

Cooperative MIMO Communications

Information Theoretical Limits and Practical Coding Strategies

PhD Thesis Dissertation

Author: Sébastien Simoens

Advisors: Josep Vidal Manzano, Olga Muñoz Medina

SPCOM Group

Technical University of Catalonia (UPC)
Jordi Girona 1-3, Campus Nord, Edifici D5
08034 Barcelona SPAIN

Release : January, 3. 2009

Summary

Early works on the capacity of the wireless relay channel date back more than 30 years. In its genuine version this channel consists of three nodes: a source transmits data to a destination with the help of a relay. In the Information Theory community, various coding strategies have been proposed and their achievable rates have been derived for the three-node relay channel and its extensions to multiple relays, multiple sources and destinations. In the Signal Processing and Wireless Communications community, various questions related to the implementation and performance of relaying have been addressed, such the diversity and multiplexing characteristics of the relay channel, distributed space-time coding, linear processing at the source, relay and destination, ...etc.

Recently, the topic of cooperative relaying has received a lot of attention in the academia and in the industry. Cooperative relaying refers to the fact that advanced coding strategies for the relay channel involve the distribution of coding and decoding functions at several nodes which cooperate in order to maximize the achievable rate between the source(s) and the destination(s). The recent interest in cooperative relaying and cooperative communications in general is motivated by the explosion of wireless internet traffic and can be summarized by the following question: can cooperative relaying substantially increase the spectral efficiency of future Broadband Wireless Access (BWA) networks?

In this thesis, we do not pretend to provide a final answer to this question, but at least we try to contribute on several aspects. The first one is the derivation of capacity bounds for the Multiple Input Multiple Output (MIMO) relay channel with full Channel State Information (CSI), i.e. in the case when all devices are equipped with multiple-antennas and have the capability to exploit channel knowledge at the transmitter side. We propose source and relay precoder optimization procedures which allow the efficient computation of the Cut Set Bound and of achievable rates for the Decode-and-Forward (DF), Compress-and-Forward (CF), and for the more recent distributed CF coding schemes. In a second part of the thesis, we try to exploit these new information-theoretic bounds in order to predict the throughput performance of future BWA networks. We review several implementation-related constraints at the device and link levels (duplexing, broadband transmission, practical modulation and coding, power constraints,

imperfect CSI, ...etc), and also at system-level (deployment topology, macroscopic propagation effects, interference). We analyze their effect analytically and/or by simulations and investigate how capacity bounds can be modified to model them.

Resumen

Los primeros trabajos sobre la capacidad del canal de *relay* (repetidor) inalámbrico datan de hace más de 30 años. En su versión básica el canal de *relay* consta de tres nodos: una fuente transmite datos a un destino con ayuda de un repetidor. En el seno de la comunidad de Teoría de la Información se han propuesto varias estrategias de codificación y se ha calculado su tasa de transmisión alcanzable para el canal básico de tres nodos y sus extensiones a múltiples repetidores, fuentes y destinos. Por otra parte, en el seno de la comunidad de Comunicaciones Inalámbricas y Procesado de Señal se ha intentado dar respuesta a varias cuestiones relativas a la implementación y prestaciones de los esquemas basados en repetidores, tales como las características de diversidad y multiplexado del canal de *relay*, esquemas de codificación espacio-tiempo distribuida, procesado lineal en la fuente, el repetidor y el destino, etc.

Recientemente el tema de repetidores cooperativos ha recibido gran atención por parte del entorno académico y también industrial. La retransmisión cooperativa se basa en estrategias avanzadas de codificación para el canal de *relay*. Estas estrategias implican la distribución de las funciones de codificación y decodificación en los nodos que cooperan con el fin de maximizar la tasa de transmisión entre la fuente(s) y el destino(s). El reciente interés en repetidores cooperativos y comunicaciones cooperativas en general viene motivado por la explosión del tráfico de internet inalámbrico y puede resumirse en la siguiente cuestión: ¿pueden los repetidores cooperativos incrementar sustancialmente la eficiencia espectral de las futuras redes de acceso inalámbrico de banda ancha, *Broadband Wireless Access* (BWA)?

En esta tesis no se pretende dar una respuesta definitiva a esta cuestión, pero sí contribuir en varios aspectos. El primero de ellos es la obtención de cotas para la capacidad del canal de *relay* con múltiples entradas y múltiples salidas, *Multiple Input Multiple Output* (MIMO), y con total información del estado del canal, *Channel State Information* (CSI), es decir, en el caso en el que todos los dispositivos están equipados con múltiples antenas y disponen de la capacidad de explotar el conocimiento del canal en el transmisor. En la tesis se proponen procedimientos de optimización para el precodificador de la fuente y del repetidor que permiten calcular de forma eficiente la denominada *Cut Set Bound* y las tasas de transmisión alcanzables para los esquemas *Decode-and-Forward* (DF), *Compress-and-Forward* (CF) y los esquemas de codificación CF distribuidos más recientes. En la segunda parte de la tesis, estas nuevas cotas basadas

en teoría de la información se explotan con el objetivo de predecir las prestaciones en términos de *throughput* de las futuras redes BWA. En la tesis se revisan varias restricciones relativas a la implementación a nivel de dispositivo y de enlace (duplexado, transmisión de banda ancha, modulación y codificación práctica, restricciones de potencia, CSI imperfecta, etc.), y también a nivel de sistema (topología de despliegue, efectos de propagación macroscópicos, interferencia). Se analiza su efecto analíticamente y/o mediante simulaciones y se investiga como modificar las cotas de la capacidad para modelarlas.

ACKNOWLEDGEMENTS

The completion of this thesis corresponds to a kind of milestone (sorry I've been for too long in E.U. projects) in my life: a new job, a new place, soon a new kid...so let me hurry up writing these acknowledgements as long as I still have some (a little) free time on this Sunday evening...

I will start by thanking the Spanish part of my family and of course my supervisors Josep Vidal and Olga Muñoz come first in the list. You provided me all I needed along these three years: advice, ideas, smart tips, timely reviews and equally important trust, freedom and excellent food during my trips to Barcelona. I take this opportunity to thank all the TSC department at UPC and especially Adrian Agustin, as a working partner in FIREWORKS and ROCKET projects. I was really impressed by the level of expertise in this department, and at the same time the warmth and kindness of everybody I met there (the list would be too long). While in Barcelona, I should also express my thanks to the CTTC. Carles Anton and my old friend Diego Bartolome invited me to Casteldefells, which gave me the opportunity to meet Aitor del Coso and to convince Christian Ibars that Aitor would be in good hands for a summertime internship in Paris. At the time we agreed that Aitor would stop working with me immediately after his return to CTTC, but fortunately things do not always happen as planned... Aitor needs at least a special acknowledgement for a whole part of this thesis is joint work with him. He turned out to be an ironman of optimization problems and he deserves well the nickname he received from Jean-Noël.

My second list of thanks will be for Motorolans, or more exactly ex-Motorolans. I have spent ten years in Motorola's Paris Lab. The first few years were just great: I was surrounded by smart and funny people, given a lot of freedom and doing research on exciting topics. Of course recently the situation for the lab became somewhat more difficult, but I will record only the best. I will first thank my successive managers Marc de Courville, Jean-Noël Patillon, and Guillaume Vivier for letting me conduct this thesis in parallel with my job, which was not always an easy task but overall I don't regret my choice and I thank you for your constant support. I also received previous technical help from Julien Reynier on Quasi Monte-Carlo methods and from Laurent Mazet on nasty bug fixing. Well, actually I would like to thank all my former colleagues (including the

above managers) because after ten years, many of you are just friends and I hope that we will be able to keep in touch even if you never read this manuscript.

I also take the opportunity to thank the IST-FIREWORKS and ICT-ROCKET projects, funded by the European commission. A large part of the research in this thesis was conducted in the framework of these two projects. In addition to the funding without which my research might never have been conducted, it was interesting and very refreshing to be involved in a cross-European effort. As Emilio said once, E.U. projects are like an Erasmus program for professionals.

Last but not least, I would like to thank my family. The fact that my father tried to explain me Einstein's relativity when I was 6 probably has a remote connection with my decision to start a PhD thesis some 23 years after. Likewise, the tons of pies, cakes and waffles that my mother made during all my studies probably generated enough calories to drive me through all these exams. Then I must thank my wife Laure and my son Leo, for always keeping me in the real world of what really matters, and for helping me find the right balance between "Dora the explorer" and "Cover and Thomas"!

October 2008.

Table of Contents

MATHEMATICAL NOTATION	21
CHAPTER 1 : INTRODUCTION.....	23
1.1 Motivation and previous work.....	23
1.2 Summary of contribution and organization of the dissertation.....	26
CHAPTER 2 : SYSTEM MODEL AND RELAYING STRATEGIES.....	31
2.1 Theoretical background on relaying and cooperation.....	31
2.1.1 Preliminary note on capacity bounds.....	32
2.1.2 The cut-set bound on the relay channel capacity.....	32
2.1.3 Relay Duplexing considerations.....	35
2.1.4 Coding Strategies for the relay channel.....	37
2.1.5 Multi-relay, multi-user and multi-way extensions.....	46
2.1.6 From out-of-band relaying to base stations cooperation.....	48
2.2 Modeling state-of-the-art broadband wireless systems.....	50
2.2.1 The fading MIMO relay channel.....	50
2.2.2 Modeling radio device constraints.....	52
2.3 Conclusions.....	58
CHAPTER 3 : CAPACITY BOUNDS FOR THE GAUSSIAN MIMO RELAY CHANNEL - A CONVEX OPTIMIZATION FRAMEWORK.....	60
3.1 Introduction and overview of our contribution.....	60
3.2 The Cut-Set Bound with full MIMO CSI.....	62
3.3 Formulating the CSB as a convex problem.....	62
3.3.1 Full-Duplex Relaying Case.....	62
3.3.2 TDD Relaying Case.....	64
3.4 Computing the CSB.....	65
3.4.1 Dual Method.....	67

