SIGNAL DESIGN FOR MIMO–OFDM SYSTEMS

by

Harry Zhi Bing Chen

M.A.Sc., The University of British Columbia, 2004
M.E., University of Science and Technology of China, 2000
B.E., Huazhong University of Science and Technology, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

January 2010

© Harry Zhi Bing Chen, 2010
Abstract

Orthogonal frequency division multiplexing (OFDM) combined with multiple-input multiple-output (MIMO) wireless technology is an attractive air-interface solution for wireless systems. The time-selective and dispersive nature of the wireless channel, however, poses challenges for signal design. To address these challenges, this dissertation investigates several physical layer aspects of signal design for MIMO–OFDM systems.

To improve the information outage rate of certain MIMO systems with low-rank channel matrices, we propose two novel channel augmentation schemes, namely, the complex virtual antenna (CVA) and the real virtual antenna (RVA) schemes. For Bell labs space–time (BLAST) type architectures, both the CVA and the RVA schemes are more robust to channel correlations than a previously proposed scheme.

For performance enhancement of coded MIMO–OFDM, we consider wrapped space-frequency coding (WSFC) and coded vertical BLAST (V–BLAST) with adaptive bit loading (ABL) and optimize both schemes. We show that bit–loaded WSFC and V–BLAST optimized for coded MIMO–OFDM can achieve error rate performances close to that of quasi–optimal MIMO–OFDM based on single value decomposition (SVD) of the channel, while their feedback requirements for loading are low.

To exploit the spatial and frequency diversity in coded MIMO–OFDM systems, we propose cyclic space–frequency (CSF) filtering, which is applicable to both traditional MIMO systems with co–located transmit antennas and cooperative diversity systems
with distributed transmit antennas. We further propose a robust CSF filtering scheme, which exploits imperfect channel state information at the transmitter (CSIT) and takes into account the reliability of the CSIT. To further improve robustness, we combine CSF filtering with space–time block coding (STBC) in the frequency domain. We also propose a linear prediction method for improving the quality of the CSIT via post–processing. For these CSF filtering schemes, design criteria are derived, closed–form solutions are given for certain special cases, and various design methods are provided for the general case. Simulation results confirm that (robust) CSF filtering is a promising solution for wireless communication systems.
# Contents

Abstract ......................................................... ii

Contents ....................................................... iv

List of Tables ................................................... ix

List of Figures .................................................... x

List of Abbreviations ........................................... xv

Notation ......................................................... xix

Acknowledgments ............................................... xxi

1 Introduction ................................................... 1
   1.1 MIMO Wireless Communications .............................. 2
   1.2 OFDM Modulation ........................................... 4
   1.3 MIMO–OFDM Systems ........................................ 6
   1.4 Thesis Contributions and Organization ..................... 8

2 Preliminaries .................................................. 11
   2.1 MIMO Channel Model ....................................... 11
2.1.1 Signal Envelope ............................................. 12
2.1.2 Characteristics of Multipath Fading ....................... 13
2.1.3 MIMO Channel Parameters ................................ 15
2.2 MIMO Techniques .............................................. 17
  2.2.1 Open–Loop Technique I – Layered Space–Time Processing . 18
  2.2.2 Open–Loop Technique II – STC .......................... 21
  2.2.3 Closed–Loop Technique I – Spatial Multiplexing via SVD . 25
  2.2.4 Closed–Loop Technique II – Combining Precoding with STBC 27
2.3 Virtual MIMO Techniques ..................................... 31
  2.3.1 Cooperative Diversity .................................... 31
  2.3.2 Channel Augmentation ................................. 32
2.4 OFDM Basics .................................................. 34
  2.4.1 Orthogonality of Subcarriers .......................... 34
  2.4.2 Guard Interval ........................................ 35
  2.4.3 BICM–OFDM ............................................ 36
  2.4.4 Adaptive Loading ...................................... 38
2.5 MIMO–OFDM .................................................. 40
  2.5.1 CDD .................................................. 40

3 Two Novel Channel Augmentation Schemes for MIMO Systems . 44
  3.1 Introduction .................................................. 44
  3.2 System Model ............................................... 46
  3.3 Channel Augmentation Schemes ............................ 47
    3.3.1 Rankin, Taylor, and Martin (RTM) Scheme ............. 47
    3.3.2 Conjugate Virtual Antenna (CVA) Scheme ............. 48
    3.3.3 Real Virtual Antenna (RVA) Scheme .................. 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Information Outage Rate</td>
<td>50</td>
</tr>
<tr>
<td>3.4.1</td>
<td>V–BLAST</td>
<td>51</td>
</tr>
<tr>
<td>3.4.2</td>
<td>D–BLAST</td>
<td>52</td>
</tr>
<tr>
<td>3.5</td>
<td>Numerical Results</td>
<td>53</td>
</tr>
<tr>
<td>3.6</td>
<td>Conclusions</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>On MIMO–OFDM with Coding and Loading</td>
<td>58</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>58</td>
</tr>
<tr>
<td>4.2</td>
<td>System Model</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>MIMO Processing for Coded MIMO–OFDM</td>
<td>61</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Wrapped Space-Frequency Coding (WSFC)</td>
<td>61</td>
</tr>
<tr>
<td>4.3.2</td>
<td>V–BLAST</td>
<td>63</td>
</tr>
<tr>
<td>4.3.3</td>
<td>SVD</td>
<td>64</td>
</tr>
<tr>
<td>4.4</td>
<td>Adaptive Bit-Loading (ABL) Schemes</td>
<td>65</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Full Loading (FL)</td>
<td>65</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Grouped Loading (GL)</td>
<td>65</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Modification of Loading for V–BLAST</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>Results and Discussion</td>
<td>67</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Optimization of WSFC and V–BLAST for MIMO–OFDM with ABL</td>
<td>68</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Performance Comparisons</td>
<td>69</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusions</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>Cyclic Space–Frequency Filtering for BICM–OFDM Systems</td>
<td>73</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>73</td>
</tr>
<tr>
<td>5.2</td>
<td>System Model</td>
<td>75</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Network Model</td>
<td>76</td>
</tr>
</tbody>
</table>
5.2.2 Transmitter ................................................. 76
5.2.3 Channel Model ............................................. 79
5.2.4 Receiver Processing ....................................... 80
5.2.5 Extension to OFDMA ....................................... 81
5.3 CSF Filter Optimization for Known $S$ .................. 82
  5.3.1 Optimization Criterion ............................... 83
  5.3.2 Upper Bound on Cost Function $f(g)$ .......... 86
  5.3.3 Closed-Form Solution for Special Cases .......... 86
  5.3.4 CSF Filter Design for the General Case .......... 88
5.4 CSF Filter Optimization for Unknown $S$ ............. 90
  5.4.1 Optimization Problem ............................... 92
  5.4.2 Upper Bound and Closed-Form Solution for Special Case 93
  5.4.3 CSF Filter Design for General Case ............ 94
5.5 Simulation Results .......................................... 96
  5.5.1 Co-located Antennas ............................... 96
  5.5.2 Distributed Antennas ............................... 101
5.6 Conclusions ................................................. 105

6 Robust Transmit Processing for BICM–OFDM Systems with Imperfect CSIT ................................................. 107
  6.1 Introduction .............................................. 107
  6.2 System Model ............................................ 109
    6.2.1 Channel Model ..................................... 110
    6.2.2 Transceiver Structure for Robust CSF .......... 110
    6.2.3 Transceiver Structure for Robust SFC–CSF .......... 112
    6.2.4 Bayesian Statistics for Imperfect CSIT .......... 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3</td>
<td>Optimization of the Robust CSF Filters</td>
<td>118</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Optimization Problem</td>
<td>119</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Solution for Unstructured $G$</td>
<td>121</td>
</tr>
<tr>
<td>6.4</td>
<td>Design of Robust CSF Filters</td>
<td>123</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Special Cases</td>
<td>123</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Solution of Problem with Additional Constraints</td>
<td>125</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Gradient Algorithm (GA)</td>
<td>127</td>
</tr>
<tr>
<td>6.5</td>
<td>Simulation Results</td>
<td>128</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Robust CSF Filtering</td>
<td>129</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Robust SFC–CSF Filtering</td>
<td>132</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Linear Prediction (LP)</td>
<td>134</td>
</tr>
<tr>
<td>6.6</td>
<td>Conclusions</td>
<td>137</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions and Future Work</td>
<td>138</td>
</tr>
<tr>
<td>7.1</td>
<td>Research Contributions</td>
<td>138</td>
</tr>
<tr>
<td>7.2</td>
<td>Future Work</td>
<td>140</td>
</tr>
<tr>
<td>A</td>
<td>Related Publications</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>144</td>
</tr>
</tbody>
</table>
5.1 Gradient algorithms (GAs) for calculation of the optimum CSF filters for known and unknown $S$, respectively. The elements of the gradient vectors $\partial f(g)/\partial g^*$ and $\partial f_{2,k}(g_S)/\partial g_S^*$ are given in (5.22) and (5.30), respectively. The adaptation constant $\mu[i]$ has to be optimized experimentally. The termination constant $\epsilon$ has a small value (e.g. $\epsilon = 10^{-6}$). $i$ denotes the iteration.
# List of Figures

2.1 V–BLAST structure. .......................................................... 19
2.2 D–BLAST structure. .......................................................... 20
2.3 D–BLAST format. ............................................................... 20
2.4 WSTC format. ................................................................. 21
2.5 System structure combining beamforming with STBC for a system with
    imperfect CSIT. ............................................................... 27
2.6 (a) Concept of CP guard interval and (b) OFDM symbol time with CP. 36
2.7 Block diagram of BICM–OFDM. .......................................... 37
2.8 Block diagram of CDD. ...................................................... 41

3.1 Outage probability vs. outage rate for V–BLAST with MMSE–SIC re-
    ceiver and different channel augmentation schemes. \( \rho = 0.9 \) and \( N_T = N_V = 4 \). ................................................................. 54
3.2 Outage rate vs. correlation coefficient \( \rho \) for V–BLAST with MMSE–SIC
    receiver and different channel augmentation schemes. \( P_{out} = 0.01 \) and
    \( N_T = N_V = 4 \). ............................................................. 55
3.3 Outage rate vs. number of transmit antennas $N_T$ for D–BLAST with MMSE–SIC receiver and the CVA scheme and OSTBCs. $\rho = 0.6$, $P_{out} = 0.01$. $N_V = 1$ represents the baseline scheme. SNRs of (a) 0 dB, (b) 12 dB, and (c) 24 dB.

4.1 Block diagram of the MIMO–OFDM system with adaptive bit loading. (I)FFT denotes (inverse) fast Fourier transform.

4.2 Example of a WSFC codeword matrix with $N_T = 3$, $d = 3$, and $K = 384$. The indices of the coded symbols are shown in the blocks.

4.3 SNR required for WSFC to achieve BER $= 10^{-3}$ vs. interleaving delay $d$ with $\bar{R} = 2$ bits and 4 bits.

4.4 BER required for V–BLAST with and without ordering vs. extra margin $\eta_{extra}$.

4.5 BER performance for coded MIMO–OFDM with and without ABL. $\bar{R} = 2$ bits, and $d = 16$ for WSFC.

4.6 SNR required to achieve BER $= 10^{-3}$ for WSFC and V–BLAST with ordering based MIMO–OFDM with different group sizes $G$ for ABL. $\bar{R} = 2$ bits, and $d = 16$ for WSFC. The center subcarrier and equivalent SNR methods are used for loading.

5.1 Example of a simple two–phase transmission with $N = 10$ potentially cooperating relay nodes. $N_T = 4$ relay nodes ($S = \{1, 3, 9, 10\}$) successfully decode the packet received from the source node (S) and forward it to the destination node (D).

5.2 Block diagram of a) transmitter with $N_T$ co–located antennas employing BICM–OFDM and CSF filtering, and b) transmitter of the $n$th relay node employing BICM–OFDM and CSF filtering.
5.3 Ratio $\alpha \triangleq f(g)/\tilde{f}(G_{opt})$ vs. $L_g$ for GA–CSF, EV–CSF, and O–CDD filters, respectively. $N_T = 4$, $\rho = 0.9$, and $L = 5$. .......................... 98

5.4 BER of BICM–OFDM with GA–CSF filtering, BM–CDD [87], and Alamouti’s STBC vs. $E_b/N_0$. OFDM, $K = 64$, BPSK, $N_T = 2$, and $\rho_L = 0$. GA–CSF filtering: $8 \times 8$ block interleaver, $L_g = 2$ for $L = 1$, $L_g = 6$ for $L = 5$. BM–CDD: $8 \times 8$ block interleaver and no interleaver. Alamouti’s STBC: $16 \times 8$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$. .......................... 99

5.5 BER of BICM–OFDMA with GA–CSF filtering, O–CDD, BM–CDD [87], and C–CDD [98] vs. $E_b/N_0$. $U = 4$, $K = 512$, 16–QAM, $N_T = 4$, $L = 5$, $\rho_L = 0$, $32 \times 16$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$. .......................... 100

5.6 Minimum coding gain $C_{\text{min}}$ of GA–CSF filtering for various $L_g$, upper bound $C_{b,k}$ for known $S$ and $L_g = 2$, and upper bound $C_{u,k}$ for unknown $S$ and $L_g = 2$ vs. $N$. $L = 1$, $N_T = N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. .......................... 102

5.7 BER of BICM–OFDM with GA–CSF, R–CSF, R–CDD, and C–CDD [86] filtering when $S$ is unknown vs. $E_b/N_0$. For comparison the BER of GA–CSF filtering when $S$ is known is also shown. $N = 10$, $L = 1$, $N_T = N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$. .......................... 103
5.8 BER of BICM–OFDM with GA–CSF and C–CDD [86] filtering when $S$ is unknown vs. $E_b/N_0$. For comparison the BER of GA–CSF filtering when $S$ is known is also shown. The effects of imperfect synchronization (IS) and $N_T = 5 \neq N_a = 2$ are considered. $N = 10$, $L = 1$, $N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$.

5.9 BER at destination node for two-hop transmission with GA–CSF, R–CSF, and C–CDD [86] filtering when $S$ is unknown vs. $E_b/N_0$. For comparison the BER of GA–CSF filtering when $S$ is known is also shown. Each relay node listens to source node with probability $p$. $N = 10$, $L = 1$, $N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$.

6.1 Block diagram of transceiver for BICM–OFDM with robust CSF.

6.2 Block diagram of transceiver for BICM–OFDM with robust SFC–CSF.

6.3 BER of BICM–OFDM with robust CSF filtering vs. $E_b/N_0$ for $L_g = 5$. $N_T = 3$, $L = 2$, and outdated CSIT.

6.4 BER of BICM–OFDM with robust CSF filtering vs. $\rho$ for $L_g = 5$. Furthermore, results for CSF filters optimized under the assumption of perfect and statistical CSIT are also shown. $N_T = 3$, $L = 2$, and outdated CSIT.

6.5 BER of BICM–OFDM with robust CSF filtering vs. $E_b/N_0$ for $L_g = 1$ and $L_g = 5$. $N_T = 3$, $L = 2$, and outdated CSIT.

6.6 BER of BICM–OFDM with robust CSF ($L_g = 3$) and robust SFC–CSF ($L_g = 2$) filtering vs. $1/$SQNR. $N_T = 3$, $L = 1$, $\rho = 0.999$, and quantized CSIT.
6.7 BER of BICM–OFDM with robust CSF ($L_g = 5$) and robust SFC–CSF
($L_g = 2$) filtering vs. $\rho$. $N_T = 3$, $L = 2$, and outdated CSIT. . . . . . . . . 134
6.8 BER of BICM–OFDM with robust CSF filtering vs. $E_b/N_0$. $N_P = 4$, $N_T = 3$, $L = 1$, $L_g = 3$, and outdated CSIT. . . . . . . . . . . . . . . . . 135
6.9 BER of BICM–OFDM with robust CSF filtering vs. $\rho$. $N_T = 2$, $L = 5$, $L_g = 6$, and outdated CSIT. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 136
List of Abbreviations

3GPP 3rd Generation Partnership Project
4G The Fourth-Generation Mobile Wireless System
ABL Adaptive Bit Loading
ADSL Asymmetric Digital Subscriber Lines
AF Amplify-and-Forward
AWGN Additive White Gaussian Noise
BER Bit–Error–Rate
BICM Bit Interleaved Coded Modulation
BLAST Bell Labs Layered Space–Time
BM Bauch and Malik
BPSK Binary Phase Shift Keying
BWA Broadband Wireless Access
CC Convolutional Code
CCB Chow, Cioffi and Bingham
CDD Cyclic Delay Diversity
CDF Cumulative Distribution Function
CDMA Code Division Multiple Access
CE Consumer Electronics
CIR Channel Impulse Response
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>Channel Model</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic Prefix</td>
</tr>
<tr>
<td>CSF</td>
<td>Cyclic Space–Frequency</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>CSIT</td>
<td>CSI at the Transmitter</td>
</tr>
<tr>
<td>CVA</td>
<td>Conjugate Virtual Antenna</td>
</tr>
<tr>
<td>DAB</td>
<td>Digital Audio Broadcasting</td>
</tr>
<tr>
<td>DD</td>
<td>Delay Diversity</td>
</tr>
<tr>
<td>D–BLAST</td>
<td>Diagonal BLAST</td>
</tr>
<tr>
<td>DF</td>
<td>Decode-and-Forward</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
</tr>
<tr>
<td>DSTC</td>
<td>Distributed Space–Time Coding</td>
</tr>
<tr>
<td>DSTF</td>
<td>Distributed Space–Time Filtering</td>
</tr>
<tr>
<td>DVB</td>
<td>Digital Video Broadcasting</td>
</tr>
<tr>
<td>EV</td>
<td>Eigenvector</td>
</tr>
<tr>
<td>FCC</td>
<td>Federal Communications Commission</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>GA</td>
<td>Gradient Algorithm</td>
</tr>
<tr>
<td>GL</td>
<td>Grouped Loading</td>
</tr>
<tr>
<td>IDFT</td>
<td>Inverse Discrete Fourier Transform</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronic Engineers</td>
</tr>
<tr>
<td>IFFT</td>
<td>Inverse Fast Fourier Transform</td>
</tr>
<tr>
<td>I</td>
<td>In-Phase</td>
</tr>
<tr>
<td>ICI</td>
<td>Inter–Carrier Interference</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>Independent and Identically Distributed</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter–Symbol Interference</td>
</tr>
<tr>
<td>LOS</td>
<td>Line of Sight</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Prediction</td>
</tr>
<tr>
<td>LMMSE</td>
<td>Linear Minimum Mean–Squared Error</td>
</tr>
<tr>
<td>LSE</td>
<td>Least-Square Error</td>
</tr>
<tr>
<td>LTE</td>
<td>(3GPP) Long Term Evolution</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>MMSE–SIC</td>
<td>MMSE Successive Interference Cancellation</td>
</tr>
<tr>
<td>MRC</td>
<td>Maximum Ratio Combining</td>
</tr>
<tr>
<td>NAC</td>
<td>Nulling and Cancelling</td>
</tr>
<tr>
<td>O–CDD</td>
<td>Optimized CDD</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>pdf</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PEP</td>
<td>Pairwise Error Probability</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
<tr>
<td>PSP</td>
<td>Per–Survivor Processing</td>
</tr>
<tr>
<td>Q</td>
<td>Quadrature</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quaternary Phase Shift Keying</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RTM</td>
<td>Rankin, Taylor, and Martin</td>
</tr>
<tr>
<td>RVA</td>
<td>Real Virtual Antenna</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>SC</td>
<td>Single Carrier</td>
</tr>
<tr>
<td>SFC</td>
<td>Space–Frequency Coding</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal–to–Interference–plus–Noise–Ratio</td>
</tr>
<tr>
<td>SIR</td>
<td>Signal–to–Interference–Ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single–Input Single–Output</td>
</tr>
<tr>
<td>SIMO</td>
<td>Single–Input Multiple–Output</td>
</tr>
<tr>
<td>SM</td>
<td>Spatial Multiplexing</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal–to–Noise Ratio</td>
</tr>
<tr>
<td>SQNR</td>
<td>Signal–to–Quantization Noise Ratio</td>
</tr>
<tr>
<td>STC</td>
<td>Space–Time Coding</td>
</tr>
<tr>
<td>STTC</td>
<td>Space–Time Trellis Codes</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>UWB</td>
<td>Ultra-Wideband</td>
</tr>
<tr>
<td>V–BLAST</td>
<td>Vertical BLAST</td>
</tr>
<tr>
<td>WL</td>
<td>Widely Linear</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>WMAN</td>
<td>Wireless Metropolitan Area Network</td>
</tr>
<tr>
<td>WSFC</td>
<td>Wrapped Space–Frequency Coding</td>
</tr>
<tr>
<td>WSTC</td>
<td>Wrapped Space–Time Coding</td>
</tr>
<tr>
<td>ZP</td>
<td>Zero Padded</td>
</tr>
</tbody>
</table>
Notation

Bold upper case and lower case letters denote matrices and vectors, respectively. The remaining notation and operators used in this thesis are listed below:

\((\cdot)^*\) Complex conjugate
\([\cdot]^{T}\) Transpose
\([\cdot]^{H}\) Hermitian transpose
\(\text{det}(\cdot)\) Matrix determinant
\(\text{diag}(\mathbf{x})\) Diagonal matrix with the elements of vector \(\mathbf{x}\) on the main diagonal
\(\Re\{\cdot\}\) Real part of a complex number
\(\Im\{\cdot\}\) Imaginary part of a complex number
\(\mathbb{E}\{\cdot\}\) Expectation
\(\text{Pr}\{\cdot\}\) Probability of some event
\(\lambda_i(\cdot)\) \(i\)th non–zero eigenvalue of a matrix
\(Q(\cdot)\) Gaussian Q–function [1]
\(r(\cdot)\) Rank of a matrix
\(tr(\cdot)\) Trace of a matrix
\(*\) Convolution operator
\(\circ\) Cyclic convolution operator
\(\otimes\) Kronecker product operator
\(I_\eta\) Identity matrix of dimension \(\eta \times \eta\)
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{0}_\eta$</td>
<td>All–zero matrix of dimension $\eta \times \eta$</td>
</tr>
<tr>
<td>$\mathbf{0}_{\eta \times 1}$</td>
<td>All–zero column vector of length $\eta$</td>
</tr>
<tr>
<td>$| \cdot |_2$</td>
<td>$L_2$ vector norm</td>
</tr>
<tr>
<td>$\lceil x \rceil$</td>
<td>Smallest integer larger than $x$</td>
</tr>
<tr>
<td>$\lfloor x \rfloor$</td>
<td>Closest integer to $x$</td>
</tr>
<tr>
<td>vec($\cdot$)</td>
<td>Vectorization of a matrix</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank my advisor, Professor Robert Schober, for his support and invaluable advice during the course of my Ph.D. study. I am much indebted for his patience and encouragement over the years. Without his support and guidance, this thesis would not be possible. I also would like to express my gratitude to Professor Lutz Lampe for many stimulating and helpful discussions and guidance on the work we have done together.

I would like to thank the other member of my doctoral committee: Professor Vincent Wong, the University examiners: Professor Jane Wang and Professor Ozgur Yilmaz, and the External examiner: Professor Gerhard Bauch, for the time and effort in evaluating my work and providing valuable feedback and suggestions.

Many thanks go to the members of the Communications Theory Group for all the fruitful discussions, group meetings, and feedback at many points.

Special thanks to my parents and my brothers for all their support during these years. My deepest gratitude to my wife for her love and patience. All my work is dedicated to my father Chen Jinhai, who passed away in May 2007.
1 Introduction

Information–theoretic studies have shown that multiple–input multiple–output (MIMO) communication architectures are able to provide extraordinary high spectral efficiencies in rich multipath environments, which are simply unattainable using conventional techniques [2–4]. Meanwhile, orthogonal frequency–division multiplexing (OFDM) is regarded as an efficient approach to combat the adverse effects of multipath fading. For improved power and bandwidth efficiency the combination of OFDM and multiple transmit and receive antennas, often referred to as MIMO–OFDM, has been extensively studied for wireless communication systems. In such a MIMO–OFDM system, spatial multiplexing, space–time (or space–frequency) coding, adaptive bit loading, and special signal processing algorithms are usually employed for signal design in order to approach the MIMO channel capacity. MIMO–OFDM has already been adopted in major wireless standards, such as the IEEE 802.11n wireless local areal networks (WLANs) [5], the IEEE 802.16e wireless metropolitan area networks (WMANs) [6], and the 3rd Generation Partnership Project (3GPP) long term evolution (LTE) [7]. In the future, MIMO–OFDM is expected to be employed in more wireless communication systems, e.g., the fourth-generation (4G) mobile wireless systems.

The remaining sections of this Introduction are organized as follows. In Section 1.1, we briefly review MIMO wireless systems. In Section 1.2, we discuss the OFDM modulation scheme. In Section 1.3, we introduce MIMO–OFDM systems, and in Section
1. Introduction

1.4, we briefly outline the contributions made in this thesis and the thesis organization.

1.1 MIMO Wireless Communications

In a MIMO wireless communication system, multiple antennas are employed at both the transmitter and the receiver. Data transmissions over the resulting MIMO channels may be implemented in different ways to obtain a spatial multiplexing gain, a spatial diversity gain, or both.

Spatial multiplexing schemes yield a throughput increase by transmitting independent information streams in parallel through the spatial channels, and the capacity is limited by the minimum of the number of transmit and the number of receive antennas. Several schemes have been proposed to exploit the spatial multiplexing gain, such as the vertical Bell Labs layered space–time (V–BLAST) [8] and the diagonal Bell labs layered space–time (D–BLAST) [3] schemes.

Diversity gain is achieved by sending the data signal over multiple independent fading paths in time, frequency, and space and by utilizing appropriate combining techniques at the receiver. Spatial diversity is particularly attractive as no bandwidth expansion is required compared to the two other commonly used diversity techniques, time diversity and frequency diversity. Spatial diversity can be further broken down into receive diversity and transmit diversity. Space–time coding (STC) schemes, such as delay diversity (DD) [9], space–time block codes (STBC) [10, 11], and space–time trellis codes (STTC) [12], are the well–known transmit diversity techniques, which lead to improved link reliability.

While the aforementioned benefits of spatial diversity or spatial multiplexing are realizable with an open–loop configuration, where the receiver alone knows the communication channel, these benefits are further enhanced in a closed–loop MIMO system,
where the transmitter also knows the channel. By exploiting channel state information (CSI) at the transmitter (CSIT), eigenbeamforming, which converts a MIMO channel into a bank of scalar channels without crosstalk from one scalar channel to the others, achieves both a spatial multiplexing gain and a diversity gain. For example, in a wireless system with four transmit and two receive antennas, CSIT can more than double the capacity at a signal–noise–ratio (SNR) of 0.5 dB and add 1.5 bps/Hz to the capacity at an SNR of 5 dB [13]. Therefore, exploiting CSIT is of great practical interests.

However, in practice, it is often not possible to obtain perfect instantaneous CSIT due to delays in the feedback channel from the receiver to the transmitter, measurement noise, and quantization noise. In this case, high performance can only be achieved if not only the imperfect CSIT itself but also the reliability of the CSIT is taken into account for transmitter design. For flat fading channels, examples for successful exploitation of imperfect CSIT can be found in e.g. [14–17]. Furthermore, the effect of linear prediction of CSIT on the performance of beamforming in flat fading channels has been considered in [18, 19].

MIMO systems require that multiple antennas are spaced from each other at least one–half of the wavelength of the transmitted signal to obtain low correlation between the spatial channels. This requirement fundamentally limits the possibility of having multiple antennas on small communication devices. It has recently been shown that node cooperation can be used to create a virtual MIMO system [20, 21], where relays help the source exploit the spatial diversity of a slow fading channel in a distributed fashion by decode-and-forward (DF) or amplify-and-forward (AF) relaying.

Research related to cooperative communication started with the introduction of the relay channel in [22]. Important bounds on the capacity of a three–node relay network for the additive white Gaussian noise (AWGN) channel were provided in [23]. The concept of user cooperation diversity was introduced in [21], where two users share
their antennas and resources to obtain a diversity gain through distributed transmission. Various cooperative transmission protocols were proposed in [20], and their performances were analyzed in terms of outage behavior in [24]. The idea of distributed space–time coding (DSTC) was introduced in [20] and various DSTC schemes [25–27] have been investigated since then.

Channel augmentation is another form of virtual MIMO techniques, where virtual receive antennas are created for performance improvement without requiring the help of other nodes. Channel augmentation is realized by means of special signal processing at both the transmitter and the receiver. In [28] the fingers of a RAKE receiver were exploited to form an augmented channel matrix. A virtual MIMO framework for single-transmitter single-receiver multipath fading channels was developed in [29] to enable maximum exploitation of channel diversity at both the transmitter and the receiver. A smart antenna system using virtual array elements was proposed in [30]. Furthermore, in [31], the information outage rate of MIMO channels was shown to be potentially improved by repeatedly transmitting the same signal and combining the received signals in such a way as to increase the effective rank of the MIMO channel.

1.2 OFDM Modulation

OFDM is a special form of multi-carrier transmission that divides a communication channel into a number of equally spaced frequency bands, over which a number of lower rate data signals are transmitted. The OFDM transmission scheme has the following key advantages.

1. OFDM is an efficient means of combating multipath fading in the wireless environment. By using the inverse fast Fourier transform (IFFT) and cyclic prefix (CP) insertion at the transmitter, together with FFT and CP removal at the receiver, OFDM
converts a frequency–selective fading channel into a number of parallel flat fading channels, thus significantly simplifying the receiver design without the need for a complicated equalizer, necessary for single–carrier transmission.

2. OFDM can potentially enhance the capacity by adapting the data rate and power to the subcarrier channel condition. This is achieved by using adaptive bit and power loading techniques, where signal constellations of different size are transmitted over subcarriers with different gains.

3. OFDM is robust to narrowband interference. The narrowband interference only affects a small portion of the subcarriers, and the corrupted symbols can possibly be recovered by using forward error correction (FEC) coding.

4. OFDM facilitates multiple access by allocating multiple users to different subcarriers in a single carrier network. This approach is referred to as orthogonal frequency division multiple access (OFDMA).

The concept of using parallel data transmission and frequency division multiplexing can be traced back to the 1950s and the 1960s [32, 33]. However, due to the implementation complexity of the oscillator and demodulator for each subcarrier required by frequency division multiplexing, the application of OFDM was limited to military systems at that time. In 1971, Weinstein and Ebert [34] introduced the idea of using a Discrete Fourier Transform (DFT) for implementation of the generation and demodulation of OFDM signals, eliminating the requirement for banks of analog subcarrier oscillators. This presented an opportunity for a simple implementation of OFDM, especially with the use of FFT. Recent advances in integrated circuit technology have made the implementation of OFDM cost effective. In the 1980s, OFDM was studied for high–speed modems and digital mobile communications. Since the 1990s, OFDM has been commercialized in asymmetric digital subscriber lines (ADSL), digital audio broadcasting (DAB), digital video broadcasting (DVB–T), and more recently several
wireless communication standards [5–7]. In these systems, the combination of OFDM and FEC coding is necessary to protect the symbols over the subcarriers experiencing deep fades, achieving a frequency diversity gain inherent in OFDM.

Conventional multicarrier systems use a fixed signal constellation and power across all subcarriers, thus the overall error probabilities are dominated by the subcarriers with the worst performance. To maximize the link capacity or to minimize the overall error probability of an OFDM system, adaptive loading schemes allocate different constellations and powers across the subcarriers according to the measured SNR. Most loading algorithms can be classified into two categories: rate–adaptive loading and margin–adaptive loading. The first type of loading aims at maximizing the overall bit rate (throughput) given a fixed amount of energy. In the second type of loading, the goal is to determine the bit and power allocation that requires the least amount of energy with a fixed number of bits per OFDM symbol. A number of algorithms have been proposed to solve the discrete loading problem in practice. Examples include the Hughes–Hartogs algorithm [35], the Chow, Cioffi and Bingham algorithm [36], the Fischer and Huber algorithm [37], and some fast algorithms proposed in [38–40].

