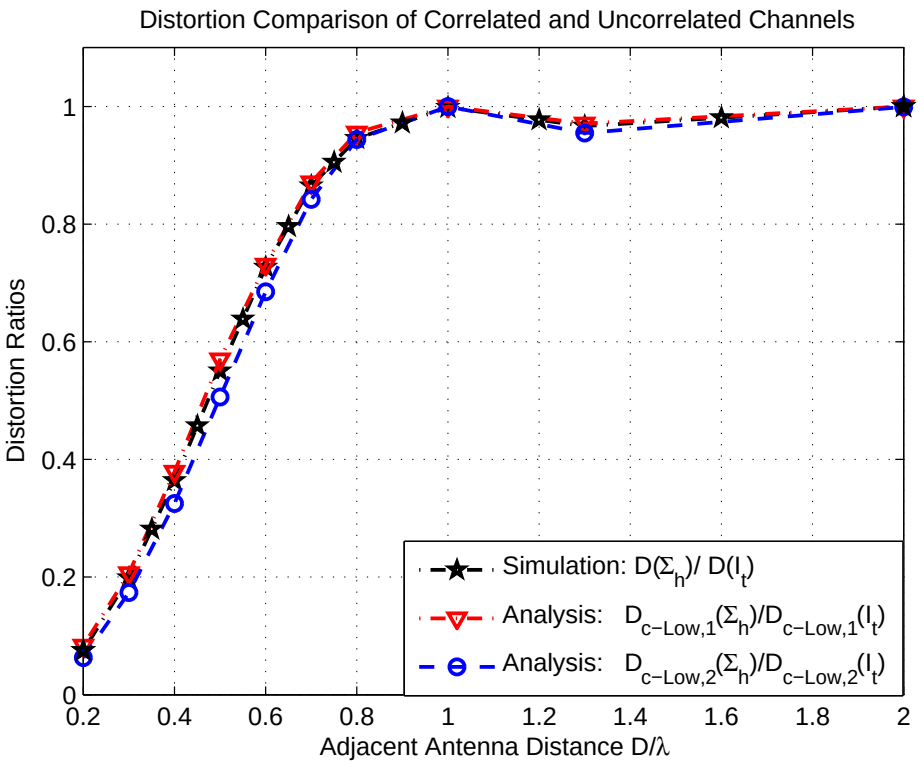

Figure 3.4: Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) versus transmit antenna spacing D/λ of a 3×1 MISO transmit beamforming system with signal to noise ratio $\rho = 5$dB under CSI feedback rate $B = 10$ bits.

channels, we also show in Fig. 3.4 the ratio of the distortion for correlated MISO channels over the distortion for i.i.d. fading channels with quantization rate $B = 10$ bits, signal to noise ratio $\rho = 5$dB, and under different channel correlations. It can be observed from the plot that the system distortion of correlated MISO channels is strictly less than that of the i.i.d. channels and the analytical result agree well with the actual simulation results.

3.7 Distortion Analysis in High-SNR and Low-SNR Regimes

3.7.1 High-SNR Distortion Analysis

In high SNR regimes, the constrained sensitivity matrix $\mathbf{W}_{\mathrm{c},\,\alpha}$ reduces to be

$$\mathbf{W}_{\mathrm{c},\,\alpha}^{\text{H-snr}}\left(\mathbf{v},\,\alpha\right)=\lim_{\rho\to\infty}\ \frac{\rho\alpha}{\ln 2\cdot(1+\rho\alpha)}\cdot I=\frac{I}{\ln 2}\ , \tag{3.63}$$

which is independent of $\mathbf{v}$, the side information information α as well as the SNR ρ. This means that 1) the encoder can discard the available side information α without any loss of system performance; 2) one single codebook is used for different system SNRs in high SNR regions. In this case, the inertial profile $\widetilde{I}_{\mathrm{opt}}(\mathbf{v},\alpha)$ and the average inertial profile $\widetilde{I}_{\mathrm{opt}}^{\mathrm{w}}(\mathbf{v},\alpha)$ also reduce to be a constant independent of the location $\mathbf{v}$ as well as side information α

$$\widetilde{I}_{\mathrm{opt}}^{\text{H-snr}}=\widetilde{I}_{\mathrm{opt}}^{\mathrm{w},\,\text{H-snr}}=\frac{(t-1)\cdot\gamma_t^{\,-\frac{1}{t-1}}}{\ln 2\cdot t}\ . \tag{3.64}$$

By substituting (3.64) into the distortion lower bound given by (2.39), the system capacity loss of i.i.d. MISO channels in high SNR regime is given by

$$\widetilde{D}_{\text{c-Low},1}^{\text{H-snr}}\big(\mathbf{\Sigma}_{\mathrm{h}}=I_t\big)\ =\ \frac{t-1}{t}\cdot 2^{-\frac{B}{t-1}}\ , \tag{3.65}$$

which is consistent with the analysis obtained in [25] based on a statistical approach. For correlated MISO fading channels, by substituting the average inertia profile $\widetilde{I}_{\mathrm{opt}}^{\mathrm{w},\,\text{H-snr}}$ given by equation (3.64) as well as the marginal pdf $p_{\mathbf{v}}\left(\mathbf{x}\right)$ given by (3.17) into the distortion bound given by (2.39), lower bound $\widetilde{D}_{\text{c-Low},1}^{\text{H-snr}}\left(\mathbf{\Sigma}_{\mathrm{h}}\right)$ can be represented by the following form,

$$\widetilde{D}_{\text{c-Low},1}^{\text{H-snr}}\left(\mathbf{\Sigma}_{\mathrm{h}}\right)=\left(\frac{(t-1)\cdot\left(\prod_{i=1}^{\mathsf{T}}\lambda_{\mathrm{h},\,i}\right)^{\frac{1}{t-1}}}{\ln 2\cdot t}\cdot\left(\beta_6\right)^{\frac{t}{t-1}}\right)\cdot 2^{-\frac{B}{t-1}}\ , \tag{3.66}$$

where the coefficient β_6 is given by

$$\beta_6=E\left[\ \frac{\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{h}}{\mathbf{h}^{\mathsf{H}}\,\mathbf{h}}\ \right]\ , \tag{3.67}$$

which is a ratio of Gaussian quadratic variables. The moments of ratios of random variables, including central quadratic forms in normal variables, were investigated in [51], and the results can described as the following integration

$$E\left[\left(\frac{X}{Y}\right)^n\right] = \Gamma(n)^{-1}\int_0^\infty v^{n-1} M_{X,Y}^{(n)}(0,-v)\; d\, v \;, \tag{3.68}$$

where $M_{X,Y}(u,v)$ is the joint moment generating function (m.g.f.) of random variables X and Y, and $M_{X,Y}^{(n)}(0,-v)$ stands for $\partial^n M_{X,Y}(u,-v)/\partial v^n$ evaluated at $u=0$. Therefore, by setting $X = \mathbf{h}^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{h}$ and $Y = \mathbf{h}^{\mathsf{H}}\,\mathbf{h}$, the joint m.g.f. of variables X and Y can be represented as

$$M_{X,Y}(u,v) = \frac{1}{\mathbf{det}\Big(I - \big(u\cdot I + v\cdot \boldsymbol{\Sigma}_{\mathrm{h}}\big)\Big)} = \left(\prod_{k=1}^{t}\left(1-u-v\cdot\lambda_{\mathrm{h},k}\right)\right)^{-1} \quad . \tag{3.69}$$

By substituting the joint m.g.f. given by equation (3.69) into the moments integration function (3.68), coefficient β_6 has the following closed-form expression

$$\beta_6 = (t-1)\sum_{i=1}^{t}\frac{\big(\ln\lambda_{\mathrm{h},\,i}\big)/\lambda_{\mathrm{h},\,i}}{\prod_{k\neq i}\big(1-\lambda_{\mathrm{h},\,k}/\lambda_{\mathrm{h},\,i}\big)} \;. \tag{3.70}$$

Finally, by substituting (3.70) into equation (3.66), distortion lower bound $\widetilde{D}_{\text{c-Low},1}^{\text{H-snr}}\left(\boldsymbol{\Sigma}_{\mathrm{h}}\right)$ can be shown to have the following closed-form expression,

$$\begin{aligned}&\widetilde{D}_{\text{c-Low},1}^{\text{H-snr}}\left(\boldsymbol{\Sigma}_{\mathrm{h}}\right)\\ &= \frac{(t-1)\cdot\Big(\prod_{i=1}^{\mathsf{T}}\lambda_{\mathrm{h},\,i}\Big)^{\frac{1}{t-1}}}{\ln 2\cdot t}\cdot\left((t-1)\sum_{i=1}^{t}\frac{\big(\ln\lambda_{\mathrm{h},\,i}\big)/\lambda_{\mathrm{h},\,i}}{\prod_{k\neq i}\big(1-\lambda_{\mathrm{h},\,k}/\lambda_{\mathrm{h},\,i}\big)}\right)^{\frac{t}{t-1}}\cdot 2^{-\frac{B}{t-1}} \;.\end{aligned} \tag{3.71}$$

3.7.2 Low-SNR Distortion Analysis

In low SNR regimes, i.e., $\rho\to 0$, the constrained sensitivity matrix $\mathbf{W}_{\mathrm{c},\,\alpha}$ reduces to be

$$\mathbf{W}_{\mathrm{c},\,\alpha}^{\text{L-snr}} = \lim_{\rho\to 0}\;\frac{2\rho\alpha}{\ln 2\cdot(1+\rho\alpha)}\cdot I = \frac{2\,\rho\alpha}{\ln 2}\cdot I \;. \tag{3.72}$$

Therefore, the inertial profile $\widetilde{I}_{\text{opt}}(\mathbf{v},\alpha)$ and the average inertial profile $\widetilde{I}^{\text{w}}_{\text{opt}}(\mathbf{v},\alpha)$ are given by

$$\widetilde{I}^{\text{L-snr}}_{\text{opt}}(\mathbf{v},\alpha) = \frac{(t-1)\cdot\gamma_t^{-\frac{1}{t-1}}\cdot\rho\alpha}{\ln 2\cdot t}\ , \qquad \widetilde{I}^{\text{w, L-snr}}_{\text{opt}}(\mathbf{v},\alpha) = \frac{(t-1)\cdot\gamma_t^{-\frac{1}{t-1}}\cdot\rho}{\ln 2\cdot\left(\mathbf{v}^{\mathsf{H}}\boldsymbol{\Sigma}_{\text{h}}^{-1}\mathbf{v}\right)}\ . \tag{3.73}$$

Similarly, by substituting (3.73) into the distortion lower bound given by (2.39), the MISO system capacity loss in low SNR regimes over both i.i.d. and correlated fading channels can be represented as:

$$\widetilde{D}^{\text{L-snr}}_{\text{c-Low,1}}\left(\boldsymbol{\Sigma}_{\text{h}}=I_t\right) = \frac{(t-1)\,\rho}{\ln 2}\cdot 2^{-\frac{B}{t-1}}\ , \tag{3.74}$$

$$\widetilde{D}^{\text{L-snr}}_{\text{c-Low,1}}\left(\boldsymbol{\Sigma}_{\text{h}}\right) = \frac{(t-1)\,\rho\cdot\gamma_t^{-\frac{t}{t-1}}\cdot\beta_3\left(t,\,\boldsymbol{\Sigma}_{\text{h}}\right)}{\ln 2\cdot\left|\boldsymbol{\Sigma}_{\text{h}}\right|}\cdot 2^{-\frac{B}{t-1}}\ , \tag{3.75}$$

where $\beta_3\left(t,\,\boldsymbol{\Sigma}_{\text{h}}\right)$ is a constant coefficient given by

$$\beta_3\left(t,\,\boldsymbol{\Sigma}_{\text{h}}\right) = \left(\int_{\mathbf{v}\,:\,\mathbf{g}(\mathbf{v})=0}\left(\mathbf{v}^{\mathsf{H}}\boldsymbol{\Sigma}_{\text{h}}^{-1}\,\mathbf{v}\right)^{-\frac{t^2-1}{t}}\,d\,\mathbf{v}\right)^{\frac{t}{t-1}}\ . \tag{3.76}$$

Moveover, when there are a large number of transmit antennas, the high-dimensional approximation of the distortion lower bound $\widetilde{D}^{\text{L-snr}}_{\text{c-Low,1}}$ can be represented by the following closed-form expression (obtained after some manipulation of equation (3.74))

$$\widetilde{D}^{\text{L-snr, H-dim}}_{\text{c-Low,1}} = \frac{\rho\cdot(t-1)\cdot\left(\prod_{i=1}^{t}\lambda_{\text{h},\,i}\right)^{\frac{1}{t-1}}}{\ln 2}\cdot 2^{-\frac{B}{t-1}}\ . \tag{3.77}$$

3.8 Distortion Comparisons between i.i.d. and Correlated channels

Through numerical evaluations, the second product term in the R.H.S of the equation (3.71) is found to be close to 1 in most cases leading to the following approximate relationship

$$\frac{\widetilde{D}^{\text{H-snr, H-dim}}_{\text{c-Low,1}}\left(\boldsymbol{\Sigma}_{\text{h}}\right)}{\widetilde{D}^{\text{H-snr, H-dim}}_{\text{c-Low,1}}\left(I_t\right)} = \frac{\widetilde{D}^{\text{L-snr, H-dim}}_{\text{c-Low,1}}\left(\boldsymbol{\Sigma}_{\text{h}}\right)}{\widetilde{D}^{\text{L-snr, H-dim}}_{\text{c-Low,1}}\left(I_t\right)} \approx \eta\left(\boldsymbol{\Sigma}_{\text{h}}\right)\ , \tag{3.78}$$

where the constant coefficient $\eta(\mathbf{\Sigma}_{\mathrm{h}})$ is given by

$$\eta(\mathbf{\Sigma}_{\mathrm{h}}) = \frac{\left(\prod_{i=1}^{\mathsf{T}} \lambda_{\mathrm{h},i}\right)^{\frac{1}{t}}}{\sum_{i=1}^{t} \lambda_{\mathrm{h},i}/t} \leq 1 \ , \tag{3.79}$$

and represents the relative capacity loss of quantizing a correlated MISO channel as compared to that of an i.i.d. MISO channel in high-SNR and low-SNR regimes with large number of antennas. This means that 1) the ratio of the geometric mean over the arithmetic mean of the eigen-values of the channel covariance matrix is a key parameter that characterizes the system performance; 2) the capacity loss of a MISO system with finite-rate CSI feedback is proportional to this ratio.

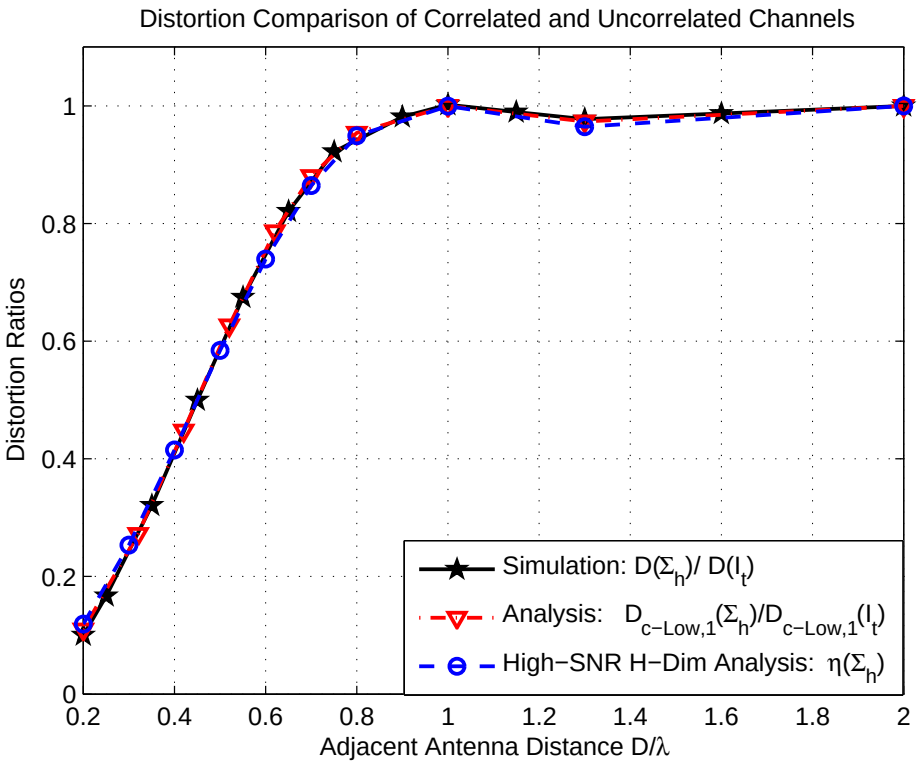

Figure 3.5: Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) versus transmit antenna spacing D/λ of a 3×1 MISO transmit beamforming system in high-SNR regime with $\rho = 20$dB.

As a numerical example, we show in Fig. 3.5 the normalized capacity loss

(distortion ratio of correlated MISO channels over i.i.d. fading channels) versus antenna spacing D/λ in high-SNR regimes with $\rho = 20$dB and quantization rate $B = 10$ bits. For comparison purpose, the ratio of the distortion lower bound, i.e. $\widetilde{D}_{\text{c-Low},1}(\mathbf{\Sigma}_{\text{h}})/\widetilde{D}_{\text{c-Low},1}(I_t)$, as well as its high-SNR and high-dimensional approximation $\eta(\mathbf{\Sigma}_{\text{h}})$ given by equation (3.79) are also included in the plot. Interestingly, it can be observed from Fig. 3.5 that the obtained high-dimensional approximation of the distortion ratio agree well with the simulation results even for cases with a small number of antennas $t = 3$.

3.9 Summary

This chapter employed high resolution quantization theory to study the effects of finite-rate quantization of the CSI on the performance of MISO systems over both i.i.d. and correlated fading channels. To be specific, tight lower bounds on the capacity loss of correlated MISO systems due to the finite-rate channel quantization were provided. Interestingly, in high-SNR and low-SNR regimes, the capacity loss of correlated MISO channels was shown to be related to that of i.i.d. fading channels by a simple multiplicative factor which is given by the ratio of the geometric mean to the arithmetic mean of the eigenvalues of the channel covariance matrix. The text of this chapter is in part a reprint of the paper which was coauthored with Bhaskar D. Rao and has been submitted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels"*.

4 Analysis of MISO CSI Quantizers with Mismatched Codebooks and Transformed Codebooks

4.1 Motivation

In Chap. 3, the analysis of MISO systems with finite-rate feedback were derived under the assumption that both the encoder and the decoder have perfect knowledge of the source distribution, distortion function, and are using the most efficient quantization algorithm. This is clearly not always true as practical constraints often result in approximations and various types of suboptimal choices in the design of feedback-based wireless communication systems. These suboptimal choices often result in various types of mismatches. Mismatched conditions arise in practical situations due to various reasons such as enabling reduced design and encoding complexity, imperfect knowledge of the source distribution, among others. As an extended application of the distortion analysis provided in Chap. 2, capacity analysis of multiple antenna systems using sub-optimal CSI quantizers is provided in this chapter. Specifically, two types of mismatched MISO CSI quantizers are investigated, which include quantizers that are designed with minimum mean square error (MMSE) criterion but the desired measure is ergodic capacity

loss (i.e. mismatched design criterion), and quantizers whose codebooks are designed with a mismatched channel covariance matrix (i.e. mismatched statistics). Moreover, MISO systems over spatially correlated fading channels with CSI quantizers using transformed codebooks are also investigated. Finally, the performance of these sub-optimal channel quantization schemes is further compared to that of the optimal CSI quantizers.

4.2 Mismatched Analysis of Quantized MISO Beamforming Systems

As an application of the mismatched analysis provided in Section 2.4, this section provides a capacity loss analysis of a finite-rate feedback-based MISO beamforming system when the CSI quantizer is mismatched and suboptimal. This is in contrast to the distortion analysis provided in Chap. 3 of MISO systems with optimal CSI quantizers wherein the codebook and the encoding algorithm were designed to perfectly match the distortion function as well as the source distribution. Imperfect codebook and suboptimal quantizer are quite prevalent in practice which makes this study interesting. Mismatched conditions may arise in practical situations due to various reasons such as enabling reduced design and encoding complexity, imperfect knowledge of the source distribution, among others.

4.2.1 Dimensionality Mismatch and Quantization Criterion Mismatch

In this subsection, we present the analysis of a suboptimal (mismatched) quantizer that directly quantizes the CSI using the mean-squared error (MSE) as the distortion measure. The results illustrate the importance of encoding the appropriate parameters as well as the distortion function of interest.

For an MMSE channel quantizer, the channel state information $\mathbf{h}$ is directly quantized and results in a conventional vector quantization problem with the source variable having $2t$ free (real) dimensions and with no encoder side in-

formation. The corresponding distortion function of the MMSE channel quantizer is given by

$$D_{\text{mis-R}}\big(\mathbf{h},\,\widehat{\mathbf{h}}\,\big) \;=\; \|\mathbf{h}-\widehat{\mathbf{h}}\|^2 \;, \tag{4.1}$$

whose sensitivity matrix $\mathbf{W}_{\text{mis-R}}$ is given by $\mathbf{W}_{\text{mis-R}} = I_{2t}$. At the transmitter, the unit norm beamforming vector $\widehat{\mathbf{v}}$ is obtained by normalizing the quantized channel vector $\widehat{\mathbf{h}}$, i.e. $\widehat{\mathbf{v}} = \widehat{\mathbf{h}}/\|\widehat{\mathbf{h}}\|$. Hence, the actual system distortion function (or the capacity loss) can be expressed in terms of vectors $\mathbf{h}$ and $\widehat{\mathbf{h}}$ as

$$D_{\text{Q-R}}\big(\mathbf{h},\,\widehat{\mathbf{h}}\,\big) = \log_2\left(1+\rho\cdot\|\mathbf{h}\|^2\right) - \log_2\left(1+\rho\cdot\frac{|\langle\,\mathbf{h},\,\widehat{\mathbf{h}}\,\rangle|^2}{\|\widehat{\mathbf{h}}\|^2}\right) \;, \tag{4.2}$$

Its corresponding sensitivity matrix can be shown to have the following form

$$\mathbf{W}\big(\widehat{\mathbf{h}}\,\big) = \frac{\rho}{\ln 2\cdot\left(1+\rho\cdot\|\widehat{\mathbf{h}}\|^2\right)}\cdot\Big(I-\boldsymbol{\Omega}\Big) \;, \tag{4.3}$$

where matrix $\boldsymbol{\Omega}\in\mathbb{R}^{2k_{\text{q}}\times 2k_{\text{q}}}$ is given by

$$\boldsymbol{\Omega} = \left[\begin{array}{c|c} \widehat{\mathbf{v}}_{\text{R}}\widehat{\mathbf{v}}_{\text{R}}^{\text{T}}+\widehat{\mathbf{v}}_{\text{I}}\widehat{\mathbf{v}}_{\text{I}}^{\text{T}} & \widehat{\mathbf{v}}_{\text{R}}\widehat{\mathbf{v}}_{\text{I}}^{\text{T}}-\widehat{\mathbf{v}}_{\text{I}}\widehat{\mathbf{v}}_{\text{R}}^{\text{T}} \\ \hline \widehat{\mathbf{v}}_{\text{I}}\widehat{\mathbf{v}}_{\text{R}}^{\text{T}}-\widehat{\mathbf{v}}_{\text{R}}\widehat{\mathbf{v}}_{\text{I}}^{\text{T}} & \widehat{\mathbf{v}}_{\text{R}}\widehat{\mathbf{v}}_{\text{R}}^{\text{T}}+\widehat{\mathbf{v}}_{\text{I}}\widehat{\mathbf{v}}_{\text{I}}^{\text{T}} \end{array}\right] \;. \tag{4.4}$$

The MMSE channel quantizer being analyzed suffers from two types of mismatches: 1) The quantizer is designed to quantize a redundant channel state information vector $\mathbf{h}$ of dimensions $2t$ instead of $2t-2$ in the optimal quantizer, which leads to a dimensionality mismatch; 2) The quantizer uses a mismatched distortion function $D_{\text{mis-R}}$ given by (4.1) as compared to D_{Q} given by equation (3.11). Since the MMSE codebook is designed to match the mismatched sensitivity matrix $\mathbf{W}_{\text{mis-R}}$, the Voronoi region of the MMSE quantizer is close to a hyper-sphere of dimension $2t$, which leads to a sub-optimal point density given by [33]

$$\lambda_{\text{mis-R}}(\mathbf{h}) = p\,(\mathbf{h})^{\frac{t}{t+1}}\cdot\left(\int p\,(\mathbf{h})^{\frac{t}{t+1}}\,d\,\mathbf{h}\right)^{-1} \;, \tag{4.5}$$

where $p(\mathbf{y})$ is the PDF of the MISO channel impulse response $\mathbf{h}$. Furthermore, from equation (2.67), the suboptimal MMSE quantizer also leads to a mismatched normalized inertial profile given by,

$$I_{\text{mis-R}}(\mathbf{h}) = \frac{(t-1)(t!)^{\frac{1}{t}}\,\rho}{\ln 2\cdot(t+1)\cdot\pi\cdot\left(1+\rho\,\|\mathbf{h}\|^2\right)} \;. \tag{4.6}$$

By substituting equations (4.5) and (4.6) into the asymptotic distortion integration given by (2.33), the average system distortion of a mismatched MMSE channel quantizer can be represented by the following form

$$\widetilde{D}_{\text{mis-R-Low,1}}\left(\mathbf{\Sigma}_{\text{h}}\right) = \frac{(t-1)\cdot\left(t\,!\cdot\left|\mathbf{\Sigma}_{\text{h}}\right|\right)^{\frac{1}{t}}\cdot\left(\frac{t+1}{t}\right)^{t}\cdot\beta_4\left(\rho,\,t,\,\mathbf{\Sigma}_{\text{h}}\right)}{\ln 2\cdot t}\cdot 2^{-\frac{B}{t}}\quad, \tag{4.7}$$

where $\beta_4\left(\rho,\,t,\,\mathbf{\Sigma}_{\text{h}}\right)$ is a constant coefficient given by

$$\beta_4 = E\left[\frac{\rho}{1+\rho\cdot(t+1)/t\cdot\mathbf{h}^{\mathsf{H}}\,\mathbf{h}}\right]\ . \tag{4.8}$$

Moreover, it can be shown that coefficient β_4 has a analytically closed-form expression. To see this, first note that β_4 can be viewed as the first order moment of a ratio of Gaussian quadratic variables, i.e. $\beta_4 = E\big[X/Y\big]$. By taking a similar approach which is utilized in Section 3.7.1, one can set $X=\rho$ and $Y = 1+\rho\cdot(t+1)/t\cdot\mathbf{h}^{\mathsf{H}}\,\mathbf{h}$, whose joint m.g.f. is given by

$$\begin{aligned} M_{X,Y}(u,v) &= \frac{\exp(u\,\rho+v)}{\mathbf{det}\Big(I-\rho\cdot(t+1)/t\cdot v\cdot\mathbf{\Sigma}_{\text{h}}\Big)} \\ &= \exp(u\,\rho+v)\cdot\left(\prod_{k=1}^{t}\left(1-\rho\cdot\frac{t+1}{t}\cdot v\cdot\lambda_{\text{h},k}\right)\right)^{-1}. \end{aligned} \tag{4.9}$$

By substituting the joint m.g.f. (4.9) into the integration (3.68) provided in [51], closed-form expression of coefficient β_4 can be obtained,

$$\begin{aligned} \beta_4\left(\rho,\,t,\,\mathbf{\Sigma}_{\text{h}}\right) = -\frac{t}{t+1}\sum_{i=1}^{t}\ &\left(\lambda_{\text{h},\,i}\prod_{k\neq i}\left(1-\frac{\lambda_{\text{h},\,k}}{\lambda_{\text{h},\,i}}\right)\right)^{-1} \\ &\exp\left(\frac{t}{\rho\,(t+1)\lambda_{\text{h},\,i}}\right)\cdot E_{\text{i}}\left(-\frac{t}{\rho\,(t+1)\lambda_{\text{h},\,i}}\right), \end{aligned} \tag{4.10}$$

with $E_{\text{i}}(\cdot)$ representing the exponential integral function.

It can be observed from (4.7) that the system distortion of the mismatched MMSE channel quantizer decays slower (with slope $-1/t$ in the exponent) than that of the optimal quantizer (with slope $-1/(t-1)\big)$. This is a significant system performance degradation especially for systems with a small number of antennas, emphasizing the importance of choosing an appropriate CSI quantization scheme.

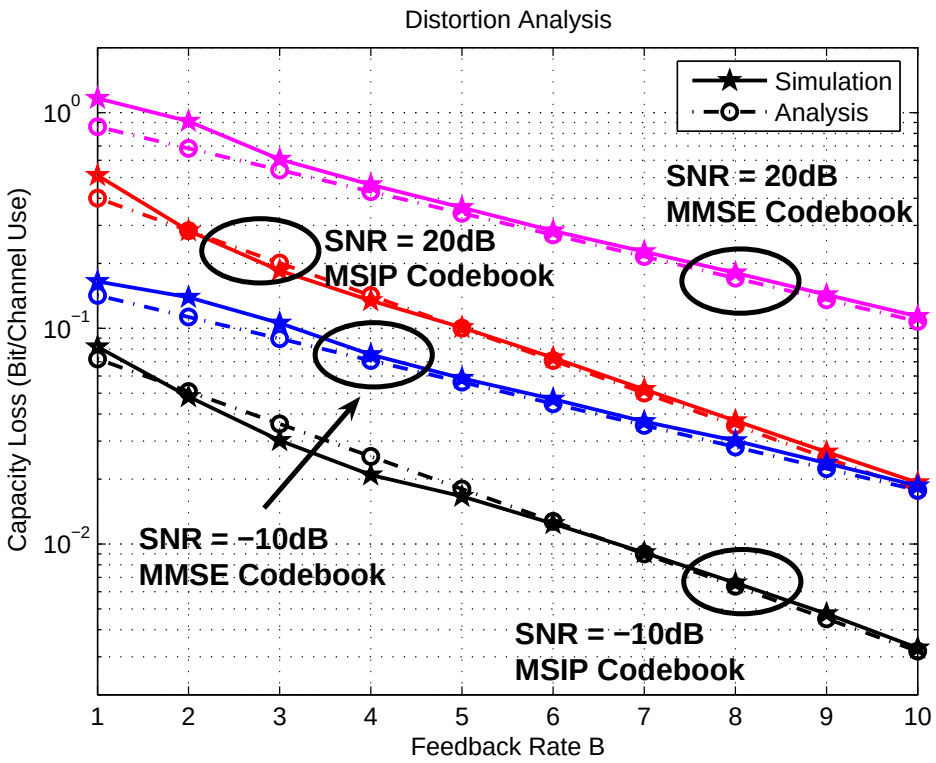

Figure 4.1: Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs MMSE quantizer).

In order to get a better understanding of the degradation caused by the mismatched MMSE channel quantizer, we plot in Fig. 4.1 the capacity loss due to the finite-rate CSI quantization versus feedback rate B for a 3×1 MISO system over correlated fading channels with adjacent antenna spacing $D/\lambda = 0.5$ and different system SNRs of $\rho = -10$, and 20 dB respectively. Codebooks are designed by using both the optimal mean-squared weighted inner-product (MSwIP) criterion proposed in [27] and the simple MMSE criterion mentioned in this section. The analytical evaluations of the system distortion lower bound $\widetilde{D}_{\text{c-Low},1}$ provide by (3.48) and the mismatched distortion $\widetilde{D}_{\text{mis-R-Low},1}$ provided by (4.7) are also included in the plot for comparisons. It can be observed from the plot that

the system performance is significantly degraded by the mismatched quantizer, especially for systems with small number of antennas. Moreover, the proposed distortion analysis is tight and matches very well the actual system capacity loss obtained from Monte Carlo simulations.

4.2.2 Source Distribution Mismatch (or Point Density Mismatch)

For the correlated MISO channels, the channel distribution depends on the covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}$, which needs to be estimated and is subject to estimation error. Moreover, it is also practically infeasible to redesign codebooks for every $\mathbf{\Sigma}_{\mathrm{h}}$, store them and use them adaptively. Therefore, in practical situations, only very limited codebooks are available and so the mismatched channel covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}}$ will cause performance degradation.

Based on the mismatched covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}}$, a sub-optimal codebook is generated with the mismatched point density given by (from equation (3.50))

$$\lambda_{\text{mis-P}}\left(\mathbf{v}\right) = \beta_1\left(\rho,\, t,\, \mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}}\right)^{-\frac{t-1}{t}} \cdot\left(\left(\mathbf{v}^{\mathsf{H}}\left(\mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}}\right)^{-1}\,\mathbf{v}\right)^{-(t+1)} {}_2F_0\left(t+1,\, 1;\, ;\, -\frac{\rho}{\mathbf{v}^{\mathsf{H}}\left(\mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}}\right)^{-1}\,\mathbf{v}}\right)\right)^{\frac{t-1}{t}} . \quad (4.11)$$

By substituting the mismatched point density $\lambda_{\text{mis-P}}$ given by (4.11) into the distortion integral (2.73), the system distortion lower bound of the source-distribution-mismatched quantizer can be obtained as

$$\widetilde{D}_{\text{mis-P-Low,1}} \;=\; \left(\int_{\mathbf{x}:\;\mathbf{g}(\mathbf{x})=0} \widetilde{I}^{\mathrm{w}}_{\mathrm{c,\,opt}}\big(\mathbf{x}\big)\cdot p_{\mathbf{v}}\left(\mathbf{x}\right)\cdot\lambda_{\text{mis-P}}\big(\mathbf{x}\big)^{-\frac{1}{t-1}}\, d\,\mathbf{x}\right)\cdot 2^{-\frac{B}{t-1}} \;. \quad (4.12)$$

As a special case, if the codebook designed for i.i.d. MISO channels is used for correlated MISO systems[1], i.e. $\mathbf{\Sigma}_{\mathrm{h}}^{\mathrm{m}} = I_t$, the mismatched point density $\lambda_{\mathrm{mis}}(\mathbf{v})$ is uniform and the asymptotic distortion of the mismatched quantizer can be represented as the following form

$$\widetilde{D}_{\text{mis-P-Low,1}}\left(\mathbf{\Sigma}_{\mathrm{h}}\right) = \frac{(t-1)\cdot\beta_5\left(\rho,\,\mathbf{\Sigma}_{\mathrm{h}}\right)}{\ln 2\cdot t}\cdot 2^{-\frac{B}{t-1}} \;, \quad (4.13)$$

[1]This can be also viewed as the case where the channel covariance matrix is completely unavailable at both the transmitter and the receiver, and hence one single codebook is used for any channel correlation.

where the constant coefficient $\beta_5\left(\rho,\, \mathbf{\Sigma}_{\rm h}\right)$ is given by

$$\beta_5 = E\left[\frac{\rho\cdot \mathbf{h}^{\mathsf{H}}\,\mathbf{h}}{1+\rho\cdot \mathbf{h}^{\mathsf{H}}\,\mathbf{h}}\right]\,, \tag{4.14}$$

which is also a ratio of Gaussian quadratic random variables. It is evident from equation (4.14) that β_5 is related to β_4 given by equation (4.8) through the following connection

$$\beta_5\left(\rho, \mathbf{\Sigma}_{\rm h}\right) = 1-\frac{1}{\rho}\,\beta_4\left(\frac{t\,\rho}{t+1}, \mathbf{\Sigma}_{\rm h}\right)\,. \tag{4.15}$$

By substituting the results given by (4.10) into equation (4.15), β_5 can be expressed as the following closed-form expression

$$\beta_5\left(\rho,\, \mathbf{\Sigma}_{\rm h}\right) = 1+\sum_{i=1}^{t}\left(\rho\,\lambda_{{\rm h},\,i}\prod_{j\neq i}\left(1-\frac{\lambda_{{\rm h},\,j}}{\lambda_{{\rm h},\,i}}\right)\right)^{-1}\cdot\exp\left(\frac{1}{\rho\,\lambda_{{\rm h},\,i}}\right)\cdot E_{\rm i}\left(\frac{-1}{\rho\,\lambda_{{\rm h},\,i}}\right). \tag{4.16}$$

As a numerical example, we demonstrate in Fig. 4.2 the capacity loss due to the finite-rate CSI quantization versus feedback rate B for the same 3×1 MISO system over correlated fading channels with adjacent antenna spacing $D/\lambda = 0.5$ and different system SNRs at $\rho = -10$, and 20 dB, respectively. Both the optimal codebooks with correct channel covariance matrix as well as the mismatched i.i.d. codebooks are employed for simulation. The analytical evaluations of the distortion lower bound $\widetilde{D}_{\text{c-Low,1}}$ provide by (3.48) and the mismatched distortion $\widetilde{D}_{\text{mis-P-Low,1}}$ provided by (4.13) are also included in the plot for comparison. It can be observed from the plot that the system performance is degraded by the mismatched i.i.d. codebook but with the same exponential decaying factor $2^{-B/(t-1)}$. Moreover, the proposed distortion analysis closely matches the system capacity loss obtained from simulations.

4.2.3 Comparisons with Other Channel Quantizers

In order to understand how the mismatched channel covariance matrix $\left(\mathbf{\Sigma}_{\rm h}^{\rm m} = I_t\right)$ affects the MISO system performance, a distortion comparison between optimal and mismatched quantizers under both correlated and i.i.d. fading

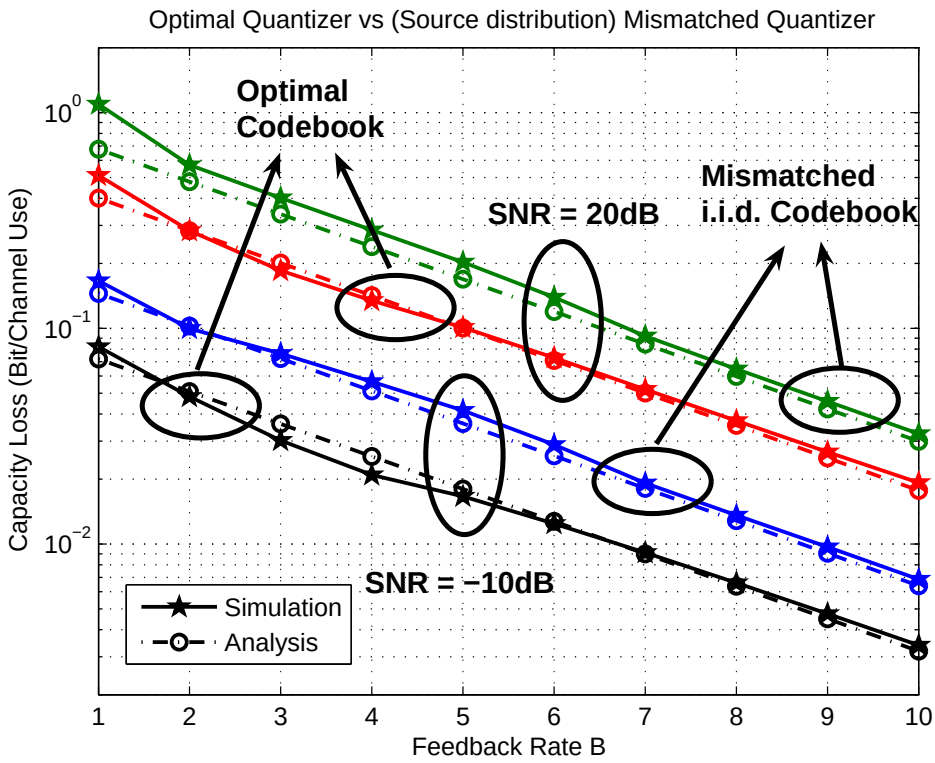

Figure 4.2: Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs Mismatched codebook for i.i.d. fading channels).

channels is formed. To be specific, we first denote $\widetilde{D}_{\text{mis-P-Low, 1}}(\mathbf{\Sigma}_{\text{h}})$ as the average distortion of a mismatched quantizer using i.i.d. codebook in a correlated environments with channel covariance matrix $\mathbf{\Sigma}_{\text{h}}$, whereas $\widetilde{D}_{\text{c-Low,1}}(\mathbf{\Sigma}_{\text{h}})$ represents the system distortion of an optimal quantizer with codebook designed to match the same correlated MISO fading channel. The following proposition establishes the relations between the two system distortions.

Proposition 4 *For a MISO system with finite-rate CSI feedback, the following inequality of the system distortions is valid for any correlated fading channels with*

covariance matrix $\mathbf{\Sigma}_h$ *satisfying* $\mathbf{tr}\big(\mathbf{\Sigma}_h\big) = t$,

$$0 < \widetilde{D}_{\textit{mis-P-Low,1}}\big(\mathbf{\Sigma}_h\big) \leq \widetilde{D}_{\textit{c-Low,1}}\big(I_t\big) \ . \tag{4.17}$$

Moreover, the mismatched system distortion $\widetilde{D}_{\textit{mis-P-Low, 1}}\left(\mathbf{\Sigma}_h\right)$ *converges to the distortion of i.i.d. MISO channels with optimal quantizers in high-SNR and low-SNR regimes, i.e.*

$$\lim_{\rho \to 0, \infty} \frac{\widetilde{D}_{\textit{mis-P-Low,1}}\big(\mathbf{\Sigma}_h\big)}{\widetilde{D}_{\textit{c-Low,1}}\big(I_t\big)} = 1 \ . \tag{4.18}$$

Proof: First, note that the asymptotic distortion of the mismatched quantizer can be represented as the following form after some manipulations

$$\widetilde{D}_{\text{mis-P-Low,1}}\left(\mathbf{\Sigma}_{\text{h}}\right) = \frac{(t-1)\cdot \beta_7\left(\rho,\, \mathbf{\Sigma}_{\text{h}}\right)}{\ln 2 \cdot t} \cdot 2^{-\frac{B}{t-1}} \ , \tag{4.19}$$

where the constant coefficient $\beta_7\left(\rho,\, \mathbf{\Sigma}_{\text{h}}\right)$ is given by the following form

$$\beta_7\left(\rho,\, \mathbf{\Sigma}_{\text{h}}\right) = E\left[\frac{\rho \cdot \|\mathbf{h}\|^2}{1+\rho\cdot\|\mathbf{h}\|^2}\right] = E\left[\frac{\rho \cdot \mathbf{h}_0^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{h}_0}{1+\rho\cdot\mathbf{h}_0^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{h}_0}\right] \ , \tag{4.20}$$

where vectors $\mathbf{h}$ and $\mathbf{h}_0$ have the following distribution

$$\mathbf{h} \sim \mathcal{N}_{\text{c}}\left(\mathbf{0}, \mathbf{\Sigma}_{\text{h}}\right) \ , \qquad \mathbf{h}_0 \sim \mathcal{N}_{\text{c}}\left(\mathbf{0}, I_t\right) \ . \tag{4.21}$$

It is evident from equation (4.20) that β_7 is invariant under the following transformation,

$$\beta_7\left(\rho,\, \mathbf{\Sigma}_{\text{h}}\right) = \beta_7\left(\rho,\, \mathbf{U}\,\mathbf{\Sigma}_{\text{h}}\mathbf{U}^{\mathsf{H}}\right) \ , \tag{4.22}$$

where $\mathbf{U}$ is any unitary matrix. Hence according to equation (4.22), if the unitary matrix $\mathbf{U}$ is set to be the eigenvectors of $\mathbf{\Sigma}_{\text{h}}$, we only need to focus our attention on the case where $\mathbf{\Sigma}_{\text{h}}$ is a diagonal matrix.

Furthermore, it is also true that β_7 is invariant to any permutations on the diagonal elements of $\mathbf{\Sigma}_{\text{h}}$,

$$\beta_7\left(\rho,\, \mathbf{\Sigma}_{\text{h}}\right) \overset{a}{=} \frac{1}{t!}\sum_{\mathbf{P}} \beta_7\left(\rho,\, \mathbf{P}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{P}\right) \overset{b}{\leq} \beta_7\left(\rho,\, \left(\frac{1}{t!}\sum_{\mathbf{P}} \mathbf{P}^{\mathsf{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{P}\right)\right) = \beta_7\left(\rho,\, I_t\right) \ , \tag{4.23}$$

where $\mathbf{P}$ is any permutation matrix, equality (a) follows the same reasoning as the invariant transformation (4.22), and (b) follows from the concavity property of function $f(x) = x/(1+x)$. At this point, by substituting the inequality (4.23) into the system distortion expression given by (4.19), inequality (4.17) can be obtained.

According to the definition of β_7 given by (4.20), the following equations can be obtained

$$\lim_{\rho\to 0} \beta_7\left(\rho,\, \mathbf{\Sigma}_{\mathrm{h}}\right) = \rho\, t\,, \qquad\qquad \lim_{\rho\to \infty} \beta_7\left(\rho,\, \mathbf{\Sigma}_{\mathrm{h}}\right) = 1\ \ , \tag{4.24}$$

which further lead to the convergence of the system distortions given by equation (4.18). ■

The above results mean that: 1) The capacity loss of a correlated MISO channel by using the mismatched quantizer is larger than that of the optimal quantizer, but still less than that of an uncorrelated MISO channel even with optimal codebook. 2) The performance of the mismatched quantizer is strongly affected by the sub-optimality caused by the mismatched codebook. In high-SNR and low-SNR regimes, mismatched CSI quantizers using i.i.d. codebooks will lead to the same "worst" system distortion $\widetilde{D}^{\text{H-snr}}_{\text{c-Low},1}\big(I_t\big)$ regardless of the actual fading channel correlations, or channel covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}$.

We plot in Fig. 4.3 the normalized capacity loss (distortion ratio over i.i.d. fading channels) of a correlated 3×1 MISO system versus antenna spacing D/λ with the mismatched i.i.d. codebooks, and with system SNR $\rho = -10, 20$ dB and quantization rate $B = 10$ bits. For comparison purpose, the ratio of the average distortion of the same MISO system and at the same correlated channel conditions but with optimal codebooks are also included in the plot. The curves provided in Fig. 4.3 further confirm the two observations made in the previous paragraph.

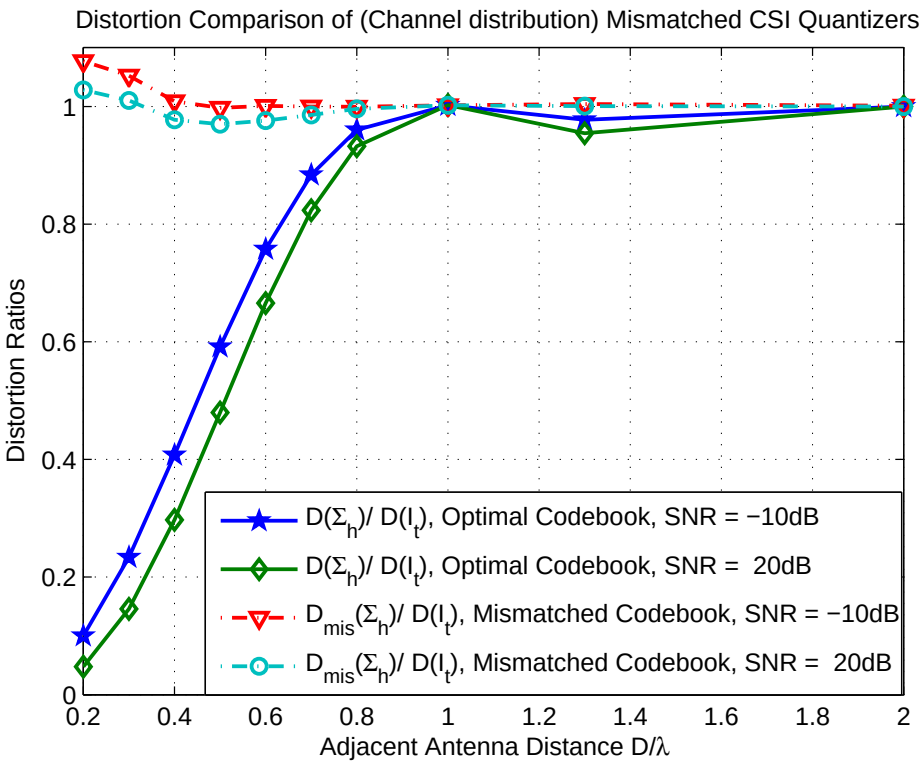

Figure 4.3: Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) comparison of a 3×1 MISO transmit beamforming with optimal and mismatched codebooks versus antenna spacing $d = D/\lambda$, in high and low SNR regimes ($\rho = -10$ and 20 dB).

4.3 Analysis of MISO Channel Quantizers with Transformed Codebook

In practically situations, the spatial correlation conditions of the fading channel responses may change during transmission process. However, for a real system, it is impossible to design different codebooks optimized for every instantiation of the channel covariance matrix and it might also be infeasible for the transmitter and receiver to store a large amount of codebooks and use them adaptively. In these cases, it is convenient to use a channel quantizer whose codebook is gener-

ated from a fixed pre-designed codebook through a transformation parameterized by the channel covariance matrix.