3.4.2 Barrier Method.....	67
3.5 Decode-and-Forward with full MIMO CSI	68
3.5.1 Non-Cooperative Decode-and-Forward	68
3.5.2 Partial Decode-and-Forward	68
3.5.3 Full Decode-and-Forward	69
3.6 Simulation Results	73
3.7 Conclusions	77
CHAPTER 4 : DISTRIBUTED COMPRESSION FOR COOPERATIVE MIMO	78
4.1 Introduction and overview of our contribution	78
4.2 Compress-and-Forward strategy for the 3-node MIMO relay channel	80
4.2.1 Signal Model and Achievable Rates	80
4.2.2 Impact of the Compression on the CF achievable rate.....	85
4.2.3 Maximizing the Achievable Rate.....	86
4.2.4 Simulation Results	91
4.2.5 Conclusions.....	94
4.3 Extension to multiple out-of-band relays or cooperative base stations.....	95
4.3.1 Distributed Compression strategy	95
4.3.2 Two cooperative BS case	102
4.3.3 Multiple (more than two) cooperative BS case	103
4.3.4 Multi-user case: sum-rate and achievable rate region	104
4.3.5 Simulation results.....	104
CHAPTER 5 : FROM CAPACITY BOUNDS TO PRACTICAL IMPLEMENTATION	114
5.1 Introduction and overview of our contribution	114
5.2 Practical Implementation of Decode-and-Forward strategies.....	116
5.2.1 Constrained Capacity Bounds	116
5.2.2 An implementation based on cooperative IR	127
5.2.3 An implementation based on superposition coding.....	135
5.2.4 Conclusions and perspectives.....	138

5.3 Practical implementation of Compress-and-Forward strategies	139
5.3.1 Constrained Capacity Bounds	139
5.4 Deployment aspects	141
5.4.1 Single-cell simulation methodology	141
5.4.2 Simulation results with 1 relay per sector	146
5.4.3 Simulation results with 2 relays per sector	148
5.4.4 Simulations in a multi-cell scenario	156
5.5 Conclusions	167
GENERAL CONCLUSIONS AND POSSIBLE FUTURE WORK	168
APPENDIX A - DIFFERENTIATION WITH RESPECT TO COMPLEX STRUCTURED MATRICES.....	172
APPENDIX B - NUMERICAL OPTIMIZATION ALGORITHMS.....	178
APPENDIX C - PROOFS OF PROPOSITIONS.....	184
APPENDIX D - EESM MODEL FOR COOPERATIVE LINKS.....	192

List of Figures

Figure 1: The max-flow min-cut upper-bound for the 3-node relay channel.....	33
Figure 2: The IEEE802.16j TDD relaying frame structure.....	35
Figure 3: A possible sub-categorization of TDD relaying protocols	36
Figure 4: A three-node relay setting with S,R and D aligned	44
Figure 5: Capacity bounds on the full-duplex Gaussian relay channel	45
Figure 6: Capacity Bounds on the Gaussian TDD relay channel.....	46
Figure 7: Cellular downlink relaying with 2 relays and 2 users.....	47
Figure 8: Diamond relay topology in downlink.....	48
Figure 9: Average Spectral Efficiency in IEEE 802.16e PUSC mode with perfect channel knowledge	57
Figure 10: Upper and lower bounds on TDD MIMO relaying channel capacity with Source and Relay precoder optimization in a 4x2x2 antenna configuration.....	75
Figure 11: Optimum vs. Sub-optimum precoder optimization for full DF strategy in a 4x2x2 antenna configuration. (Legend: N=No optimization, V=Vector Optimization, M=Matrix Optimization).....	76
Figure 12: Sub-optimum source precoder during the first slot for full DF strategy: variation of the time-sharing and power balancing vs. SNR on the Source-Relay link....	77
Figure 13: Source coding of Gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the decoder.....	83
Figure 14: Average CF achievable rate and comparison with DF strategies and capacity upper-bounds.....	92
Figure 15: Capacity bounds with and without transmit covariance optimization.....	93
Figure 16: Comparison of CF with other capacity bounds in a cellular uplink scenario with fixed relaying ($\gamma_0 = 0\text{dB}$, $\gamma_2 = 20\text{dB}$, $N_S = 2$, $N_R = N_D = 4$).....	94
Figure 17: Symmetric Diamond out-of-band relay deployment topology with 2 parallel relays	105
Figure 18: Average achievable rate of the DF protocol and Channel capacity in a symmetric diamond Relay channel at low SNR on the MS-RS links and high backhaul rate, single user, $N_S = 2$, $N_R = 2$	106
Figure 19: Required backhaul capacity to achieve 90% of the uplink capacity as a function of the SNR on the MS-RS link in a symmetric diamond topology, single user, $N_S = 2$, $N_R = 2$	107

Figure 20: Same as Figure 18 and Figure 19 in a $N_S = 2$ $N_R = 4$ antenna configuration, two-users.....	109
Figure 21: Same as Figure 18 and Figure 19 in a $N_S = 2$ $N_R = 2$ antenna configuration, two users.	111
Figure 22: CDF of the achievable rate for DF protocols in a $N_S = N_R = 2$ configuration at $(\gamma_0 = 5\text{dB}, \gamma_1 = 20\text{dB}, \gamma_2 = 10\text{dB})$, for a broadband channel (SCME typical urban model) and for a single-carrier Rayleigh i.i.d. model. Top: $N_D = 1$, Bottom: $N_D = 2$	117
Figure 23: Impact of achievable rate degradation parameters (SNR degradation Γ and maximum rate per QAM symbol R_{\max}) on the achievable rate of cooperative DF protocols in the multiple-antenna case ($N_S = N_R = N_D = 2$, $\gamma_0 = 10\text{dB}$, $\gamma_2 = 15\text{dB}$) for 4 increasing levels of degradation ($\Gamma = 0\text{dB}$, $R_{\max} = +\infty$), $(\Gamma = 4\text{dB}, R_{\max} = 10)$, $(\Gamma = 4\text{dB}, R_{\max} = 6)$, $(\Gamma = 4\text{dB}, R_{\max} = 4)$	120
Figure 24: Impact of using a 2-bit precoder codebook in cooperative FDF Protocol III, in a $1 \times 1 \times 1$ antenna configuration with $\gamma_0 = \gamma_2 = 0\text{dB}$ (Legend: N=No optimization, V=Vector Optimization, M=Matrix Optimization, C=Quantized Precoder Codebook).	126
Figure 25: Same as Figure 24 in a $2 \times 2 \times 1$ antenna configuration with 2, 4, 5 bit precoder codebook size.....	127
Figure 26 CTC encoder.....	128
Figure 27 CTC constituent encoder	128
Figure 28 Block diagram of subpacket generation for CTC	129
Figure 29: Sequence of coded bits transmitted by the relay for different cooperative IR strategies (the bits corresponding to the solid line are always transmitted by the relay, the bits corresponding to the dashed line are optionally transmitted depending on the relay code rate).....	130
Figure 30: Average Throughput vs SINR γ_2 on R-D link of cooperative IR strategies, assuming $\gamma_1 = 30\text{dB}$ (S-R link). Top: $\gamma_0 = 10\text{dB}$ (S-D link) Bottom: $\gamma_0 = 5\text{dB}$ (S-D link).....	132
Figure 31: Comparison of cooperative IR throughput and degraded achievable rate assuming $\gamma_1 = 30\text{dB}$ (S-R link) and $\gamma_0 = 10\text{dB}$ (S-D link).	134

Figure 32: Average Throughput performance of cooperative DF with superposition coding and with cooperative IR v1 strategy, assuming $\gamma_1 = 30\text{dB}$ (S-R link). Top $\gamma_0 = 10\text{dB}$ (S-D link) Bottom: $\gamma_0 = 5\text{dB}$ (S-D link).....	137
Figure 33: Comparison of average throughput and average degraded achievable rate (Same assumptions as Figure 32, top).....	138
Figure 34: Single-cell deployment topology in a downlink with 4x4x2 antenna configuration	141
Figure 35: Generation of 10000 points uniformly distributed in $[0;1] \times [0;1]$ (left: Low Discrepancy Sequence, right: Matlab Pseudo-Random generator).....	146
Figure 36: Average capacity of Direct Link (left) and 2-hop non-cooperative DF (right) vs. MS position in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$	147
Figure 37: Average rate of cooperative partial DF Protocol III (left) and cooperation gain vs. MS position in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$	147
Figure 38: Average capacity of various strategies vs. MS position on the BS-RS axis (top) and inter-sector separation axis (bottom) in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$	148
Figure 39: Coverage Improvement by Cooperative DF. Top: Average achievable rate without cooperation vs. MS position; Center: Average achievable rate with Cooperation vs. MS position; Bottom: Cooperation gain vs. MS position, $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, CSIR only, $B=20\text{MHz}$	150
Figure 40: Average achievable rate of Cooperative DF vs. MS position on BS-RS axis (top) and on sector separation axis (bottom) in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, CSIR only, $B=20\text{MHz}$	151
Figure 41: Average capacity on BS-RS axis (left) and on sector separation axis (right) in a $N_{BS} = N_{RS} = N_{MS} = 2$ antenna configuration, no CSIT, $B=20\text{MHz}$, 2 relays/sector.....	152
Figure 42: Impact of shadowing correlation on the achievable rate of cooperative DF (on the BS-RS axis at 90% coverage probability). Top: No Correlation; Bottom: Correlation of 0.5. $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration.....	154
Figure 43: Effect of full CSI on cooperative DF Protocols. Average achievable rate on the sector separation axis at 90% coverage probability in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$	

antenna configuration, with 2 relays per sector. Shadowing Correlation = 0.5. Top: CSIR only; Bottom: full CSI	155
Figure 44: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/3 multi-cell deployment without relays	159
Figure 45: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 multi-cell deployment without relays	160
Figure 46: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 single-cell deployment without relays	161
Figure 47: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/3 multi-cell deployment with 2 relays per sector	162
Figure 48: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 multi-cell deployment with 2 relays per sector.....	163
Figure 49: CDF of the SINR (top) and of the peak spectral efficiency (bottom) versus user location in a 1/3/3 single-cell deployment with 2 relays per sector	164
Figure 50: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/1 multi-cell deployment with 2 relays per sector	165
Figure 51: CDF of the SINR (top) and of the peak spectral efficiency (bottom) versus user location in a 1/3/1 single-cell deployment with 2 relays per sector	166
Figure 52: Rate-distortion coding of gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the encoder and decoder.....	187
Figure 53: BLER vs EESM SNR_{eff} curves on AWGN and Actual Channel snapshots for code rates 1/3 to 5/6 and constellations QPSK, 16QAM and 64QAM.....	193
Figure 54: Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. S-R: 64QAM, R=5/6 R-D: 64QAM, R=1/2, Cooperative IR v1.	196
Figure 55: Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. S-R: 16QAM, R=3/4 R-D: 16QAM, R=3/4, Cooperative IR v1.	197
Figure 56 Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. BS-RS: 16QAM, R=3/4 RS-MS: 64QAM, R=1/2, Cooperative IR v1.	198