1.3 MIMO–OFDM Systems

Most MIMO schemes were initially proposed for flat fading channels and cannot be directly applied to frequency–selective channels, which are the most practical channels for broadband wireless communications. On the other hand, OFDM provides an efficient approach to combat the adverse effect of multipath spread. A combined MIMO–OFDM approach can benefit from the advantages of both OFDM and MIMO to enable spatial multiplexing gains and diversity gains over frequency–selective channels.

Since OFDM results in a number of parallel flat–fading channels, existing MIMO
techniques proposed for single-carrier transmission can be applied to each subcarrier. Spatial multiplexing is performed by transmitting independent data streams on a subcarrier-by-subcarrier basis with the total transmit power split uniformly across antennas and subcarriers. Due to a large number of subcarriers, the computational complexity of spatial multiplexing for MIMO-OFDM could be extremely high. In [41], an interpolation-based method was introduced to alleviate this problem by taking advantages of the correlation between subcarriers.

Exploiting the presence of the spatial diversity offered by multiple antennas, STC relies on simultaneous coding across space and time to achieve a diversity gain. OFDM offers frequency diversity by coding and interleaving across subcarriers and appropriate decoding algorithms. A straightforward way to realize space-frequency diversity is to apply existing space-time codes to code across space and frequency (subcarrier) [42, 43]. However, space-frequency coding (SFC), which spreads the symbols across space and frequency together with bit-interleaved coded modulation (BICM) [44], achieves space and frequency diversity in a more systematic fashion [45–48].

A particularly simple and efficient form of SFC is cyclic delay diversity (CDD) which is a straightforward extension of linear DD introduced in [9] for single-carrier transmission to multi-carrier systems. In CDD the same signal is transmitted over different antennas with different cyclic delays [49, 50]. CDD increases the frequency selectivity of the channel without increasing the length requirements for the CP. In fact, CDD transforms spatial diversity into additional frequency diversity which can be picked up by an appropriate FEC coding scheme. In this context, BICM is particularly interesting due to its simplicity and high performance [51]. In fact, it has been shown that CDD with BICM can exploit the full spatial and spectral diversity of the channel if the combination of cyclic delay, interleaver, and free distance of the code is carefully chosen [51].

The detailed literature review relevant to specific prior work will be given in Chap-
1. Introduction

1.4 Thesis Contributions and Organization

This thesis considers several issues regarding the design of MIMO–OFDM systems from the point of view of signal processing. The goal is to design a high performance MIMO–OFDM system with affordable computational complexity. The main contributions of this thesis are as follows.

1. We propose two novel channel augmentation schemes for MIMO systems, which can achieve a better outage performance than a previously proposed scheme with the same low computational complexity.

2. We propose several modifications for MIMO–OFDM combined with adaptive bit-loading (ABL) and FEC coding that improve the frame and bit error rates.

3. We propose cyclic space–frequency (CSF) filtering for BICM–OFDM systems with multiple co–located or distributed transmit antennas to achieve a spatial and frequency diversity gain.

4. We propose robust CSF filtering for MIMO–OFDM, which exploits imperfect CSIT and takes into account the reliability of the CSIT via a Bayesian model.

While the first contribution is related to generic MIMO systems, the remaining contributions are pertinent to MIMO–OFDM systems. In the following, we provide a brief overview of the remainder of this thesis.

In Chapter 2, we first introduce some background material and important results on MIMO–OFDM that are relevant to our proposed signal design schemes in Chapters 3–6.

In Chapter 3, we present the proposed channel augmentation schemes for MIMO systems. Recently, Rankin, Taylor, and Martin (RTM) [31] proposed a simple channel augmentation scheme for MIMO systems by repeatedly transmitting the same symbols...
in several time slots and appropriate combining techniques at the receiver. This scheme was shown to be able to improve the outage rate of slow fading MIMO channels. Inspired by [31] we propose two channel augmentation schemes having the same low complexity as the RTM scheme. In the first scheme, the transmitted symbols are both repeated and complex conjugated leading to the conjugate virtual antenna (CVA) scheme. In the second, the transmitted symbols are constrained to be real-valued leading to the real virtual antenna (RVA) scheme. We show that the RTM, CVA, and RVA schemes can achieve a higher outage rate than the baseline scheme without augmentation for both V–BLAST and D–BLAST. The novel RVA and CVA schemes are shown to be more robust to channel correlation than the RTM scheme.

Chapter 4 deals with MIMO–OFDM combined with ABL and FEC coding. We investigate “simple” coding schemes and their combination with ABL for MIMO–OFDM. In particular, we consider wrapped space–frequency coding (WSFC) and coded V–BLAST with ABL and optimize both schemes to mitigate error propagation inherent in the detection process. Simulation results show that bit–loaded WSFC and V–BLAST optimized for coded MIMO–OFDM achieve excellent error rate performances, close to that of quasi–optimal MIMO–OFDM based on singular value decomposition (SVD) of the channel, while their feedback requirements for loading are low.

In Chapter 5, we introduce CSF filtering for OFDM systems using BICM and multiple transmit antennas for fading mitigation, and discuss the extension to OFDMA systems. CSF filtering is a simple form of SFC and may be viewed as a generalization of CDD where the cyclic delays are replaced by CSF filters. CSF is applicable to both traditional MIMO systems with co–located transmit antennas and cooperative diversity systems with DF relaying and distributed transmit antennas. Similar to CDD and in contrast to other SFC schemes, CSF filtering does not require any changes to the receiver compared to single–antenna transmission. Based on the asymptotic PEP of the overall system
1. Introduction

We derive an optimization criterion for the CSF filters for co-located and distributed transmit antennas, respectively. We show that the optimum CSF filters are independent of the interleaver if they do not exceed a certain length. If this length is exceeded, the adopted interleaver has to be carefully taken into account in the CSF filter design. For several special cases we derive closed-form solutions for the optimum CSF filters and for the general case we provide various CSF filter design methods. Our simulation results show that CSF filtering can achieve significant performance gains over existing CDD schemes.

Chapter 6 is concerned with robust CSF filtering, which exploits imperfect CSIT and takes into account the reliability of the CSIT via a Bayesian model. To further improve robustness, we combine CSF filtering with STBC in the frequency domain and refer to the resulting scheme as SFC–CSF filtering. We also propose a linear prediction method for improving the quality of the CSIT via post-processing. Based on an upper bound on the worst-case pairwise error probability we formulate an optimization problem for the robust CSF filters which can be solved exactly for certain special cases. For the general case, we obtain an approximate solution by solving a related problem with additional constraints. This approximate solution can be further improved with a gradient algorithm for the original problem. Simulation results confirm the excellent performance of BICM–OFDM with robust CSF and SFC–CSF filtering and the effectiveness of the proposed CSIT post-processing method.

Finally, Chapter 7 summarizes the contributions of this thesis and outlines areas of future research.
2 Preliminaries

This chapter introduces some relevant background material related to MIMO–OFDM systems. We present in Section 2.1 MIMO fading channel models that will be considered for the rest of this thesis. Various MIMO signal processing schemes have been developed according to the nature of the available CSI at the transmitter and the receiver. We discuss these MIMO techniques in Section 2.2. In some situations, multiple antennas are not available, virtual MIMO techniques can be used to achieve a diversity gain or to improve outage performance. We introduce virtual MIMO techniques in Section 2.3. OFDM is a popular method for transmission over frequency–selective channels. We present the basic principle of OFDM as well as a related performance enhancement method, namely, adaptive loading, in Section 2.4. OFDM is often combined with MIMO to form a so–called MIMO–OFDM architecture for improved power and bandwidth efficiency. We discuss MIMO–OFDM and focus on the existing SFC schemes in Section 2.5.

2.1 MIMO Channel Model

The key characteristics of the wireless channel are the variations of the channel strength over time and frequency. These variations are often classified into scales [52, 53]: large and small. Large scale fading captures path loss and shadowing, which result from sig-
2. Preliminaries

Inertial attenuation with distance and random blockage by large objects, such as hills and buildings. Small scale fading captures the variation arising from signals of multiple random paths adding constructively and destructively. Since signal processing for wireless communication usually exploits the small scale channel variations, this section focuses on small scale channel models.

2.1.1 Signal Envelope

Small-scale variation of a channel causes a rapid fluctuation of the received signal envelope. The in-phase (I) and quadrature (Q) components of the received signal are sums of many independent random variables because of multipath propagation. From the central limit theorem, both the I and Q components of the channel gain can be approximated as Gaussian random variables \( X_I \) and \( X_Q \) with variance \( \sigma^2 \) [52]. The envelope of the channel response with a non-zero mean resulting from a direct path, \( Z = \sqrt{X_I^2 + X_Q^2} \), is a Ricean distributed random variable given by [53]

\[
p(z) = \frac{z}{\sigma^2} \exp \left( -\frac{z^2 + s^2}{2\sigma^2} \right) I_0 \left( \frac{zs}{\sigma^2} \right), \quad \text{for } z \geq 0, \tag{2.1}
\]

where \( 2\sigma^2 \) is the average power in the multipath components and \( s^2 \) is the power of the direct path component. The function \( I_0(x) \) is the zeroth order modified Bessel function of the first kind defined as

\[
I_0(x) \triangleq \frac{1}{2\pi} \int_0^{\pi} e^{x \cos \theta} d\theta. \tag{2.2}
\]

The Ricean factor \( K_R \) is defined as the ratio of the direct path power and the average multipath power

\[
K_R \triangleq \frac{s^2}{2\sigma^2}. \tag{2.3}
\]
We can write the Ricean distribution in terms of parameter $K_R$ as

$$p(z) = \frac{2z(K_R + 1)}{\Omega_p} \exp\left(-K_R - \frac{(K_R + 1)z^2}{\Omega_p}\right) I_0\left(2z \sqrt{\frac{K_R(K_R + 1)}{\Omega_p}}\right), \quad \text{for } z \geq 0, \quad (2.4)$$

where $\Omega_p = 2\sigma^2 + s^2$ is the total average power of the direct path and the multipath. If $K_R = 0$, both $X_I$ and $X_Q$ have zero means. As a result, a Ricean random variable becomes a Rayleigh random variable with

$$p(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad \text{for } z \geq 0. \quad (2.5)$$

### 2.1.2 Characteristics of Multipath Fading

When an ideal impulse is transmitted over a time–varying multipath fading channel, there will be two effects on the received signal. First, the received signal may appear as a train of pulses with different delays and magnitudes due to the different lengths and attenuation factors of different signal paths. Second, the multiple paths are varying with time because of the random nature of the wireless channel. Thus, the number of received pulses, the delay between them, and their magnitudes may vary over time. The impulse response of the channel captures both of these effects, and exhibits spectral and temporal selectivity.

It is usually convenient to represent the channel in a discrete–time model for digital signal processing. The commonly used tapped–delay–line model is given by [52]

$$y[k] = \sum_{l=0}^{L-1} h[k; l] x[k - l] + n[k], \quad (2.6)$$

where $y[k], x[k],$ and $n[k]$ are the received signal, the transmitted signal, and the noise samples at sampling instant $k$, respectively, and $h[k; l]$ represents the sampled time-
varying channel impulse response (CIR) of finite memory $L$.

Spectral selectivity is caused by the presence of multipath. The channel frequency response varies with frequency, leading to selectivity. The longer the multipath delays, the more frequency selective the channel becomes. A measure for this selectivity is the delay spread, defined as the range of multipath spread in the channel, $T_m = \tau_{\text{max}} - \tau_{\text{min}}$, where $\tau_{\text{max}}$ and $\tau_{\text{min}}$ are the maximum and minimum multipath delays, respectively. The channel coherence bandwidth $B_c$ is accordingly defined as $B_c \approx 1/T_m$ [1].

Temporal selectivity, on the other hand, is caused by motion of the transmitter, the receiver, or the scatterers in the channel. These motions cause a transmitted single tone to be spread in frequency at the receiver. This phenomenon is also referred to as Doppler effect. Suppose a receiver is moving at a constant speed $v$ and the arrival angle of plane waves incident on the receiver is $\theta$, then, the Doppler shift $f_d$ is defined as

$$f_d = \frac{v \cos \theta}{\lambda} = \frac{vf_c \cos \theta}{c},$$

(2.7)

where $\lambda$ is the wavelength, $f_c$ is the signal frequency, and $c$ is the speed of light. The maximum Doppler frequency shift $f_D$ occurs when $\cos \theta = 1$, given by $f_D = vf_c/c$.

If the multiple paths are widely distributed in angles, it is reasonable to assume that the arrival angles are uniformly distributed over $[0, \ 2\pi)$. With the uniform arrival angle assumption, the Doppler spectrum is given as [1]

$$S(f) = \frac{1}{\pi \sqrt{f_D^2 - (f - f_c)^2}}, \quad f_c - f_D \leq f \leq f_c + f_D,$$

(2.8)

The time correlation function of the $l$th tap can be derived from the Doppler spectrum as

$$\rho[m; l] = \mathcal{E} \{h[k; l]h^*[k + m; l]\} = J_0(2\pi f_D m \Delta t),$$

(2.9)
where $\mathcal{E}\{\cdot\}$ is the mean of its argument, $J_0(x)$ is the zeroth order Bessel function of the first kind, and $\Delta_t$ is the sampling interval between two signal measurements.

An indicator of the temporal selectivity is the channel coherence time, defined as the time interval over which the channel remains strongly correlated. Clearly, a slowly changing channel has a large coherence time. Since the coherence time is a statistically defined quantity, an approximate relation to the maximum Doppler shift is $T_c \approx 1/f_D$ [1].

For OFDM systems, we usually assume that the channel is constant for at least one OFDM symbol processing window, leading to the following simplified channel model [52]

$$y[k] = \sum_{l=0}^{L-1} h[l]x[k - l] + n[k],$$

(2.10)

with time–invariant CIR $h[l]$.

### 2.1.3 MIMO Channel Parameters

The MIMO wireless channel is formed by using multiple antennas at the transmitter and the receiver. Multiple antennas add an additional degree of freedom to wireless systems with the existing temporal and spectral dimensions. The MIMO channel is often represented in a matrix form. In a system with $N_T$ transmit and $N_R$ receive antennas, the $l$th tap of the channel at sampling instant $k$ can be represented as a matrix $H[l]$ of size $N_R \times N_T$,

$$H[l] = \begin{bmatrix}
  h_{11}[l] & h_{12}[l] & \cdots & h_{1N_T}[l] \\
  h_{21}[l] & h_{22}[l] & \cdots & h_{2N_T}[l] \\
  \vdots & \vdots & \ddots & \vdots \\
  h_{N_R1}[l] & h_{N_R2}[l] & \cdots & h_{N_RN_T}[l]
\end{bmatrix},$$

(2.11)
where $h_{ij}[l], 1 \leq i \leq N_R, 1 \leq j \leq N_T$, is the channel gain from the $j$th transmit element to the $i$th receive element, and can be modeled as a complex Gaussian random process. These channel gains, however, can be correlated and have different means. The channel can be decomposed into a fixed part and a variable part as

$$H[l] = H_m[l] + \tilde{H}[l],$$

(2.12)

where $H_m[l] = \mathcal{E}\{H[l]\}$ is the channel mean, and $\tilde{H}[l] = H[l] - H_m[l]$ is a zero-mean complex Gaussian random matrix.

Even at the same time and frequency, the channels resulting from different transmit–receive antenna pairs could experience different fading, which is related to antenna spacing, the angle of departure of signals, and the angle of arrival of signals. The channel covariance captures the spatial correlation among all the transmit and receive antennas. For convenience we collect the CIR coefficients in vector $h_{nr} \triangleq [h_{nr}[0] h_{nr}[1] \ldots h_{nr}[L-1]]^T$ and define the $N = N_R N_T L$ dimensional vector $h \triangleq [h_{11}^T h_{21}^T \ldots h_{N_T N_R}^T]^T$ containing all CIR coefficients. The mean and the covariance matrix of $h$ are defined as

$$m_h \triangleq \mathcal{E}\{h\}$$

(2.13)

$$C_{hh} \triangleq \mathcal{E}\{(h - m_h)(h - m_h)^H\}.$$  

(2.14)

$C_{hh}$ is a positive semidefinite Hermitian matrix and is often assumed to have a separable Kronecker structure [54, 55], i.e., the transmission and reception fade independently, and tap correlation can be separated from transmit and receive antenna correlation. In this case, the channel covariance can now be decomposed as

$$C_{hh} = \frac{1}{1 + K_R} C_{N_R} \otimes C_{N_T} \otimes C_L,$$

(2.15)
where $C_{NR}$, $C_{NT}$, and $C_{CL}$ denote the receive antenna, the transmit antenna, and the CIR coefficient covariance matrices, respectively, and $K_R$ is the Ricean factor.

The channel auto-covariance sequence captures both channel spatial and temporal correlations. The channel auto-covariance sequence is defined as

$$C_{hh}[m] = \mathcal{E}\{h[k]h^H[k + m]\}, \quad (2.16)$$

which depends on the time difference $m$. Here, $h[k]$ is the channel sampled at instant $k$. It is often assumed that the temporal correlation is identical for all antenna elements, thus the spatial and temporal correlation effects are separable, and the channel auto-covariance becomes

$$C_{hh}[m] = \rho[m]C_{hh}. \quad (2.17)$$

### 2.2 MIMO Techniques

Assuming ideal CSI at the receiver, work on multiple antennas can be classified into two main classes: open-loop and closed-loop, according to the availability of CSIT. The former does not require CSI while the latter uses a feedback channel to send CSI acquired at the receiver back to the transmitter for signal design.

Among the open-loop techniques, layered space–time architectures [3, 56, 57] have been proposed to achieve high spectral efficiencies while STC [10, 12] has been conceived mainly for exploiting diversity gains to improve link reliability. Compared with open-loop techniques, closed-loop techniques have an SNR advantage due to the array gain. If the transmitter only knows the channel state partially, precoding combined with STC [14] can achieve high data rates. In the ideal case of perfect CSI available at both sides of the communication link, the system can be transformed into a set of parallel scalar
channels by SVD, followed by waterfilling power allocation to exploit the full channel capacity [4].

Since initial MIMO research efforts focused on narrowband transmission, we first consider a system with \( N_T \) transmit antennas and \( N_R \) receiver antennas over a flat fading channel. The \( N_R \times 1 \) vector of received signal samples at the output of the receive antennas is given by

\[
y = H x + n,
\]

where \( x \) is the \( N_T \times 1 \) vector of modulation symbols transmitted with unit energy in parallel by the \( N_T \) transmit antennas, \( H \) is the normalized \( N_R \times N_T \) channel matrix with \( \mathcal{E}\{\text{tr}[H]\} = N_T N_R \), where \( \text{tr}[\cdot] \) denotes the trace of a matrix, and \( n \) is the \( N_R \times 1 \) AWGN vector with each noise element having variance \( \sigma_n^2 \). The block fading channel is assumed to be perfectly known at the receiver.

2.2.1 Open–Loop Technique I – Layered Space–Time Processing

In layered space–time architectures, multiple coding streams (layers), each of which corresponds to a different transmit antenna, are transmitted simultaneously to exploit the spatial multiplexing gain of a MIMO system without CSIT. Among the layered space–time processing schemes are V–BLAST, D–BLAST, and wrapped space–time coding (WSTC), which will be introduced in this section.

The V–BLAST architecture proposed in [56] uses a vertical layered coding structure as illustrated in Fig. 2.1. Since optimal maximum–likelihood (ML) detection might be too complicated, V–BLAST typically relies upon suboptimal low–complexity signal processing at the receiver. The received signal vector \( y \) is first multiplied with a front–end filter matrix \( F^H \), which is obtained according to the zero–forcing (ZF) or the minimum
mean square error (MMSE) criterion,

\[ v = F^H y. \]  \hfill (2.19)

The resulting observable samples are then processed with nulling and cancelation successively, where once a layer is successfully decoded, it is subtracted from the received vector, thus its interference to the remaining layers is canceled. However, if one layer is decoded incorrectly, this error propagates to all remaining layers. Therefore, the order of detection of different layers is crucial for the error probability due to the risk of error propagation. In general, the layers with the lowest error probability should be decoded first \[8, 58–60\].

In V–BLAST, each coded layer extends horizontally in the space–time grid and is placed vertically above each other. Since there is no coding across different transmit antennas, V–BLAST does not exploit full diversity in the channel, and its performance is dominated by its worst stream. Therefore, V–BLAST is a suboptimal architecture for slow–fading channels, and outage occurs whenever one of these streams is in a deep fade and cannot support the rate of the stream.

Significant improvement of V–BLAST can be achieved by coding across different antennas. This is the basic idea of D–BLAST \[3\]. Fig. 2.2 illustrates the high level system diagram of D–BLAST.

In D–BLAST, multiple codewords are staggered so that each codeword spans multi-
ple antennas, but the symbols sent simultaneously by the different antennas belong to different codewords. Each layer of D–BLAST is striped diagonally across the space–time grid. Since each diagonal layer constitutes a complete codeword, decoding is performed layer–by–layer as shown in Fig. 2.3.

From the figure, we observe that in the initialization and termination phase some of the antennas have to be silent. Therefore, D–BLAST suffers from a rate loss. Nevertheless, as the number of layers increases, D–BLAST approaches the theoretical outage performance of MIMO systems. In this respect, D–BLAST is an outage-optimal architecture.

The length of each D–BLAST codeword is the product of the delay and the number of transmit antennas $N_T$. For a given $N_T$, a large delay is required in order to have a long codeword, which leads to a smaller rate loss. On the other hand, a small delay might pose a problem for using trellis codes with large number of states. To solve this dilemma, WSTC was proposed in [57].
Unlike D–BLAST, which has several codewords, WSTC uses only a single encoder to produce a single codeword, which is diagonally interleaved to form a special codeword matrix as depicted in Fig. 2.4. At the receiver, low–complexity suboptimal decoding is performed combining decision–feedback equalization and per–survivor processing (PSP), similar to the nulling and canceling procedure in D–BLAST and V–BLAST. The interference is canceled by using tentative decisions from the survivor history in the Viterbi decoder. In WSTC, the interleaver delay can be freely chosen. For the limiting case with zero delay, WSTC has a vertical structure. For a long codeword, the codeword matrix can be concatenated to fill the leading and tailing triangles of zeros, which entails no rate loss.

\subsection{Open–Loop Technique II – STC}

STC achieves a diversity gain by introducing temporal and spatial dependencies into the signals transmitted from different antennas without increasing the total transmit power or the transmission bandwidth. The most attractive form of STC is STBC due to its simplicity, where the information data stream is encoded in blocks.

The simplest example of STC is the DD scheme [9, 61], where the signal transmitted at sampling instant \( k \) from the \( n \)th antenna is \( x_n[k] = x[k - \Delta_n] \) for \( 2 \leq n \leq N_T \) and
$x_1[k] = x[k]$, where $\Delta_n$ denotes the time delay (in symbol periods) on the $n$th transmit antenna. Assuming a single receive antenna, the received signal is given by

$$y[k] = \sum_{n=1}^{N_T} x_n[k] \cdot h[k] + n[k] = x[k] \cdot \left( h[k] + \sum_{n=2}^{N_T} h[k - \Delta_n] \right) + n[k], \quad (2.20)$$

where $\ast$ denotes the convolution operation, $h[k]$ is the CIR, and $n[k]$ is the AWGN. From (2.20) it is clear that delay diversity transforms spatial diversity into multipath diversity that can be exploited through equalization. For flat-fading channels, we can choose $\Delta_n = (n-1)$ and achieve full spatial diversity with order $N_T$ using ML detection.

We now consider general coherent STBCs. Suppose the information data stream is divided into groups of $L_n$ source symbols $x_0, x_1, \ldots, x_{L_n-1}$ and these symbols are encoded as a STBC matrix

$$B = \begin{bmatrix}
    b_1[0] & b_2[0] & \ldots & b_{N_T}[0] \\
    b_1[1] & b_2[1] & \ldots & b_{N_T}[1] \\
    \vdots & \vdots & \ddots & \vdots \\
    b_1[T-1] & b_2[T-1] & \ldots & b_{N_T}[T-1]
\end{bmatrix}, \quad (2.21)$$

where $b_n[t]$, $n = 1, \ldots, N_T$, $t = 0, \ldots, T-1$, is the encoded symbol transmitted by antenna $n_T$ at time $t$. The encoded symbols $b_n[t]$ are normalized such that the total energy transmitted by the $N_T$ antennas in any symbol interval $t$ is equal to 1. The code rate $R$ is defined as

$$R = \frac{L_n}{T}. \quad (2.22)$$

Assuming coherent detection and perfect channel estimation at the receiver, the PEP $P_e$ of mistaking an erroneous codeword $B_1$ for the transmitted codeword $B_0$, $B_0 \neq B_1$
is upper bounded by [12]

\[ P_e \leq \prod_{i=1}^{r(A)} \frac{1}{\lambda_i(A)} \left( \frac{E_s}{4N_0} \right)^{-r(A)} \] (2.23)

at high SNRs. Here, matrix \( A \) is defined as \( A \triangleq E^H E \), \( E \triangleq B_0 - B_1 \), and \( \lambda_i(A) \) and \( r(A) \) denote the \( i \)th non-zero eigenvalue of \( A \) and the rank of \( A \), respectively. The coding gain and diversity gain are defined as

\[ G \triangleq \left( \prod_{i=1}^{r(A)} \lambda_i(A) \right)^{1/r(A)} \quad \text{and} \quad d \triangleq r(A) \] (2.24)

respectively. The objective of code design is to make both the diversity gain and the coding gain as large as possible. At high SNRs, the upper bound in (2.23) is dominated by the diversity gain \( d \), which determines the asymptotic slope of the bit-error-rate (BER) curve on the log-log scale. On the other hand, the coding gain specifies the relative offset of the BER curves of different codes.

In general, the decoding complexity of the ML detection of STBC is exponential in \( L_n \). In addressing the decoding complexity, Alamouti discovered an ingenious STBC scheme for transmission with two antennas. In Alamouti’s scheme, the source symbols are grouped in pairs, where two symbols \( \{x_0, x_1\} \) are simultaneously transmitted in the first time slot by antennas 1 and 2, respectively. In the next time slot, antenna 1 transmits \( -x_1^* \) and antenna 2 transmits \( x_0^* \). This imposes an orthogonal spatial–temporal structure on the transmitted symbols. Since two symbols are transmitted in two symbol intervals, the rate of this STBC is \( R = 1 \). The STBC matrix can be written as

\[ B = \begin{bmatrix} x_0 & x_1 \\ -x_1^* & x_0^* \end{bmatrix} \] (2.25)
and the received signals in the two symbol intervals can be collected in the received vector $r \triangleq [r[0] \ r[1]]^T$ given by

$$r = Bh + n,$$

(2.26)

where $h \triangleq [h_1 \ h_2]^T$, $n \triangleq [n[0] \ n[1]]^T$, and $h_\nu$, $\nu = \{1, 2\}$ are the fading gains between antenna $\nu$ and the receive antenna and AWGN, respectively. In particular, the received signals in the two time slots can be expressed as

$$r[0] = h_1x_0 + h_2x_1 + n[0] \quad \text{and} \quad r[1] = -h_1^*x_1^* + h_2^*x_0^* + n[1],$$

(2.27)

respectively. We define a new received vector $y \triangleq [r[0] \ r^*[1]]^T$ such that (2.27) can now be written as

$$y = Hx + \tilde{n},$$

(2.28)

where $x \triangleq [x_0 \ x_1]^T$, $\tilde{n} \triangleq [n_0 \ n_1^*]^T$, and

$$H \triangleq \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}.$$

(2.29)

Assuming coherent detection and perfect channel estimation, the ML estimate $\hat{x}$ for the information carrying vector $x$ is given by

$$\hat{x} = \arg\min_{x \in X^2} \{ \| y - Hx \|^2_2 \},$$

(2.30)

where $X^2$ is the set of all possible signal vectors for $x$. Taking advantage of the fact that $H^H H = (|h_1|^2 + |h_2|^2)I_2$, where $I_2$ is the $2 \times 2$ identity matrix, we obtain

$$\hat{y} \triangleq H^H y = (|h_1|^2 + |h_2|^2)x + \tilde{n},$$

(2.31)
where \( \hat{n} \triangleq H^H \tilde{n} \). It can be easily verified that the elements of \( \hat{n} \) are still i.i.d. Therefore, the decision rule in (2.30) is equivalent to

\[
\hat{x} = \arg\min_{x \in \mathcal{X}^2} \left\{ \| \hat{y} - (|h_1|^2 + |h_2|^2)x \|_2^2 \right\}.
\] (2.32)

By using simple linear combining, the joint ML decoding for symbols \( \{x_0, x_1\} \) in (2.30) is decoupled into two individual decoding problems. The decoding complexity increases only linearly with the code size \( L \), thus, orthogonal STBCs (OSTBCs) are very attractive due to their linear decoding complexity.

For real constellations and square matrices \( B \), it was proven in [11] that rate one \((R = 1)\) codes can exist only for \( N_T = 2, 4, \) or \( 8 \) antennas. However, for complex constellations, orthogonal designs with \( R = 1 \) only exists for the case of \( N_T = 2 \). Given the number of transmit antennas \( N_T = 2m - 1 \), or \( N_T = 2m \), where \( m \) is a natural number, the maximum possible rate for complex orthogonal designs was shown to be \( R = \frac{m+1}{2m} \) [62].

### 2.2.3 Closed–Loop Technique I – Spatial Multiplexing via SVD

If the CSI is perfectly known at the transmitter and the receiver, by performing SVD [4], the channel matrix \( H \) can be written as

\[
H = U \Sigma V^H,
\] (2.33)

where \( U \) and \( V^H \) are \( N_R \times N_R \) and \( N_T \times N_T \) unitary matrices, respectively. The entries of the \( N_R \times N_T \) diagonal matrix \( \Sigma \), \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{N_{\min}} \geq 0 \) are the sorted non-negative
singular values with $N_{\text{min}} = \min(N_T, N_R)$. Thus, (2.18) can be rewritten as

$$\mathbf{y} = \mathbf{U} \Sigma \mathbf{V}^H \mathbf{x} + \mathbf{n},$$

(2.34)

and the channel can be expressed as

$$\mathbf{H} = \sum_{i=1}^{N_{\text{min}}} \sigma_i \mathbf{u}_i \mathbf{v}_i^H,$$

(2.35)

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. Let $\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{y} \triangleq [\tilde{y}_1 \tilde{y}_2 \cdots \tilde{y}_{N_R}]^T$, $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x} \triangleq [\tilde{x}_1 \tilde{x}_2 \cdots \tilde{x}_{N_T}]^T$, and $\tilde{\mathbf{n}} = \mathbf{U}^H \mathbf{n} \triangleq [\tilde{n}_1 \tilde{n}_2 \cdots \tilde{n}_{N_R}]^T$, then the channel in (2.34) is equivalent to

$$\tilde{\mathbf{y}} = \Sigma \tilde{\mathbf{x}} + \tilde{\mathbf{n}},$$

(2.36)

where $\tilde{\mathbf{n}}$ still has the same distribution as $\mathbf{n}$. Thus, by applying the weighting matrices $\mathbf{V}$ and $\mathbf{U}^H$ at the transmitter and the receiver respectively, we obtain equivalent parallel single–input–single–output (SISO) channels

$$\tilde{y}_i = \sigma_i \tilde{x}_i + \tilde{n}_i, \quad 0 \leq i \leq N_{\text{min}} - 1,$$

(2.37)

where $\tilde{y}_i$, $\tilde{x}_i$, and $\tilde{n}_i$ are the received symbols, the transmitted symbols, and the AWGN noise samples over the $i$th equivalent SISO channel, respectively.