4.3.1 Problem Formulation

To be specific, suppose $\mathcal{C}_0$ is the optimal codebook designed for the i.i.d. MISO fading channels. When the elements of the fading channel response $\mathbf{h}$ are correlated, i.e. $\mathbf{h} \sim \mathcal{N}_{\mathrm{c}}(\mathbf{\Sigma}_{\mathrm{h}})$, it is evident that codebook $\mathcal{C}_0$ is no longer optimal. In order to compensate the mismatch between $\mathcal{C}_0$ and the current channel statistics, a transformed codebook $\mathcal{C}$ can be generated by the following manner,

$$\mathcal{C} = \left\{ \mathbf{F}\big(\widehat{\mathbf{v}}\big) \;\middle|\; \widehat{\mathbf{v}} \in \mathcal{C}_0 \right\} \quad , \tag{4.25}$$

where $\mathbf{F}(\cdot)$ is a general non-linear transformation that depends on the channel statistics. Optimization of the transformation $\mathbf{F}(\cdot)$ turns out to be difficult, and hence a simple sub-optimal transformation,

$$\mathbf{F}(\widehat{\mathbf{v}}) = \frac{\mathbf{G}\,\widehat{\mathbf{v}}}{\|\mathbf{G}\,\widehat{\mathbf{v}}\|} \quad , \tag{4.26}$$

was proposed in [23] [52] where $\mathbf{G} \in \mathbb{C}^{t\times t}$ is a fixed matrix depends on the channel covariance matrix $\mathbf{\Sigma}_{\mathrm{h}}$. In the next subsection, distortion analysis of CSI-quantizers with transformed codebooks is provided.

4.3.2 Distortion Analysis of Transformed Codebooks

First of all, according to the codebook transformation given by (4.26), the transformed point density function $\lambda_{\text{c-tr}}(\mathbf{v})$ can be obtained as the following form, from equation (2.76),

$$\lambda_{\text{c-tr}}(\mathbf{v}) = \gamma_t^{-1} \cdot |\mathbf{\Sigma}|^{-1} \cdot \left(\mathbf{v}^{\mathsf{H}}\mathbf{\Sigma}^{-1}\,\mathbf{v}\right)^{-t} \quad , \qquad \mathbf{\Sigma} = \mathbf{G}\cdot\mathbf{G}^{\mathsf{H}} \quad . \tag{4.27}$$

which is equivalent to the PDF of a unit-norm complex vector $\mathbf{x}/\|\mathbf{x}\|$ with $\mathbf{x}$ having complex Gaussian distribution $\mathbf{x} \sim \mathcal{N}_{\mathrm{c}}(\mathbf{0},\, \mathbf{\Sigma})$. It is evident that the transformed point density given by (4.27) does not match to the optimal point density function

$\lambda^*(\mathbf{v})$ given by (3.50) in the general case. However, for MISO systems with a large number of antennas and in high-SNR and low-SNR regimes, it can be shown that the optimal point density $\lambda^*(\mathbf{v})$ reduces to be the source distribution $p_{\mathbf{v}}(\mathbf{x})$ given by the following form

$$\lim_{t\to\infty}\ \lambda^*(\mathbf{x}) = p_{\mathbf{v}}(\mathbf{x}) = \gamma_t^{-1}\cdot|\mathbf{\Sigma}_{\mathrm{h}}|^{-1}\cdot\left(\mathbf{x}^{\mathsf{H}}\mathbf{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{x}\right)^{-t}\ . \tag{4.28}$$

In this case, by choosing matrix $\mathbf{G}$ as a product $\mathbf{G} = \mathbf{U}\,\mathbf{\Lambda}^{\frac{1}{2}}$ with matrices $\mathbf{U}$ and $\mathbf{\Lambda}$ form the eigen-value decomposition of the channel covariance matrix $\mathbf{\Sigma}_{\mathrm{h}} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^{\mathsf{H}}$, one can generate a transformed codebook with optimal point density $\lambda_{\text{c-tr}}(\mathbf{v}) \approx \lambda^*(\mathbf{v})$. Hence, there is no distortion loss caused by the point density mismatch, though the system also suffers from the oblongitis of the Voronoi shape.

By substituting the transformation given by (4.26) into equation (2.81), the inertial profile of the transformed codebook with sub-optimal encoder $\mathcal{Q}_{\text{sub}}$ (or encoding process) can be represented as

$$\widetilde{I}_{\text{c-sub}}\left(\mathbf{v}\,;\,\alpha\right) = \frac{\gamma_t^{-\frac{1}{t-1}}\cdot\rho\alpha\cdot\left(\mathbf{v}^{\mathsf{H}}\,\mathbf{\Sigma}^{-1}\,\mathbf{v}\right)}{t\cdot\ln 2\cdot(1+\rho\alpha)}\cdot\mathbf{tr}\Big(\left(I-\mathbf{v}\,\mathbf{v}^{\mathsf{H}}\right)\cdot\mathbf{\Sigma}\Big) \geq \widetilde{I}_{\text{c-opt}}\left(\mathbf{v}\,;\,\alpha\right)\ . \tag{4.29}$$

where $\widetilde{I}_{\text{c-opt}}\left(\mathbf{v}\,;\,\alpha\right)$ is the optimal inertia profile given by equation (3.39). It is evident from (4.29) that except unitary rotations of the i.i.d. codebook, any non-trivial transformation of the codebook will lead to mismatched Voronoi shape and hence causes inertial profile loss. Therefore, a codebook transformation that compromises both the point density loss and the inertial profile loss is favored.

Finding the optimal codebook transformation $\mathbf{F}$ that minimizes the system distortion turns out to be a difficult problem. In this chapter, instead of optimizing the overall distortion w.r.t. matrix $\mathbf{G}$, we provide a distortion analysis of MISO systems with transformed CSI-quantizers using codebooks generated by the heuristic choice $\mathbf{\Sigma}_{\mathrm{h}} = \mathbf{G}\cdot\mathbf{G}^{\mathsf{H}}$ (or[2] $\mathbf{G} = \mathbf{U}\,\mathbf{\Lambda}^{\frac{1}{2}}$). To be specific, by substituting the transformed point density (4.27) and transformed inertia profile (4.29) into

[2]Note that the codebook transformation is not unique. Any right unitary rotation $\mathbf{G}\cdot\mathbf{P}$ on matrix $\mathbf{G}$, with $\mathbf{P}\cdot\mathbf{P}^{\mathrm{H}} = I$, can generate another codebook transformation (or codebook) with the same performance.

distortion integral (2.89), the corresponding upper and lower bounds of the average system distortion of a MISO CSI-quantizer with transformed codebook can be expressed in the following forms

$$\widetilde{D}_{\text{c-tr-Low}} = \frac{(t-1)\cdot|\mathbf{\Sigma}_{\text{h}}|^{\frac{1}{t-1}}}{\ln 2\cdot t}\cdot E\left[\frac{\rho\cdot\left(\mathbf{h}^{\text{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\mathbf{h}\right)^{\frac{t}{t-1}}}{\left(1+\rho\cdot\|\mathbf{h}\|^2\right)\cdot\|\mathbf{h}\|^{\frac{2}{t-1}}}\right]\cdot 2^{-\frac{B}{t-1}}\ , \tag{4.30}$$

$$\widetilde{D}_{\text{c-tr-Upp}} = \frac{|\mathbf{\Sigma}_{\text{h}}|^{\frac{1}{t-1}}}{\ln 2\cdot t}\cdot E\left[\frac{\rho\cdot\left(\mathbf{h}^{\text{H}}\mathbf{\Sigma}_{\text{h}}^{-1}\mathbf{h}\right)^{\frac{2t-1}{t-1}}\cdot\left(t\cdot\|\mathbf{h}\|^2-\mathbf{h}^{\text{H}}\mathbf{\Sigma}_{\text{h}}\,\mathbf{h}\right)}{\left(1+\rho\cdot\|\mathbf{h}\|^2\right)\cdot\|\mathbf{h}\|^{\frac{4t-2}{t-1}}}\right]\cdot 2^{-\frac{B}{t-1}}\ . \tag{4.31}$$

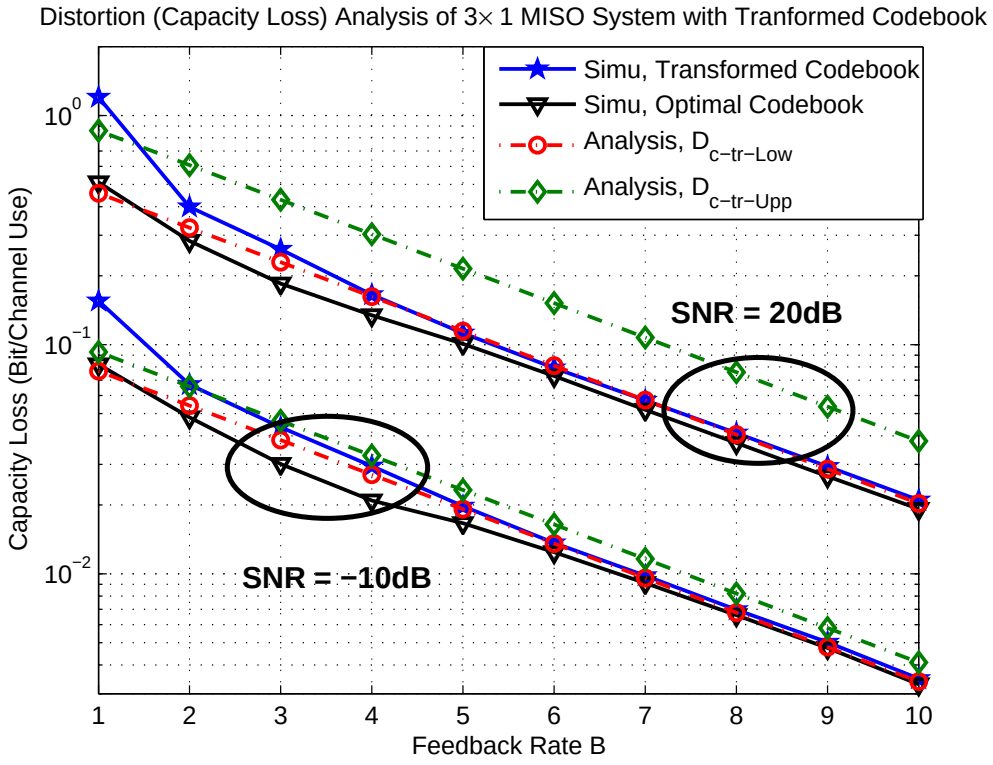

Figure 4.4: Capacity loss of a 3×1 correlated MISO system with normalized antenna spacing $d = D/\lambda = 0.5$ versus CSI feedback rate B using different channel quantization codebooks (Optimal codebook vs Transformed codebook).

Some numerical experiments were conducted to get a better feel for the

utility of the bounds. Fig. 4.4 shows the system capacity loss due to the finite-rate quantization of the CSI versus feedback rate B for a 3×1 MISO system over correlated Rayleigh fading channels under different system SNRs at $\rho = -10$ and 20 dB, respectively. The spatially correlated channel is simulated by the correlation model in [50]: A linear antenna array with antenna spacing of half wavelength, i.e. $D/\lambda = 0.5$, uniform angular-spread in $[-30°, 30°]$ and angle of arrival $\phi = 0°$. Simulation results of both the optimal designed codebook using the minimal mean-squared weighted inner product (MSwIP) criterion proposed in [26] as well as the sub-optimal transformed codebook are plotted. For comparison purpose, distortion lower bound $\widetilde{D}_{\text{c-tr-Low}}$ given by (4.30) and the distortion upper bound $\widetilde{D}_{\text{c-tr-Upp}}$ given by (4.31) are also included in the plot. It can be observed from Fig. 4.4 that the distortion lower bound $\widetilde{D}_{\text{c-tr-Low}}$ is tight and the performance of the CSI quantizer with transformed codebook is close to that of the optimal codebooks.

4.3.3 Comparison with Optimal MISO CSI-Quantizers

In order to see sub-optimality caused by codebook transformation, one would like to compare the system performance in terms of the average distortion of quantizers using transformed codebooks with that of the optimally designed codebooks. Interestingly, in high-SNR and low-SNR regimes with a large number transmit antennas t, the average system distortion of CSI quantizers with transformed codebook can be upper and lower bounded by some multiplicative factors of the optimal quantization distortion.

Proposition 5 *For MISO systems with a large number of transmit antennas, i.e. $t \to \infty$, the following inequalities are satisfied*

$$1 \overset{a}{=} \frac{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-tr-Low}}}{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-Low,1}}} \leq \frac{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-tr}}}{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-Low,1}}} \leq \frac{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-tr-Upp}}}{\widetilde{D}^{\textit{H-dim, H-SNR}}_{\textit{c-Low}}} \overset{b}{\leq} c_1 \ , \tag{4.32}$$

$$1 \overset{a}{=} \frac{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-tr-Low}}}{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-Low,1}}} \leq \frac{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-tr}}}{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-Low,1}}} \leq \frac{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-tr-Upp}}}{\widetilde{D}^{\textit{H-dim, L-SNR}}_{\textit{c-Low,1}}} \overset{b}{\leq} c_2 \ , \tag{4.33}$$

where constant coefficients c_1 and c_2 are given by

$$c_1 = \left(\frac{\delta(t-2)}{\lambda_{h,1} \cdot \lambda_{h,2}} - (t-1)(t-2) \sum_{i=1}^{t} \frac{\big(\ln \lambda_{h,i}\big)/\lambda_{h,i}^2}{\prod_{k \neq i} \big(1 - \lambda_{h,k}/\lambda_{h,i}\big)} \right) \Big/ c_1^{\frac{t}{t-1}} , \tag{4.34}$$

$$c_2 = (t-1) \sum_{i=1}^{t} \frac{\big(\ln \lambda_{h,i}\big)/\lambda_{h,i}}{\prod_{k \neq i} \big(1 - \lambda_{h,k}/\lambda_{h,i}\big)} \ . \tag{4.35}$$

Proof: First of all, in high-SNR regimes, distortion bounds $\widetilde{D}_{\text{c-tr-Low}}^{\text{H-dim, H-SNR}}$ and $\widetilde{D}_{\text{c-tr-Upp}}^{\text{H-dim, H-SNR}}$ can be represented as the following forms

$$\widetilde{D}_{\text{c-tr-Low}}^{\text{H-dim, H-SNR}} = \left(\frac{(t-1) \cdot |\mathbf{\Sigma}_{\text{h}}|^{\frac{1}{t-1}} \cdot \beta_2}{\ln 2 \cdot t} \right) \cdot 2^{-\frac{B}{t-1}} \approx \widetilde{D}_{\text{c-Low,1}}^{\text{H-dim, H-SNR}} , \tag{4.36}$$

$$\widetilde{D}_{\text{c-tr-Upp}}^{\text{H-dim, L-SNR}} \leq \left(\frac{(t-1) \cdot |\mathbf{\Sigma}_{\text{h}}|^{\frac{1}{t-1}} \cdot \beta_3}{\ln 2 \cdot t} \right) \cdot 2^{-\frac{B}{t-1}} = \left(\beta_3 \cdot \beta_2^{-\frac{t}{t-1}} \right) \widetilde{D}_{\text{c-Low,1}}^{\text{H-dim, H-SNR}}, \tag{4.37}$$

where coefficients β_2 and β_3 can be expressed as the expected powers of the ratios of Gaussian quadratic variables, which is given by

$$\beta_2 = E\left[\frac{\mathbf{h}^{\mathsf{H}} \mathbf{\Sigma}_{\text{h}}^{-1} \mathbf{h}}{\mathbf{h}^{\mathsf{H}} \mathbf{h}} \right] , \qquad \beta_3 = E\left[\left(\frac{\mathbf{h}^{\mathsf{H}} \mathbf{\Sigma}_{\text{h}}^{-1} \mathbf{h}}{\mathbf{h}^{\mathsf{H}} \mathbf{h}} \right)^2 \right] . \tag{4.38}$$

The moments of ratios of random variables, including central quadratic forms in normal variables, were investigated in [51], and the results can described as the following integration

$$E\left[\left(\frac{X}{Y} \right)^n \right] = \Gamma(n)^{-1} \int_0^{\infty} v^{n-1} M_{X,Y}^{(n)}(0, -v) \, d\, v \ , \tag{4.39}$$

where $M_{X,Y}(u,v)$ is the joint moment generating function (m.g.f.) of random variables X and Y, and $M_{X,Y}^{(n)}(0,-v)$ stands for $\partial^n M_{X,Y}(u,-v)/\partial v^n$ evaluated at $u = 0$. Therefore, by setting $X = \mathbf{h}^{\mathsf{H}} \mathbf{\Sigma}_{\text{h}}^{-1} \mathbf{h}$ and $Y = \mathbf{h}^{\mathsf{H}} \mathbf{h}$, the joint m.g.f. of variables X and Y can be represented as

$$M_{X,Y}(u,v) = \frac{1}{\mathbf{det}\Big(I - \big(u \cdot I + v \cdot \mathbf{\Sigma}_{\text{h}} \big) \Big)} = \left(\prod_{k=1}^{t} (1 - u - v \cdot \lambda_{\text{h},k}) \right)^{-1} . \tag{4.40}$$

By substituting the joint m.g.f. given by equation (4.40) into the integration (4.39) with $n = 1$, coefficient β_2 can be obtained as the following closed-form expression after some manipulations,

$$\beta_2 = (t-1)\sum_{i=1}^{t} \frac{\big(\ln \lambda_{\mathrm{h},\,i}\big)/\lambda_{\mathrm{h},\,i}}{\prod_{k\neq i}\big(1-\lambda_{\mathrm{h},\,k}/\lambda_{\mathrm{h},\,i}\big)} \ . \tag{4.41}$$

Finally, by substituting (4.41) into equation (4.36), equality (a) of (4.32) is proved. With similar reasoning, by substituting the joint m.g.f. (3.69) into equation (4.39) with $n = 2$, coefficient β_3 is obtained. Correspondingly, a closed-form expression of coefficient $c_1 = \beta_3 \cdot \beta_2^{-\frac{t}{t-1}}$, which is given by equation (4.34), can also be obtained, and inequality (b) of (4.32) is proved.

Similarly, in Low-SNR regimes, distortion bounds $\widetilde{D}_{\text{c-tr-Low}}^{\text{H-dim, L-SNR}}$ and $\widetilde{D}_{\text{c-tr-Upp}}^{\text{H-dim, L-SNR}}$ can be represented as the following forms

$$\widetilde{D}_{\text{c-tr-Low}}^{\text{H-dim, L-SNR}} = \left(\frac{(t-1)\cdot|\boldsymbol{\Sigma}_{\mathrm{h}}|^{\frac{1}{t-1}}\cdot\beta_4\cdot\rho}{\ln 2}\right)\cdot 2^{-\frac{B}{t-1}} = \beta_4\cdot\widetilde{D}_{\text{c-Low,1}}^{\text{H-dim, L-SNR}}\ , \tag{4.42}$$

$$\widetilde{D}_{\text{c-tr-Upp}}^{\text{H-dim, L-SNR}} \leq \left(\frac{(t-1)\cdot|\boldsymbol{\Sigma}_{\mathrm{h}}|^{\frac{1}{t-1}}\cdot\beta_5\cdot\rho}{\ln 2}\right)\cdot 2^{-\frac{B}{t-1}} = \beta_5\cdot\widetilde{D}_{\text{c-Low,1}}^{\text{H-dim, L-SNR}}\ , \tag{4.43}$$

where the coefficients β_4 and β_5 are given by

$$\beta_4 = E\left[\frac{\mathbf{h}^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{h}}{t}\right], \qquad \beta_5 = E\left[\frac{\big(\mathbf{h}^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{h}\big)^2}{t\cdot\mathbf{h}^{\mathsf{H}}\,\mathbf{h}}\right]\ . \tag{4.44}$$

From equation (4.44), it is evident that $\beta_4 = 1$, and hence the equality (a) of equation (4.33) can be proved. Moreover, by extending the results of the moments of the quadratic forms provided in [51], the following expectation can be obtained after some manipulations

$$E\left[\frac{X^2}{Y}\right] = \int_0^{\infty} \frac{\partial^2 M_{X,Y}(u,-v)}{\partial\, u^2}\bigg|_{u=0} d\,v\ , \tag{4.45}$$

Therefore, by setting $X = \mathbf{h}^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathrm{h}}^{-1}\,\mathbf{h}$ and $Y = \mathbf{h}^{\mathsf{H}}\,\mathbf{h}$, and substituting the joint m.g.f. given by equation (3.69) into the integration (4.45), coefficient β_5 can be

obtained. It is equivalent to coefficient c_2 given by equation (4.35), and hence the inequality (b) of (4.33) can be proved.

■

Note from proposition 5 that constants c_1 and c_2 can be viewed as the upper bounds of the penalty paid for using a transformed codebook instead of optimal design. Further verified by the numerical example shown below, c_1 and c_2 are slightly greater than 1 for most channels that are not "highly" correlated. This means that the intuitive choice of $\mathbf{F}$ given in [23] [52] is a fairly good solution especially for cases when the channel covariance matrix having relative small condition numbers.

We plot in Fig. 4.5 the distortion ratio of correlated fading channels over i.i.d. fading channels (normalized capacity loss) of a 3×1 MISO system versus antenna spacing D/λ with both optimal and transformed codebooks. The average system signal to noise ratio is $\rho = -10$ dB, and the quantization resolution is $B = 10$ bits per channel update. For comparison purpose, the ratio of the distortion bounds, i.e. $\widetilde{D}_{\text{c-tr-Low}}(\mathbf{\Sigma}_{\text{h}})/\widetilde{D}_{\text{c-tr-Low}}(I_t)$ and $\widetilde{D}_{\text{c-tr-Upp}}(\mathbf{\Sigma}_{\text{h}})/\widetilde{D}_{\text{c-tr-Upp}}(I_t)$, are also included in the plot. It can be observed from Fig. 4.5 that the analytical bounds agree well with the obtained simulation results.

In order to demonstrates the tightness of the distortion bounds $\widetilde{D}_{\text{c-tr-Upp}}$ and $\widetilde{D}_{\text{c-tr-Low}}$ in high-SNR and low-SNR regimes, Fig. 4.6 plots the constant coefficient c_1 and c_2 versus the number of transmit antennas t for correlated MISO channels with adjacent antenna spacing $D/\lambda = 0.5$. From the plot, it can be observed that the performance degradation caused by the transformed codebook is less than 10% in low-SNR regimes and 22% in high-SNR regimes for MISO systems with more than 10 transmit antennas.

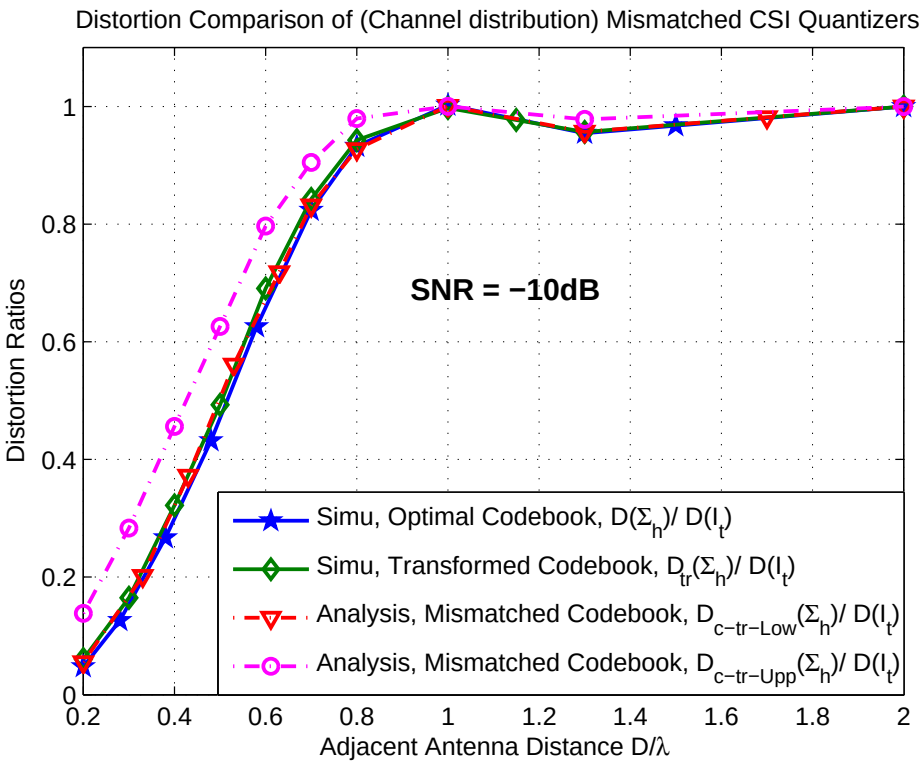

Figure 4.5: Normalized capacity loss (w.r.t. the capacity loss of uncorrelated fading channels) comparison of a 3×1 MISO transmit beamforming with optimal and transformed codebooks versus antenna spacing $d = D/\lambda$, in low SNR regimes ($\rho = -10$ dB).

4.4 Summary

In this chapter, the analysis of MISO systems with finite-rate feedback was extended to sub-optimal CSI-quantizers using mismatched and transformed codebooks. In particular, two types of mismatched MISO CSI quantizers were investigated: quantizers designed with MMSE criterion, and quantizers whose codebooks are designed with a mismatched channel covariance matrix. Moreover, capacity analysis of a feedback-based MISO system over correlated fading channels using channel quantizers with codebooks transformed from i.i.d. channel environments were also investigated. Bounds on the channel capacity loss of the

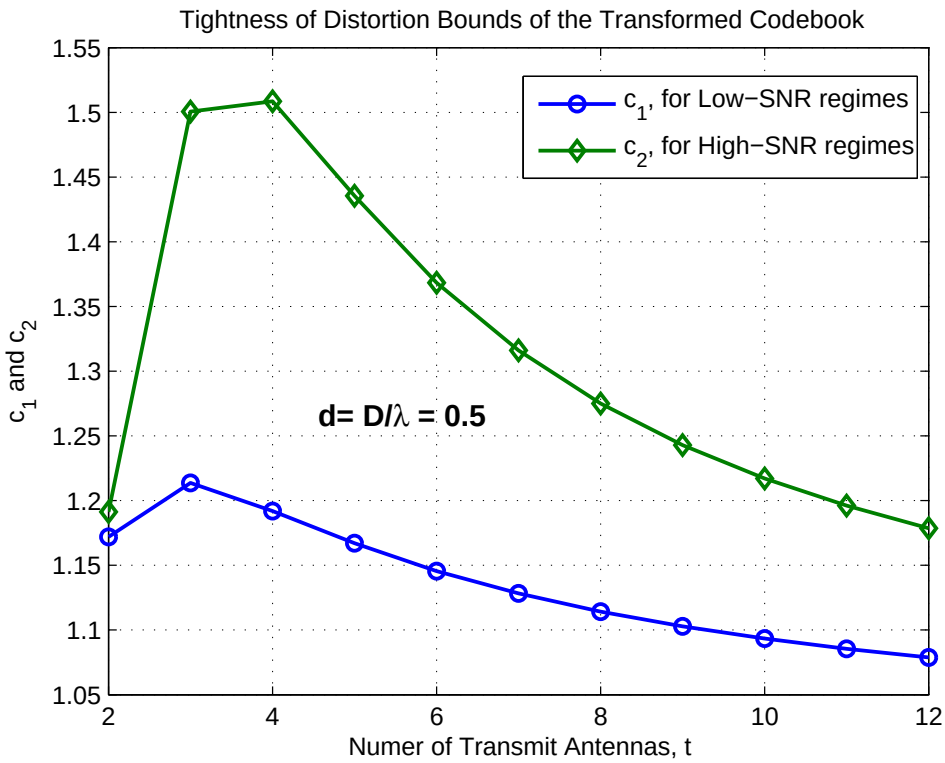

Figure 4.6: Demonstration of the tightness of the distortion bounds $D_{\text{c-tr-Low}}$ and $D_{\text{c-tr-Upp}}$ for a MISO system using transformed codebook over correlated fading channels with different number of transmit antennas of antenna spacing $d = D/\lambda = 0.5$.

MISO systems using mismatched and transformed codebooks were provided and compared to that of the optimal quantizers. Finally, numerical and simulation results were presented and they confirm the accuracy of the obtained theoretical distortion bounds. The text of this chapter as well as Chap. 3, in part and under some rearrangements, are reprints of papers which were coauthored with Bhaskar D. Rao and have been submitted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels"*, and in *IEEE Journal on Selected Areas in Communications* under the title *"Analysis*

of vector quantizers using transformed codebook with application to feedback-based multiple antenna systems" respectively.

5 Capacity Analysis of MIMO Systems with Finite-Rate CSI Feedback

5.1 Motivation

In this chapter, the analysis of CSI-feedback-based multiple antenna systems is further extended to MIMO fading channels. Compared to the MISO systems investigated in Chap. 3 and 4, where the CSI information is only a vector, the quantization objective in the MIMO context is a matrix. As a consequence, more complicated transmit pre-coding schemes and even more involved system capacity analysis are required. By employing the high resolution quantization framework described in Chap. 2, the effects of finite-rate CSI feedback on the performance of MIMO systems over i.i.d. Rayleigh flat fading channels are studied in this chapter. Specifically, tight lower bounds on the capacity loss of MIMO systems over fading channels due to the finite-rate channel quantization are provided. Moreover, MIMO CSI-quantizers using mismatched codebooks that only optimized for high-SNR and low-SNR regimes are investigated. As an application of the obtained distortion analysis, the performance of MIMO systems using multi-mode spatial multiplexing transmission schemes with finite-rate CSI feedback is also provided.

5.2 System Model of MIMO Systems with Finite Rate CSI Feedback

5.2.1 Fading Channel Model

We consider a MIMO system with t transmit antennas and r receive antennas, signaling through a frequency flat Rayleigh fading channel. The channel model can be represented as

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n} \ , \tag{5.1}$$

where $\mathbf{y}$ is the received signal, $\mathbf{n}$ is the additive complex Gaussian noise with distribution $\mathcal{N}_c\big(\mathbf{0},\, I_r\big)$, and $\mathbf{H}$ is the MIMO channel response of size $r \times t$ with each of its element having independent complex Gaussian distribution with zero mean and unit variance. The transmitted signal vector $\mathbf{x}$ is normalized to have a power constraint given by $E\big[\|\mathbf{x}\|^2\big] = \rho$, with ρ representing the average signal to noise ratio at each receive antenna. With probability one, the MIMO channel matrix $\mathbf{H}$ has rank m equal to the minimum number of the transmit and receive antennas, i.e. $m = \min(t,\, r)$. The singular value decomposition (SVD) of matrix $\mathbf{H}$ is denoted as $\mathbf{H} = \mathbf{U}_{\mathrm{H}}\, \mathbf{\Sigma}_{\mathrm{H}}^{\frac{1}{2}}\, \mathbf{V}_{\mathrm{H}}^{\mathsf{H}}$, where $\mathbf{U}_{\mathrm{H}} \in \mathbb{C}^{r \times m}$ and $\mathbf{V}_{\mathrm{H}} \in \mathbb{C}^{t \times m}$ are orthonormal column matrices and $\mathbf{\Sigma}_{\mathrm{H}} = \mathbf{diag}\Big[\lambda_1, \lambda_2, \cdots, \lambda_m\Big]$ is a diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$ representing the sorted eigen-values of matrix $\mathbf{H}\,\mathbf{H}^{\mathsf{H}}$.

5.2.2 Finite-Rate Channel State Information Feedback

In this chapter, the channel state information $\mathbf{H}$ is assumed to be perfectly known at the receiver but only partially available at the transmitter through a finite-rate feedback link of B bits per channel update between the transmitter and receiver. More specifically, a quantization codebook $\mathcal{C} = \big\{\widehat{\mathbf{F}}_1, \cdots, \widehat{\mathbf{F}}_N\big\}$, which is composed of unit-norm[1] transmit pre-coding matrices, is assumed known to both the receiver and the transmitter. Based on the channel realization $\mathbf{H}$, the receiver selects the best code point $\widehat{\mathbf{F}}$ from the codebook and sends the corresponding index

[1] A matrix norm is defined to be the standard Froebinius norm given by $\|\mathbf{A}\| = \mathbf{tr}\left(\mathbf{A}\,\mathbf{A}^{\mathsf{H}}\right)$.

back to the transmitter. At the transmitter, the unit-norm matrix $\widehat{\mathbf{F}}$ is employed as the pre-coding matrix, and the channel model can be represented as

$$\mathbf{y} = \mathbf{H} \cdot \left(\widehat{\mathbf{F}} \cdot \mathbf{s}\right) + \mathbf{n}, \qquad E\left[\mathbf{s} \cdot \mathbf{s}^{\mathsf{H}}\right] = \rho \cdot I_n \ . \tag{5.2}$$

With perfect channel state information available at the transmitter, which corresponds to the case of infinite feedback rate $B = \infty$, it is optimal to choose $\mathbf{F} = \mathbf{V}_{\mathrm{H}} \mathbf{P}^{\frac{1}{2}}$ as the transmit pre-coding matrix, where $\mathbf{V}_{\mathrm{H}}$ is the right singular matrix of $\mathbf{H}$, and the diagonal matrix $\mathbf{P}$ represents the optimal water-filling power allocation matrix given by the following form

$$\mathbf{P} = \mathbf{diag}\Big[p_1, p_2, \cdots, p_m\Big], \qquad p_i = \left(\mu - \frac{1}{\rho\,\lambda_i}\right)^{\dagger}, \quad 1 \le i \le m \ , \tag{5.3}$$

where "†" is an operation defined as $a^{\dagger} = \max(a, 0)$, and μ is a constant coefficient (water-level) that satisfies $\sum_{i=1}^{m} p_i = 1$. When the feedback rate B is limited, the quantized pre-coding matrix $\widehat{\mathbf{F}}$ can also be represented as a product given by $\widehat{\mathbf{F}} = \widehat{\mathbf{V}}_{\mathrm{H}} \widehat{\mathbf{P}}^{\frac{1}{2}}$, where $\widehat{\mathbf{V}}_{\mathrm{H}}$ represents the quantized sigular matrix and $\widehat{\mathbf{P}}$ represents the quantized power allocation. It is evident that quantizing the power allocation reduces the bit budget for the finite-rate representation of beamforming matrix $\widehat{\mathbf{V}}_{\mathrm{H}}$. Therefore, we focus our attention only on MIMO systems using transmit pre-coders with equal power allocations among the spatial beams. Furthermore, in order to compensate the equal power allocation scheme when compared to the optimal water-filling solution, a multi-mode spatial-multiplexing (MMSM) strategy was proposed in [26], whose performance was shown to be close to that of systems using pre-coders with optimal power allocations.

Notice that in practical systems, the channel error and feedback delay also exist in the reverse link, which will impact the overall system performance. However, this chapter assumes the feedback is error-free and delay-less, and focuses solely on the effect of finite-rate quantization of the channel state information.

5.2.3 Practical Transmit Pre-Coders with Quantized CSIT

Under practical situations, in order to simplify the codebook design complexity or simply reduce the amount of feedback information, a sub-optimal CSI quantization scheme using orthonormal pre-coding matrix with equal power allocation on each of its spatial beams is utilized in [25]. To be specific, first denote the following decomposition of matrices $\mathbf{U}_{\mathrm{H}}$, $\mathbf{\Sigma}_{\mathrm{H}}$ and $\mathbf{V}_{\mathrm{H}}$, given by

$$\mathbf{U}_{\mathrm{H}} = \Big[\mathbf{U}\ \mathbf{U}'\Big], \qquad \mathbf{V}_{\mathrm{H}} = \Big[\mathbf{V}\ \mathbf{V}'\Big], \qquad \mathbf{\Sigma}_{\mathrm{H}} = \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{H},n} & 0 \\ 0 & \mathbf{\Sigma}' \end{bmatrix}, \tag{5.4}$$

where $\mathbf{U}$ is of size $r \times n$, $\mathbf{V}$ is of size $t \times n$, and $\mathbf{\Sigma}_{\mathrm{H},n}$ is of size $n \times n$ containing the first (largest) n eigen-values of $\mathbf{H}\,\mathbf{H}^{\mathsf{H}}$ with $1 \le n \le m$. For a transmit pre-coder only using n spatial beams, orthonormal matrix $\mathbf{V}$ becomes the channel quantization objective, and the equivalent MIMO channel can be represented as

$$\mathbf{y} = \mathbf{H} \cdot \left(\frac{1}{\sqrt{n}}\,\widehat{\mathbf{V}} \cdot \mathbf{s}\right) + \mathbf{n}\ , \tag{5.5}$$

where $\widehat{\mathbf{V}}$ is the quantized pre-coding matrix from codebook $\mathcal{C} = \big\{\widehat{\mathbf{V}}_1, \cdots \widehat{\mathbf{V}}_N\big\}$, whose elements are orthonormal column matrices of sizes $t \times n$.

When the channel state information is perfectly known at the transmitter, the system capacity[2] by using a n-beam transmit pre-coding scheme with equal power allocation is given by

$$\begin{aligned} C_{\mathrm{perf}} &= E\Big[C_{\mathrm{p}}\big(\mathbf{H}\big)\Big] \\ &= E\bigg[\log_2 \det\Big(I + \frac{\rho}{n} \cdot \mathbf{V}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\,\mathbf{H}\,\mathbf{V}\Big)\bigg] = E\bigg[\log_2 \det\Big(I + \frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}\Big)\bigg]. \end{aligned} \tag{5.6}$$

When the feedback link is restricted to a finite rate (B bits per channel update), the system capacity with finite-rate CSI feedback can be represented as

$$C_{\mathrm{quan}} = E\Big[C_{\mathrm{q}}\big(\mathbf{H},\,\widehat{\mathbf{V}}\big)\Big] = E\bigg[\log_2 \det\Big(I + \frac{\rho}{n} \cdot \widehat{\mathbf{V}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\,\widehat{\mathbf{V}}\Big)\bigg]\ , \tag{5.7}$$

[2]The system capacity here refers to the mutual information rate of a specific setting. The actual capacity without the restriction of equal power allocation among the spatial bemas is presumably larger than the aforementioned mutual information rate.

where the quantized beamforming matrix $\widehat{\mathbf{V}}$ is a function of the current channel realization $\mathbf{H}$, i.e. $\widehat{\mathbf{V}} = \widehat{\mathbf{V}}\big(\mathbf{H}\big) \in \mathcal{C}$. Therefore, the performance of a CSI-feedback-based MIMO system can be characterized by the system capacity loss C_{Loss} due to the finite-rate quantization of the CSI. It is defined as the expectation of the instantaneous capacity loss, i.e.

$$C_{\text{Loss}} = C_{\text{perf}} - C_{\text{quan}} = E\Big[C_{\text{L}}\big(\mathbf{H}, \widehat{\mathbf{V}}\big)\Big] \ , \tag{5.8}$$

where the instantaneous capacity loss $C_{\text{L}}(\mathbf{H}, \widehat{\mathbf{V}})$ is given by the following form

$$\begin{aligned} C_{\text{L}}\big(\mathbf{H}, \widehat{\mathbf{V}}\big) &= C_{\text{p}}\big(\mathbf{H}\big) - C_{\text{q}}\big(\mathbf{H}, \widehat{\mathbf{V}}\big) \\ &= \log_2 \det\left(I + \frac{\rho}{n}\mathbf{\Sigma}_{\text{H},n}\right) - \log_2 \det\left(I + \frac{\rho}{n}\widehat{\mathbf{V}}^{\mathsf{H}}\mathbf{V}_{\text{H}}\mathbf{\Sigma}_{\text{H}}\mathbf{V}_{\text{H}}^{\mathsf{H}}\widehat{\mathbf{V}}\right) \ . \end{aligned} \tag{5.9}$$

This performance metric was also used in [26] and [27].

5.3 Analysis of MIMO Pre-coders with Quantized CSI

As an extended application of the general distortion analysis provided in [44] and [45], we provide in this section the performance analysis of a finite-rate CSI-feedback-based MIMO system using transmit precoding schemes with equal power allocation on multiple spatial beams.

5.3.1 Formulating the MIMO CSI-Quantizer as A General Vector Quantization Problem

Similar to the case of MISO systems with finite-rate feedback provided in [53], one can formulate the CSI-feedback-based MIMO system into a fixed-rate general vector quantization problem by utilizing the framework provided in [45]. To be specific, the source variable to be quantized is the right singular matrix $\mathbf{V}$ of the fading channel response $\mathbf{H}$. It is a complex matrix of size $t \times n$, which contains $k_{\text{q}} = 2tn$ real dimensions. The system distortion function D_{Q} is chosen to be the instantaneous capacity loss C_{L} given by equation (5.9), whose second order Taylor series expansion is given by the following lemma.

Lemma 5 *The system distortion function D_Q (or C_L) can be approximated by the following second order Taylor series expansion,*

$$D_Q\big(\mathbf{V}, \widehat{\mathbf{V}}\,;\, \mathbf{\Sigma}_{H,n}\big) = \log_2 \det\Big(I + \frac{\rho}{n}\mathbf{\Sigma}_{H,n}\Big) - \log_2 \det\Big(I + \frac{\rho}{n}\,\mathbf{\Sigma}_H \cdot \mathbf{V}_H^{H}\widehat{\mathbf{V}}\widehat{\mathbf{V}}^{H}\mathbf{V}_H\Big)$$
$$\approx \frac{1}{\ln 2} \cdot \mathbf{tr}\left(\mathbf{V}^{H}\Big(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{H}\Big)\,\mathbf{V} \cdot \frac{\rho}{n}\,\mathbf{\Sigma}_{H,n}\,\Big(I + \frac{\rho}{n}\,\mathbf{\Sigma}_{H,n}\Big)^{-1}\right) \quad . \tag{5.10}$$

Proof: It is noted that the distortion function D_{Q} (or capacity loss C_{L}) given by equation (5.9) is a real-valued function of complex variable $\mathbf{V}_{\mathrm{H}}$. We therefore utilize the Wirtinger calculous [47] to obtain the complex derivative and complex Hessian matrix of the distortion function with respect to $\mathbf{V}_{\mathrm{H}}$.

Let us first consider a real-valued complex function $f(\mathbf{X})$ given by the following form

$$f\big(\mathbf{X};\, \mathbf{A}_1,\, \mathbf{A}_2\big) = \log_2 \det\Big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\Big)\,, \tag{5.11}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are semi-definite complex Hessian matrices. According to the definitions given in [48], the generalized complex derivative of function $f(\mathbf{X})$ can be obtained by the following form,

$$\mathbf{d}_{\mathrm{f}} = \left[\frac{\partial}{\partial\,\mathbf{x}^{\star}}\,f\big(\mathbf{X}\big)\right]^{\mathsf{T}} = \frac{1}{\ln 2}\;\mathbf{vec}\Big(\,\mathbf{A}_2\,\mathbf{X}\cdot\big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\big)^{-1}\cdot\mathbf{A}_1\Big)\;, \tag{5.12}$$

where $\mathbf{x} = \mathbf{vec}\big(\mathbf{X}\big)$. Furthermore, the complex Hessian matrices of $f(\mathbf{X})$ can also be obtained as,

$$\begin{aligned}\mathbf{\Phi}_{\mathrm{f}} = \frac{\partial\,\mathbf{d}_{\mathrm{f}}}{\partial\,\mathbf{x}} \;&=\; \frac{1}{\ln 2}\Big(\mathbf{A}_1^{\mathsf{T}}\Big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\Big)^{-\mathsf{T}}\Big)\\ &\quad\otimes\Big(\mathbf{A}_2 - \mathbf{A}_2\mathbf{X}\Big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\Big)^{-1}\mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\Big)\quad,\end{aligned} \tag{5.13}$$

$$\begin{aligned}\mathbf{\Phi}'_{\mathrm{f}} = \frac{\partial\,\mathbf{d}_{\mathrm{f}}}{\partial\,\mathbf{x}^{\star}} \;&=\; \frac{-1}{\ln 2}\Bigg(\Big(\mathbf{A}_1^{\mathsf{T}}\Big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\Big)^{-\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{A}_2^{\mathsf{T}}\Big)\\ &\quad\otimes\Big(\mathbf{A}_2\mathbf{X}\cdot\Big(I + \mathbf{A}_1\mathbf{X}^{\mathsf{H}}\mathbf{A}_2\mathbf{X}\Big)^{-1}\mathbf{A}_1\Big)\Bigg)\cdot\mathbf{P},\end{aligned} \tag{5.14}$$

where $\mathbf{P}$ (of size $tn \times tn$) is a permutation matrix defined as

$$\mathbf{P} = \sum_{r=1}^{t} \sum_{s=1}^{n} \mathbf{E}_{rs} \otimes \mathbf{E}_{sr} \quad , \tag{5.15}$$

where $\mathbf{E}_{rs}$ (of size $t \times n$) and $\mathbf{E}_{sr}$ (of size $n \times t$) are elementary matrices which have unity in the $(r,s)^{\text{th}}$ or $(s,r)^{\text{th}}$ position and all other elements are zero.

The distortion function D_{Q} (or capacity loss C_{L}) given by equation (5.9) can also be represented as the following form

$$D_{\text{Q}} = C_{\text{L}} = f\left(\mathbf{V}_{\text{H}};\ \begin{bmatrix} \frac{\rho}{n}\,\mathbf{\Sigma}_{\text{H},n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},\ I_t\right) - f\left(\mathbf{V}_{\text{H}};\ \frac{\rho}{n}\,\mathbf{\Sigma}_{\text{H}}\,,\ \widehat{\mathbf{V}}\widehat{\mathbf{V}}^{\mathsf{H}}\right) \quad . \tag{5.16}$$

After some manipulations, the complex derivative of function D_{Q} can be obtained

$$\mathbf{d}_{\text{Q}} = \left[\frac{\partial}{\partial\, \mathbf{v}_{\text{H}}^{\star}}\, D_{\text{Q}}\big(\mathbf{V}_{\text{H}}\big)\right]^{\mathsf{T}} \Bigg|_{\mathbf{V}_{\text{H}}=\widehat{\mathbf{V}}_{\text{H}}} = \mathbf{0}\,, \quad \widehat{\mathbf{V}}_{\text{H}} = \left[\widehat{\mathbf{V}},\, \mathbf{V}'\right], \quad \mathbf{v}_{\text{H}}^{\star} = \mathbf{vec}\,(\mathbf{V}_{\text{H}}) \ , \tag{5.17}$$

Moreover, according to equations (5.13) and (5.14), the Hessian matrices of D_{Q} can also be obtained as

$$\mathbf{\Phi}_{\text{Q}} = \frac{\partial\, \mathbf{d}_{\text{Q}}}{\partial\, \mathbf{v}_{\text{H}}} \Bigg|_{\mathbf{V}_{\text{H}}=\widehat{\mathbf{V}}_{\text{H}}} = \frac{1}{\ln 2} \begin{bmatrix} \widetilde{\mathbf{\Sigma}}_{\text{H},n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \otimes \left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right), \tag{5.18}$$

and

$$\mathbf{\Phi}'_{\text{Q}} = \frac{\partial\, \mathbf{d}_{\text{Q}}}{\partial\, \mathbf{v}_{\text{H}}^{\star}} \Bigg|_{\mathbf{V}_{\text{H}}=\widehat{\mathbf{V}}_{\text{H}}} = \mathbf{0} \ , \tag{5.19}$$

where matrix $\widetilde{\mathbf{\Sigma}}_{\text{H},n}$ is given by

$$\widetilde{\mathbf{\Sigma}}_{\text{H},n} = \frac{\rho}{n}\mathbf{\Sigma}_{\text{H},n} \cdot \left(I + \frac{\rho}{n}\mathbf{\Sigma}_{\text{H},n}\right)^{-1} \quad . \tag{5.20}$$

It is shown in [48] that a real-valued complex function has the following second order Taylor series expansion,

$$\begin{aligned} D_{\text{Q}}(\mathbf{V}_{\text{H}}) = {} & D_{\text{Q}}\big(\widehat{\mathbf{V}}_{\text{H}}\big) \ + \ 2\Re\left[\big(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\big)^{\mathsf{H}} \cdot \mathbf{d}_{\text{Q}}\right] \\ & + \Re\Big[\big(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\big)^{\mathsf{H}} \mathbf{\Phi}_{\text{Q}} \left(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\right) \ + \ \big(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\big) \mathbf{\Phi}'_{\text{Q}} \left(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\right)^{\star}\Big] \ + \ \text{h.o.t} \ , \end{aligned} \tag{5.21}$$

where vector $\widehat{\mathbf{v}}_{\text{H}} = \mathbf{vec}\big(\widehat{\mathbf{V}}_{\text{H}}\big)$. By substituting equation (5.17) and (5.18) into the Taylor series expansion give by (5.21), one can obtain the following result

$$\begin{aligned} D_{\text{Q}} &\approx \frac{1}{\ln 2}\left(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\right)^{\mathsf{H}} \cdot \left(\begin{bmatrix} \widetilde{\boldsymbol{\Sigma}}_{\text{H},n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \otimes \left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right)\right) \cdot \left(\mathbf{v}_{\text{H}} - \widehat{\mathbf{v}}_{\text{H}}\right) \\ &= \frac{1}{\ln 2}\,\mathbf{tr}\left(\left(\mathbf{V} - \widehat{\mathbf{V}}\right)^{\mathsf{H}} \cdot \left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right) \cdot \left(\mathbf{V} - \widehat{\mathbf{V}}\right) \cdot \widetilde{\boldsymbol{\Sigma}}_{\text{H},n} \right) \\ &= \frac{1}{\ln 2}\left(\mathbf{v} - \widehat{\mathbf{v}}\right)^{\mathsf{H}} \cdot \left(\widetilde{\boldsymbol{\Sigma}}_{\text{H},n} \otimes \left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right)\right) \cdot \left(\mathbf{v} - \widehat{\mathbf{v}}\right) . \end{aligned} \tag{5.22}$$

It can be observed from equation (5.22) that D_{Q} (up to the second order approximation) is a function between $\mathbf{V}$ and $\widehat{\mathbf{V}}$, which is only parameterized by $\boldsymbol{\Sigma}_{\text{H},n}$. ■

It can be observed from equation (5.10) that the distortion function D_{Q} between $\mathbf{V}$ and $\widehat{\mathbf{V}}$ is only parameterized by $\boldsymbol{\Sigma}_{\text{H},n}$, under the second order approximation. The diagonal elements of matrix $\boldsymbol{\Sigma}_{\text{H},n}$ represent the first n non-zero eigen-values of matrix $\mathbf{H}^{\mathsf{H}}\,\mathbf{H}$. Therefore, the encoder side information in this case can be denoted as $\mathbf{z} = \boldsymbol{\Sigma}_{\text{H},n}$ of n degrees of freedom.