Acronyms

AF	Amplify-and-Forward
AMC	Adaptive Modulation and Coding
BC	Broadcast Channel
BRC	Broadcast Relay Channel
BICM	Bit-Interleaved Coded Modulation
BLER	Block Error Rate
BS	Base Station
BWA	Broadband Wireless Access
CPU	Central Processing Unit
CSB	Cut Set Bound
CF	Compress-and-Forward
CKLT	Conditional Karhunen-Loève Transform
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the Transmitter
CTC	Convolutional Turbo Code
DAS	Distributed Antenna System
DCF	Distributed Compress-and-Forward
DCF-JD	DCF with Joint Decoding
DCF-SD	DCF with Sequential Decoding
DF	Decode-and-Forward
DL	Downlink
DPC	Dirty Paper Coding
DMC	Discrete Memoryless Channel
DMT	Diversity Multiplexing Trade-off
EES	Exponential Effective SNR
EESM	Exponential Effective SNR Mapping
FCF	Full Compress-and-Forward
FD	Full Duplex
FDF	Full Decode-and-Forward

FDD	Frequency Division Duplex
GPM	Gradient Projection Method
HARQ	Hybrid ARQ
IR	Incremental Redundancy
KLT	Karhunen-Loève Transform
LDS	Low Discrepancy Sequence
LR	Linear Relaying
MAC	Multiple Access Channel
MARC	Multiple Access Relay Channel
MCP	Multicell Central Processing
MCS	Modulation and Coding Scheme
MIESM	Mutual Information Effective SNR Mapping
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
ML	Maximum Likelihood
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
MS	Mobile Station
NAF	Non-Orthogonal Amplify-and-Forward
NCDF	Non-Cooperative Decode-and-Forward
OAF	Orthogonal Amplify-and-Forward
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PCF	Partial Compress-and-Forward
PDF	Partial Decode-and-Forward
PER	Packet Error Rate
PSD	Positive Semi-Definite
PUSC	Partial Usage of Subcarriers (see IEEE802.16e standard)
QF	Quantize and Forward
QMC	Quasi Monte-Carlo
RMSE	Root Mean Squared Error
RS	Relay Station
SC	Superposition Coding

SCME	Spatial Channel Model Extended
SDM	Spatial Division Multiplexing
SIMO	Single Input Multiple Output
STBC	Space Time Block Code
SWC	Slepian Wolf Coding
TDD	Time Division Duplex
UL	Uplink
VAA	Virtual Antenna Array
VMIMO	Virtual MIMO
WZ	Wyner-Ziv

MATHEMATICAL NOTATION

$\{1, \dots, N\} \setminus \mathcal{G}$	Sequence of integers ranging from 1 to N excluding those in the set \mathcal{G}
$\{\mathbf{a}_i\}_1^N$	Sequence of vectors \mathbf{a}_i for index i ranging from 1 to N.
$\langle \mathbf{A}, \mathbf{B} \rangle$	Inner product for complex vectors and matrices equal to $\text{tr}(\mathbf{A}^H \mathbf{B})$
$(\cdot)^+$	Operator which returns $a^+ = \max(a, 0)$ if $a \in \mathbb{R}$ and $\mathbf{a}^+ = (a_1^+, \dots, a_N^+)$ if $\mathbf{a} \in \mathbb{R}^N$
$(\cdot)^T, (\cdot)^*, (\cdot)^H$	Transpose, Conjugate and Hermitian-transpose operators
\geq	Component-wise ordering ($\mathbf{a} \geq \mathbf{b}$ means that each component of $\mathbf{a} - \mathbf{b}$ is non-negative)
\succeq	Ordering on the Positive Semi-Definite cone ($\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PSD)
$\mathbf{0}_{N \times P}, \mathbf{0}_N$	Null matrices of respective size $N \times P$ and $N \times N$
a, A	Scalars
\mathbf{a}	Column Vector
\mathbf{a}^{-1}	Component-wise inversion of a column vector $\mathbf{a}^{-1} \triangleq (a_1^{-1}, \dots, a_N^{-1})^T$ where $\mathbf{a} \in \mathbb{R}_*^N$
\mathbf{A}	Matrix
$\mathcal{C}(\mathbf{X}, \mathbf{H})$	Given $\mathbf{X} \in \mathcal{S}_+^M$ and \mathbf{H} an $N \times M$ complex matrix, $\mathcal{C}(\mathbf{X}, \mathbf{H}) \triangleq \log_2 \mathbf{I}_M + \mathbf{H}\mathbf{X}\mathbf{H}^H $ represents the capacity of a point-to-point MIMO channel with source covariance \mathbf{X} , channel matrix \mathbf{H} and noise covariance \mathbf{I}_M . Note that the function $f: \mathbf{X} \rightarrow \mathcal{C}(\mathbf{X}, \mathbf{H})$ is concave in \mathbf{X} .
$\text{diag}(\mathbf{a})$	Operator which generates an $N \times N$ diagonal matrix \mathbf{A} from a length-N column vector such that $A_{i,i} = a_i$
$\text{diag}(\{\mathbf{A}_i\}_1^N)$	Block-diagonal matrix created from the sequence of N matrices \mathbf{A}_i .
$I(\mathbf{x}; \mathbf{y} \mathbf{z})$	Conditional Mutual Information between \mathbf{x} and \mathbf{y} given \mathbf{z} .
\mathbf{I}_N	Identity matrix of size $N \times N$
$\text{mat}(\cdot)$	Operator which generates an $N \times N$ matrix \mathbf{A} from a column-vector \mathbf{a} such that $A_{i,j} = a_{(j-1)N+i}$ and $\text{vec}(\text{mat}(\mathbf{a})) = \mathbf{a}$
$\mathbb{R}_*, \mathbb{R}_+, \mathbb{R}_{++}$	Set of non-zero (resp. non-negative and strictly positive) real numbers
\mathcal{S}_+^k	Cone of Positive Semi-Definite matrices of size $k \times k$

$\text{vec}(\cdot)$ Operator which generates a column vector \mathbf{a} from a matrix \mathbf{A} by stacking the columns of \mathbf{A} by increasing column order.

Chapter 1: Introduction

1.1 Motivation and previous work

In my opinion, an important difference between research in the industry and academia is in the nature of questions that we are asked to solve. When I started to work on cooperative relaying and tried to convince my hierarchy that this was a technology to investigate, I was immediately asked how much spectral efficiency increase it could bring to future BWA systems and the question that came immediately after was whether a hybrid deployment of relays and Base Stations would reduce the cost of a cellular network for the same coverage and spectral efficiency?

In order to address these definitely too ambitious questions, my battle plan was the following: I would rely on capacity bounds, because they were very successful in providing insight into the performance of point-to-point MIMO links (see e.g. [T99][TV05]). This good understanding of their performance supported their recent introduction into BWA and WLAN standards such as IEEE802.16e [16e05] and 802.11n [11n08]. The second step in my plan would be to insert link-level capacity bounds into a system-level simulator to take into account macroscopic effects such as interference and finally I could provide answers to my managers. However, I quickly realized that the capacity bounds for the relay channel published in the literature did not suit my needs:

- The vast majority of these bounds (e.g. [LW04][GMZ06][HZ05]) were derived assuming single-antenna devices in a narrowband flat fading channel. However, state-of-the-art BWA systems in 2005 were already based on MIMO-OFDM broadband transmission, with at least two antennas at the BS and considering dual antenna handsets in a very near future. We found a few papers (e.g. [WZH05][LV05][MVA07]) deriving capacity bounds for the MIMO relay channel but for reasons explained in the next bullet they did not completely answer our problem.
- State-of-the-art systems are based on TDD or FDD duplexing, but again the majority of information-theoretic papers on the relay channel were assuming full-duplex operation (e.g. [WZH05][LVH05]). The few papers which

considered half-duplex relaying were either for the single-antenna case (e.g. [HZ05]) or for linear relaying (e.g. [MVA07]). Unfortunately, linear relaying is not spectrally efficient in half-duplex relaying (at least at the link-level) as explained in Chapter 2 and we wanted to investigate the performance of more spectrally efficient strategies such as Decode-and-Forward (DF) and Compress-and-Forward (CF).

- The optimization of capacity bounds for the MIMO relay channel with full CSI was still an open issue although it had been partially addressed in [WZH05] and [MVA07]. However, state-of-the-art BWA systems already had the capability to exploit CSI at the transmitter-side in order to do beamforming [16e05] or even to transmit multiple spatial streams by Singular Value Decomposition of the Channel (SVD-MIMO) [11n08].

We therefore decided to focus our initial efforts on the derivation of capacity bounds for the MIMO relay channel with full CSI, with a special emphasis on DF and CF strategies in the half-duplex case.

During our investigations on CF for the MIMO relay channel, we realized that our work could be extended to the topic of Base Stations cooperation. This topic was only emerging at the time I started this thesis, but by the time I am finishing it seems to receive a lot of attention (see e.g. [GHS06][FKV06][MF07][SSS08]). The goal of BS cooperation, a.k.a. coordinated networks, is to jointly process the signals transmitted from or received at a group of BSs instead of a single BS, thus forming a large VAA [DDA02] and subsequently removing co-channel interference. The distributed compression framework on which we relied for our derivation of CF achievable rates could be applied to the compression of the received observations at a set of cooperative BSs. Just like in our studies on relaying, practical requirements helped us differentiate our contribution:

- State-of-the-art BSs have multiple antennas. Therefore the rates derived in [SSS08] in the single-antenna Gaussian case are not directly applicable to our scenario.
- The cellular backhaul rate is limited and cannot be assumed infinite. Therefore an efficient compression is needed. Although the backhaul assumptions considered in [MF07] were realistic-enough, they did not address the problem of reducing the backhaul rate by advanced compression techniques.

We therefore conducted a derivation of achievable rates with distributed compression for cooperative MIMO uplink under a backhaul rate constraint.

The next step to take was to bridge theoretical capacity bounds with actual throughput simulations. In order to extend our MIMO bounds to the OFDM-MIMO case, we could rely on an approach similar to that of [BGP02] in which it is shown that OFDM-MIMO capacity can be expressed as a sum of MIMO capacity terms. However, we also had to take into account various practical constraints:

- There is a gap between an achievable rate with Gaussian signaling and an actual throughput with real-world modulation and coding. For this purpose, we tried to apply a modification of capacity formula similar to that of [CCB95], including an SNR degradation and a maximum bit rate limitation. We had to validate our modified capacity bounds by comparing them to actual throughput curves. To that aim, we studied a real implementation of cooperative relaying and predicted its throughput by the EESM methodology [E03] after validation by a link simulator compliant to a state-of-the-art BWA standard [16e05]. We knew that we could rely on EESM because it had been successfully applied to the reliability combining of codewords in [BSC04] and to MIMO-OFDM throughput prediction [SRS05].
- Though in a first step we could assume that perfect knowledge of all channels was available at each node, more realistic CSI assumptions have to be considered towards real implementation, including statistical and quantized CSI. A lot of attention has been paid to these topics for point-to-point MIMO (e.g. [JVG01][LHS03][LH05]) but the extension to cooperative MIMO relaying was (and remains) an open field of research.
- Actual transmit power constraints in state-of-the-art systems often differ from the literature where a sum-power constraint is assumed over all transmitting devices or over all antennas of a transmitting device. We therefore ensured that our transmit precoder optimization procedures could include not only sum-power but also per-antenna power constraints, as well as spectral mask constraints for broadband OFDM transmission.