Assuming the components of $\mathbf{x}$ are i.i.d., the capacity of the channel [4] is proven to be

$$C = \sum_{i=0}^{N_{\text{min}}-1} \log \left( 1 + \frac{P_i^* \sigma_i^2}{N_0} \right),$$

(2.38)

where $N_0$ is the noise energy per complex symbol time and $P_i^*$ is the power allocated to the $i$th SISO channel. This capacity is achieved by allocating power according to the
waterfilling algorithm \[4\]:

\[
P^*_i = \left( \mu - \frac{N_0}{\sigma^2_i} \right)^+ ,
\]

where \(\mu\) is chosen to meet the total power constraint \(\sum_i P^*_i = P_{\text{total}}\) and \(\sigma_i\) corresponds to the gain of an eigenmode of the channel.

### 2.2.4 Closed–Loop Technique II – Combining Precoding with STBC

While the channel can be measured directly at the receiver with sufficient accuracy, the transmitter must obtain channel information indirectly, using either channel reciprocity or feedback, often accompanied by a delay, measurement noise, and quantization errors. Thus, in practice, the transmitter often obtains imperfect CSIT or partial CSIT, which leads to performance degradation in a wireless system. However, imperfect CSIT can be exploited together with the reliability of the CSIT to achieve high performance \[14\].

One example of successful exploitation of imperfect CSIT is a system configuration combining a precoder with STBC, where the STBC exploits channel diversity and the linear precoder exploits the CSIT. The STBC and precoder combination, therefore, is robust to channel fading and can exploit the available CSIT at the same time. Fig. 2.5 shows the structure of a system combining a linear precoder with STBC \[15\].

Let the \(T \times N_T\) matrix \(B\) be the block codeword in the STBC, then the received
signal block is
\[ Y = HFB + N, \] (2.40)

where \( F \) is the linear precoding matrix satisfying the power constraint \( \text{tr}(FF^H) = 1 \) and \( N \) is the AWGN matrix. ML detection is performed at the receiver to obtain the estimate of the codeword
\[ \hat{B} = \arg\min_{\hat{B} \in B} ||Y - HFB||_F^2, \] (2.41)

where \( B \) is the STBC codebook.

The PEP of detecting the transmitted codeword \( B \) as another codeword \( \hat{B} \) can be upper bounded by
\[ P(B \rightarrow \hat{B}) \leq \exp\left( -\frac{||HF(B - \hat{B})||_F^2}{4\sigma_n^2} \right). \] (2.42)

This upper bound can be rewritten as
\[ f(H, A, F) = \exp\left( -\frac{\rho}{4} \text{tr}(HFAF^H \hat{H}^H) \right), \] (2.43)

where \( \rho = P/\sigma_n^2 \) is the SNR and \( A \) is defined as \( \frac{1}{P}(B - \hat{B})(B - \hat{B})^H \) with the average sum transmit power \( P \).

The imperfect CSIT and the true CSI are assumed to be jointly complex Gaussian, and the pdf of the true channel vector \( h \) conditioned on the side information \( \gamma \) is given by [14]
\[ p_{h|\gamma}(h|\gamma) = \frac{\exp\left( -(h - m_{h|\gamma})^H C_{hh|\gamma}^{-1}(h - m_{h|\gamma}) \right)}{\pi^N \det(C_{hh|\gamma})}, \] (2.44)

where \( m_{h|\gamma} = \mathcal{E}\{h|\gamma\} \) is the conditional mean vector and \( C_{hh|\gamma} = \mathcal{E}\{(h - m_{h|\gamma})(h - m_{h|\gamma})^H|\gamma\} \) is the conditional covariance matrix.
2. Preliminaries

The optimal precoder for the case with uncorrelated antennas and the transmitter knowing the channel means is provided in [14] for OSTBCs. The precoder design for the more general cases with antenna correlation remains an open question. For the special case with only transmit antenna correlation and the CSIT in the form of an arbitrary mean, [15] formulated the precoder optimization problem and provided a solution.

Suppose the channel mean $H_m$ and the transmit correlation matrices $C_{N_T}$ are known to the transmitter. The receive antennas are assumed to be uncorrelated. Averaging (2.43) over the channel statistics, we obtain the following bound on the average PEP [15]:

$$P_e \leq \exp\left(\frac{\text{tr}(H_m W^{-1} H_m^H)}{\det(W)^{N_R}}} \right) \det(C_{N_T})^{N_R} \exp\left(\text{tr}(H_m W^{-1} H_m^H)\right),$$  \hspace{1cm} (2.45)

where $W = \frac{\eta}{4} C_{N_T} F A F^H C_{N_T} + C_{N_T}$. Taking the logarithm of this bound and ignoring the constant terms leads to the following objective function

$$J(F) = \text{tr}(H_m W^{-1} H_m^H) - N_R \log \det(W)$$  \hspace{1cm} (2.46)

which is convex in the matrix variable $W$. For OSTBC, $A$ is a scaled identity matrix, which helps to simplify the optimization problem into the following form:

$$\min_F J(F)$$

s.t. \hspace{1cm} $W = \eta_0 C_{N_T} F F^H C_{N_T} + C_{N_T}$, \hspace{1cm} $\text{tr}(F^H F) = 1$

where $\eta_0 = 1/(4\mu_0\rho)$.

To solve the optimization problem, define $\Phi(W) = C_{N_T}^{-1} W C_{N_T}^{-1} - C_{N_T}^{-1}$. Now, the
constraints in (2.47) can be rewritten as

\[
\text{tr}(\Phi) = \eta_0 \quad (2.48a)
\]
\[
\Phi \succeq 0. \quad (2.48b)
\]

Inequality (2.48b) means that \( \Phi \) is a positive semi–definite matrix. By using the Lagrangian dual method, the solution for \( W \) is obtained as

\[
W = \frac{1}{2\nu} C_{N_T}(N_R I_{N_T} + \Psi^{1/2})C_{N_T},
\]

where \( \Psi = N_R^2 I_{N_T} + 4\nu C_{N_T}^{-1} H_m H_m^H C_{N_T}^{-1} \). Here, \( \nu \) is a parameter to be computed based on the transmit power constraint. \( \Phi \) can now be expressed as a function of \( \nu \)

\[
\Phi(\nu) = \frac{1}{2\nu} (N_R I_{N_T} + \Psi^{1/2}) - C_{N_T}^{-1}.
\]

The full–mode solution is obtained by solving for \( \nu \), assuming \( \Phi \) is full rank. If the full–mode solution does not meet the requirement of \( \Phi \succeq 0 \), then the weakest eigenmode is dropped, forcing its smallest eigenvalue to be zero. This procedure of dropping the weakest eigenvalues continues until both constraints in (2.48a) and (2.48b) are met simultaneously. When \( \nu \) is determined, the optimum solution for \( F \) is given by

\[
F = \frac{1}{\sqrt{\eta_0}} U \Sigma^{1/2} V^H,
\]

where the matrices \( U \) and \( \Sigma \) contain the left singular vectors and the singular values of \( \Phi \), respectively. Here, \( V \) can be an arbitrary unitary matrix. From (2.51) the optimal precoder is not unique and \( V \) can be identity matrix for simplicity.
2.3 Virtual MIMO Techniques

In the previous section, we have discussed various processing techniques for multiple antennas. However, in applications such as wireless sensor network and *ad hoc* network, implementation of multiple antennas is not always possible due to space limitation or cost consideration. Virtual MIMO techniques provide us with an alternative means to achieve a diversity gain or to improve the outage performance.

2.3.1 Cooperative Diversity

The basic idea of cooperative diversity is to allow a collection of nodes to share their antennas to emulate a virtual antenna array and exploit spatial diversity in wireless fading channels. In a cooperative network, the source node transmits data symbols to the destination nodes with the help of relays. A diversity order equal to the number of cooperative nodes can be achieved.

Various cooperative transmission protocols have been proposed [20, 24]. The two-phase protocol is the most popular one, where in the first phase, the source node broadcasts the information to the relay nodes. Then, in the second phase, the relay nodes either decode–and–forward (DF) or amplify–and–forward (AF) the received signals to the destination. At the destination, the received signals are combined to obtain a diversity gain via certain signal processing techniques. For example, the distributed space–time coding (DSTC) [20, 63, 64] and distributed space–time filtering (DSTF) [65] schemes employ the two-phase protocol, and these codes have to be applied at the transmitters.

The relay nodes can be active or non–active depending on whether they participate in the cooperation or not. Furthermore, depending on the level of cooperation between the relay nodes, two different approaches can be distinguished. In centralized schemes, the cooperative nodes know their partners, and numerous existing STC schemes for
co-located antennas can be directly applied. In decentralized schemes, the cooperating nodes are not required to know their partners. Thus, decentralized schemes may be more flexible and practical, especially in application such as wireless sensor networks where sensors decide independently about their own actions.

### 2.3.2 Channel Augmentation

Channel augmentation is another class of virtual MIMO techniques, where virtual antennas are created to increase the effective rank of the channel matrix by means of special signal processing methods. Channel augmentation can be employed to improve the outage performance of a wireless system. For instance, in a communication link, where the number of transmit antennas exceeds the number of receive antennas, i.e. \( N_R < N_T \), layered space–time architectures such as V–BLAST cannot be applied straightforwardly since these schemes typically require \( N_T \leq N_R \). In this case, an augmented MIMO channel can be created to obtain more virtual receive antennas than the original channel has.

In the recently proposed RTM scheme [31], the information outage rate is improved by simple signal processing of independently faded received signals at the receiver. Suppose a standard uncorrelated MIMO Rayleigh fading channel model with received vector \( \mathbf{y} \) modeled as

\[
\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{n},
\]

where \( \mathbf{x} \) is the transmitted data with maximum power \( P = \mathcal{E}\{\mathbf{x}^H \mathbf{x}\} \), \( \mathbf{H} \) is the \( N_R \times N_T \) channel matrix, and \( \mathbf{n} \) is the noise vector. \( \mathbf{H} \) and \( \mathbf{n} \) are independent and both have i.i.d. elements taken from a complex zero–mean Gaussian distribution with unit variance. The channel is assumed to be known only at the receiver. At the transmitter, the same data are sent repeatedly in several time slots, which are assumed to fade independently. For
instance, if the data are transmitted in two time slots, the received signals are given by

\[ y_1 = H_1 x + n_1, \]  
\[ y_2 = H_2 x + n_2, \]

where \( H_\nu \) and \( n_\nu, \nu = \{1, 2\} \), are the channel matrix and the noise vector in time slot 1 and 2, respectively, and they all mutually independent. At the receiver, to increase the effective number of receive antennas, the signals received in two time slots are stacked on top each other to form

\[
z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} x + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = G x + v \]

where \( G \) has size \( 2N_T \times N_R \) and the SNR is still the same as in (2.52). Thus, if \( N_T \leq N_R/2 \), the channel rank has been doubled. However, the ergodic capacity is not improved since the data have been transmitted twice. In fact, the ergodic capacity is even slightly reduced [31].

The information rate of the augmented channel \( G \), which is known only at the receiver, is given by

\[
I = \frac{1}{N_V} \log_2 \left( \det \left( I + \frac{P}{N_T} GG^H \right) \right),
\]

where \( N_V \) is the number of slots for repeating the same signals. The virtual antenna technique increases the size and rank of the augmented channel matrix with a corresponding decrease in data rate. Simulation results show that the improvement in outage rate is more evident at low outage probabilities (\( \leq 10^{-2} \)) than at higher outage probabilities (\( \geq 10^{-1} \)).

Inspired by the RTM scheme, we have proposed two novel channel augmentation
schemes, which can achieve a higher outage rate than the RTM scheme with the same low complexity. We will introduce both schemes in detail in Chapter 3.

2.4 OFDM Basics

The principle of OFDM is to divide a high rate data stream into a number of lower rate streams that are transmitted simultaneously over a number of channels (subcarriers) that experience flat fading. These subcarriers are chosen to be orthogonal to each other so that no guard bands are required. The resulting channel is a sum of orthogonal subchannels. OFDM is thus capable of achieving a high spectral efficiency, while at the same time being able to combat multipath fading. Furthermore, a CP is inserted in every OFDM symbol to eliminate inter–symbol interference (ISI) and inter–carrier interference (ICI).

2.4.1 Orthogonality of Subcarriers

In an OFDM system, the frequency spacing between adjacent subcarriers is \( \Delta f = 1/(KT) \), where the number of the subcarriers is \( K \) and the overall symbol rate is \( 1/T \). With this frequency spacing, the subchannels can maintain orthogonality, although the subchannels overlap. Therefore, ideally, there is no intersymbol and intercarrier interference in OFDM systems.

The discrete–time representation of the baseband transmitted signals for one OFDM symbol period can be expressed as

\[
s[n] = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} a[k] e^{j2\pi kn/K}, \quad 0 \leq n \leq K - 1,
\]  

(2.57)

where \( a[k] \) is the transmitted data symbol over the \( k \)th subcarrier. Obviously, the com-
plexity of the modulation/demodulation increases with increasing the number of subcarriers. However, instead of carrying out the modulation/demodulation on a subcarrier by subcarrier basis, all OFDM subcarriers are modulated and demodulated in a single IFFT/FFT step efficiently. For example, using the radix-2 algorithm, a $K$-point IFFT requires only $(K/2) \log_2 K$ complex multiplications. The number of complex multiplication in the IFFT can be further reduced by using a radix-4 algorithm [66].

At the receiver, a spectral decomposition of the received time domain samples is performed employing a $K$-point FFT, and the recovered data symbols are restored in serial order and demultiplexed.

### 2.4.2 Guard Interval

If the delay spread of a frequency–selective fading channel is relatively large compared to the sampling interval, ISI will inevitably occur. To eliminate this undesirable effect, a guard interval is introduced for each OFDM symbol. The guard time is chosen larger than the expected delay spread so that no OFDM symbol is interfered by its previous symbol.

Typically, OFDM systems employ a CP guard interval $T_g$ to protect against ISI as well as maintaining the orthogonality among subcarriers. The CP is formed by padding the start of the OFDM symbol using a copy of some fraction of the samples from the end of the symbol as shown in Fig. 2.6. This ensures that delayed replicas of the OFDM symbol always have an integer number of cycles within the FFT intervals, as long as the delay is smaller than the guard time. Therefore, multipath signals with delays smaller than the guard time do not cause ISI. At the receiver, the prefix samples are discarded, and the remainder of the samples are converted back into the frequency domain using the FFT. In this way, the CP transforms linear convolution to cyclic convolution in the
time domain, which is converted to a simple multiplication in the frequency domain by the FFT at the receiver. In the IEEE 802.16e standard [6], the length of CP is chosen between 1/32 to 1/4 of an OFDM symbol interval.

Some systems have adopted zero–padding (ZP) instead of CP. In ZP, the end of OFDM symbols are postfixed with a number of “0” guard samples. The transmitter keeps silent during ZP interval and thus does not consume energy. Another benefit of ZP is that the spectrum of the ZP signal does not have “spikes”, which normally appear in an OFDM system with CP due to its periodic nature. The second benefit is especially favored by some systems with power spectrum density (PSD) limits, e.g., the ultra–wideband (UWB) system. However, the use of ZP introduces additional complexity of signal processing at the receiver side [67]. Currently most systems employ CP as the guard interval, thus we consider only CP guards in this thesis.

2.4.3 BICM–OFDM

It has been shown that code diversity can be improved by bit–wise interleaving [68]. Caire et al. presented the theory behind BICM in [57]. Their work provided tools to evaluate the performance of BICM with tight bounds and design guidelines. BICM can be combined with OFDM, where coding and modulation are separated by a bit-level
interleaver. For the decoding of the received signal, it is necessary to obtain a soft-bit metric for the input of the Viterbi decoder.

Fig. 2.7 shows a block diagram of BICM–OFDM. We consider a system with only one transmit and one receiver antenna (SISO). Each OFDM symbol has $K$ subcarriers. At the transmitter, a convolution encoder with minimum free distance $d_{\text{free}}$ is used to generate the binary code. The $N_c$ encoded bits $c_{k'}$, $1 \leq k' \leq N_c$, are then interleaved and mapped onto symbols $X[k], 0 \leq k \leq K - 1$. The symbols are taken from a constellation $\mathcal{X}$ with constellation size $|\mathcal{X}| = 2^m$. A CP of appropriate length is then added to each OFDM symbol before transmission. At the receiver, after CP removal and FFT operation, the received signal is given by

$$Y[k] = H[k]X[k] + N[k], \quad 0 \leq k \leq K - 1, \quad (2.58)$$

where $H[k]$ and $N[k]$ are the channel gain and the AWGN for the $k$th subcarrier, respectively. For frequency selective channels with impulse response taps $h[l], 0 \leq l \leq L - 1$, $H[k]$ can be computed by

$$H[k] = \sum_{l=0}^{L-1} h[l] e^{j2\pi kl/K}, \quad 0 \leq k \leq K - 1. \quad (2.59)$$
Based on the received signal, the demapper computes the metric of each coded bit \([44, 48]\)

\[
\lambda_i(Y[k], c_k') = \min_{X \in \mathcal{X}_i} \left\{ |Y[k] - H[k]|^2 X^2 \right\},
\]

(2.60)

where \(\mathcal{X}_i^b\) denotes the subset of all symbols \(X \in \mathcal{X}\) whose label has the value \(b \in \{0, 1\}\) in position \(i\). The metrics are de-interleaved and then sent to a Viterbi decoder. It was shown in [69] that for an \(L\)-tap frequency selective channel, BICM-OFDM can achieve a diversity order of \(\min(d_{\text{free}}, L)\) in a SISO system.

### 2.4.4 Adaptive Loading

When CSIT is available, the transmit power and the bits to be transmitted for each subcarrier can be adapted according to the CSIT in order to increase data rate or decrease transmit power consumption. The scheme employed to adaptively allocate transmit power and bits for each subcarrier, known as adaptive loading, is an important aspect of the design of an OFDM system over ADSL, and has recently been extended to wireless OFDM systems. Different strategies can be applied for adaptive loading [35–37, 70].

It is well known that the waterfilling approach is the capacity–achieving power allocation algorithm for a spectrally shaped channel [71]. However, the waterfilling algorithm assumes infinite granularity in constellation size, which is not realizable in practical communication systems. One finite–granularity loading algorithm, known as Hughes–Hartogs algorithm [72], mimics the waterfilling approach and yields a near optimal solution by using an incremental strategy. The Hughes–Hartogs algorithm creates a table of incremental energy required to transmit one additional bit on each of the subcarriers. Then, at each step, one more bit is added to the subcarrier that requires the least incremental energy, until all bits are assigned or the total power is exceeded depending on
the predefined criterion. The extensive sorting needed renders this algorithm impractical for an OFDM system with a large number of subcarriers. The more efficient algorithms proposed by Li et al. [70] and Piazzo [40] also use incremental strategy.

One of the solutions that omit the extensive sorting in the Hughes–Hartogs algorithm is the capacity–based strategy in [36], which approximates the channel capacity of the subcarriers. Let $P[k]$ and $\sigma_n^2[k]$ be the transmit power and the noise variance on subcarrier $k$, respectively. Then, the rate $r[k]$ for this subcarrier is given by

$$r[k] = \log_2 \left( 1 + \frac{P[k]}{\sigma_n^2[k] \cdot \Gamma \gamma} \right), \quad (2.61)$$

where $\Gamma$ is a fixed bias and $\gamma$ is a parameter iteratively decided such that the sum rate over all subcarrier, $\sum_k r[k]$, equals the target rate $R_{\text{total}}$. $r[k]$ has to be rounded to the nearest integer for bit allocation in such a way that the overall quantization error is minimal. The total transmit power $P_{\text{total}}$ is then distributed among the subcarriers such that the error performances of all subcarriers are identical and $\sum_k P[k] = P_{\text{total}}$.

The problem with the capacity–based strategy is that the power and bit allocation which approximates the capacity is not necessarily the optimum solution for a system with a specific coding and modulation scheme. Bits and power can also be allocated using closed–form expression for the probability of error, given the SNR. For instance, the probability of error for uncoded QAM on subcarrier $k$ is given by [1]

$$P(M_k, \gamma_k) = 4 \left( 1 - \frac{1}{\sqrt{M_k}} \right) Q \left( \frac{3\gamma_k}{M_k - 1} \right) \left( 1 - \left( 1 - \frac{1}{\sqrt{M_k}} \right) Q \left( \frac{3\gamma_k}{M_k - 1} \right) \right), \quad (2.62)$$

where $M_k$ and $\gamma_k$ are the constellation size and the average symbol energy, respectively.
2. Preliminaries

$Q(\cdot)$ is the $Q$-function defined as

$$Q(x) \triangleq \frac{1}{2\pi} \int_{x}^{\infty} e^{-t^2/2} dt. \quad (2.63)$$

Subject to a total rate and power constraint, Fischer and Huber [37] distribute the bits and power to minimize the error probability on each subcarrier. $\log_2(M_k)$ gives a non-integer number of bits. After simplifying approximations, $M_k$ can be determined and rounded to the nearest integer.

2.5 MIMO–OFDM

The main motivation behind using MIMO–OFDM is the fact that OFDM modulation turns a frequency–selective MIMO channel into a set of parallel flat MIMO channels. This greatly simplifies the multi–channel equalization at the receiver as for each subcarrier only a constant matrix has to be inverted. While spatial multiplexing increases spectral efficiency, SFC is used to improve link reliability through a diversity gain. SFC spreads the symbols across space and frequency to exploit space and frequency diversity in a systematic way. In this section, we will introduce CDD, a particular simple and efficient form of SFC.

2.5.1 CDD

The block diagram of CDD is depicted in Fig. 2.8. The information bits are encoded by a forward error correction (FEC) encoder, and are then interleaved and mapped on quadrature amplitude modulation (QAM) or phase-shift keying (PSK) symbols, which are assigned to the $K$ subcarriers of an OFDM symbol. A $K$-point IFFT converts these symbols from the frequency domain into the time domain. The time domain symbols
are denoted as $\tilde{x}[k], 0 \leq k \leq K - 1$. Before inserting the CP as guard interval, the time-domain OFDM symbol is shifted cyclically by $\Delta_n, 1 \leq n \leq N_T$, for transmit antenna element $n$. This yields the antenna specific signals:

$$x_n[k] = \frac{1}{\sqrt{K}} \tilde{x}[(k - \Delta_n) \mod K], \quad 0 \leq k \leq K - 1, \quad 1 \leq n \leq N_T. \quad (2.64)$$

We can simply set the cyclic shift for the first antenna to $\Delta_1 = 0$. In contrast to the cyclic shift $\Delta_n$ in the case of CDD, conventional DD simply shifts the original data signals in time domain by antenna specific delays.

For simplicity, we consider a single receive antenna here. The CIR from transmitter $n$ to the receive antenna is given by

$$q_n[k] = \sum_{l=0}^{L-1} h_n[l] \delta[k - l], \quad (2.65)$$

where the length of the channel is $L$ and $h_n[l]$ is the channel gain of the $l$th tap for transmit antenna $n$. We assume that the channel remains constant during one OFDM symbol interval.

At the receiver, the CP is first removed before FFT is performed. The resulting
signal can be expressed by

\[ R[k] = H_{eq}[k]X[k] + N[k], \quad 0 \leq k \leq K - 1, \]  

where \( N[k] \) is an AWGN sample and the channel coefficient \( H_{eq}[k] \) is given by

\[ H_{eq}[k] = \sum_{n=1}^{N_T} \sum_{l=0}^{L-1} h_n[l] e^{-j2\pi k(l - \Delta_n)/K}. \]  

Eq. (2.67) shows that CDD transforms the multiple input channel into an equivalent single input channel \( H_{eq}[k] \) with increased frequency selectivity. The spatial diversity has been transformed into frequency diversity, which can be picked up by FEC coding. For the commonly used convolutional code (CC), the minimum free distance \( d_{\text{free}} \) determines the maximum diversity level, where ideally successive bits experience independent fading. For this purpose, BICM and CCs with a large \( d_{\text{free}} \) are favored in order to exploit maximum frequency diversity.

Selection of the cyclic delay is an important issue that affects the performance of CDD. The cyclic delay should be chosen such that an FEC decoder can exploit the full spatial diversity inherent in the channel. To preserve the maximum diversity, the equivalent channel taps in (2.67) should not contain sums of taps at different delays, therefore, the cyclic delays should satisfy

\[ \Delta_n \geq \Delta_{n-1} + L, \quad 2 \leq n \leq N_T. \]  

A safe choice for the cyclic delay is [51]

\[ \Delta_n = \frac{(n - 1)K}{N_T}, \quad 1 \leq n \leq N_T, \]
which yields the largest possible delay difference and low correlation between adjacent subcarriers.

Compared with OSTBC, CDD has the following major advantages:

1. CDD enjoys low complexity at both the transmitter and receiver. For OSTBC, an extra IFFT at each transmit antenna and a diversity combiner at the receiver are required. CDD uses only one IFFT at the transmitter independent of the number of transmit antennas, and there is no additional processing apart from channel estimation at the receiver.

2. CDD provides a very flexible space–frequency coding approach for any number of transmit antennas. The number of transmit antennas can even be changed without changing the codes used and the receiver structure. Thus, CDD can be easily incorporated into existing communication systems. For OSTBC, the number of transmit antennas can not be easily changed once it is determined.

3. There is no rate loss for CDD compared to a single transmit antenna OFDM. In contrast, OSTBCs suffer from a rate loss in a system using more than two transmit antennas, cf. Section 2.2.2.

The performance of CDD can be improved by its generalized version, generalized CDD (GCDD) [73], or CSF filtering [74], which is one of the main contributions of this thesis. We will discuss CSF filtering in detail in Chapters 5 and 6.
3 Two Novel Channel Augmentation Schemes for MIMO Systems

3.1 Introduction

In this chapter, we introduce the proposed two novel channel augmentation schemes for MIMO systems. Channel augmentation schemes belong to a class of virtual MIMO techniques, where special signal processing methods are applied at the transmitter and the receiver to create virtual antennas, leading to improved outage performance.

Generally, if the number of transmit antennas significantly exceeds the number of receive antennas, STC [12] is usually preferred to BLAST–type processing. However, recently, Rankin, Taylor, and Martin (RTM) [31] have shown that the information outage rate of MIMO channels can be potentially improved by repeatedly transmitting the same BLAST–type signal. This is especially true for the case where the number of transmit antennas is larger than the number of receive antennas. In fact, stacking the received signal vectors corresponding to the same transmitted symbols results in an augmented

---

1The material in this chapter was presented in part at IEEE Wireless Communications and Networking Conference, Hong Kong, China, March 2007.
MIMO channel with an increased number of virtual receive antennas. This also facilitates the application of suboptimum processing techniques typically used in BLAST receivers, cf. Section 2.2.1. The concept of creating virtual receive antennas is not new and has been shown to be an effective technique for performance improvement in other contexts before, cf. [29, 75].

In this chapter, we introduce two novel channel augmentation techniques and make the following contributions.

- The first novel scheme is similar to the RTM technique. However, instead of simply repeating the transmitted symbols, the symbols are also complex conjugated leading to the conjugate virtual antenna (CVA) scheme. The CVA scheme has the same complexity as the RTM scheme but yields a better performance if the channels seen by the repeated symbols are correlated.

- The second novel scheme restricts the transmitted symbols to be real–valued. By processing both the real and the imaginary parts of the received signal the number of receive antennas is essentially doubled. We refer to this approach as real virtual antenna (RVA) scheme. We will also investigate the relation of the RVA scheme to so–called widely linear (WL) processing [76, 77].

- While [31] only considered the information outage rate for optimum transmitter and receiver processing, we investigate the impact of channel augmentation on the information outage rates of various optimum and suboptimum transmitter and receiver structures. In particular, we consider V–BLAST and D–BLAST with linear minimum mean–squared error (LMMSE) and MMSE successive interference cancellation (MMSE–SIC) receivers.

- Our simulation results confirm that channel augmentation is an efficient approach
to increase the information outage rate for both optimum and suboptimum transmitter/receiver processing and show that the CVA and the RVA schemes achieve a better performance than the RTM scheme in correlated fading.

The remainder of this chapter is organized as follows. In Section 3.2, the considered system model is introduced. In Section 3.3, the RTM scheme is briefly reviewed and the two proposed channel augmentation schemes are presented. The outage performance of all three schemes is discussed for different transmitter and receiver structures in Section 3.4. Simulation results are presented in Section 3.5, and some conclusions are drawn in Section 3.6.

### 3.2 System Model

We consider a coded MIMO system with $N_T$ transmit antennas and $N_R$ receive antennas. In the $k$th coding frame the received $N_R \times 1$ vectors $y_k$ can be expressed as

$$y_k = H_k x_k + n_k,$$

(3.1)

where the $N_T \times 1$ vectors $x_k$ and $n_k$ contain the transmitted data symbols and rotationally symmetric i.i.d. AWGN samples with variance $\sigma_n^2$, respectively.

The channel matrix $H_k$ has size $N_R \times N_T$, where the $(m, n)$th element $h_{mn}[k]$ represents the path gain between the $n$th transmit antenna and the $m$th receive antenna. We assume that $H_k$ does not change during a coding frame (quasi–static channel) and that the path gains $h_{mn}[k]$ are spatially uncorrelated and complex Gaussian distributed with variance $\sigma_h^2 \equiv \mathcal{E}\{|h_{mn}[k]|^2\} = 1$. Furthermore, the path gains corresponding to different coding frames may be correlated. Since different coding frames may correspond to different frequencies or different time slots or both, we use here a generic exponen-
3. Two Novel Channel Augmentation Schemes for MIMO Systems

The power of the transmitted symbol vector is constrained to $E\{x_k^H x_k\} = P$. Furthermore, since we consider the scenario in which the instantaneous CSI is only available to the receiver but not to the transmitter, the total power is equally distributed over all transmit antennas.

In V–BLAST the $N_T$ elements of all $x_k$ corresponding to the same coding frame are separately encoded. In contrast, in D–BLAST all elements of $x_k$ are jointly encoded, cf. Section 2.2.1. In this chapter, we will use the information outage rate as a performance measure for comparing different channel augmentation schemes. Furthermore, since we are primarily interested in an information theoretical comparison, we assume that the elements of $x_k$ are i.i.d. real or complex Gaussian distributed.

3.3 Channel Augmentation Schemes

We first briefly review the RTM scheme before we present two novel channel augmentation schemes, which have the same low computational complexity.

3.3.1 Rankin, Taylor, and Martin (RTM) Scheme

In the RTM scheme [31], also cf. Section 2.2.4, the same complex Gaussian symbol vector $x$ is transmitted in $K_f$ coding frames to create $N_V = N_R K_f$ virtual receive antennas, i.e., $x_k = x$, $1 \leq k \leq K_f$. By stacking the related received vectors $y_k$, $1 \leq k \leq K_f$, we obtain

$$y = Gx + n.$$  \hfill (3.2)
where

\[ y \triangleq [y_1^T \ y_2^T \ldots \ y_{K_f}^T]^T, \quad (3.3) \]

\[ n \triangleq [n_1^T \ n_2^T \ldots \ n_{K_f}^T]^T, \quad (3.4) \]

\[ G \triangleq [H_1^T \ H_2^T \ldots \ H_{K_f}^T]^T. \quad (3.5) \]

Eqs. (3.2) and (3.5) show that the RTM scheme creates an augmented \( N_V \times N_T \) channel matrix \( G \) corresponding to \( N_V \) virtual receive antennas at the expense of a \( K_f \)-fold decrease in data rate. It was shown in [31] that for many practically relevant cases, the increased rank of the channel matrix outweighs the decrease in data rate as far as the information outage rate is concerned.