In contrast to the conventional quantization problems, the source variable to be quantized in this case is subject to some constraints. First of all, according to the SVD definition, matrix $\mathbf{V}$ has orthonormal column vectors, i.e.

$$\mathbf{V}^{\mathsf{H}}\,\mathbf{V} = I_n\,, \qquad \mathbf{V} \in \mathbb{C}^{t\times n}\,, \tag{5.23}$$

which corresponds to n^2 independent constrained equations. Furthermore, the distortion function D_{Q} can also be rewritten in the following form after some manipulations

$$\begin{aligned} &D_{\text{Q}}\big(\mathbf{V}, \widehat{\mathbf{V}}\,;\, \boldsymbol{\Sigma}_{\text{H},n}\big) \\ &\quad = \frac{1}{\ln 2} \cdot \mathbf{tr}\left(\left(I - \left(\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\right) \cdot \left(\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\right)^{\mathsf{H}}\right) \cdot \frac{\rho}{n}\,\boldsymbol{\Sigma}_{\text{H},n} \left(I + \frac{\rho}{n}\,\boldsymbol{\Sigma}_{\text{H},n}\right)^{-1}\right). \end{aligned} \tag{5.24}$$

From the above equation, it can be observed that the distortion function D_{Q} depends only on the matrix product $\mathbf{V}^{\mathsf{H}}\,\widehat{\mathbf{V}}$ and the encoder side information $\boldsymbol{\Sigma}_{\text{H},n}$. Hence, the distortion function can be denoted as $D_{\text{Q}}\big(\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\,;\, \boldsymbol{\Sigma}_{\text{H},n}\big)$. Moveover,

it also can be shown that D_{Q} is invariant under any unitary rotations on matrix $\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}$ at the right hand side, i.e.

$$D_{\mathrm{Q}}\big(\,\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\,\mathbf{Q}_{\mathrm{R}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\,\big) = D_{\mathrm{Q}}\big(\,\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\,\big)\ , \tag{5.25}$$

where $\mathbf{Q}_{\mathrm{R}}$ is an arbitrary unitary matrix. Suppose matrix product $\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}$ has a unique R-Q decomposition given by the following form

$$\mathbf{B} = \mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}} = \mathbf{R}_{\mathrm{p}}\cdot\mathbf{Q}_{\mathrm{p}}\ , \tag{5.26}$$

where $\mathbf{Q}_{\mathrm{p}}$ is a unitary matrix and $\mathbf{R}_{\mathrm{p}}$ is an upper triangle matrix with real diagonal elements. Hence, for any realizations of $\mathbf{V}$ (or $\mathbf{B}$), there always exists a unitary rotation $\mathbf{Q}_{\mathrm{R}} = \mathbf{Q}_{\mathrm{p}}^{\mathsf{H}}$ such that $\mathbf{B}\,\mathbf{Q}_{\mathrm{R}}$ is an upper triangle matrix. Therefore, with out loss of generality, one can impose on matrix $\mathbf{B}$ the following constrained conditions, such that for points $\mathbf{V}$ in the small neighborhood of $\widehat{\mathbf{V}}$ matrix $\mathbf{B}$ is an upper triangle matrix with real diagonal elements, i.e.

$$\measuredangle\, b_{i,\,i} = 0\,, \qquad\qquad b_{\substack{i,j\\ i<j}} = 0\ , \tag{5.27}$$

where $b_{i,\,j}$ is the $(i,\,j)^{\mathrm{th}}$ element of matrix $\mathbf{B}$. Note that the above constraints on $\mathbf{V}$ given in equation (5.27) count for another n^2 independent constrained equations.

According to the constraints given by (5.23) and (5.27), there are total $k_{\mathrm{c}} = 2n^2$ independent constrained conditions (equations), and the number of free dimensions of matrix $\mathbf{V}$ reduces to be $k'_{\mathrm{q}} = \big(2tn - 2n^2\big)$. These constrained conditions can be further represented as the following concise manner, which is denoted as a multi-dimensional real function $\mathbf{g}(\mathbf{V})$ given by

$$\mathbf{g}\big(\mathbf{V}\big) = \Big[\mathbf{g}_1^{\mathsf{T}}\big(\mathbf{V}\big),\,\mathbf{g}_2^{\mathsf{T}}\big(\mathbf{V}\big)\Big]^{\mathsf{T}} = \mathbf{0}\,, \tag{5.28}$$

where vectors $\mathbf{g}_1\big(\mathbf{V}\big)$ and $\mathbf{g}_2\big(\mathbf{V}\big)$ can be represented by

$$\mathbf{g}_i\big(\mathbf{V}\big) = \Big[g_{i,1}\big(\mathbf{V}\big), g_{i,2}\big(\mathbf{V}\big), \cdots, g_{i,n^2}\big(\mathbf{V}\big)\Big], \quad i = 1, 2\ , \tag{5.29}$$

whose element functions $g_{i,j}(\cdot)$ are given by the following form

$$\begin{aligned}
&g_{1,(i-1)n+i} = \mathbf{v}_i^{\mathsf{H}}\mathbf{v}_i - 1\,, && g_{2,(i-1)n+i} = j\left(\mathbf{v}_i^{\mathsf{H}}\widehat{\mathbf{v}}_i - \widehat{\mathbf{v}}_i^{\mathsf{H}}\mathbf{v}_i\right), \quad 1 \le i \le n \\
&g_{1,(i-1)n+k} = \mathbf{v}_i^{\mathsf{H}}\mathbf{v}_k + \mathbf{v}_k^{\mathsf{H}}\mathbf{v}_i\,, && g_{2,(i-1)n+k} = j\left(\mathbf{v}_i^{\mathsf{H}}\mathbf{v}_k - \mathbf{v}_k^{\mathsf{H}}\mathbf{v}_i\right), \quad 1 \le i < k \le n \\
&g_{1,(i-1)n+k} = \mathbf{v}_i^{\mathsf{H}}\widehat{\mathbf{v}}_k + \widehat{\mathbf{v}}_k^{\mathsf{H}}\mathbf{v}_i\,, && g_{2,(i-1)n+k} = j\left(\mathbf{v}_i^{\mathsf{H}}\widehat{\mathbf{v}}_k - \widehat{\mathbf{v}}_k^{\mathsf{H}}\mathbf{v}_i\right), \quad 1 \le k < i \le n.
\end{aligned} \tag{5.30}$$

In the above equation, vectors $\mathbf{v}_i$ and $\widehat{\mathbf{v}}_i$ are the i^{th} column of matrices $\mathbf{V} = \Big[\mathbf{v}_1, \cdots, \mathbf{v}_n\Big]$ and $\widehat{\mathbf{V}} = \Big[\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_n\Big]$ respectively.

5.3.2 High-Resolution Distortion Analysis

In most communication problems, the CSI is usually represented as a complex vector or complex matrix. However, the high-resolution analysis provided in [44] and [45] is only suitable for real vectors. Hence, in most of the cases, the complex CSI is first converted to real vectors by expanding its real and imaginary parts. Fortunately, in some special cases including the MIMO CSI quantization problem investigated in this section, the distortion analysis can be performed in the complex domain directly. A detailed distortion analysis of finite-rate quantization of complex source variables is provided in Section 2.7, which includes the necessary and sufficient conditions that guarantee a concise distortion analysis in the complex domain.

According to the second order Taylor series expansion (5.10) provided in Lemma 5, distortion function D_{Q} can also be represented by the following form

$$D_{\mathrm{Q}}\big(\mathbf{V}, \widehat{\mathbf{V}}\,;\, \mathbf{\Sigma}_{\mathrm{H},n}\big) = \big(\mathbf{v} - \widehat{\mathbf{v}}\big)^{\mathsf{H}} \cdot \mathbf{W}\big(\,\widehat{\mathbf{V}}\,;\, \mathbf{\Sigma}_{\mathrm{H},n}\,\big) \cdot \big(\mathbf{v} - \widehat{\mathbf{v}}\big) \tag{5.31}$$

where the (complex) unconstraint sensitivity matrix $\mathbf{W}\big(\,\widehat{\mathbf{V}}\,;\, \mathbf{\Sigma}_{\mathrm{H},n}\,\big)$ is given

$$\mathbf{W}\big(\,\widehat{\mathbf{V}}\,;\, \mathbf{\Sigma}_{\mathrm{H},n}\,\big) = \frac{1}{\ln 2}\,\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n} \otimes \Big(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\Big)\,,\ \widetilde{\mathbf{\Sigma}}_{\mathrm{H},n} = \frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n} \cdot \Big(I + \frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}\Big)^{-1}. \tag{5.32}$$

In order to obtain the constrained sensitivity matrix, the derivative of the constrained equation $\mathbf{g}(\cdot)$ needs to be derived. First of all, it is clear that $\mathbf{g}(\cdot)$ given

in equation (5.30) is a multi-dimensional real function of size $2n^2 \times 1$. According to the method of Wirtinger calculus [48], we can first obtain the following partial derivative of function $\mathbf{g}(\cdot)$ w.r.t. vector $\mathbf{v}^*$,

$$\frac{\partial}{\partial\, \mathbf{v}^*}\, \mathbf{g}_1(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = \big(I+\mathbf{P}\big)\cdot\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{T}}\Big)\,, \tag{5.33}$$

$$\frac{\partial}{\partial\, \mathbf{v}^*}\, \mathbf{g}_2(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = j\big(I-\mathbf{P}\big)\cdot\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{T}}\Big)\ , \tag{5.34}$$

where $\mathbf{P} \in \mathbb{R}^{\,n^2\times n^2}$ is a sparse matrix with its elements given by

$$p_{\,(i-1)n+k,\,m} = \begin{cases} 1 & \text{for} \quad m=(k-1)n+i \quad \& \quad i<k\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{5.35}$$

where $1 \le i,k \le n$ and $1 \le m \le n^2$. After similar manipulations, the partial derivative of $\mathbf{g}(\cdot)$ w.r.t. vector $\mathbf{v}$ can also be obtained, which is given by

$$\frac{\partial}{\partial\, \mathbf{v}}\, \mathbf{g}_1(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = \big(I+\mathbf{P}\big)\cdot\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{H}}\Big)\,, \tag{5.36}$$

$$\frac{\partial}{\partial\, \mathbf{v}}\, \mathbf{g}_2(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = -j\big(I-\mathbf{P}\big)\cdot\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{H}}\Big)\ . \tag{5.37}$$

Therefore, by defining $\widetilde{\mathbf{v}} = \big[\,\mathbf{v}^{\mathsf{T}},\, \mathbf{v}^{\mathsf{H}}\big]^{\mathsf{T}}$ and $\widetilde{\mathbf{v}}' = \big[\,-\mathbf{v}^{\mathsf{T}},\, \mathbf{v}^{\mathsf{H}}\big]^{\mathsf{T}}$, the partial derivatives of function $\mathbf{g}(\cdot)$ w.r.t. vectors $\widetilde{\mathbf{v}}$ and $\widetilde{\mathbf{v}}'$ can be obtained as the following form

$$\frac{\partial}{\partial\, \widetilde{\mathbf{v}}}\, \mathbf{g}_1(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = \big(I+\mathbf{P}\big)\cdot\Big[\ -\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{H}}\Big)\ \Big|\ \Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{T}}\Big)\ \Big]\,, \tag{5.38}$$

$$\frac{\partial}{\partial\, \widetilde{\mathbf{v}}}\, \mathbf{g}_2(\mathbf{V})\bigg|_{\mathbf{V}=\widehat{\mathbf{V}}} = j\big(I-\mathbf{P}\big)\cdot\Big[\ -\Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{H}}\Big)\ \Big|\ \Big(I_n \otimes \widehat{\mathbf{V}}^{\mathsf{T}}\Big)\ \Big]\ , \tag{5.39}$$

which satisfies the necessary and sufficient condition give by equation (2.109) in Section 2.7.2. According to Proposition 2, the constrained sensitivity matrix of the CSI-quantized MIMO system is given by

$$\mathbf{W}_{\mathrm{c}}\big(\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big) = \mathbf{V}_2^{\mathsf{H}}\cdot\mathbf{W}\big(\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big)\cdot\mathbf{V}_2\ , \tag{5.40}$$

where matrix $\mathbf{W}\big(\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big)$ is the unconstrained sensitivity matrix given by equation (5.32), and $\mathbf{V}_2$ is an orthonormal matrix with its columns constituting an orthonormal basis of the null space $\mathcal{N}\big(\frac{\partial}{\partial \mathbf{v}}\,\mathbf{g}_1(\mathbf{v})\big)$, which is given by

$$\mathbf{V}_2 = I_n \otimes \widehat{\mathbf{V}}_2\ , \tag{5.41}$$

with $\widehat{\mathbf{V}}_2$ being an orthonormal matrix with its columns constituting an orthonormal basis of the null space $\mathcal{N}\big(\widehat{\mathbf{V}}^{\mathsf{H}}\big)$. After some manipulations, it can be shown that the constrained sensitivity matrix can be represented by the following form,

$$\mathbf{W}_{\mathrm{c}}\big(\,\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\,\big) = \frac{1}{\ln 2}\;\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n} \otimes I_{(t-n)} \quad . \tag{5.42}$$

By substituting equation (5.42) into the hyper-ellipsoidal approximation given by (2.110), the optimal inertial profile is tightly lower bounded by the following form

$$\widetilde{I}_{\mathrm{c,opt}}\big(\,\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\,\big) = \frac{(tn-n^2)\cdot\left|\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n}\right|^{1/n}}{\ln 2\cdot(tn-n^2+1)\cdot\gamma_0^{\,1/(tn-n^2)}}\;, \tag{5.43}$$

where γ_0 is a constant given by

$$\gamma_0 = \frac{\pi^{tn-n^2}}{(tn-n^2)!}\;. \tag{5.44}$$

It can be observed from equations (5.42) and (5.43) that the constrained sensitivity matrix as well as its corresponding normalized inertial profile are independent of the location $\widehat{\mathbf{V}}$.

When the elements of the channel matrix $\mathbf{H}$ are assumed to have i.i.d. complex Gaussian distributions, it is shown in [49] that matrix $\mathbf{V}$ is independent of the side information $\mathbf{\Sigma}_{\mathrm{H},n}$. According to the definition, matrix $\mathbf{V}$ belongs to the set of all $t\times n$ $(t\geq n)$ complex matrices with orthonormal columns, which is called the complex Stiefel manifold, denoted as $\mathcal{V}_{n,t} = \big\{\mathbf{V}\,:\mathbf{V}^{\mathsf{H}}\mathbf{V} = I_n\big\}$. The volume of the complex Stiefel manifold is found in [54], which is given by

$$\mathrm{Vol}\big(\mathcal{V}_{n,t}\big) = \int_{\mathcal{V}_{n,t}} d\mathbf{V} = \frac{2^n\pi^{nt}}{\widetilde{\Gamma}_n(t)}\;, \tag{5.45}$$

where $\widetilde{\Gamma}_n(\cdot)$ is the complex multivariate gamma function given by $\widetilde{\Gamma}_n(t) = \pi^{n(n-1)/2}\prod_{k=1}^{n}\Gamma(t-k+1)$. Therefore, for random matrix $\mathbf{V}$ uniformly distributed over $\mathcal{V}_{n,t}$, the joint density function for $\mathbf{V}$ is simply given by the following form

$$p\big(\mathbf{V}\big) = \mathrm{Vol}\big(\mathcal{V}_{n,t}\big)^{-1}, \qquad \mathbf{V}\in\mathcal{V}_{n,t}\;. \tag{5.46}$$

Since we are interested in the case of constrained source variables, where $\mathbf{V}$ is subject to constrained condition (5.30), the joint density function given by

(5.46) can not be directly applied. If we denote the constrained source space by $\mathbb{Q}$, then the complex Stiefel manifold $\mathbf{V}_{n,t}$ can be represented by expanding space $\mathbb{Q}$ under a unitary rotation $\mathbf{Q}_{\mathrm{R}}$, given by

$$\mathcal{V}_{n,t} = \Big\{ \mathbf{V} \cdot \mathbf{Q}_{\mathrm{R}} \ : \ \mathbf{V} \in \mathbb{Q}, \ \mathbf{Q}_{\mathrm{R}} \in \mathcal{V}_{n,n} \Big\} \ . \tag{5.47}$$

Therefore, the probability density function of the constrained source $\mathbf{V}$ variable is given by the following form

$$p\big(\mathbf{V}\big) = \int_{\mathcal{V}_{n,n}} \mathrm{Vol}\big(\mathcal{V}_{n,t}\big)^{-1} \, d\,\mathbf{Q}_{\mathrm{R}} = \frac{(t-1)\tilde{!}}{\pi^{tn-n^2}(t-n-1)\tilde{!}\,(n-1)\tilde{!}} \ , \tag{5.48}$$

where we define $k\tilde{!} \triangleq \prod_{i=1}^{k} i!$.

Therefore, by substituting the obtained pdf given by (5.48) into the definition of the average inertial profile given by (2.36), the average normalized inertial profile of the CSI-quantized MIMO system can be obtained as

$$\widetilde{I}^{\mathrm{w}}_{\mathrm{c,opt}}\big(\, \mathbf{V} \,;\, \mathbf{\Sigma}_{\mathrm{H},n} \,\big) = \frac{(tn-n^2)\cdot\beta_1}{\ln 2 \cdot (tn-n^2+1)\cdot \gamma_0^{1/(tn-n^2)}} \ , \tag{5.49}$$

where the constant coefficient β_1 is given by

$$\beta_1 = E\left[\left|\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n}\right|^{1/n}\right] = E\left[\left(\prod_{i=1}^{n} \frac{\rho\,\lambda_i/n}{1+\rho\,\lambda_i/n}\right)^{1/n}\right] \ , \tag{5.50}$$

where $\lambda_1, \cdots, \lambda_n$ are the largest n eigenvalues of matrix $\mathbf{H}^{\mathsf{H}}\,\mathbf{H}$. Finally, the asymptotic distortion (or the system capacity loss) of a finite-rate CSI-quantized MIMO system with spatially equal power allocated transmit beamforming scheme is given by the following form

$$\begin{aligned} &C_{\mathrm{Loss}} = D \geq \widetilde{D}_{\mathrm{Low}} \\ &= \left(\frac{(tn-n^2)\cdot\beta_1}{\ln 2 \cdot (tn-n^2+1)} \cdot \Big(\frac{(tn-n^2)!\,(t-n-1)\tilde{!}\,(n-1)\tilde{!}}{(t-1)\tilde{!}} \Big)^{\frac{1}{tn-n^2}} \right) \cdot 2^{-\frac{B}{tn-n^2}} \ . \end{aligned} \tag{5.51}$$

The optimal point density function $\lambda^*(\mathbf{V})$ that achieves the minimal system distortion is a uniform distribution, which is given by

$$\lambda^*(\mathbf{V}) = \frac{(t-1)\tilde{!}}{\pi^{tn-n^2}(t-n-1)\tilde{!}\,(n-1)\tilde{!}} \, , \qquad \text{for} \quad \mathbf{V} \in \mathcal{V}_{n,t} \quad \& \quad \mathbf{g}\big(\mathbf{V}\big) = \mathbf{0} \ . \tag{5.52}$$

5.3.3 Interesting Observations of the Distortion Lower Bounds

Based on the expressions of the average distortion lower bound $\widetilde{D}_{\text{Low}}$ given by (5.51), the following observations can be made.

1. MIMO system with finite-rate CSI feedback using quantized transmit beamforming scheme is a special case of the analysis provided in Section 5.3.2. In this case, $n = 1$ and $\mathbf{V} \in \mathbb{C}^{t\times 1}$ is the dominant eigenvector of $\mathbf{H}^{\mathsf{H}}\,\mathbf{H}$. The average system distortion can be lower bounded by

$$\widetilde{D}_{\text{Low}}^{\text{BF}} = \left(\frac{(t-1)\cdot\beta_2}{\ln 2\cdot t}\right)\cdot 2^{-\frac{B}{t-1}}\,, \qquad \beta_2 = E\left[\frac{\rho\,\lambda_1}{1+\rho\,\lambda_1}\right]\,, \tag{5.53}$$

where λ_1 is the largest eigenvalue of matrix $\mathbf{H}^{\mathsf{H}}\,\mathbf{H}$. By utilizing the statistical properties of the largest eigenvalues of a central Wishart matrix given in [55], coefficient β_2 can be expressed in a closed-form expression.

To be specific, it was shown in [55] that when the elements of the channel matrix $\mathbf{H}$ are i.i.d. complex Gaussian distributed with zero mean and unit variance, the probability density function of the maximum eigenvalue λ_1 of the Wishart matrix $\mathbf{H}^{\mathsf{H}}\,\mathbf{H}$ is given by the following form

$$f_{\lambda_1}(u) = \frac{1}{\prod_{i=1}^{m}(m-i)!\,(l-i)!}\cdot\frac{d}{d\,u}\det\big(\mathbf{S}(u)\big)\,, \tag{5.54}$$

where $m = \min(t,r)$ and $l = \max(t,r)$. Matrix $\mathbf{S}(u)$ is an $m\times m$ Hankel matrix with its $(i,j)^{\text{th}}$ element given by $S_{i,j}(u) = \Gamma(l-m+i+j-1,u)$, where the incomplete gamma function $\Gamma(k+1,u)$ for $k = 0,1,2,\cdots$, and $u > 0$ has the representation

$$\Gamma(k+1,u) = \int_0^u x^k\,\exp(-x)\;\,d\,x = k!\left(1-e^{-u}\sum_{i=0}^{k}\frac{u^i}{i!}\right)\,. \tag{5.55}$$

The density function $f_{\lambda_1}(u)$ can be written as a finite linear combination of elementary gamma pdfs, i.e.

$$f_{\lambda_1}(u) = \sum_{i=1}^{m}\sum_{j=l-m}^{(l+m)i-2i^2} d_{i,j}\,\left(\frac{i^{j+1}\cdot u^j\cdot e^{-iu}}{j!}\right)\,, \tag{5.56}$$

where $d_{i,j}$ is given by

$$d_{i,j} = \frac{j! \cdot c_{i,j}}{i^{j+1} \cdot \Big(\prod_{i=1}^{m} (m-i)!\,(l-i)! \Big)} \ , \tag{5.57}$$

and $c_{i,j}$ is the coefficient of in front of term $e^{-iu}\,u^j$ when expanding the matrix determinant $\big|\mathbf{S}(u)\big|$. By substituting the density function $f_{\lambda_1}(u)$ into the expectation of β_2, one can obtain the following result

$$\begin{aligned} \beta_2 &= \int_0^{\infty} \left(\frac{\rho\, u}{1+\rho\, u} \right) f_{\lambda_1}(u)\ d\, u \\ &= \sum_{i=1}^{m} \sum_{j=l-m}^{(l+m)i-2i^2} d_{i,j} \cdot \left(\frac{(j+1)\cdot\rho}{i} \right) \cdot {}_2F_0\Big(j+2,\, 1;\, ;\, -\frac{\rho}{i} \Big) \ . \end{aligned} \tag{5.58}$$

As a special case of a 2×4 MIMO system with $t = 4$, $r = 2$, β_2 is given by the following form

$$\begin{aligned} \beta_2 = \rho \cdot \Big(& 12\ {}_2F_0\big(4,1;\,;\,-\rho\big) - 24\,{}_2F_0\big(5,1;\,;\,-\rho\big) \Big) + 20\,{}_2F_0\big(6,1;\,;\,-\rho\big) \\ & -\frac{3}{4}\,{}_2F_0\left(4,1;\,;\,-\frac{\rho}{2}\right) - \frac{3}{4}\,{}_2F_0\left(5,1;\,;\,-\frac{\rho}{2}\right) - \frac{5}{16}\,{}_2F_0\left(6,1;\,;\,-\frac{\rho}{2}\right) \Big) \ . \end{aligned} \tag{5.59}$$

2. MISO system with t transmit antennas and single receive antenna is an even more special case, i.e. $r = n = 1$, where the average system distortion reduces to be the following form

$$\widetilde{D}_{\text{Low}}^{\text{MISO}} = \left(\frac{t-1}{\ln 2} \cdot\ {}_2F_0\big(t+1,\,1;\,;\,-\rho\big) \cdot \rho \right) \cdot 2^{-\frac{B}{t-1}} \ . \tag{5.60}$$

This result is consistent to the analysis provided in [26] and [45].

3. In high-SNR regimes, $\beta_1 \to 1$ and the average system distortion can be represented by

$$\begin{aligned} & \widetilde{D}_{\text{Low}}^{\text{H-SNR}} \\ & = \left(\frac{tn-n^2}{\ln 2 \cdot (tn-n^2+1)} \Big(\frac{(tn-n^2)!\,\widetilde{(t-n-1)!}\,\widetilde{(n-1)!}}{\widetilde{(t-1)!}} \Big)^{\frac{1}{tn-n^2}} \right) \cdot 2^{-\frac{B}{tn-n^2}} \end{aligned} \tag{5.61}$$

It can be shown that in high-SNR regimes, the average distortion (or system capacity loss) of a $t \times r$ MIMO system using a precoder with n beams (with equal power allocation) is exactly the same as that of the same system using a precoder with $(t-n)$ beams. This means that for a MIMO system with t transmit antennas, quantizing the first n singular vectors (matrix $\mathbf{V}$) is equivalent to quantizing the rest $(t-n)$ singular vectors (matrix $\mathbf{V}_2$). In another word, quantizing the orthonormal matrix $\mathbf{V}$ under the constrained condition given by (5.30) is the same as quantizing the projection matrix $\mathbf{V}\mathbf{V}^{\mathsf{H}}$ (or $\mathbf{V}_2\mathbf{V}_2^{\mathsf{H}}$) with $2(tn-n^2)$ degrees of freedom.

4. The average system distortion decreases exponentially with a factor of $2^{-B/(tn-n^2)}$, where the exponential component is inverse proportional to the degrees of freedom of the source variable to be quantized, which is equal to $(2tn-n^2)$. It is interesting to note that the MIMO system has the maximum number of free parameters to quantize when the number of spatial beams used by the transmit pre-coders equals to half of the transmit antennas, i.e. $n = t/2$, but not the minimal number of the transmit and receive antennas $\min(t, r)$. When $n = t/2$, the average system distortion function give by equation (5.51) has the minimal exponential slope $2^{-2B/n^2}$.

5.3.4 Analysis of CSI-Quantizers using Mismatched High-SNR and Low-SNR Codebooks

Revisit of the Codebook Design of the Transmit Pre-coding Matrices

In order to obtain an in-depth understanding of MIMO CSI-quantizers using various codebooks, let us recall some codebook design criterions proposed in [26]. First of all, a generalized mean squared weighted inner product (MSwIP) criterion was proposed in the context that it minimizes the system capacity loss. This criterion can be represented by the following form

$$\max_{\mathcal{C},\mathcal{Q}} E\left[\left\|\widehat{\mathbf{V}}^{\mathsf{H}}\mathbf{V}\cdot\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n}^{\frac{1}{2}}\right\|^2\right] , \qquad \widehat{\mathbf{V}} = \mathcal{Q}(\mathbf{H}) , \tag{5.62}$$

where the maximization is w.r.t. to both the codebook $\mathcal{C}$ as well as the encoder algorithm $\mathcal{Q}$. It is not hard to show that the codebook design criterion given by (5.62) is equivalent to the following criterion

$$\max_{\mathcal{C},\mathcal{Q}} E\left[\mathbf{tr}\Big(\ \Big(I - \big(\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\big)\cdot\big(\mathbf{V}^{\mathsf{H}}\widehat{\mathbf{V}}\big)^{\mathsf{H}}\ \Big)\cdot\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n}\Big)\right]\ , \qquad \widehat{\mathbf{V}} = \mathcal{Q}(\mathbf{H})\ , \quad (5.63)$$

which is directly related to the distortion function D_{Q} considered in this chapter given by equation (5.10).

A drawback of the generalized MSwIP design method is that the codebook is optimized for a particular system SNR ρ. Multiple codebooks are needed for MIMO systems operating in an environment with a wide SNR range. Therefore, two alternative codebook design criterions were also proposed in [26], which do not depend on the system SNR. The first design criterion is called high-SNR criterion, where $\rho \to \infty$ and $\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n} \to I_n$. The optimized high-SNR codebook is designed to maximize the following expectation

$$\max_{\mathcal{C},\mathcal{Q}} E\left[\left\|\widehat{\mathbf{V}}^{\mathsf{H}}\mathbf{V}\right\|^2\right]\ , \qquad \widehat{\mathbf{V}} = \mathcal{Q}(\mathbf{H})\ , \quad (5.64)$$

which is related to the following high-SNR distortion function

$$D_{\mathrm{Q}}^{\text{H-snr}}\big(\mathbf{V},\widehat{\mathbf{V}}\big) = \frac{1}{\ln 2}\cdot\mathbf{tr}\Big(\mathbf{V}^{\mathsf{H}}\left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right)\mathbf{V}\Big)\ . \quad (5.65)$$

Similarly, in the low-SNR regimes, $\rho \to 0$ and $\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n} \to \frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}$. Hence, the low-SNR codebook design criterion is given by

$$\max_{\mathcal{C},\mathcal{Q}} E\left[\left\|\widehat{\mathbf{V}}^{\mathsf{H}}\mathbf{V}\cdot\mathbf{\Sigma}_{\mathrm{H},n}^{\frac{1}{2}}\right\|^2\right]\ , \qquad \widehat{\mathbf{V}} = \mathcal{Q}(\mathbf{H})\ , \quad (5.66)$$

which is again is related to the following low-SNR distortion function

$$D_{\mathrm{Q}}^{\text{L-snr}}\big(\mathbf{V},\widehat{\mathbf{V}}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big) = \frac{1}{\ln 2}\cdot\mathbf{tr}\Big(\mathbf{V}^{\mathsf{H}}\left(I - \widehat{\mathbf{V}}\,\widehat{\mathbf{V}}^{\mathsf{H}}\right)\mathbf{V}\cdot\frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}\Big)\ . \quad (5.67)$$

Mismatched Analysis of High-SNR and Low-SNR Codebooks

By utilizing the mismatched analysis provided in Section 2.4, we provide in this subsection a distortion (or capacity) analysis of MIMO CSI-quantizers

using high-SNR and low-SNR codebooks. First, by the extending the second order expansion results given by (5.42), the complex constrained sensitivity matrix $\mathbf{W}_{\mathrm{c}}^{\text{H-snr}}\big(\mathbf{V}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big)$, which corresponds to distortion function $D_{\mathrm{Q}}^{\text{H-snr}}$ given by (5.65), can be represented by the following form in high-SNR regimes,

$$\mathbf{W}_{\mathrm{c}}^{\text{H-snr}}\big(\mathbf{V}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big) = \frac{1}{\ln 2}\cdot I_{tn-n^2} \ . \tag{5.68}$$

By substituting the mismatched (high-SNR) sensitivity matrix (5.68) into equation (2.67), the mismatched inertial profile of the high-SNR codebook can be obtained as

$$\widetilde{I}_{\text{mis-D}}^{\text{H-snr}}\big(\,\mathbf{V}\,;\,\widetilde{\mathbf{\Sigma}}_{\mathrm{H,n}}\,\big) = \frac{(tn-n^2)\cdot\mathbf{tr}\Big(\widetilde{\mathbf{\Sigma}}_{\mathrm{H},n}\Big)}{\ln 2\cdot(tn-n^2+1)\cdot n\cdot\gamma_0^{1/(tn-n^2)}} \ . \tag{5.69}$$

Moreover, since the optimal point density given by (5.52) is a uniform distribution that does not depend on the system SNR, there is not point density mismatch for quantizers using the high SNR codebook. Finally, by substituting equation (5.69) and (5.52) into the distortion integral given by (2.70), the average system distortion of a MIMO CSI quantizer using high-SNR codebook is given by

$$\widetilde{D}_{\text{mis}}^{\text{H-snr}} = \left(\frac{(tn-n^2)\cdot\beta_3}{\ln 2\cdot(tn-n^2+1)}\Big(\frac{(tn-n^2)!\,(t-n-1)\tilde{!}\,(n-1)\tilde{!}}{(t-1)\tilde{!}}\Big)^{\frac{1}{tn-n^2}}\right)2^{-\frac{B}{tn-n^2}}, \tag{5.70}$$

where coefficient β_3 is given by

$$\beta_3 = E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\rho\,\lambda_i/n}{1+\rho\,\lambda_i/n}\right] \ . \tag{5.71}$$

By utilizing similar derivations, the low-SNR constrained sensitivity matrix can also be obtained

$$\mathbf{W}_{\mathrm{c}}^{\text{L-snr}}\big(\mathbf{V}\,;\,\mathbf{\Sigma}_{\mathrm{H},n}\big) = \frac{\rho}{\ln 2\cdot n}\cdot\mathbf{\Sigma}_{\mathrm{H},n}\otimes I_{t-n} \ , \tag{5.72}$$

which leads the following mismatched inertial profile

$$\widetilde{I}_{\text{mis-D}}^{\text{L-snr}}\big(\,\mathbf{V}\,;\,\widetilde{\mathbf{\Sigma}}_{\mathrm{H,n}}\,\big) = \frac{(tn-n^2)\cdot\Big|\frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}\Big|^{1/n}\cdot\mathbf{tr}\Big(I+\frac{\rho}{n}\mathbf{\Sigma}_{\mathrm{H},n}\Big)^{-1}}{\ln 2\cdot(tn-n^2+1)\cdot n\cdot\gamma_0^{1/(tn-n^2)}} \ . \tag{5.73}$$

Once again, by substituting equation (5.73) and (5.52) into the distortion integral (2.70), the average system distortion of a MIMO CSI quantizer using low-SNR codebook is given by

$$\widetilde{D}_{\text{mis}}^{\text{L-snr}} = \left(\frac{(tn-n^2)\cdot\beta_4}{\ln 2\cdot(tn-n^2+1)} \left(\frac{(tn-n^2)!\,(t-n-1)\tilde{!}\,(n-1)\tilde{!}}{(t-1)\tilde{!}} \right)^{\frac{1}{tn-n^2}} \right) 2^{-\frac{B}{tn-n^2}}, \tag{5.74}$$

where coefficient β_4 is given by

$$\beta_4 = E\left[\left(\prod_{i=1}^{n} \frac{\rho}{n}\lambda_i \right)^{1/n} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\rho\,\lambda_i/n} \right) \right] \quad . \tag{5.75}$$

As a direct results of the above mismatched analysis, MIMO CSI-quantizers using mismatched high-SNR and low-SNR codebooks give rise to the following performance losses:

$$L_{\text{H-snr}} = \frac{\widetilde{D}_{\text{mis}}^{\text{H-snr}}}{\widetilde{D}_{\text{Low}}} = \frac{\beta_3}{\beta_1}, \qquad L_{\text{L-snr}} = \frac{\widetilde{D}_{\text{mis}}^{\text{L-snr}}}{\widetilde{D}_{\text{Low}}} = \frac{\beta_4}{\beta_1} \quad . \tag{5.76}$$

The performance losses $L_{\text{H-snr}}$ and $L_{\text{L-snr}}$ defined in (5.76) can be viewed as a capacity penalty by using the mismatched high-SNR and low-SNR codebooks instead of the optimal codebook designed to match a specific SNR point. Both losses can be shown to be greater than one, i.e. $L_{\text{H-snr}}$, $L_{\text{L-snr}} \geq 1$, and are independent of the quantization resolution (feedback rate) B.

5.3.5 Analysis of MIMO Pre-Coding Schemes with Multi-Mode spatial Multiplexing Strategy

In order to compensate the degradations due to the equal power allocation among the spatial beams used by the transmit pre-coder, the multi-mode spatial multiplexing (MMSM) scheme was proposed in [26], where the number of active spatial beams adopted by the transmitter is adjusted adaptively accordingly to the current system SNR. As an example, we plot in Fig. 5.1 the normalized capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. fading channels with finite-rate CSI feedback of $B = 8$ bits per channel update. The normalized MIMO capacity

is defined to be the ratio of the system capacity with quantized CSI to that of a system using optimal transmit pre-coder with ideal CSI at the transmitter. The proposed multi-mode transmission strategy is employed for the simulation, where the MIMO transmit precoder used for each mode has n active spatial beams with equal power allocation. For this particular case, there are total $\min(t, r) = 3$ modes available for the current MIMO system, i.e. $1 \leq n \leq 3$. The codebooks of the CSI quantizer used at each mode are generated by the so-called generalized mean-squared weighted inner-product (MSwIP) criterion proposed in [26]. It can be observed from Fig. 5.1 that by switching the modes based on the SNR, one can therefor make the best of each modes and the system capacity of using the MMSM scheme is the maximum capacity of all the available modes.

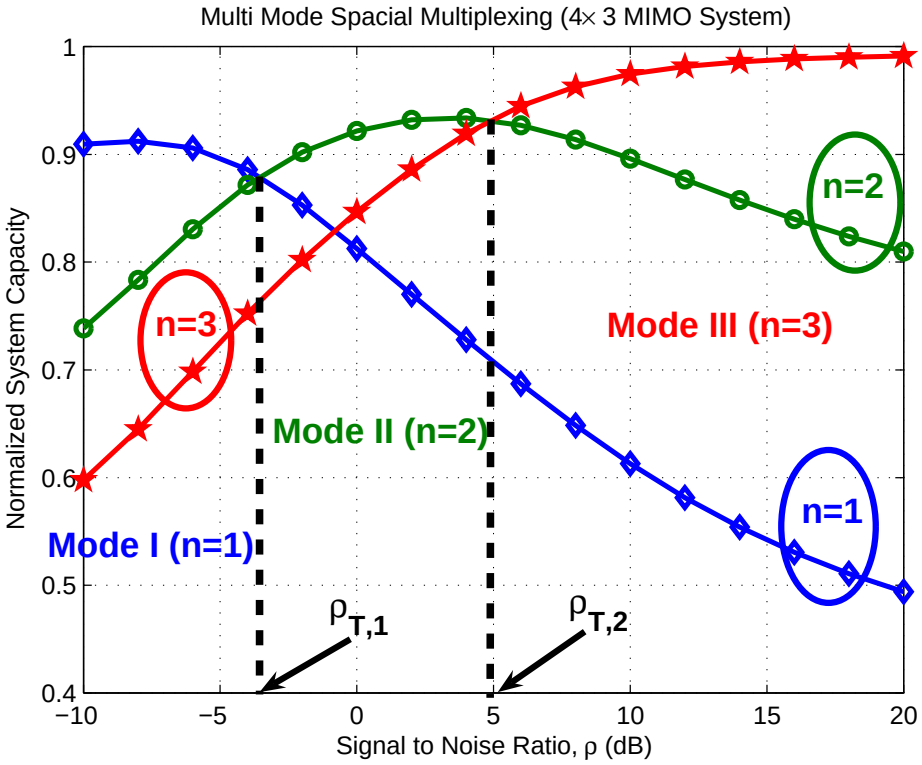

Figure 5.1: Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels with finite-rate CSI feedback ($B = 8$), and using multi-mode spatial multiplexing transmission schemes.

As a direct result of the high-rate analysis obtained in Section 5.3.4, the system capacity of a $t \times r$ MIMO system with finite-rate CSI feedback of B bits per channel update, and using MMSM transmission scheme with high-SNR codebooks can be represented by the following form

$$C_{\text{MMSM}} = \max_{1 \le n \le \min(t,r)} \left(E\left[\sum_{i=1}^{n} \log_2 \left(1 + \frac{\rho}{n} \lambda_i \right) \right] - \alpha_n \cdot 2^{-\frac{B}{tn-n^2}} \right) \quad , \tag{5.77}$$

where α_n is a coefficient that depends on t, r, n and ρ, which is given by

$$\alpha_n(\rho) = \frac{(tn - n^2) \cdot \beta_3}{\ln 2 \cdot (tn - n^2 + 1)} \cdot \Big(\frac{(tn-n^2)! \, \tilde{(t-n-1)!} \, \tilde{(n-1)!}}{\tilde{(t-1)!}} \Big)^{\frac{1}{tn-n^2}} \quad , \tag{5.78}$$

with β_1 given by equation (5.71). Consequently, for a particular operating SNR of the system, which is assumed to change at a much slower rate than the channel itself, the best transmission mode is given by

$$n_{\text{opt}} = \arg \max_{1 \le n \le \min(t,r)} \left(E\left[\sum_{i=1}^{n} \log_2 \left(1 + \frac{\rho}{n} \lambda_i \right) \right] - \alpha_n \cdot 2^{-\frac{B}{tn-n^2}} \right) \quad . \tag{5.79}$$

Based on the analysis result of the MMSM pre-coder given by (5.77) and (5.79), the boundary points of the mode transitions (such as $\rho_{\text{T},1}$ and $\rho_{\text{T},2}$ in Fig. 5.1) can be calculated analytically without actual simulations.

5.4 Numerical and Simulation Results

5.4.1 High-Rate Capacity Analysis

Some numerical experiments are now presented to provide a better feel for the utility of the distortion analysis. Fig. 5.2 shows the capacity loss due to the finite-rate quantization of the CSI versus feedback rate B for a 4×2 MISO system ($t = 4$, $r = 2$) over i.i.d. Rayleigh fading channels under different system SNRs at $\rho = -10, 0$ and 20 dB, respectively. The transmit precoder used for the MIMO system has $n = 2$ spatial beams with equal power allocations. The codebook of the CSI quantizer is generated the generalized mean-squared weighted inner-product (MSwIP) criterion [26]. The distortion lower bounds $\widetilde{D}_{\text{Low}}$ given by equation (5.51)

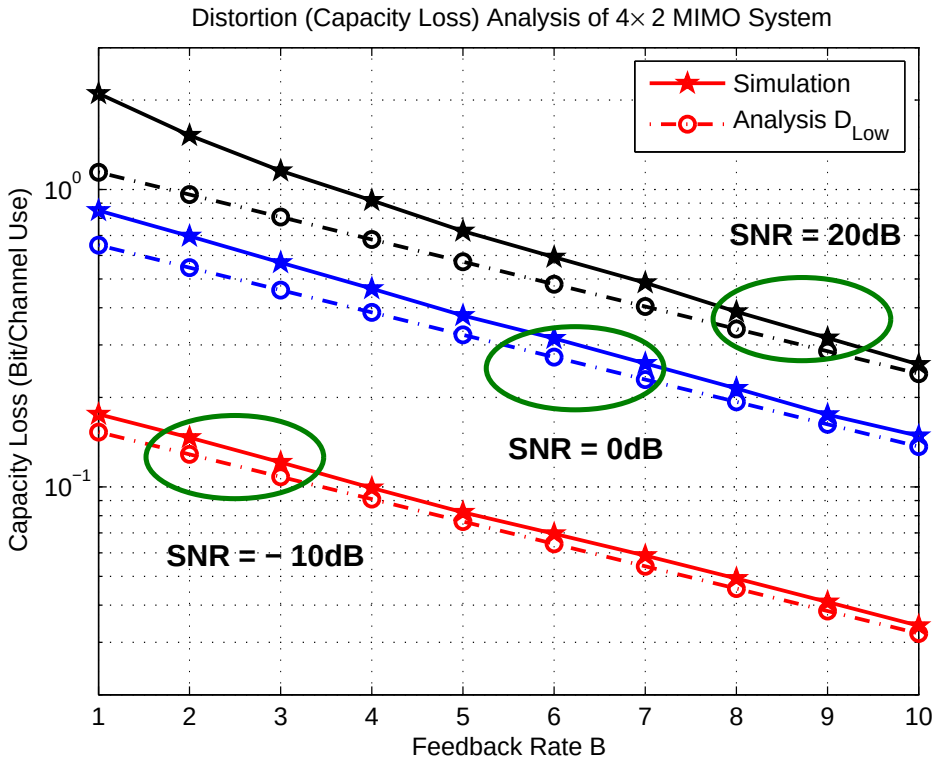

Figure 5.2: Capacity loss versus CSI feedback rate B of a 4×2 MIMO system ($t = 4$, $r = 2$ and $n = 2$) over i.i.d. Rayleigh fading channels and with signal to noise ratio $\rho = -10, 0$ and 20dB.

are also included in the plot for comparisons. It can be observed from the plot that the proposed distortion (or system capacity loss) lower bounds are tight and predict very well the actual system capacity loss obtained from Monte Carlo simulations.

5.4.2 Analysis of Mismatched High-SNR and Low-SNR Codebooks

In order to understand the performance degradation caused by the mismatched CSI-quantizers using high-SNR and low-SNR codebooks, we plot in Fig. 5.3 the performance losses $L_{\text{H-snr}}$ and $L_{\text{L-snr}}$ versus the system SNR ρ of a 4×3 MIMO system ($t = 4$, $r = 3$) with finite-rate CSI feedback of $B = 8$ bits per channel update. The performance loss $L_{\text{H-snr}}$ and $L_{\text{L-snr}}$) represents the ratio of the aver-

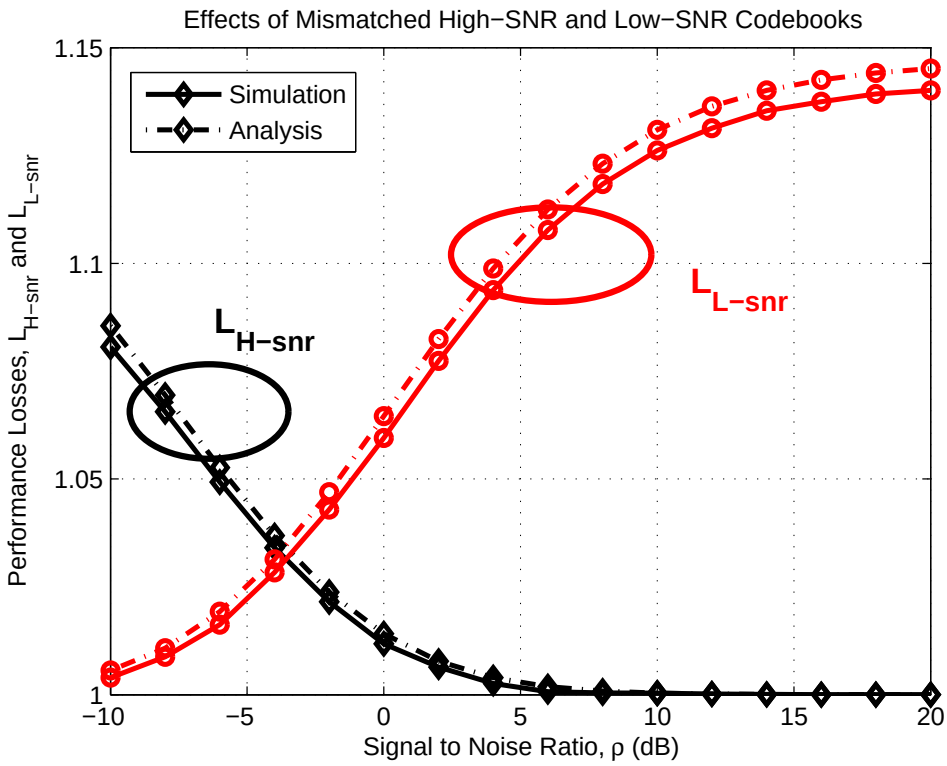

Figure 5.3: Performance losses ($L_{\text{H-snr}}$ and $L_{\text{L-snr}}$) versus signal to noise ratio ρ of a 4×3 MIMO system ($t = 4$, $r = 3$, and $n = 2$ over i.i.d. Rayleigh fading channels with feedback rate $B = 8$ bits per channel update.

age system distortion of a mismatched quantizer to that of the optimal quantizer, whose definition is given by equation (5.76). The transmit pre-coder used for the MIMO system has $n = 2$ spatial beams with equal power allocations. The codebook of the CSI quantizer is also generated the generalized MSwIP criterion. For comparison purpose, the ratios of the distortion bounds, i.e. $\widetilde{D}_{\text{mis}}^{\text{H-snr}}/\widetilde{D}_{\text{Low}}$ and $\widetilde{D}_{\text{mis}}^{\text{L-snr}}/\widetilde{D}_{\text{Low}}$, are also included in the plot. It can be observed from Fig. 5.3 that the obtained performance losses (or system distortion ratios) agree well with the simulation results.