Finally, the last step we had to take was to design a system-level simulator and to draw conclusions from simulations. Here, the main challenge was the complexity vs.

realism trade-off. A preliminary requirement was to find efficient computation procedures for our link-level capacity bounds. This forced us to come back to our link-level capacity bounds and apply advanced optimization techniques. We knew that we could rely on some strong references in the literature such as [B99] for non-linear programming, [BV04] for convex optimization, and also [HG07] for the computation of gradients in closed-form. We relied on even more recent tools kindly provided by the authors of [HP08] to exploit the structure of our matrices in order to further reduce the optimization complexity. At the system-level, one challenge was the large number of realizations of the shadowing over which we had to collect the throughput statistics for each possible user location. With the help of a colleague, we investigated how Quasi Monte-Carlo [S77] simulations could reduce the number of random variable trials without compromising the accuracy of our throughput estimates.

1.2 Summary of contribution and organization of the dissertation

In Chapter 2, we present the various assumptions which together make up the design and evaluation framework for this PhD thesis. After clarifying our notations, we review some theoretical background on relaying and cooperation. We introduce the cut-set bound on capacity, motivate our focus on TDD relaying and present three TDD relaying protocols (I, II and III) as in [NBK04]. We then review coding strategies (DF, CF and LR) for the relay channel and provide their achievable rates on the Gaussian scalar relay channel. One original result in this chapter is the proof in §2.1.4.1.1 that superposition coding at the source cannot increase the achievable rate of DF relaying in TDD. We briefly review extensions of the classical one-way 3-node relay channel to handle multiple relays and multiple users, and we clarify the relationship between in-band relaying, out-of-band relaying and BS cooperation. Thereafter, we introduce our assumptions on radio device constraints (e.g. transmit power constraints) and radio propagation, to which our CSI assumptions are directly connected. Finally, we present the degraded capacity and EESM methods for throughput prediction.

In Chapter 3, we derive capacity bounds for the one-way three-node Gaussian MIMO relay channel with full CSI. We show that the cut-set bound can be formulated in the full-duplex and TDD cases as a convex optimization problem, which yields a tighter capacity upper-bound than previously published in [WZH05]. We present efficient

procedures based on duality and interior point algorithms to compute it. We show that achievable rates for the DF strategies with either partial or full decoding at the relay can be computed by reusing the same convex optimization procedures as for the cut-set bound. We then design lower-complexity sub-optimum precoders with a specific structure for the source and relay. This design results in either a closed-form expression or a reduction of the problem dimensions at the expense of a slightly lower achievable rate. Finally, we perform a comparative analysis of the capacity bounds in a simulation scenario which matches as much as possible a cellular downlink case with fixed relaying. We show that thanks to full CSI large capacity gains can be achieved by cooperative beamforming, and we also observe that the partial DF strategy achieves a rate very close to capacity in this downlink scenario. The work in this chapter is published in [SMVC08], and extended in the submission [SMVC08b] by including the convex formulation in the TDD case, the use of patterned derivatives and a discussion on implementation constraints.

In Chapter 4, we derive achievable rates for partial CF relaying on the three-node Gaussian MIMO relay channel with full CSI. The achievable rates are obtained in §4.2 by applying recent results on distributed compression of Gaussian sources [GDV04][GDV06] to the specific case of partial CF coding strategy of [HZ05]. The compression at the relay consists of a linear transform (the Conditional Karhunen Loeve Transform) followed by parallel Wyner-Ziv coding.

- We analyze the effect of compression on the achievable rate of partial CF and derive a closed-form expression for the optimum Wyner-Ziv coding rates. We show that these rates differ from those of the rate-distortion trade-off derived in [GDV06].
- We show that an optimum decoding order exists for the messages transmitted by the Source and Relay, and this can be used to simplify the optimization of the source and relay covariance matrices. Finally, an iterative procedure is proposed (§4.2.3.3) which jointly optimizes the compression, the transmit covariance matrices and the time resource allocation.
- Simulations are performed in both uplink and downlink cellular scenarios which illustrate the phenomena mentioned above and a comparison with other capacity bounds is performed (§4.2.4).

The first two bullet points are published in [SMV07], while the third bullet is included in the submission [SMVC08c].

In §4.3 we apply a distributed Compress-and-Forward strategy to multiple parallel out-of-band multi-antenna relays or equivalently to a coordinated MIMO cellular network. Our work in §4.3 relies mainly on the distributed coding schemes of [DW04] and [SSS08], in which the signals received at each BS are partially decoded and compressed before being processed by a Central Processing Unit. Our contribution essentially consists in a computation of achievable rates in the multiple antenna case:

- In §4.3.1 we instantiate the results in [SSS08] for a Gaussian multiple-antenna setting with Gaussian codebook, and formulate the achievable rate as an optimization problem with respect to compression noise covariance matrices. In particular, we show that the problem is simplified under a backhaul sum-rate constraint.
- In §4.3.2 we show that the compression noise distribution which maximizes the achievable rate in the 3-node case corresponds to the Transform Coding approach introduced in [GDV04][GDV06] with the WZ coding rate allocation of [SMV07] that is derived in §4.2.3.1.
- Achievable rates are derived in the case of multiple parallel relays in §4.3.3 and an achievable rate region in the multi-user case is derived in §4.3.4.
- Finally, these theoretical results are illustrated by simulations under either per-link or total backhaul rate constraints in §4.3.5.

The above four bullets are the subject of several publications [CS08a][CS08b] and submissions [CS08c][CS08d]

In Chapter 5, we review various issues which arise when a practical implementation of DF and CF is considered in a state-of-the-art broadband wireless access network such as IEEE802.16 [16j07][16m06]. Our contribution is the following:

- In §5.2 we review the implementation of cooperative DF relaying.
 - First, we show that the capacity bounds that we derived in the previous chapters can be extended to model MIMO-OFDM transmission and various transmit power, modulation and coding constraints.
 - In §5.2.1.4 we study the effect of imperfect CSI. We propose some modifications to the achievable rate optimization problem in order to handle

the case of statistical CSI and we verify that quantized precoder codebooks can also be applied to cooperative relaying.

- We conduct a detailed study of two practical implementations of cooperative DF Protocol I based on the convolutionally turbo-coded mode of IEEE802.16e.
 - The first implementation is a cooperative Incremental Redundancy strategy. We derive the parameters of an EESM error predictor for cooperative IR and compute its throughput performance under a target error rate. We verify that the throughput envelope can be well approximated by the degraded achievable rate which is obtained by simple modifications of the information-theoretic formulas of previous chapters.
 - However, the peak rate of cooperative IR may be limited if the set of MCS does not allow very high spectral efficiencies per modulation symbol. In such situations, we show that a strategy which performs superposition coding during the first slot of the TDD protocol can overcome the peak rate saturation problem.
- In §5.3, we review some implementation constraints for the CF strategy. We show that as for DF, the capacity bounds can be extended to handle practical constraints such as MIMO-OFDM transmission. We also briefly describe how practical Wyner-Ziv coding can be realized and what performance can be expected.
- In §5.4 we conduct some system-level simulations to check whether the observations from link-level simulations can match practical deployment scenarios. We review the main simulation parameters and introduce the principle of Quasi Monte-Carlo simulations, before running some simulations in single-cell and multi-cell downlink scenarios to assess how cooperative DF strategies can increase the cellular throughput.
 - In the single-cell scenario we illustrate the effect of shadowing and relay density. We show that cooperative partial DF Protocol III is the most efficient and allows a large increase of achievable rate in the vicinity of the RS and at cell edge. When full CSI is available, even larger gains are

achievable by cooperative DF protocols II and III, as predicted by link-level simulations.

- In the multi-cell scenario, we model additional effects such as inter-sector and inter-cell interference. We show that a careful positioning of RSs in the deployment is required if RSs cannot handle a connection to multiple BSs. We study the potential gains of non-orthogonal resource allocation with a spatial reuse of the relay time-frequency slot and show that it allows a large increase of spectral efficiency. Moreover, spatial reuse is possible with Protocol I but cannot be directly implemented with Protocol III. Therefore, Protocol I can be preferred in many cases at the system-level although it is outperformed by Protocol III at the link level.

These system-level simulation results have been only partially published in [FIR07c] and [VLK07].

Chapter 2: System Model and Relaying Strategies

A huge amount of theoretical and practical work on relaying and cooperative transmission has been published since the early works by Van der Meulen [V71], Cover and El Gamal [CEG79]. In [KGG05], Kramer et al. review of past and recent information-theoretic work on the relay channel. In this chapter, we attempt to summarize the results that make the background for our investigations and we introduce and justify the various system assumptions which are made in this thesis. First, the landmark papers on coding strategies for the relay channel are introduced, which are the focus of subsequent chapters of this report. Some specific aspects related to relaying are then discussed: duplexing, multi-relay deployment, multi-user transmission. Some topics are also addressed which can be considered as borderline but are strongly connected to relaying such as out-of-band relaying and base stations cooperation. A quick application of the previously introduced concepts to the Gaussian SISO relay channel is then presented. Next, various assumptions related to the radio channel and devices are reviewed and finally we discuss how to bridge information-theoretic analysis with link-level and system level actual performance.

2.1 Theoretical background on relaying and cooperation

The 3-node relay channel was introduced by Van der Meulen [V71], but we start our literature review with Cover and El Gamal's landmark paper [CEG79] "Capacity theorems for the relay channel" which introduces most of the concepts that are used in subsequent studies on the so-called one-way relay channel. This channel involves three nodes: a source (S), a relay (R) and a Destination (D). The general coding problem at the source and at the relay aims at maximizing the information rate from S to D. Cover and El Gamal consider that devices are interconnected by a Discrete Memoryless Channel. Moreover, they also assume that the relay is full-duplex, i.e. it can transmit and receive at the same time on the same time-frequency resource. In Theorems 1 and 4 of [CEG79], a block-Markov coding strategy nowadays termed "Decode-and-Forward" (DF) is introduced and is shown to be capacity-achieving when the channel is physically

degraded, i.e. when the signal received at D is a degraded (e.g. noisy) version of the signal received at R. Another coding strategy is introduced in Theorem 6, and is called Compress-and-Forward (CF), Quantize and Forward or Estimate-and-Forward in the literature. For the general (non-degraded) relay channel, [CEG79] only provides an upper-bound on the capacity, which is often termed cut-set bound or max-flow min-cut bound.