### 3.3.2 Conjugate Virtual Antenna (CVA) Scheme

The CVA scheme is similar to the RTM scheme. However, instead of simply repeating the same signal vector \( K_f \) times, we also allow complex conjugation, i.e., \( x_{2k-1} = x \), \( 1 \leq k \leq K_f/2 \), and \( x_{2k} = x^* \), \( 1 \leq k \leq K_f/2 \), where \( K_f \) is assumed to be even. In order to get a simple equivalent channel model, the received vectors corresponding to the complex conjugated transmitted vectors are complex conjugated at the receiver. Therefore, the resulting augmented channel model is still given by (3.2). However, now the definitions

\[ y \triangleq [y_1^T \ y_2^H \ y_3^T \ldots \ y_{K_f}^H]^T, \quad (3.6) \]

\[ n \triangleq [n_1^T \ n_2^H \ n_3^T \ldots \ n_{K_f}^H]^T, \quad (3.7) \]

\[ G \triangleq [H_1^T \ H_2^H \ H_3^T \ldots \ H_{K_f}^H]^T. \quad (3.8) \]
are valid. Since the elements of $n_k$, $1 \leq k \leq K_f$, are i.i.d. rotationally symmetric complex Gaussian random variables, the noise vectors $n$ in (3.5) and (3.8) have the same statistical properties. Similarly, if the channel matrices $H_k$ in different coding frames are uncorrelated, the augmented channel matrices $G$ in (3.5) and (3.8) have the same statistical properties and the RTM and CVA schemes are equivalent and result in the same information outage rate. However, when the channel matrices $H_k$, $1 \leq k \leq K_f$, are correlated, i.e., $\rho > 0$, the CVA scheme is superior as will be shown in Section 3.5.

### 3.3.3 Real Virtual Antenna (RVA) Scheme

In the RVA scheme, the transmit symbol vector is constrained to have real-valued entries. To create $N_V = K_f N_R$ virtual antennas, we transmit the real-valued vector $x_R K_f/2$ times, where $K_f$ is assumed to be even. Since compared to the RTM and the CVA schemes only half of the number of time/frequency slots are required, we assume that only odd time/frequency slots $k$ are used for transmission to minimize the correlation of neighboring slots. The even slots can then be used for transmission of independent data. At the receiver, we stack the real parts and the imaginary parts of the received vectors $y_{2k-1}$, $1 \leq k \leq K_f/2$, and obtain the augmented channel model\(^2\)

$$
y = Gx_R + n_R, \tag{3.9}
$$

\(^2\)We note that the RVA scheme only leads to an augmented channel if the $H_k$, $1 \leq k \leq K_f$, are not purely real or purely imaginary. However, since their elements are complex Gaussian random variables, $H_k$, $1 \leq k \leq K_f$, are purely real or purely imaginary with probability zero.
where

\[
y \triangleq \begin{bmatrix} \Re\{y_1\}^T \Im\{y_1\}^T & \ldots & \Im\{y_{K_f-1}\}^T \end{bmatrix}^T,
\]
\[
n_R \triangleq \begin{bmatrix} \Re\{n_1\}^T \Im\{n_1\}^T & \ldots & \Im\{n_{K_f-1}\}^T \end{bmatrix}^T,
\]
\[
G \triangleq \begin{bmatrix} \Re\{H_1\}^T \Im\{H_1\}^T & \ldots & \Im\{H_{K_f-1}\}^T \end{bmatrix}^T.
\]

We note that the elements of the noise vector \( n_R \) are real–valued Gaussian random variables with variance \( \sigma_n^2 / 2 \).

In contrast to the RTM and CVA schemes, the RVA scheme only uses \( K_f / 2 \) time slots for transmission and thus makes more efficient use of the available resources in this respect. On the other hand, the transmission of real–valued data vectors instead of complex–valued data vectors uses only half of the available degrees of freedom offered by the channel.

The fact that real–valued transmit signals lead to an augmented channel matrix has been observed before in the WL signal processing literature, cf. e.g. [77, 78]. However, there the augmented channel is used to improve the performance of linear and non–linear equalizers whose filters are optimized based on second–order statistics. In contrast, here the purpose of the augmentation is to improve the information outage rate regardless of the receiver structure.

### 3.4 Information Outage Rate

In this section, we investigate the information outage rates of the three channel augmentation schemes introduced in the previous section. We consider both V–BLAST and D–BLAST architectures at the transmitter and at the receiver MMSE–SIC processing is assumed.
3.4.1 V–BLAST

In V–BLAST, the $N_T$ data streams transmitted over the $N_T$ antennas are independently encoded. Since the channel is unknown at the transmitter, each of the $N_T$ data streams has a rate of $R/N_T$ if the total data rate is $R$.

We first consider the channel model in (3.2), i.e., we assume RTM or CVA is used. In MMSE–SIC, for detection of data stream $i$, $1 \leq i \leq N_T$, the contributions of previously detected data streams $j < i$ are subtracted from $y$ and the resulting signal is filtered with a linear filter vector $f_i, 1 \leq i \leq N_T$, of length $N_V$. If we denote the $N_T$ elements of $x$ by $x_i, 1 \leq i \leq N_T$, and assume that the decisions on the $i - 1$ previously detected data streams (layers) are correct, the MMSE estimate for $x_i$ is given by \[ \hat{x}_i = f_i^H g_i x_i + \sum_{j=i+1}^{N_T} f_i^H g_j x_j + f_i^H n, \] (3.13)

where $g_i, 1 \leq i \leq N_T$, denotes the $i$th column of the augmented channel matrix $G$. The optimum MMSE filter vectors $f_i$ can be easily calculated, cf. e.g. [52]. Based on (3.13) it can be shown that the signal–to–interference–plus–noise ratio (SINR) of the $i$th transmitted signal stream is given by

\[ \text{SINR}_i = \frac{P}{N_T} g_i^H \left( \frac{P}{N_T} \sum_{j=i+1}^{N_T} g_j g_j^H + \gamma \sigma_n^2 I_{N_V} \right)^{-1} g_i, \] (3.14)

where $\gamma = 1$. If the RVA scheme is used, (3.14) remains still valid if $\gamma = 1$ is replaced by $\gamma = 1/2$ and the vectors $g_i, 1 \leq i \leq N_T$, are the columns of the augmented channel matrix $G$ defined in (3.12). In general, the performance of MMSE–SIC depends on the order in which the $N_T$ data streams are processed. In this chapter, we use the optimum ordering which maximizes the minimum \( \text{SINR}_i \), $1 \leq i \leq N_T$. This optimum ordering is found by calculating $\min_{1 \leq i \leq N_T} \{ \text{SINR}_i \}$ for all possible orderings and selecting the one
which yields the maximum value for the minimum SINR.

From (3.13) we observe that for a given realization of $G$ the interference–plus–noise term is Gaussian distributed. Furthermore, for the RTM and the CVA schemes the $K_f$–fold repetition of the complex vector $x$ leads to a $K_f$–fold decrease in data rate. The same is true for the RVA scheme due to the $K_f/2$–fold repetition and the restriction to real–valued vectors $x_R$. Therefore, recalling that $x_i$, $1 \leq i \leq N_T$, are i.i.d. Gaussian symbols and assuming ideal decoding of $x_j$, $j < i$, the information rate for the $i$th data stream is

$$R_i = \frac{1}{K_f} \log_2(1 + \text{SINR}_i), \quad 1 \leq i \leq N_T, \quad (3.15)$$

for all three schemes. The instantaneous SINR and thus the instantaneous information rate $R_i$, $1 \leq i \leq N_T$, vary randomly with the channel. Since in V–BLAST the $N_T$ data streams are individually decoded, the decoding is successful only if the information rate $R_i$ supported by the channel is larger than the allocated rate $R/N_T$ for all $1 \leq i \leq N_T$. Therefore, the outage probability of V–BLAST with MMSE–SIC receiver is given by

$$P_{\text{out}} = \Pr \left\{ \min_{1 \leq i \leq N_T} \{R_i\} < R/N_T \right\}. \quad (3.16)$$

This will be used in Section 3.5 to compare the different channel augmentation schemes.

### 3.4.2 D–BLAST

In D–BLAST, the $N_T$ data streams transmitted over the $N_T$ antennas are jointly encoded and at the receiver joint decoding is applied. Therefore, an outage only occurs if the overall rate $R$ exceeds the sum of the $N_T$ information rates $R_i$, $1 \leq i \leq N_T$. Assuming that the rate loss due to initialization and termination is negligible, $R_i$ is given by (3.14) and (3.15), and the total information rate is $R = \sum_{i=1}^{N_T} R_i$. It is straightforward to show
that $R$ can also be expressed as [52]

$$R = \frac{1}{K_f} \log_2 \left( \det \left( I_{N_V} + \frac{P}{\gamma \sigma_n^2 N_T} GG^H \right) \right),$$

(3.17)

where $\gamma = 1$ and $\gamma = 1/2$ for RTM, CVA and RVA, respectively. Depending on the channel augmentation scheme $G$ is given by (3.5), (3.8), or (3.12). We note that the information rate in (3.17) is identical to the maximum information rate achievable with joint maximum–likelihood decoding as in [31].

The outage probability for a MIMO system with D–BLAST and MMSE–SIC receiver is given by

$$P_{out} = \Pr \{ R < R \},$$

(3.18)

where $R$ is given by (3.17).

### 3.5 Numerical Results

In this section, we evaluate the outage probabilities in (3.16) and (3.18) via Monte–Carlo simulation using 1,000,000 channel realizations. We compare the various channel augmentation schemes with the case when no channel augmentation is used. We refer to the latter case as the “baseline” scheme. Since the baseline scheme may be interpreted as a special case of the RTM or the CVA schemes with $K_f = 1$, (3.16) and (3.18) can still be used for calculation of the outage probability. Since the gains achievable with channel augmentation are largest if only one receive antenna is available, we assume $N_R = 1$ throughout this section, i.e., $N_V = K_f$ is valid.

In Fig. 3.1, we compare the outage probabilities of the different schemes for V–BLAST with MMSE–SIC receiver assuming $N_T = 4$ transmit antennas and different (average) SNRs. For RTM, CVA, and RVA $N_V = 4$ virtual antennas were created and
the correlation coefficient was $\rho = 0.9$. The baseline scheme does not perform well, especially at reasonably low outage probabilities (e.g. $P_{\text{out}} = 0.01$) since the number of transmit antennas exceeds the number of receive antennas. For $P_{\text{out}} < 0.2$ all channel augmentation schemes lead to significant performance improvements for all considered SNRs. However, the proposed CVA scheme outperforms the RTM scheme which is more affected by channel correlation. Although the rate loss due to repetition is smaller for the RVA scheme than for the RTM and CVA schemes, RVA performs better than the RTM scheme only at very high outage probabilities. This suggests that the restriction to real–valued transmit symbols causes a larger decrease in the information rate than repetition.

Figure 3.1: Outage probability vs. outage rate for V–BLAST with MMSE–SIC receiver and different channel augmentation schemes. $\rho = 0.9$ and $N_T = N_V = 4$. 
In Fig. 3.2, we investigate the impact of correlation on the various channel augmentation schemes if again V–BLAST with MMSE–SIC receiver is employed. \( N_T = N_V = 4 \), \( P_{\text{out}} = 0.01 \), and SNRs of 12 and 24 dB are assumed. For the considered case, all three schemes outperform the baseline scheme even for high correlations. While for \( \text{SNR} = 12 \) dB RTM and CVA have practically the same performance, for \( \text{SNR} = 24 \) dB the higher robustness of CVA against channel correlation becomes obvious. Since for the RVA scheme with \( N_V = 4 \) only one repetition is necessary, it is even more robust against channel correlations than the CVA scheme which requires three repetitions with complex conjugation. However, for the considered scenario the outage rate of RVA is still lower than that of CVA even for high correlations.

![Figure 3.2: Outage rate vs. correlation coefficient \( \rho \) for V–BLAST with MMSE–SIC receiver and different channel augmentation schemes. \( P_{\text{out}} = 0.01 \) and \( N_T = N_V = 4 \).](image)
The effect of the number of transmit antennas and the number of virtual receive antennas on outage performance is considered in Fig. 3.3 for optimum processing, i.e., for D–BLAST and an MMSE–SIC receiver. For comparison we have also included the outage rate for maximum–rate full–diversity OSTBCs [62]. An outage probability of \( P_{\text{out}} = 0.01 \), SNRs of (a) 0 dB, (b) 12 dB, and (c) 24 dB, and \( \rho = 0.6 \) were assumed. For channel augmentation the proposed CVA scheme was adopted and the curves for \( N_V = 1 \) correspond to the baseline scheme without augmentation. Fig. 3.3 shows that for the optimum transmitter and receiver structures the optimum value for \( N_V \) strongly depends on the SNR but is almost independent of \( N_T \). In particular, \( N_V = 6 \) and \( N_V = 2 \) are optimum for SNR = 0 dB and SNR = 12 dB for all considered \( N_T \). For SNR = 24 \( N_V = 2 \) is optimum for \( N_T > 2 \), while for \( N_T = 2 \) the baseline scheme with \( N_V = 1 \) performs slightly better. While OSTBCs achieve the same outage rate as the baseline scheme for \( N_T = 2 \), their performance deteriorates for \( N_T > 2 \) due to their inherent rate loss [62]. We note, however, that the decoding complexity of OSTBCs is lower than that of the MMSE–SIC receiver required for D–BLAST with channel augmentation.

### 3.6 Conclusions

In this chapter, we have presented two simple channel augmentation schemes referred to as CVA and RVA. We compared both schemes with the previously proposed RTM scheme for the V–BLAST and D–BLAST architectures. Our results show that both the CVA and the RVA schemes are more robust to channel correlations than the RTM scheme. Furthermore, while all considered channel augmentation schemes can achieve significantly higher outage rates than the baseline scheme without augmentation, the CVA scheme outperformed the RVA and RTM schemes for all considered cases. Therefore, the performance gains compared to the baseline scheme are larger for the suboptimum
V–BLAST structure than for the optimum (and more complex) D–BLAST structure.
4 On MIMO–OFDM with Coding and Loading

4.1 Introduction

In the previous chapter, we have introduced two novel channel augmentation schemes for MIMO systems. Starting from this chapter, we will focus on several aspects of signal design specific for MIMO–OFDM systems. We first present ABL schemes for coded MIMO–OFDM in this chapter.

MIMO–OFDM schemes with ABL have been extensively studied recently [70, 79–82] assuming different levels of CSI at the transmitter (perfect CSI at the receiver is assumed). Kühn et al. [79] and Li et al. [70] presented results for coded MIMO–OFDM with ABL based on the SVD of the MIMO channel in case of full CSI. In [80] eigen–beamforming is applied to uncoded transmission when only partial CSI is available. V–BLAST [3] processing is often employed if CSI is not available at the transmitter [81, 82]. This kind of MIMO processing without the need for CSI at the transmitter is particularly interesting for moderately mobile applications as envisaged in e.g. IEEE 802.16e OFDM systems.

\footnote{The material in this chapter was presented in part at IEEE International Conference on Computer Communications and Networks, Hawaii, USA, August 2007.}
In this chapter, we study pragmatic schemes for coded and bit–loaded MIMO–OFDM which do not require CSI for MIMO processing at the transmitter and only a low–rate feedback channel to perform ABL. Our contributions can be summarized as follows.

• We propose the application of wrapped space–frequency coding (WSFC), which is the space–frequency counterpart of wrapped space–time coding (WSTC) devised in [57], as efficient coding scheme for MIMO–OFDM without CSI. WSFC retains the simplicity of V–BLAST, but alleviates the problem of error propagation by means of a special formatting of coded symbols to transmitted symbols. Furthermore, we optimize the WSFC decision delay for the considered application.

• For V–BLAST with ABL we devise a simple method to increase the performance margin of the symbols corresponding to the antennas decoded first such that error propagation is mitigated.

• While compared to SVD–based MIMO–OFDM, WSFC–based and V–BLAST–based MIMO–OFDM requires already only little feedback for ABL, we explore further feedback reduction using subchannel grouping methods.

• We present simulation results for practically relevant systems and channel parameters that show that MIMO–OFDM with optimized WSFC and V–BLAST and reduced feedback for ABL achieve similar performances, which closely approach that of SVD–based MIMO–OFDM.

The remainder of this chapter is organized as follows. Section 4.2 provides the system model for MIMO–OFDM. Section 4.3 presents WSFC applied to MIMO–OFDM, and reviews V–BLAST and SVD. Section 4.4 introduces adaptive bit–loading schemes and describes the proposed modification for bit–loaded V–BLAST. Section 4.5 presents the simulation results and discussion, and finally Section 4.6 gives the conclusions.
4.2 System Model

The OFDM system under consideration is equipped with \( N_T \) transmit and \( N_R \) receive antennas and we assume \( N_T \leq N_R \). The number of subcarriers is \( K \). The block diagram of the OFDM system with MIMO signal processing and adaptive bit loading is shown in Fig. 4.1.

Figure 4.1: Block diagram of the MIMO–OFDM system with adaptive bit loading. (I)FFT denotes (inverse) fast Fourier transform.

At the transmitter, source bits are first encoded with a binary convolutional encoder and possibly interleaved (see below). The coded bits are fed into the ABL unit, which allocates bits to each of the \( K \) OFDM subcarriers and \( N_T \) antennas. While ABL and MIMO transmission are described in detail below, it should be noted that ABL requires feedback from the receiver only in the form of a vector of integers which specifies the number of bits assigned to each subcarrier and antenna. The amount of feedback for loading is thus much smaller than that for providing full CSI to the transmitter as required for MIMO processing with SVD.

Denoting \( \mathbf{x}_k \triangleq [x_{k,0} \ldots x_{k,N_T-1}]^T \) the \( N_T \)-dimensional vector transmitted over \( N_T \)
antennas and subcarrier \( k, 0 \leq k < K \), and assuming standard OFDM transmission and reception, the corresponding \( N_R \)-dimensional received vector is given by

\[
y_k = H_k x_k + n_k, \quad 0 \leq k < K,
\]

with the \( N_R \times N_T \) channel matrix \( H_k \) and the additive spatially and spectrally white Gaussian noise \( n_k \). The MIMO–OFDM channel is assumed to be block fading, i.e., the channel does not change during one coding block, but may vary from one block to another. We assume perfect CSI at the receiver (ideal channel estimator in Fig. 4.1).

### 4.3 MIMO Processing for Coded MIMO–OFDM

We now introduce the WSFC scheme for MIMO–OFDM with ABL (Section 4.3.1) and briefly review V–BLAST-based (Section 4.3.2) and SVD-based MIMO–OFDM (Section 4.3.3), cf. e.g. [79, 81].

#### 4.3.1 Wrapped Space-Frequency Coding (WSFC)

WSFC is the straightforward extension of WSTC devised in [57] for single-carrier space-time transmission to MIMO–OFDM. The coded bit stream is divided into \( N_T \) layers assigned to \( N = (K - (N_T - 1)d)N_T \) transmit symbols such that if \( c_j, 0 \leq j < K \), denotes the \( j \)th symbol mapped from the encoder output, then \( x_{k,i} = c_{N_T(k-id)+i} \) for \( id \leq k < K - (N_T - 1 - i)d \) and \( x_{k,i} = 0 \) otherwise, \( 0 \leq i < N_T \). The parameter \( d \) is the so-called interleaving delay. This formatting “wraps” the codeword around the space-frequency plane, skewed by the delay \( d \). Fig. 4.2 shows an example of a WSFC codeword matrix with \( N_T = 3 \), \( d = 3 \), and \( K = 384 \) (cf. [57, Fig. 2] for WSTC). Note that \( d = 3 \) is chosen only for illustration. The actual value for \( d \) needs to be optimized.
for the best tradeoff between rate loss due to zero-symbols $x_{k,i} = 0$ and error propagation (see Section 4.5).

The skewness of the space-frequency arrangement of data symbols enables decoding with per-survivor processing (PSP) at the receiver. The received vectors are first processed with linear matrix-filters $F_k$ to form the vectors

$$\mathbf{v}_k = \left[ v_{k,0} \ldots v_{k,N_T-1} \right]^T$$

$$= F_k^H y_k$$

$$= B_k x_k + n'_k,$$

where $B_k = F_k^H H_k$ is the so-called feedback matrix and $n'_k = [n'_{k,0} n'_{k,1} \ldots n'_{k,N_T-1}]^T$ is the additive noise. This filtering is performed in the “MIMO Signal Processing” block of Fig. 4.1. Usual choices for the matrix $F_k$ in MIMO processing are the whitened matched filter, for which $B_k$ would be upper triangular and $n_k$ would be spatially white Gaussian noise, or the unbiased MMSE filter, in which case the elements of $n_k$ are correlated, cf. [57, Section III]. Here, we consider the unbiased MMSE filter for its usually superior performance and we approximate $n'_k$ as AWGN for the decoder design. Then, denoting
the element of $B_k$ in row $i$ and column $l$ by $b_{k,i,l}$, the samples

$$d_{k,i} = v_{k,i} - \sum_{l=i+1}^{N_T-1} b_{k,i,l} \hat{x}_{k,l}$$

(4.5)

are used as input information about $c_{N_T(k-id)+i}$ for the standard Viterbi decoder. The decisions $\hat{x}_{k,l}$ are taken from the survivor history of the decoder, whose depth is proportional to $d$. Hence, the effect of error propagation is alleviated with increasing $d$. In case of correct decisions, we have

$$d_{k,i} = b_{k,i,i} x_{k,i} + n'_{k,i}$$

(4.6)

and thus $N_T$ equivalent channels with gains $b_{k,i,i}$, $0 \leq i < N_T$, for each subcarrier $k$.

### 4.3.2 V–BLAST

V–BLAST for MIMO–OFDM can be regarded as a special case of WSFC with $d = 0$ and cancellation is performed using immediate decisions $\hat{x}_{k,i}$. However, different from WSFC, the order of detection, i.e., the sequence of values of $i$ in which decisions about $\hat{x}_{k,i}$ are made, can be modified to mitigate error propagation (see results in Section 4.5).

Applying a permutation matrix $\pi_k$ to the channel matrix $H_k$ to account for ordering, we obtain

$$v_k = B_k \pi_k^T x_k + n'_k.$$  

(4.7)

The conventional ordering strategy is to successively maximize the effective channel gains after cancelling, $|b_{k,i,i}|$, for $i = N_T - 1, N_T - 2, \ldots, 0$ in order to minimize error propagation. This greedy algorithm was proven to maximize the minimum channel gain and thus the SNR [83].

For V–BLAST with ABL, the optimum loading algorithm will distribute the rate
to the equivalent channels such that their error rates are approximately equal (see Section 4.4). In particular, the loading algorithm will always assign symbols from larger signal constellations to spatial channels with higher gains without considering error propagation. Hence, the optimum decoding order for V–BLAST with ABL could be different from that for V–BLAST without ABL. In fact, it has been found in [84] that the near-optimal decoding order for V–BLAST with ABL is obtained if the greedy algorithm chooses the transmit antenna with the smallest equivalent gain among the remaining unassigned antennas (cf. also [82]). In Section 4.4.3, we will describe a modification of V–BLAST with ABL to further reduce error propagation.

In addition to ordering, with V–BLAST coded bits are interleaved before mapping to (non-binary) signal points, which is not possible in case of WSFC due to the strict correspondence between coded bits and space-frequency transmit symbols.

### 4.3.3 SVD

By performing SVD, the channel matrix $H_k$ can be written as

$$H_k = U_k \Sigma_k V_k^H,$$  \hspace{1cm} (4.8)

where $U_k$ and $V_k^H$ are unitary matrices. The entries of the diagonal matrix $\Sigma_k$, $\sigma_{k,0} \geq \sigma_{k,1} \geq \ldots \geq \sigma_{k,N_T-1} \geq 0$, are the sorted non-negative singular values of $H_k$. In SVD-based MIMO transmission, the matrices $V_k$ and $U_k^H$ are applied to $x_k$ at the transmitter and $y_k$ at the receiver, respectively. This generates $N_T$ parallel channels with gains $\sigma_{k,i}$, $0 \leq i < N_T$, for each subcarrier $k$. As for V–BLAST, coding with bit-interleaving can be applied. We note that, different from WSFC and V–BLAST, full knowledge of $H_k$ is necessary to perform SVD-based transmission.
4.4 Adaptive Bit-Loading (ABL) Schemes

A number of loading algorithms have been proposed for single-antenna OFDM systems, cf. [35–40], and most of them achieve quite similar performance-complexity tradeoffs. In this chapter, we are interested in constant throughput and thus apply the margin-adaptive loading algorithm by Chow, Cioffi and Bingham (CCB) [36], whose information-theoretic capacity criterion seems to be a good match for coded transmission (although the codes considered in Section 4.5 do not operate at the capacity limit). However, numerical results not shown here indicate that the choice of the particular loading algorithm is not critical for coded MIMO–OFDM.

Since the MIMO processing schemes described in the previous section lead to an overall system with $N_T K$ parallel channels (assuming perfect cancellation for WSFC and V–BLAST), the CCB algorithm can be directly applied. We first consider two versions of loading with different feedback requirements and computational complexities and then describe a modification of the loading algorithm to account for error propagation in V–BLAST.

4.4.1 Full Loading (FL)

This scheme allocates bits to all $N_T K$ equivalent channels individually without distinguishing between spectral or spatial dimensions.

4.4.2 Grouped Loading (GL)

This scheme forms groups of equivalent channels with similar channel gains and the loading algorithm considers all channels within a group as identical. Since the channel gains $b_{k,i,i}$ (WSFC/V–BLAST) and $\sigma_{k,i}$ (SVD) are typically highly correlated along the frequency axis (index $k$) but strongly vary in the spatial domain (index $i$), grouping
of $G$ adjacent subcarriers corresponding to the same transmit antenna is proposed. $G$ is referred to as the group size. To provide the loading algorithm with a group representative, we consider two methods:

**Center subcarrier method**

The center subcarrier of the group (or one of the center subcarriers if $G$ is even) represents the group.

**Equivalent SNR method**

A virtual channel whose SNR equals

$$\text{SNR}_{eq} = \Gamma \gamma \left[ \prod_{l=0}^{G-1} \left( 1 + \frac{\text{SNR}_l}{\Gamma \gamma} \right)^{1/G} - 1 \right], \quad (4.9)$$

where $\text{SNR}_l$ is the SNR for the $l$th subcarrier in the group, $\Gamma$ is the “SNR gap” and $\gamma$ is the system performance margin iteratively updated by the CCB algorithm, cf. [36]. Eq. (4.9) directly derives from averaging the capacities (see [36, Eq. (1)]) associated with the subcarriers in the group.

Since GL reduces the required amount of feedback by a factor of $G$, it is a very interesting alternative, especially for WSFC/V–BLAST OFDM, which do not require CSI at the transmitter for MIMO processing. A virtual channel whose capacity equals the mean of the capacities of the channels in the group represents the group.

### 4.4.3 Modification of Loading for V–BLAST

As described in Section 4.3.2, the effect of error propagation in V–BLAST with ABL is mitigated by sorting the spatial subchannels in the order of increasing channel gains. It seems, however, advisable to also take error propagation into account when actually...
performing the loading. More specifically, we propose to increase the performance margin for the symbols of the antennas decoded first in the bit loading algorithm, which makes the tentative decisions of V–BLAST more reliable. To this end, we introduce a parameter, the extra margin $\eta_i$, $0 \leq i < N_T$, and make the following modification to the CCB algorithm. We replace Eq. (1) of [36] with

$$b_{k,i} = \log_2 \left( 1 + \frac{\text{SNR}_{k,i}}{\Gamma \gamma \eta_i} \right),$$

where $b_{k,i}$, SNR$_{k,i}$, and $\eta_i$ are the number of bits allocated, the SNR, and the extra performance margin of the ordered $i$-th symbol on subcarrier $k$, respectively.

If we set $\eta_0 = 1$, then $\eta_i$, $0 < i < N_T$, become the extra margins relative to the last detected symbol for a certain subcarrier $k$. The remaining task is to find the $\eta_i$ that minimizes the overall error rate. Since the parameter space increases exponentially with $N_T$, we propose the pragmatic choice $\eta_i = (\eta_{\text{extra}})^i$, $0 \leq i < N_T$, where $\eta_{\text{extra}} \geq 1$ is the only parameter to be optimized. This will be done in the next section based on simulated performances.

### 4.5 Results and Discussion

We now present and discuss simulation results for the different MIMO–OFDM schemes with ABL. We adopt the following system parameters from the IEEE 802.16e standard [6]: OFDM with 3.5 MHz bandwidth and 512 subcarriers of which 384 are active; rectangular $M$–QAM constellations with $M = 2^i$, $0 \leq i \leq 8$, and Gray labelling of signal points; convolutional encoder with generator polynomials $(171,133)_8$. We further assume $N_T = N_R = 2$ as a relevant example, and the ITU–R vehicular channel model A [85]. In all cases, the average data rate per active subcarrier is fixed to $\bar{R} = 2$ bits and
\( \bar{R} = 4 \) bits, respectively.

4.5.1 Optimization of WSFC and V–BLAST for MIMO–OFDM with ABL

First, we consider the optimization of WSFC. Fig. 4.3 shows the SNR \( E_s/N_0 \) required for a BER of \( 10^{-3} \) versus the interleaving delay \( d \). While increasing \( d \) leads to more accurate tentative decisions it also incurs a larger rate loss due to initialization and termination of WSFC encoding. In order to keep the overall rate unchanged, more bits have to be allocated to subcarriers not affected by initialization and termination, which has a negative effect on BER performance. In the case of \( \bar{R} = 2 \) bits, the system achieves the best performance when the delay lies in the range from \( d = 16 \) to \( d = 32 \). For \( \bar{R} = 4 \) bits, the best performance is obtained between \( d = 8 \) and \( d = 28 \). Hence, \( d = 16 \) is a universally good choice and used in the following. We note, however, that somewhat smaller (larger) delays may be optimal for OFDM with fewer (more) than 384 subcarriers due to the more (less) pronounced rate loss for fixed \( d \).

Next, we consider the optimization of bit-loaded V–BLAST with ordering, where the symbol assigned to the spatial channel with the smallest gain will be decoded first. Fig. 4.4 shows the SNR \( E_s/N_0 \) required for BERs of \( 10^{-3} \) and \( 10^{-4} \) versus the extra margin \( \eta_{\text{extra}} \) for the case of \( \bar{R} = 2 \) bits. The curves for V–BLAST without ordering are also included as references. Using an extra margin, \( \eta_{\text{extra}} > 1 \), leads to more reliable tentative decisions, however, it also makes the symbols corresponding to the antenna detected last more error-prone. It can be seen that the optimum extra margins are approximately at \( \eta_{\text{extra}} = 2 \sim 2.5 \) and that optimization with respect to \( \eta_{\text{extra}} \) provides gains of 0.7 dB at BER = \( 10^{-3} \) and 1.3 dB at BER = \( 10^{-4} \), respectively. This is quite remarkable considering that the improvement due to ordering (i.e. \( \eta_{\text{extra}} = 1 \)) is only
4. On MIMO–OFDM with Coding and Loading

Figure 4.3: SNR required for WSFC to achieve BER = $10^{-3}$ vs. interleaving delay $d$ with $\bar{R} = 2 \text{ bits and 4 bits.}$

0.25 and 0.9 dB, respectively.