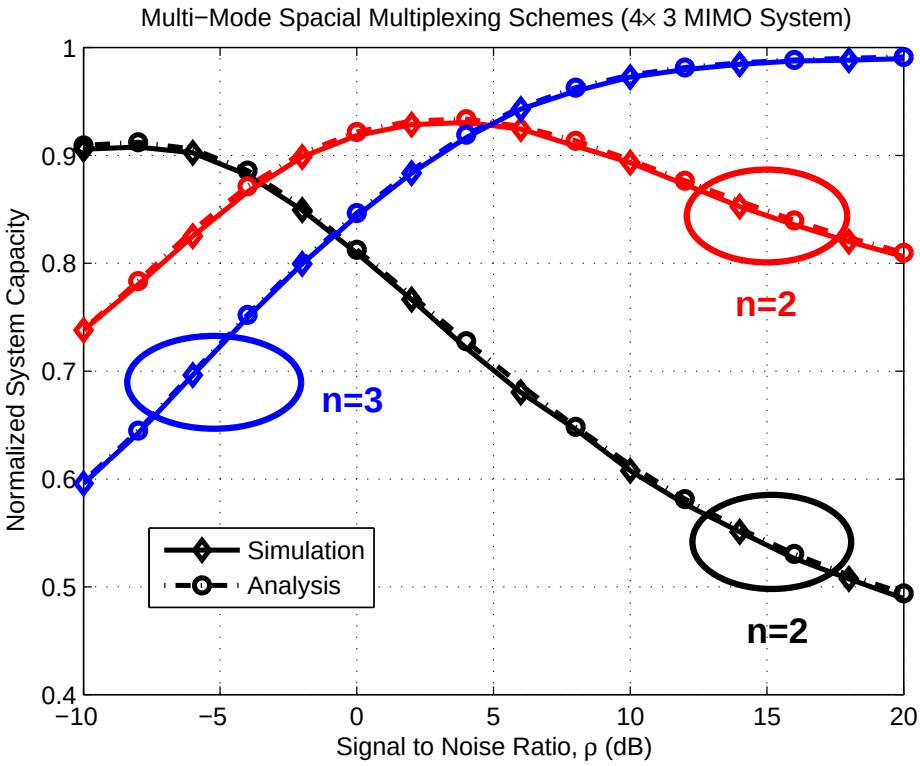

Figure 5.4: Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels with feedback rate $B = 8$ bits per channel update, and using multi-mode spatial multiplexing (MMSM) transmission schemes.

5.4.3 Performance of Multi-Mode spatial Multiplexing Schemes

In order to see the utility of the proposed distortion analysis to MIMO systems using multi-mode spatial multiplexing transmission schemes, we demonstrate in Fig. 5.4 the normalized capacity of the same 4×3 MIMO system ($t = 4$, $r = 3$), which is described in Section 5.3.5, over i.i.d. Rayleigh fading channels with $B = 8$ bits CSI feedback. The MIMO pre-coder again employs the MMSM scheme, with total three modes available ($n = 1, 2, 3$). Both the capacity analysis given by equation (5.77) as well as the results obtained from Monte Carlo simulations are shown in Fig. 5.4. It can be observed from the plot that the proposed capacity analysis closely matches the simulation results, where the two curves almost fall

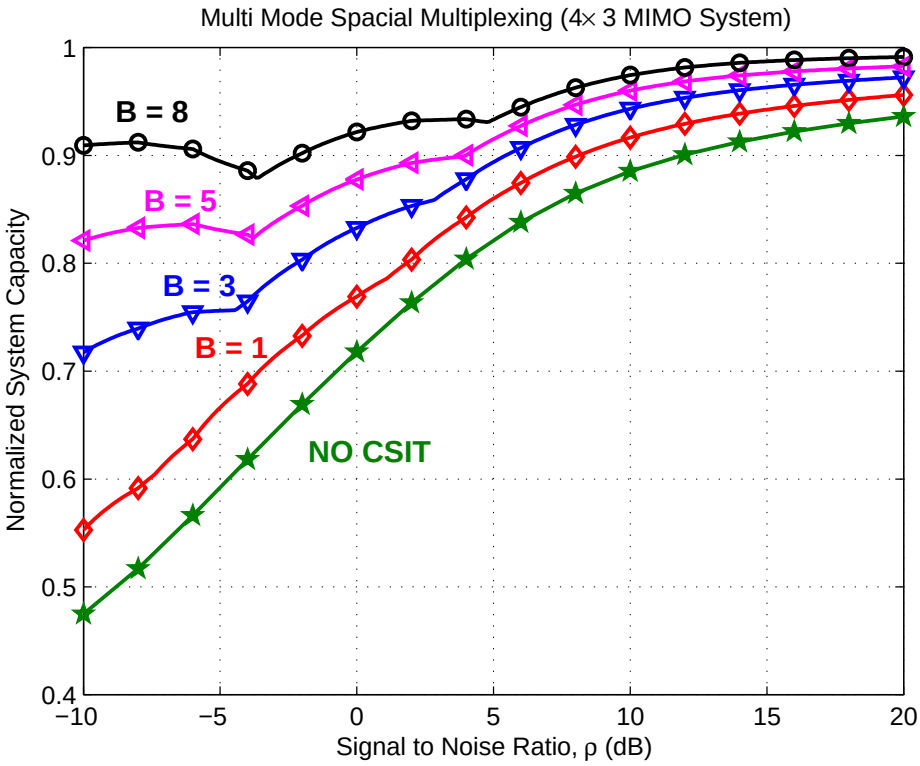

Figure 5.5: Normalized system capacity of a 4×3 MIMO system ($t = 4$, $r = 3$) over i.i.d. Rayleigh fading channels using MMSM transmission scheme, and with several different CSI feedback rate ($B = 1, 3, 5, 8$ bits per channel update).

on top of each other.

We also demonstrate in Fig. 5.5 the analytical results of the normalized system capacity of the same 4×3 MIMO system using MMSM transmission schemes but with different rate of CSI feedback of $B = 1, 3, 5, 8$ bits per channel update. For the sake of comparison, we also include in the plot the normalized capacity of MIMO system with no CSI feedback, which corresponds to the case where no CSIT is available and the MIMO transmitter sends independent data stream on each of its antennas with equal power allocations. It can observed from Fig. 5.5 that the system capacity improves significantly as the feedback rate B increases. To be specific, we can see that with a feedback rate of $B = 8$ bits, a

4×3 MIMO system with MMSM scheme can almost achieve %90 of the capacity of a system with ideal CSIT. Compared with the total free dimensions of the original CSI information $\mathbf{H}$, which is 24, only 1/3 bits per dimension is needed for a properly designed MIMO CSI feedback scheme. Therefore, as a rough conclusion, in order to achieve a performance in terms of capacity close to that of systems with ideal CSIT, only limited CSI feedback rate per dimension is required.

5.5 Summary

This chapter employs a high resolution quantization framework to study the effects of finite-rate quantization of the channel state information (CSI) on the performance of MIMO systems over i.i.d. Rayleigh flat fading channels. Specifically, tight lower bounds on the capacity loss of MIMO systems due to the finite-rate channel quantization were provided. The obtained analytical results reveal an interesting fact that the system capacity loss decreases exponentially as the ratio of the quantization rate to the total number of degrees of freedom of the channel state information to be quantized. Moreover, MIMO CSI-quantizers with mismatched codebooks that only optimized for high-SNR and low-SNR regimes were investigated. The performance analysis of the sub-optimal CSI-quantizer analytically quantifies the penalties caused by the mismatched codebooks. As a further application of the obtained distortion analysis, the performance of MIMO systems using multi-mode spatial multiplexing transmission schemes with finite-rate CSI feedback were also provided. Finally, numerical and simulation results were presented which confirm the tightness of theoretical distortion bounds. The text of this chapter is in part a reprint of the paper which was coauthored with Bhaskar D. Rao and has been accepted in *Proceedings of IEEE Asilomar Conference 2006*, and will be submitted for publication in *IEEE Transactions on Signal Processing* under the title *"Analysis of MIMO systems with finite-rate channel state information feedback"*.

6 Capacity Analysis of MIMO Systems with Unknown Channel State Information

6.1 Motivation

In order to meet the increasing demands of high speed data services required by next-generation communication systems, considerable effort is being expended to develop advanced system architectures and algorithms that can support high data rate communications. Using multiple antennas at both the transmitter and the receiver is one of the most promising techniques that can offer significant increases in channel capacity of a communication system in a wireless fading environment [2, 3, 5, 56]. However, the capacity gains reported are based on the assumption that the fading channel coefficients between each transmit and receive antenna pairs are perfectly known at the receiver at no cost, which is not a reasonable assumption for most practical communication systems especially for fast fading channels.

Recently, several researchers have considered the capacity of non-coherent fading channels, where neither the transmitter nor the receiver has channel state information. Abou-Faycal et al. provide in [57] that the capacity achieving distribution for the single-input single-output (SISO) unknown fading channel is discrete and with a finite number of mass points. Lapidoth and Moser provide in [58] the

asymptotically tight upper and lower capacity bounds for this unknown SISO channel and Chen et al. provide in [59] the optimal input distribution for the unknown Gaussian Markov channel. Marzetta and Hochwald provide in [6] the capacity analysis of an unknown block fading MIMO channel with a finite coherent time interval T. They showed that there is no benefit in making the number of transmit antennas M greater than the length of the coherent time T, and that the capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix and a $T \times M$ random diagonal matrix with real, nonnegative diagonal elements. Zheng and Tse [7] extend the SISO asymptotic results to the MIMO case, and compute the asymptotic capacity of the non-coherent MIMO channel at high signal to noise ratios in terms of M, N and T and show that the capacity gain is $M^\star(1-M^\star/T)$ bits per second per hertz for every 3-dB increase in SNR, where the optimal number of transmit antennas is $M^\star = \min(M, N, \lfloor T/2 \rfloor)$. Furthermore, Liang and et al. study in their recent paper [60] the non-coherent channel capacity for time-selective fading channels, and provide the asymptotic first order term of the high SNR expansion of the capacity.

All the theoretical analysis so far provides the fundamental transmission limits of the non-coherent channel, which shed interesting insight into what is feasible. However, in practice finding the optimal input distribution to achieve capacity is an involved task which requires difficult numerical optimization. Furthermore, the known block fading signaling schemes for non-coherent MIMO channels generally fall into into two categories: MIMO differential modulation [61]– [63] and unitary space-time modulation (USTM) scheme [64]– [70]. Both of these schemes cannot approach the non-coherent MIMO capacity limit due to their suboptimal code structure and in the later case (USTM) only achieves asymptotic (or the diversity) optimality in high SNR regimes and suffers from exponential decoding complexity. Therefore, we adopt a more pragmatic approach and focus on systems that are able to take advantages of the existing channel estimation algorithms

and the powerful forward error correction coding techniques, like turbo or LDPC codes. In this context, Hassibi and Hochwald propose in [71] a channel model that separates one coherent block into two phases: training and data. Based on the two phase channel model, as well as by applying the MMSE channel estimation algorithm, they provide a capacity lower bound for the unknown MIMO channel. Through the analysis of the proposed capacity lower bound, they demonstrate how training affects the capacity and show that the optimal number of training symbols is equal to $T_\tau = M$ when the training and data powers are allowed to vary.

The capacity lower bound provided in [71] assumes the channel estimation (linear minimal mean square error (LMMSE) channel estimator) is obtained by only using the training symbols, and therefore does not make use of the channel information contained in the received data symbols. Consequently, the lower bound is pessimistic. Our aim in this chapter is to develop an improved lower bound that better represents the system capacity behavior and to carry out optimization over the different system parameters to further extend and confirm the observations made in [71]. To this end, we first propose a mutual information upper bound for the unknown MIMO channel under the assumption that the input distribution is restricted to a certain structure and form but without assuming any specific channel estimation algorithm. The proposed upper bound is shown to have a fast convergence rate to the system mutual information as the number of transmit antennas M increases, and then becomes an improved capacity lower bound of the unknown MIMO channel. Through the analysis of the mutual information upper bounds (or the improved capacity lower bounds) with respect to different system parameters, we show that the orthogonal pilots structure not only minimizes the mean square estimation error, but also maximizes the mutual information rate upper bound. We also prove that the mutual information upper bound is a monotonically decreasing function with respect to the number of pilot symbols T_τ. Furthermore, numerical evaluation of the upper bounds also show that there is an insignificant rate increment when T_τ decreases below the number of transmit

antennas M. By setting the number of pilot symbols T_τ equal to M, which is a good trade-off point between achievable capacity and system complexity, we show that there is no benefit in making the number of transmit antennas M greater than N. Numerical results also show that there is an insignificant amount of rate gain by utilizing optimal power allocations between pilot and data symbols compared to equal power allocation schemes in moderate to high signal to noise ratio (SNR) ranges.

6.2 System Model

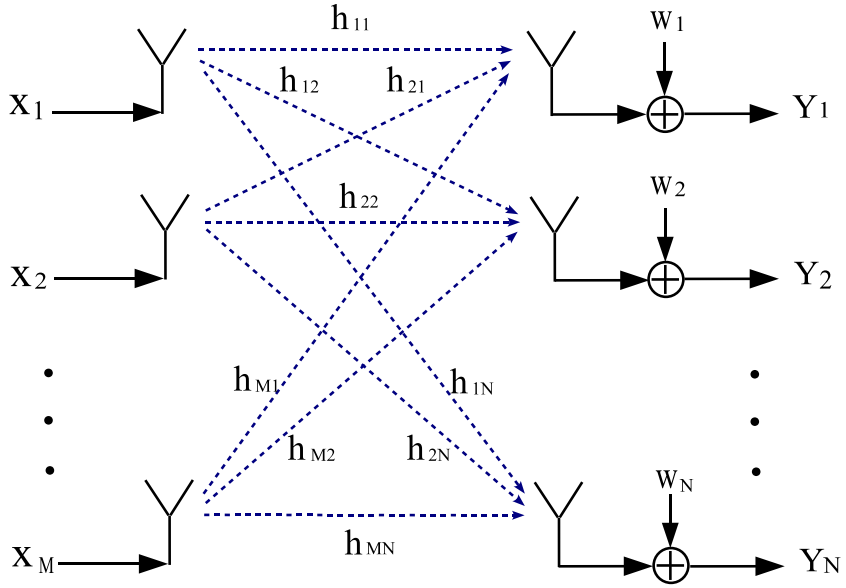

Figure 6.1: MIMO system model composed of M transmit antennas and N receive antennas

We consider a MIMO system with M transmitter antennas and N receive antennas, signaling through a frequency flat fading channel with i.i.d channel coefficient between each transmit and receive antenna pairs. The system model is

illustrated in Fig. 6.1. Furthermore, the MIMO channel is assumed to be block fading, where the fading coefficient $\mathbf{H}$ remains static within a coherent time interval of T symbol periods, and varies independently from one coherent time block to another. Hence, the signal model can be written in the following form

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{H} + \mathbf{w} \ , \tag{6.1}$$

where $\mathbf{Y}$ is a $T \times N$ received complex signal matrix, $\mathbf{X}$ is a $T \times M$ transmitted complex signal matrix, $\mathbf{H}$ a $M \times N$ complex channel matrix, and $\mathbf{w}$ is a $T \times N$ matrix of additive Gaussian noise. Both matrix $\mathbf{H}$ and $\mathbf{w}$ have zero mean unit variance independent complex Gaussian entries. We also assume that the entries of the transmitted signal matrix $\mathbf{X}$ have the following average power constraint,

$$\frac{1}{T} \cdot E\big[\mathbf{tr}\big(\mathbf{X}^H\mathbf{X}\big)\big] = \rho \ , \tag{6.2}$$

where ρ is the average received signal to noise ratio at each receive antenna. According to the above non-coherent MIMO fading channel setup (6.1), the channel transitional probability (or the conditional probability density of the received signal $\mathbf{Y}$ given the transmitted signal $\mathbf{X}$) is given by [6],

$$p\big(\mathbf{Y}\big|\mathbf{X}\big) = \frac{\exp\left(-\mathbf{tr}\Big\{\Big[I_T + \mathbf{X}\mathbf{X}^H\Big]^{-1} \cdot \mathbf{Y}\mathbf{Y}^H\Big\}\right)}{\pi^{TN}\det^N\Big[I_T + \mathbf{X}\mathbf{X}^H\Big]} \ . \tag{6.3}$$

Therefore, the capacity (or the maximum mutual information rate) of the unknown MIMO channel can be represented as the following optimization problem,

$$C = \max_{p(\mathbf{X})\in\mathbb{P}_\rho} \ \frac{1}{T}\ I(\mathbf{X};\mathbf{Y}) \ , \qquad \mathbb{P}_\rho = \Big\{p(\mathbf{X})\Big|E\big[\mathbf{tr}\big(\mathbf{X}^H\mathbf{X}\big)\big]/T \le \rho\Big\} \ . \tag{6.4}$$

It was shown in [6] that the capacity of the unknown channel described in equation (6.4) is achieved when the $T \times M$ transmitted signal matrix $\mathbf{X}$ is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix and a certain $T \times M$ random matrix that is diagonal, real, and nonnegative. However, finding the optimal input distribution is an involved task. Furthermore, there are no known space-time codes that can approach this capacity.

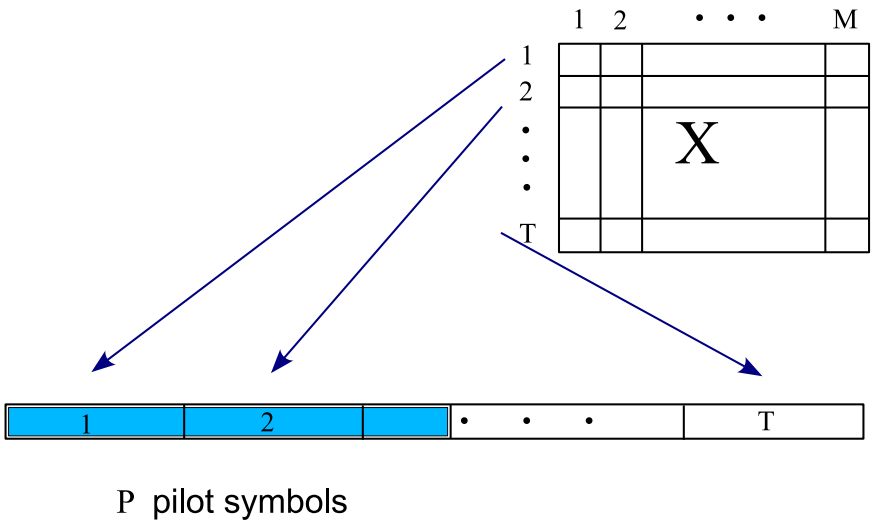

Figure 6.2: Transmitted symbol structure of the MIMO system

In this chapter, we restrict our attention to a conventional MIMO system with the input signal matrix having a two-phase (training followed by data) structure proposed and utilized for capacity analysis in [71]. The symbol structure of the transmitted signal $\mathbf{X}$ is illustrated in Fig. 6.2, where the first p symbols are training pilots, followed by $(TM - p)$ data symbols. For the sake of simplicity, we only consider the case where the pilot numbers $p = T_\tau \times M$, a multiple of the number of transmit antennas M. Hence, the transmitted signal $\mathbf{X}$ can be separated into two sub-matrices: training followed by data, which is represented as

$$\mathbf{X} = \begin{bmatrix} \left(\rho_\tau / M\right)^{\frac{1}{2}} \cdot \mathbf{S}_\tau \\ \left(\rho_d / M\right)^{\frac{1}{2}} \cdot \mathbf{X}_d \end{bmatrix}, \tag{6.5}$$

where $\mathbf{S}_\tau$ is the fixed pilot symbols and $\mathbf{X}_d$ is the information bearing data symbols, whose structures are given by

$$\begin{aligned} \mathbf{S}_\tau &= \left[\mathbf{s}_1^H, \cdots, \mathbf{s}_{T_\tau}^H\right]^H, & \mathbf{S}_\tau &\in \mathbb{C}^{T_\tau \times M}, \\ \mathbf{X}_d &= \left[\mathbf{x}_1{}^H, \cdots, \mathbf{x}_{T_d}{}^H\right]^H, & \mathbf{X}_d &\in \mathbb{C}^{T_d \times M}. \end{aligned} \tag{6.6}$$

Conservation of time and energy leads to the following constraints,

$$\begin{aligned} \mathbf{tr}\big(\mathbf{S}_\tau^H \cdot \mathbf{S}_\tau\big) = MT_\tau, \qquad & E_{\mathbf{X}_d}\Big[\mathbf{tr}\big({\mathbf{X}_d}^H \cdot \mathbf{X}_d\big)\Big] = MT_d \ , \\ T = T_\tau + T_d, \qquad & \rho T = \rho_\tau T_\tau + \rho_d T_d \ . \end{aligned} \tag{6.7}$$

In the following sections, we provide analysis of the system mutual information rate (or system capacity) based on the two-phase input signal structure described above.

6.3 Improved Capacity Lower Bound of Unknown MIMO Channels

6.3.1 Upper Bound of System Mutual Information Rate

The non-coherent nature of the MIMO fading channel imposes great challenge in measuring the system capacity and we therefore focus our attention on analyzing the mutual information rate of the unknown MIMO system with input signal having a two-phase structure described in Section 6.2. However, even under this simplified signaling structure, the optimal input distribution of data $\mathbf{X}_d$ that maximizes the mutual information between input $\mathbf{X}_d$ and output $\mathbf{Y}$ is also analytically intractable. In principle, the optimal input distribution of $\mathbf{X}_d$ depends on the system parameters such as pilots structure $\mathbf{S}_\tau$, data and training slot allocations (T_d, T_τ), power allocations (ρ_τ, ρ_d), as well as number of transmit and receiver antennas (M, N), and can be evaluated only through involved numerical optimizations. In this chapter, we simplify the problem by designing (or optimizing) the system parameters for a given input distribution. For the sake of analytical tractability, we assume in this chapter that the input data matrix $\mathbf{X}_d$ has a multivariate Gaussian distribution, denoted as $g(\mathbf{X}_d)$, with each of its row vectors $\mathbf{x}_i$ being i.i.d. Gaussian distributed, i.e.

$$g(\mathbf{X}_d) \qquad \Longleftrightarrow \qquad \Big\{\mathbf{x}_i \sim \mathcal{N}_c\big(\mathbf{0}\,,\, I_M\big), \quad E\Big[\mathbf{x}_i^H \cdot \mathbf{x}_j\Big] = \delta_{i,j} \cdot I_M \Big\} \ . \tag{6.8}$$

It is the same assumption of the input distribution as in [71], which is equivalent as saying that the transmit antennas send independent data streams with equal power allocation.

In order to avoid possible confusions with the actual channel capacity C, we denote R as the mutual information rate between $\mathbf{X}$ and $\mathbf{Y}$ when the input signal has a two-phase structure with data signal $\mathbf{X}_d$ having distribution $g(\mathbf{X}_d)$ (denoted as $\mathbf{X}_d \sim g(\mathbf{X}_d)$), i.e.

$$R = \frac{1}{T}\, I\big(\mathbf{X};\mathbf{Y}\big) = \frac{1}{T}\, I\big(\mathbf{X}_d;\mathbf{Y}\big)\ . \tag{6.9}$$

It is obvious that the by restricting the input distribution (6.8) to a certain structure and form, the mutual information rate R between input signal $\mathbf{X}$ and output signal $\mathbf{Y}$ is not optimized and is always a lower bound of the unknown MIMO capacity C. However, the restricted input signal structure does allow us to obtain a tractable upper bound of the mutual information rate R as described in the following proposition.

Proposition 6 *The mutual information rate R between $\mathbf{X}$ and $\mathbf{Y}$ with input distribution $g(\mathbf{X}_d)$ is upper bounded by,*

$$\begin{aligned} R \le \overline{R} = \frac{N}{T}\Bigg(& \log_2\Big|I_M + \frac{\rho_\tau}{M}\mathbf{S}_\tau^H\mathbf{S}_\tau\Big| + T_d\cdot\log_2(1+\rho_d) \\ & -E_{\mathbf{X}_d}\bigg[\log_2\Big|I_M + \frac{\rho_\tau}{M}\mathbf{S}_\tau^H\mathbf{S}_\tau + \frac{\rho_d}{M}\mathbf{X}_d^H\mathbf{X}_d\Big|\bigg]\Bigg), \end{aligned} \tag{6.10}$$

where the expectation $E_{\mathbf{X}_d}\big[\,\cdot\,\big]$ is taken with respect to data $\mathbf{X}_d$.

Proof: First, conditioned on any input data sequences $\mathbf{X}_d$ (or $\mathbf{X}$), $\mathbf{vec}\big(\mathbf{Y}\big)$ is a Gaussian distributed vector[1] of zero mean and variance

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}} &= \mathbf{Cov}\Big(\mathbf{vec}\big(\mathbf{Y}\big)\big|\mathbf{X}\Big) = \big(I_N\otimes\mathbf{X}\big)\cdot\big(I_N\otimes I_T\big)\cdot\big(I_N\otimes\mathbf{X}^H\big) + \big(I_N\otimes I_T\big) \\ &= I_N\otimes\Big(\mathbf{X}\mathbf{X}^H + I_T\Big)\ . \end{aligned} \tag{6.11}$$

[1]For a matrix $\mathbf{X}$ of size $m\times n$, $\mathbf{vec}\big(\mathbf{X}\big)$ is the $mn\times 1$ vector defined as $\mathbf{vec}\big(\mathbf{X}\big) = \Big[\mathbf{x}_1^T,\cdots,\mathbf{x}_n^T\Big]^T$, where $\mathbf{x}_i$, $i=1,\cdots,n$ is the i^{th} column of $\mathbf{X}$.

Taking expectation of (6.11) with respect to $\mathbf{X}_d$, the covariance matrix of $\mathbf{vec}\big(\mathbf{Y}\big)$ is obtained as,

$$\begin{aligned}\mathbf{\Sigma_Y} &= \mathbf{Cov}\Big(\mathbf{vec}\big(\mathbf{Y}\big)\Big) = E_{\mathbf{X_d}}\Big[\mathbf{Cov}\Big(\mathbf{vec}\big(\mathbf{Y}\big)\big|\mathbf{X}\Big)\Big] \\ &= I_N \otimes \Big(E_{\mathbf{X}_d}\Big[\mathbf{X}\mathbf{X}^H\Big] + I_T\Big) = I_N \otimes \left(\left[\begin{array}{c|c} \frac{\rho_\tau}{M}\cdot \mathbf{S}_\tau\mathbf{S}_\tau^H & \mathbf{0} \\ \hline \mathbf{0} & \rho_d \cdot I_{T_d} \end{array}\right] + I_T \right).\end{aligned} \tag{6.12}$$

Due to the fact that Gaussian distribution has the maximum entropy among any vector distributions with the same covariance matrix, entropy $h(\mathbf{Y})$ can be upper bounded by

$$h\big(\mathbf{Y}\big) \le \log_2\Big((\pi e)^{NT}\cdot\big|\mathbf{\Sigma_Y}\big|\Big)\ . \tag{6.13}$$

Therefore, we have the following mutual information upper bound

$$\begin{aligned} R &= \frac{1}{T}\,I\Big(\mathbf{X};\mathbf{Y}\Big) = \frac{1}{T}\Big(h\big(\mathbf{Y}\big) - h\big(\mathbf{Y}\big|\mathbf{X}\big)\Big) \\ &\overset{a}{=} \frac{1}{T}\left(h\big(\mathbf{Y}\big) - E_{\mathbf{X}_d}\Big[\log_2\Big((\pi e)^{NT}\cdot\big|\mathbf{\Sigma_{Y|X}}\big|\Big)\Big]\right) \\ &\overset{b}{\le} \frac{1}{T}\left(\log_2\Big((\pi e)^{NT}\cdot\big|\mathbf{\Sigma_Y}\big|\Big) - E_{\mathbf{X}_d}\Big[\log_2\Big((\pi e)^{NT}\cdot\big|\mathbf{\Sigma_{Y|X}}\big|\Big)\Big]\right) \\ &= \frac{N}{T}\left(\log_2\Big|I_M + \frac{\rho_\tau}{M}\mathbf{S}_\tau^H\mathbf{S}_\tau\Big| + T_d\cdot\log_2(1+\rho_d)\right. \\ &\qquad \left. -E_{\mathbf{X}_d}\Big[\log_2\Big|I_M + \frac{\rho_\tau}{M}\mathbf{S}_\tau^H\mathbf{S}_\tau + \frac{\rho_d}{M}\mathbf{X}_d^H\mathbf{X}_d\Big|\Big]\right), \end{aligned} \tag{6.14}$$

where the second term of equality (a) is from a direct expansion of the conditional entropy according to the definition, and inequality (b) is from (6.13). ■

As an example, we demonstrate in Fig. 6.3 the actual non-coherent MIMO system mutual information R (obtained by Monte Carlo simulation) and the proposed mutual information upper bound $\overline{R}$ of a 2×2 MIMO system over unknown fading channels. For comparison purpose, we also include in the plot the MMSE-based capacity (or mutual information rate) lower bound provided in [71]. The 2×2 non-coherent MIMO system considered here has signal to noise ratio

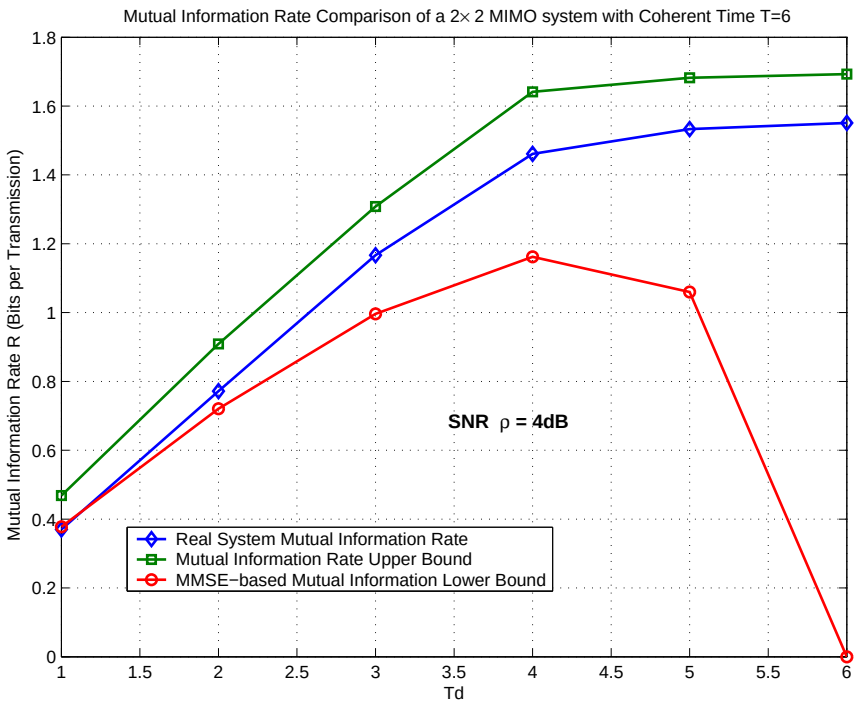

Figure 6.3: Mutual information rate comparison (between actual system mutual information rate, MMSE-based capacity (or mutual information) lower bound, and the proposed mutual information rate upper bound) of a 2×2 MIMO system with channel coherent time $T = 6$ and signal to noise ratio $\rho_\tau = \rho_d = 4dB$.

$\rho_\tau = \rho_d = 4dB$, and with channel coherent time $T = 6$. From Fig. 6.3, we can observe that the mutual information upper bound $\overline{R}$ can be viewed as a shifted version (vertical direction) of the actual mutual information R and hence represents the non-coherent MIMO mutual information rate more accurately than the MMSE-based lower bound. From the plot, we also notice that the MMSE-based lower bound matches to the system mutual information rate more closely when $T\tau$ is large (or T_d is small), which is due to improved channel estimation made possible by the presence of a large number of training symbols. However, as the number of training symbol T_τ decreases (or T_d increases), the MMSE-based lower bound diverges from the system mutual information rate (as well as the capacity) due to its two-phase processing limitation (training followed by detection). This is the high

capacity (or information rate) region that is likely to be of interest. Therefore, the proposed mutual information upper bound $\overline{R}$ can be viewed as an improved measurement of the system mutual information compared to the MMSE-based lower bound, and hence provides a better representation of the mutual information (or capacity) behavior of MIMO systems with unknown channel state information.

6.3.2 Tight Mutual Information Rate Upper Bound Leads to Improved Capacity Lower Bound

In order to study the tightness of the mutual information rate upper bound, we now examine the nature of the approximation made in arriving at (6.14) (inequality (b)). The received signal $\mathbf{Y}$ can be viewed as a continuous multivariate Gaussian mixture consists of an uncountably infinite number of Gaussian components, whose distribution is given by,

$$p\big(\mathbf{Y}\big) = E_{\mathbf{X}_d}\left[\frac{\exp\Big(-\mathbf{tr}\Big\{\big[I_T + \mathbf{X}\mathbf{X}^H\big]^{-1}\cdot\mathbf{Y}\mathbf{Y}^H\Big\}\Big)}{\pi^{TN}\det^N\big[I_T+\mathbf{X}\mathbf{X}^H\big]}\right] \ . \tag{6.15}$$

Therefore, the tightness of the upper bound provided in Proposition 6 depends on the closeness of the above mixture density to the Gaussian density with the same covariance matrix. By expanding the received signal structure in the following manner

$$\mathbf{Y} = \mathbf{X}\mathbf{H} + \mathbf{W} = \mathbf{V} + \mathbf{W} = \left[\begin{array}{c}\mathbf{V}_1\\ \mathbf{V}_2\end{array}\right] + \mathbf{W} \ , \tag{6.16}$$

where

$$\mathbf{V} = \mathbf{X}\mathbf{H}, \qquad \mathbf{V}_1 = \sqrt{\frac{\rho_\tau}{M}}\cdot\mathbf{S}_\tau\mathbf{H}, \qquad \mathbf{V}_2 = \sqrt{\frac{\rho_d}{M}}\cdot\mathbf{X}_d\mathbf{H} \ , \tag{6.17}$$

it is evident that the only non-Gaussian part in $\mathbf{Y}$ is matrix $\mathbf{V}_2$, which is a product of two Gaussian random matrices with every element $v_{i,j}$ a weighted sum of M independent complex Gaussian products, i.e.

$$v_{i,j} = \sqrt{\frac{\rho_d}{M}}\cdot\sum_{k=1}^{M} x_k\cdot h_k, \qquad x_k,\, h_k \sim \mathcal{N}_c\big(0,1\big) \ . \tag{6.18}$$

According to the central limit theorem, each element $v_{i,j}$ of matrix $\mathbf{V}_2$ converges to the Gaussian distribution when the number of transmit antennas M increases. Analytically speaking, both the characteristic function $\Psi_v(u)$ of a single element $v_{i,j}$ and the joint characteristic function $\Psi_{\mathbf{V}}(\mathbf{U})$ of the entire matrix $\mathbf{V}$ are proved to converge to Gaussian (and multivariate Gaussian) characteristic functions with the same variance (and covariance matrix). Further due to its particular product structure, it is proved in the following proposition that the convergence rate in this case is much faster than that of an arbitrary sum of independent variables (or matrices).

Proposition 7 *The joint characteristic function of the Gaussian product matrix* $\mathbf{V} = \mathbf{XH}$*, where both* $\mathbf{H}$ *and* $\mathbf{X}$ *are multivariate Gaussian distributed and have structures described in Section 6.2, converges to the characteristic function of Gaussian random matrix with the same mean and variance. The convergence rate* $O\big(1/M\big)$ *is faster than that of an arbitrary sum of independent variables (or matrices) with rate* $O\big(\sqrt{1/M}\big)$*.*

Proof: First of all, one can extend the definition of joint characteristic function to the case of complex random vectors,

$$\begin{aligned}\Phi_{\mathbf{z}}(\mathbf{u}) &= E_{\mathbf{z}}\Big[\exp\big(j\mathbf{u}_r^T\cdot\mathbf{z}_r + j\mathbf{u}_i^T\cdot\mathbf{z}_i\big)\Big] \\ &= E_{\mathbf{z}}\Big[\exp\Big(j\cdot\Re\big[\mathbf{u}^H\mathbf{z}\big]\Big)\Big] = E_{\mathbf{z}}\Big[\exp\Big(j\cdot\frac{\mathbf{u}^H\mathbf{z}+\mathbf{z}^H\mathbf{u}}{2}\Big)\Big]\ ,\end{aligned} \tag{6.19}$$

where $\mathbf{z}$ is a complex random vector and $\mathbf{u}$ is the corresponding multi-dimensional variable of the characteristic function. Both $\mathbf{u}$ and $\mathbf{z}$ are complex vectors of the same size, and can be decomposed (real and imaginary part) in the following manner

$$\mathbf{z} = \mathbf{z}_r + j\cdot\mathbf{z}_i\ , \qquad \mathbf{u} = \mathbf{u}_r + j\cdot\mathbf{u}_i\ . \tag{6.20}$$

Similarly, the concept can also be extended to the case of complex random matrices, i.e.

$$\begin{aligned}\Psi_{\mathbf{Z}}\big(\mathbf{U}\big) &= E_{\mathbf{Z}}\Big[\exp\big(j\cdot\mathbf{vec}(\mathbf{U}_r)^T\cdot\mathbf{vec}(\mathbf{Z}_r)+j\cdot\mathbf{vec}(\mathbf{U}_i)^T\cdot\mathbf{vec}(\mathbf{Z}_i)\big)\Big]\\ &= E_{\mathbf{Z}}\Big[\exp\Big(j\cdot\Re\big[\mathbf{vec}(\mathbf{U})^H\cdot\mathbf{vec}(\mathbf{Z})\big]\Big)\Big]\\ &= E_{\mathbf{Z}}\Big[\exp\Big(j\cdot\Re\Big[\mathbf{tr}\big(\mathbf{U}^H\mathbf{Z}\big)\Big]\Big)\Big]\ ,\end{aligned}\tag{6.21}$$

where $\mathbf{Z}$ is a complex random vector and $\mathbf{U}$ is the corresponding matrix variable of the characteristic function. It is also well known that the joint characteristic function of a Gaussian random vector $\mathbf{z}$ is given by

$$\Phi_{\mathbf{z}}(\mathbf{u})=E_{\mathbf{z}}\Big[\exp\Big(j\cdot\Re\big[\mathbf{u}^H\mathbf{z}\big]\Big)\Big]=\exp\Big(-\frac{\mathbf{u}^H\boldsymbol{\Sigma}\ \mathbf{u}}{4}+j\cdot\Re\big[\mathbf{u}^H\boldsymbol{\mu}\big]\Big)\ ,\tag{6.22}$$

where vector $\mathbf{z}$ has Gaussian distribution $\mathbf{z}\sim\mathcal{N}_c\big(\boldsymbol{\mu},\boldsymbol{\Sigma}\big)$.

By representing matrices $\mathbf{H}$ and $\mathbf{X}$ in the following manner,

$$\mathbf{H}=\Big[\mathbf{h}_1\ \mathbf{h}_2\ \cdots\ \mathbf{h}_{_M}\Big]^H,\ \mathbf{X}=\frac{1}{\sqrt{M}}\cdot\Big[\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_{_M}\Big]\ ,\ \mathbf{x}_i=\begin{bmatrix}\sqrt{\rho_\tau}\cdot\mathbf{s}_i\\ \sqrt{\rho_d}\cdot\mathbf{x}_{\mathbf{d}i}\end{bmatrix},\tag{6.23}$$

matrix product $\mathbf{V}=\mathbf{XH}$ can be rewritten as the following summation form, given by

$$\mathbf{V}=\mathbf{XH}=\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\mathbf{V}_i=\frac{1}{\sqrt{M}}\sum_{i=1}^{M}\mathbf{x}_i\cdot\mathbf{h}_i^H,\qquad\mathbf{V}_i=\mathbf{x}_i\cdot\mathbf{h}_i^H\ .\tag{6.24}$$

According to definition (6.21), the joint characteristic function of $\mathbf{V}$ (or $\mathbf{vec}(\mathbf{V})$) is given by

$$\begin{aligned}\Psi_{\mathbf{V}}(\mathbf{U}) &= E_{\mathbf{V}}\Big[\exp\Big(j\cdot\Re\Big[\mathbf{tr}\big(\mathbf{U}^H\mathbf{V}\big)\Big]\Big)\Big]\\ &= E_{\mathbf{V}}\Big[\exp\Big(\frac{j}{\sqrt{M}}\cdot\sum_{i=1}^{M}\Re\Big[\mathbf{tr}\big(\mathbf{U}^H\mathbf{V}_i\big)\Big]\Big)\Big]\\ &\overset{a}{=} \Pi_{i=1}^{M}\ E_{\mathbf{V}_i}\Big[\exp\Big(\frac{j}{\sqrt{M}}\cdot\Re\Big[\mathbf{tr}\big(\mathbf{U}^H\mathbf{V}_i\big)\Big]\Big)\Big]\\ &= \Pi_{i=1}^{M}\ \Psi_{\mathbf{V}_i}\Big(\frac{\mathbf{U}}{\sqrt{M}}\Big)\ ,\end{aligned}\tag{6.25}$$

where (a) is due to the fact that random matrices $\{\mathbf{V}_i\}$ are independently distributed.

According to the product structure given by (6.24), matrix $\mathbf{V}_i$ is conditionally Gaussian distributed (conditioned on vector $\mathbf{h}_i$), and can be represented as

$$\mathbf{vec}\big(\mathbf{V}_i\big|\mathbf{h}_i\big) \sim \mathcal{N}_c\big(\boldsymbol{\mu}_c,\, \boldsymbol{\Sigma}_c\big) \ , \tag{6.26}$$

where mean $\boldsymbol{\mu}_c$ and covariance matrix $\mathbf{V}_c$ are given by

$$\begin{aligned} \boldsymbol{\mu}_c &= \mathbf{vec}\left(\left[\frac{\sqrt{\rho_\tau} \cdot \mathbf{s}_i \cdot \mathbf{h}_i^H}{\mathbf{0}} \right] \right), \\ \boldsymbol{\Sigma}_c &= \mathbf{Cov}\Big(\mathbf{vec}\big(\mathbf{V}_i\big)\big|\mathbf{h}_i\Big) = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \rho_d \cdot I_{_{T_d}} \end{array} \right] \otimes \Big(\mathbf{h}_i \cdot \mathbf{h}_i^H\Big) \ . \end{aligned} \tag{6.27}$$

Therefore, by substituting the conditional mean and covariance matrix (6.27) into equation (6.22), the conditional characterisitic function $\Psi_{\mathbf{V}_i|\mathbf{h}_i}\big(\mathbf{U}\big)$ can be obtained as,

$$\Psi_{\mathbf{V}_i|\mathbf{h}_i}\big(\mathbf{U}\big) = \exp\Big(- \mathbf{h}_i^H \mathbf{U}^H \boldsymbol{\Sigma}_1 \, \mathbf{U} \, \mathbf{h}_i \, + j \cdot \Re\big[\sqrt{\rho_\tau} \cdot \mathbf{h}_i^H \mathbf{U}^H \mathbf{s}_i \,\big]\Big) , \tag{6.28}$$

where matrix $\boldsymbol{\Sigma}_1$ is given by

$$\boldsymbol{\Sigma}_1 = \frac{1}{4} \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \rho_d \cdot I_{_{T_d}} \end{array} \right] \ . \tag{6.29}$$

After some manipulations, the joint characteristic function of the matrix product $\mathbf{V}_i$ is obtained

$$\begin{aligned} \Psi_{\mathbf{V}_i}\big(\mathbf{U}\,\big) &= E_{\mathbf{h}_i}\Big[\Psi_{\mathbf{V}_i|\mathbf{h}_i}\big(\mathbf{U}\big)\Big] \\ &= \frac{\exp\left(- \mathbf{tr}\left[\Big(\mathbf{U}^H \boldsymbol{\Sigma}_1 \mathbf{U} + I\Big)^{-1} \cdot \Big(\mathbf{U}^H \boldsymbol{\Sigma}_{2,i}\, \mathbf{U}\Big)\right] \right)}{\Big|\mathbf{U}^H \boldsymbol{\Sigma}_1 \mathbf{U} + I\Big|}, \end{aligned} \tag{6.30}$$

where $\boldsymbol{\Sigma}_{2,i}$ is given by

$$\boldsymbol{\Sigma}_{2,i} = \frac{1}{4} \left[\begin{array}{c|c} \rho_\tau \cdot \mathbf{s}_i \mathbf{s}_i^H & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \ . \tag{6.31}$$

By substituting equation (6.31) into (6.25), we can finally obtain the joint characteristic function of $\mathbf{V}$, represented as

$$\Psi_{\mathbf{V}}(\mathbf{U}) = \frac{\exp\left(-\mathbf{tr}\left[\left(\mathbf{U}^H\boldsymbol{\Sigma}_1\mathbf{U}/M + I\right)^{-1}\cdot\left(\mathbf{U}^H\boldsymbol{\Sigma}_2\,\mathbf{U}/M\right)\right]\right)}{\left|\mathbf{U}^H\boldsymbol{\Sigma}_1\mathbf{U}/M + I\right|^M}\ , \tag{6.32}$$

where $\boldsymbol{\Sigma}_2$ is given by

$$\boldsymbol{\Sigma}_2 = \sum_{i=1}^{M}\boldsymbol{\Sigma}_{2,i} = \frac{1}{4}\left[\begin{array}{c|c} \rho_\tau\cdot\mathbf{S}_\tau\mathbf{S}_\tau^H & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0}\end{array}\right]\ . \tag{6.33}$$

When the number of transmit antennas M is large, we can further have the following convergence property,

$$\begin{aligned}
&\Psi_{\mathbf{V}}(\mathbf{U}) \\
&= \exp\left(-\mathbf{tr}\left[\left(\frac{\mathbf{U}^H\boldsymbol{\Sigma}_1\mathbf{U}}{M}+I\right)^{-1}\left(\frac{\mathbf{U}^H\boldsymbol{\Sigma}_2\,\mathbf{U}}{M}\right)\right] - M\log\left|\frac{\mathbf{U}^H\boldsymbol{\Sigma}_1\mathbf{U}}{M}+I\right|\right) \\
&= \exp\left(-\mathbf{tr}\left(\frac{\mathbf{U}^H\boldsymbol{\Sigma}_2\,\mathbf{U}}{M}\right)\left(1+O(\frac{1}{M})\right) - M\left(\frac{\mathbf{tr}\left(\mathbf{U}^H\boldsymbol{\Sigma}_1\mathbf{U}\right)}{M}+O\left(\frac{1}{M^2}\right)\right)\right) \\
&= \exp\left(-\frac{\mathbf{tr}\Big(\mathbf{U}^H\boldsymbol{\Sigma}_3\,\mathbf{U}\Big)}{4}+O\Big(\frac{1}{M}\Big)\right),
\end{aligned} \tag{6.34}$$

where $\boldsymbol{\Sigma}_3$ is given by

$$\boldsymbol{\Sigma}_3 = \left[\begin{array}{c|c} \frac{\rho_\tau}{M}\cdot\mathbf{S}_\tau\mathbf{S}_\tau^H & \mathbf{0} \\ \hline \mathbf{0} & \rho_d\cdot I_{T_d}\end{array}\right]\ . \tag{6.35}$$

It is also clear that matrix $\mathbf{V}$ has mean and variance given by

$$\boldsymbol{\mu}_V = \mathbf{0}\,,\qquad \boldsymbol{\Sigma}_V = \mathbf{Cov}\Big(\mathbf{vec}\big(\mathbf{V}\big)\Big) = \boldsymbol{\Sigma}_3\otimes I\ . \tag{6.36}$$

For multivariate Gaussian matrix $\mathbf{V}'$ having the same mean and covariance matrix, i.e.

$$\mathbf{vec}\big(\mathbf{V}'\big)\sim\mathcal{N}_c\Big(\mathbf{0},\boldsymbol{\Sigma}_V\Big)\ , \tag{6.37}$$

its joint characteristic function $\Psi_{\mathbf{V}'}$ can be represented as

$$\Psi_{\mathbf{V}'}\big(\mathbf{U}\big) = \exp\left(-\frac{\mathbf{vec}(\mathbf{U})\cdot\boldsymbol{\Sigma}_V\cdot\mathbf{vec}(\mathbf{U})^H}{4}\right) = \exp\left(-\frac{\mathbf{tr}\Big(\mathbf{U}^H\boldsymbol{\Sigma}_3\,\mathbf{U}\Big)}{4}\right). \tag{6.38}$$

Therefore, from the obtained characteristic functions (6.34) and (6.38), we can observe that $\Psi_{\mathbf{V}}(\mathbf{U})$ converges to Gaussian characteristic function $\Psi_{\mathbf{V}'}(\mathbf{U})$ with convergence rate $O(1/M)$. And this convergence rate is much faster than that of an arbitrary sum of independent variables (vectors or matrices) with rate only $O(1/\sqrt{M})$.