2.1.1 Preliminary note on capacity bounds

Before providing expressions for capacity bounds, it is important to clarify the notations used in this document. Capacity bounds are in general established [CT91] by random coding techniques and the use of joint typicality and the Asymptotic Equipartition Property, or by strong typicality. Both require infinite length codewords because they rely on either the weak or the strong (for strong typicality) law of large numbers. For instance the codewords transmitted on a Discrete Memoryless Relay Channel by a source S and a relay R can be denoted as length- n sequences of random variables $\{X_i^S\}_1^n$ and $\{X_i^R\}_1^n$ drawn i.i.d. from the set $\mathcal{X}_S^n \times \mathcal{X}_R^n$ where \mathcal{X}_S and \mathcal{X}_R are discrete sets of symbols, and the capacity theorems are obtained by growing n to infinity. At least on DMC and Gaussian channels, the capacity bounds are ultimately expressed as a function of the probability density function (in the Gaussian channel case) or probability mass function (in the DMC case) of the source and relay symbols. In this thesis, we mainly focus on the optimization of the signal distributions and we rely on information-theoretic results derived elsewhere for the proof of convergence to the capacity bound. Therefore, unless explicitly stated, we denote the source and relay codewords by x_S and x_R . Furthermore, we write the joint distribution $p(x_S, x_R)$ in which we do not distinguish the random variable and its realization. In the multi-antenna case we write $p(\mathbf{x}_S, \mathbf{x}_R)$ where \mathbf{x}_S and \mathbf{x}_R are two random vectors of length N_S and N_R , the number of antennas at S and R.

2.1.2 The cut-set bound on the relay channel capacity

The cut-set upper-bound on the full-duplex three-node relay channel capacity is derived in [CT91] and [CEG79]. With the notations of §2.1.1, the CSB reads:

$$C_{FD} \leq C_{CSB,FD} = \max_{p(x_S, x_R)} \min \left(I(x_S; y_D, y_R | x_R), I(x_S, x_R; y_D) \right) \quad (2.1)$$

where y_D, y_R are the symbols received at D and $I(x_S; y_D, y_R | x_R)$ denotes the conditional mutual information between x_S and (y_R, y_D) given x_R . It can be noticed that (2.1) can be obtained by a straightforward application of the max-flow min-cut upper bound on the capacity of any m -node network given in Theorem 14.10.1 of [CT91] and which states that the rates R^{ij} are achievable if there exists a joint p.d.f. $p(x_1, x_2, \dots, x_m)$ such that

$$\sum_{i \in S, j \in S^c} R^{ij} \leq I(x^{(S)}; y^{(S^c)} | x^{(S^c)}) \quad (2.2)$$

where the sum is performed over all the possible partitions of the nodes into complementary sets S and S^c such that the sources are in S and the destinations are in S^c . The equation (2.2) states that the sum-rate between all the sources and the destinations is upper-bounded by the minimum mutual information between the signals transmitted by the nodes in S and the signals received by the nodes in S^c given the knowledge of the signals transmitted by the nodes in S^c . As illustrated on Figure 1, there are two cuts that separate S from D in the 3-node relay channel, which leads to equation (2.1). The cut that separates S on one side and (R, D) on the other side is called the broadcast cut and the cut that separates (S, R) from D is called the MAC cut. However, one should pay attention that the capacity of the relay channel is not equal to the minimum between the capacity of the S -(R, D) broadcast channel and the (S, R) - D MAC channel, which are both computable.

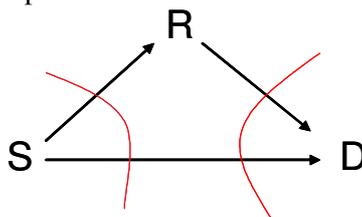


Figure 1: The max-flow min-cut upper-bound for the 3-node relay channel

In [CEG79], it is shown that the cut-set bound is tight on the physically degraded relay channel and on the general relay channel with feedback. We do not consider these two cases because the physically degraded relay channel cannot model a real 3-node

wireless relay channel¹, and because feedback in [CEG79] means that S and R must know perfectly the observations at R and D, which is again unrealistic.

Also note that the cut-set bound (2.2) is also valid for continuous sources, as highlighted in remark 28 of [KGG05]. In [HZ05] the cut-set bound on the general full-duplex Gaussian relay channel is given, but to our knowledge the details of the associated computation are published for the first time by El Gamal in Appendix A of [GMZ06], i.e. more than 25 years after [CEG79]! Let assume circularly symmetric white Gaussian noise of unit variance at R and D, and denote by H_0 , H_1 and H_2 the complex channel gains on the S-D, S-R and R-D links. In the Gaussian relay channel these gains are assumed fixed, and Gaussian codebooks are always assumed [CEG79][CT91][HZ05] because it can be proven that they maximize at least the DF achievable rates and the CSB [CEG79]. The links can therefore be characterized by their signal to noise ratios $\gamma_0 = |H_0|^2 P_S$, $\gamma_1 = |H_1|^2 P_S$ and $\gamma_2 = |H_2|^2 P_R$, where $P_S \triangleq E[|x_S|^2]$ and $P_R \triangleq E[|x_R|^2]$. The cut-set bound can be computed as:

$$C_{FD} \leq \max_{0 \leq \rho \leq 1} \min \left\{ \begin{array}{l} \log(1 + (1 - \rho)(\gamma_0 + \gamma_1)), \\ \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \end{array} \right\} \quad (2.3)$$

where $\rho \triangleq |E[x_S x_R^*]|^2 / (P_S P_R)$ is the correlation between the source and relay codewords. In [HZ05], the authors distinguish the synchronous and asynchronous relay cases. In the synchronous case, the complex channel is known at each node and the source and relay can transmit coherently to the receiver. In the asynchronous case, the author introduces on H_2 an unknown random phase uniformly distributed on $[0; 2\pi[$ and proves that in this case, the maximum in (2.3) is achieved when $\rho = 0$, i.e. the source and relay transmit uncorrelated codewords. In this report we prefer to distinguish the case where channel knowledge is available at each node from the case where channel is only available at the receiver side. Indeed, in OFDM systems it is relatively easy to achieve accurate frequency synchronization between S and R to within a small fraction of the subcarrier spacing and time synchronization to within a small fraction of the cyclic prefix. However, it is very challenging to achieve full channel knowledge at each node when one of the nodes is mobile. Therefore, although S and R may be synchronized, they

¹ In a physically degraded relay channel, the signal received at the destination is a random degradation of the signal received at the relay. This means that all the information is contained in the signal transmitted by the relay.

may not be able to transmit coherently unless full channel knowledge is available at each node. Also notice that $\rho = 1$ is optimum when $\gamma_1 \rightarrow +\infty$ and $\rho = 0$ is optimum when $\gamma_2 \rightarrow +\infty$ when other SNRs are fixed. This limit behaviour can be easily explained once the Decode-and-Forward and Compress-and-Forward strategies are introduced, which is the topic of the next sections.

2.1.3 Relay Duplexing considerations

The practical realization of a full-duplex relay seems challenging. Indeed, a large isolation of the transmit and receive chains needs to be achieved. Otherwise, a strong signal may loop back from the transmitter into the receiver. If the RF front-end does not have enough dynamic range, saturation may occur. But even if it does have enough dynamic range, this interfering component has to be removed by e.g. echo-cancellation techniques. Achieving a large isolation is feasible by separating the transmit and receive antennas by several meters, provided there is enough space on the relay site. If the transmit and receive antennas are close, then directional antennas can be used and the front-to-back ratio shall be large-enough to avoid the saturation problems mentioned before.

2.1.3.1 Half-duplex relaying protocols

For the reasons mentioned above, half-duplex relays are therefore very frequently considered when it comes to practical implementation (e.g.[16j07]). In half-duplex relaying the relay transmission and reception are scheduled on separate time-frequency resource. Both TDD and FDD relaying are technically feasible, although TDD relay implementation seems more straightforward [T05]. For instance in the IEEE802.16j Task Group, a frame structure allowing TDD relaying is considered, as illustrated on Figure 2.

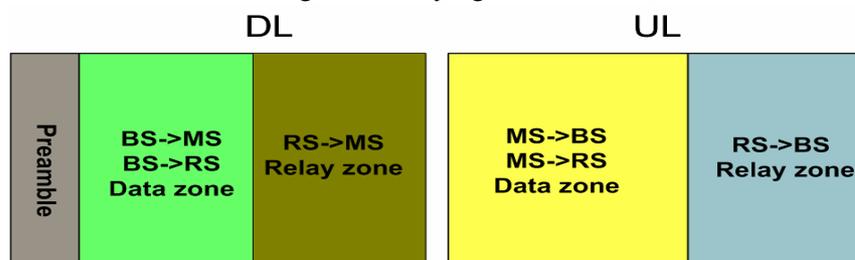


Figure 2: The IEEE802.16j TDD relaying frame structure

In this thesis we consider three TDD relaying protocols as in [NBK04], and define them as illustrated on Figure 3:

- Protocol I (P1): the source is not allowed to transmit in the relay-transmit slot. The destination receiver is active during the two slots and can therefore combine the signals received from the source and relay.
- Protocol II (P2) assumes that the Source and Relay transmit simultaneously during the second slot, but the destination receiver is switched off during the relay-receive slot. This protocol is typically useful if cooperative relaying is introduced in existing standards with backward compatibility requirements (e.g. IEEE802.16j): in this case, existing space-time coding schemes (e.g. Alamouti STBC) can be distributed on the antennas of the Base Station and Relay Station to realize downlink cooperation without having to modify standard-compliant Mobile Stations.
- Protocol III (P3) assumes that the Source and Relay are allowed to transmit simultaneously during the second slot, and that the Destination is allowed to listen to the first slot and combine the signals from both slots.

P1	Slot 1	Slot 2	P2	Slot 1	Slot 2	P3	Slot 1	Slot 2
Src.	Tx		Src.	Tx	Tx	Src.	Tx	Tx
Rel.	Rx	Tx	Rel.	Rx	Tx	Rel.	Rx	Tx
Dest.	Rx	Rx	Dest.		Rx	Dest.	Rx	Rx

Figure 3: A possible sub-categorization of TDD relaying protocols

Note also that another categorization of TDD protocols is introduced in [YE07] and [K04], in which static vs. random and fixed vs. dynamic TDD protocols are considered. In random protocols, the time-sharing between the relay-transmit and receive phase is a random variable that is used to convey some information, whereas it is deterministic for a given channel realization in static protocols. Moreover, a static protocol can be dynamic if the value of the time-sharing variable depends on the channel realization. In this report we consider only static protocols, which may be dynamic when the time-sharing parameter can be optimized as a function of the Channel State Information. We therefore define a variable $t \in [0;1]$ and consider a two-slot TDD² protocol where the relay

² Note that in FDD a separation of the time into two time slots is also performed. For instance in the FDD-DL the relay receives from the BS at the higher frequency during a first slot and transmits to the MS at the higher frequency during a second slot.

receives during the first slot of duration t and transmits during the second slot of duration $1-t$. We consider three random variables $x_S^{(1)}$, $x_S^{(2)}$ and $x_R^{(2)}$ where the superscript (i) with $i \in \{1,2\}$ denotes the slot in which the signal was transmitted.