### 4.5.2 Performance Comparisons

We now compare the performances of coded MIMO–OFDM based on SVD, V–BLAST, and WSFC. To separate the different effects, (i) V–BLAST without ordering, (ii) V–BLAST with ordering, and (iii) V–BLAST with ordering and optimal $\eta_{\text{extra}}$ are considered. Note that bit-interleaving is applied for V–BLAST but not for WSFC.

Fig. 4.5 shows the BER results for $\bar{R} = 2 \text{ bits with and without ABL.}$ As expected, SVD with ABL yields the best performance among all the schemes and its bit-loading gain is more than 8.4 dB at BER = $10^{-4}$. Interestingly, SVD without loading is inferior to WSFC, which can be attributed to the large variations of the subchannel gains in case of SVD. WSFC with ABL approaches the performance of SVD within 1.2 dB, and
its loading gain is 1.5 dB at BER $= 10^{-4}$. WSFC clearly outperforms V–BLAST, which confirms the effectiveness of the interleaving delay $d$. If the detection order is optimized, the performance of V–BLAST with ABL is 2.1 dB worse than that of WSFC. If the proposed additional margin $\eta_{\text{extra}}$ is applied for ABL, the SNR gap between V–BLAST and WSFC decreases to 0.8 dB at BER $= 10^{-4}$.

Finally, we consider the performance if GL is applied for the example of $\bar{R} = 2$ bits. For V–BLAST (with ordering), ABL without and with extra margin is performed. Fig. 4.6 shows the results in terms of the SNR required to achieve a BER of $10^{-3}$ for group sizes of $G = [1, 2, 4, 8]$. The SNR values for transmission without loading are also given as a reference. It can be seen that WSFC is more robust to the suboptimality due to grouping than V–BLAST. The larger deterioration for V–BLAST should be attributed to the aggravated error propagation when employing non-ideal loading. This
Figure 4.5: BER performance for coded MIMO–OFDM with and without ABL. $\bar{R} = 2$ bits, and $d = 16$ for WSFC.

Effect is alleviated in case of WSFC due to the interleaving delay $d > 0$. For WSFC the SNR-penalties compared to $G = 1$ are $[0.05, 0.16, 0.53]$ dB when using the center subcarrier and only $[0.02, 0.09, 0.3]$ dB when using the equivalent SNR for ABL. The latter criterion is apparently advantageous for WSFC and losses of e.g. 0.09 and 0.3 dB are fairly small given the reduction in feedback required for loading by factors of 4 and 8, respectively. Interestingly, for V–BLAST the center-subcarrier criterion yields better performances, which shows that one should not blindly apply a certain criterion for ABL with grouping of subcarriers.

We conclude that both optimized WSFC and V–BLAST achieve power efficiencies close to that of SVD-based MIMO–OFDM with ABL, and WSFC is somewhat advantageous if the feedback channel required for ABL has a very limited capacity.
4. On MIMO–OFDM with Coding and Loading

Figure 4.6: SNR required to achieve BER = $10^{-3}$ for WSFC and V–BLAST with ordering based MIMO–OFDM with different group sizes $G$ for ABL. $R = 2$ bits, and $d = 16$ for WSFC. The center subcarrier and equivalent SNR methods are used for loading.

4.6 Conclusions

In this chapter, we have studied coded MIMO–OFDM with ABL. We have proposed WSFC for MIMO–OFDM and a modified loading for V–BLAST to mitigate the problem of error propagation. Furthermore, we have considered ABL with subcarrier grouping based on two criteria to reduce the feedback load. The presented simulation results have shown notable gains due to WSFC and V–BLAST optimization, and that WSFC and V–BLAST perform fairly close to the benchmark case of SVD, which requires full CSI at the transmitter. We thus conclude that the devised WSFC-based and V–BLAST–based MIMO–OFDM with ABL are attractive solutions for power and bandwidth-efficient transmission for scenarios with small feedback rates like in e.g. IEEE 802.16e systems.
5 Cyclic Space–Frequency Filtering for BICM–OFDM Systems

5.1 Introduction

In Chapter 4, we have studied coded MIMO–OFDM with ABL, which requires small feedback of instant parameters for ABL from the receiver. In this chapter, we introduce cyclic space–frequency (CSF) filtering, a SFC scheme, which exploits the space and frequency diversity of MIMO–OFDM without requiring instant feedback from the receiver.

OFDM converts a frequency–selective fading channel into a number of parallel flat fading channels. As mentioned before, the negative effects of flat fading can be reduced with multiple transmit and multiple receive antennas and application of some form of SFC, cf. e.g. [45, 48]. Thereby, the multiple transmit antennas may either belong to the transmitting user (source node) or to other users who act as relay nodes to form a virtual MIMO system [20]. In this chapter, we refer to the former and the latter scenarios as transmission with co–located and distributed antennas, respectively. In the distributed case, we assume a two–hop protocol with DF relays [20]. A particularly simple and

\footnote{The material in this chapter was presented in part at IEEE Vehicular Technology Conference (VTC–Fall), Baltimore, USA, October 2007, and in part at VTC–Spring, Singapore, May 2008.}
efficient form of SFC, which is applicable to both co-located [49, 50] and distributed [86] transmit antennas, is CDD, cf. Section 2.5. CDD is typically combined with BICM [44] which is capable of exploiting the additional frequency diversity created by CDD and widely used in existing wireless standards [87].

The proposed CSF filtering schemes can be applied for BICM–OFDM systems with multiple co-located or distributed transmit antennas. CSF filtering may be viewed as a generalization of CDD where the cyclic delays are replaced by CSF filters. Unlike in conventional SFC schemes, the CSF filters are not defined over a finite field but over the field of complex numbers. The contributions made in this chapter can be summarized as follows:

• Assuming BICM at the transmitter and Viterbi decoding at the receiver, we develop optimization criteria for CSF filtering with co-located and distributed transmit antennas, respectively.

• Based on these criteria we present closed-form solutions for the optimum CSF filters for several special cases and numerical methods for efficient CSF filter design for the general case, where closed-form solutions do not exist.

• We show that BICM–OFDM with CSF filtering can exploit the full spatial and spectral diversity of wireless systems with co-located or distributed antennas.

• We also discuss the extension of our results to OFDMA systems.

There is a large body of literature on finite field based SFC schemes for co-located, e.g. [45, 46, 88, 89], and distributed, e.g. [65, 90–93], transmit antennas. In contrast to these schemes, the proposed CSF filtering scheme, if properly designed, does not require any changes to the receiver compared to single-antenna transmission. This is a major advantage in networks where devices with different numbers of antennas have to
coexist and facilitates the upgrade of existing wireless systems. For example, to provide transmit diversity in the downlink only the access points have to be equipped with CSF filters and multiple antennas but the user devices do not have to be modified. Recently, several attempts have been made to optimize CDD for co–located antennas. In [87], CDD with large cyclic delays is advocated and the interleaver is optimized accordingly. Precoding for CDD was proposed in [94] and optimized for uncoded transmission. The idea of replacing the cyclic delays with cyclic filters has been proposed independently in [95].\textsuperscript{2} However, [95] advocates a relatively complex simulation–based approach to filter optimization whereas we derive a simple closed–form optimality criterion. The optimization of CDD for cooperative OFDM transmission with distributed antennas was discussed in [50]. We finally note that there are several related papers which optimize linear DD and/or linear space–time filtering for \textit{uncoded single–carrier} transmission with co–located [96, 97] and distributed [65, 98] antennas.

The remainder of this chapter is organized as follows. In Section 5.2, the system models for co–located and distributed CSF filtering are introduced. The optimization of the CSF filters when the set of transmit antennas is known and not known is presented in Sections 5.3 and 5.4, respectively. In Section 5.5, simulation results are provided, and Section 5.6 concludes this chapter.

5.2 System Model

In this section, we introduce the considered network, transmitter, channel, and receiver models.

\textsuperscript{2}Both the conference version of this chapter and [95] were presented at IEEE VTC–Fall 2007.
5.2.1 Network Model

We consider a BICM–OFDM system with $K$ subcarriers, $N_R$ receive antennas, and $N_T$ co–located or distributed transmit antennas.

*Co–located Antennas:* In this case, we assume a simple point–to–point transmission where transmitter and receiver are equipped with $N_T$ and $N_R$ antennas, respectively.

*Distributed Antennas:* We consider a network with one source node, one destination node, and $N$ relay nodes. Each relay node is assigned a number $n \in \mathcal{N}, \mathcal{N} \triangleq \{1, 2, \cdots, N\}$. Furthermore, the source and relay nodes are equipped with a single antenna, whereas the destination node has $N_R$ antennas. We adopt a commonly used two–phase protocol [20]. During the first phase, the source node broadcasts a data packet to the relay nodes, and there is no transmission from the source to the destination node. During the second phase (relay phase), only those $N_T \leq N$ relay nodes which can successfully decode the packet received from the source in the first phase forward the packet to the destination node, cf. Fig. 5.1. The relay nodes participating in the second phase of transmission are collected in the set of active nodes $\mathcal{S} = \{n_1, \ldots, n_{N_T}\} \subseteq \mathcal{N}$, $|\mathcal{S}| = N_T$. We consider both coordinated and uncoordinated relaying. In coordinated relaying, the transmitting relay nodes know $\mathcal{S}$ (i.e., they know which other nodes are active), whereas in uncoordinated relaying, the relay nodes do not know $\mathcal{S}$. Uncoordinated relaying requires less overhead since the active relay nodes do not have to communicate with each other.

5.2.2 Transmitter

Block diagrams of the proposed transmitter structure for co–located and distributed antennas are shown in Figs. 5.2a) and b), respectively.
Figure 5.1: Example of a simple two-phase transmission with $N = 10$ potentially cooperating relay nodes. $N_T = 4$ relay nodes ($S = \{1, 3, 9, 10\}$) successfully decode the packet received from the source node (S) and forward it to the destination node (D).

Figure 5.2: Block diagram of a) transmitter with $N_T$ co-located antennas employing BICM–OFDM and CSF filtering, and b) transmitter of the $n$th relay node employing BICM–OFDM and CSF filtering.
**Co-located Antennas:** The bit stream is encoded with a convolutional encoder (CC) which has rate $R$ and free distance $d_{\text{free}}$. The $K \log_2(M)$ coded bits $c_{k'}$, $1 \leq k' \leq K \log_2(M)$ are interleaved and mapped onto symbols $X[k]$, $1 \leq k \leq K$, taken from an $M$-ary symbol constellation $\mathcal{X}$ such as $M$-ary phase-shift keying ($M$-PSK) or $M$-ary quadrature amplitude modulation ($M$-QAM). The effect of the interleaver $\Pi$ can be modeled by $k' \rightarrow (k, i)$ where $k'$ denotes the original index of coded bit $c_{k'}$, and $k$ and $i$ denote the index of symbol $X[k]$ and the position of $c_{k'}$ in the label of $X[k]$, respectively. We assume that the interleaver is chosen such that coded bits $c_{k'_1}$ and $c_{k'_2}$ with $|k'_1 - k'_2| \leq d_{\text{free}}$ are mapped onto different symbols which are transmitted over different subcarriers. This is necessary to exploit the full diversity offered by the channel, cf. [48]. Next, an IFFT is performed which converts frequency-domain symbols $X[k]$, $1 \leq k \leq K$, into time-domain symbols $x[t]$, $1 \leq t \leq K$.

At transmit antenna $n$, $1 \leq n \leq N_T$, the symbols $x[t]$ are passed through CSF filter $g_n \triangleq [g_n[0] \ g_n[1] \ldots g_n[L_g - 1]]^T$ of length $L_g \leq K$. The vector $g \triangleq [g_1^T \ldots g_{N_T}^T]^T$ containing all CSF filter coefficients is normalized to $||g||^2 = 1$. The CSF filter output at antenna $n$ is given by

$$s_n[t] = g_n[t] \otimes x[t] = \sum_{\nu=0}^{L_g-1} g_n[\nu] x[(t - \nu)_K], \quad (5.1)$$

where $\otimes$ denotes the cyclic convolution operator. We note that conventional CDD with delay $D$ may be interpreted as a special case of the CSF filtering operation described in (5.1) with $g_n[\nu] = 1/\sqrt{N_T}$ for $\nu = (n - 1)D$ and $g_n[\nu] = 0$ otherwise. Before the symbols $s_n[t]$ are transmitted by antenna $n$, a CP is added. We note that the CSF filtering in (5.1) has no implications for the required length of the CP which – similar to single-antenna transmission and CDD – is determined only by the length of the CIR. Furthermore, similar to CDD, and in contrast to other SFC schemes, only one IFFT
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

has to be performed at the transmitter. However, CSF filtering has a higher complexity than CDD since it requires $N_T$ cyclic filtering operations whereas for CDD only $N_T$ cyclic delays are necessary.

**Distributed Antennas:** The processing at active node $n \in \mathcal{S}$ is identical to the processing at antenna $n$ in the co–located antenna case, i.e., the transmit signal (before the CP is added) is given by (5.1). All active relay nodes employ identical CCs, interleavers, and mappers but in general different CSF filters $g_n$. In case of coordinated relaying the CSFs $g_n$, $n \in \mathcal{S}$, are optimized for the given set $\mathcal{S}$, cf. Section 5.3. In contrast, for uncoordinated relaying, the CSF filters $g_n$, $1 \leq n \leq N$, are optimized for all nodes in $\mathcal{N}$ and do not depend on $\mathcal{S}$ (which is assumed to be unknown), cf. Section 5.4. We adopt the power constraint $\|g_S\|^2 = 1$, where $g_S \triangleq [g_{n1}^T \ldots g_{nN_T}^T]^T$, for coordinated relaying. For uncoordinated relaying a stricter per–node power constraint $\|g_n\|^2 = 1/N_T$, $n \in \mathcal{S}$, is adopted since $\mathcal{S}$ is not known.

For the following, we formally define $\mathcal{S} \triangleq \{1, 2, \ldots, N_T\}$ for the co–located antenna case, which enables a unified treatment of the channel models and the receiver processing for CSF filtering with co–located and distributed transmit antennas.

### 5.2.3 Channel Model

We consider a frequency–selective channel with CIR $h_{nr}[l], 0 \leq l \leq L - 1$, of length $L$ between transmit antenna $n$ and receive antenna $r$. For convenience we collect the corresponding CIR coefficients in vector $h_{nr} \triangleq [h_{nr}[0] \ldots h_{nr}[L - 1]]^T$. The CIR coefficients are modeled as possibly correlated zero–mean complex Gaussian random variables (Rayleigh fading).

We define the vector containing all CIR coefficients as $h_S \triangleq [h_{n1}^T h_{n2}^T \ldots h_{nN_T}^T]^T$. For the channel autocorrelation matrix we adopt the popular Kronecker model, i.e.,
\( C_{hh} \triangleq \mathcal{E}\{h_S h_S^H\} = C_{NR} \otimes C_{NT} \otimes C_L \), where \( C_{NR} \), \( C_{NT} \), and \( C_L \) denote the receive antenna, the transmit antenna, and the temporal autocorrelation matrices, respectively.

Temporal correlation may be introduced by transmitter and receiver filtering. For co-located transmit antennas the off-diagonal elements of \( C_{NT} \) will be in general non-zero due to spatial correlation. In contrast, for distributed transmit antennas \( C_{NT} \) will be a diagonal matrix but its main diagonal elements may not be identical because of non-identical relay-destination distances.

For both co-located and distributed transmit antennas we assume the channel to be constant for the duration of at least one OFDM symbol and the autocorrelation matrix to have full rank, i.e., \( r(C_{NR}) = N_R \), \( r(C_{NT}) = N_T \), and \( r(C_L) = L \).

### 5.2.4 Receiver Processing

At the receiver, CSF filtering does not require any modifications compared to CDD and single-antenna transmission. Assuming that the CP has a length of at least \( L - 1 \), after removal of the CP and FFT the received signal at antenna \( r \), \( 1 \leq r \leq N_R \), can be expressed as

\[
R_r[k] = H_{eq,r}[k]X[k] + N_r[k], \quad 1 \leq k \leq K, \tag{5.2}
\]

where the channel coefficient \( H_{eq,r}[k] \) is given by

\[
H_{eq,r}[k] = \sum_{n \in \mathcal{S}} H_{nr}[k]g_n[k],
\]

with \( H_{nr}[k] \triangleq \sum_{l=0}^{L-1} h_{nr}[l] e^{-j2\pi kl/K} \) and \( G_n[k] \triangleq \sum_{l=0}^{L_g-1} g_n[l] e^{-j2\pi kl/K} \), and \( N_r[k] \) is AWGN with variance \( \sigma_n^2 \triangleq \mathcal{E}\{|N_r[k]|^2\} \). Next, optimum maximum ratio combining of the received signals in each subcarrier is performed. The combined signal

\[
R[k] = \sum_{r=1}^{N_R} H_{eq,r}^*[k]R_r[k], \quad 1 \leq k \leq K, \tag{5.3}
\]
is then used for bit metric calculation for the coded bit $c_k'$ [44, 48]

$$
\lambda^i(R[k], c_k') = \min_{X \in X^i_k} \left\{ \left| R[k] - \sum_{r=1}^{N_R} |H_{eq,r}[k]|^2 X \right|^2 \right\}, \tag{5.4}
$$

where $X^i_b$ denotes the subset of all symbols $X \in X$ whose label has the value $b \in \{0, 1\}$ in position $i$. The bit metrics are de–interleaved and Viterbi decoded [44, 48].

**Remark:** Eq. (5.2) shows that CSF filtering transforms the MIMO channel into an equivalent single–input multiple–output (SIMO) channel with frequency response $H_{eq,r}[k]$ for receive antenna $r$, $1 \leq r \leq N_R$. This corresponds to effective SIMO CIRs $h_{eq,r}[l]$ of length $L_{eq} = L + L_g - 1$. Note that, in general, $H_{eq,r}[k]$ will contain more spectral degrees of freedom than $H_{nr}[k]$ because of the CSF filtering. The receiver can directly estimate the equivalent overall CIRs $h_{eq,r}[l]$ using one of the numerous existing techniques for channel estimation in SIMO OFDM systems, cf. e.g. [99]. Direct estimation of $h_{eq,r}[l]$ is particularly convenient for distributed antennas since the receiver does not have to know the set of active nodes $S$. However, even for co–located antennas it may be advisable to directly estimate the $N_R L_{eq}$ coefficients of the equivalent SIMO channel instead of the $N_T N_R L$ coefficients of the MIMO channel as long as $L_{eq} \leq N_T L$. Thus, as far as channel estimation is concerned, small $L_g$ are preferable since they result in short equivalent CIRs.

### 5.2.5 Extension to OFDMA

Similar to CDD [86, 87], CSF filtering can be easily extended to OFDMA systems. In the following, we briefly point out the necessary modifications.

**Co–located Antennas:** For co–located antennas, the data streams of $U$ users are separately encoded, interleaved, mapped, and assigned to mutually exclusive sets of $K/U$ subcarriers. Subsequently a joint IFFT is applied to all $K$ frequency domain
symbols, cf. [87, Fig. 6]. The resulting time-domain sequence $x[t]$, $1 \leq t \leq K$, is applied to the inputs of the $N_T$ CSF filters as shown in Fig. 5.2a).

**Distributed Antennas:** In the distributed case, $U$ source nodes communicate with up to $U$ destination nodes at the same time using mutually exclusive sets of subcarriers [86]. The $n$th relay node tries to decode the received packets of all $U$ users, and forwards the successfully decoded packets to the destination node(s) using OFDMA. Thereby, the packets received from different sources are separately encoded, interleaved, mapped, and assigned to mutually exclusive sets of $K/U$ subcarriers. The subcarriers of unsuccessfully decoded source nodes remain unmodulated. After the IFFT, CSF filtering is applied as shown in Fig. 5.2b).

The received signals $R_r[k]$, $1 \leq r \leq N_R$, belonging to different users/source nodes are processed separately using the standard procedure outlined in Section 5.2.4. The CSF filter optimization for BICM–OFDM presented in Sections 5.3 and 5.4 is – with minor modifications – also applicable to BICM–OFDMA. We will point out these modifications wherever necessary.

### 5.3 CSF Filter Optimization for Known $S$

In this section, we assume that $S$ is known, which applies to co-located antennas and coordinated relaying. A unified treatment is possible by adopting again $S = \{1, 2, \ldots, N_T\}$ for the co-located antenna case. We assume that the correlation matrices $C_{NT}$ and $C_L$ are known and optimize the CSF filters accordingly. For simplicity of notation, in this section, we replace $g_S$ with $g$, since $S$ is fixed and known.
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

5.3.1 Optimization Criterion

The proposed optimization criterion is based on the worst–case PEP of the considered transmission system. From (5.3) we observe that for a given subcarrier and a given channel realization the overall system including maximum ratio combining is equivalent to a SISO system with

\[ \text{SNR}[k] = \frac{|H_{eq}[k]|^2}{\sigma_n^2}, \quad 1 \leq k \leq K, \]

where

\[ |H_{eq}[k]|^2 \triangleq \sum_{r=1}^{N_R} |H_{eq,r}[k]|^2 = \sum_{r=1}^{N_R} \left| \sum_{n \in S} H_{nr}[k] G_n[k] \right|^2. \quad (5.5) \]

Therefore, we can exploit the results presented in [48, Section IV], where the PEP of BICM–OFDM was recently derived. In particular, exploiting [48, Eq. (7)] and \( Q(x) \leq \frac{1}{2} e^{-x^2/2} \) the PEP of two codewords \( c \) and \( \hat{c} \) conditioned on \( h_S \) and \( g \) can be upper bounded as

\[ P(c \rightarrow \hat{c}|h_S, g) \leq \frac{1}{2} \exp \left( -\frac{d_{\text{min}}^2}{4\sigma_n^2} \sum_{k,d_{\text{free}}} |H_{eq}[k]|^2 \right), \quad (5.6) \]

where \( d_{\text{min}} \) denotes the minimum Euclidean distance of the signal constellation \( \mathcal{X} \) and the sum in the argument of the exponential function in (5.6) is over \( d_{\text{free}} \) distinct subcarriers \( K \triangleq \{ k_1, k_2, \ldots, k_{d_{\text{free}}} \} \).\(^3\) For the following, it is helpful to rewrite this sum as

\[ \sum_{k,d_{\text{free}}} |H_{eq}[k]|^2 = \sum_{k,d_{\text{free}}} \left| \sum_{r=1}^{N_R} \sum_{n \in S} w[k] G_n h_{nr} \right|^2 = h_S^H (I_{N_R} \otimes G^H W G) h_S, \quad (5.7) \]

where \( w[k] \triangleq [1, e^{j2\pi k/K}, \ldots, e^{j2\pi k(L_{eq} - 1)/K}]^T \), \( W \triangleq \sum_{k,d_{\text{free}}} w[k] w^H[k], \ G \triangleq [G_{n1} \ldots G_{n_{N_T}}] \), and \( G_n \) is an \( L_{eq} \times L \) column–circular matrix with vector \( [g_n^T \ 0_{L-1}^T]^T \) in its first column.

Using (5.7) in (5.6) and averaging the PEP with respect to \( h_S \) yields

\[ P(c \rightarrow \hat{c}|g) \leq \frac{1}{2} \left[ \det \left( I_{N_R L_{eq}} + \frac{d_{\text{min}}^2}{4\sigma_n^2} (C_{N_R} \otimes W G(C_{N_T} \otimes C_L) G^H) \right) \right]^{-1}. \quad (5.8) \]

\(^3\)We note that in OFDMA the possible sets \( K \) are user dependent.
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

Introducing the definition \( M \triangleq G(C_N \otimes C_L)G^H \) and assuming \( \sigma_n^2 \ll 1 \), we can rewrite (5.8) as

\[
P(c \rightarrow \hat{c}|g) \leq \frac{1}{2} \left[ \det(C_{N_R}) \right]^{-r(WM)} \left[ \prod_{i=1}^{r(WM)} \lambda_i(WM) \right]^{-N_R} \left( \frac{d_{\min}^2}{4\sigma_n^2} \right)^{-d}, \quad (5.9)
\]

where \( d \triangleq N_R r(WM) \) denotes the diversity gain. Assuming \( d_{\text{free}} \leq K \) and \( L_{eq} \leq K \), \(^4\) we obtain \( r(W) = \min\{d_{\text{free}}, L_{eq}\} \), cf. e.g. [48, Appendix], and \( r(M) = \min\{L_{eq}, N_T L\} \). For the rank of the product \( WM \) the inequality

\[
r(W) + r(M) - L_{eq} \leq r(WM) \leq \min\{r(W), r(M)\} \quad (5.10)
\]

holds [100], where the upper bound is achievable if \( W \) and/or \( M \) are not singular. We conclude from (5.9) that the achievable diversity gain with CSF filtering is

\[
d = N_R \cdot \min\{d_{\text{free}}, L_{eq}, N_T L\}. \quad (5.11)
\]

The coding gain of BICM–OFDM with CSF filtering can be defined as

\[
C(g) \triangleq \left( \prod_{i=1}^{r(WM)} \lambda_i(WM) \right)^{1/r(WM)}, \quad (5.12)
\]

cf. (5.9). To minimize the asymptotic PEP in (5.9), the CSF filters have to simultaneously maximize the coding gain \( C(g) \) and the diversity gain \( d \). Note that (5.11) implies that for given \( d_{\text{free}}, N_T, \) and \( L \) the maximum possible diversity gain can be exploited with CSF filters of length \( L_{g,\text{min}} = \min\{(N_T - 1)L + 1, d_{\text{free}} - L + 1\} \). Increasing \( L_g \) (and consequently \( L_{eq} \)) further does not improve \( d \) but may increase \( C(g) \) and will sig-

\(^4\)If OFDMA is employed, we have to assume \( d_{\text{free}} \leq K/U \) and \( L_{eq} \leq K/U \), respectively.
nificantly impact the design of the optimal $\mathbf{g}$. Therefore, we consider two different cases for $L_g$ in the following.

**Case 1**\( [L_g \leq L_{g,\min}] \): In this case, \( L_{\text{eq}} \leq \min\{d_{\text{free}}, N_T L\} \) and \( d = N_R L_{\text{eq}} \) are valid and we can rewrite the coding gain as \( C(\mathbf{g}) = (\det(\mathbf{W}) \det(\mathbf{M}))^{1/L_{\text{eq}}} \), which implies that we can optimize the interleaver (which influences \( \mathbf{W} \)) and the CSF filters separately. In particular, for a given interleaver it suffices to maximize the cost function

\[
  f(\mathbf{g}) \triangleq \det(\mathbf{M}) = \det(\mathbf{G}(\mathbf{C}_{N_T} \otimes \mathbf{C}_L)\mathbf{G}^H)
\]  

subject to the power constraint \( ||\mathbf{g}||^2 = 1 \).

**Case 2**\( [L_g > L_{g,\min}] \): In this case, we obtain \( L_{\text{eq}} > \min\{d_{\text{free}}, N_T L\} \) and \( d = N_R \min\{d_{\text{free}}, N_T L\} \), and (5.12) cannot be further simplified. Therefore, the interleaver and the CSF filters have to be optimized jointly. Based on (5.10) it can be shown that for \( \min\{d_{\text{free}}, N_T L\} < L_{\text{eq}} \leq \max\{d_{\text{free}}, N_T L\} \) maximum–rank \( \mathbf{W} \) and \( \mathbf{M} \) guarantee a diversity order of \( d = N_R \min\{d_{\text{free}}, N_T L\} \). This means if we choose to optimize/design \( \mathbf{W} \) and \( \mathbf{M} \) separately, we are guaranteed to achieve full diversity but the coding gain may not be maximized. On the other hand, for \( L_{\text{eq}} > \max\{d_{\text{free}}, N_T L\} \) we obtain from (5.10) \( N_R (N_T L + d_{\text{free}} - L_{\text{eq}}) \leq d \leq N_R \min\{d_{\text{free}}, N_T L\} \), which means that in this case maximum–rank \( \mathbf{W} \) and \( \mathbf{M} \) do not even guarantee full diversity. This latter case was considered for the CDD scheme in [87], where it was indeed found that the interleaver and the cyclic delays have to be designed jointly.

Considering the above discussion, we will concentrate on Case 1 \( (L_g \leq L_{g,\min}) \) in the remainder of this section. This allows us to optimize the CSF filters independent from the interleaver, which is a major advantage. Note that this choice does not compromise the diversity gain since \( L_g = L_{g,\min} \) exploits the maximum diversity gain \( d = N_R \min\{d_{\text{free}}, N_T L\} \) of the BICM–OFDM system. Choosing \( L_g > L_{g,\min} \) yields
additional coding gains which, however, tend to be small if $S$ is known, cf. discussion of Fig. 5.6.

### 5.3.2 Upper Bound on Cost Function $f(g)$

In this section, we derive an upper bound on $f(g)$. For this purpose we assume that $G$ is a general $L_{eq} \times N_T L$ matrix (without a pre-defined structure), which only has to fulfill the power constraint $\text{tr}\{G^H G\} = L$. For clarity we denote the corresponding cost function by $\tilde{f}(G) = \det(G(C_{N_T} \otimes C_L)G^H)$. It can be shown that the maximum value of $\tilde{f}(G)$ is given by

$$
\tilde{f}(G_{\text{opt}}) = \left(\frac{L}{L_{eq}}\right)^{L_{eq}} \prod_{i=1}^{L_{eq}} \lambda_i(C_{N_T} \otimes C_L) \quad (5.14)
$$

with optimum matrix $G_{\text{opt}} = \sqrt{\frac{L}{L_{eq}}} U_{L_{eq}}^H$, where the unitary $N_T L \times L_{eq}$ matrix $U_{L_{eq}}$ contains the eigenvectors (EVs) of $C_{N_T} \otimes C_L$ corresponding to the $L_{eq}$ largest eigenvalues.

Since the CSF filter matrix $G$ does not only have to fulfill a power constraint but the sub-matrices $G_{nm}$, $1 \leq m \leq N_T$, also have to be cyclic, cf. Section 5.3.1, $f(g)$ can be bounded as

$$
f(g) \leq \left(\frac{L}{L_{eq}}\right)^{L_{eq}} \prod_{i=1}^{L_{eq}} \lambda_i(C_{N_T} \otimes C_L) \quad (5.15)
$$

for any $g$ with $||g||^2 = 1$. This (in general not achievable) bound will be used in the following to establish optimality of CSF filters for certain special cases.

### 5.3.3 Closed–Form Solution for Special Cases

In general, a closed–form solution for the optimum CSF filter vector $g_{\text{opt}}$ does not seem to exist. However, there are three special cases for which closed–form solutions do exist.
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

**Case 1 \([L = 1]\):** For \(L = 1\) (frequency–nonselective channel) matrix \(G\) does not have a circular structure and is given by \(G = [g_{n_1} \ldots g_{n_N}]\). Therefore, the results from Section 5.3.2 can be directly used to obtain the optimum matrix \(G_{\text{opt}} = \frac{1}{\sqrt{L_g}} U_{L_g}^H\), where the unitary \(N_T \times L_g\) matrix \(U_{L_g}\) contains the EVs of \(C_{N_T}\) corresponding to the \(L_g\) largest eigenvalues. The optimum CSF filter vectors are simply given by the columns of \(G_{\text{opt}}\) and achieve the upper bound in (5.15).