For a special case when $T_\tau = 0$, $T_d = 1$, and $N = 1$, matrix product $\mathbf{V}$ reduces to be a scalar random variable v, i.e.

$$v = \frac{\rho_d}{\sqrt{M}} \sum_{i=1}^{M} x_i \cdot h_i \ , \qquad x_i, h_i \sim \mathcal{N}_c(0,1) \ , \tag{6.39}$$

with characteristic function Ψ_v given by

$$\begin{aligned} \Psi_v(u) &= \Big(1 + \frac{\rho_d \cdot |u|^2}{4M}\Big)^{-M} \\ &= \exp\Big(- \frac{\rho_d \cdot |u|^2}{4} + \sum_{i=2}^{\infty} \frac{(-\rho_d \cdot |u|^2/4)^i}{(i-1) \cdot M^{i-1}} \Big) \\ &= \exp\Big(- \frac{\rho_d \cdot |u|^2}{4} + O\big(\frac{1}{M}\big) \Big) \ . \end{aligned} \tag{6.40}$$

According to the central limit theory, it is clear that v converges to Gaussian distribution as M increases. Further due to its particular product structure, its characteristic function $\Psi_v(u)$ converges to Gaussian characteristic function with rate $O(1/M)$ as shown in (6.40).

■

As an example, we demonstrate in Fig. 6.4 the histogram of the real part of random variable $v_{i,j}$ with different number of transmit antennas M. From the plot, it is quite clear that $v_{i,j}$ has a fast convergence rate to the Gaussian distribution with its histogram almost falling on top of the Gaussian PDF when $M \geq 3$. We also show in Fig. 6.5 the differential entropy of random matrix $\mathbf{V}$ versus the number of transmit antennas M of a $M \times 2$ unknown MIMO system with equal power allocation. The entropy is evaluated at an average signal to noise ratio of $\rho = 4dB$, with channel coherent time $T = 4$, and with training and data slot allocation $(T_\tau, T_d) = (2, 2)$. Simulation results shown in Fig. 6.5 further

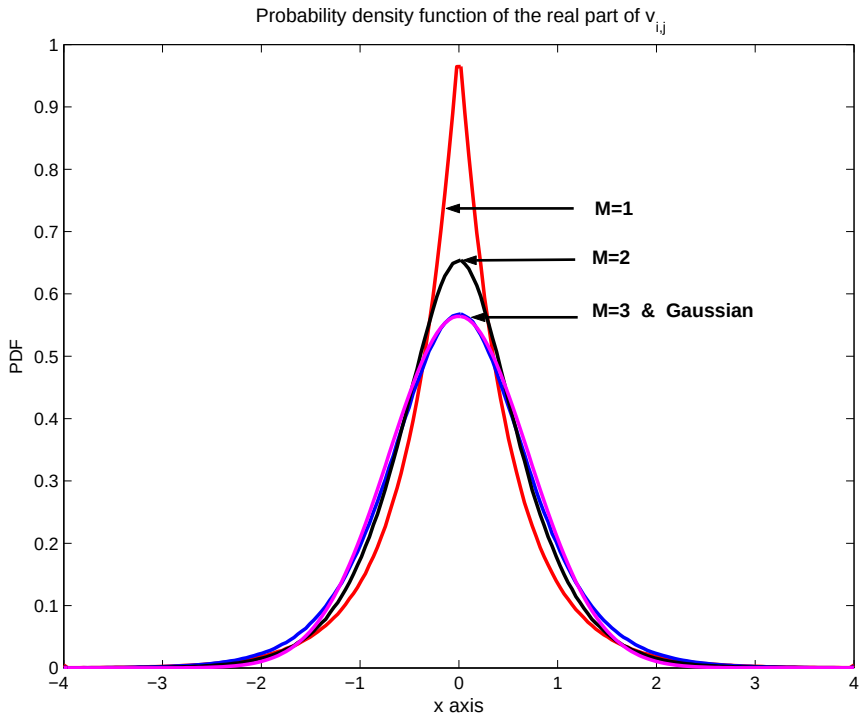

Figure 6.4: Histogram of $\Re[v_{i,j}]$ with different number of transmit antennas $M = 1, 2, 3$.

confirm the fast convergence rate of the Gaussian mixture matrix $\mathbf{V}$ to a Gaussian distribution from an information theoretical perspective, which leads to a tight mutual information upper bound.

From the channel model represented in (6.16), we know that the distribution of the received signal $\mathbf{Y}$ is dominated by the Gaussian noise $\mathbf{W}$ at low SNR range ($\rho_\tau, \rho_d \ll 1$). Hence, the upper bound is expected to be tight and can represent the mutual information rate R accurately. When the MIMO system is in moderate to high SNR regimes, according to the above convergence discussion and simulation results, the entropy of the Gaussian mixture density is tightly upper bounded by the Gaussian entropy (inequality (b) in (6.14)) with a moderate number of transmit antennas M. Hence, the mismatch between the upper bound $\overline{R}$ and the actual mutual information rate R can rapidly decrease to an insignificant

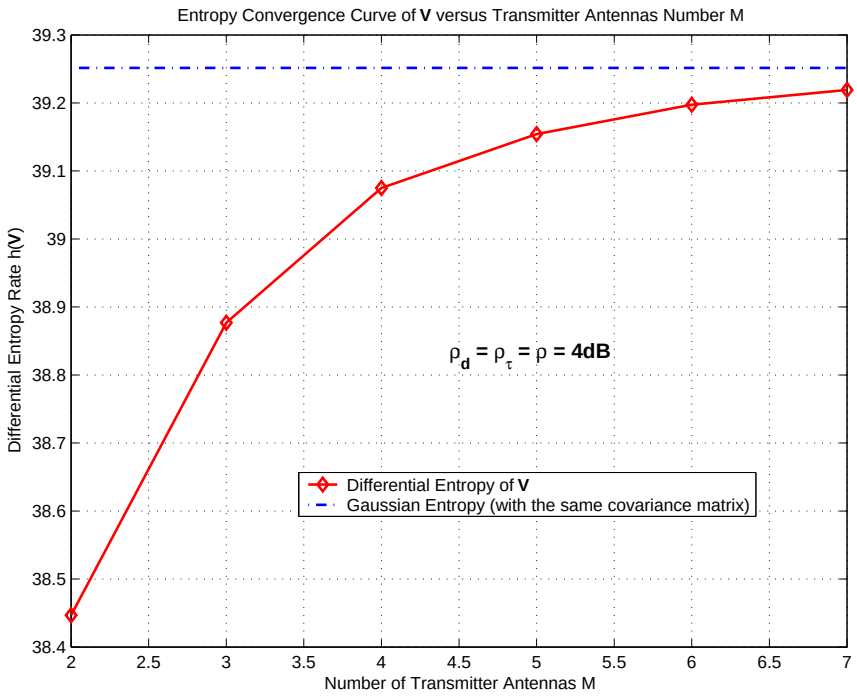

Figure 6.5: Differential entropy $h(\mathbf{V})$ of Gaussian product matrix $\mathbf{V}$ versus the number of transmit antennas M.

amount as the number of transmit antennas M increases. Further due to the sub-optimality of the input signal structure, the unknown MIMO channel capacity C is always lower bounded by the mutual information rate R ($C \geq R$). Therefore, it is reasonable to assume that when the number of transmit antennas M is larger than a certain threshold M_{th}, channel capacity C of the non-coherent MIMO system is lower bounded by the mutual information rate upper bound $\overline{R}$, i.e.

$$C \geq \overline{R}\,, \qquad M \geq M_{th}\ , \tag{6.41}$$

where M_{th} is a threshold that depends on system parameters T_τ, T_d, ρ_τ, ρ_d, and N.

According to inequality (6.41), the proposed mutual information upper bound $\overline{R}$ is a valid and improved capacity lower bound of the unknown MIMO system under certain conditions (with moderate to high transmit antennas M

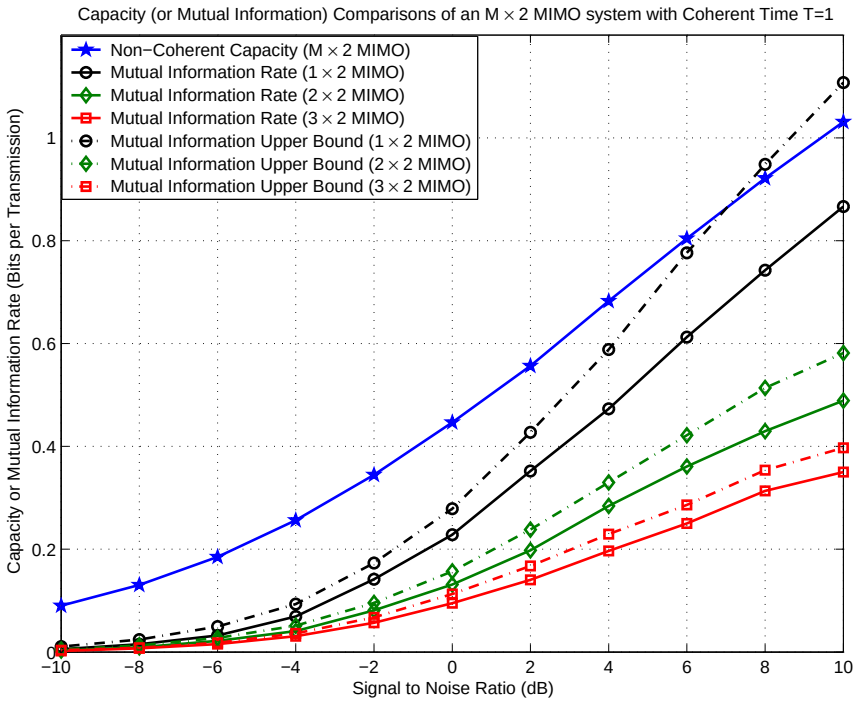

Figure 6.6: Capacity and mutual information comparison between non-coherent MIMO channel capacity C, the actual system mutual information rate R, and the proposed mutual information upper bound $\overline{R}$ of an $M \times 2$ MIMO system with channel coherent time $T = 1$ and signal to noise ratio $\rho_\tau = \rho_d = 4dB$.

or within low SNR regimes). It is hence reasonable to maximize (or optimize) the mutual information upper bounds (or capacity lower bounds) with respect to different system parameters, which is provided in Section 6.4. As an example, we demonstrate in Fig. 6.6 the capacity comparison curves of the non-coherent MIMO channel capacity C, the proposed mutual information rate upper bound $\overline{R}$, and the actual system mutual information rate R. The simulation is performed on a $M \times 2$ MIMO system over an unknown fast fading channel with coherent time $T = 1$. The reason for choosing a small MIMO system over a rapidly changing channel is that the optimal input distribution can be reduced to a complex scalar having uniformly distributed phase and a magnitude with a discrete probability density [6], which makes the optimization of the input distribution numerically

tractable (in contrast to the intractable higher dimensional density optimizations when $\min(T, M) \geq 2$).

From the plot of this particular example, we can observe that the mutual information upper bound $\overline{R}$ becomes a valid and hence improved capacity lower bound (as compare to the MMSE-based lower bound) of the unknown MIMO channel when the number of transmit antennas M is greater than 1. The only problem arises when M is extremely small ($M = 1$ in this case) in high SNR regimes and can be attributed to insufficient entropy convergence.

6.4 Analysis of the Mutual Information Upper Bound

Having obtained the mutual information rate upper bound (or the improved capacity lower bound under certain conditions) in Section 6.3, we provide in this section the detailed analysis (or optimization) of the mutual information bounds with respect to different system parameters such as channel coherent intervals, training and data slot allocations, power allocations, number of active transmit and receive antennas, as well as pilot structures.

6.4.1 Optimization of Pilot Structures

The most commonly used pilots have an orthogonal structure. They are optimal in a sense that they minimize the mean square error of the linear MMSE channel estimator [72]. More specifically, the minimal mean square error (MMSE) channel estimation as well as its error covariance matrix for the unknown MIMO channel are given by

$$\mathbf{vec}(\widehat{\mathbf{H}}) = \mathbf{Cov}^{-1}\Big(\mathbf{vec}(\mathbf{H}), \mathbf{vec}(\mathbf{Y})\Big) \cdot \mathbf{Cov}^{-1}\Big(\mathbf{vec}(\mathbf{Y}), \mathbf{vec}(\mathbf{Y})\Big) \cdot \mathbf{vec}(\mathbf{Y}) \ , \tag{6.42}$$

and

$$\begin{aligned}\mathbf{Cov}\Big(\mathbf{vec}(\widetilde{\mathbf{H}}), \mathbf{vec}(\widetilde{\mathbf{H}})\Big) = \mathbf{Cov}\Big(\mathbf{vec}(\mathbf{H}), \mathbf{vec}(\mathbf{H})\Big) - \mathbf{Cov}\Big(\mathbf{vec}(\mathbf{H}), \mathbf{vec}(\mathbf{Y})\Big) \\ \cdot \mathbf{Cov}^{-1}\Big(\mathbf{vec}(\mathbf{Y}), \mathbf{vec}(\mathbf{Y})\Big) \cdot \mathbf{Cov}\Big(\mathbf{vec}(\mathbf{Y}), \mathbf{vec}(\mathbf{H})\Big) \ ,\end{aligned} \tag{6.43}$$

where $\widetilde{\mathbf{H}} = \mathbf{H} - \widehat{\mathbf{H}}$ is the channel estimation error. After some manipulation, we can obtain the following result,

$$\begin{aligned} \mathbf{vec}\big(\widehat{\mathbf{H}}\big) &= I_N \otimes \left(\sqrt{\frac{\rho_\tau}{M}} \mathbf{S}_\tau^H \Big(\frac{\rho_\tau}{M} \mathbf{S}_\tau \mathbf{S}_\tau^H + I_{T_\tau} \Big)^{-1} \right) \cdot \mathbf{vec}\big(\mathbf{Y}\big) \ , \\ \widehat{\mathbf{H}} &= \sqrt{\frac{\rho_\tau}{M}} \mathbf{S}_\tau^H \Big(\frac{\rho_\tau}{M} \mathbf{S}_\tau \mathbf{S}_\tau^H + I_{T_\tau} \Big)^{-1} \cdot \mathbf{Y} \ , \end{aligned} \tag{6.44}$$

and

$$\mathbf{C}_{\widetilde{\mathbf{H}},\widetilde{\mathbf{H}}} = \mathbf{Cov}\Big(\mathbf{vec}\big(\widetilde{\mathbf{H}}\big), \mathbf{vec}\big(\widetilde{\mathbf{H}}\big) \Big) = I_N \otimes \big(I_M + \frac{\rho_\tau}{M} \mathbf{S}_\tau^H \mathbf{S}_\tau \big)^{-1} \ . \tag{6.45}$$

From (6.45), it is obvious that the mean square error of the channel estimation

$$\mathbf{tr}\big(\mathbf{C}_{\widetilde{\mathbf{H}},\widetilde{\mathbf{H}}}\big) = N \times \mathbf{tr}\Big(\big(I_M + \frac{\rho_\tau}{M} \mathbf{S}_\tau^H \mathbf{S}_\tau \big)^{-1} \Big) \ , \tag{6.46}$$

is minimized when the non-zero eigenvalues of $\mathbf{S}_\tau^H \mathbf{S}_\tau$ are all equal. Therefore, the following orthogonal pilot structure, represented as

$$\begin{aligned} \mathbf{S}_\tau^H \mathbf{S}_\tau &= T_\tau \cdot I_M, \qquad T_\tau \geq M \ , \\ \mathbf{S}_\tau \mathbf{S}_\tau^H &= M \cdot I_{T_\tau}, \qquad T_\tau < M \ . \end{aligned} \tag{6.47}$$

minimizes the MIMO MMSE channel mean square estimation error.

Although orthogonal pilot structure given by (6.47) minimizes the MMSE estimation error, it does not necessarily imply maximization of the system mutual information rate. In order to obtain the optimal pilot structure with respect to the mutual information upper bound $\overline{R}$, we utilize the following concavity property.

Proposition 8 *The mutual information upper bound obtained in Proposition 6 is concave with respect to matrix* $\mathbf{Q} = \mathbf{S}_\tau^H \mathbf{S}_\tau$*, i.e.,*

$$\lambda \cdot \overline{R}\big(\mathbf{Q_1}\big) + (1-\lambda) \cdot \overline{R}\big(\mathbf{Q_2}\big) \leq \overline{R}\Big(\lambda \cdot \mathbf{Q_1} + (1-\lambda) \cdot \mathbf{Q_2} \Big), \qquad 0 \leq \lambda \leq 1 \ . \tag{6.48}$$

Proof: First, the partial derivative of equation (6.10) with respect to matrix $\mathbf{Q}$ is given by,

$$\frac{\partial \overline{R}}{\partial \mathbf{Q}} = \frac{N \rho_\tau}{\ln 2 \cdot TM} \left(\Big(I_M + \frac{\rho_\tau}{M} \mathbf{Q} \Big)^{-1} - E_{\mathbf{X}_d} \left[\Big(I_M + \frac{\rho_\tau}{M} \mathbf{Q} + \frac{\rho_d}{M} \mathbf{X}_d^H \mathbf{X}_d \Big)^{-1} \right] \right) \ , \tag{6.49}$$

and the corresponding gradient can be represented as

$$\frac{\partial \overline{R}}{\partial \mathbf{vec}\big(\mathbf{Q}\big)} = \mathbf{vec}^H\Big(\frac{\partial \overline{R}}{\partial \mathbf{Q}}\Big) \ . \tag{6.50}$$

Based on (6.50), the Hessian of the upper bound $\overline{R}$ can be obtained,

$$\frac{\partial^2 \overline{R}}{\partial^2 \mathbf{vec}\big(\mathbf{Q}\big)} \ = \ -\frac{N\rho_\tau^2}{\ln 2 \cdot TM^2} \cdot E_{\mathbf{X}_d}\Big[\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-2} \otimes \boldsymbol{\Sigma}_2^{-1}\Big] \ , \tag{6.51}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are given by

$$\boldsymbol{\Sigma}_1 = I_M + \frac{\rho_\tau}{M}\mathbf{Q} \ , \qquad \boldsymbol{\Sigma}_2 = I_M + \frac{\rho_\tau}{M}\mathbf{Q} + \frac{\rho_d}{M}\mathbf{X}_d^H\mathbf{X}_d \ . \tag{6.52}$$

It can be shown that the Hessian matrix is negative semi-definite, due to the fact that

$$\Big(I_M + \frac{\rho_\tau}{M}\mathbf{Q}\Big)^{-1} > \Big(I_M + \frac{\rho_\tau}{M}\mathbf{Q} + \frac{\rho_d}{M}\mathbf{X}_d^H\mathbf{X}_d\Big)^{-1} \ , \tag{6.53}$$

and hence the mutual information rate upper bound is concave with respect to $\mathbf{Q}$.

■

As a direct result of Proposition 8, we have the following optimal pilot structure.

Proposition 9 *The optimal pilot structure, which maximizes the mutual information rate upper bound (6.10), satisfies the following orthogonal conditions*

$$\mathbf{Q} = \mathbf{S}_\tau^H\mathbf{S}_\tau = \frac{MT_\tau}{\min(T_\tau, M)}\left[\begin{array}{c|c} I_{\min(T_\tau,M)} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \ , \tag{6.54}$$

which is equivalent to (6.47).

Proof: First, substituting (6.52) into (6.10), the mutual information upper bound can be represented as

$$\overline{R} = \frac{N}{T}\bigg(T_d \cdot \log_2(1+\rho_d) + \log_2\big|\boldsymbol{\Sigma}_1\big| - E_{\mathbf{X}_d}\Big[\log_2\big|\boldsymbol{\Sigma}_2\big|\Big]\bigg) \ . \tag{6.55}$$

The following equality is true

$$\begin{aligned} &E_{\mathbf{X}_d}\Big[\log_2\big|\mathbf{\Sigma}_2(\mathbf{Q})\big|\Big] = E_{\mathbf{X}_d}\Big[\log_2\big|\mathbf{U}^H\mathbf{\Sigma}_2\mathbf{U}\big|\Big] \\ &= E_{\mathbf{X}_d}\left[\log_2\left|I_M + \frac{\rho_\tau}{M}\mathbf{U}^H\mathbf{Q}\,\mathbf{U} + \frac{\rho_d}{M}(\mathbf{X}_d\mathbf{U})^H(\mathbf{X}_d\mathbf{U})\right|\right] \\ &\overset{a}{=} E_{\mathbf{X}_d}\left[\log_2\left|I_M + \frac{\rho_\tau}{M}\mathbf{U}^H\mathbf{Q}\,\mathbf{U} + \frac{\rho_d}{M}\mathbf{X}_d^H\mathbf{X}_d\right|\right] = E_{\mathbf{X}_d}\Big[\log_2\big|\mathbf{\Sigma}_2\big(\mathbf{U}^H\mathbf{Q}\mathbf{U}\big)\big|\Big]\,, \end{aligned} \tag{6.56}$$

where $\mathbf{U}$ is any unitary matrix, and (a) follows from the fact that $\mathbf{X}_d\mathbf{U}$ has the same distribution as $\mathbf{X}_d$. Further due to the fact that

$$\big|\mathbf{\Sigma}_1(\mathbf{Q})\big| = \big|\mathbf{U}^{\mathbf{H}}\mathbf{\Sigma}_1\mathbf{U}\big| = \big|I_M + \frac{\rho_\tau}{M}\mathbf{U}^H\mathbf{Q}\mathbf{U}\big| = \big|\mathbf{\Sigma}_1(\mathbf{U}^H\mathbf{Q}\mathbf{U})\big|, \tag{6.57}$$

the mutual information upper bound (6.55) is hence invariant under the following transformation

$$\overline{R}(\mathbf{Q}) = \overline{R}(\mathbf{U}^H\mathbf{Q}\mathbf{U})\ . \tag{6.58}$$

If the unitary matrix $\mathbf{U}$ is set to be composed of the eigenvectors of $\mathbf{Q}$, then according to (6.58) we only need to focus our attention on the case where $\mathbf{Q}$ is a diagonal matrix.

Furthermore, it is also true that any permutations on the non-zero diagonal elements of $\mathbf{Q}$ will not change the upper bound,

$$\overline{R}\big(\mathbf{Q}\big) \overset{a}{=} \overline{R}\big(\mathbf{P}^H\mathbf{Q}\,\mathbf{P}\big) = \frac{1}{K!}\sum_{\mathbf{P}}\overline{R}\big(\mathbf{P}^H\mathbf{Q}\,\mathbf{P}\big) \overset{b}{\leq} \overline{R}\Big(\frac{1}{K!}\sum_{\mathbf{P}}\mathbf{P}^H\mathbf{Q}\,\mathbf{P}\Big)\ , \tag{6.59}$$

where $K = \min(T_\tau, M)$ and $\mathbf{P}$ is any permutation matrix that permutes the first K rows (or columns), equality (a) follows the same reasoning as the invariant transformation (6.58), and (b) follows from the concavity property of the upper bound. At this point, it is evident that the optimal pilot, which achieves the maximum mutual information upper bound, has an orthogonal structure given by (6.54). ■

Therefore, although starting from different perspectives, orthogonal pilots structure not only minimize the estimation mean square error, but also maximize

the mutual information rate upper bounds. Substituting the optimal structure (6.54) into the equation (6.10), we obtain the following mutual information upper bound

$$\overline{R} = \frac{N}{T}\left(T_d \cdot \log_2(1+\rho_d) - E_{\mathbf{X}_d}\left[\log_2 \left| I_M + \frac{\rho_d}{M} \mathbf{\Lambda}^{-1} \mathbf{X}_d^H \mathbf{X}_d \right| \right] \right), \tag{6.60}$$

where $\mathbf{X}_d$ is of size $\mathbb{C}^{T_d \times M}$, and $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda} = \left[\begin{array}{c|c} \left(1 + \frac{\rho_\tau T_\tau}{\min(T_\tau, M)}\right) \cdot I_{\min(T_\tau, M)} & \mathbf{0} \\ \hline \mathbf{0} & I_{M - \min(T_\tau, M)} \end{array} \right] \ . \tag{6.61}$$

By utilizing the known results of the joint probability density function (PDF) of the eigenvalues of a Wishart matrix [73], the mutual information upper bound can be further expressed in a concise form given by the following proposition, which facilitates the evaluation of the bounds.

Proposition 10 *In general, except for the case* $\big(0 < T_d, T_\tau < M\big)$*, mutual information upper bound* $\overline{R}$ *can be given by*

$$\overline{R} = \left\{ \begin{array}{ll} \frac{N}{T}\Big(T_d \cdot \log_2(1+\rho_d) - F\Big(\min\big(T_d, M\big), \max(T_d, M), \, f_1(\cdot) \Big) \Big) & \text{if } T_\tau \geq M \\ \frac{N}{T}\Big(T_d \cdot \log_2(1+\rho_d) - F_1\Big(T_d, T_\tau, M, \rho \Big) \Big) & \text{if } T_d \geq M > T_\tau > 0 \ . \\ \frac{N}{T}\Big(T_d \cdot \log_2(1+\rho_d) - F\Big(\min\big(T_d, M\big), \max(T_d, M), \, f_2(\cdot) \Big) \Big) & \text{if } T_\tau = 0 \end{array} \right. \tag{6.62}$$

where mapping F *is given by*

$$F\big(m, n, f(\cdot)\big) = \int_0^\infty f(\lambda) \sum_{k=0}^{m-1} \frac{k!}{(n-m+k)!} \Big[L_k^{n-m}(\lambda) \Big]^2 \lambda^{n-m} e^{-\lambda} d\lambda \ , \tag{6.63}$$

where $L_k^{n-m}(\lambda)$ *is the associated Laguerre polynomial [74] of order* k*, given by*

$$L_k^{n-m}(\lambda) = \frac{1}{k!} e^\lambda \lambda^{m-n} \cdot \frac{d^k}{d\lambda^k}\big(e^{-\lambda} \lambda^{n-m+k} \big) \ , \tag{6.64}$$

and functions $f_1(\cdot)$ *and* $f_2(\cdot)$ *are given by*

$$\begin{aligned} f_1(\lambda) &= \log_2\big(1 + \frac{\rho_d / M}{1 + \rho_\tau T_\tau / \min(T_\tau, M)} \lambda\big) \ , \\ f_2(\lambda) &= \log_2\big(1 + \frac{\rho_d}{M} \lambda\big) \ . \end{aligned} \tag{6.65}$$

And mapping F_1 is given by

$$F_1\Big(T_d, T_\tau, M, \rho\Big) = \int_0^\infty \cdots \int_0^\infty \sum_{i=1}^{M} f_2(\lambda_i) \cdot \frac{1}{K}\big|\mathbf{\Lambda}\big|^{T_d} \prod_{i=1}^{M} \lambda_i^{T_d - M} \cdot \prod_{i<j}^{M} (\lambda_i - \lambda_j)^2 \times {}_0\tilde{F}_0\Big(-\mathbf{\Lambda}, \mathbf{diag}\big(\lambda_1, \cdots, \lambda_M\big)\Big) \cdot d\lambda_1 \cdot d\lambda_2 \cdots d\lambda_M, \tag{6.66}$$

where ${}_0\tilde{F}_0\Big(\mathbf{A}, \mathbf{B}\Big)$ is the hypergeometric function of Hermitian matrix arguments, which is defined in terms of a series involving zonal polynomials [75]. Normalization factor K is given by

$$K = M! \cdot \prod_{i=1}^{M} (T_d - i)! \cdot (M - i)! \ . \tag{6.67}$$

Proof: Since $\mathbf{X}_d$ has a multivariate normal distribution, which is denoted as $\mathcal{N}_{M,T_d}\big(\mathbf{0}, I_M\big)$, it is obvious that $\mathbf{X}_d^H \mathbf{X}_d$ or $\mathbf{X}_d \mathbf{X}_d^H$ have Wishart distributions [73] given by

$$\begin{aligned} \mathbf{X}_d^H \mathbf{X}_d &\sim W_M\big(T_d, I_M\big) \qquad \text{if } M \le T_d \\ \mathbf{X}_d \mathbf{X}_d^H &\sim W_{T_d}\big(M, I_{T_d}\big) \qquad \text{if } M > T_d \ . \end{aligned} \tag{6.68}$$

It is well known that the joint probability density function of the eigenvalues of the Wishart matrix $\mathbf{S} \sim W_n\big(m, \mathbf{\Sigma}\big)$ is given by

$$p_{\lambda_s}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \frac{1}{K} \prod_{i=1}^{n} \lambda_i^{m-n} e^{-\lambda_i} \prod_{i<j} (\lambda_i - \lambda_j)^2 \ , \tag{6.69}$$

where K is given by

$$K = n! \cdot \prod_{i=1}^{m} (n - i)! \cdot (m - i)! \ . \tag{6.70}$$

And the marginal distributions of the unordered eigenvalues of $\mathbf{X}_d^H \mathbf{X}_d$, given by Telatar in [5], is

$$p_{\lambda_i}(\lambda_i) = \frac{1}{\min(T_d, M)} \sum_{k=0}^{\min(T_d, M) - 1} \frac{k!}{\Big(k + |T_d - M|\Big)!} \cdot \Big[L_k^{|T_d - M|}(\lambda_i)\Big]^2 \lambda_i^{|T_d - M|} e^{-\lambda_i}. \tag{6.71}$$

Hence, substituting the marginal p.d.f. function (6.71) into the mutual information upper bound (6.60), we can obtain the concise form expression for the case where $T_\tau \geq M$ and $T_\tau = 0$. When $T_d \geq M > T_\tau > 0$, the mutual information upper bound can be rewritten as,

$$\overline{R} = \frac{N}{T}\left(T_d \cdot \log_2(1+\rho_d) - E_{\mathbf{X}_d}\left[\log_2\left|I_M + \frac{\rho_d}{M}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{X}_d^H\mathbf{X}_d\mathbf{\Lambda}^{-\frac{1}{2}}\right|\right]\right) , \qquad (6.72)$$

where matrix $\mathbf{X}'_d = \mathbf{X}_d\mathbf{\Lambda}^{-\frac{1}{2}}$ follows a multivariate normal distribution given by $\mathcal{N}_{M,T_d}\big(\mathbf{0}, \mathbf{\Lambda}^{-1}\big)$, and ${\mathbf{X}'_d}^H\mathbf{X}'_d$ has a Wishart distribution $W_M\big(T_d, \mathbf{\Lambda}^{-1}\big)$, whose eigenvalues have a joint p.d.f given by

$$\begin{aligned} &p_\lambda(\lambda_1, \lambda_2, \cdots, \lambda_M) \\ &\quad = \frac{1}{K}\big|\mathbf{\Lambda}\big|^{T_d}\,{}_0\tilde{F}_0\Big(-\mathbf{\Lambda}, \mathbf{diag}\big(\lambda_1, \cdots, \lambda_M\big)\Big) \cdot \prod_{i=1}^{M}\lambda_i^{T_d-M} \cdot \prod_{i<j}^{M}(\lambda_i - \lambda_j)^2 . \end{aligned} \qquad (6.73)$$

Plugging (6.73) into the mutual information upper bound (6.60), we can obtain the expression for the case $T_d \geq M > T_\tau > 0$. When $0 < T_d, T_\tau < M$, matrix ${\mathbf{X}'_d}^H\mathbf{X}'_d$ has a so-called pseudo Wishart distribution, whose eigenvalues do not have a simple close form p.d.f., and hence the proposed upper bound can not be expressed in a concise form. ■

6.4.2 Optimization of Pilot and Data Slot Allocations (under Equal Power Assumptions)

For some communication systems, it might not be possible to vary the power during the training slots and data slots. Hence the mutual information upper bound, assuming training symbol and data symbol share the same power, is obtained by substituting the power allocations ($\rho_\tau = \rho_d = \rho$) into (6.60). The upper bound is further optimized with respect to the data allocation scheme (T_τ, T_d), and we have the following important result concerning the dependence on T_d, the number of data symbols.

Proposition 11 *Mutual information rate upper bounds under equal power allocation schemes are monotonically increasing with respect to the number of data slots T_d, i.e.*

$$\begin{aligned} \overline{R}(T_d = k) &\geq \overline{R}(T_d = k-1), \qquad k \leq T - M, \\ \overline{R}(T_d = T) &\geq \overline{R}(T_d = T - M) \ . \end{aligned} \tag{6.74}$$

Proof: We begin with the first part of (6.74) when $T_\tau \geq M$, where equation (6.60) is reduced to

$$\overline{R} = \frac{N}{T}\left(T_d \cdot \log_2(1+\rho) - E_{\mathbf{X}_d}\left[\log_2 \left| I_M + \frac{\rho/M}{1+\rho T_\tau/M} \cdot \mathbf{X}_d^H \mathbf{X}_d \right| \right] \right), \tag{6.75}$$

where we further separate $\mathbf{X}_d$ into $\mathbf{X}'_d$ and $\mathbf{x}_{T_d}$,

$$\mathbf{X}_d = \begin{bmatrix} \mathbf{X}'_d \\ \mathbf{x}_{T_d} \end{bmatrix}, \qquad \mathbf{X}'_d \in \mathbb{C}^{T_d - 1 \times M}, \qquad \mathbf{x}_{T_d} \in \mathbb{C}^{1 \times M} \ . \tag{6.76}$$

Then we have the following inequality,

$$\begin{aligned} & E_{\mathbf{X}_d}\left[\log_2 \left| I_M + \frac{\rho'}{M} \mathbf{X}_d^H \mathbf{X}_d \right| \right] = E_{\mathbf{X}_d}\left[\log_2 \left| I_M + \frac{\rho'}{M} \left({\mathbf{X}'_d}^H \mathbf{X}'_d + \mathbf{x}_{T_d}^H \mathbf{x}_{T_d} \right) \right| \right] \\ & \overset{a}{\leq} E_{\mathbf{X}'_d}\left[\log_2 \left| I_M + \frac{\rho'}{M} \left({\mathbf{X}'_d}^H \mathbf{X}'_d + E_{\mathbf{x}_{T_d}}\left[\mathbf{x}_{T_d}^H \mathbf{x}_{T_d} \right] \right) \right| \right] \ , \end{aligned} \tag{6.77}$$

where (a) follows from the fact that $\log|\cdot|$ is a concave function. Therefore, using assumption (6.8), we can obtain the following result for the mutual information

rate difference

$$\begin{aligned}
\Delta\overline{R} &= \overline{R}(T_d = k) - \overline{R}(T_d = k-1) \\
&= \frac{N}{T}\Bigg(\log_2(1+\rho) - E_{\mathbf{X}_d}\Big[\log_2 \Big| I_M + \frac{\rho \cdot \mathbf{X}_d{}^H \mathbf{X}_d}{M + \rho T_\tau} \Big| \Big] \\
&\quad + E_{\mathbf{X}'_d}\Big[\log_2 \Big| I_M + \frac{\rho \cdot \mathbf{X}'_d{}^H \mathbf{X}'_d}{M + \rho(T_\tau+1)} \Big| \Big] \Bigg) \\
&\overset{a}{\geq} \frac{N}{T}\Bigg(\log_2(1+\rho) - E_{\mathbf{X}'_d}\Big[\log_2 \Big| I_M + \frac{\rho \cdot \mathbf{X}'_d{}^H \mathbf{X}'_d + I_M}{M + \rho T_\tau} \Big| \Big] \\
&\quad + E_{\mathbf{X}'_d}\Big[\log_2 \Big| I_M + \frac{\rho \cdot \mathbf{X}'_d{}^H \mathbf{X}'_d}{M + \rho(T_\tau+1)} \Big| \Big] \Bigg) \\
&= \frac{N}{T}\Bigg(\log_2(1+\rho) - M \log_2 \Big(\frac{1 + \rho(T_\tau+1)/M}{1 + \rho T_\tau / M} \Big) \Bigg) \\
&= \frac{N}{T} \cdot \Delta f(\rho) \ , \qquad (6.78)
\end{aligned}$$

where (a) follows from inequality (6.77) and scaler function $\Delta f(\rho)$ represented as

$$\Delta f(\rho) = \log_2(1+\rho) - M \log_2 \Big(\frac{1 + \rho(T_\tau+1)/M}{1 + \rho T_\tau / M} \Big) \geq 0 \ , \qquad (6.79)$$

is a positive function due to the following fact

$$\begin{aligned}
\Delta f(0) &= 0, \\
\Delta' f(\rho) &= \frac{1}{1+\rho} - \frac{T_\tau + 1}{1 + \rho(T_\tau+1)/M} + \frac{T_\tau}{1 + \rho T_\tau / M} > 0 \ . \qquad (6.80)
\end{aligned}$$

For the second part of (6.74), let us first denote $\Delta\overline{R}$ as the following mutual information rate difference,

$$\Delta\overline{R} = \overline{R}\big(T_d = T\big) - \overline{R}\big(T_d = T - M\big) = \frac{1}{T}\Bigg(M \cdot \log_2(1+\rho) \\
- E_{\mathbf{X}_1}\Big[\log_2 \Big| I_M + \frac{\rho}{M} \cdot \mathbf{X}_1^H \mathbf{X}_1 \Big| \Big] + E_{\mathbf{X}_2}\Big[\log_2 \Big| I_M + \frac{\rho}{M(1+\rho)} \cdot \mathbf{X}_2^H \mathbf{X}_2 \Big| \Big] \Bigg), \qquad (6.81)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are of sizes

$$\mathbf{X}_1 \in \mathbb{C}^{T \times M}, \qquad \mathbf{X}_2 \in \mathbb{C}^{(T-M) \times M} \ . \qquad (6.82)$$

We can further separate $\mathbf{X}_1$ into the following form,

$$\mathbf{X}_1 = \left[\frac{\mathbf{X}_1'}{\mathbf{x}}\right], \qquad \mathbf{X}_1' \in \mathbb{C}^{T-M\times M}, \qquad \mathbf{x} \in \mathbb{C}^{M\times M} \ . \tag{6.83}$$

Then the following inequality is obtained,

$$\begin{aligned}
&E_{\mathbf{X}_1}\left[\log_2\left|I_M + \frac{\rho}{M}\cdot\mathbf{X}_1^H\mathbf{X}_1\right|\right] = E_{\mathbf{X}_1}\left[\log_2\left|I_T + \frac{\rho}{M}\cdot\mathbf{X}_1\mathbf{X}_1^H\right|\right] \\
&= E_{\mathbf{X}_1'}\left[E_{\mathbf{x}}\left[\log_2\left|I_T + \frac{\rho}{M}\cdot\mathbf{X}_1\mathbf{X}_1^H\right|\right]\right] \overset{a}{\leq} E_{\mathbf{X}_1'}\left[\log_2\left|I_T + \frac{\rho}{M}\cdot E_{\mathbf{x}}\left[\mathbf{X}_1\mathbf{X}_1^H\right]\right|\right] \\
&= E_{\mathbf{X}_1'}\left[\log_2\left|I_T + \frac{\rho}{M}\cdot\left[\begin{array}{c|c}\mathbf{X}_1'\mathbf{X}_1'^H & \mathbf{0}\\ \hline \mathbf{0} & M\cdot I_M\end{array}\right]\right|\right] \\
&= M\cdot\log_2(1+\rho) + E_{\mathbf{X}_1'}\left[\log_2\left|I_M + \frac{\rho}{M}\cdot\mathbf{X}_1'^H\mathbf{X}_1'\right|\right] \ ,
\end{aligned} \tag{6.84}$$

where (a) follows from the fact the $\log_2|\cdot|$ is a concave function. Therefore, substituting (6.84) into (6.81), we can establish the following inequality

$$\begin{aligned}
\Delta\overline{R} = \frac{1}{T}\Bigg(&M\cdot\log_2(1+\rho) + E_{\mathbf{X}_2}\left[\log_2\left|I_M + \frac{\rho}{M(1+\rho)}\cdot\mathbf{X}_2^H\mathbf{X}_2\right|\right] \\
&-E_{\mathbf{X}_1}\left[\log_2\left|I_M + \frac{\rho}{M}\cdot\mathbf{X}_1^H\mathbf{X}_1\right|\right]\Bigg) \geq \frac{1}{T}\Bigg(E_{\mathbf{X}_2}\left[\log_2\left|I_M + \frac{\rho}{M(1+\rho)}\cdot\mathbf{X}_2^H\mathbf{X}_2\right|\right] \\
&E_{\mathbf{X}_1'}\left[\log_2\left|I_M + \frac{\rho}{M}\cdot\mathbf{X}_1'^H\mathbf{X}_1'\right|\right]\Bigg) = 0 \ .
\end{aligned} \tag{6.85}$$

■

According to the above proposition, we have shown that the mutual information rate upper bound is monotonically increasing with respect to T_d up to $T_d \leq T - M$, and all of them are upper bounded by the rate where there is no training at all. As an example, we show in Fig. 6.7 the mutual information upper bounds versus the number of data slots T_d of a 6×6 MIMO system under equal power allocation schemes. The mutual information bounds are evaluated at an average signal to noise ratio of $\rho = 4dB$, and with several different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20,$ which parameterize the upper bound curves.

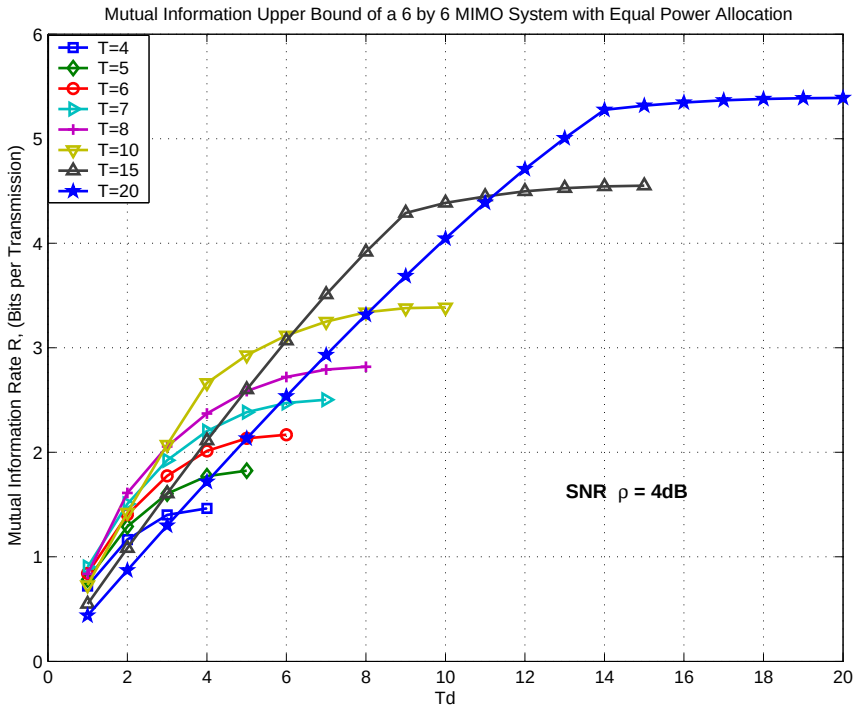

Figure 6.7: Mutual information rate upper bound of a 6×6 MIMO system under equal power allocation scheme of SNR $\rho = 4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$

As illustrated by the numerical results in the plot, the mutual information rate upper bound is indeed monotonically increasing with respect to T_d, even for the case $T_d > T - M$. However, the rate gain is insignificant after T_d grows beyond $T - M$ especially when T is larger than M. Therefore, $T_d = T - M, T > M$ is a good trade-off point between the achievable information rate and system complexity, where the mutual information upper bound can be reduced to

$$\overline{R}(\rho) = N\left(T_d \cdot \log_2(1+\rho) - E_{\mathbf{X}_d}\left[\log_2\left|I_M + \frac{\rho/M}{1+\rho}\mathbf{X}_d^H\mathbf{X}_d\right|\right]\right),\ T > M = T_\tau. \quad (6.86)$$

6.4.3 Optimization of the Number of Active Transmit Antennas

Since the channel state information is not known to the transmitter nor to the receiver, using a large number of transmit antennas will definitely intro-

duce more channel uncertainties and hence prevent us from correctly decoding the transmitted information. From another point of view, for a fixed channel coherent time interval T, introducing more transmit antennas means we have to allocate more time slots to training symbols and sacrifice the data rates and hence the channel capacity. Therefore, we provide in the following proposition the choice of the appropriate number of active transmit antennas.

Proposition 12 *Under equal power allocation schemes, and with pilots number T_τ equal to the number of transmit antennas M, mutual information rate upper bound is monotonically decreasing with respect to the number of transmit antennas M, i.e.*

$$\overline{R}(M+1) \leq \overline{R}(M), \qquad M \geq N, \quad T > N \ . \tag{6.87}$$

Proof: First notice that the second term in equation (6.86) is actually the ergodic capacity for a MIMO system composed of M transmit antennas and T_d receive antennas, where channel state information is perfectly known at the receiver. The equivalent signal to noise ratio is equal to $\rho' = \rho/(1+\rho)$. Due to the fact that MIMO channel capacity with perfect CSIR is a monotonically increasing function with respect to the number of transmit antennas M, we can have the following inequality,

$$E_{\mathbf{X}_1}\left[\log_2\left|I_{M+1}+\frac{\rho/(M+1)}{1+\rho}\cdot\mathbf{X}_1^H\mathbf{X}_1\right|\right] \geq E_{\mathbf{X}_2}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot{\mathbf{X}_2}^H\mathbf{X}_2\right|\right], \tag{6.88}$$

where matrices $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are of sizes given by,

$$\mathbf{X}_1 \in \mathbb{C}^{T_d\times(M+1)},\ \mathbf{X}_2 \in \mathbb{C}^{T_d\times M},\ \mathbf{X}_3 \in \mathbb{C}^{(T_d+1)\times M},\ \mathbf{x} \in \mathbb{C}^{1\times M},\ \mathbf{X}_3 = \left[\frac{\mathbf{x}}{\mathbf{X}_2}\right], \tag{6.89}$$

with each of their elements follow i.i.d. zero mean complex Gaussian distribution with unit variance. Furthermore, since $\log|\cdot|$ is a concave function, the following

inequality can be obtained

$$
\begin{aligned}
E_{\mathbf{X}_3}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot{\mathbf{X}_3}^H\mathbf{X}_3\right|\right] &= E_{\mathbf{X}_3}\left[\log_2\left|I_{T_d+1}+\frac{\rho/M}{1+\rho}\cdot\mathbf{X}_3{\mathbf{X}_3}^H\right|\right] \\
&\le E_{\mathbf{X}_2}\left[\log_2\left|I_{T_d+1}+\frac{\rho/M}{1+\rho}\cdot E_{\mathbf{x}}\left[\mathbf{X}_3{\mathbf{X}_3}^H\right]\right|\right] \\
&= E_{\mathbf{X}_2}\left[\log_2\left|I_{T_d+1}+\frac{\rho/M}{1+\rho}\cdot\left[\begin{array}{c|c} M & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{X}_2\mathbf{X}_2^H \end{array}\right]\right|\right] \\
&= \log_2\left(1+\frac{\rho}{1+\rho}\right)+E_{\mathbf{X}_2}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot{\mathbf{X}_2}^H\mathbf{X}_2\right|\right]. \qquad (6.90)
\end{aligned}
$$

Therefore, for a fixed coherent time interval T, the mutual information upper bound (6.86) has the following inequality,

$$
\begin{aligned}
&\overline{R}(M+1)=\frac{N}{T}\left(T_d\cdot\log_2(1+\rho)-E_{\mathbf{X}_1}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot\mathbf{X}_1^H\mathbf{X}_1\right|\right]\right) \\
&\le\frac{N}{T}\left(T_d\cdot\log_2(1+\rho)-E_{\mathbf{X}_2}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot\mathbf{X}_2^H\mathbf{X}_2\right|\right]\right) \\
&\le\frac{N}{T}\left(T_d\cdot\log_2(1+\rho)+\log_2\left(1+\frac{\rho}{1+\rho}\right)-E_{\mathbf{X}_3}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot\mathbf{X}_3^H\mathbf{X}_3\right|\right]\right) \\
&\le\frac{N}{T}\left((T_d+1)\cdot\log_2(1+\rho)-E_{\mathbf{X}_3}\left[\log_2\left|I_M+\frac{\rho/M}{1+\rho}\cdot\mathbf{X}_3^H\mathbf{X}_3\right|\right]\right)=\overline{R}(M)\,. \qquad (6.91)
\end{aligned}
$$

Although inequality (6.91) is true for any $M>0$, it is only reasonable to maximize the above mutual information upper bound $\overline{R}(M)$ up to $M\ge M_{th}$. This is because the upper bound $\overline{R}$ provides tight measurement of the system mutual information rate R and serves as a valid capacity lower bound of the unknown MIMO system only when M is beyond a certain threshold M_{th}. Generally speaking, it is very difficult to determine M_{th} by a close form expression and can be resolved only through numerical simulations. However, a straightforward condition that is necessary of leading to a tight mutual information upper bound (and hence a valid capacity lower bound) is that the number of independent random variables of the mixture Gaussian distribution $\mathbf{Y}$ should at least be larger than that of the

approximated Gaussian distribution, given by

$$M_{th} \times N + T_d \times M_{th} \geq T \times N \ , \tag{6.92}$$

which is equivalent to the following condition after some manipulations,

$$M \geq M_{th} \geq \frac{TN}{T_d + N} \qquad \Longleftrightarrow \qquad M \geq N \ . \tag{6.93}$$

Therefore, for a MIMO communication system having N receive antennas and transmitting over an unknown channel with coherent time interval T ($T > N$) with equal training and data power allocation, an appropriate choice of the system design parameters would be

$$T_\tau = M = N, \qquad T_d = T - N \ . \tag{6.94}$$

■

We demonstrate in Fig. 6.8 the mutual information rate upper bounds versus the number of active transmit antennas of a $M \times 6$ unknown MIMO system under equal power allocations. The MIMO system shown in the plot has data and training allocation scheme (T_d, T_τ) given by $(T - M, M)$. The mutual information upper bounds are evaluated at an average SNR of $\rho = 4dB$, and with several different coherent time intervals $T = 8, 10, 12, 14, 16, 18, 20$. As illustrated in Fig. 6.8, the upper bound (6.86) is monotonically decreasing with respect to the number of active transmit antennas $M = T_\tau$, and hence there is no benefit in using transmit antennas greater than N when $T > N$.