In [HZ05], the cut-set bound on the static TDD relay channel capacity is expressed as:

$$C_{TDD} \leq \max_{p(x_S^{(1)}, x_S^{(2)}, x_R^{(2)})} \min \left\{ \begin{aligned} &tI(x_S^{(1)}; y_R^{(1)}, y_D^{(1)} | x_R^{(1)} = 0) + (1-t)I(x_S^{(2)}; y_D^{(2)} | x_R^{(2)}), \\ &tI(x_S^{(1)}; y_D^{(1)} | x_R^{(1)} = 0) + (1-t)I(x_S^{(2)}, x_R^{(2)}; y_D^{(2)}) \end{aligned} \right\} \quad (2.4)$$

In the Gaussian case, (2.4) becomes:

$$C_{TDD} \leq \max_{\substack{0 \leq t \leq 1 \\ 0 \leq \rho \leq 1}} \min \left\{ \begin{aligned} &t \log(1 + (\gamma_0 + \gamma_1)) + (1-t) \log(1 + (1-\rho)\gamma_0), \\ &t \log(1 + \gamma_0) + (1-t) \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \end{aligned} \right\} \quad (2.5)$$

Note that in (2.5) it is assumed that the transmit power at a given device remains fixed, whereas in [HZ05] it can be subject to a further optimization under an average power over the two slots. In this report, unless specified, we will assume that devices operate under a maximum transmit power constraint, and in this case transmitting at full power during the two slots maximizes the CSB. This means that in Protocol 3 a larger total power is transmitted during the second slot, compared to Protocol 1. Sometimes in the document we will investigate the effect of constraining the total source plus relay transmit power during the second slot not to exceed the maximum source transmit power.

2.1.4 Coding Strategies for the relay channel

In order to introduce cooperative relaying strategies, it is interesting to study the behaviour of the cut-set bounds (2.3) and (2.5) in the two limit cases when $\gamma_1 \rightarrow +\infty$ and when $\gamma_2 \rightarrow +\infty$.

$$\begin{aligned} \lim_{\gamma_1 \rightarrow +\infty} C_{FD} &= \lim_{\gamma_1 \rightarrow +\infty} C_{TDD} = \max_{0 \leq \rho \leq 1} \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \\ &= \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\gamma_0\gamma_2}) \end{aligned} \quad (2.6)$$

$$\begin{aligned} \lim_{\gamma_2 \rightarrow +\infty} C_{FD} &= \lim_{\gamma_2 \rightarrow +\infty} C_{TDD} = \max_{0 \leq \rho \leq 1} \log(1 + (1-\rho)(\gamma_0 + \gamma_1)) \\ &= \log(1 + (\gamma_0 + \gamma_1)) = C_{SIMO} \end{aligned} \quad (2.7)$$

The physical interpretation of (2.6) is the following: when the Source to Relay link has infinite capacity, then both the TDD and FD cut-set bound converge to the 2×1 MISO capacity with per-antenna power constraint. In this case, the relay can successfully decode any message transmitted by the source in an infinitely short fraction t of the total time as long as it contains a finite number of information bits per symbol. In the second time slot of duration $1-t \rightarrow 1$, S and R can then transmit this message coherently to the destination using Maximum Ratio Transmission (MRT), which is well known to achieve the capacity of the MISO channel and in this case the correlation ρ of the two codewords tends to 1. Such a strategy is called Decode-and-Forward and is therefore capacity achieving in the limit case when $\gamma_1 \rightarrow +\infty$. Likewise when $\gamma_2 \rightarrow +\infty$ (2.7) the cut-set bound converges to 1×2 SIMO channel capacity. In this case, a capacity-achieving strategy consists in performing rate-distortion source encoding of the signal observed at the Relay. This encoding can be modeled by the addition of an uncorrelated white Gaussian noise of variance equal to the quadratic distortion and the latter can be made arbitrarily small even if the duration $1-t$ of the second slot becomes infinitely small. Thus, the source message can be decoded from the Source and Relay observations, the latter being reconstructed with negligible distortion at the Destination. In the literature, this strategy is called Quantize-and-Forward (QF), Estimate-and-Forward (EF) or Compress-and-Forward (CF). In this report, we will use the CF acronym, and we may further categorize CF strategies according to the type of source coding that is used. Finally, note that hybrid strategies have been proposed (see e.g. Theorem 7 in [CEG79] and [SSS08]) which combine the DF and CF strategies, but we will not focus on them in this thesis.

2.1.4.1 DF strategies

A DF strategy for the full-duplex relay channel is considered in [CEG79] and shown to achieve the following rate:

$$R_{DF,FD} = \max_{p(x_S, x_R)} \min \left(I(x_S; y_R | x_R), I(x_S, x_R; y_D) \right) \quad (2.8)$$

As discussed in [KGG05], this rate is achieved by a Block-Markov superposition irregular encoding (i.e. codebooks of different size) strategy and successive decoding in

[CEG79], but an easier derivation employs Block-Markov superposition regular encoding and backward or sliding window decoding. On the Gaussian relay channel (2.8) becomes:

$$R_{DF,FD} = \max_{0 \leq \rho \leq 1} \min \left\{ \begin{array}{l} \log(1 + (1 - \rho)\gamma_1), \\ \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \end{array} \right\} \quad (2.9)$$

In the following, we call Full DF (FDF) a strategy in which the relay has to decode all the message, and Partial DF (PDF) a strategy in which the relay only has to decode a part of the message. The achievable rate of PDF in the FD case is given in equation (13) of [KGG05]. The (regular) coding strategy that achieves this rate employs superposition coding at the source of two messages: one is decoded by the relay, and the other one is decoded only by the destination. The PDF strategy achieves the rate

$$R_{PDF,FD} = \max_{p(x_S, x_R, u)} \min \left(I(u; y_R | x_R) + I(x_S; y_D | u, x_R), I(x_S, x_R; y_D) \right) \quad (2.10)$$

The computation of the PDF achievable rate in the Gaussian case is performed in [GMZ06]:

$$R_{PDF,FD} = \max_{0 \leq \rho \leq 1} \min \left\{ \begin{array}{l} \max(\log(1 + \gamma_0), \log(1 + (1 - \rho)\gamma_1)), \\ \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \end{array} \right\} \quad (2.11)$$

Comparing (2.11) with (2.9), it can be observed that PDF and FDF achieve the same rate on the full-duplex Gaussian relay channel, except in the case when $\gamma_0 > \gamma_1$ where PDF degenerates into direct transmission from S to D, skipping the relay, but such a case does not have a practical interest, because it can be addressed by an adaptive selection of the best transmission strategy as a function of the CSI on the three links.

Contrary to the FD case, the achievable rate of the PDF and FDF strategies on the static TDD relay channel are different. Let first consider TDD Protocol 3 as defined in §2.1.3.1. A PDF strategy for the Gaussian TDD relay channel is proposed in [HZ05]: the Source transmits a first message ω_0 at a rate R_0 using a signal $x_S^{(1)}(\omega_0)$ during the first slot. The relay decodes $\hat{\omega}_0$ and transmits $x_R^{(2)}(\hat{\omega}_0)$ during the second slot while S transmits a new message ω_1 by superposition coding: $x_S^{(2)}(\omega_0, \omega_1) = x_{S,0}^{(2)}(\omega_0) + x_{S,1}^{(2)}(\omega_1)$. Because we assume a synchronized scenario, the signals $x_R^{(2)}(\omega_0)$ and $x_{S,0}^{(2)}(\omega_0)$ are correlated in order to cooperate by performing a Maximum Ratio Transmission whereas ω_1 is mapped onto an independent signal $\mathbf{x}_{S,1}^{(2)}(\omega_1)$ transmitted at rate R_1 using superposition coding. The destination successively decodes $\hat{\omega}_0$ and $\hat{\omega}_1$. The achievable rate of this strategy is

derived in [HZ05] and the derivation can also be found as a special case of the proof of Proposition 2.1 in Appendix C.1:

$$R_{PDF,P3} = \max_{\substack{0 \leq t \leq 1 \\ 0 \leq \rho \leq 1}} \min \left\{ \begin{array}{l} t \log(1 + \gamma_1) + (1-t) \log(1 + (1-\rho)\gamma_0), \\ t \log(1 + \gamma_0) + (1-t) \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho\gamma_0\gamma_2}) \end{array} \right\} \quad (2.12)$$

The FDF strategy achieves the following rate for Protocol 3:

$$R_{FDF,P3} = \max_{0 \leq t \leq 1} \min \left\{ \begin{array}{l} t \log(1 + \gamma_1), \\ t \log(1 + \gamma_0) + (1-t) \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{\gamma_0\gamma_2}) \end{array} \right\} \quad (2.13)$$

It is important to introduce some notations for some simple strategies which will serve as references in all the report. First, the capacity of the S-D, S-R and R-D links:

$$C_{SD} = \log(1 + \gamma_0) \quad C_{SR} = \log(1 + \gamma_1) \quad C_{RD} = \log(1 + \gamma_2) \quad (2.14)$$

The achievable rate of the FDF strategy for Protocol 1 is:

$$R_{FDF,P1} = \max_{0 \leq t \leq 1} \min \{ t C_{SR}, t C_{SD} + (1-t) C_{RD} \} \quad (2.15)$$

It can be easily checked that if $C_{SD} < C_{RD}$ and $C_{SD} < C_{SR}$ then the optimum rate is:

$$R_{FDF,I} = \frac{C_{SR} C_{RD}}{C_{SR} + C_{RD} - C_{SD}} \quad (2.16)$$

Otherwise $R_{FDF,P1}$ cannot exceed C_{SD} which means that it is better not to relay. Finally, we will also consider in this report the Non-Cooperative DF relaying (NCDF) strategy, due to its practical importance. It can be defined as a variant of Protocol I in which the destination only receives during the second slot. In this case the achievable rate is:

$$R_{NC,DF} = \max_{t \in [0,1]} \min \{ t C_{SR}, (1-t) C_{RD} \} \quad (2.17)$$

The optimum time sharing is:

$$\hat{t} = \frac{C_{RD}}{C_{SR} + C_{RD}} \quad (2.18)$$

And the achievable rate is:

$$R_{NC,DF} = \frac{C_{SR} C_{RD}}{C_{SR} + C_{RD}} \quad (2.19)$$

Note that from (2.19) it is clear that the achievable rate of NCDF is upper-bounded by the minimum between the capacity of the first hop link and that of the second hop link:

$$R_{NC,DF} \leq \min(C_{SR}, C_{RD}) \quad (2.20)$$

As suggested in the introduction of this section, all cooperative DF strategies become capacity achieving for $\gamma_1 \rightarrow +\infty$. However, when $\gamma_0 = \gamma_1$ even PDF cannot outperform direct source to destination transmission, therefore other strategies have to be considered in order to benefit from relaying in such situations, and this is the purpose of the CF strategy presented in the next section.