**Case 2 \([L_g = (N_T - 1) + 1]\):** In this case, \(L_{\text{eq}} = N_T L\) and conventional CDD (C–CDD) [96] with delay \(D = L\) results in \(g_{nm}[^{\nu}1] = 1/\sqrt{N_T}\) for \(\nu = (m - 1) L\) and \(g_{nm}[^{\nu}2] = 0\) otherwise, \(1 \leq m \leq N_T\), i.e., \(G = \frac{1}{\sqrt{N_T}} I_{N_T L}\). Therefore, we obtain for C–CDD

\[
f(g) = \frac{1}{N_T L} \det(C_{N_T} \otimes C_L), \quad (5.16)
\]

which is identical to the upper bound in (5.15) for this case since \(\det(C_{N_T} \otimes C_L) = \prod_{i=1}^{N_T L} \lambda_i(C_{N_T} \otimes C_L)\). This shows that, surprisingly, C–CDD is optimum for \(L_g = (N_T - 1)L + 1\) independent of the correlation matrix \(C_{N_T} \otimes C_L\). Note that we implicitly have assumed that \(L_{g,\text{min}}\) is limited by \(N_T L\) and not by \(d_{\text{free}}\), i.e., \(d_{\text{free}} \geq N_T L\). Furthermore, \(L_g < (N_T - 1)L + 1\) may be preferable in practice since short equivalent channels facilitate channel estimation. If \(d_{\text{free}} < N_T L\) and/or \(L_g < (N_T - 1)L + 1\), C–CDD is not optimum and the CSF filters have to be optimized.

**Case 3 \([L_g = (\kappa - 1)L + 1, \kappa \in \{1, 2, \ldots, N_T\}]\):** We define the matrix whose columns are the CSF filter vectors and the EV of \(C_{N_T}\) corresponding to the \(n\)th largest eigenvalue \(\lambda_n(C_{N_T})\) as \(\tilde{G} \triangleq [g_{n_1} g_{n_2} \ldots g_{n_N}]\) and \(u_n\), respectively, and choose \(\tilde{G}\) as

\[
\tilde{G} = \frac{1}{\sqrt{\kappa}} \left[ u_1 \ 0_{N_T \times (L-1)} \ u_2 \ 0_{N_T \times (L-1)} \ \ldots \ u_{\kappa} \right]^T. \quad (5.17)
\]
It can be shown that the CSF filters defined by (5.17) yield a cost function of
\[ f(g) = \frac{1}{\kappa L} \left( \det(C_L) \right)^\kappa \left( \prod_{i=1}^\kappa \lambda_i(C_{NT}) \right)^L. \] (5.18)

On the other hand, with \( L_{eq} = \kappa L \) we obtain from (5.15)
\[ f(g) \leq \frac{1}{\kappa L} \prod_{i=1}^{\kappa L} \lambda_i(C_{NT} \otimes C_L). \] (5.19)

Furthermore, \( \lambda_i(C_{NT} \otimes C_L) = \lambda_n(C_{NT}) \lambda_l(C_L), 1 \leq i \leq N_T L, 1 \leq n \leq N_T, 1 \leq l \leq L \) holds [101]. Therefore, for \( 1 \leq \kappa < N_T \), \( f(g) \) in (5.18) achieves the upper bound in (5.19) if and only if \( \lambda_\kappa(C_{NT}) \lambda_L(C_L) \geq \lambda_{\kappa+1}(C_{NT}) \lambda_1(C_L) \) is true. In this case, the columns of matrix \( \tilde{G} \) contain the coefficients of the optimum CSF filters. We note that our results in Section 5.5 suggest that this design is also optimum if \( \lambda_\kappa(C_{NT}) \lambda_L(C_L) < \lambda_{\kappa+1}(C_{NT}) \lambda_1(C_L) \). However, in this case, we cannot prove optimality since the upper bound is not achieved. In contrast, for \( \kappa = N_T \), (5.19) is always identical to the upper bound regardless of the eigenvalues of \( C_{NT} \) and \( C_L \), i.e., the CSF filters according to (5.17) are optimum.

### 5.3.4 CSF Filter Design for the General Case

Due to the particular structure of \( G \) it does not seem to be possible to find a closed-form solution for \( g \) in general. Furthermore, since \( f(g) \) is not a concave function, the application of well-known standard tools from the convex optimization literature does not seem to be possible. Therefore, we consider three numerical design methods in the following.
1) **O–CDD**: For optimized CDD (O–CDD), we choose $g_n[\nu] = 1/\sqrt{N_T}$ for $\nu = D_n < L_g$ and $g_n[\nu] = 0$ otherwise, $n \in S$. The delays $D_n$ are optimized according to

$$\{D_{\text{opt},n_1}, \ldots, D_{\text{opt},n_{N_T}}\} = \text{argmax}_{D_n, n \in S} \{f(g)\},$$  \hspace{1cm} (5.20)

which requires a brute force search over all $L_g^{N_T}$ possible delay combinations $0 \leq D_n < L_g, n \in S$. Similar to conventional (unoptimized) CDD, for O–CDD the transmit signal only has to be delayed and multiplications are not necessary. However, compared to conventional CDD, O–CDD offers more flexibility in the choice of the delays.

2) **EV–CSF Filters**: Motivated by Case 3 in Section 5.3.3 we propose the following (in general suboptimum) CSF filter matrix

$$\tilde{G} = \frac{1}{\sqrt{\kappa}} \left[ u_1 0_{N_T \times D_1} u_2 0_{N_T \times D_2} \ldots 0_{N_T \times D_{\kappa-1}} u_\kappa \right]^T,$$  \hspace{1cm} (5.21)

where $\kappa \triangleq \lceil (L_g - 1)/L \rceil + 1$ with $\lceil \cdot \rceil$ representing the smallest integer larger than the argument and $u_n, 1 \leq n \leq \kappa$, denotes the EVs corresponding to the $\kappa$ largest eigenvalues of $C_{N_T}$. Furthermore, the $D_n, 0 \leq D_n \leq L - 1, 1 \leq n \leq \kappa - 1$, meet the constraint

$$\sum_{n=1}^{\kappa-1} D_n + \kappa = L_g$$

and are optimized for maximization of $f(g)$ in a similar fashion as the delays in (5.20). If the conditions outlined in Case 3 in Section 5.3.3 are met, the EV–CSF filter design in (5.21) is optimum, otherwise it is suboptimum.

3) **GA–CSF Filters**: If $g$ is not required to have a particular structure, the gradient algorithm (GA) in Table 5.1 (known $S$) can be used to optimize $g$ directly. The elements of the gradient vector $\partial f(g)/\partial g^* \triangleq [\partial f(g)/\partial g_{n_1}[0] \partial f(g)/\partial g_{n_1}[1] \ldots \partial f(g)/\partial g_{n_{N_T}}[L_g - 1]]^H$ required for the adaptation in Table 5.1 are given by

$$\frac{\partial f(g)}{\partial g_n[\nu]} = \det(M) \text{tr} \left( M^{-1} E_{\nu}(C_{N_T} \otimes C_L) G^H \right),$$  \hspace{1cm} (5.22)
where the entries of $L_{eq} \times N_T L$ matrix $E_{nu}$ are equal to 1 at those positions where $G$ has entries $g_n[\nu]$ and zero otherwise. Because of its relatively involved form, we cannot prove convergence of the GA to the global optimum of $f(g)$. However, our extensive simulations have shown that different random initializations of the algorithm lead to the same $f(g)$ after the algorithm has converged, cf. also Fig. 5.3.

Remark: We note that for co–located antennas the complexity of the proposed CSF design methods (e.g. speed of convergence of the GA) is of minor importance. Since the statistical properties of the channel are typically constant, the CSF filters can be pre–computed off–line. For coordinated relaying the CSF filters have to be updated whenever $S$ changes. For this purpose, the CSF filters may be pre–computed and stored for all possible $S$ or calculated on–line before data transmission starts.

## 5.4 CSF Filter Optimization for Unknown $S$

In this section, we assume that $\mathcal{N}$, $C_L$, and all possible correlation matrices $C_{NT}$ are known for CSF filter optimization. However, the set of antennas (nodes) $S$ participating in the current transmission are not known (uncoordinated relaying). Thus, the relay nodes $n \in \mathcal{N}$ can only be assigned fixed, pre–optimized CSF filters $g_n$ which are independent of the current $S$. The set of all filters $g_n$, $n \in \mathcal{N}$, is denoted by $\mathcal{G} \triangleq \{g_1, g_2, \ldots, g_N\}$. Under these circumstances the goal of the CSF filter optimization is to design sets $\mathcal{G}$ of CSF filters which achieve a good performance regardless of which nodes are active. In this section, for clarity we add the subscript $S$ to $g$, $G$, and $M$ to emphasize their dependence on the set of active nodes. Furthermore, we add subscript $\mathcal{K}$ to matrix $W$ to emphasize its dependence on set $\mathcal{K}$ which contains the subcarrier indices corresponding to $d_{free}$ consecutive outputs of the convolutional encoder.
Table 5.1: Gradient algorithms (GAs) for calculation of the optimum CSF filters for known and unknown $S$, respectively. The elements of the gradient vectors $\partial f(g)/\partial g^*$ and $\partial f_{2,|\mathcal{K}|}(g_S)/\partial g^*_S$ are given in (5.22) and (5.30), respectively. The adaptation constant $\mu[i]$ has to be optimized experimentally. The termination constant $\epsilon$ has a small value (e.g. $\epsilon = 10^{-6}$). $i$ denotes the iteration.

<table>
<thead>
<tr>
<th>Known $S$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Initialization:</strong></td>
<td>Let $i = 0$ and initialize the CSF filter vector $g[i]$ with e.g. O–CDD or EV–CSF filters.</td>
</tr>
<tr>
<td><strong>2 Update CSF filter vector:</strong></td>
<td>$\tilde{g}[i + 1] = g[i] + \mu[i] \frac{\partial f(g[i])}{\partial g^*[i]}$</td>
</tr>
<tr>
<td><strong>3 Normalization:</strong></td>
<td>$g[i + 1] = \tilde{g}[i + 1]/</td>
</tr>
<tr>
<td><strong>4 Termination:</strong></td>
<td>If $</td>
</tr>
<tr>
<td><strong>5 End:</strong></td>
<td>$g[i + 1]$ is the desired CSF filter vector.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown $S$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Initialization:</strong></td>
<td>Let $i = 0$ and generate a random initial set of complex vectors $g_n[i]$, $1 \leq n \leq N$, with $</td>
</tr>
<tr>
<td><strong>2 Find worst set of nodes $S_{\min}[i]$ and the worst set of subcarriers $\mathcal{K}_{\min}[i]$:</strong></td>
<td>${S_{\min}[i], \mathcal{K}<em>{\min}[i]} = \arg\min</em>{S,\mathcal{K}} {f_{2,</td>
</tr>
<tr>
<td><strong>3 Update CSF filter vector:</strong></td>
<td>$\tilde{g}<em>{S</em>{\min}[i]}[i + 1] = g_{S_{\min}[i]}[i] + \mu[i] \frac{\partial f_{2,</td>
</tr>
<tr>
<td><strong>4 Normalization:</strong></td>
<td>$g_n[i + 1] = \tilde{g}_n[i + 1]/(\sqrt{N_a}</td>
</tr>
<tr>
<td><strong>5 Termination:</strong></td>
<td>If $</td>
</tr>
<tr>
<td><strong>6 End:</strong></td>
<td>$g_n[i + 1]$, $1 \leq n \leq N$, is the desired set $G_{\text{grad}}$.</td>
</tr>
</tbody>
</table>
5.4.1 Optimization Problem

To make the problem more manageable, we assume for optimization of $G$ that the number of active nodes $N_T = N_a$ is fixed and known. In a practical transmission, the actual number of active nodes $N_T$ may differ from the design parameter $N_a$, of course. If we exclude the non-diversity case $N_T = 1$, $N_T = 2$ may be considered as the worst-case scenario and more active nodes can only increase the diversity of the channel. Therefore, it is reasonable to adopt $N_a = 2$ for CSF filter design.

Similar to Section 5.3 the optimization is based on the PEP. More precisely, our goal is to minimize the maximum asymptotic PEP in (5.9) over all possible sets $S$ with $|S| = N_a$. The diversity gain $d$ is also limited by (5.11) if $S$ is unknown, i.e., the maximum diversity gains for known and unknown $S$ are identical. Furthermore, to maximize the coding gain (5.12), depending on $L_g$, we use the cost functions

\[
    f_1(g_S) = \det(G_S(C_{N_T} \otimes C_L)G_S^H), \quad L_g \leq L_{g,\text{min}},
\]

and

\[
    f_2(g_S) = \det(I + \gamma W_K G_S(C_{N_T} \otimes C_L)G_S^H), \quad L_g > L_{g,\text{min}},
\]

where $\gamma$ can be interpreted as an SNR (cf. (5.8)). For $\gamma \gg 1$ $f_2(g_S)$ is equivalent to the coding gain $C(g_S)$ in (5.12) but is easier to optimize. Similar to the case when $S$ is known, choosing $L_g > L_{g,\text{min}}$ complicates the optimization since the properties of the interleaver have to be taken into account via $W_K$. However, unlike when $S$ is known, when $S$ is unknown, substantial gains are possible with $L_g > L_{g,\text{min}}$ (cf. Fig. 5.6). Therefore, we also consider $L_g > L_{g,\text{min}}$ in this section. Based upon the above discussion,
we formally state our optimization problem as

$$G_{\text{opt}} = \arg\max_{\mathcal{G}} \left\{ \min_{|S|=N_a} f_\nu(g_S) \right\}, \quad \nu \in 1, 2,$$

subject to $||g_n||^2 = 1/N_a, \quad 1 \leq n \leq N,$ (5.26)

i.e., the optimum set of CSF filters $G_{\text{opt}}$ maximizes the minimum coding gain for a given $L_g$.

### 5.4.2 Upper Bound and Closed–Form Solution for Special Case

In this section, we concentrate on the case $L_g \leq L_{g,\text{min}}$. We note that (5.15) is also an upper bound for $f_1(g_S), \ g_n \in G_{\text{opt}}, \ n \in S$. However, this bound may be very loose in general. In the following, we will show that a tighter bound and a closed–form solution to (5.25), (5.26) exist for $N_a = N_T = 2, \ L = 1, \text{ and } L_g = 2$. For this special case, $f_1(g_S)$ can be re–written as

$$f_1(g_S) = \frac{1}{N_T^{N_T}} \det(C_{N_T}) (1 - |\rho_S|^2),$$

where $\rho_S \triangleq N_T g_n^H g_{n_2}$ denotes the correlation between $g_{n_1}$ and $g_{n_2}$. Therefore, in this case, (5.25), (5.26) is equivalent to minimizing the maximum correlation between any two vectors in $G$. The maximum correlation can be bounded as [102]

$$\rho_{\text{max}} \triangleq \max_S \{ |\rho_S| \} \geq \max \left\{ \sqrt{\frac{N - 2}{2(N - 1)}}, \ 1 - \frac{2}{N} \right\}.$$ (5.28)

Therefore, an upper bound for the minimum of the cost function over all sets $S$ is given by

$$f_{1,\text{min}} \triangleq \min_S \{ f_1(g_S) \} \leq \frac{1}{N_T^{N_T}} \det(C_{N_T}) \min \left\{ \frac{N}{2(N - 1)}, \frac{4}{N} \left( 1 - \frac{1}{N} \right) \right\}.$$ (5.29)
Sets of vectors with minimum maximum correlation $\rho_{\text{max}}$ (and consequently maximum $f_{1,\text{min}}$) are called Grassmannian frames [103]. For $N \leq L_g^2 = 4$ these Grassmannian frames achieve the bounds in (5.28) and (5.29) with equality. For the construction of the corresponding optimum sets for $N = 3$ and $N = 4$ we refer to [103]. For $N > L_g^2$ the bounds in (5.28) and (5.29) are not achievable and Grassmannian frames are difficult to construct. In this case, numerical design methods are an attractive alternative, see [103] and next subsection.

5.4.3 CSF Filter Design for General Case

In this section, we propose suboptimum and optimum CSF filter design methods if $S$ is not known.

1) Distributed C–CDD: In distributed C–CDD, the delay at node $n$ is equal to $D_n = [(n - 1)(L_g - 1)/(N - 1)]$, $1 \leq n \leq N$, where $\lceil \cdot \rceil$ denotes the closest integer to its argument. For the special case $L_g = K$ this choice for the CDD delays was proposed in [86] for an amplify–and–forward protocol. However, with the decode–and–forward protocol considered in this chapter, distributed C–CDD fails to capture the full diversity of the channel in general. The minimum delay difference between two nodes is $\Delta D = \min_{2 \leq n \leq N}\{D_n - D_{n-1}\}$. Therefore, assuming for example that the diversity gain is not limited by $d_{\text{free}}$ and $N_T = 2$, it is easy to see that distributed C–CDD cannot exploit the full diversity of the channel if $L > \Delta D$.

2) GA–CSF Filters: Similar to the case of known $S$, a GA can be used to optimize $G$ based on $f_1(g_S)$ and $f_2(g_S)$, respectively. We refer to the corresponding set of CSF filter vectors as $G_{\text{grad}}$. The elements of the gradient vector $\partial f_1(g_S)/\partial g_S^*$ are already given in
(5.22), and the elements of the gradient vector $\frac{\partial f_2(g_S)}{\partial g^*_S}$ can be obtained as

$$
\frac{\partial f_{2,K}(g_S)}{\partial g_n[\nu]} = \gamma f_{2,K}(g_S) \text{tr} \left( (I_{eq} + \gamma W_K M_S)^{-1} W_K E_n[\nu](C_{Ny} \otimes C_L)G_S^{ll} \right),
$$

$n \in S, 0 \leq \nu < L_{eq}$, (5.30)

where we have added the subscript "$K$" in $f_{2,K}(g_S)$ to emphasize its dependence on $K$.

The GA with $f_{2,K}(g_S)$ as cost function is given in Table 5.1. Similar to the case where $S$ is known, we cannot prove global convergence of the proposed algorithm. However, in our (extensive) simulations the sets $G_{\text{grad}}$ obtained with the algorithm for different random initializations always resulted in similar performances (cf. also Fig. 5.6). If OFDMA is employed, the search for the worst set of subcarriers $K_{\text{min}}[i]$ has to be performed over the $U$ sets of subcarriers corresponding to the $U$ users. On the other hand, if $L_g \leq L_{g,\text{min}}$, $f_{2,K}(g_S)$ can be replaced by $f_1(g_S)$ in the GA in Table 5.1, which becomes independent of $K$. Therefore, the algorithm can be simplified and in Step 2 only the worst set of nodes $S_{\text{min}}[i]$ but not the worst set of subcarriers has to be found, which reduces the complexity of the algorithm considerably.

**Remark:** We would like to point out that the speed of convergence and complexity of the GA are of minor importance. If the positions of the source, relay, and destination nodes are fixed, the channel statistics are time–invariant and the set of CSF filters has to be computed only once. If the nodes are mobile, the channel statistics will change and the CSF filters should be updated accordingly.

Finally, we point out that for the special case $S = N$ the optimization problems for known and unknown $S$ are identical. Therefore, the GA in Table 5.1 for unknown $S$ with $S = N$ can be used to find the optimum CSF filters for known $S$ and a given interleaver if $L_g > L_{g,\text{min}}$, cf. Fig. 5.6.

3) **R–CSF Filters:** If nodes frequently enter and leave the network, random CSF (R–
CSF) filters are an attractive alternative to deterministic CSFs. In this case, each relay node generates an R–CSF filter before it encodes and forwards the received message to the destination node. If the R–CSF filters are taken independently from some suitable distribution, matrix $G_S$ will have maximum rank for all possible $S$ with probability one. Therefore, for $L_g \leq L_{g,\text{min}}$ R–CSF filters can exploit the full available diversity. Note that this is not necessarily true if $L_g > L_{g,\text{min}}$, cf. Case 2 in Section 5.3.1. In this chapter, we consider R–CSF filters whose coefficients are uniformly distributed on a complex hypersphere. Similar random designs have been proposed for distributed space–time codes in [104].

4) R–CDD: Similar to the R–CSF filters in 3), each relay node $n$ may simply apply a random delay $D_n \in \{0, 1, \ldots, L_g - 1\}$ before data transmission. This scheme is referred to as R–CDD. Since two relay nodes will pick the same delay with non–zero probability, R–CDD does not achieve the maximum diversity gain of the BICM–OFDM system.

5.5 Simulation Results

In this section, we present simulation results for CSF filtering with co–located and distributed antennas, respectively. For all BER results shown, we adopted the quasi–standard convolutional code $(171, 133)_8$ with rate $R = 1/2$, constraint length 7, and free distance $d_{\text{free}} = 10$. Since the CSF design is not affected by the number of receive antennas, we assume $N_R = 1$ throughout this section.

5.5.1 Co–located Antennas

In this section, we consider co–located antennas with $[C_{NT}]_{i,j} = \rho^{|i-j|}$ and $[C_L]_{i,j} = \rho_L^{|i-j|}$, where $\rho$ and $\rho_L$ denote the spatial and temporal correlation coefficients, respectively. The GA–CSF filters employed in this section were generated with the GA given...
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

in Table 5.1 (known $S$), which was initialized with the O–CDD solution. The adaptation step size of the GA was $\mu[i+1] = \lambda \mu[i]$, where $\mu[0] = 0.1$ and $\lambda = 0.999$.

First we compare the CSF filter designs from Section 5.3.4. In particular, we consider the cost function $f(g)$ normalized with the upper bound $\tilde{f}(G_{opt})$ in (5.14). The resulting ratio $\alpha \triangleq \frac{f(g)}{\tilde{f}(G_{opt})}$ is shown in Figs. 5.3a) and b) for $\rho_L = 0.0$ and $\rho_L = 0.6$, respectively, as a function of $L_g$ for GA–CSF, EV–CSF, and O–CDD filters. $N_T = 4$, $\rho = 0.9$, and $L = 5$ are valid. As expected from the results in Section 5.3.3, for $L_g = (N_T - 1)L + 1 = 16$ EV–CSF filters and O–CDD, which is equivalent to C–CDD with $D = L = 5$ in this case, achieve the upper bound, i.e., $\alpha = 1$. Furthermore, for $\rho_L = 0$, $\lambda_k(C_{N_T})\lambda_L(C_L) \geq \lambda_{k+1}(C_{N_T})\lambda_1(C_L)$ holds for $k = \{1, 2, 3\}$. Therefore, the EV–CSF filters also achieve the upper bound for $L_g = (k - 1)L + 1, k = \{1, 2, 3\}$, cf. Sections 5.3.3, 5.3.4, Fig. 5.3a). The fact that the GA–CSF filters also achieve the upper bound in this case indicates that the GA indeed finds the optimum solution. For $\rho_L = 0.6$ and $k = \{2, 3\}$, $\lambda_k(C_{N_T})\lambda_L(C_L) < \lambda_{k+1}(C_{N_T})\lambda_1(C_L)$. Therefore, the EV–CSF filters cannot achieve the upper bound, cf. Case 3 in Section 5.3.3. However, the fact that the EV–CSF filters achieve the same performance as the GA–CSF filters suggests that the EV–CSF filters are still optimum. Finally, we note that GA–CSF and EV–CSF filters outperform O–CDD for all $L_g$ except for the extreme cases $L_g = 1$ and $L_g = 16$.

In Fig. 5.4, we consider the BER of BICM–OFDM with GA–CSF filtering vs. $E_b/N_0$ ($E_b$: received energy per information bit, $N_0$: power spectral density of underlying continuous–time passband noise process) and compare it with that of the CDD scheme proposed by Bauch and Malik (BM–CDD) in [87] and with BICM–OFDM concatenated with Alamouti’s STBC. We assume BPSK, $K = 64$, $N_T = 2$, $\rho_L = 0$, $\rho = \{0.0, 0.9\}$, and $L = \{1, 5\}$. For CSF filtering full–diversity–achieving filter lengths of $L_g = 2$ and $L_g = 6$ are assumed for $L = 1$ and $L = 5$, respectively. Furthermore, $8 \times 8$ and $16 \times 8$
block interleavers are adopted for CSF filtering and the STBC scheme, respectively. For BM–CDD, which uses a delay of $D = K/2 = 32$, we show results for the $8 \times 8$ interleaver and without interleaving. For $L = 1$ BM–CDD without interleaving achieves the same performance as CSF filtering. As explained in [87], for $L = 1$ the interleaver is not beneficial to BM–CDD and leads even to a small performance degradation. In contrast, for $L = 5$ interleaving and no interleaving is preferable for BM–CDD if $\rho = 0.0$ and $\rho = 0.9$, respectively, which underlines the importance of a channel–dependent joint optimization of interleaver and delay in BM–CDD. For $L = 5$ and $\rho = 0.9$ CSF filtering yields a gain of more than 2 dB at BER $= 10^{-5}$ even if interleaving is not applied in BM–CDD. While for $L = 1$ Alamouti’s code and GA–CSF achieve a similar performance, for $L = 5$ Alamouti’s code yields a performance gain compared to GA–CSF. Note, however, that unlike CSF filtering Alamouti’s code requires a change to the receiver structure.

Figure 5.3: Ratio $\alpha \triangleq f(g)/\tilde{f}(G_{opt})$ vs. $L_g$ for GA–CSF, EV–CSF, and O–CDD filters, respectively. $N_T = 4$, $\rho = 0.9$, and $L = 5$. 
5. Cyclic Space–Frequency Filtering for BICM–OFDM Systems

Figure 5.4: BER of BICM–OFDM with GA–CSF filtering, BM–CDD [87], and Alamouti’s STBC vs. $E_b/N_0$. OFDM, $K = 64$, BPSK, $N_T = 2$, and $\rho_L = 0$. GA–CSF filtering: $8 \times 8$ block interleaver, $L_g = 2$ for $L = 1$, $L_g = 6$ for $L = 5$. BM–CDD: $8 \times 8$ block interleaver and no interleaver. Alamouti’s STBC: $16 \times 8$ block interleaver. Convolutional code with $(171, 133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$.

compared to single–antenna transmission.

In Fig. 5.5, we show the BER of user 1 of a BICM–OFDMA system with GA–CSF filtering, O–CDD, BM–CDD, and C–CDD. We assume $U = 4$, $K = 512$, 16–QAM, $N_T = 4$, $L = 5$, $\rho_L = 0$, $\rho = \{0.6, 0.9\}$, and a $32 \times 16$ block interleaver. User $u \in \{1, 2, 3, 4\}$ is assigned subcarriers $u + 4k$, $0 \leq k \leq 127$. Since $N_T L > d_{\text{free}}$, the diversity gain is limited by the code. Therefore, for the GA–CSF filters and the O–CDD filters we adopted $L_g = L_{g,\text{min}} = 6$. In contrast, for BM–CDD and C–CDD we chose $D = K/(N_T U) = 32$ [87] and $D = L = 5$ [98], respectively, resulting in $L_g > L_{g,\text{min}}$ in both cases. Fig. 5.5 shows that the large delays of BM–CDD and C–CDD are not beneficial if the diversity gain is limited by the code. BM–CDD even suffers from a loss
in diversity (see Case 2 in Section 5.3.1 for an explanation). For $\rho = 0.6$ the proposed O–CDD scheme with optimized delays yields performance gains of 1.8 dB and 7 dB compared to C–CDD and BM–CDD at $BER = 10^{-5}$, respectively. An additional gain of 0.3 dB is possible with GA–CSF filtering. We note that the performance of both C–CDD and BM–CDD could be improved by optimizing the interleaver for the given delays.\(^5\)

![Figure 5.5: BER of BICM–OFDMA with GA–CSF filtering, O–CDD, BM–CDD [87], and C–CDD [98] vs. $E_b/N_0$. $U = 4$, $K = 512$, 16–QAM, $N_T = 4$, $L = 5$, $\rho_L = 0$, $32 \times 16$ block interleaver. Convolutional code with (171, 133)s, $R = 1/2$, and $d_{\text{free}} = 10$.](image)

\(^5\)We have not included results for the scheme in [95] in Figs. 5.4 and 5.5 since the simulation–based CSF filter optimization outlined in [95] is difficult to reproduce and the designs given in [95] were optimized for a convolutional turbo code, whereas we consider convolutional codes.
5.5.2 Distributed Antennas

We assume that the source–relay and the relay–destination channels are i.i.d., respectively. This implies that the distances between the relay nodes are much smaller than the source–relay and the relay–destination distances, respectively. Furthermore, we assume perfect synchronization throughout this section. The GA–CSF filters were optimized with the GA specified in Table 5.1 (unknown $S$) with $N_a = 2$, $\gamma = 15$, and adaptation step size $\mu[i + 1] = \lambda \mu[i]$, where $\mu[0] = 0.1$ and $\lambda = 0.9999$. We considered a system with $K = 128$, 16–QAM, and a $16 \times 32$ block interleaver throughout this section.

In Fig. 5.6, we consider the case $L = 1$ and $N_T = 2$ and compare the minimum coding gain $C_{\text{min}} \triangleq \min_S \{C(g_S)\}$ of sets of GA–CSF filters of various lengths $L_g$ with the corresponding upper bounds for $L_g = 2$ and known and unknown $S$ which are given by $C_{b,k} \triangleq \frac{1}{2} \sqrt{\det(W)}$ and $C_{b,u} \triangleq \sqrt{f_{1,\text{min}} \det(W)}$, respectively, cf. (5.12), (5.15), (5.29). For $L_g = L_{g,\text{min}} = 2$ the GA–CSF filters achieve the upper bound $C_{b,u}$ for $N \leq L_g^2 = 4$ as expected from Section 5.4.2. For $N > 4$ the upper bound cannot be achieved in principle, but the minimum coding gain of the GA–CSF filters closely follows the upper bound, which confirms the effectiveness of the proposed GA. For $L_g > L_{g,\text{min}} = 2$ considerable additional performance gains are possible for large $N$. For $N = 2$ $S = \mathcal{N}$ holds and the second GA in Table 5.1 calculates the GA–CSF filters for known $S$ and $L_g > L_{g,\text{min}}$, cf. Section 5.4.3. Fig. 5.6 shows that for known $S$ (i.e., $N = 2$) the performance gains achievable with $L_g > L_{g,\text{min}}$ are small which also validates our focus on $L_g \leq L_{g,\text{min}}$ in Section 5.3.

In Fig. 5.7, we show the BER of BICM–OFDM with GA–CSF, R–CSF, R–CDD, and C–CDD [86] filtering (as defined in Section 5.4.3) when $S$ is not known for the second phase of transmission. We assume that $N_T = N_a = 2$ out of $N = 10$ relay nodes forward the (error–free) decoded packet to the destination node over a flat fading
channel ($L = 1$) and the BER is averaged over all possible relay node combinations. For comparison we also show the performance of GA–CSF filtering when $S$ is known. Fig. 5.7 shows that CSF filtering with optimized filters (GA–CSF) and random filters (R–CSF) achieves full diversity. As expected from Fig. 5.6, increasing $L_g$ from 2 to 5 yields significant performance gains in both cases. For $L_g = 5$ GA–CSF with unknown $S$ closely approaches the performance of GA–CSF with known $S$. Both R–CDD and C–CDD achieve only diversity gain one since, in R–CDD, two active relays will select the same delay with non–zero probability and, in C–CDD, $L_g > L_{g,\text{min}}$ is valid in general but the interleaver is not optimized accordingly.

In Fig. 5.8, we investigate the impacts of imperfect synchronization (IS) and $N_T \neq N_a$, respectively, on the performance of GA–CSF filtering with $L_g = 6$ optimized for...
Figure 5.7: BER of BICM–OFDM with GA–CSF, R–CSF, R–CDD, and C–CDD [86] filtering when $\mathcal{S}$ is unknown vs. $E_b/N_0$. For comparison the BER of GA–CSF filtering when $\mathcal{S}$ is known is also shown. $N = 10$, $L = 1$, $N_T = N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. Convolutional code with $(171, 133)$, $R = 1/2$, and $d_{\text{free}} = 10$.