6.4.4 Optimization of Power Allocations (ρ_τ, ρ_d) between Training and Data Symbols

For communication systems where the power allocation can be varied between training and data symbols, the optimal mutual information upper bound is obtained by solving the following constrained optimization problem

$$\overline{R}_{opt}\big(T_\tau, T_d\big) = \max_{(\rho_\tau, \rho_d)} \overline{R}\big(\rho_\tau, \rho_d, T_\tau, T_d\big), \quad \rho_\tau T_\tau + \rho_d T_d = \rho T \ . \tag{6.95}$$

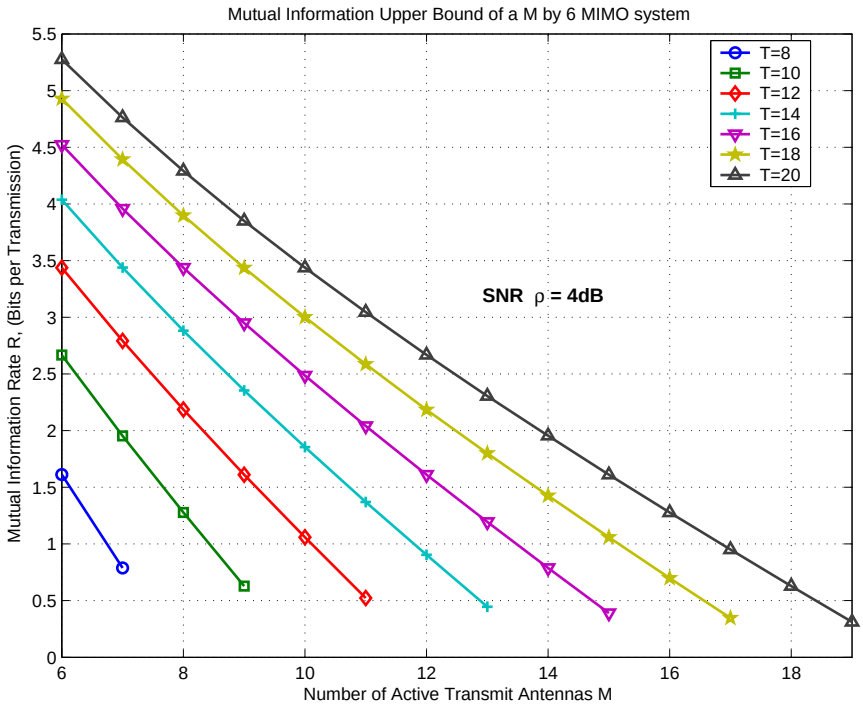

Figure 6.8: Mutual information upper bound of a $M \times 6$ MIMO system with SNR $\rho = 4dB$ and $T_\tau = M$, under different coherent time intervals $T = 8, 10, 12, 14, 16, 18, 20$

As an example, we show in Fig. 6.9 the mutual information upper bounds versus the the number of data slots T_d of a 6×6 unknown MIMO system under optimal power allocation strategies. The upper bounds are evaluated at an average SNR of $\rho = 4dB$, and with several different channel coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$ that parameterize the curves. From Fig. 6.9, we can observe that the mutual information rate upper bounds under optimal power allocations have the same monotonically increasing property and behave very similarly to the mutual information upper bounds with equal power allocations.

We also demonstrate in Fig. 6.10 the optimal power $\rho_d^\star$ allocated to data symbols versus the data slot number T_d for the same MIMO system with different coherent time intervals T. From the plot, we can observe that the optimal

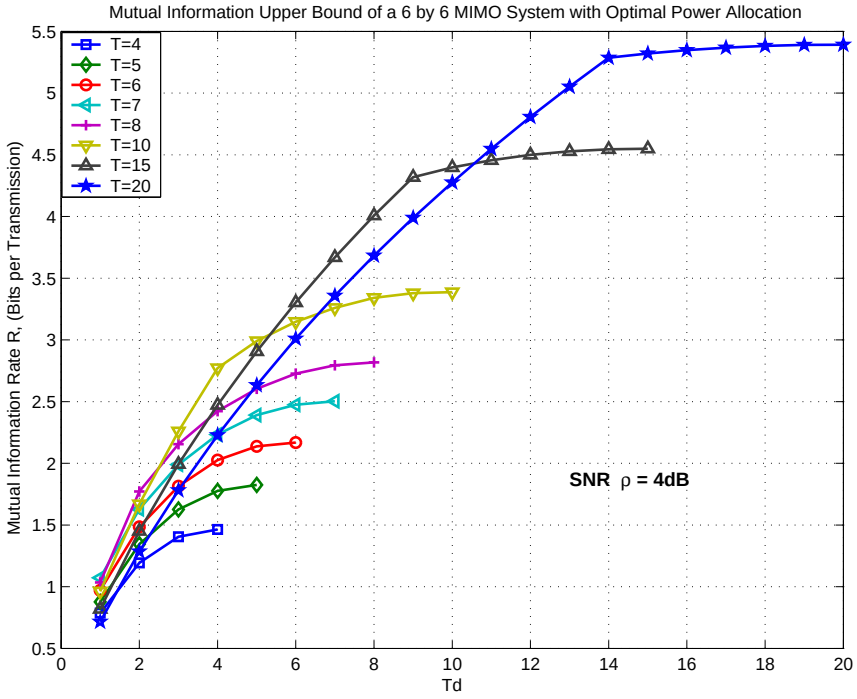

Figure 6.9: Mutual information upper bound of a 6×6 MIMO system under optimal power allocation scheme of SNR $\rho = 4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$

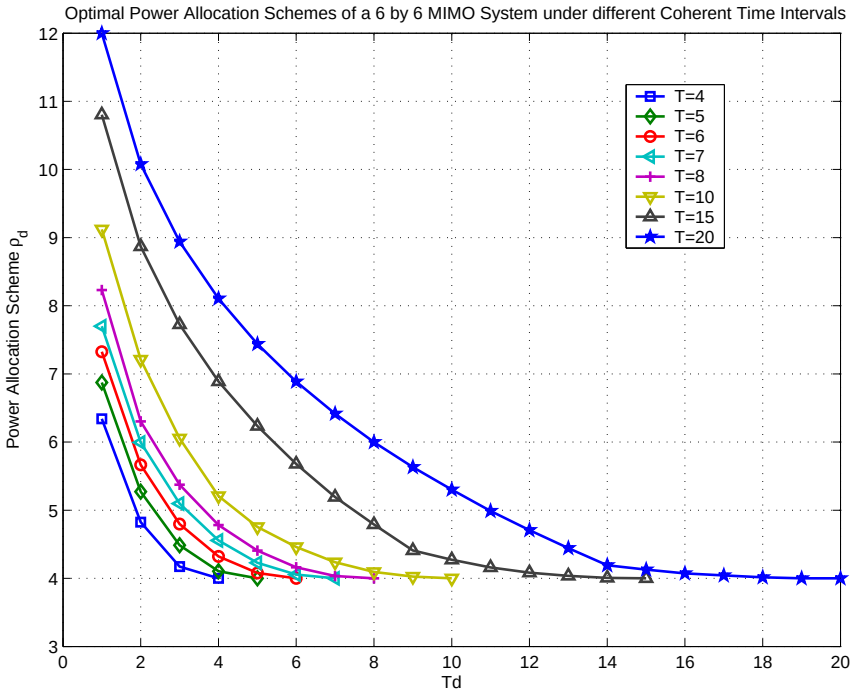

Figure 6.10: Optimal power allocation of a 6×6 MIMO system with SNR $\rho = 4dB$, under different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15,$ and 20

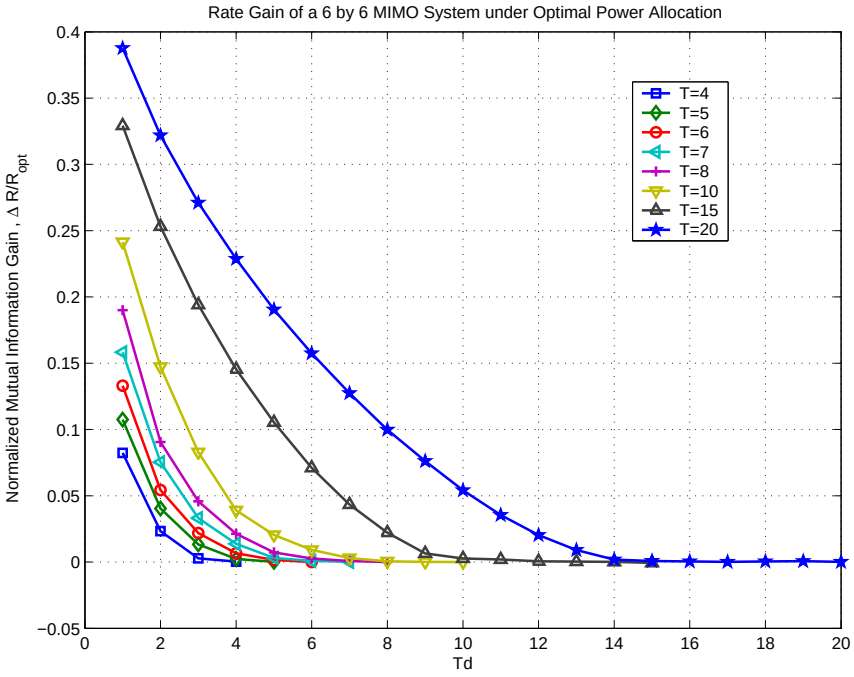

Figure 6.11: Mutual information rate gain of optimal power allocations over equal power allocations of a 6×6 MIMO system with SNR $\rho = 4dB$, under different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$

allocation scheme $\left(\rho_d^\star, \rho_\tau^\star\right)$ always has the following fact

$$\rho_d^\star \geq \rho \geq \rho_\tau^\star \ , \tag{6.96}$$

which is equivalent as saying that the average data power is always larger than the average training power. Furthermore, the optimal power allocation $\left(\rho_\tau^\star, \rho_d^\star\right)$ is quite close to equal power allocation scheme when T_τ is small (or T_d is large), especially in the regime where $T_\tau \leq M$ (or $T_d \geq T - M$).

As a comparison between optimal and equal power allocation schemes, we show in Fig. 6.11 the normalized mutual information rate gain $\left(\Delta R/R_{opt}\right)$ versus the number of data slots T_d for the same unknown MIMO system with different coherent time intervals T. It can be observed from the plot that there is an insignificant amount of rate loss by using equal power allocation schemes, which are much easier for implementation, as compared to applying optimal power allocations. The loss of the mutual information rate is negligible in the regime where $T_\tau \leq M$ (or $T_d \geq T - M$), which is the high capacity region of interest.

6.4.5 Low SNR Regimes

According to the mutual information upper bound given by (6.60), we can obtain a concise closed form approximation of the upper bound in the low SNR regime, given by

$$\begin{aligned} \overline{R} &\approx \frac{N}{\ln 2 \cdot TM} \cdot \left(\rho_d T_d \cdot \rho_\tau T_\tau + \mu\left(\rho_d T_d\right)^2\right) \\ &= \frac{N}{\ln 2 \cdot TM} \cdot \rho_d T_d \cdot \left(\rho T - (1-\mu) \cdot \rho_d T_d\right) \ , \end{aligned} \tag{6.97}$$

where μ is given by

$$\mu = \frac{1}{2M} \sum_{i=1}^{M} E\left[\lambda_i^2\right] \geq \frac{1}{2} \ , \tag{6.98}$$

with $\{\lambda_i\}_{i=1}^M$ being the eigenvalues of the Wishart matrix $\mathbf{X}_d^H \mathbf{X}_d$. Therefore, optimization (w.r.t to power and data slot allocation) of the mutual information upper bound (6.97) in the low SNR regime reduces to be a constraint quadratic

maximization problem, with solution given by

$$(\rho_d T_d)^\star = \rho T \iff \left(\rho_d^\star, \rho_\tau^\star\right) = (\rho,\, 0),\ \left(T_d^\star, T_\tau^\star\right) = (T,\, 0),\ \overline{R}^\star = \frac{N \cdot \mu \cdot \rho^2 T}{\ln 2 \cdot M}. \tag{6.99}$$

Furthermore, the mutual information upper bound $\overline{R}$ under equal power allocation schemes at low SNR regime can be obtained as

$$\overline{R} \approx \frac{N\rho^2}{\ln 2 \cdot M} \cdot T_d \cdot \left(1 - \frac{(1-\mu) \cdot T_d}{T}\right) \ . \tag{6.100}$$

It is evident from (6.100) that $\overline{R}$ is a monotonically increasing function w.r.t the number of data slots T_d. As a comparison to the low SNR approximation of the MMSE-based unknown MIMO capacity lower bound provided in [71], which is given by

$$\underline{C}^\star \approx \frac{N \cdot \rho^2 T}{4 \ln 2 \cdot M}\,, \qquad (\rho_d T_d)^\star = (\rho_\tau T_\tau)^\star = \frac{1}{2}\rho T \ , \tag{6.101}$$

the optimal mutual information upper bound $\overline{R}^\star$ (or the improved capacity lower bound with moderately large M) has a rate gain

$$\frac{\overline{R}^\star}{\underline{C}^\star} = 4\mu \geq 2 \ . \tag{6.102}$$

We can observe from (6.97), (6.100), and (6.101) that both $\overline{R}$ and $\underline{C}$ decay as ρ^2 at low power ranges. However, the true unknown channel capacity, which does not require training to achieve, decay as ρ rather than ρ^2 reported in [76] [57]. Therefore, it implies that not only the two-phase signal processing scheme (training and using channel estimate as if it were correct) is highly suboptimal when ρ is small (as stated in [71]), the suboptimal structure and its distribution of the input signal is the main reason that causes capacity (or rate) loss.

6.5 Numerical and Simulation Results

Although some very important numerical results of the mutual information upper bound have already been demonstrated in the previous sections, we are going to provide in this section a few more numerical examples to support the remaining obtained results.

6.5.1 Orthogonal Pilot Structure

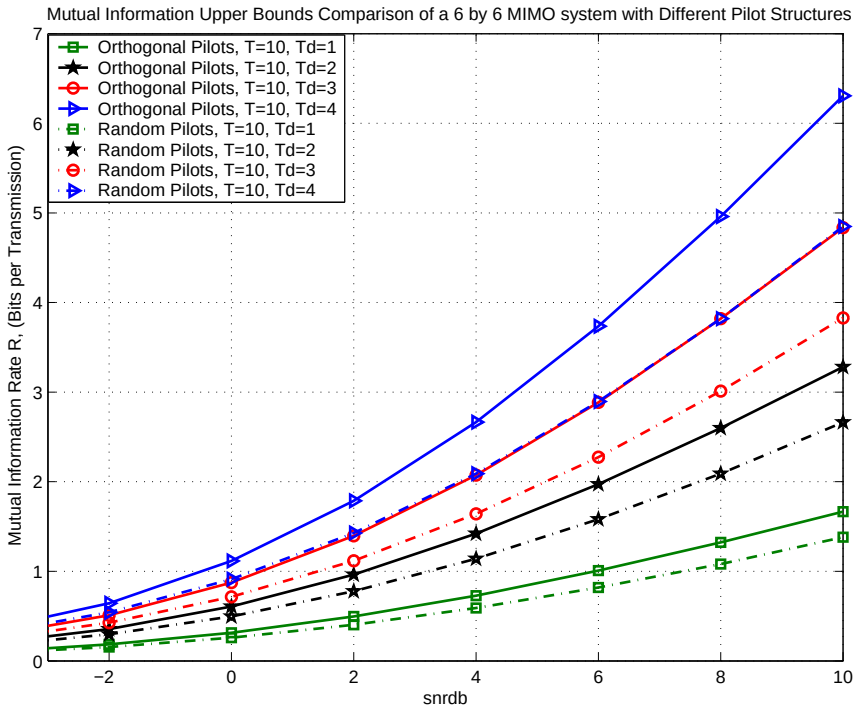

Figure 6.12: Mutual information upper bound comparison between orthogonal pilot structures and random pilot structures under equal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$, and data interval $T_d = 1, 2, 3, 4$.

We know from Section 6.4.1 that the orthogonal pilot structure not only minimizes the mean square estimation error, but also maximizes the mutual information rate upper bound (6.10). Fig. 6.12 is a demonstration of the sensitivity of the mutual information upper bound with respect to different pilots structures. We compare in the plot the mutual information upper bounds of a MIMO system using orthogonal pilot structure with those of using random pilot structure having the same training power. The upper bounds are evaluated assuming a 6×6 unknown MIMO system with coherence time interval $T = 10$, and for varying data slot number $T_d = 1, 2, 3, 4$. As can be observed from Fig. 6.12, the mutual infor-

mation upper bounds of using random pilots, which are denoted as doted curves, are inferior to those of applying orthogonal pilots. There is significant rate gain by using orthogonal pilot structures as compared with random pilots when T_d is large, which are the high information rate curves of interest.

6.5.2 Pilot and Data Slot Allocations (under Equal Power Assumptions)

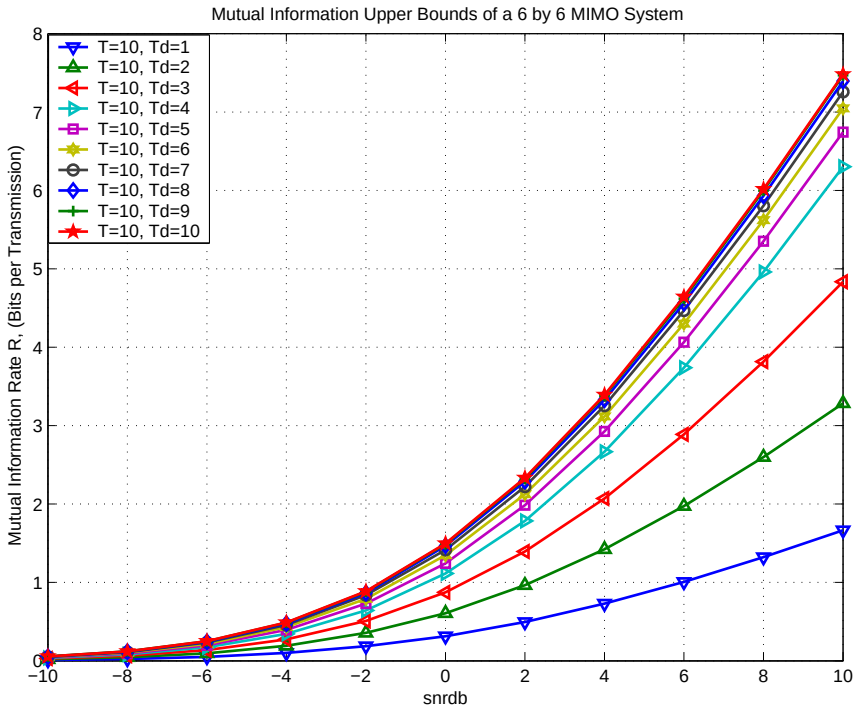

Figure 6.13: Mutual information rate upper bound of 6×6 MIMO system with coherent time interval $T = 10$, with different data slots allocation $T_d = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$

As is shown in Proposition 11, the mutual information upper bound is monotonically increasing with respect to the number of data slots T_d. We demonstrate in Fig. 6.13 the mutual information rate upper bounds versus the average SNR ρ. The mutual information bounds are evaluated assuming a 6×6 unknown MIMO system with channel coherence time $T = 10$, and for varying data slot

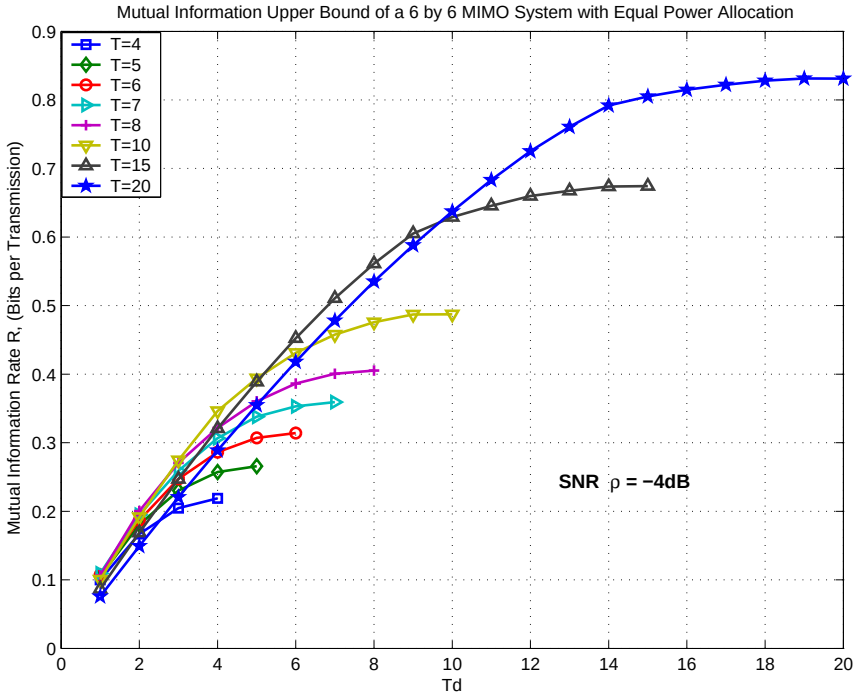

Figure 6.14: Mutual information rate upper bound of a 6×6 MIMO system under equal power allocation scheme of SNR $\rho = -4dB$, and with different coherent time intervals $T = 4, 5, 6, 7, 8, 10, 15, 20$

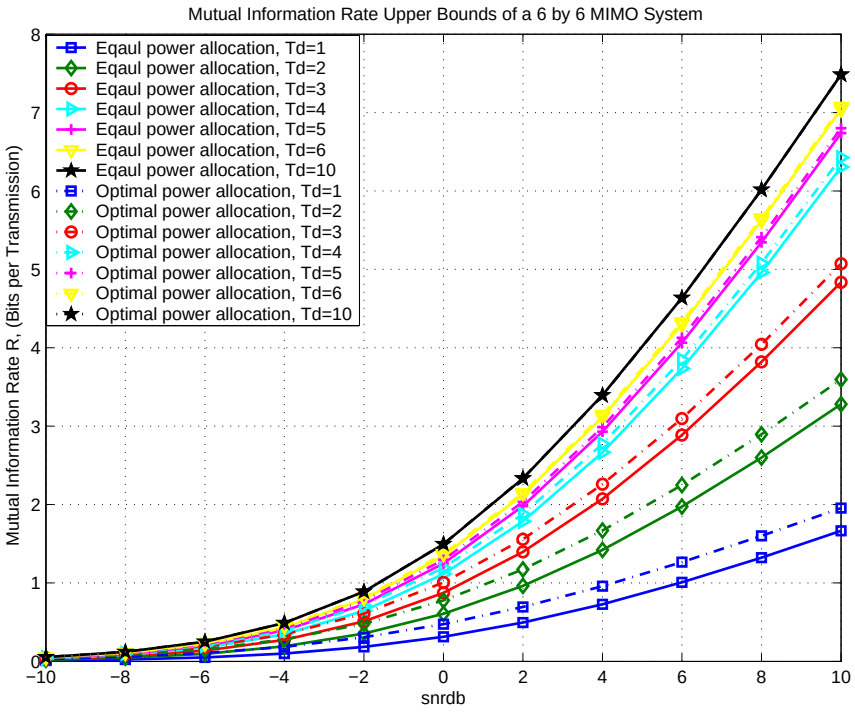

Figure 6.15: Mutual information upper bound comparison between equal power allocation and optimal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$ and different data slot allocation $T_d = 1, 2, 3, 4, 5, 6, 10$

allocations $T_d = 1, 2, \cdots, 10$. As is expected, we can easily observe from the plot that the information rate is monotonically increasing, and the rate increment is insignificant when $T_d \geq T - M$.

As a complement to Fig. 6.7, which demonstrates the monotonically increasing property of the mutual information upper bounds versus T_d for a moderate SNR of $\rho = 4dB$, we show in Fig. 6.14 the mutual information upper bounds under the same system settings in a low SNR environment of $\rho = -4dB$. It can be observed from the plot that very similar monotonically increasing property of the mutual information rate upper bounds exists even in a low SNR regime.

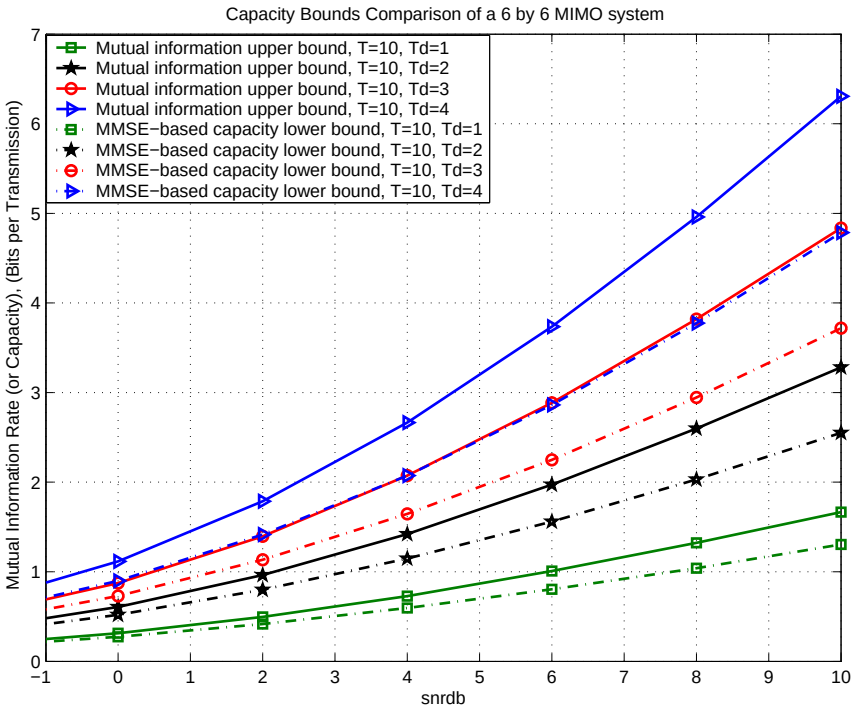

Figure 6.16: Comparison between the MMSE-based capacity lower bounds and the improved capacity lower bounds (mutual information upper bounds) under equal power allocation schemes of a 6×6 MIMO system with coherent time intervals $T = 10$, and data interval $T_d = 1, 2, 3, 4$

6.5.3 Power Allocations between Training and Data Symbols

We demonstrate in Fig. 6.15 the mutual information rate upper bound comparison between equal power and optimal power allocation schemes. The upper bounds are evaluated versus the average SNR ρ for a 6×6 unknown MIMO system with channel coherence time $T = 10$, and for varying data slot allocations $T_d = 1, 2, 3, 4, 5, 6, 10$. As can be observed from the plot, the information rate gain achieved by using optimal power allocation is insignificant especially when $T_d \geq T - M$, which is the high rate (capacity) region of interest.

6.5.4 Comparison with Other Capacity Analysis Results

In order to compare the capacity analysis results obtained in this chapter with previous publications, we show in Fig. 6.16 both the proposed mutual information upper bounds (or the improved capacity lower bounds) as well as the MMSE-based capacity lower bounds provided in [71] for the same unknown MIMO system. Again, the capacity bounds are evaluated assuming a 6×6 MIMO system with channel coherence time $T = 10$, for varying data slot allocations $T_d = 1, 2, 3, 4$. It can be observe from the plot that the two bounds have a significant capacity difference when the number of the data symbols T_d is large, which corresponds to the high capacity achieving data slot allocations of interest.

6.6 Summary

In this chapter, we studied the intrinsic role of training symbols in a MIMO communication system. First, we propose a system mutual information upper bound, which is tight and becomes a valid capacity lower bound of the unknown MIMO channel when we have a moderate number of transmit antennas M. Through the analysis (or optimization) of the proposed upper bound with respect to different system parameters, we show that orthogonal pilot structure is optimal in a sense that it not only minimizes the mean square estimation error, but also maximizes the proposed mutual information upper bound. We also prove that under equal power allocations, the mutual information upper bound is a monotonically increasing function with respect to the number of data slots T_d. Through numerical evaluations, we also demonstrate that the rate increment is insignificant when T_d is larger than $T - M$, suggesting a training duration of M time slots provides excellent trade-off between complexity and performance. By setting $T_\tau = M$, followed by further analysis on the proposed bounds, we show that there is no benefit in making the number of transmit antennas M greater than N. Furthermore, in an optimal power allocation scheme, the power allocated

to data symbols ρ_d is always larger than that of the training symbols ρ_τ. Also, the information rate gain by applying optimal power allocations is insignificant compared with equal power allocation schemes in high mutual information rate regimes where $T_\tau \leq M$ (or $T_d \geq T - M$). The text of chapter is in part a reprint of the paper which was coauthored with Bhaskar D. Rao and has been submitted for publication in *IEEE Transactions on Information Theory* under the title *"A study of limits on training via capacity analysis of MIMO systems with unknown channel state information"*.

7 Design of LDPC-Coded MIMO Systems With Unknown Block Fading Channels

7.1 Motivation

Communication systems using multiple antennas at both the transmitter and the receiver have recently received increased attention due to their ability to provide great capacity increases in a wireless fading environment [5] [2]. However, MIMO capacity analysis and system design is often based on the assumption that the fading channel coefficient between each transmit and receive antenna pair is perfectly known at the receiver. This is not a realistic assumption for most practical communication systems especially in fast fading channels.

For communication systems with unknown channel state information (CSI) at both ends, conventional receivers usually have a two-phase structure, data-aided channel estimation using the preset training symbols followed by coherent data detection by treating the estimated channel as the actual channel coefficients. Due to the importance of channel estimator, which directly determines estimation quality and hence the overall system performance, various MIMO channel estimation algorithms have been studied [77]– [79]. However, conventional channel estimators form estimates based only on the training symbols, thereby failing to make use of the channel information contained in the received data symbols.

Consequently, the two-phase model limits the performance and can not approach the MIMO channel capacity (or the maximum achievable information rate), especially in a fast fading environment (with small channel coherence time). Possible solutions to the above problem include use of blind source signal separation algorithms [80]– [82], MIMO differential modulation [61]– [63], and unitary space-time modulation (USTM) [64]– [70]. However, none of these schemes can approach the non-coherent MIMO capacity limit due to their sub-optimal code structure, and in the later case, USTM, only asymptotic (or the diversity) optimality is achieved in high SNR regimes and the approach suffers from exponential decoding complexity.

In order to achieve better spectral efficiency than the conventional data-aided estimation algorithm that uses large number of training symbols for accurate channel estimation, the so-called code-aided joint channel estimation and data detection algorithms have recently received much attention. By treating the unknown channel as unobserved (or missing) data, ML sequence estimation of the coded data frames using the EM algorithm was proposed by Georghiades [83] and Kaleh [84] over single input single output fading channels and extended to MIMO channels by Cozzo [85]. Alternatively, several recent publications [86]- [88] have developed EM-based algorithms that can iteratively improve the channel estimate based on the soft extrinsic information from the outer soft decoder, and the schemes work well in an iterative receiver structure.

In this chapter, we focus on the design of practical LDPC-coded MIMO systems employing a soft iterative receiver structure consisting of three soft decoding component blocks, a soft MIMO detector and two soft LDPC component decoders (variable node and check node decoders). At the component level, we first propose a soft optimal MIMO detector, which can generate soft log likelihood ratio (LLR) of each coded bit under the condition of unknown CSIR without forming any explicit channel estimate. Based on the proposed soft optimal detector, we develop two simplified sub-optimal MIMO detectors with polynomial and log polynomial decoding complexities. In addition, motivated by the EM-based

detection algorithm in [88], we also propose in the MIMO context a modified EM-based detector that completely removes the positive feedback between the input and output extrinsic information and provides much better performance compared to the direct EM-based detector that has strong correlations. By analyzing the mutual information transfer characteristic [89] of the proposed soft MIMO detectors, system performance of different MIMO detection algorithms are analyzed and compared under various channel conditions. At the structural level, inspired by the turbo iterative principle [90], the LDPC-coded MIMO receiver is constructed in an unconventional manner where the soft MIMO detector and LDPC variable node decoder form one super soft-decoding unit and the LDPC check node decoder forms the other component of the iterative decoding scheme. Utilizing the proposed receiver structure, tractable extrinsic information transfer functions of the component soft decoders are obtained, which lead to a simple and efficient LDPC code degree profile optimization algorithm. This algorithm is shown to have global optimality and guaranteed convergence from any initialization, and is an improvement over the sub-optimal manual curve fitting technique proposed in [91]. Numerical and simulation results of the LDPC-coded MIMO system using the optimized degree profile further confirm the advantages of the proposed design approach for the coded MIMO system.

The rest of the chapter is organized as follows. Section 7.2 describes the LDPC-coded MIMO system structure as well as the unknown block fading channel model. Section 7.3 proposes several different soft MIMO detectors that can be used as the building blocks for the turbo iterative MIMO receivers. In section 7.4, the receiver design of the coded MIMO systems is addressed in detail, which includes the overall receiver structure in Section 7.4.1, the extrinsic mutual information transfer characteristic analysis in Section 7.4.2, and the LDPC code degree profile optimization algorithm in Section 7.4.3. In Section 7.5, the simulation results of the LDPC-coded MIMO system under various channel conditions are presented. Finally, conclusions are drawn in Section 7.6.

7.2 System Model

7.2.1 MIMO transmitter structure

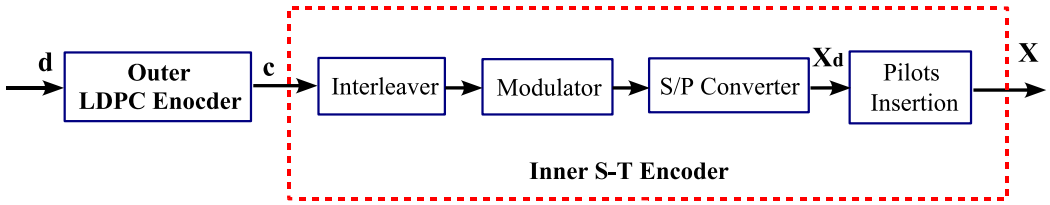

Figure 7.1: Transmitter model of LDPC-coded MIMO systems

We consider a MIMO system with M transmit antennas and N receive antennas signaling through a frequency flat fading channel with independent channel propagation coefficient between each transmit and receive antenna pair. As illustrated in Fig. 7.1, a block of k binary information bits denoted $\mathbf{d} = \{d_1, \cdots, d_k\}$ is first encoded by an outer LDPC encoder with code rate $R_{\text{outer}} = k/n$ into a codeword $\mathbf{c} = \{c_1, \cdots, c_n\}$ of length n. The codeword $\mathbf{c}$ is further segmented into L consecutive sub-blocks $\mathbf{C}_i$ of length K. Each sub-block $\mathbf{C}_i$ is then encoded by the inner space-time encoder into a coherent space-time *sub-frame* $\mathbf{X}_i$. This encoder is composed of an interleaver, modulator, serial-to-parallel converter, and a pilot insertion operator. The symbol structure of each sub-frame $\mathbf{X}_i$ is illustrated in Fig. 7.2, where the first p symbols are training pilots, followed by $(TM - p)$ data symbols. For the sake of simplicity, we only consider the case where both the number of pilot symbols $\big(p = T_\tau \times M\big)$ and the number of data symbols $\big(TM - p = T_d \times M\big)$ are multiples of the transmit antenna number M. We further denote the average signal to noise ratio (SNR) of pilot symbols by ρ_τ and data symbols by ρ_d. Hence, the transmitted signal $\mathbf{X}_i$ can be partitioned into two sub-matrices: training followed by data, which is represented as

$$\mathbf{X}_i = \begin{bmatrix} \big(\rho_\tau/M\big)^{\frac{1}{2}} \cdot \mathbf{X}_\tau \\ \big(\rho_d/M\big)^{\frac{1}{2}} \cdot \mathbf{X}_{d,\,i} \end{bmatrix}, \tag{7.1}$$

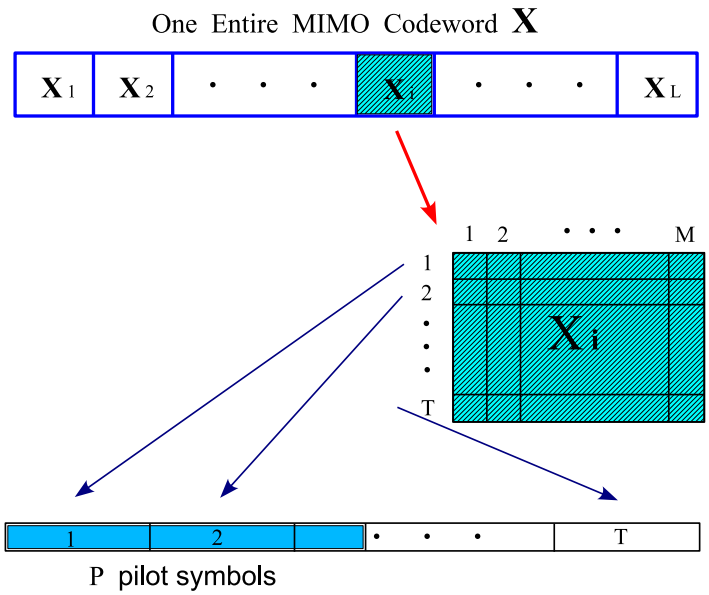

Figure 7.2: Transmitted symbol structure of the coded MIMO system

where $\mathbf{X}_\tau \in \mathbb{C}^{T_\tau \times M}$ are the fixed pilot symbols sent over T_τ time intervals, and $\mathbf{X}_{d,\,i} \in \mathbb{C}^{T_d \times M}$ are the information bearing data symbols sent over T_d transmission intervals. Each element of the transmitted data signal $\mathbf{X}_{d,\,i}$ is a member of a finite complex alphabet $\mathcal{X}$ of size $|\mathcal{X}|$. One entire MIMO codeword $\mathbf{X}$ consists of $l = LTM$ complex symbols, which are transmitted from M transmit antennas and across L consecutive coherent sub-frames of length TM symbols.

It is assumed that the fading coefficient matrix $\mathbf{H}_i$ remains static within each coherent sub-block and varies independently from one sub-block to another. Hence, the signal model can be written as

$$\mathbf{Y}_i = \mathbf{X}_i \cdot \mathbf{H}_i + \mathbf{w}_i, \qquad 1 \leq i \leq L \ , \tag{7.2}$$

where $\mathbf{Y}_i$ is a $T \times N$ received complex signal matrix, $\mathbf{X}_i$ is a $T \times M$ transmitted complex signal matrix, $\mathbf{H}_i$ is an $M \times N$ complex channel matrix, and $\mathbf{w}_i$ is a $T \times N$ matrix of additive noise matrix. Both matrices $\mathbf{H}_i$ and $\mathbf{w}_i$ are assumed to have zero mean unit variance independent complex Gaussian entries. We also assume that the entries of the transmitted signal matrix $\mathbf{X}_i$ have, on average, the following

power constraint,

$$\frac{1}{T} \cdot E\big[\mathbf{tr}\big(\mathbf{X}_i^H \mathbf{X}_i\big)\big] = \rho \ . \tag{7.3}$$

where ρ is the average signal to noise ratio at each receive antenna. Conservation of time and energy leads to the following constraints,

$$\begin{aligned} \mathbf{tr}\big(\mathbf{X}_\tau^H \cdot \mathbf{X}_\tau\big) = MT_\tau, \qquad & E_{\mathbf{X}_{d,\,i}}\Big[\mathbf{tr}\big(\mathbf{X}_{d,\,i}^H \cdot \mathbf{X}_{d,\,i}\big)\Big] = MT_d \ , \\ T = T_\tau + T_d, \qquad & \rho T = \rho_\tau T_\tau + \rho_d T_d \ . \end{aligned} \tag{7.4}$$

Due to the insignificant capacity gain resulting from using optimal power allocation between training and data symbols as reported in [71] [92], equal power allocation is assumed in this chapter, with

$$\rho_\tau = \rho_d = \rho \ . \tag{7.5}$$

7.2.2 MIMO receiver structure

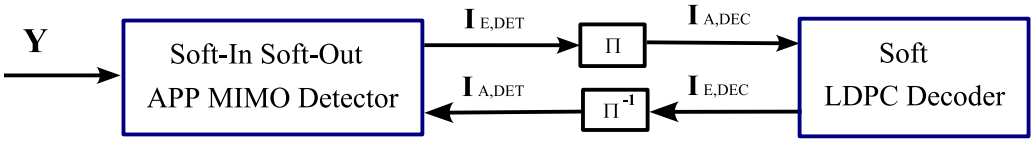

Figure 7.3: Conventional receiver structure of LDPC-coded MIMO systems

The MIMO receiver decodes the transmitted information bits $\mathbf{d}$ based on received signal matrices $\big\{\mathbf{Y}_i\big\}_{i=1}^{L}$ without knowing any instantaneous channel state information $\big\{\mathbf{H}_i\big\}_{i=1}^{L}$. The channel statistical distribution $p\big(\mathbf{H}_i\big)$ is assumed to be known both to the receiver and to the transmitter throughout the chapter. We know that even with ideal CSI, the optimal decoding algorithm for this system has an exponential complexity. Hence the near-optimal iterative receiver structure based on turbo principle [90] becomes a promising alternative.

As a standard iterative decoding procedure, the structure of the LDPC-coded MIMO receiver is demonstrated in Fig. 7.3. It consists of two important

components, an inner soft-input soft-output MIMO detector (or MIMO demodulator) and an outer soft LDPC decoder, which form a bipartite graph structure [93]. Soft log likelihood ratio of each transmitted bit is passed forward and backward between these two soft decoders with increasing accuracy as the number of the iterations increase. At each iteration, the MIMO detector forms soft extrinsic information of each coded bit based on the received symbols $\left\{\mathbf{Y}_i\right\}_{i=1}^{L}$ and the apriori information coming from the soft LDPC decoder through proper interleaving, and serves as the apriori information for the LDPC decoder in the next iteration. Convergence is reached after certain number of the iterations and decoded bits are hence obtained.

Notice that the LDPC decoder in Fig. 7.3 is itself composed of two component soft decoders (variable node decoder and check node decoder), and the entire MIMO receiver can be viewed as a complicated graph code structure. Therefore, any bipartite separation other than the conventional structure can lead to an alternative iterative decoder. We utilize in Section 7.4.1 an unconventional MIMO receiver structure, which combines the soft MIMO detector and LDPC variable node decoder together as a super component soft-decoder. The proposed receiver structure has great design advantages that can easily lead to an efficient LDPC code degree profile optimization algorithm as shown in Section 7.4.3.

7.3 Soft-Input Soft-Output MIMO Detector

As described in Section 7.2.2, the soft-input soft-output MIMO detector is an important decoding component of the MIMO receiver, and plays an important role in determining the performance of the entire coded MIMO system. Regular communication systems with unknown channel state information typically employ a two-stage decoding procedure, which consists of channel estimation followed by coherent decoding based on the estimated channel parameters. However, conventional channel estimators perform estimation based only on the training pi-

lots, thereby failing to make use of the channel information contained in the data symbols. Due to the mismatch between the actual and estimated channel, system performance suffers severe degradation especially in a communication environment with low signal to noise ratio, or limited training pilots in fast fading channels.

In this section, several better MIMO detectors which include the soft MIMO detector, EM-based MIMO detector, as well as their modified versions are proposed that offer an effective tradeoff between detection complexity and performance. The MIMO detection algorithms proposed in this section are block-based in the sense that the data detections are performed within each coherent fading block. By considering the channel coefficient correlations between adjacent coherent blocks, one could achieve even better performance by performing data detection on several adjacent coherent blocks together. In this case, the data detection algorithm has higher computational complexity and depends heavily on the correlations of the fading channel, and is beyond the scope of this chapter. Therefore for simplicity, it is reasonable to use a block fading channel model in this situation and the performance penalty of the simple block-based MIMO detection algorithm would be small by properly tuning the channel coherence time T according to the actual channel correlations.

For the sake of simplicity, subscript (or time index) i, denoting the i^{th} coherent block, is dropped in this section while describing the block-wise soft MIMO detection algorithms. To be specific, we denote $\mathbf{X} = \left[\mathbf{X}_\tau^T, \mathbf{X}_d^T\right]^T$, $\mathbf{H}$, and $\mathbf{Y} = \left[\mathbf{Y}_\tau^T, \mathbf{Y}_d^T\right]^T$ as the transmitted signal, channel matrix, and received signal in each coherent block, respectively. Furthermore, sub-matrices $\mathbf{X}_\tau$, $\mathbf{X}_d$, $\mathbf{Y}_\tau$, and $\mathbf{Y}_d$ have the following structures, i.e.

$$\begin{aligned} \mathbf{X}_\tau &= \left[\mathbf{x}_{\tau,1}^T, \cdots, \mathbf{x}_{\tau,T_\tau}^T\right]^T, \qquad & \mathbf{X}_d &= \left[\mathbf{x}_{d,1}^T, \cdots, \mathbf{x}_{d,T_d}^T\right]^T, \\ \mathbf{Y}_\tau &= \left[\mathbf{y}_{\tau,1}^T, \cdots, \mathbf{y}_{\tau,T_\tau}^T\right]^T, \qquad & \mathbf{Y}_d &= \left[\mathbf{y}_{d,1}^T, \cdots, \mathbf{y}_{d,T_d}^T\right]^T, \end{aligned} \tag{7.6}$$

where $\mathbf{x}_{\tau,k}$, $\mathbf{x}_{d,k}$, $\mathbf{y}_{\tau,k}$, and $\mathbf{y}_{d,k}$ represent complex row vectors of size $1 \times M$. Similarly, the binary sub-codeword $\mathbf{C}$ that maps to the transmitted signal $\mathbf{X}$ can also

be decomposed into

$$\mathbf{C} = \Big[\mathbf{c}_1^T, \cdots, \mathbf{c}_{T_d}^T\Big]^T, \qquad \mathbf{c}_k\Big|_{k=1}^{T_d} \in \mathbb{B}^{1\times M\cdot \log_2|\mathcal{X}|} \ , \tag{7.7}$$

where $\mathbb{B}$ is binary set $\{0,1\}$ and each row $\mathbf{c}_k$ represents the corresponding binary information that maps to $\mathbf{x}_{d,k}$.