2.1.4.1.1 Can superposition coding further increase the rate?

We consider an extension of the partial DF strategy in which superposition coding is performed in both slots. Three messages $\omega_{d,1}$, $\omega_{d,2}$ and ω_r are sent to the destination at respective rates $R_{d,1}$, $R_{d,2}$ and R_r . The message ω_r is called the relayed message, while the other two are called direct messages because they are not forwarded by the relay. During the first slot, S transmits $\omega_{d,1}$ and ω_r via superposition coding. The relay first decodes $\omega_{d,1}$ from $\mathbf{y}_R^{(1)}$ and removes the contribution of this message from its observation before decoding ω_r . During the second slot, the relayed message ω_r is used by S and R to cooperate while S sends the second direct message $\omega_{d,2}$ via superposition coding. The destination starts by decoding ω_r from $\mathbf{y}_D^{(1)}$ and $\mathbf{y}_D^{(2)}$, and removes its contribution from the observation before decoding $\omega_{d,2}$.

Proposition 2.1: *Superposition coding during the first slot cannot increase the achievable rate of the partial DF strategy on the single-antenna Gaussian TDD relay channel.*

Proof: See Appendix C.1.

This result is not straightforward, and it illustrates well the fact that the relay channel shall not be treated as a BC followed by a MAC. Indeed, SC is the optimum coding strategy for the scalar Gaussian broadcast channel [CT91], therefore it could have been expected to increase the rate when applied to the first hop of the relay channel.

2.1.4.2 CF strategies

The CF strategy discussed in the introduction of §2.1.4 when $\gamma_2 \rightarrow +\infty$ can actually be improved. Indeed, a CF strategy with a larger achievable rate is introduced in Theorem 6 of [CEG79]. It relies on Wyner-Ziv coding of the relay observation. In [WZ76], Wyner and Ziv compute the rate-distortion function for source coding of discrete sources with side information at the decoder. When the decoder has the knowledge of a signal correlated with the source, the latter can be encoded at a lower rate for a given distortion. In [W78], Wyner generalizes this work to continuous sources, and in particular to Gaussian sources with quadratic distortion. In [CEG79], the authors exploit the fact that the observations at R and D are correlated, since they are both noisy versions of the same signal transmitted by S. Therefore, Wyner-Ziv coding can be applied to perform rate-distortion coding of the relay observation. We will discuss this source coding strategy in more details in Chapter 4, but for the moment we will assume that the relay observation y_R can be compressed to a certain message ω_R and the destination has a reconstruction function $\hat{y}_R = f(\omega_R, y_D)$. The general expression for the CF achievable rate in the FD case is:

$$\begin{aligned}
 R_{CF,FD} &= \max_{p(x_S, x_R, y_R, y_D, \hat{y}_R)} I(x_S; \hat{y}_R, y_D | x_R) \\
 \text{s.t.} \quad & I(\hat{y}_R; y_R | y_D, x_R) \leq I(x_R; y_D) \\
 \text{and} \quad & p(x_S, x_R, y_R, y_D, \hat{y}_R) = p(x_S) p(x_R) p(\hat{y}_R | x_R, y_R) p(y_R, y_D | x_S, x_R)
 \end{aligned} \tag{2.21}$$

The proof of (2.21) is quite involved technically, but the outcome lends itself to interpretation. The term in (2.21) that shall be maximized corresponds to the left hand side of the CSB in equation (2.1), i.e. the broadcast cut, except that the relay observation is replaced by the reconstructed relay observation. The optimization over the joint probability distribution is constrained by the fact that the rate at which the relay observation can be encoded cannot exceed the rate at which the relay can reliably send the message ω_R to the destination.

In [KGG05] and [HZ05], the achievable rates of the CF strategy on the FD and TDD Gaussian relay channels are derived:

$$R_{CF,FD} = \log \left(1 + \gamma_0 + \frac{\gamma_1}{1 + \eta_{FD}} \right) \tag{2.22}$$

where η_{FD} is called the compression noise variance and is equal to:

$$\eta_{FD} = \frac{1 + \gamma_0 + \gamma_1}{\gamma_2} \quad (2.23)$$

In TDD, the achievable rate is:

$$R_{CF,TDD} = t \log \left(1 + \gamma_0 + \frac{\gamma_1}{1 + \eta_{TDD}} \right) + (1-t) \log(1 + \gamma_0) \quad (2.24)$$

with

$$\eta_{TDD} = \frac{1 + \gamma_0 + \gamma_1}{(1 + \gamma_0) \left(\left(1 + \frac{\gamma_2}{1 + \gamma_0} \right)^{\frac{1-t}{t}} - 1 \right)} \quad (2.25)$$

The rate (2.24) is achieved in [HZ05] using a Partial CF strategy (PCF): during the first slot, S transmits a message ω_0 at rate R_0 by means of the codeword $x_S(\omega_0)$, the relay compresses its observation $y_R^{(1)}$ to the index ω_1 , and transmits $x_R^{(2)}(\omega_1)$ to D during the second slot, while S transmits a new message ω_2 at rate R_2 via a new codeword $x_S^{(2)}(\omega_2)$. The achievable rate (2.24) is computed as the sum $R_0 + R_2$. More details on the source and channel decoding strategy will be provided in Chapter 4.

2.1.4.3 LR strategies

The Amplify-and-Forward (AF) strategy has been known and used for a long time, for instance by conventional non-regenerative satellite systems: the satellite receives the signal from an earth station, amplifies it and retransmits it towards the earth, without attempting to decode it. In [LW04], Laneman considers TDD protocols and a strategy that is called Orthogonal Amplify-and-Forward (OAF), which corresponds to Protocol 1 in §2.1.3.1. In this case, the relay quantizes its observation during the first slot with enough accuracy so that the distortion is negligible, the quantized samples are stored and retransmitted during the second slot. In this case, the achievable rate is:

$$R_{OAF,TDD} = \frac{1}{2} \log \left(1 + \gamma_0 + \frac{\gamma_1 \gamma_2}{1 + \gamma_1 + \gamma_2} \right) \quad (2.26)$$

Notice the presence of the $1/2$ factor in (2.26) due to the time-sharing parameter that cannot be optimized in AF. When $\gamma_1 \rightarrow +\infty$, the achievable rate with OAF cannot exceed one-half of the MISO capacity, contrary to DF which is capacity achieving. Likewise, when $\gamma_2 \rightarrow +\infty$, R_{OAF} cannot exceed one-half of the SIMO channel capacity, and is therefore outperformed by CF which is capacity achieving. In [AV06], it is shown

that the negative effect of the constrained time-sharing parameter can be alleviated at the system level, when several AF relays transmitting to different destinations operate in parallel on the same time-frequency resource and the relay-destination pairs are sufficiently separated. Non-orthogonal AF (NAF) strategy, which corresponds to Protocol 3 in §2.1.3.1 is addressed in [GMZ06] as a special case of Linear Relaying (LR). In LR, the relay retransmits a (causal) linear combination of previous observation blocks. Thus AF is a special case of LR in which only the previous block is retransmitted. In [GMZ06] the authors show that NAF, although not the best among all LR strategies, may outperform DF and CF at low SNR γ_1 on the Source-Relay link. Yet, in the 3-node TDD case, LR suffers from the same $\frac{1}{2}$ penalty as AF which makes it achieve a rate significantly lower than both DF and CF and this is the reason why we do not consider it further in this thesis.

2.1.4.4 A case study: the Gaussian 3-node relay channel

In this section, we illustrate the various capacity bounds expressed so far in this report by considering a simple 3-node Gaussian relay channel, in which S, R and D are aligned, with $\gamma_0 = 0dB$. The Source and Relay transmit powers are set to 20dBm and the noise power at D is set to -90dBm. We assume a log-distance path-loss model with 1 meter breakpoint distance, and a path loss exponent of 2.6 beyond the breakpoint distance. On the plots, distances are normalized by the Source-Destination distance $d(S-D)$, which is equal to 490m.

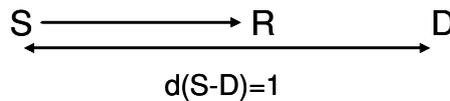


Figure 4: A three-node relay setting with S,R and D aligned

In this case, the capacity of the Source-Destination channel, measured in bit per channel use, is $C_{SD} = \log(1+1) = 1$. On Figure 5, the capacity bounds on the FD Gaussian relay channel are plotted. The following observations can be made:

- The MISO and SIMO bounds are valid only when R is in the immediate neighborhood of S or D. The CSB is the only valid upper bound for all relay locations.
- The DF is the best strategy when R is close to S and CF is the best strategy when R is close to D. Moreover, a mixed strategy which selects between DF and CF depending

on the relay location would operate in the worst case at less than 0.3 bit per channel use from the CSB, i.e. about 85% of the relay channel capacity can be achieved by this strategy.