$N_a = 2$ and perfectly synchronized relay nodes. We assume $N = 10$ and $L = 2$, and consider again only the second phase of transmission. In case of imperfect node synchronization the transmit signal of node $n$ before the CP is added is given by $s_n[t + \tau_n]$, $n \in \mathcal{S}$, where delay $\tau_n$ is randomly chosen from $\{0, 1, 2, 3\}$. The delays of different nodes are independent. Fig. 5.8 shows that GA–CSF is robust against synchronization errors. C–CDD [86] even slightly benefits from imperfect node synchronization. Fig. 5.8 also shows that the performance of CSF filtering significantly improves if $N_T = 5$ nodes are active even if the GA–CSF filters were optimized for $N_T = 2$ active nodes.

In Fig. 5.9, we show the BER at the destination node for a two–hop transmission with $N = 10$ and $L = 1$. The energy per received bit at the relays in the first hop is identical to the energy per received bit $E_b$ at the destination and each relay listens to the source
with probability $p$. $p < 1$ may be caused by power saving protocols or devices that are permanently turned off. Only listening relays which can successfully decode the packet (corresponding to one OFDM symbol) received from the source participate in the second phase of transmission using GA–CSF, R–CSF, or C–CDD [86] filtering. If no relay can decode the source packet, the packet is retransmitted. Note that this decoding failure is not included in the BER curves in Fig. 5.9. For comparison, we also show the BER of GA–CSF filtering with known $S$ and for the case where only a single relay retransmits the packet ($N_T = 1$). In contrast to Fig. 5.7, the number of active relay nodes $N_T$ is now a random variable and only a single node will be active with a certain probability. For high SNR the latter case dominates the BER which explains why asymptotically
all BER curves are parallel to the curve for $N_T = 1$. However, CSF filtering and CDD still achieve significant performance gains compared to $N_T = 1$. For GA–CSF filtering the performance loss caused by the lack of knowledge of $S$ is remarkably small (i.e., less than one dB) for all considered cases.

![Graph](image)

Figure 5.9: BER at destination node for two–hop transmission with GA–CSF, R–CSF, and C–CDD [86] filtering when $S$ is unknown vs. $E_b/N_0$. For comparison the BER of GA–CSF filtering when $S$ is known is also shown. Each relay node listens to source node with probability $p$. $N = 10$, $L = 1$, $N_a = 2$, $K = 128$, 16–QAM, and $16 \times 32$ block interleaver. Convolutional code with $(171,133)_8$, $R = 1/2$, and $d_{\text{free}} = 10$.

5.6 Conclusions

In this chapter, we have proposed CSF filtering for BICM–OFDM systems with multiple co–located or distributed transmit antennas. CSF filtering may be viewed as an extension of well–known CDD where the cyclic delays are replaced by CSF filters. Similar to
CDD and unlike other SFC schemes, CSF filtering requires only one IFFT at the transmitter and no modifications at the receiver compared to single-antenna transmission. Based on the PEP of the overall system, we have derived optimization criteria for the CSF filters when the set of transmitting antennas is known and not known, respectively. For certain special cases we have provided closed-form solutions for the resulting optimization problems, and we have proposed various CSF filter design methods for the general case. We have shown that the proposed CSF filter designs are independent of the interleaver if the CSF filters do not exceed a certain length $L_{g,\min}$. If this length is exceeded, the interleaver has to be taken into account for CSF filter design to ensure full diversity. Our simulation results have shown that CSF filtering can achieve significant performance gains over various existing CDD designs at the expense of a small increase in complexity.
6 Robust Transmit Processing for BICM–OFDM Systems with Imperfect CSIT

6.1 Introduction

In the previous chapter, we have introduced CSF filtering for MIMO-OFDM which exploits statistical CSIT to improve system performance. In this chapter, we consider robust CSF filtering for BICM–OFDM systems with imperfect instantaneous CSIT, taking into account the reliability of the CSIT for transmitter design.

For flat fading channels, examples for successful exploitation of imperfect CSIT can be found in e.g. [14]–[17]. Furthermore, the effect of linear prediction of CSIT on the performance of beamforming in flat fading channels has been considered in [18, 19]. In [105], it has been shown that imperfect CSIT can also lead to significant performance improvements in frequency–selective channels. In [105], uncoded single–carrier modulation was considered which requires complex equalizers at the receiver. A more elegant approach to deal with frequency selectivity is OFDM which essentially converts

\[ \text{The material in this chapter was presented in part at IEEE International Conference on Communications, Dresden, Germany, June 2009.} \]
the frequency–selective channel into a number of parallel frequency–flat channels. To exploit the inherent diversity of frequency–selective channels, OFDM is often combined with BICM [44, 48].

If CSIT is not available, CDD or OSTBCs may be used in combination with BICM–OFDM to exploit both space and frequency diversity [48, 51]. Furthermore, statistical CSIT can be exploited if the simple cyclic delays in CDD are replaced by CSF filters [74], also cf. Chapter 5. For perfect CSIT, capacity maximizing CSF filter designs were considered in [106]. The performance of beamforming in BICM–OFDM systems with full and partial CSIT was investigated in [107] without taking into account the reliability of the CSIT for beamformer design. Furthermore, a robust adaptive modulation scheme for BICM–OFDM with imperfect CSIT was proposed in [108]. Per–subcarrier beamforming combined with space–time block coding was considered for uncoded (i.e., without FEC coding) MIMO–OFDM systems with imperfect CSIT in [109]. In [87], the channel capacity was used as performance measure to evaluate the achievable performance of MIMO–OFDM systems exploiting imperfect CSIT.

In this chapter, we consider robust CSF filtering for BICM–OFDM systems with imperfect CSIT and investigate the possibility of further improving robustness by combining CSF filtering with an OSTBC. The resulting scheme is referred to as space–frequency coded (SFC) filtering since the OSTBC is applied in the frequency domain. The contributions made in this chapter can be summarized as follows:

- We develop a unified mathematical description for the received signal for CSF and SFC–CSF filtering.

- Using this unified description, we derive an upper bound for the worst–case PEP based on a Bayesian model for the reliability of the CSIT, which is general enough to account for outdated and quantized CSIT, as well as CSIT whose quality has
been improved by linear prediction.

- We derive an optimization criterion for the robust CSF filters by exploiting the upper bound on the PEP.

- We show that for flat fading the resulting optimization problem can be solved using the tools from [15]. For frequency–selective fading an exact solution to the problem can only be found in special cases.

- For frequency–selective fading, we modify the original problem by introducing additional constraints and show that this modified problem can be solved exactly.

- For frequency–selective fading, we devise a gradient algorithm, which is initialized with the solution to the modified problem, in order to find an approximate solution to the original problem.

- Our simulation results show that the proposed robust CSF and SFC–CSF filtering schemes effectively exploit imperfect CSIT and that linear prediction is an effective means to improve the quality of imperfect CSIT.

The remainder of this chapter is organized as follows. In Section 6.2, the system model is established. The optimization problem is stated in Section 6.3, and corresponding robust CSF filter designs are presented in Section 6.4. Simulation results are given in Section 6.5, and some conclusions are drawn in Section 6.6.

### 6.2 System Model

We consider a BICM–OFDM system with $K$ subcarriers, $N_T$ transmit antennas, and $N_R$ receive antennas. In this section, we introduce the models for the channel, the CSF and SFC–CSF transceivers, and the CSIT.
6.2.1 Channel Model

The channel between transmit antenna \( n \) and receive antenna \( r \) is characterized by CIR \( h_{nr}[l], 0 \leq l < L \), of length \( L \). For convenience we collect the corresponding CIR coefficients in vector \( h_{nr} \triangleq [h_{nr}[0] \ h_{nr}[1] \ldots h_{nr}[L-1]]^T \). The CIR is assumed to be approximately constant within one OFDM symbol but may vary slowly between OFDM symbols. The CIR coefficients are modeled as possibly correlated complex Gaussian random variables (Ricean fading).

We define the \( N = N_R N_T L \) dimensional vector \( h \triangleq [h_{11}^T \ h_{21}^T \ldots h_{N_T N_R}^T]^T \) containing all CIR coefficients. The mean and the covariance matrix of \( h \) are denoted by \( m_h \triangleq \mathcal{E}\{h\} \) and \( C_{hh} \triangleq \mathcal{E}\{(h - m_h)(h - m_h)^H\} \), respectively. For the covariance matrix we adopt the popular Kronecker model\(^2\)

\[
C_{hh} = \frac{1}{1 + K_R} C_{N_R} \otimes C_{N_T} \otimes C_L,
\]

where \( C_{N_R}, C_{N_T}, \) and \( C_L \) denote the receive antenna, the transmit antenna, and the CIR coefficient covariance matrices, respectively, and \( K_R \) is the Ricean factor. Similar to [15] we assume in the following that, because of rich scattering and sufficient antenna spacing, the correlation at the receiver is negligible, i.e., \( C_{N_R} = I_{N_R} \).

6.2.2 Transceiver Structure for Robust CSF

The transceiver structure for robust CSF is shown in Fig. 6.1. The transceiver signal processing for robust CSF is similar to that for CSF in Section 5.2.2. Here, we repeat the processing for clarity. As customary in BICM–OFDM, the output bits of a binary convolutional encoder (minimum free distance \( d_{\text{free}} \)) are interleaved and mapped onto

\(^2\)We note that this model is suitable for Rayleigh fading but may not be always accurate for Ricean fading [110].
symbols $X[k] \in \mathcal{X}$, $0 \leq k < K$, where $\mathcal{X}$ denotes an $M$–ary symbol alphabet such as $M$–PSK or $M$–QAM. The effect of the interleaver can be modeled by the mapping $k' \rightarrow (k, i)$, where $k'$ denotes the original index of coded bit $c_{k'}$ and $k$ and $i$ denote the index of symbol $X[k]$ and the position of $c_{k'}$ in the label of $X[k]$, respectively. As usual, we assume that the interleaver is chosen such that coded bits $c_{k'_1}$ and $c_{k'_2}$ with $|k'_1 - k'_2| \leq d_{\text{free}}$ are mapped onto different symbols which are transmitted over different subcarriers [48]. An IDFT converts symbols $X[k], 0 \leq k < K$, into time–domain symbols $x[t], 0 \leq t < K$.

At transmit antenna $n, 1 \leq n \leq N_T$, the symbols $x[t]$ are passed through a CSF filter with impulse response $g_n[l], 0 \leq l < L_g$, of length $L_g \leq K$. The CSF filter output at antenna $n$ can be expressed as

$$s_n[t] = g_n[t] \odot x[t] = \sum_{l=0}^{L_g-1} g_n[l] x[(t - l)_K]. \quad (6.2)$$

After a CP of length greater than $L-2$ has been added, the $s_n[t], 0 \leq t < K, 1 \leq n \leq N_T$, are transmitted over the channel.

At the receiver, CSF filtering does not require any modifications compared to single–antenna transmission. After removal of the CP and DFT, the received signal at antenna
6. Robust Transmit Processing for BICM-OFDM Systems with Imperfect CSIT

$r, 1 \leq r \leq N_R$, can be expressed as

$$R_r[k] = H_{\text{eq},r}[k]X[k] + N_r[k], \quad 0 \leq k < K,$$

(6.3)

where the channel coefficient $H_{\text{eq},r}[k]$ is given by $H_{\text{eq},r}[k] = \sum_{n=1}^{N_T} H_{nr}[k]G_n[k]$, with $H_{nr}[k] \triangleq \sum_{l=0}^{L-1} h_{nr}[l]e^{-j2\pi kl/K}$ and $G_n[k] \triangleq \sum_{l=0}^{L_g-1} g_n[l]e^{-j2\pi kl/K}$, and $N_r[k]$ is AWGN with variance $\sigma_n^2 \triangleq \mathbb{E}\{|N_r[k]|^2\}$. Eq. (6.3) shows that CSF filtering transforms the $N_T \times N_R$ MIMO channel into an equivalent $1 \times N_R$ SIMO channel with frequency response $H_{\text{eq},r}[k]$ for receive antenna $r, 1 \leq r \leq N_R$. This corresponds to effective SIMO CIRs $h_{\text{eq},r}[l]$ of length $L_{\text{eq}} = L + L_g - 1$.

Next, optimal maximum ratio combining (MRC) of the received signals in each subcarrier is performed. The combined signal

$$R[k] = \sum_{r=1}^{N_R} H_{\text{eq},r}^*[k]R_r[k], \quad 0 \leq k < K,$$

(6.4)

is then used for bit metric calculation for the coded bit $c_{k'}$ [44, 48]

$$\lambda^i(R[k], c_{k'}) = \min_{X \in \mathcal{X}_b^i} \left\{ \left| R[k] - \sum_{r=1}^{N_R} |H_{\text{eq},r}[k]|^2 X \right|^2 \right\}^2,$$

(6.5)

where $\mathcal{X}_b^i$ denotes the subset of all symbols $X \in \mathcal{X}$ whose label has the value $b \in \{0, 1\}$ in position $i$. The bit metrics are de–interleaved and Viterbi decoded [44, 48].

6.2.3 Transceiver Structure for Robust SFC-CSF

Now, we turn our attention to the transceiver structure for robust SFC-CSF, cf. Fig. 6.2. In this case, $P$ bit streams are convolutionally encoded, interleaved, and mapped onto $M$-ary symbols $X^{(p)}[k], 1 \leq p \leq P, 0 \leq k < K/C$, where $K/C$ is an integer. Next,
symbols $X^{(p)}[k], 1 \leq p \leq P$, are encoded with an $N_S \times C$ OSTBC [10, 11] of rate $P/C$ to produce the $N_S \leq N_T$ symbol streams $X_\nu[k], 1 \leq \nu \leq N_S, 0 \leq k < K$. Although any OSTBC can be employed for SFC–CSF, in the following, for simplicity of exposition and practical relevance, we will concentrate on Alamouti’s STBC ($N_S = P = C = 2$) [10]. In this case, we obtain

$$X_1[2k] = X^{(1)}[k], X_2[2k] = X^{(2)}[k], X_1[2k + 1] = -(X^{(2)}[k])^*, \text{ and } X_2[2k + 1] = (X^{(1)}[k])^*, \ 0 \leq k < K/C.$$

Symbol streams $X_\nu[k], 1 \leq \nu \leq N_S$, are subsequently transformed via IDFT into time–domain symbols $x_\nu[t], 1 \leq \nu \leq N_S, 0 \leq t < K$. The time–domain symbols $x_\nu[t]$ are filtered with a CSF–matrix filter having coefficients $g_{\nu n}[l], 1 \leq \nu \leq N_S, 1 \leq n \leq N_T, 0 \leq l < L_g$. The resulting filter output at antenna $n$ is given by

$$s_n[t] = \sum_{\nu=1}^{N_S} g_{\nu n}[l] \otimes x_\nu[t] = \sum_{\nu=1}^{N_S} \sum_{l=0}^{L_g-1} g_{\nu n}[l] x_\nu[(t - l)K]. \quad (6.6)$$

Similar to the CSF filtering case, $s_n[t]$ is transmitted from antenna $n, 1 \leq n \leq N_T$, after a CP of length greater than $L - 2$ has been added. At the receiver, the CP is removed and the DFT is applied to the received signal. The resulting frequency domain signal can be expressed as

$$R_r[k] = \sum_{\nu=1}^{N_S} H_{eq,\nu r}[k] X_\nu[k] + N_r[k], \ 0 \leq k < K, 1 \leq r \leq N_R, \quad (6.7)$$

with $H_{eq,\nu r}[k] \triangleq \sum_{n=1}^{N_T} H_{nr}[k] G_{\nu n}[k]$ and $G_{\nu n}[k] \triangleq \sum_{l=0}^{L_g-1} g_{\nu n}[l] e^{-j2\pi kl/K}$. Subsequently, SF decoding is performed for each receive antenna. For example, for Alamouti’s STBC we obtain

$$R^{(1)}_r[k] = H^*_{eq,1 r}[2k] R_r[2k] + H_{eq,2 r}[2k] R^*_r[2k + 1], \quad (6.8)$$

$$R^{(2)}_r[k] = H^*_{eq,2 r}[2k] R_r[2k] - H_{eq,1 r}[2k] R^*_r[2k + 1]. \quad (6.9)$$
Assuming that the frequency response $H_{eq,vr}[k]$ is constant over $C$ neighboring subcarriers (i.e., $H_{eq,vr}[k] = H_{eq,vr}[k + c], 1 \leq c < C$), which is approximately valid for sufficiently narrow subcarrier spacing, (6.8) and (6.9) can be written as

$$R_r^{(p)}[k] = |H_{eq,r}[k]|^2 X^{(p)}[k] + \tilde{N}_r^{(p)}[k], \quad 1 \leq p \leq P, \quad 0 \leq k < K/C,$$

(6.10)

where $|H_{eq,r}[k]|^2 \triangleq \sum_{\nu=1}^{N_S} |H_{eq,\nu r}[Ck]|^2$ and $\tilde{N}_r^{(p)}[k]$ is AWGN with variance $|H_{eq,r}[k]|^2 \sigma_n^2$.

For each symbol stream $p, 1 \leq p \leq P$, the SF-decoded symbols are combined to obtain

$$R^{(p)}[k] = \sum_{r=1}^{N_R} R_r^{(p)}[k] = \sum_{r=1}^{N_R} |H_{eq,r}[k]|^2 X^{(p)}[k] + \sum_{r=1}^{N_R} \tilde{N}_r^{(p)}[k].$$

(6.11)

Since the $P$ signal streams $R^{(p)}[k], 1 \leq p \leq P$, are decoded separately, for simplicity of notation, we will drop in the following superscript “$(p)$” in $R^{(p)}[k]$ and $X^{(p)}[k]$. For decoding of the $P$ information bit streams the same bit metric as in (6.5) can be applied, where $\sum_{r=1}^{N_R} |H_{eq,r}[k]|^2 = \sum_{r=1}^{N_R} \sum_{\nu=1}^{N_S} |H_{eq,\nu r}[Ck]|^2$. In other words, SFC–CSF transforms the $N_T \times N_R$ MIMO channel into a $1 \times N_R N_S$ SIMO channel. The corresponding equivalent CIRs $h_{eq,\nu r}[k]$ have again length $L_{eq} = L + L_g - 1$ and are the inverse Fourier transforms of $H_{eq,\nu r}[k], 1 \leq \nu \leq N_S, 1 \leq r \leq N_R$.

The SFC–CSF scheme described above employs space–time coding in the frequency domain, cf. Fig. 6.2. We note, however, that the extension of CSF filtering to the case where the space–time code is employed in the time domain is also possible, cf. [111].

---

3We note that the assumption that $H_{eq,vr}[k]$ is constant over $C$ neighboring subcarriers is made only to simplify the analysis and the optimization of the CSF filters. However, for the simulation results shown in Section 6.5, this restriction was not imposed and the simulated channels change from one subcarrier to the next.
6.2.4 Bayesian Statistics for Imperfect CSIT

We assume that the channel state information is perfectly known at the receiver, whereas an imperfect estimate of the channel is available at the transmitter via a dedicated feedback channel, cf. Figs. 6.1 and 6.2. The true channel vector $h$ and the instantaneous CSIT vector $\gamma$ are assumed to be jointly complex Gaussian. In this case, the probability density function $p_{h|\gamma}(h|\gamma)$ of the true channel conditioned on $\gamma$ is also complex Gaussian,

$$p_{h|\gamma}(h|\gamma) = \frac{\exp \left( - (h - m_{h|\gamma})^H C^{-1}_{hh|\gamma} (h - m_{h|\gamma}) \right)}{\pi^N \det(C_{hh|\gamma})}.$$  \hfill (6.12)

The conditional mean vector $m_{h|\gamma}$ and the conditional covariance matrix $C_{hh|\gamma}$ can be calculated as [112]

$$m_{h|\gamma} = \mathbb{E}\{h|\gamma\} = m_h + C_{h\gamma} C^{-1}_{\gamma\gamma} (\gamma - m_\gamma),$$  \hfill (6.13)

$$C_{hh|\gamma} = \mathbb{E}\{(h - m_{h|\gamma})(h - m_{h|\gamma})^H | \gamma\} = C_{hh} - C_{h\gamma} C^{-1}_{\gamma\gamma} C_{\gamma h},$$  \hfill (6.14)
where $C_{\gamma\gamma}$ and $C_{h\gamma}$ are the covariance matrix of $\gamma$ and the cross-covariance matrix of $h$ and $\gamma$, respectively. Furthermore, $C_{\gamma h} = C_{h\gamma}^H$ and $m_\gamma = \mathcal{E}\{\gamma\}$ hold. The Bayesian model in (6.12) contains the special cases of purely statistical CSIT ($m_{h|\gamma} = m_h$, $C_{h h|\gamma} = C_{hh}$) and perfect CSIT ($m_{h|\gamma} = \gamma$, $C_{h h|\gamma} = 0_{NN}$). While the considered Bayesian model is quite general, we concentrate in the following on the practically relevant cases of outdated and quantized CSIT. We also consider the possibility of improving the quality of the CSIT by linear prediction.

1) Outdated and Quantized CSIT: Outdated and quantized CSIT can be modeled as [14, 16, 105]

$$\gamma = \rho(h - m_h) + \sqrt{1 - \rho^2} h_{\text{aux}} + m_h + e,$$

(6.15)

where correlation coefficient $\rho$ may be smaller than one due to a delay in the feedback channel and the error vector $e$ is caused by quantization of $h - m_h$.\footnote{We assume that the mean $m_h$ changes very slowly with time compared to the diffuse component $h - m_h$. Therefore, $m_h$ can be quantized with high resolution and the related quantization error is negligible.} True channel vector $h$, auxiliary vector $h_{\text{aux}}$, and $e$ are mutually statistically independent. Furthermore, $h_{\text{aux}}$ is a zero-mean Gaussian random vector with covariance matrix $C_{hh}$. For the classical Jakes’ model [1], the correlation coefficient is given by $\rho = J_0(2\pi f_D T)$, where $J_0(\cdot)$, $f_D$, and $T$ denote the zeroth order Bessel function of the first kind, the maximum Doppler frequency, and the delay in the feedback channel, respectively. Assuming scalar quantization, since the elements of $h - m_h$ are zero-mean Gaussian random variables, the elements of $e$ can be modeled as independent zero-mean Gaussian random variables, whose variances are proportional to the variances of the respective elements of $h$ [105,
Therefore, based on (6.1), the covariance matrix of $e$ can be modeled as

$$C_{ee} = \frac{\xi_e}{1 + K_R} I_{NR} \otimes D_{NT} \otimes D_L,$$

(6.16)

where $D_{NT} \triangleq \text{diag}\{d_{T,1}, d_{T,2}, \ldots, d_{T,N_T}\}$ and $D_L \triangleq \text{diag}\{d_{L,1}, d_{L,2}, \ldots, d_{L,L}\}$ are diagonal matrices, whose main diagonal entries are identical to the main diagonal entries of $C_{NT}$ and $C_L$, respectively. Note that $\xi_e$ and the main diagonal entries of $D_{NT}$ and $D_L$ determine the variance of the quantization error. For example, the variance of the first element of $e$ is given by $\sigma_{e,1}^2 = \xi_e d_{T,1} d_{L,1} / (1 + K_R)$.

2) Linear Prediction: If $N_P$ outdated and quantized CSIT vectors $h[\tau - i] - m_h + e[\tau - i], 1 \leq i \leq N_P,$ are available, the quality of the CSIT at time $\tau$ may be improved by post-processing. Here, $h[\tau]$ and $e[\tau]$ denote the channel vector and the quantization error vector at time $\tau$, respectively. The processed CSIT vector can be expressed as

$$\gamma = \sum_{i=1}^{N_P} q_i(h[\tau - i] - m_h + e[\tau - i]) + m_h,$$

(6.17)

where $q_i$ denotes the $i$th coefficient of an $N_P$th order linear predictor. Let $q \triangleq [q_1 \ q_2 \ldots \ q_{N_P}]^T$, $R \triangleq \mathcal{E}\{\tilde{H}_e^H \tilde{H}_e\}$, $r \triangleq \mathcal{E}\{\tilde{H}_e^H h[\tau]\}$, $\tilde{H}_e \triangleq [h[\tau-1] + e[\tau-1] \ldots \ h[\tau-N_P] + e[\tau-N_P]]$, and $\tilde{h}[\tau] \triangleq h[\tau] - m_h$. With these definitions it is easy to show that the predictor coefficient vector minimizing the mean square error $\mathcal{E}\{||h[\tau] - \tilde{H}_e q||^2\}$ is given by

$$q = R^{-1} r.$$

(6.18)

The processed CSIT vector can be modeled as

$$\gamma = \chi_h(h - m_h) + \sqrt{\chi_h^2 - \chi_{h_{aux}}^2} h_{aux} + m_h + \chi_e e,$$

(6.19)
where $h$, $h_{\text{aux}}$, and $e$ have been defined in the previous subsection. Furthermore, a comparison of (6.17) and (6.19) reveals that $\chi_h \triangleq r^H R^{-1} r / \sigma_h^2$, $\chi_0 \triangleq r^H R^{-1} R_{hh} R^{-1} r / \sigma_h^2$, and $\chi_e \triangleq \sqrt{q^H q}$, where $R_{hh} \triangleq \mathcal{E}\{\hat{H}^H \hat{H}\}$, $\hat{H} \triangleq [\hat{h}[\tau - 1] \ldots \hat{h}[\tau - N_P]]$, and $\sigma_h^2 \triangleq \mathcal{E}\{\hat{h}^H[\tau] \hat{h}[\tau]\}$.

From (6.19) we obtain $m_{\gamma} = m_h$, $C_{h\gamma} = \chi_h^* C_{hh}$, and $C_{\gamma\gamma} = \chi_0^2 C_{hh} + \chi_e^2 C_{ee}$.

Using these results in (6.13) and (6.14), the conditional mean vector and the conditional covariance matrix for outdated and quantized CSIT can be expressed as

$$m_{h|\gamma} = m_h + \chi_h^* C_{hh} (C_{hh} + C_{ee})^{-1} (\gamma - m_h) \quad \text{and} \quad C_{hh|\gamma} = I_{N_R} \otimes C,$$  \hfill (6.20)

where we have introduced the definition

$$C \triangleq \frac{1}{1 + K_R} (C_{N_T} \otimes C_L) (I_{N_T} - |\chi_h|^2 (|\chi_0|^2 C_{N_T} \otimes C_L + \xi_e \sigma_e C_{N_T} \otimes D_L)^{-1} (C_{N_T} \otimes C_L)).$$ \hfill (6.21)

We note that (6.20) and (6.21) are also valid for outdated and quantized CSIT without linear prediction if we set $\chi_h = \rho$, $\chi_0 = 1$, and $\chi_e = 1$. Furthermore, the special cases of outdated CSIT without quantization error and quantized CSIT without feedback delay are obtained by letting $\chi_h = \rho$, $\chi_0 = 1$, and $\chi_e = 0$ and $\chi_h = 1$, $\chi_0 = 1$, and $\chi_e = 1$, respectively. In other words, (6.19)–(6.21) constitute a unified representation of the different imperfect CSIT models considered in this chapter.

### 6.3 Optimization of the Robust CSF Filters

In this section, we derive the worst–case PEP of BICM–OFDM with CSF and SFC–CSF filtering taking into account imperfect CSIT. Based on this result we formulate the CSF filter optimization problem. To enable a unified representation of robust CSF and robust
SFC–CSF filtering, in the sequel, we formally treat CSF as a special case of SFC–CSF with \( C = P = 1, N_S = 1 \), and \( g_{n1} = g_n, 1 \leq n \leq N_T \).

### 6.3.1 Optimization Problem

The proposed optimization criterion for the robust CSF filters is based on the expected value of the worst-case PEP with respect to the Bayesian statistics \( p_{h|\gamma}(h|\gamma) \). The PEP of two code words \( c \) and \( \hat{c} \) conditioned on \( h \) and \( g \triangleq [g_1^T \ldots g_{NS}^T]^T \), \( g_\nu \triangleq [g_{1\nu} \ldots g_{N_T\nu}]^T \), \( g_{n\nu} \triangleq [g_{n\nu}[0] \ldots g_{n\nu}[L_g - 1]]^T \), can be upper bounded as [69, 74]

\[
P(c \rightarrow \hat{c}|h, g) \leq \frac{1}{2} \exp \left( -\frac{d_{\text{min}}^2}{4\sigma_n^2} \sum_{k,d\text{free}} |H_{eq}[k]|^2 \right), \tag{6.22}
\]

where \( |H_{eq}[k]|^2 \triangleq \sum_{r=1}^{NR} \sum_{\nu=1}^{NS} |H_{eq,\nu r}[Ck]|^2 \), \( d_{\text{min}} \) denotes the minimum Euclidean distance of the signal constellation \( \mathcal{X} \), and the sum in the argument of the exponential function in (6.22) is over \( d_{\text{free}} \) distinct subcarriers \( \mathcal{K} \triangleq \{k_1, k_2, \ldots, k_{d_{\text{free}}}\} \). For the following, it is helpful to rewrite this sum as

\[
\sum_{k,d_{\text{free}}} |H_{eq}[k]|^2 = \sum_{k,d_{\text{free}}} \sum_{r=1}^{NR} \sum_{\nu=1}^{NS} \left| \sum_{n=1}^{NT} a^H[k] G_{n\nu} h_{nr} \right|^2
\]

\[
= h^H(I_{NR} \otimes G^H(I_{NS} \otimes A)G)h, \tag{6.23}
\]

where \( a[k] \triangleq [e^{j2\pi Ck/K} \ldots e^{j2\pi Ck(L_e q - 1)/K}]^T \), \( A \triangleq \sum_{k,d_{\text{free}}} a[k] a^H[k] \), \( G \triangleq [G_1^T \ldots G_{NS}^T]^T \), \( G_\nu \triangleq [G_{1\nu} \ldots G_{N_T\nu}] \), and \( G_{n\nu} \) is an \( L_{eq} \times L \) column–circular matrix with vector \([g_{n\nu}^T 0_{L-1}^T]^T\) in its first column, i.e., \( G \) is an \( N_S L_{eq} \times N_T L \) matrix. Defining \( \alpha \triangleq \frac{d_{\text{min}}^2}{4\sigma_n^2} \) and averaging the PEP with respect to the Bayesian statistics \( p_{h|\gamma}(h|\gamma) \) depending on
the CSIT vector $\gamma$ yields

$$P(c \to \hat{c}|\gamma, g) = \int_{-\infty}^{\infty} P(c \to \hat{c}|h, g)p_{h|\gamma}(h|\gamma)dh$$

$$\leq \exp\left(-\frac{m_{h|\gamma}^H(C_{hh|\gamma}^{-1}-W^{-1})m_{h|\gamma}}{2 \det(I_{N_TN_RL} + \alpha C_{hh|\gamma}S)}\right),$$

(6.24)

where $\bar{W} \triangleq C_{hh|\gamma} + \alpha C_{hh|\gamma}SC_{hh|\gamma}$ and $S \triangleq I_{NR} \otimes G^H(I_{NS} \otimes A)G$. For typical interleavers, $A$ can be well approximated by the scaled identity matrix $d_{free}I_{Leq}$ [48]. Using this approximation and exploiting (6.21), we can rewrite (6.24) as

$$P(c \to \hat{c}|\gamma, g) \leq \exp\left(-\frac{tr\{M^H(C^{-1} - W^{-1})M\}}{2(\det(I_{N_TL} + \beta CG^H)_{NR}}\right),$$

(6.25)

where $W \triangleq \beta CG^HGC + C$, $\beta \triangleq d_{free}\alpha$, and $M$ is an $N_TL \times N_R$ matrix whose element in row $\mu$ and column $\nu$ is the $[(\nu - 1)N_TL + \mu]$th element of vector $m_{h|\gamma}$. Taking the logarithm of (6.25) and ignoring irrelevant constant terms leads to the cost function

$$J(G) = tr\{M^HW^{-1}M\} - N_R \log \det(W).$$

(6.26)

The corresponding optimization problem for the robust CSF filter vector can be formulated as

$$\min_G J(G)$$

s.t.

$$W = \beta CG^HGC + C$$

$$\text{tr}\{G^HG\} = N_RL,$$

(6.27)

where the second constraint limits the transmit power to enable a fair comparison of different schemes. Unfortunately, problem (6.27) is not a convex optimization probl-
6. Robust Transmit Processing for BICM–OFDM Systems with Imperfect CSIT

Furthermore, although problem (6.27) appears to be very similar to problem [15, Eq. (15)], (6.27) is much more difficult to solve in general because of the special structure of $G$, cf. definition of $G$. This special structure makes an exact solution of problem (6.27) in general intractable.