7.3.1 Optimal soft MIMO detector

First, according to the channel model (7.2), the conditional probability density of the received signal matrix $\mathbf{Y}$ given the transmitted signal matrix $\mathbf{X}$ is given by [6]

$$p\big(\mathbf{Y}\big|\mathbf{X}\big) = \frac{\exp\left(-\mathbf{tr}\Big\{\Big[I_T + \mathbf{X}\mathbf{X}^H\Big]^{-1}\cdot \mathbf{Y}\mathbf{Y}^H\Big\}\right)}{\pi^{TN}\det^N\Big[I_T + \mathbf{X}\mathbf{X}^H\Big]} \ . \tag{7.8}$$

It is evident from the above transitional probability that the unknown MIMO channel is actually a memoryless vector channel and hence the optimal MIMO detector does not necessarily need to form a specific channel estimate.

In order to obtain the a posteriori probability of each coded bit, the a priori probability of the input signal matrix $\mathbf{X}$ is first calculated as

$$p\big(\mathbf{X}\big) = p\big(\mathbf{X}_d\big) = p(\mathbf{C}) = \prod_{k=1}^{T_d} p(\mathbf{x}_k) = \prod_{k=1}^{T_d} p(\mathbf{c}_k) = \prod_{k=1}^{T_d}\prod_{j=1}^{M\log_2|\mathcal{X}|} p(c_{k,j}) \ , \tag{7.9}$$

where each element of matrix $\mathbf{X}_d$ is a member of a complex alphabet $\mathcal{X}$ of size $|\mathcal{X}|$, and corresponding to $\log_2|\mathcal{X}|$ LDPC-coded bits. Therefore, the log likelihood ratio of each LDPC coded bit is given by

$$L_{\text{pos}}(c_{k,j}) = \log\left(\frac{p\big(c_{k,j}=1\big|\mathbf{Y}\big)}{p\big(c_{k,j}=0\big|\mathbf{Y}\big)}\right) = \log\left(\frac{\Sigma_{\mathbf{X}\in\mathcal{D}_{k,j}^+}\ p\Big(\mathbf{Y}\Big|\mathbf{X}\Big)\cdot p\big(\mathbf{X}\big)}{\Sigma_{\mathbf{X}\in\mathcal{D}_{k,j}^-}\ p\Big(\mathbf{Y}\Big|\mathbf{X}\Big)\cdot p\big(\mathbf{X}\big)}\right) \ , \tag{7.10}$$

where $1\le k\le T_d$, $1\le j\le M\cdot\log_2|\mathcal{X}|$, and $\mathcal{D}_{k,j}^+$ ($\mathcal{D}_{k,j}^-$) is the set of $\mathbf{X}$ for which the $(k,j)^{\text{th}}$ bit $c_{k,j}$ of the LDPC coded sub-block $\mathbf{C}$ is "$+1$" ("-1"). Finally,

by subtracting the input a priori information from the obtained a posterior log likelihood ratio, the soft extrinsic information of each coded bit is obtained as

$$L_{\text{ext}}(c_{k,j}) = L_{\text{pos}}(c_{k,j}) - L_{\text{app}}(c_{k,j}), \qquad L_{\text{app}}(c_{k,j}) = \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right), \quad (7.11)$$

where $L_{\text{app}}(c_{k,j})$ is the a priori information of the coded bit $c_{k,j}$ from the last iteration. Notice that there is no channel estimation stage in the soft MIMO detector described above, and therefore the proposed detection algorithm does not depend on the unknown channel state $\mathbf{H}$ but only on its underlying statistical distribution. Furthermore, the optimality of the proposed soft MIMO detection algorithm is restricted within the component level and does not depend on the overall receiver structure of the coded-MIMO system.

7.3.2 Sub-optimal soft MIMO detector

The optimal soft MIMO detection algorithm proposed in Section 7.3.1 provides the optimal extrinsic LLR values of each coded bit. However, the summation in both the numerator and the denominator of equation (7.10) consists of 2^{K-1} items, with K $\left(=T_d M \log_2 |\mathcal{X}|\right)$ increasing linearly with number of data slots T_d (or coherence time T). It has an unaffordable exponential complexity for practical communication systems, especially when the coherence time T is large. Hence, we propose a sub-optimal MIMO detector in this section with complexity increasing linearly with T_d.

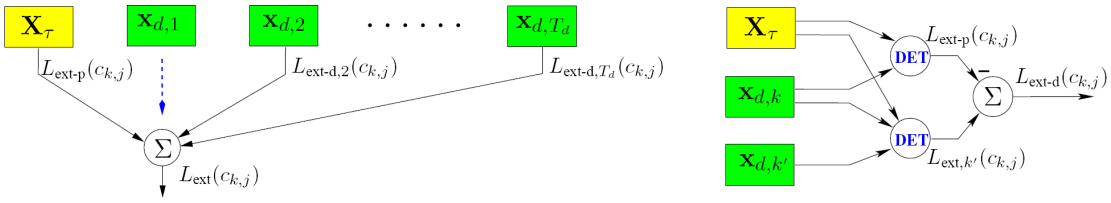

Figure 7.4: Sub-optimal soft MIMO detector structure

Notice that the optimal extrinsic LLR value of bit $c_{k,j}$ depends on the input a priori information as well as the channel observations of the entire coherent

block. Taking another point of view, the obtained extrinsic LLR is a combination of all the input information through the utilization of the proposed algorithm (7.10) in an implicit manner. Therefore, instead of performing soft MIMO detection in one operation, we can extract partial extrinsic information by processing only two rows of the data matrix $\mathbf{X}_d$ at a time, and then combining different partial extrinsic information to form the final extrinsic LLR. As illustrated in (the right side of) Fig. 7.4, in order to combine information from coded rows $\mathbf{x}_{d,k}$ and $\mathbf{x}_{d,k'}$, we first perform the optimal MIMO detection algorithm on the following reduced size *sub-coherent* block

$$\mathbf{X}_{[k,k']} = \Big[\mathbf{X}_\tau{}^T, \mathbf{x}_{d,k}^T\,, \mathbf{x}_{d,k'}^T\Big]^T, \qquad \mathbf{Y}_{[k,k']} = \Big[\mathbf{Y}_\tau{}^T, \mathbf{y}_{d,k}^T\,, \mathbf{y}_{d,k'}^T\Big]^T\ . \tag{7.12}$$

Therefore, the partial extrinsic LLR value $L_{\text{ext},\,k'}\big(c_{k,j}\big)$ of bit $c_{k,j}$ obtained from the a priori information of row $\mathbf{c}_k$, $\mathbf{c}_{k'}$, and channel observation $\mathbf{Y}_{[k,k']}$ is given by

$$L_{\text{ext},\,k'}\big(c_{k,j}\big) = \log\left(\frac{\Sigma_{\mathbf{X}_{[k,k']}\in\mathcal{D}_{k,j}^{+}} p\Big(\mathbf{Y}_{[k,k']}\Big|\mathbf{X}_{[k,k']}\Big)\, p\big(\mathbf{X}_{[k,k']}\big)}{\Sigma_{\mathbf{X}_{[k,k']}\in\mathcal{D}_{k,j}^{-}} p\Big(\mathbf{Y}_{[k,k']}\Big|\mathbf{X}_{[k,k']}\Big)\, p\big(\mathbf{X}_{[k,k']}\big)}\right) - \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right), \tag{7.13}$$

where

$$1 \le k, k' \le T_d, \qquad 1 \le j \le M\log_2|\mathcal{X}|\ ,$$

and $\mathcal{D}_{k,j}^{+}$ ($\mathcal{D}_{k,j}^{-}$) is the set of $\mathbf{X}_{[k,k']}$ for which bit $c_{k,j}$ is "$+1$" ("-1"). By the same reasoning, partial extrinsic information of bit $c_{k,j}$, related to (and contained in) the a priori information of $\mathbf{c}_k$ and channel observations $\mathbf{Y}_\tau$ and $\mathbf{y}_{d,k}$ can also be obtained by performing optimal detection on the following *sub-coherent* block

$$\mathbf{X}_{[k]} = \Big[\mathbf{X}_\tau^T, \mathbf{x}_{d,k}^T\Big]^T, \qquad \mathbf{Y}_{[k]} = \Big[\mathbf{Y}_\tau^T, \mathbf{y}_{d,k}^T\Big]^T\ , \tag{7.14}$$

with the corresponding extrinsic LLR value given by

$$L_{\text{ext-p}}\big(c_{k,j}\big) = \log\left(\frac{\Sigma_{\mathbf{X}_{[k]}\in\mathcal{D}_{k,j}^{+}} p\Big(\mathbf{Y}_{[k]}\Big|\mathbf{X}_{[k]}\Big)\cdot p\big(\mathbf{X}_{[k]}\big)}{\Sigma_{\mathbf{X}_{[k]}\in\mathcal{D}_{k,j}^{-}} p\Big(\mathbf{Y}_{[k]}\Big|\mathbf{X}_{[k]}\Big)\cdot p\big(\mathbf{X}_{[k]}\big)}\right) - \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right), \tag{7.15}$$

where

$$1 \le k \le T_d, \qquad 1 \le j \le M\log_2|\mathcal{X}|\ .$$

Having obtained extrinsic information $L_{\text{ext},\,k'}\big(c_{k,j}\big)$ and $L_{\text{ext-p}}\big(c_{k,j}\big)$, one can obtain by the following substraction,

$$L_{\text{ext-d},\,k'}\big(c_{k,j}\big) = L_{\text{ext},\,k'}\big(c_{k,j}\big) - L_{\text{ext-p}}\big(c_{k,j}\big)\ , \tag{7.16}$$

the extrinsic information of bit $c_{k,j}$ extracted solely from the channel observation $\mathbf{y}_{d,k'}$ and the a priori information of $\mathbf{c}_{k'}$. In contrast to the situation of perfect channel state information at the receiver (CSIR) where $L_{\text{ext}}(c_{k,j})$ only depends on the a priori knowledge of $\mathbf{c}_k$ and observation $\mathbf{y}_{d,k}$, a non-zero extrinsic information of $c_{k,j}$ can be obtained from the a priori knowledge of $\mathbf{c}_{k'}$ and observation $\mathbf{y}_{k'}$ (with $k' \neq k$) in an unknown MIMO fading environment. An intuitive explanation of the above difference can be made by viewing $\mathbf{c}_{k'}$ as partially fixed pilots based on the input a priori information. Therefore, better channel knowledge is learned (although no explicit channel estimation exists), which translates into a better a posterior probability of $c_{k,j}$. Hence, a non-zero partial extrinsic information solely from the a priori probability of $\mathbf{c}_{k'}$ and the channel observation $\mathbf{y}_{k'}$ is obtained.

Due to the assumption that the input a priori information of different bits are independent, all the partial extrinsic information $L_{\text{ext-d},k'}(c_{k,j})$ and $L_{\text{ext-p}}(c_{k,j})$ can be viewed as being close to independent. As illustrated in (the left side of) Fig. 7.4, the final output extrinsic information $L_{\text{ext}}(c_{k,j})$ is obtained by summing all the independent partial extrinsic information obtained from different coded rows $\mathbf{c}_{k'}$ and pilot observations, i.e.

$$\begin{aligned} L_{\text{ext}}\big(c_{k,j}\big) &= \sum_{\substack{k'=1\\k'\neq k}}^{T_d} L_{\text{ext-d},\,k'}\big(c_{k,j}\big) + L_{\text{ext-p}}\big(c_{k,j}\big) \\ &= \sum_{\substack{k'=1\\k'\neq k}}^{T_d} L_{\text{ext},\,k'}\big(c_{k,j}\big) - (T_d-2)\cdot L_{\text{ext-p}}\big(c_{k,j}\big)\ , \end{aligned} \tag{7.17}$$

where

$$1 \le k \le T_d, \quad 1 \le j \le M\log_2|\mathcal{X}|\ .$$

A summation of $2^{2M\log_2|\mathcal{X}|}$ terms is required to extract the partial extrinsic information $L_{\text{ext},k'}(c_{k,j})$ in equation (7.13) and $2^{M\log_2|\mathcal{X}|}$ terms for $L_{\text{ext-p}}(c_{k,j})$ in

equation (7.15). Therefore, in order to obtain the output soft extrinsic LLR values, a total number of $\left((T_d-1)\cdot 2^{2M\log_2|\mathcal{X}|}+2^{M\log_2|\mathcal{X}|}\right)$ terms of probability summation is required for each coded bit, as opposed to $2^{T_dM\log_2|\mathcal{X}|}$ terms in the original optimal soft MIMO detector. Furthermore, the proposed sub-optimal soft MIMO detection algorithm can be easily generalized by extracting partial extrinsic information through combining more than two (E in general) rows of the sub-codeword $\mathbf{C}$ together. By choosing different combination size of $2 \le E \le T_d$, a group of sub-optimal MIMO detectors can be constructed which offer a varying degree of detection complexity to system performance tradeoff.

7.3.3 Sub-optimal butterfly soft MIMO detector

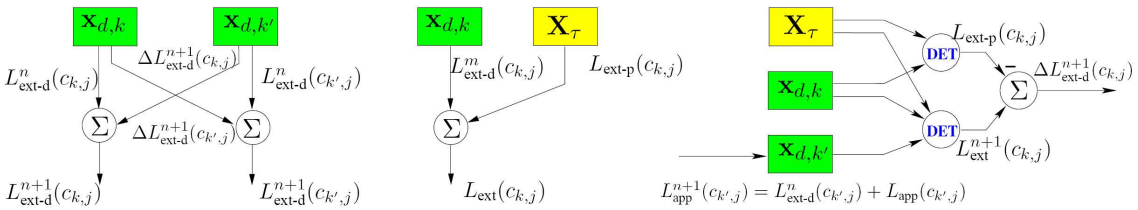

Figure 7.5: Sub-optimal soft MIMO detector using butterfly structure

Motivated by the fast Fourier transform (FFT) algorithm, we can further reduce the complexity of the soft MIMO detector to $\left(\log_2 T_d\cdot 2^{2M\log_2|\mathcal{X}|}+2^{M\log_2|\mathcal{X}|}\right)$ terms of summation per coded bit by using a sub-optimal butterfly MIMO detector structure as illustrated in Fig. 7.5. It is first assumed that the number of the data slots $T_d = 2^m$, a power of 2. If not, we can appropriately zero-pad the transmitted signal matrix $\mathbf{X}$. As demonstrated in (the left part of) Fig. 7.5, the sub-optimal butterfly detection algorithm obtains the extrinsic information through a multi-level structure similar to the fast Fourier transform, where the extrinsic information is accumulated from level to level. Specifically, if the partial extrinsic LLR value of coded bit $c_{k,j}$ at the n^{th} level is $L^{n}_{\text{ext-d}}\left(c_{k,j}\right)$, then the extrinsic LLR value of the $(n+1)^{\text{th}}$ level can be updated by the following form, which is illustrated in (the

right part of) of Fig. 7.5,

$$L_{\text{ext-d}}^{n+1}\big(c_{k,j}\big) = L_{\text{ext-d}}^{n}\big(c_{k,j}\big) + \Delta L_{\text{ext-d}}^{n+1}\big(c_{k,j}\big), \qquad 0 \le n \le m-1 \ , \tag{7.18}$$

where the second term $\Delta L_{\text{ext-d}}^{n+1}\big(c_{k,j}\big)$ of equation (7.18) represents the additional partial extrinsic information obtained from the information of coded bits $\mathbf{c}_{k'}$, with sub-codeword row index k' given by

$$k' = \begin{cases} k + 2^{m-n-1} & \text{if} \quad k \ (\text{mod} \ 2^{m-n}) < 2^{m-n-1} \\ k - 2^{m-n-1} & \text{if} \quad k \ (\text{mod} \ 2^{m-n}) \ge 2^{m-n-1} \end{cases} \ . \tag{7.19}$$

Similar to the extraction algorithm provided in (7.16), $\Delta L_{\text{ext-d}}^{n+1}\big(c_{k,j}\big)$ is given by the following form

$$\Delta L_{\text{ext-d}}^{n+1}\big(c_{k,j}\big) = L_{\text{ext}}^{n+1}\big(c_{k,j}\big) - L_{\text{ext-p}}\big(c_{k,j}\big) \ , \tag{7.20}$$

where $L_{\text{ext-p}}\big(c_{k,j}\big)$ is given by equation (7.15), and partial extrinsic information $L_{\text{ext}}^{n+1}\big(c_{k,j}\big)$ is obtained by performing optimal soft MIMO detection on the sub-coherent block $\mathbf{X}_{[k,k']}$ and $\mathbf{Y}_{[k,k']}$ with modified input a priori information, i.e.

$$L_{\text{ext}}^{n+1}\big(c_{k,j}\big) = \log\left(\frac{\Sigma_{\mathbf{X}_{[k,k']}\in\mathcal{D}_{k,j}^{+}} p\Big(\mathbf{Y}_{[k,k']}\Big|\mathbf{X}_{[k,k']}\Big)\, p_{\text{app}}^{n+1}\big(\mathbf{X}_{[k,k']}\big)}{\Sigma_{\mathbf{X}_{[k,k']}\in\mathcal{D}_{k,j}^{-}} p\Big(\mathbf{Y}_{[k,k']}\Big|\mathbf{X}_{[k,k']}\Big)\, p_{\text{app}}^{n+1}\big(\mathbf{X}_{[k,k']}\big)}\right) - \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right). \tag{7.21}$$

Furthermore, the modified a priori probability $p_{\text{app}}^{n+1}\big(\mathbf{X}_{[k,k']}\big)$ in equation (7.21) is a combination of the a priori probability of $\mathbf{c}_k$ and $\mathbf{c}_{k'}$ as well as the n^{th} level extrinsic information of $\mathbf{c}_{k'}$, which can be represented as

$$p_{\text{app}}^{n+1}\big(\mathbf{X}_{[k,k']}\big) = p\big(\mathbf{c}_k\big)\cdot p\big(\mathbf{c}_{k'}\big)\cdot p_{\text{ext}}^{n}\big(\mathbf{c}_{k'}\big) = \prod_{j=1}^{M\log_2|\mathcal{X}|} p\big(c_{k,j}\big)\cdot p\big(c_{k',j}\big)\cdot p_{\text{ext}}^{n}\big(c_{k',j}\big) \ , \tag{7.22}$$

where $p_{\text{ext}}^{n}\big(c_{k',j}\big)$ is given by

$$p_{\text{ext}}^{n}\big(c_{k',j}\big) \ = \ \frac{\exp\Big(c_{k',j}\cdot L_{\text{ext-d}}^{n}\big(c_{k',j}\big)\Big)}{1+\exp\Big(L_{\text{ext-d}}^{n}\big(c_{k',j}\big)\Big)} \ . \tag{7.23}$$

Therefore, $\Delta L^{n+1}_{\text{ext-d}}\big(c_{k,j}\big)$ can be viewed as the partial extrinsic information obtained solely from the a priori information of $\mathbf{c}_{k'}$, channel observation $\mathbf{y}_{k'}$, and its extrinsic information at the n^{th} level.

Starting from the initial condition $L^0_{\text{ext-d}}(c_{k,j}) = 0$, the extrinsic information $L^n_{\text{ext-d}}(c_{k,j})$ of each coded bit is accumulated at each level by absorbing additional partial extrinsic information through the sub-coherent block combining process. As illustrated in (the middle part of) Fig. 7.5, the final soft extrinsic LLR value of each coded bit is formed by combining the extrinsic LLR information at the m^{th} (lowest) level with the extrinsic information obtained from pilot observations, which is given by

$$L_{\text{ext}}\big(c_{k,j}\big) = L^m_{\text{ext-d}}\big(c_{k,j}\big) + L_{\text{ext-p}}\big(c_{k,j}\big) \qquad 1 \le k \le T_d, \qquad 1 \le j \le M \log_2 |\mathcal{X}| \quad (7.24)$$

Note that both the sub-optimal structure in Section 7.3.2 as well as the sub-optimal butterfly MIMO detector in the previous subsection are modifications of the optimal soft MIMO detection algorithm provided in Section 7.3.1. The two sub-optimal MIMO detection algorithms provided in Section 7.3.2 and 7.3.3 have the following structural differences. First, the sub-optimal MIMO detector in Section 7.3.2 forms extrinsic information through a *linear* combining structure, where there are a total of $(T_d - 1)$ partial extrinsic information terms (each corresponding to the partial extrinsic LLR obtained from other rows k'); each term is computed by performing optimal detection on the sub-coherent block given by (7.13)-(7.16). On the other hand, the sub-optimal butterfly MIMO detector in Section 7.3.3 performs data detection by employing a multi-level structure, where the extrinsic information is distributed at succeeding levels until all the input a priori information and the channel observations are combined and exchanged between all different rows.

7.3.4 Modified EM-based MIMO detector

The soft MIMO detector and its two sub-optimal modifications proposed in previous sections perform data detection without forming any specific channel

estimate. However, forming a channel estimation followed by coherent MIMO detection is in some cases a promising alternative especially when there are enough training pilots. Besides, the estimated channel state information $\widehat{\mathbf{H}}$ can be easily fedback to the transmitter for better power allocation and spectral shaping of the channel coding.

Recently, a lot of attention has been focused on turbo MAP EM estimators, which can take into account not only the training pilots but also the a priori information of the coded bits from the outer soft LDPC decoder. As reported in [87] [88], the proposed turbo EM estimator provides better performance than the conventional MMSE-based channel estimator and works well in an iterative decoding algorithm, especially when T_d is large. However, there exists positive feedback between the input and output soft LLR values which can cause severe performance degradation of the coded MIMO system. Therefore, we propose in this section a modified EM-based MIMO detector that avoids positive feedback and results in better performance than the direct EM-based detection algorithm. Mutual information transfer characteristic of the modified EM-based detector as well as the corresponding simulation results provided in Section 7.4 and 7.5 further confirm our claims of superiority of the new detector.

To start with the detection algorithm, let us first look at the conventional MAP EM estimator, whose objective is to find the channel estimation $\widehat{\mathbf{H}}$ that maximizes a posterior probability

$$\widehat{\mathbf{H}} = \arg \max_{\mathbf{H}} p\big(\mathbf{H}\big|\mathbf{Y}\big) = \arg \max_{\mathbf{H}} p\big(\mathbf{Y}\,,\mathbf{H}\big) \ , \tag{7.25}$$

which is intractable by direct maximization. Hence by taking the transmitted data signal matrix $\mathbf{X}_d$ (or $\mathbf{X}$) as the unobserved (or missing) data, the following iterative expectation maximization (EM) algorithm (similar to [87]) is applied .

- **E-step:**

$$Q\big(\mathbf{H}\big|\widehat{\mathbf{H}}^{(n)}\big) = E_{\mathbf{X}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{Y}}\Big[-\log p\big(\mathbf{H},\mathbf{Y}\big|\mathbf{X}\big)\Big] \ . \tag{7.26}$$

After some manipulations, we have the following concise form

$$Q\big(\mathbf{H}\big|\widehat{\mathbf{H}}^{(n)}\big) = E_{\mathbf{X}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{Y}}\Big[\mathbf{tr}\Big(\mathbf{H}^H\mathbf{H} + \big(\mathbf{Y}-\mathbf{X}\mathbf{H}\big)^H\big(\mathbf{Y}-\mathbf{X}\mathbf{H}\big)\Big)\Big]$$
$$= \mathbf{tr}\Big(\mathbf{H}^H\mathbf{R}\ \mathbf{H} + \mathbf{Y}^H\mathbf{Y} - \big(\mathbf{Y}^H\mathbf{U}\mathbf{H} + \mathbf{H}^H\mathbf{U}^H\mathbf{Y}\big)\Big), \qquad (7.27)$$

where $\mathbf{R}$ is given by

$$\mathbf{R} = E_{\mathbf{X}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{Y}}\big[\mathbf{X}^H\mathbf{X}\big] + I_M$$
$$= \frac{\rho}{M}\cdot\Big(\sum_{j=1}^{T_d}\sum_{\mathbf{x}_{d,k}\in\mathcal{X}^M} p\big(\mathbf{x}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{y}_{d,k}\big)\cdot\mathbf{x}_{d,k}^H\,\mathbf{x}_{d,k} + \mathbf{X}_\tau^H\mathbf{X}_\tau\Big) + I_M\,, \qquad (7.28)$$

and $\mathbf{U}$ is given by

$$\begin{aligned}\mathbf{U} &= E_{\mathbf{X}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{Y}}\big[\mathbf{X}\big] = \sqrt{\frac{\rho}{M}}\cdot\Big[\mathbf{X}_\tau^T,\,\boldsymbol{\mu}_1^T,\cdots,\boldsymbol{\mu}_{T_d}^T\Big]^T,\\ \boldsymbol{\mu}_k &= \sum_{\mathbf{x}_{d,k}\in\mathcal{X}^M} p\big(\mathbf{x}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{y}_{d,k}\big)\cdot\mathbf{x}_{d,k}\,. \end{aligned} \qquad (7.29)$$

The a posterior probability $p\big(\mathbf{x}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{y}_{d,k}\big)$ is given by

$$p\big(\mathbf{x}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{y}_{d,k}\big) = \frac{p\big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{x}_{d,k}\big)\cdot p\big(\mathbf{x}_{d,k}\big)}{\sum_{\mathbf{x}_{d,k}\in\mathcal{X}^M} p\big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{x}_{d,k}\big)\cdot p\big(\mathbf{x}_{d,k}\big)}\ , \qquad (7.30)$$

with $p\big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{x}_{d,k}\big)$ and $p\big(\mathbf{x}_{d,k}\big)$ given as

$$p\big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}^{(n)},\mathbf{x}_{d,k}\big) = \frac{1}{\pi^N}\exp\Big(-\big\|\mathbf{y}_{d,k} - \sqrt{\frac{\rho}{M}}\cdot\widehat{\mathbf{H}}^{(n)}\mathbf{x}_{d,k}\big)\big\|^2\Big)\ , \qquad (7.31)$$

where $p\big(\mathbf{x}_{d,k}\big)$ is given by

$$p\big(\mathbf{x}_{d,k}\big) = \prod_{j=1}^{M\log_2|\mathcal{X}|} p\big(c_{k,j}\big)\ . \qquad (7.32)$$

- **M-step:**

$$\widehat{\mathbf{H}}^{(n+1)} = \arg\ \min_{\mathbf{H}}\ Q\big(\mathbf{H}\big|\widehat{\mathbf{H}}^{(n)}\big)\ . \qquad (7.33)$$

After some manipulations, the updated channel estimation $\widehat{\mathbf{H}}^{(n+1)}$ is obtained as

$$\widehat{\mathbf{H}}^{(n+1)} = \mathbf{R}^{-1}\cdot\mathbf{U}^H\cdot\mathbf{Y}\ . \qquad (7.34)$$

- **Initialization:**

  We use the conventional minimal mean square error (MMSE) channel estimator for initialization, which is given by

$$\widehat{\mathbf{H}}^{(0)} = \sqrt{\frac{\rho}{M}} \cdot \mathbf{X}_{\tau}^{H} \cdot \big(\frac{\rho}{M}\mathbf{X}_{\tau}\mathbf{X}_{\tau}^{H} + I_{T_{\tau}}\big)^{-1} \cdot \mathbf{Y}_{\tau} \ . \tag{7.35}$$

  Since the EM iteration is embedded within the large iterative decoding loop of the soft MIMO receiver, we can also take the estimated channel $\mathbf{H}$ from the last decoding iteration as an EM initialization. Compared with the simple MMSE estimator, the obtained estimation from the last decoding iteration (through an EM algorithm) provides a better initialization since additional a priori information of the coded bits is used. Therefore, by using the alternative initialization, EM algorithm is able to begin at a better starting point and hence results in smaller number of EM iterations.

Maximum a posterior channel estimation is obtained when the MAP EM algorithm converge to $\widehat{\mathbf{H}}$ after certain number of iterations. Hence the soft extrinsic information of each coded bit $c_{k,j}$ is provided by taking $\widehat{\mathbf{H}}$ as the true channel coefficients followed by coherent MIMO detection,

$$L_{\text{ext}}\big(c_{k,j}\big) = \log\left(\frac{\Sigma_{\mathbf{x}_{d,k}\in\mathcal{D}_{k,j}^{+}}\ p\Big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}},\mathbf{x}_{d,k}\Big)\cdot p\big(\mathbf{x}_{d,k}\big)}{\Sigma_{\mathbf{x}_{d,k}\in\mathcal{D}_{k,j}^{-}}\ p\Big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}},\mathbf{x}_{d,k}\Big)\cdot p\big(\mathbf{x}_{d,k}\big)}\right) - \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right) \ , \tag{7.36}$$

where

$$1 \le k \le T_d, \qquad 1 \le j \le M\log_2|\mathcal{X}| \ .$$

$\mathcal{D}_{k,j}^{+}$ ($\mathcal{D}_{k,j}^{-}$) is the set of $\mathbf{x}_{d,k}$ for which bit $c_{k,j}$ is "$+1$" ("-1"), and probabilities $p\Big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}},\mathbf{x}_{d,k}\Big)$ and $p\big(\mathbf{x}_{d,k}\big)$ are given by (7.31) and (7.32) respectively.

It is well known that short girth in the LDPC Tanner graph is one of the major performance bottleneck for short length LDPC code design [94] [95], where positive feedback of the iterative LLR values generated by the existing short length loops directly affects the iterative message passing algorithm. Similarly, positive

feedback caused by the correlations between input and output extrinsic information of the MIMO detector will also cause severe system performance degradation. Therefore, considerable effort has been made in various detection algorithms to avoid the same information from counting twice, or to avoid the output extrinsic LLR values from containing any input a priori information.

Unfortunately, if we study the conventional direct EM-based soft MIMO detection algorithm carefully, we will find that the estimated channel coefficient does depend on the a priori information of the entire sub-codeword $\mathbf{C}$. To be specific, channel estimation $\widehat{\mathbf{H}}$ can be represented as a function given by

$$\widehat{\mathbf{H}} = \widehat{\mathbf{H}}\Big(\mathbf{Y}, \mathbf{A}_{\text{L-app}}\Big) \ , \tag{7.37}$$

where $\mathbf{A}_{\text{L-app}} \in \mathbb{R}^{T_d \times M \log_2 |\mathcal{X}|}$ is the a priori information matrix with each element $a_{k,j}$ equal to the a priori LLR value $L_{\text{app}}(c_{k,j})$. Therefore, the extrinsic information obtained by equation (7.36) contains the input a priori information through $\widehat{\mathbf{H}}$, in a sense that $L_{\text{ext}}(c_{k,j})$ depends on $L_{\text{app}}(c_{k,j})$, even though the a priori LLR value is already subtracted from the log a posterior value as demonstrated by the second term. In order to eliminate input-output correlations introduced by the direct channel estimation $\widehat{\mathbf{H}}$, which is a function of $L_{\text{app}}\big(c_{k,j}\big)$, we propose a modified EM channel estimation algorithm that uses only part of the a priori information (a subset of matrix $\mathbf{A}_{\text{L-app}}$) of the sub-codeword $\mathbf{C}$. If we denote $\mathcal{E}$ as a subset of $\{1, 2, \cdots, T_d\}$ that includes k, the partial a priori information matrix can be formed by the following weighting operation

$$\mathbf{A}^{\mathcal{E}}_{\text{L-app}} = \mathbf{diag}\big(\mathbf{s}\big) \cdot \mathbf{A}_{\text{L-app}} \ , \tag{7.38}$$

where the selecting vector $\mathbf{s}$ of size $1 \times T_d$ is given by

$$\mathbf{s} = \big[s_1, s_2, \cdots, s_{T_d}\big], \qquad s_j = \begin{cases} 1 & \text{if } j \notin \mathcal{E} \\ 0 & \text{if } j \in \mathcal{E} \end{cases} \ . \tag{7.39}$$

The modified channel estimation is hence obtained by applying the same APP EM algorithm by using $\mathbf{A}^{\mathcal{E}}_{\text{L-app}}$ as the input a priori information matrix instead, i.e.

$$\widehat{\mathbf{H}}_{\mathcal{E}} = \widehat{\mathbf{H}}\Big(\mathbf{Y}, \mathbf{A}^{\mathcal{E}}_{\text{L-app}}\Big) \ . \tag{7.40}$$

The modified estimation $\widehat{\mathbf{H}}_{\mathcal{E}}$ can therefore be used to perform coherent detections for coded rows $\mathbf{c}_k$ with index $k \in \mathcal{E}$,

$$L_{\text{ext}}\big(c_{k,j}\big) = \log\left(\frac{\Sigma_{\mathbf{x}_{d,k}\in\mathcal{D}_{k,j}^{+}}\; p\Big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}_{\mathcal{E}},\mathbf{x}_{d,k}\Big)\cdot p\big(\mathbf{x}_{d,k}\big)}{\Sigma_{\mathbf{x}_{d,k}\in\mathcal{D}_{k,j}^{-}}\; p\Big(\mathbf{y}_{d,k}\big|\widehat{\mathbf{H}}_{\mathcal{E}},\mathbf{x}_{d,k}\Big)\cdot p\big(\mathbf{x}_{d,k}\big)}\right) - \log\left(\frac{p\big(c_{k,j}=1\big)}{p\big(c_{k,j}=0\big)}\right). \tag{7.41}$$

Let us further assume that the entire set $\{1,2,\cdots,T_d\}$ can be decomposed into the following disjoint sets with the same size, i.e.

$$\{1,2,\cdots,T_d\} = \bigcup_{n}\mathcal{E}_n, \qquad \mathcal{E}_n\bigcap\mathcal{E}_{n'} = \phi, \qquad \big|\mathcal{E}_n\big| = S_E \ . \tag{7.42}$$

Instead of having only one EM estimation in the direct EM-based detector, $\lceil T_d/S_E\rceil$ separate EM estimations are to be completed during one entire soft decoding iteration in the modified EM-based detector. Note that $\mathcal{E}_n = \{n\}$ and $\mathcal{E}_n = \phi$ correspond to special cases: $\mathcal{E}_n = \{n\}$ has the maximum detection complexity, but takes into account all available a priori information from the outer soft decoder, while on the other hand $\mathcal{E}_n = \phi$ corresponds to the case of conventional direct EM-based detection algorithm. For a short complexity analysis, we know that within each EM estimation, there are total $N_{sum}^{EM} = \bar{I}\cdot\big(3T_d\cdot 2^{M\log_2|\mathcal{X}|}\big)$ summation operations, where $\bar{I}$ is the average number of iterations required by the convergence of the EM algorithm. Therefore, the average number of the summations for each coded bit in the modified EM-based MIMO detector is

$$N_{sum}^{DEC} = \Big(3\bar{I}\cdot\big\lceil T_d/S_E\big\rceil + 1\Big)\cdot 2^{M\log_2|\mathcal{X}|} \ . \tag{7.43}$$

The EM channel estimation algorithms proposed in this section can make full use of the soft a priori information of the coded bits from the outer LDPC decoder, and hence provide better (and more accurate) channel estimations. From another point of view, the MAP EM estimator is generally equivalent to extending the pilot structure to the entire transmitted signal matrix $\mathbf{X}$. Instead of limiting the pilots to $\mathbf{X}_\tau$, the receiver treats $\mathbf{X}_d$ as partially fixed pilots as well especially when the LLR ratios are getting significantly improved as a result of the messages being updated constantly through the iterations.

Finally, a brief comparison of the pilots size required by the different MIMO detection algorithms is as follow. First, we note that the proposed optimal soft MIMO detector as well as its two sub-optimal modifications are able to provide soft data detections with arbitrary number of pilot symbols and only need a small number of pilots in order to remove detection ambiguity (in the first decoding iteration). The modified EM-based detector only requires a small number ($T_\tau \geq M$) of pilots for the initialization of the EM estimation. Therefore, these four soft MIMO detection algorithms provide a wide range of trade-offs between complexity and performance and can work in different MIMO fading environments and support various training sizes.

7.4 Design of LDPC-coded MIMO Systems

Conventionally the coded MIMO receiver is obtained by connecting the inner soft MIMO detector and the outer LDPC decoder to form one large iterative decoding loop. As evident from Fig. 7.6, the overall MIMO receiver actually consists of two iterative decoding loops. In the outer loop, the soft MIMO detector forms extrinsic information of each coded bit $\left\{\mathbf{C}_i\right\}_{i=1}^{L}$ based on the received signal $\left\{\mathbf{Y}_i\right\}_{i=1}^{L}$ as well as the input a priori knowledge from the LDPC decoder, and serves as the input a priori information for the LDPC decoder in the next iteration. The soft LDPC decoder has an inner iterative decoding loop that is composed of a variable node decoder, a check node decoder, and two connecting edge interleavers. The soft extrinsic information, which describes the uncertainty of each coded bit, is iteratively exchanged in the outer loop between the MIMO detector and LDPC decoder as well as in the inner loop between variable node and check node decoders inside the LDPC decoder.

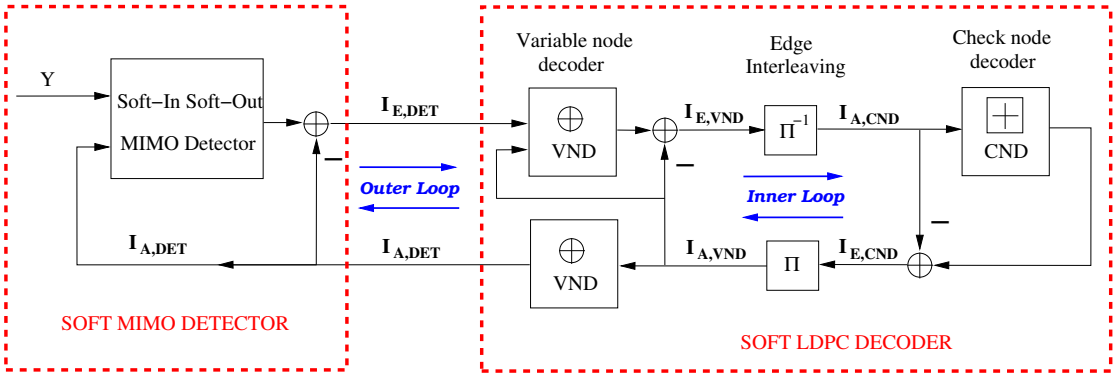

Figure 7.6: Conventional receiver structure of LDPC-coded MIMO systems

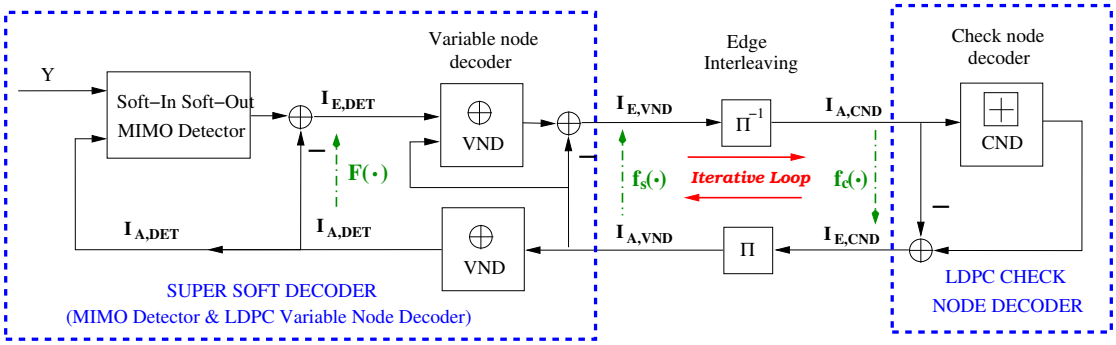

Figure 7.7: New receiver structure of LDPC-coded MIMO systems

7.4.1 Receiver structure of the LDPC-coded MIMO systems

In this chapter, we structure the MIMO receiver differently by combining the soft MIMO detector and LDPC variable node decoder as a super soft-decoder, a form also utilized in [91]. As illustrated in Fig. 7.7, the decoding loop is formed by exchanging extrinsic information between the super decoder and the LDPC check node decoder iteratively. Compared with the conventional iterative MIMO receiver (named as bit-interleaved coded modulation with iterative decoding (BICM-ID) algorithm) shown in Fig. 7.6, the new receiver structure has two advantages. First, the proposed receiver structure has only one iterative decoding loop and hence achieves smaller decoding complexity compared to the two iterative loops (inner LDPC decoder loop and outer "MIMO detector $\rightleftarrows$ LDPC decoder"

loop) in the conventional BICM-ID structure. Second, the proposed structure has the advantage of enabling the extrinsic information transfer characteristic function of the soft component decoders to have tractable forms. By fully exploiting the closed form EXIT functions, a simple and efficient LDPC code degree profile optimization algorithm with proven global optimality and guaranteed convergence is proposed in Section 7.4.3, which is superior to the sub-optimal manual curve fitting technique [89] [91].

7.4.2 Analysis of extrinsic information transfer characteristics

In order to understand as well as design the iterative decoding systems having bipartite graph structures, we use the extrinsic information transfer characteristic of the soft MIMO detector and LDPC decoder, which was proposed by Brink in [89], to analyze the convergence behavior of the iterative decoding schemes of the coded MIMO system.

1. Brief introduction on EXIT-chart

   We briefly describe in this section the EXIT-chart technique proposed in [89]. For readers who are familiar with the topic, please skip to Section 7.4.2-2 directly. The extrinsic information transfer (EXIT) function is used to the describe the input-output (a priori information versus extrinsic information) relations of the soft component decoders from an information theoretical perspective. Taking the component soft decoders in Fig. 7.7 as an example, the corresponding EXIT functions of the super soft-decoder and LDPC check node decoder can be described by the following mapping (also depicted in Fig. 7.7 accordingly),

$$I_{\mathrm{E,VND}} = f_s\big(I_{\mathrm{A,VND}}\big)\,, \qquad I_{\mathrm{E,CND}} = f_c\big(I_{\mathrm{A,CND}}\big)\,, \tag{7.44}$$

   where $I_{\mathrm{A,VND}}$ represents the mutual information between the coded bit x and the input a priori information of the super soft-decoder, and $I_{\mathrm{E,VND}}$, $I_{\mathrm{E,CND}}$,

as well as $I_{\mathrm{A,CND}}$ are similarly defined. According to the iterative decoding structure, where the output extrinsic information from one component decoder is treated as a priori input to the other one, the mutual information between the extrinsic LLR values and the coded bits is updated through the following evolution,

$$I^{k}_{\mathrm{E,CND}} = f_c \circ f_s\Big(I^{k-1}_{\mathrm{E,CND}}\Big), \qquad f_c \circ f_s(\cdot) = f_c\big(f_s(\cdot)\big)\ , \tag{7.45}$$

with index k indicating the k^{th} decoding iteration and the initialization is given by $I^{0}_{\mathrm{E,CND}} = 0$. As an example, we demonstrate in Fig. 7.8 the EXIT functions f_s and f_c (with x and y axis flipped) of the component soft decoders as well as the decoding trajectory of an LDPC-coded MIMO system. The 2×2 MIMO system considered has BPSK modulation, uses optimal soft MIMO detector at the receiver, and transmits over a fading channel with coherence time $T = 6$, training number $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$. The outer LDPC code is a regular $(3, 6)$ code with codeword length 8×10^4. We can observe from Fig. 7.8 that, as long as the EXIT chart curve f_s is above curve f_c^{-1} (with the x and y axis flipped), i.e.

$$f_s(x) \geq f_c^{-1}(x), \qquad 0 \leq x \leq 1\ , \tag{7.46}$$

the decoding trajectory is able to make its zigzag way until reaching the successful decoding point $(1, 1)$. Therefore, it can serve as a convergence criterion of the iterative decoding algorithm for the LDPC code design purpose.

2. EXIT characteristic of the soft MIMO detector

According to the results provided in [89] [91], the extrinsic mutual information $I_{\mathrm{E,DET}}$ between the transmitted bit x and the output LLR values $L_{\mathrm{ext}}(x)$, which measures the information contents of the output extrinsic LLR values, can be represented as

$$\begin{aligned} I_{\mathrm{E,DET}}(\rho; \sigma_A^2) &= I\Big(L_{\mathrm{ext}}(x)\,; x\Big) \\ &= \frac{1}{2} \sum_{x=\pm 1} \int_{-\infty}^{\infty} \log_2 \Big(\frac{2 p_{_E}(\xi|x)}{p_{_E}(\xi|x=+1) + p_{_E}(\xi|x=-1)} \Big) \cdot p_{_E}(\xi|x)\, d\xi, \end{aligned} \tag{7.47}$$

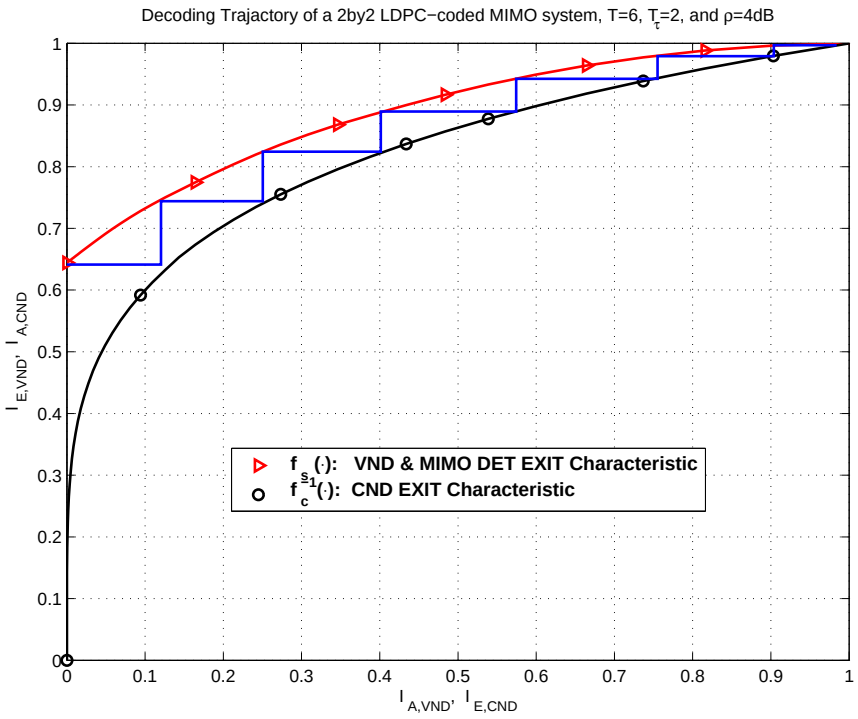

Figure 7.8: Decoding trajectory of a regular $(3, 6)$ LDPC-coded MIMO system with optimal soft MIMO detector over a 2×2 unknown MIMO channel with coherence time $T = 6$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$.

where distribution $p_{_E}(\xi|x = \pm 1)$ is obtained through Monte Carlo simulation (histogram measurements) by setting the system SNR equal to ρ and the input a priori LLR values conditioned on the transmitted bit x have a Gaussian distribution given by

$$L_{\text{app}}(x) = x \cdot n, \qquad x \in \{+1, -1\}, \qquad n \sim \mathcal{N}\Big(\frac{2}{\sigma_A^2} \,.\frac{4}{\sigma_A^2}\Big) \tag{7.48}$$

Therefore, the extrinsic mutual information $I_{\text{E,DET}}$ depends both on the system SNR ρ and the noise variance level σ_A^2 of the input a priori information. By viewing ρ as an index parameter, the EXIT function of the soft MIMO detector is given by the following form

$$I_{\text{E,DET}} = I_{\text{E,DET}}\Big(\rho\,;\ \sigma_A^2 = J^{-1}(I_{\text{A,DET}})\Big) \triangleq F\big|_{\rho}(I_{\text{A,DET}})\ , \tag{7.49}$$

where function $J(\cdot)$ is given by (equivalent to equation (24) in the Appendix of [91]),

$$\begin{aligned} J\big(\sigma_A^2\big) &= I\Big(I_{\text{app}}(x)\,;\, x\Big) \\ &= \frac{1}{\ln 2}\left(\frac{1}{\sigma_A^2} - \int_{-\infty}^{\infty} \frac{\sigma_A}{\sqrt{2\pi}} \cdot \ln\cosh(y) \cdot \exp\Big(-\frac{(\sigma_A^2 \cdot y - 1)^2}{2\sigma_A^2}\Big)\, dy\right)\ . \end{aligned} \tag{7.50}$$

Furthermore, input mutual information $I_{\text{A,DET}}$ of the soft MIMO detector is related to mutual information $I_{\text{A,VND}}$ through the following equation for a variable node of degree d_v

$$I_{\text{A,DET}} = J\Big(d_v \cdot J^{-1}\big(I_{\text{A,VND}}\big)\Big)\ . \tag{7.51}$$

3. EXIT characteristics of the LDPC variable node and check node decoders

Following the same reasoning as given in [91], the extrinsic mutual information transfer characteristic of a variable node of degree d_v is given by the following form

$$I_{\text{E,VND}}\Big(I_{\text{A,VND}},\, d_v\Big) = J\Big((d_v - 1) \cdot J^{-1}(I_{\text{A,VND}}) + J^{-1}(I_{\text{E,DET}})\Big)\ . \tag{7.52}$$

According to the duality properties [96] of the EXIT curves between single parity check codes and repetition codes over binary erasure channels, the mutual information transfer characteristic of a degree d_c check node over binary input Gaussian output channels can be well approximated as

$$I_{\mathrm{E,CND}}(I_{\mathrm{A,VND}}) \approx 1 - I_{\mathrm{E,REP}}(1 - I_{\mathrm{A,VND}}) = 1 - J\Big((d_c - 1) \cdot J^{-1}\big(1 - I_{\mathrm{A,VND}}\big)\Big). \tag{7.53}$$

7.4.3 LDPC code optimization

Following the methodology given in [89] [91], the EXIT functions of the super MIMO soft-decoder (combination of the LDPC variable node decoder and soft MIMO detector) can be obtained as

$$I_{\mathrm{E,VND}} = f_s\big(I_{\mathrm{A,VND}}\big) = \sum_{i=1}^{D_v} \lambda_i \cdot I_{\mathrm{E,VND}}\Big(I_{\mathrm{A,VND}}, d_{v,\,i}\Big)$$
$$= \sum_{i=1}^{D_v} \lambda_i \cdot J\Bigg(\big(d_{v,\,i} - 1\big) \cdot J^{-1}\big(I_{\mathrm{A,VND}}\big) + J^{-1}\Bigg(F\Big|_{\rho}\Big(J\Big(d_{v,\,i} \cdot J^{-1}\big(I_{\mathrm{A,VND}}\big)\Big)\Big)\Bigg)\Bigg). \tag{7.54}$$

where λ_i is the fraction of the variable nodes having edge degree $d_{v,\,i}$, and D_v is the number of different variable node degrees. Similarly according to (7.53), the check nodes of the LDPC code have a transfer characteristic given by the following form

$$I_{\mathrm{E,CND}} = f_c\big(I_{\mathrm{A,CND}}\big) \approx 1 - \sum_{i=1}^{D_c} \rho_i \cdot J\Big(\big(d_{c,\,i} - 1\big) \cdot J^{-1}\big(1 - I_{\mathrm{A,CND}}\big)\Big) \ , \tag{7.55}$$

where ρ_i is the fraction of the check nodes having edge degree $d_{c,\,i}$, and D_c is the number of different check node degrees.

Following the successful decoding (convergence) criterion provided in [89], the degree profile optimization problem can be reduced to the following maximization problem by taking the LDPC code rate R_{outer} as the objective

$$\max_{\{\lambda_i, \rho_i\}} R_{\mathrm{outer}} = \max_{\{\lambda_i, \rho_i\}} \left(1 - \frac{\sum_{i=1}^{D_c} \rho_i / d_{c,\,i}}{\sum_{i=1}^{D_v} \lambda_i / d_{v,\,i}}\right), \tag{7.56}$$

under linear constraints given by

$$I_{\mathrm{E,VND}}(I_{\mathrm{A,VND}}) \geq I_{\mathrm{A,CND}}(I_{\mathrm{E,CND}}) = I_{\mathrm{A,CND}}(I_{\mathrm{A,VND}}),$$
$$\sum_{i=1}^{D_v} \lambda_i = 1, \quad \sum_{i=1}^{D_c} \rho_i = 1, \qquad 0 \leq \lambda_i, \rho_i \leq 1 \ . \tag{7.57}$$

Utilizing the closed form EXIT functions of the component soft decoders given by (7.54) and (7.55), we propose an efficient LDPC code degree profile optimization algorithm in the following, which is composed of two simple linear optimization steps.

- **Variable node degree profile optimization:**

  For a fixed check node degree profile $\{\rho_i^k\}$ from the k^{th} iteration, the optimal variable node degree profile $\{\lambda_i^{k+1}\}$ is given by

  $$\{\lambda_i^{k+1}\} = \arg \max_{\{\lambda_i\}} \sum_{i=1}^{D_v} \lambda_i / d_{v,\,i} \ , \tag{7.58}$$

  under the constraints

  $$f_s\Big(f_c(a_n)\Big) \geq a_n \ ,$$
  $$\sum_{i=1}^{D_v} \lambda_i = 1, \quad 0 \leq \lambda_i \leq 1, \qquad 1 \leq n \leq N, \tag{7.59}$$

  where $\big\{a_n \big| a_n \in [0,1]\big\}$ is a set of specified constraint points, and N is the total number of constraints on the curve.

- **Check node degree profile optimization:**

  For a fixed variable node degree profile $\{\lambda_i^{k+1}\}$ from the $(k+1)^{\mathrm{th}}$ iteration, the optimal check node degree profile $\{\rho_i^{k+1}\}$ is given by

  $$\{\rho_i^{k+1}\} = \arg \min_{\{\rho_i\}} \sum_{i=1}^{D_c} \rho_i / d_{c,\,i} \ , \tag{7.60}$$

  under the constraints

  $$f_c\Big(f_s(a_n)\Big) \geq a_n \ ,$$
  $$\sum_{i=1}^{D_v} \rho_i = 1, \quad 0 \leq \rho_i \leq 1, \qquad 1 \leq n \leq N, \tag{7.61}$$

where a_n and N are similarly defined as before.

- **Initialization:**

  In general, we can start with any feasible degree profiles. Based on our experience from numerical simulations, we find that it is always a good choice to start with a regular check node degree d_c.

If we stack the LDPC code degree profile $\{\lambda_i, \rho_i\}$ into a super vector $\eta = [\lambda_1, \cdots, \lambda_{D_v}, \rho_1, \cdots, \rho_{D_c}]^T$. We can see that the objective R_{outer} given in equation (7.56) is a concave function with respect to η and that all the constraints given in (7.57) are linear. Hence, the above degree optimization problem has only one unique optimal solution. Due to the non-decreasing property of the proposed iterative maximization algorithm, it is guaranteed to converge to the global maximum solution $\eta^\star$ from any initialization point. Therefore, in contrast to the sub-optimal manual curving fitting technique proposed in [91], the above iterative LDPC optimization algorithm provides much better performance and can serve as an efficient tool for coded MIMO system design.

7.5 Numerical and Simulation Results

7.5.1 Elimination of positive feedback in EM-based MIMO detectors

We demonstrate in Fig. 7.9 the extrinsic information transfer functions of the EM-based and modified EM-based soft MIMO detectors over an unknown 2×2 MIMO channel with coherence time interval $T = 6$ and 18, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$. BPSK modulation is assumed for all the simulation results in this section unless explicitly mentioned. For comparison purpose, the mutual information transfer characteristics of the simple MMSE-based MIMO detector and detector with ideal CSIR are also included in the plot.

We can observe from the plot that for direct EM-based MIMO detector, output extrinsic mutual information $I_{\text{E,DET}}$ is even greater than that of the detec-

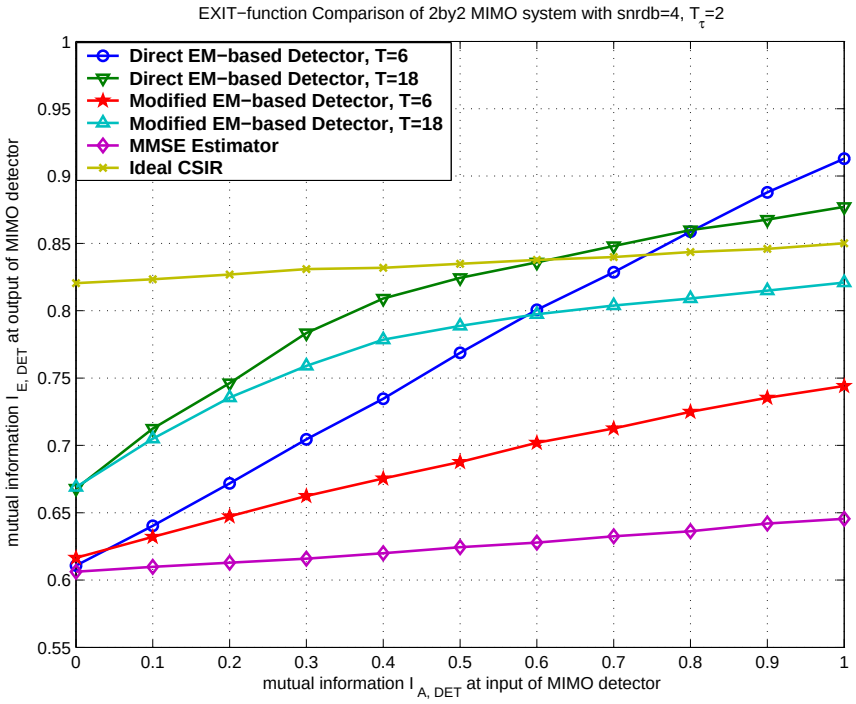

Figure 7.9: Extrinsic information transfer characteristic of the EM-based MIMO detectors over a 2×2 unknown MIMO channel with training length $T_\tau = 2$ and signal to noise ratio $\rho = 4dB$.

tor with ideal CSIR in high $I_{\mathrm{A,DET}}$ ranges, which directly indicates the existing positive feedback between the input and output extrinsic information. Such strong correlations between the output extrinsic information $L_{\mathrm{ext}}(x)$ and the input a priori information $L_{\mathrm{app}}(x)$ will cause a severe performance degradation as verified by the simulations results provided in Section 7.5.4. As expected, the modified EM-based MIMO detectors successfully eliminate the correlation and achieve significant performance gain compared to the simple MMSE-based detector especially when the coherence time interval T is large.

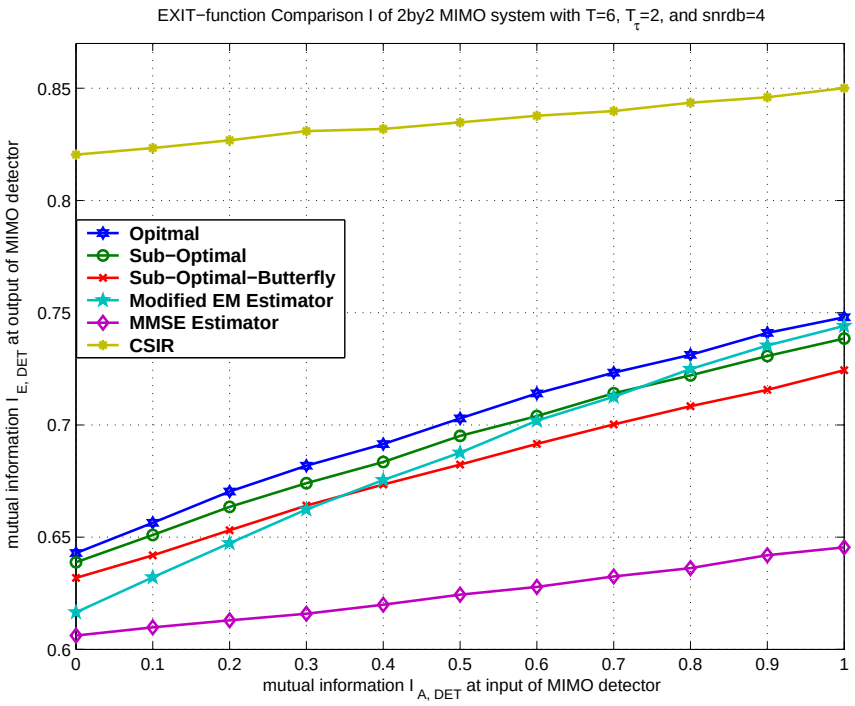

Figure 7.10: Comparison of the extrinsic information transfer characteristic of different MIMO detectors in a 2×2 unknown MIMO system with coherence time $T = 6$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$.

7.5.2 EXIT function comparison between different soft MIMO detectors

As reported in [96], the area below the transfer function $I_{\mathrm{E,DET}} = F(I_{\mathrm{A,DET}})$ well approximates the maximum achievable rate of the outer LDPC encoder. Hence, the extrinsic information transfer characteristic $F(\cdot)$ can be easily used to compare and evaluate the performance of different soft MIMO detectors. We demonstrate in Fig. 7.10 the extrinsic information transfer functions of different MIMO detectors described in Section 7.3 under the same 2×2 unknown MIMO channel with coherence time $T = 6$, training length $T_\tau = 2$, and system SNR $\rho = 4dB$. For comparison purpose again, the mutual information transfer characteristics of the simple MMSE-based MIMO detector and detector with ideal CSIR

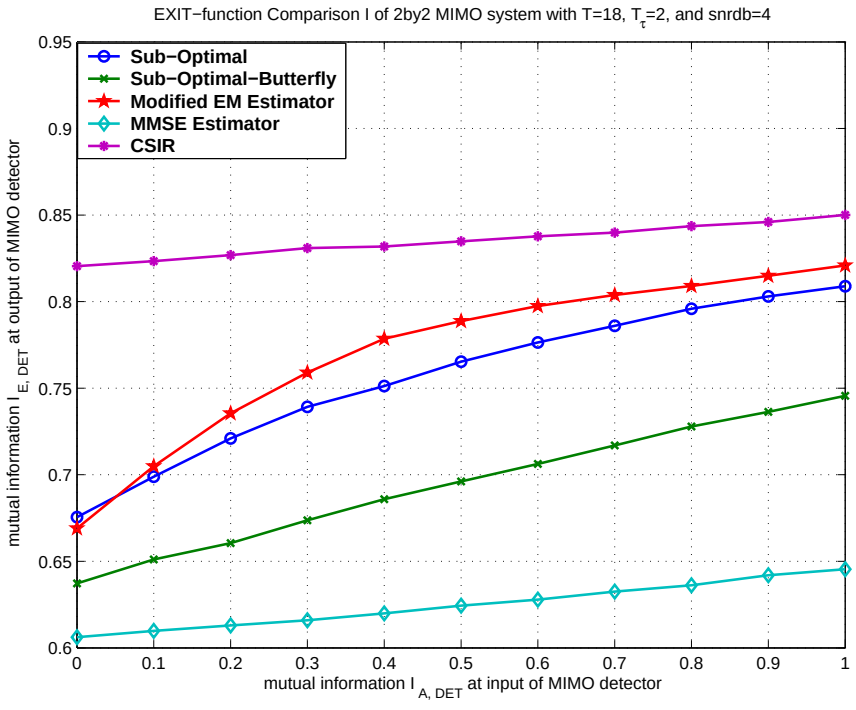

Figure 7.11: Comparison of the extrinsic information transfer characteristic of different MIMO detectors in a 2×2 unknown MIMO system with coherence time $T = 18$, training length $T_\tau = 2$, and signal to noise ratio $\rho = 4dB$.

are also included. As can be observed from the plot, all four soft MIMO detectors have comparable performance in a small coherence time T channel environment, i.e. $T \leq 10$. All of them achieve significant performance gain over the simple MMSE-based detector but are far away from the MIMO detector with ideal CSIR.

Furthermore, although the optimal soft MIMO detector is the best among all MIMO detectors under the same channel condition, it is not always affordable for practical communications systems due to its complexity especially when the coherence time T is large. Therefore, sub-optimal soft MIMO detectors as well as the EM-based detectors turn out to be promising alternatives for their excellent trade-offs between complexity and performance over moderate to slow fading channels. As illustrated in Fig. 7.11, the extrinsic information transfer functions

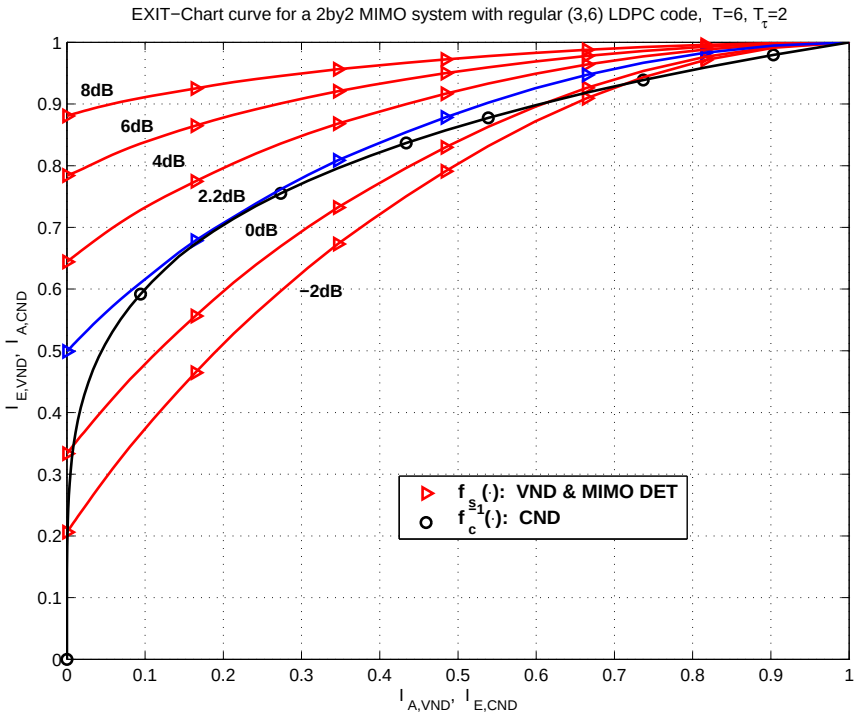

Figure 7.12: EXIT-chart curve of a 2×2 regular $(3,6)$ LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$, and training number $T_\tau = 2$ using optimal soft MIMO detectors, under different signal to noise ratios $\rho = -2, 0, 2.2, 4, 6, 8\ dB$.

of these sub-optimal MIMO detectors are compared over the same 2×2 unknown MIMO channel with large coherence time $T = 18$. In this case (with large T), the modified EM-based MIMO detector outperforms other sub-optimal detectors and tends to approach the performance of the MIMO detector with ideal CSIR.

7.5.3 LDPC code degree profile optimization

The analysis of the mutual information transfer characteristic provided in Section 7.4.2 not only enables us to analyze the system performance, but also provides a powerful design approach for the LDPC code optimization. We demonstrate in Fig. 7.12 the EXIT-chart curves of a 2×2 regular $(3, 6)$ LDPC-coded

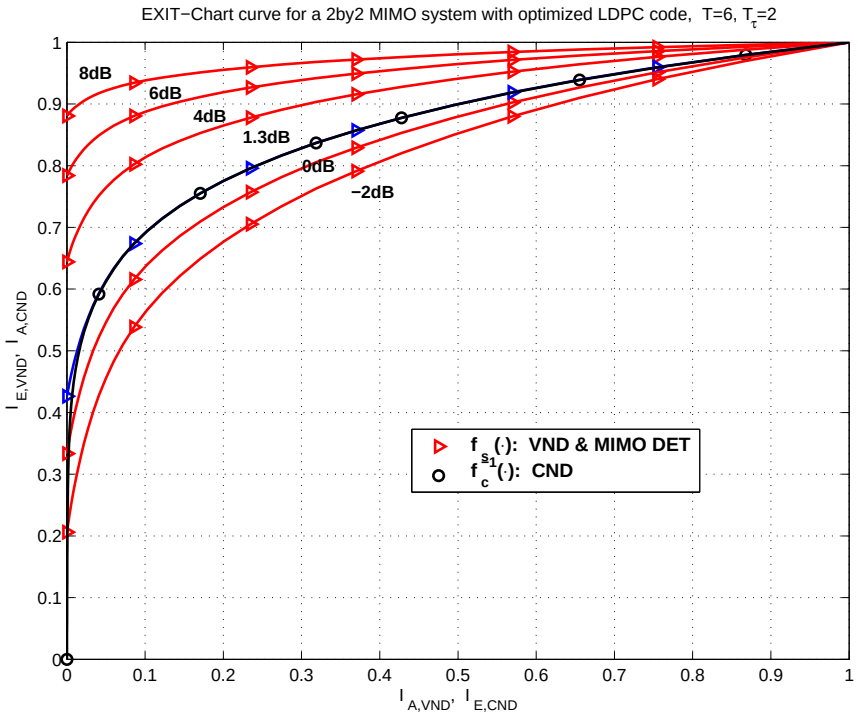

Figure 7.13: EXIT-chart curve of a 2×2 optimized LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$, and training number $T_\tau = 2$ using optimal soft MIMO detectors, under different signal to noise ratios $\rho = -2, 0, 1.3, 4, 6, 8\ dB$.

MIMO system with codeword length 8×10^4. The simulation is carried out over an unknown fading channel with coherence time $T = 6$ and training number $T_\tau = 2$, using optimal soft MIMO detectors, and under several different system SNRs. From the plot, we can observe that $2.2dB$ is the minimal SNR that can avoid curve intersection and hence leads to successful decoding, which is further confirmed by the cliff region shown in the real simulation result of Fig. 7.14. We also illustrate in Fig. 7.13 the EXIT-chart curves for the optimized LDPC-coded MIMO system with outer code rate $R_{\text{outer}} = 1/2$ and codeword length 8×10^4 under the same system settings. It can be observed from the plot that after applying LDPC code optimization, the two mutual information transfer functions match

each other perfectly (almost fall on top of each other), and achieve about $0.9dB$ gain in performance.

7.5.4 Overall coded MIMO system performance

We demonstrate in Fig. 7.14 the bit error rate of an LDPC-coded MIMO system over unknown fading channels. For the sake of simulation simplicity, we consider a small 2×2 MIMO system over a relatively fast unknown fading channel with coherence time $T = 6$. According to the non-coherent MIMO capacity analysis provided in [71] [92], the number of training symbols T_τ is set equal to the number of transmit antennas M, in a sense to maximizes the system capacity (or mutual information rate). The outer LDPC code is a regular $(3, 6)$ code with code rate $R_{\text{outer}} = 1/2$, and codeword length 8×10^4. By taking into account the pilots cost, the overall system coding rate is $R_{\text{overall}} = 1/3$ bits per transmission. We can observe from Fig. 7.14 that over 1.5dB performance gain can be achieved by using optimal soft MIMO detectors than the simple MMSE-based detector. The two sub-optimal MIMO detectors as well as the modified EM-based soft MIMO detector also provide significant performance gain, and at the same time maintain affordable decoding complexity. On the other hand, due to the existing positive feedback, the direct EM-based MIMO detector has a 2dB performance degradation compared to the modified EM-based detector and performs even worse than the simple MMSE-based detector.

Using the optimization algorithm provided in Section 7.4.3, the optimal LDPC code degree profiles (with outer code rate $R_{\text{outer}} = 1/2$ and codeword length 8×10^4) for the coded MIMO system using the different soft MIMO detection algorithms are obtained and used in the overall performance simulation. We consider the same 2×2 coded MIMO system used in Fig. 7.14 that transmits over the same unknown fading channel with coherence time $T = 6$ and pilot number $T_\tau = 2$ for simulations. The probability of bit error of the LDPC-coded MIMO system with optimized LDPC code degree profile is shown in Fig. 7.15. Compared with

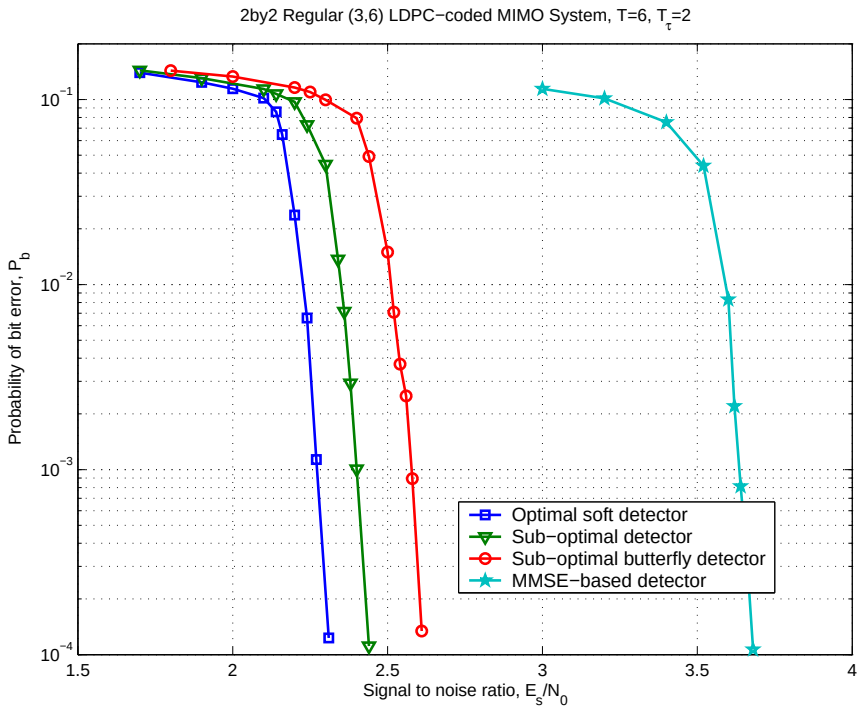

Figure 7.14: Probability of bit error of a 2×2 regular $(3, 6)$ LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$ and training number $T_\tau = 2$ using several different soft MIMO detectors.

Fig. 7.14, we can achieve about 0.6dB performance gain by using the optimized LDPC degree profile as opposed to the simple regular $(3, 6)$ LDPC code. Additional simulation results, not shown here, indicate that an even more significant performance gain can be achieved by the proposed LDPC code optimization approach if higher modulation format (such as QPSK or 16-QAM) is used, or if the coherence time interval T is larger. Under these channel conditions, the extrinsic information transfer functions (7.54) and (7.55) of a regular LDPC-coded MIMO system are very dissimilar to each other and this emphasizes the importance of the proposed curve fitting technique.

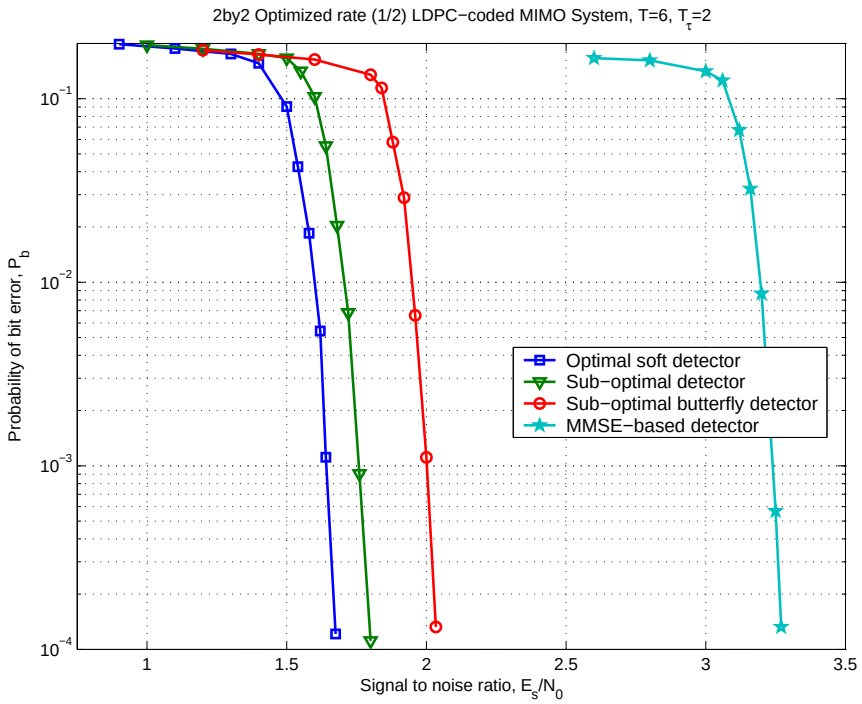

Figure 7.15: Probability of bit error of a 2×2 optimized LDPC-coded MIMO system over a unknown fading channel with coherence time $T = 6$ and training number $T_\tau = 2$ using several different soft MIMO detectors.

7.6 Summary

In this chapter, we developed a practical LDPC-coded MIMO system over a flat fading wireless environment with channel state information unavailable both at the transmitter and the receiver. We first proposed several soft-input soft-output MIMO detectors, including one optimal soft MIMO detector, two sub-optimal soft detectors, and a modified EM-based MIMO detector, whose performances are much better than the conventional MMSE-based detectors and offer an effective tradeoff between complexity and performance. By analyzing the extrinsic information transfer characteristic of the soft MIMO detectors, performance of the coded MIMO system using different MIMO detection algorithms are analyzed and compared under various channel conditions. Motivated by the turbo iterative

principle, the LDPC-coded MIMO receiver is constructed in an unconventional manner where the soft MIMO detector and LDPC variable node decoder form one super soft-decoding unit, and the LDPC check node decoder forms the other component of the iterative decoding scheme. The proposed receiver structure has lower decoding complexity and further leads to tractable EXIT functions of the component soft decoders. Based on the obtained closed form EXIT functions, a simple and efficient LDPC code degree profile optimization algorithm is developed with proven global optimality and guaranteed convergence from any initialization. Finally, numerical and simulation results of the LDPC-coded MIMO system using the optimized degree profile further confirm the advantage of using the proposed design approach for the coded MIMO system. The text of this chapter is in part a reprint of the material which was coauthored with Bhaskar D. Rao and has been published in *IEEE Transactions on Signal Processing* under the title *"LDPC-coded MIMO systems with unknown block fading channels: soft MIMO detector design, channel estimation, and code optimization"*.

8 Conclusions and Future Work

Using multiple antennas at both the transmitter and the receiver is one of the most promising techniques that can offer significant increases in channel capacity of a communication system in a wireless fading environment. However, the performance of the MIMO system depends heavily upon the availability of the channel state information (CSI) at the transmitter (CSIT) and at the receiver (CSIR). In this dissertation, we focus our attention on design and analysis of MIMO systems over wireless fading channels with practical CSI assumptions. The contributions of this dissertation can be broadly classified into the following two parts.

8.1 Analysis of MIMO Systems with Finite-Rate Feedback

The first part, which includes Chap. 2 - Chap. 5, considers the development of a general framework for the analysis of multiple antenna systems with finite-rate feedback, in the sense that the CSI is quantized at the receiver and conveyed back to the transmitter through a rate-constrained reverse link.

By connecting the channel quantization problem to classical high resolution quantization theory, Chap. 2 developed a general framework for the analysis of multiple antenna systems with finite-rate CSI feedback. The main contributions of this chapter are listed below.

- The problem of finite-rate quantized communication system was formulated as a general fixed-rate vector quantization problem with encoder side infor-

mation, non-mean square distortion functions, and constrained source variables.

- Tight lower and upper bounds of the average asymptotic (high quantization rate) distortion of the proposed general vector quantization problem as well as the sufficient conditions for the achievability of the distortion bounds were provided.

- The proposed distortion analysis was extended to the important problem of sub-optimal quantizers with mismatched distortion functions, source statistics, and quantization criteria. Bounds on the average distortion of these different mismatched quantizers were provided.

- The framework was further extended to provide analysis for a generalized vector quantizer with transformed codebook. Bounds on the average system distortion of this class of quantizers were also provided.

- Finally, asymptotic distortion analysis of complex source variables as well as source variables with constrained parameterizations was also provided.

The proposed general methodology provides a powerful analytical tool to study a wide range of finite-rate feedback systems. Chap. 3 provides a detailed capacity analysis of MISO systems with finite rate CSI feedback over both i.i.d and spatially correlated fading channels, and the main contributions are listed below.

- As an extended application of the general distortion analysis, tight lower bounds on the capacity loss of both spatially i.i.d. and correlated MISO systems due to the finite-rate channel quantization were provided.

- In high-SNR and low-SNR regimes, analytically closed form expressions of the MISO system capacity loss were provided.

- The capacity loss of correlated MISO channels was shown to be related to that of i.i.d. fading channels by a simple multiplicative factor which is given

by the ratio of the geometric mean to the arithmetic mean of the eigenvalues of the channel covariance matrix.

The general framework developed in Chap. 2 is versatile enough and enables the analysis of sub-optimal MISO CSI quantizers with mismatched codebooks and quantizers with transformed codebooks, which were provided in Chap. 4. The main contributions of this chapter are listed below.

- Two types of mismatched MISO CSI quantizers were investigated: quantizers whose codebooks are designed with MMSE criterion but the distortion measure is the ergodic capacity loss (i.e. mismatched design criterion), and quantizers with codebook designed with a mismatched channel covariance matrix (i.e. mismatched statistics).

- Bounds on the channel capacity loss of the mismatched codebooks were provided and compared to that of the optimal quantizers.

- Upper and lower bounds on the capacity loss of MISO systems transmitting over spatially correlated fading channels but using CSI quantizers whose codebook is transformed from spatially i.i.d. fading channels were also provided.

- It was further proved that the average distortion of CSI quantizers with transformed codebooks can be upper and lower bounded by some multiplicative factors of the distortion of optimal quantizers. This factor was shown to be close to one for fading channels whose channel covariance matrix has small to moderate condition numbers.

In Chap. 5, the capacity analysis was further extended to MIMO systems with finite rate CSI feedback. The main contributions are listed below.

- Tight lower bounds on the capacity loss of MIMO systems with finite-rate CSI feedback transmitting over spatially i.i.d. Rayleigh flat fading channels were provided.

- MIMO CSI-quantizers with mismatched codebooks that only optimized for high-SNR and low-SNR regimes were investigated.
- Capacity analysis of MIMO systems using multi-mode spatial multiplexing transmission schemes with finite-rate CSI feedback were also provided.

8.2 Design and Analysis of MIMO Systems with Unknown CSI

The second part of this dissertation, which includes Chap. 6 and Chap. 7, is focused on the design and analysis of MIMO systems over fading channels with CSI unavailable both at the transmitter and at the receiver.

To be specific, a capacity analysis of MIMO systems with unknown CSI assumption was provided in Chap. 6. The main contributions of this chapter are listed below.

- We proposed a system mutual information upper bound, which is shown to be tight when the number of transmit antennas M is moderately large or the system SNR is small, thereby leading to a valid capacity lower bound of the unknown MIMO channel.
- The obtained analysis result is well suited for predicting capacities of systems utilizing joint channel estimation and data detection algorithms with iterative decoding structures.
- By analyzing the proposed capacity lower bound with respect to different system parameters, we reenforce the advantages of using an orthogonal pilot structure, which not only minimizes the mean square estimation error, but also maximizes the proposed mutual information upper bound.
- It was shown that the mutual information upper bound is a monotonically decreasing function with respect to the number of pilot symbols T_τ. Numerical evaluation of the upper bound further demonstrated the fact that

the rate gain is insignificant when the number of pilot symbols T_τ decreases below M, suggesting a training duration of M time slots provides excellent trade-off between complexity and performance.

- With the setting $T_\tau = M$, it was also shown that there is no benefit in making the number of transmit antennas M greater than N.

- Further numerical results demonstrated that only limited rate gain can be achieved by using optimal power allocation between training and data symbols compared to the simple equal power allocation scheme.

Based on the capacity analysis results, design of practical LDPC-coded MIMO systems under the same unknown CSI assumption was provided in Chap. 7, and the main contributions are listed below.

- We first proposed several soft-input soft-output MIMO detectors, including one optimal soft MIMO detector, two sub-optimal soft detectors, and a modified EM-based MIMO detector, whose performances are much better than the conventional MMSE-based detectors and offer an effective tradeoff between complexity and performance.

- Motivated by the turbo iterative principle, the LDPC-coded MIMO receiver was constructed in an unconventional manner where the soft MIMO detector and LDPC variable node decoder form one super soft-decoding unit, and the LDPC check node decoder forms the other component of the iterative decoding scheme.

- The proposed receiver structure has lower decoding complexity and further leads to tractable EXIT functions of the component soft decoders.

- Based on the obtained closed form EXIT functions, a simple and efficient LDPC code degree profile optimization algorithm was developed with proven global optimality and guaranteed convergence from any initialization.

8.3 Future Work

In this section, I summarize some of the possible extensions of this dissertation. Most of the extensions focus on continuing the problem of design and analysis of multiple antenna systems with finite-rate CSI feedback.

8.3.1 CSI Quantization with Practical Assumptions

As part of my dissertation, I have looked at the problem of design and analysis of MIMO systems with finite-rate CSI feedback. However, for the sake of simplicity, we adopted some ideal assumptions which include: the feedback channel is delay-less and error-free, perfect CSIR is available at the receiver without channel estimation error, and the fading channel keeps static within the coherent block (block-fading model). Therefore, for practical communication systems, several open problems remain in this area, which are discussed in the following three directions.

Feedback Channel Error

In practice, the reverse link is not perfect and subject to noise. Hence, the partial CSI at the transmitter is distorted by two types of errors: quantization error and feedback error. Consequently, the CSI quantization schemes need to be re-derived in this case to minimize the degradation caused by these errors.

Channel Estimation Error

The CSI at the receiver is always obtained by some form of channel estimations. Inevitably, it is subject to estimation error in the sense that CSIR is not perfect. Hence, quantizing an imperfect CSI (distorted source variable) is also a very challenging problem and requires further investigation.

Time-Varying Fading Channels

It is well known that block-fading is a simplified channel model for systems in a relatively slow fading environment. For fast fading channels, the actual channel might already change into a very different state before the quantized CSI arrive at the transmitter. In these situations, it is more important to track the channel rather than accurately quantize it. Some tentative solutions include: quantizing the CSI difference between two consecutive states, quantizing the actual physical parameters of the time-varying fading channel (such as angle to arrival, angular spread, cluster number, and etc.), and utilizing proper channel prediction techniques.

8.3.2 Practical Quantizer Design for CSI Feedback

Most of the CSI quantization schemes proposed so far are only limited to VQ-based techniques, where brute-force searching algorithm is assumed in the channel quantizer. To be specific, the receiver (or the channel quantizer) forms quantization index by simply comparing the performance of every possible beamforming vectors or pre-coding matrices in the codebook. These kind of channel quantizers work fine for MIMO systems with small number of antennas. However, for moderate to large systems or MIMO multicarrier systems over frequency-selective fading channels, the CSI to be quantized has a large number of free dimensions. Take for example, for an MISO system with 5 transmit antennas, the CSI is a complex vector of size 5×1. If we quantize each of the real CSI dimensions by 2 bits (with total 10 real dimensions described by 20 bits), the entire CSI codebook consists of 10^6 beamforming vectors, which leads to 10^6 comparisons in the brute-force search encoding algorithm. Therefore, design of practical CSI quantizers by utilizing the rich source coding techniques, such as various product quantizers, structured quantizers, sphere and ellipse wrapped codes and etc., would be very beneficial. It is very interesting in the sense that not only the communication problem finds promising solutions but also the source coding theory extends

its applications.

8.3.3 Precoder Design for Multiuser MIMO with Partial CSIT

Consider a MIMO downlink broadcast channel in a wireless fading environment. The CSI at the base-station is usually assumed perfect available in most of the current literatures. However, for an FDD system without channel reciprocity, every user has to feedback its own channel state to the base station before any precoding or scheduling algorithm take place in the transmitter side. Scaled by the multiuser number, usually in an order of hundreds, the amount of CSI information need to be fedback increases dramatically in the multiuser environment. Hence, it is an urgent problem to design multiuser MIMO transmit precoders that can efficiently use the partial CSI as well as design user-side CSI feedback schemes with more strict rate-constraint. Moreover, multiuser scheduling algorithms with finite-rate CSI feedback and other cross layer design problems are also worth investigating.

Bibliography

[1] J. Winters, "On the capacity of radio communication systems with diversity in a rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 871–878, June 1987.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar 1998.

[4] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Bell Labs Technical Journal*, June 1995.

[5] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europian Trans. Telecomm. (ETT)*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, Jan. 1999.

[7] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.

[8] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. on Information Theory*, vol. 47, pp. 1273–1289, May 2001.

[9] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *IEEE International Symposium on Communications 2001*, June 2001, pp. 2266–2270.

[10] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Information Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[11] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. on Wireless Communications*, vol. 3, no. 5, pp. 1626–1636, Sept. 2004.

[12] E. A. Jorsweick and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. on Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[13] M. Skoglund and G. Jongren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 395–405, Apr. 2003.

[14] K. N. Lau, Y. Liu, and T. A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. on Communications*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

[15] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1423–1436, Oct. 1998.

[16] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Massachusetts, 1992.

[17] K. K. Mukkavilli, A. Sabharwal, E.Erkip, and B.Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[18] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Information Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[19] D. J. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. on Signal Processing*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[20] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian space," *Experimental Math.*, , no. 5, pp. 139–159, 1996.

[21] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. on Communications*, vol. 51, no. 7, pp. 1102–1110, July 2003.

[22] P. Xia, S. Zhou, and G. B. Giannakis, "Multiantenna adaptive modulation with beamforming based on bandwidth- constrained feedback," *IEEE Trans. on Communications*, vol. 53, no. 3, pp. 526–536, Mar. 2005.

[23] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. on Signal Processing*, 2006 (to appear).

[24] J. Roh and B. D. Rao, "Performance analysis of multiple antenna systems with VQ-based feedback," in *38th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2004, pp. 1978–1982.

[25] J. Roh and B. D. Rao, "Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. on Information Theory*, 2006 (to appear).

[26] J. Roh and B. D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. on Signal Processing*, 2006 (to appear).

[27] J. Roh, *Multiple-Antenna Communication with Finite Rate Feedback*, Ph.D. thesis, Univ. of California, San Diego, 2005.

[28] D. J. Love and R. W. Heath Jr., "Multi-mode precoding for MIMO wireless systems," *IEEE Trans. on Signal Processing*, 2005 (to appear).

[29] C. R. Murthy and B. D. Rao, "A vector quantization based approach for equal gain transmission," in *Proc. IEEE Globecom 2005*, St. Louis, MO, Nov. 2005, pp. 2528–2533.

[30] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Trans. on Information Theory*, vol. 46, no. 7, pp. 2697–2704, Nov. 2000.

[31] E. Martinian, G. W. Wornell, and R. Zamir, "Source coding with distortion side information at the encoder," in *IEEE Data Compression Conference, 2004. Proceedings. DCC 2004*, Snowbird, Utah, Mar. 2004, pp. 172–181.

[32] W. R. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, vol. 27, pp. 446–472, July 1948.

[33] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. on Information Theory*, vol. 25, pp. 373–380, July 1979.

[34] W. R. Gardner and B. D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 367–381, Sept. 1995.

[35] S. Na and D. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. on Information Theory*, vol. 41, no. 4, pp. 886–900, July 1995.

[36] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. on Communications*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

[37] R. M. Gray, *Source Coding Theory*, Norwell, MA: Kluwer, 1990.

[38] W. R. Gardner, *Modeling and Quantization Techniques for Speech Compression Systems*, Ph.D. thesis, University of California at San Diego, La Jolla, 1994.

[39] C. Helstrom, *Probability and Stochastic Processes for Engineers*, New York, Macmillan, 1984.

[40] J. Conway and N. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. on Information Theory*, vol. 28, pp. 211–226, Mar. 1982.

[41] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. on Information Theory*, vol. 28, pp. 129–137, Mar. 1982.

[42] J. A. Bucklew, "Companding and random quantization in several dimensions," *IEEE Trans. on Information Theory*, vol. 27, no. 2, pp. 207–211, Mar. 1981.

[43] P. L. Zador, "Asymptotic quantization error of continuous signals and quantization dimension," *IEEE Trans. on Information Theory*, vol. 28, no. 2, pp. 139–148, Mar. 1982.

[44] J. Zheng, E. R. Duni, and B. D. Rao, "Analysis of multiple antenna systems with finite-rate feedback using high resolution quantization theory," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2006, pp. 73–82.

[45] J. Zheng, E. R. Duni, and B. D. Rao, "Analysis of multiple antenna systems with finite rate feedback using high resolution quantization theory," *submitted to IEEE Trans. on Signal Processing*, Nov. 2005.

[46] J. Huang and P. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. on Communications*, vol. 11, pp. 289–296, Sept. 1963.

[47] R. Remmert, *Theory of Complex Functions*, Springer-Verlag, New York, 1991.

[48] K. Kreutz-Delgado, "Lecture supplement on complex vector calculus," *[Online] Available http://dsp.ucsd.edu/∼kreutz /PEI05.html*, Apr. 2006.

[49] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[50] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, pp. 1049–1057, Nov. 1994.

[51] M. C. Jones, "On moments of ratios of quadratic forms in normal variables," *Statistics & Probability Letters*, vol. 6, pp. 129–136, Nov. 1987.

[52] D. J. Love and R. W. Heath Jr., "Grassmannian beamforming on correlated MIMO channels," in *IEEE Globecom 2004*, Dallas, TX, Dec. 2004, vol. 1, pp. 106–110.

[53] J. Zheng and B. D. Rao, "Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels," *IEEE Trans. on Signal Processing*, in preparation.

[54] A. Edelman, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[55] P. A. Dighe, R. K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. on Communications*, vol. 51, pp. 694–703, Apr. 2003.

[56] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. on Information Theory*, vol. 44, pp. 744–765, Mar 1998.

[57] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 47, pp. 1290–1301, Mar 2001.

[58] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2426–2467, Oct 2003.

[59] R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR Rayleigh fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 3390–3396, Dec 2004.

[60] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 3095–3110, Dec 2004.

[61] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1169–1174, July 2000.

[62] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. on Information Theory*, vol. 46, pp. 2567–2578, Nov 2000.

[63] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. on Communications*, vol. 48, pp. 2041–2052, Dec 2000.

[64] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 46, pp. 543–564, Mar 2000.

[65] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. on Information Theory*, vol. 46, pp. 1962–1973, Sept 2000.

[66] D. Agrawal, T. J. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. on Information Theory*, vol. 47, pp. 2618–2626, Sept 2001.

[67] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for noncoherent space-time communication on the block-Rayleigh-fading channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 1186–1194, May 2002.

[68] V. Tarokh and I.-M. Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. on Information Theory*, vol. 48, pp. 3112–3117, Dec 2002.

[69] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of noncoherent space-time constellations," *IEEE Trans. on Information Theory*, vol. 49, pp. 2332–2351, Oct 2003.

[70] W. Zhao, G. Leus, and G. B. Giannakis, "Algebraic design of unitary space-time constellations," in *IEEE International Conference on Communications 2003, Anchorage, AK*, May 2003, pp. 3180–3184.

[71] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.

[72] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. on Wireless Communications*, 2005 (to appear).

[73] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*, Chapman & Hall/CRC, Boca Raton, Fla., 2000.

[74] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, corrected and enlarged ed., New York, 1980.

[75] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, 1964.

[76] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, pp. 2619–2692, Oct 1998.

[77] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52, pp. 546–557, 2004.

[78] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. on Signal Processing*, vol. 51, pp. 2882–2890, 2003.

[79] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras, "Blind MIMO FIR channel identification based on second-order spectra correlations," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1668–1674, 2003.

[80] H. Sahlin and H. Broman, "MIMO signal separation for FIR channels: a criterion and performance analysis," *IEEE Trans. on Signal Processing*, vol. 48, pp. 642–649, 2000.

[81] S. Amari and J. F. Cardoso, "Blind source separation-semiparametric statistical approach," *IEEE Trans. on Signal Processing*, vol. 45, pp. 2692–2700, 1997.

[82] C. Jutten and J. Heuralt, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[83] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. on Communications*, vol. 45, pp. 300–308, 1997.

[84] G. Kaleh, "Joint parameter estimation and symbol detection for linear and nonlinear unknown channels," *IEEE Trans. on Communications*, vol. 42, pp. 2406–2413, July 1994.

[85] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. on Communications*, vol. 51, pp. 1266–1270, Aug. 2003.

[86] J. J. Boutros, F. Boixadera, and C. Lamy, "Bit-interleaved coded modulations for multiple-input multiple-output channels," in *IEEE International Symposium on Spread Spectrum Techniques and Applications*, Sept. 2000, vol. 1, pp. 123–126.

[87] M. Gonzalez-Lopez, J. Miguez, and L. Castedo, "Turbo aided maximum likelihood channel estimation for space-time coded systems," in *IEEE International Symposium on PIMRC 2002*, Sept. 2002, pp. 364–368.

[88] H. Wymeersch, F. Simoens, and M. Moeneclaey, "Code-aided joint channel estimation and frame synchronization for MIMO systems," in *Workshop on Signal Processing Advances in Wireless Communications (SPAWC'04)*, July 2004.

[89] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Communications*, vol. 49, pp. 1727–1737, Oct. 2001.

[90] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *IEEE International Symposium on Turbo Codes and Related Topics*, Sept. 1997.

[91] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. on Communications*, vol. 52, pp. 670–678, Apr. 2004.

[92] J. Zheng and B. D.Rao, "Capacity analysis of MIMO systems with unknown channel state information," in *IEEE Information Theory Workshop 2004*, San Antonio, Oct. 2004.

[93] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. on Information Theory*, vol. 27, pp. 533–547, Sept. 1981.

[94] L. Wei, "Several properties of short LDPC codes," *IEEE Trans. on Communications*, vol. 52, pp. 721–727, May 2004.

[95] Y. J. Ko and J. H. Kim, "Girth conditioning for construction of short block length irregular LDPC codes," *Electronics Letters*, vol. 40, pp. 187–188, Feb 2004.

[96] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *to appear in IEEE Trans. on Information Theory*, 2004.