### 6.3.2 Solution for Unstructured $G$

Although $G$ has in general a special structure, it is useful for the further development to solve problem (6.27) first for unstructured $G$. In this chapter, we mean by “unstructured” that all elements of $G$ can be freely chosen. In this case, the solution to (6.27) can be found using a modified version of the method in [15], which we shall refer to as Vu’s method in this chapter. To this end, we define matrix $\Phi \triangleq C^{-1} W C^{-1} - C^{-1} = \beta G^H G$.

We can now transform the non–convex problem (6.27) into the convex problem

$$\min_{\Phi} \tilde{J}(\Phi)$$

s.t. \[
\begin{align*}
\text{tr}\{\Phi\} &= \eta \\
\Phi &\succeq 0 \\
\text{rank}\{\Phi\} &\leq N_S L_{eq}
\end{align*}
\] where $\tilde{J}(\Phi) = J(G)$ and $\eta \triangleq \beta N_R L$. While the first two constraints in (6.28) were also considered in [15], the rank constraint on $\Phi$ is new but necessary since here $G$ is not a square matrix in general. Nevertheless, Vu’s method is still applicable with a minor modification. Without reproducing the details from [15] we only briefly outline in the following how the optimal unstructured matrix $G_{u,\text{opt}}$ is computed.

For this purpose, we define

$$\Phi(\nu) \triangleq \frac{1}{2\nu} (S I_{N_T L} + \Psi^{1/2}(\nu)) - C^{-1}$$

(6.29)
where $\nu$ is a parameter, $S \triangleq N_R$, and

$$
\Psi(\nu) \triangleq S^2 I_{NTL} + 4\nu C^{-1} M M^H C^{-1}.
$$

Furthermore, we order the eigenvalues $\sigma_i(\nu)$ of $\Phi(\nu)$ in descending order, i.e., $\sigma_i(\nu) \geq \sigma_{i+1}(\nu)$, $1 \leq i < NT_L$. To meet the first constraint in (6.28), $\nu$ has to fulfill

$$
\sum_{i=1}^{s} \sigma_i(\nu) = \eta,
$$

where $s$ is the largest integer smaller than or equal to $NSL_{eq}$ for which $\sigma_i(\nu) \geq 0$, $1 \leq i \leq s$, holds. This condition guarantees that the second and third constraints in (6.28) are also fulfilled. For a given $s$, $\nu$ can be easily found with a simple one dimensional search. The main difference to [15] is that we restrict $s$ to $s \leq NSL_{eq}$ because of the rank constraint on $\Phi$. Once $s$ and $\nu$ are determined, the optimum unstructured CSF matrix can be obtained from

$$
G_{u,opt} = \frac{1}{\sqrt{\beta}} \Sigma^{1/2} U^H,
$$

where $\Sigma$ is an $NSL_{eq} \times NSL_{eq}$ diagonal matrix having the $s$ largest eigenvalues of $\Phi(\nu)$ and $NSL_{eq} - s$ zeros in the main diagonal and the rows of $NT_L \times NSL_{eq}$ matrix $U$ contain the eigenvectors corresponding to the $NSL_{eq}$ largest eigenvalues.

The solution for unstructured matrices $G$ developed in this section will be used to find the exact solution for flat fading ($L = 1$) and approximate solutions for frequency-selective fading ($L > 1$) in the next section. Furthermore, $J(G_{u,opt})$ can be used as a lower bound for the cost function achievable with structured $G$. 

6. Design of Robust CSF Filters

In this section, we discuss the optimization of the robust CSF filters for imperfect CSIT. We first show that problem (6.27) can be solved exactly in special cases, and discuss subsequently approximate solutions for the general case.

6.4.1 Special Cases

In the following, we discuss four cases where the exact solution to problem (6.27) can be found or the problem has been already solved in the literature.

1) Frequency–flat channel ($L = 1$): For $L = 1$, matrix $G$ is given by

\[ G = \begin{bmatrix}
g_{11} & g_{21} & \cdots & g_{N1}
g_{12} & \cdots & \cdots & \cdots 
g_{1N_S} & g_{2N_S} & \cdots & g_{N_TN_S}
\end{bmatrix}, \quad (6.33) \]

i.e., it does not have a special structure and all its elements can be freely chosen. Therefore, we can directly apply the method described in Section 6.3.2 to find the exact solution $G_{\text{opt}} = G_{u,\text{opt}}$ to problem (6.27).

2) Perfect CSIT ($\chi_h = \chi_0 = 1$, $\chi_e = 0$): In this case, $C_{hh|\gamma} = 0_{N \times N}$ and $m_{h|\gamma} = \gamma = h$. After some manipulations, (6.23) can be rewritten as

\[ \sum_{k,d_{\text{free}}} |H_{eq}[k]|^2 = d_{\text{free}} \mathbf{g}^H (I_{NS} \otimes \sum_{r=1}^{NR} H_{r}^H H_{r}) \mathbf{g}, \quad (6.34) \]

where we have used $A \approx d_{\text{free}} I_{L_{eq}}$, $H_r \triangleq [H_{1r} \ldots H_{N_{Tr}}]$, and $H_{nr}$ is an $L_{eq} \times L_g$ column–circulant matrix with vector $[h_{nr}^T \mathbf{0}^T_{L_g-1}]^T$ in its first column. With this result,
the PEP in (6.22) simplifies to

$$P(c \rightarrow \hat{c} | \gamma, g) \leq \frac{1}{2} \exp \left( -\beta g^H Z g \right),$$

(6.35)

where $Z \triangleq I_{N_S} \otimes \sum_{r=1}^{N_R} H_r^H H_r$. In order to minimize the PEP, we have to maximize $g^H Z g$. It can be easily shown that in this case the optimal $g$ is simply that eigenvector of $Z$ which corresponds to its largest eigenvalue. Note that because of the special structure of $Z$, for perfect CSIT, $g_1 = g_2 = \ldots = g_{N_S}$, i.e., the CSF filter vectors for all branches of the OSTBC are equal.

3) High SNR ($\sigma_n^2 \rightarrow 0$): Assuming for the moment that $CG^H G$ has maximum rank $G_d \triangleq N_T L$, we obtain for $\sigma_n^2 \rightarrow 0$ (which implies $\beta \rightarrow \infty$) from (6.25)

$$P(c \rightarrow \hat{c} | \gamma, g) \leq \frac{\exp \left( -\text{tr} \{ M^H C C^{-1} M \} \right)}{2 \beta^{G_d N_R} \left( \det(CG^H G) \right)^{N_R}}. \quad (6.36)$$

Since $\beta$ is proportional to the SNR, (6.36) reveals that for high SNR and full rank $CG^H G$ on a double-logarithmic scale all error rate curves have a slope of $-N_T N_R L$, i.e., the diversity gain is $N_T N_R L$. Furthermore, for PEP minimization it suffices to maximize $\det(CG^H G)$ under a power constraint, i.e., the optimum $G$ is independent of the instantaneous CSIT $\gamma$. Note that these observations are valid regardless of the quality of the CSIT as long as the CSIT is not perfect. However, the minimum finite SNR required for the above (asymptotic) conclusions to be meaningful rapidly increases with the quality of the CSIT. In fact, our results in Section 6.5 clearly show that even for moderately reliable CSIT the optimum CSF filters strongly depend on the instantaneous CSIT for SNRs of practical interest. Nevertheless, it is still of interest to note that the maximum diversity gain of $N_T N_R L$ can only be achieved if $G^H G$ has full rank, which requires $N_T L \leq N_S L_{eq}$. Thus, the minimum CSF filter length necessary for full diversity
is

\[ L_{g,\text{min}} \triangleq \lceil L(N_T/N_S - 1) \rceil + 1. \]  (6.37)

For example, for \( N_T = 3 \) and \( L = 2 \), CSF filters of lengths \( L_{g,\text{min}} = 5 \) and \( L_{g,\text{min}} = 2 \) are required to achieve full diversity for pure CSF filtering and SFC–CSF filtering with \( N_S = 2 \), respectively. The shorter filter lengths for SFC–CSF filtering are due to the fact that in this case some of the spatial diversity can be exploited by the space–frequency code, whereas for pure CSF filtering the CSF filters have to convert the entire spatial diversity into frequency diversity before it can be exploited by the convolutional code.

4) Completely unreliable instantaneous CSIT (\( \chi_h = 0 \)): In this case only statistical CSIT (i.e., \( m_h, C_{hh}, \) and \( C_{ee} \) are known) is available. For the special case \( \chi_e = 0 \) (no quantization error), \( N_S = 1 \) (CSF without OSTBC), Rayleigh fading (\( m_h = 0_N \)), and high SNR the corresponding optimization problem in (6.27) is equivalent to the problem solved in Chapter 5. However, for the general case of Ricean fading, quantization error, and/or SFC–CSF filtering a solution to (6.27) is not known even for completely unreliable instantaneous CSIT.

Since in general an exact solution to (6.27) is not possible, we introduce in the next subsection additional constraints on \( g \) which facilitate the solution of the problem.

## 6.4.2 Solution of Problem with Additional Constraints

The basic idea of the proposed approach is to introduce additional constraints such that Vu’s method becomes again applicable. To this end, we choose \( L_g = \xi L + 1, \xi \geq 0, \) and make the common assumption that for a given transmit–receive antenna pair the CIR coefficients are uncorrelated, i.e., \( C_L = I_L \), which also implies \( D_L = I_L \). In this case,
the conditional covariance matrix can be expressed as \( C_{hh|\gamma} = I_{N_R} \otimes \tilde{C} \otimes I_L \) with
\[
\tilde{C} \triangleq \frac{1}{1 + K_R} C_{NT} (I_{NT} - |\chi_h|^2 (|\chi_0|^2 C_{NT} + \xi_e |\chi_e|^2 D_{NT})^{-1} C_{NT}). \quad (6.38)
\]
Furthermore, we constrain all elements of \( g_{n\nu}, 1 \leq n \leq N_T, 1 \leq \nu \leq N_S \), to be zero except for \( g_{n\nu}[\mu L], 0 \leq \mu \leq \xi \). With these constraints we can express \( G \) as \( G = \tilde{G} \otimes I_L \), where \( \tilde{G} \triangleq [\tilde{G}_1^T \ldots \tilde{G}_{N_S}^T]^T, \tilde{G}_\nu \triangleq [g_{1\nu} \ldots g_{N_T\nu}], \) and \( \tilde{g}_{n\nu} \triangleq [g_{n\nu}[0] \ g_{n\nu}[1] \ldots g_{n\nu}[\xi L]]^T \).

Using these new expressions for \( G \) and \( C \) the cost function \( J(\tilde{G}) \), which is now only a function of the \( N_S(\xi + 1) \times N_T \) matrix \( \tilde{G} \), can be simplified to
\[
J(\tilde{G}) = \text{tr}\{\tilde{M}^H \tilde{W}^{-1} \tilde{M}\} - N_R L \log \det(\tilde{W}), \quad (6.39)
\]
where \( \tilde{W} \triangleq \beta \tilde{C} \tilde{G}^H \tilde{G} \tilde{C} + \tilde{C} \). \( \tilde{M} \) is an \( N_T \times N_R L \) matrix whose element in row \( \mu \) and column \( \nu \) is the \((\nu - 1)N_T + \mu \)th element of vector \( m_{h|\gamma} \).

Consequently, we can now define \( \tilde{\Phi} \triangleq \tilde{C}^{-1} \tilde{W} \tilde{C}^{-1} - \tilde{C}^{-1} = \beta \tilde{G}^H \tilde{G} \) and formulate the new optimization problem
\[
\min_{\tilde{\Phi}} \tilde{J}(\tilde{\Phi}) \quad (6.40)
\]
\[
\begin{align*}
\text{s.t.} \quad & \text{tr}\{\tilde{\Phi}\} = \tilde{\eta} \\
& \tilde{\Phi} \succeq 0 \\
& \text{rank}\{\tilde{\Phi}\} \leq N_S(\xi + 1)
\end{align*}
\]
where \( \tilde{J}(\tilde{\Phi}) = J(\tilde{G}) \) and \( \tilde{\eta} \triangleq \beta N_R \). Since \( \tilde{G} \) is unstructured, i.e., all its elements can be freely chosen, problem (6.40) can be solved exactly using the modified version of Vu’s method outlined in Section 6.3.2. Compared to Section 6.3.2 we only have to replace \( \eta \) by \( \tilde{\eta} \), redefine \( S \) as \( S = N_R L \), and note that \( s \) in (6.31) is now the largest integer smaller or equal \( N_S(\xi + 1) \). Once the optimal \( \tilde{G}_{opt} \) is obtained, the corresponding coefficient
vectors $g_{n\nu}$, $1 \leq n \leq N_T$, $1 \leq \nu \leq N_S$, can be easily constructed. We note that, due to the additional constraint, which forces certain elements of $g_{n\nu}$ to zero, these coefficient vectors are not the solution to the original problem (6.28). Nevertheless, the constrained solution still achieves high performance in general and can be further improved using the gradient algorithm described in the next section.

### 6.4.3 Gradient Algorithm (GA)

In this section, we provide a GA which can be used to improve an initial estimate $g_{\text{init}}$ for the optimum vector $g$. Based on (6.26) we obtain for the gradient of $J(G)$

$$
\frac{\partial J(G)}{\partial g^*} \triangleq \begin{pmatrix} \frac{\partial J(G)}{\partial g_{11}[0]} & \frac{\partial J(G)}{\partial g_{11}[1]} & \cdots & \frac{\partial J(G)}{\partial g_{N_T N_S}[L_g - 1]} \end{pmatrix}^H,
$$

$$
\frac{\partial J(G)}{\partial g_{n\nu}[l]} = -\beta m_{h[l]}^H P Q P m_{h[l]} - \beta \text{tr}\{P C_{h[l]} H Q\},
$$

where $P \triangleq (I_N + \beta C_{h[l]} H S)^{-1}$ and $Q \triangleq I_N \otimes (E_{n\nu}[l] G)$. The entries of $N_S L_{eq} \times N_T L$ matrix $E_{n\nu}[l]$ are equal to 1 at those positions where $G$ has entries $g_{n\nu}[l]$ and zero otherwise. In iteration $i$, the CSF filter vector is updated as follows:

$$
g_{0[i + 1]} = g[i] - \mu[i] \frac{\partial J(G[i])}{\partial g^*[i]},
$$

$$
g[i + 1] = \frac{g_{0[i + 1]}^H g_{0[i + 1]} g_{0[i + 1]}}{\sqrt{g_{0[i + 1]}^H g_{0[i + 1]} g_{0[i + 1]}}}.
$$

where $\mu[i]$ is a step size and $G[i]$ is constructed from $g[i]$ in the same way as $G$ is from $g$. The above GA has to be performed for every CSIT vector $\gamma$ for a given initial vector $g[0] = g_{\text{init}}$ before data transmission. Since problem (6.27) is not convex in general, there is no guarantee that the GA finds the global optimal solution. However, if the step size is appropriately chosen, the GA will find that local optimum which is closest
to the initial vector \( g_{\text{init}} \). Therefore, the choice of \( g_{\text{init}} \) is crucial for the performance of the GA. Since the solution to the constraint problem (6.40) provides already a “good” approximate solution to the original problem (6.27), we use the solution to the constraint problem for initialization of the GA. For the step size, we found that a decreasing step size \( \mu[i + 1] = \sigma \mu[i] \) with \( 0 < \sigma < 1 \) achieves good results for the problem at hand.

### 6.5 Simulation Results

In this section, we present simulation results for robust CSF and SFC–CSF filtering with Alamouti’s OSTBC. Throughout this section, we assume \( N_R = 1 \) and CIRs with independent Rayleigh fading taps having identical variances. The transmit antennas are mutually correlated and we assume \( \mathbf{C}_{N_T} \), \( i,j \) = \( \rho_{N_T}^{[i-j]} \), where \( \rho_{N_T} = 0.6 \) denotes the spatial correlation coefficient. For all BER results shown, we adopted the quasi–standard convolutional code \((177, 133)_8\) with rate \( R = 1/2 \) and free distance \( d_{\text{free}} = 10 \), 4–PSK modulation, and a DFT size of \( K = 128 \). For interleaving we adopt the \( 16 \times 16 \) block interleaver from [115]. All results shown in this section were averaged over 200,000 realizations of the fading channels. The adaptation step size of the GA was chosen as \( \mu[i + 1] = \sigma \mu[i] \), where \( \mu[0] = 0.02 \) and \( \sigma = 0.95 \), and the number of iterations was limited to 100. In order to be able to exploit the maximum diversity gain offered by the channel, we use \( L_g = L_{g,\text{min}} \), cf. (6.37), throughout this section unless specified otherwise.

For clarity of presentation, we will first discuss the performance of robust CSF filtering in Subsection 6.5.1, and subsequently investigate the benefits of additional space–frequency coding and linear prediction in Subsections 6.5.2 and 6.5.3, respectively.
6.5.1 Robust CSF Filtering

In this subsection, we consider outdated CSIT, i.e., $\chi_e = 0$, $\chi_0 = 1$, $\chi_h = \rho$, and $N_T = 3$ and CIRs of length $L = 2$.

In Fig. 6.3, the BER of BICM–OFDM with robust CSF filtering vs. $E_b/N_0$ ($E_b$: received energy per information bit; $N_0$: power spectral density of underlying passband noise process) is shown for a CSF filter length of $L_g = 5$. We show results for robust CSF filters obtained by solving constrained problem (6.40) using the modified version of Vu’s method, cf. Section 6.4.2. The corresponding BER curves are labeled with ”VU”. Furthermore, we show results for the case where the solution to the constrained problem is improved with the GA introduced in Section 6.4.3. These BER curves are labeled with ”VU+GA”. As can be observed from Fig. 6.3, the improvement achievable with the GA increases as $\rho$ increases, i.e., as the CSIT becomes more reliable. Fig. 6.3 also shows that outdated CSIT enables significant performance gains compared to the case when only statistical CSIT is available (corresponding to $\rho = 0$).

In the remainder of this section, for $L \geq 2$ the CSF filters are designed using the GA initialized by the solution for constrained problem (6.40).

In Fig. 6.4, we show the BER of BICM–OFDM with robust CSF filtering as a function of the CSIT reliability parameter $\rho$. Furthermore, we also show the BERs for the cases when CSF filters optimized for perfect ($\rho = 1$) and statistical ($\rho = 0$) CSIT are employed. In the latter case, robust CSF filtering is equivalent to the scheme in Chapter 5. As expected, the robust CSF filters approach the performance of the CSF filters optimized for perfect and statistical CSIT for $\rho \to 1$ and $\rho \to 0$, respectively. However, since the CSF filters optimized for statistical CSIT cannot exploit outdated CSIT, their performance is independent of $\rho$. On the other hand, CSF filters optimized for perfect CSIT do not exploit any reliability.
information and treat the CSIT as if it was perfect even if this is not the case for \( \rho < 1 \). Thus, the performance of CSF filters optimized for perfect CSIT degrades significantly as \( \rho \) decreases. For example, for \( E_b/N_0 = 6 \) dB and \( \rho < 0.8 \), CSF filters optimized for perfect CSIT are even outperformed by CSF filters optimized for statistical CSIT. Fig. 6.4 underlines the importance of taking into account both the CSIT and its reliability for CSF filter design.

In Fig. 6.5, we compare the performance of robust CSF filters of lengths \( L_g = 1 \) and \( L_g = L_{g,\text{min}} = 5 \) for different \( \rho \). For sufficiently large SNR \( L_g = 5 \) performs better than \( L_g = 1 \) as expected. For low SNR this is not necessarily the case since the robust CSF filters were optimized for the dominant error event which causes the worst–case PEP at high SNR. For low SNR more error events would have to be taken into account for CSF optimization, which however, may not lead to tractable results. For the considered SNR
range, the achievable gains with longer CSF filters decrease with increasing \( \rho \). This can be explained by the fact that for unreliable CSIT (small \( \rho \)) the spatial diversity of the system can only be exploited if it is converted into frequency diversity which is only possible for \( L_g > 1 \), cf. Section 6.4.1.3. Thus, for small \( \rho \) (i.e., \( \rho = 0 \) and \( \rho = 0.9 \)) the higher diversity gain (steeper slope of BER curves) for \( L_g = 5 \) can be clearly observed in Fig. 6.5. On the other hand, for highly reliable CSIT (\( \rho = 0.99 \)) the spatial diversity of the system can be exploited to a large extend through beamforming also for \( L_g = 1 \) and the curves for both values of \( L_g \) are parallel for the considered range of SNRs. Note, however, that for very high SNRs (and consequently very low BERs that cannot be simulated), also for \( \rho = 0.99 \) the BER curve for \( L_g = 5 \) has a steeper slope than that for \( L_g = 1 \), cf. Section 6.4.1.3.
6. Robust Transmit Processing for BICM–OFDM Systems with Imperfect CSIT

In Fig. 6.6, we investigate the benefits of robust SFC–CSF filtering as a function of the inverse signal–to–quantization noise ratio (SQNR). We assume $N_T = 3$, $\rho = 0.999$, $L = 1$, and $C_{ee} = \frac{\xi}{1+KR}I_N$. Note that in this case the GA does not have to be used since for $L = 1$ the modified version of Vu’s method finds the optimum CSF filters. The SQNR is defined as $\text{SQNR} \triangleq \frac{\sigma_h^2}{\sigma_e^2}$, where $\sigma_h^2$ and $\sigma_e^2 = \frac{\xi}{(1+KR)}$ denote the variances of the fading gain and the quantization error, respectively. Fig. 6.6 shows that SFC–CSF filtering outperforms pure CSF filtering for unreliable CSIT (i.e., large $1/$SQNR). This shows that extracting the available spatial diversity directly via the OSTBC is more efficient than converting it first into frequency diversity and extracting it then via the convolutional code. On the other hand, for reliable CSIT pure CSF filtering can
also directly exploit the spatial diversity via beamforming and thus, achieves practically the same performance as SFC–CSF filtering. For comparison, we also show results for CSF filtering optimized for statistical CSIT as proposed in Chapter 5 and SFC without CSIT. Even for large 1/SQNR robust CSF filtering achieves a slightly better performance than the CSF filtering scheme from Chapter 5, since, unlike the proposed robust CSF filtering scheme, the scheme in Chapter 5 only takes into account the statistics of the fading channel but not the statistics of the quantization noise.

![Figure 6.6: BER of BICM–OFDM with robust CSF ($L_g = 3$) and robust SFC–CSF ($L_g = 2$) filtering vs. 1/SQNR. $N_T = 3$, $L = 1$, $\rho = 0.999$, and quantized CSIT.](image)

In Fig. 6.7, we compare pure CSF and SFC–CSF filtering for outdated CSIT, $N_T = 3$, and $L = 2$. We observe again that SFC–CSF filtering is superior for unreliable CSIT. On the other hand, for $\rho \geq 0.95$ both scheme yield almost the same performance. For small $\rho$ CSF filtering and SFC–CSF filtering approach the performance of CSF filtering.
with statistical CSIT and SFC without CSIT, respectively.

Figure 6.7: BER of BICM–OFDM with robust CSF ($L_g = 5$) and robust SFC–CSF ($L_g = 2$) filtering vs. $\rho$. $N_T = 3$, $L = 2$, and outdated CSIT.

We conclude from Figs. 6.6 and 6.7 that the higher transceiver complexity of SFC–CSF filtering is justified for applications where the CSIT quality may be occasionally low. On the other hand, if the CSIT is always of high quality, SFC–CSF filtering does not offer any advantage over pure CSF filtering, and the latter is preferable because of its lower complexity and simpler implementation.

### 6.5.3 Linear Prediction (LP)

In this subsection, we evaluate the benefits of post–processing the CSIT via LP. For this purpose, we assume pure CSF filtering and outdated CSIT.

In Fig. 6.8, we show the BER of robust CSF filtering with and without LP as a
function of $E_b/N_0$ for different values of $\rho$. $N_P = 4$, $N_T = 3$, $L = 1$, $L_g = 3$ are valid. As can be observed from Fig. 6.8, LP enables significant performance gains for medium to high correlation factors $\rho$. For example, the performance of robust CSF filtering with LP for $\rho = 0.6$ is better than that of robust CSF filtering without LP for $\rho = 0.9$. Furthermore, robust CSF filtering with LP achieves for $\rho = 0.8$ practically the same performance as robust CSF filtering without LP for $\rho = 0.99$. This clearly demonstrates that LP is a suitable technique for improving the quality of imperfect CSIT. LP cannot improve the quality of the CSIT for completely unreliable CSIT ($\rho = 0$) and (almost) fully reliable CSIT ($\rho = 0.99$). In the former case, successive observations of the CSIT are uncorrelated and LP cannot achieve a gain, and in the latter case, the CSIT is already close to perfect and no further improvement is possible.

In order to better illustrate the dependence of the usefulness of LP on $\rho$, in Fig. 6.9,
we show the BER of robust CSIT with and without LP as a function of $\rho$. Here, we assume $N_T = 2$, $L = 5$, and $L_g = 6$. Fig. 6.9 shows that LP is an effective means to extend the range of $\rho$ values for which robust CSF filtering is beneficial. Thereby, the gain achievable with LP increases with increasing predictor order $N_P$ until saturation is reached. For example, in order to achieve $\text{BER} < 10^{-5}$ for $E_b/N_0 = 10$ dB, $\rho \geq 0.98$ and $\rho \geq 0.9$ are required for robust CSF without LP and robust CSF with LP with $N_P = 2$, while for robust CSF with LP with $N_P = 3$ and $N_P = 4$, $\rho \approx 0.75$ is sufficient to achieve the same performance.
6.6 Conclusions

Robust CSF and SFC-CSF filtering for BICM-OFDM with imperfect CSIT have been considered in this chapter. Furthermore, a linear prediction approach for improving the quality of imperfect CSIT has been proposed. Based on a unified Bayesian reliability model for the CSIT and a unified mathematical representation of CSF and SFC-CSF filtering an upper bound on the worst-case PEP has been developed and a corresponding optimization problem for the CSF filters has been formulated. We have shown that for certain special cases this optimization problem can be solved exactly. For the general case, we have proposed an efficient design procedure for the robust CSF filters consisting of the solution of a related optimization problem with additional constraints and a GA. Simulation results have shown that for imperfect CSIT robust CSF filtering outperforms both CSF filtering optimized for statistical CSIT and CSF filtering optimized for perfect CSIT, while robust SFC-CSF outperforms space-frequency coded BICM-OFDM. Furthermore, linear prediction has been shown to be an effective means for improving the quality of imperfect CSIT.
7 Conclusions and Future Work

In this final chapter, we summarize our results and highlight the contributions of this dissertation. We also suggest topics and open problems for further research.

7.1 Research Contributions

This dissertation as a whole has focused on several relevant aspects of signal design for MIMO–OFDM, namely: (1) channel augmentation schemes to improve system outage performance of MIMO systems; (2) MIMO–OFDM combined with ABL and FEC coding to enhance BER performance; (3) CSF filtering for BICM–OFDM systems with multiple co-located or distributed transmit antennas to achieve a spatial and frequency diversity gain; and (4) robust CSF filtering for MIMO–OFDM, which exploits imperfect CSIT and takes into account the reliability of the CSIT.

In Chapter 2, we have reviewed various open-loop and closed-loop MIMO processing techniques to achieve a spatial multiplexing gain or/and a spatial diversity gain, presented the basic principle of OFDM modulation and adaptive loading for performance enhancement, and introduced SFC for MIMO–OFDM to exploit a frequency diversity gain on top of a spatial diversity gain with an emphasis on CDD.

In Chapter 3, we have presented two simple channel augmentation schemes, namely, CVA and RVA. We compared both schemes with the previously proposed RTM scheme
for the V–BLAST and D–BLAST architectures. Our results have shown that both the CVA and the RVA schemes are more robust to channel correlations than the RTM scheme. Furthermore, while all considered channel augmentation schemes can achieve significantly higher outage rates than the baseline scheme without augmentation, the CVA scheme outperformed the RVA and RTM schemes for all considered cases.

In Chapter 4, we have investigated coded MIMO–OFDM with ABL. Specifically, we have considered the application of WSFC and V–BLAST to MIMO–OFDM with ABL and then optimized both schemes. For ABL, we have considered subcarrier grouping based on two criteria to reduce the feedback load. We have shown via simulations that optimized WSFC and V-BLAST perform fairly close to the benchmark case of SVD, which requires full CSI at the transmitter.

In Chapter 5, we have proposed CSF filtering for BICM–OFDM systems with multiple co–located or distributed transmit antennas. Based on the PEP of the overall system, we have derived optimization criteria for the CSF filters when the set of transmitting antennas is known and not known, respectively. For certain special cases we have provided closed–form solutions for the resulting optimization problems, and we have proposed various CSF filter design methods for the general case. We have studied the impact of the interleaver on the proposed CSF filter designs. Our simulation results have shown that CSF filtering can achieve significant performance gains over various existing CDD designs at the expense of a small increase in complexity.

In Chapter 6, we have considered robust CSF and SFC–CSF filtering for BICM–OFDM with imperfect CSIT, taking advantages of the reliability of the CSIT. We further have proposed a linear prediction approach for improving the quality of imperfect CSIT. The optimization problems for the robust CSF and SFC–CSF filtering filters have been formulated in a unified form based on the worst–case PEP. For certain special cases this optimization problem can be solved exactly. For the general case, we have proposed an
efficient design procedure for the robust CSF filters. Simulation results have confirmed the excellent performance of the proposed robust CSF filtering for imperfect CSIT and linear prediction as an effective means for improving the quality of imperfect CSIT.

7.2 Future Work

There are several immediate extensions of the work presented in this dissertation. We present a (by no means comprehensive) list.

The first interesting open question is the optimal technique to create virtual antennas. In the RTM scheme, the same data are transmitted multiple times, which is essentially a repetition code. The proposed CVA and RVA schemes introduce the conjugated signals and pure real signals in the transmission, which are shown to be more robust to channel correlations than the RTM scheme. Therefore, whether there exists a better code and how to design such code in a systematical fashion is an interesting question.

We have studied WSFC–based and V–BLAST–based MIMO–OFDM with ABL. In Chapter 4, we optimized both schemes and compared their BER performance through simulations. The error rate analysis for coded MIMO–OFDM with ABL would be an interesting topic to explore. Some preliminary work on this topic has already been performed [116]. In addition, the possible use of other FEC coding schemes, such as Turbo and low–density parity–check (LDPC) codes, is also of interest.

A further open area is to apply antenna selection to the transmitter and the receiver for MIMO–OFDM. While a substantial gain in data rate and link reliability can be achieved with a MIMO system, the use of multiple antennas also increases the hardware cost because of multiple radio frequency (RF) chains. As the price of antennas is low and digital processing becomes even cheaper, the use of fewer RF chains than antenna elements appears a natural solution. Antenna selection is a low–cost low–complexity
options to capture many of the advantages of MIMO systems [117, 118]. Therefore, the application of antenna selection to the transmitter in CSF filtering is an interesting topic for future research.
A Related Publications

The following publications are based on the research conducted for this thesis.

**Journal Papers**


Conference Papers


Bibliography