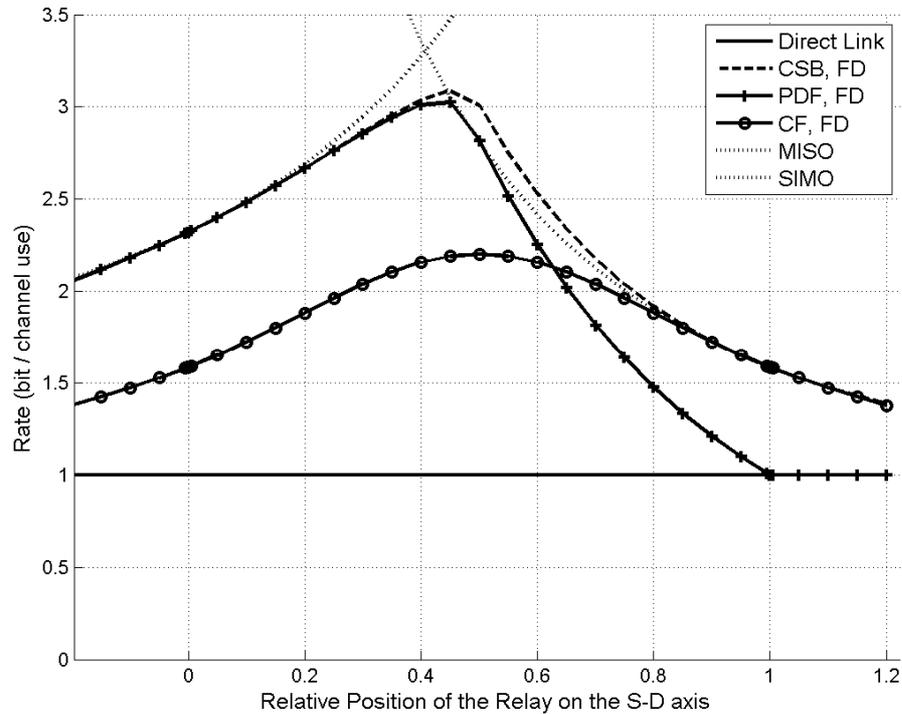


Figure 5: Capacity bounds on the full-duplex Gaussian relay channel

On Figure 6, the achievable rates of the NCDF, FDF, PDF and PCF strategies are compared to the TDD and FD cut-set bound. The following observations can be made:

- A large loss is incurred due to half-duplex relaying. This loss is the largest when R is half-way between S and D where it reaches almost 1 bit per channel use, which represents one-third of the full-duplex capacity. Moreover, even though the TDD achievable rates are theoretically equal to the FD achievable rates when R is infinitely close to either S or D, there is always a non-negligible gap of a few tenths of bit per channel use between full-duplex and TDD relaying when the distances are greater than 1m. This gap is due to the fact that the capacity of a point-to-point link grows logarithmically with the SNR, and even if at 1m distance the SNR is large the capacity of this link cannot be assumed infinite.

- The NCDF strategy brings an improvement over direct Source-Destination transmission but cannot outperform C_{SD} when R is infinitely close to S. The optimum position of R for NCDF is half-way between S and D. This is a coincidence, as the optimum relay location is in general different for arbitrary Source and Relay transmit power.
- The PDF and FDF are capacity achieving when R is infinitely close to S, whereas the NCDF strategy is not.
- The PDF strategy is optimum when R is close to S, and the PCF is optimum when R is close to D. A mixed strategy which selects the best of PDF and PCF can thus operate at less than 0.3 bit per channel use from the TDD cut-set bound.

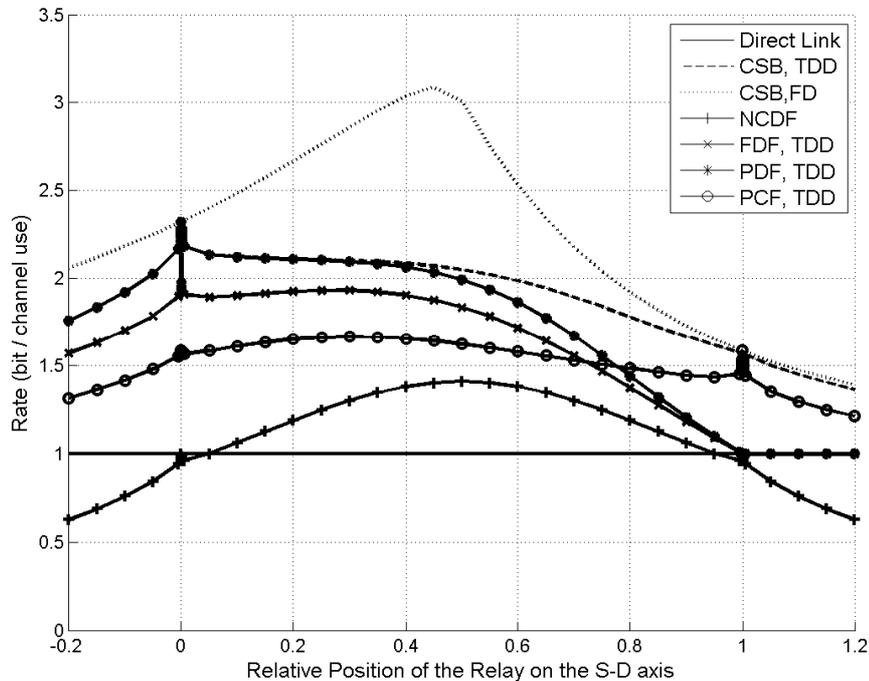


Figure 6: Capacity Bounds on the Gaussian TDD relay channel

2.1.5 Multi-relay, multi-user and multi-way extensions

The 3-node relay channel described so far can be viewed as a building block for more complex deployment topologies. Let for instance consider the deployment on Figure 7: the devices can be grouped into two sets of 3 nodes: (BS, RS_1 , MS_1) and (BS,

RS_2, MS_2). As long as transmissions of nodes belonging to these sets are scheduled on orthogonal time-frequency resource, the relays do not interfere (the dashed arrows represent potential interference) and the capacity bounds for the 3-node relay channel can be used as inputs to the resource allocation algorithm. Such a resource allocation strategy is explained in [AV07]. Likewise, considering the “diamond” [XS07] topology of Figure 8, the study of cooperative beamforming performed in the context of 3-node relaying can be almost straightforwardly applied to model the cooperative beamforming of RS_1 and RS_2 towards the MS.

However, larger rates can be achieved by considering coding strategies specifically designed for the network topology. Some example topologies which are addressed in the literature include the parallel relay channel, the relay channel with more than two hops, the Multiple Access Relay Channel and the Broadcast relay channel. A good survey of recent information-theoretic results for these topologies is given in [KGG05] and [C08], and we will not further discuss them for the sake of brevity because most of our results focus on the three node relay channel. However, in §4.3 we will consider topologies with more than 3-nodes in the special context of out-of-band relaying and BS cooperation which are the purpose of the next section.

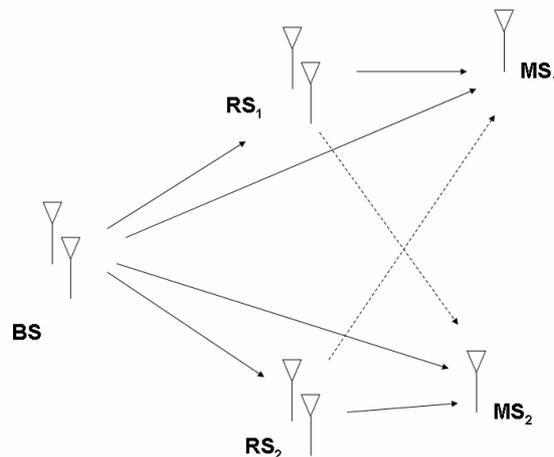


Figure 7: Cellular downlink relaying with 2 relays and 2 users

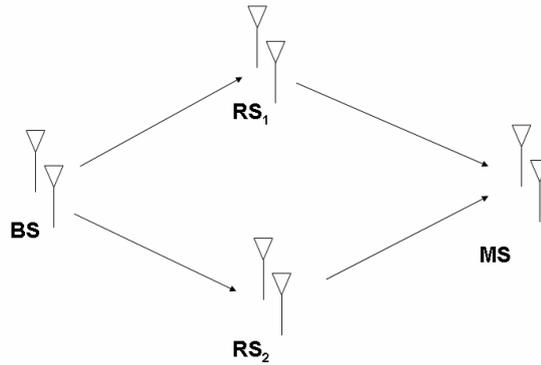


Figure 8: Diamond relay topology in downlink

In addition to relay topologies involving multiple Sources (MARC), multiple destinations (BRC) or multiple relays (e.g. parallel relay channel), an active research topic considers multiple Source-Destination pairs. A simple example is the 3-node two-way relay channel [RW07], which aims at increasing the system spectral efficiency when downlink and uplink traffic are not too asymmetric. Again, this scenario and associated coding strategies are beyond the scope of this thesis and we will not discuss them further.

2.1.6 From out-of-band relaying to base stations cooperation

Most of the literature considers so-called “in-band” relaying, in which the source and relay transmit at the same carrier frequency. Their transmissions can then either be separated in time and frequency via orthogonal scheduling, or on the contrary S and R may be scheduled on the same time-frequency resource, and cooperatively beamform to the destination. However, in a cellular deployment a fixed RS can be shared by a large number of MS, and in this case the link between the BS and the RS may well become the bottleneck of the MARC (in uplink) or the BRC (in downlink). Out-of-band relaying is a potential solution for this problem. It consists in assigning distinct carrier frequencies to the communications involving the MS and to those involving only infrastructure equipment. The carrier frequency assigned to the communications involving the MS can be termed “access carrier frequency” and the other one can be termed “backhaul carrier frequency”. The MS will be designed to transmit and receive only on the access

frequency, whereas the BS and RS shall be able to transmit and receive on both frequencies, and two transceivers are thus needed. This makes out-of-band relaying a more expensive solution at first glance.

Under the assumptions explained above, the specific features of cooperative out-of-band relaying vs. cooperative in-band relaying are the following:

- TDD Protocols 1 and 3 are not feasible in downlink and Protocols 2 and 3 are not feasible in uplink, because the MS cannot operate on the backhaul frequency. Moreover, in many cases it makes sense to assume that the BS and RS are able to transmit and receive on the backhaul and access frequencies simultaneously. If so, there is no need for a time-slotted cooperation protocol.
- Since a cellular BS and RS are infrastructure equipment, they can be designed to transmit and receive on a large bandwidth. Moreover, if the RS is mounted on a lamp post or rooftop, it is likely (but not always) in LOS with the BS. Hence, the capacity of the backhaul (BS-RS) link can be much larger than that of the access links (BS-MS and RS-MS) in out-of-band relaying. This has an impact on the relative performance of cooperative coding strategies.

Achievable rate calculations for out-of-band relaying will be the same whether the backhaul is wireless or wired. If we extend out-of-band relaying to multiple parallel relays, this naturally leads us to the topic of BS cooperation, which is addressed in §4.3 of this thesis. As pointed out in [ACH07], a key challenge for future BWA systems will be to overcome inter-cell interference. Indeed, state-of-the-art cellular deployments are typically based on frequency reuse factors between 2 or 3, and the target for future networks is reuse-1 in order to maximize the spectral efficiency. Cooperative BS transmission and reception, also called cellular network coordination, is viewed as the ultimate (but also the highest complexity) solution to maximize the system spectral efficiency. A reuse-1 coordinated cellular network is essentially the same as a set of multiple parallel out-of-band relays linked to a BS. In the literature on coordinated cellular network, the BS is often called a “Central Processing Unit” (CPU) or “main BS” and the RS is either termed BS or “receiving agent” as in [SSS08]. In the uplink, the signal transmitted from a set of MSs is received by a set of cooperative BSs interconnected by a fixed rate backhaul. The decoding is performed by the CPU which can be located at one of the BS sites. In the downlink, the cooperative BSs transmit

synchronously to a set of receiving MSs. If the backhaul capacity is large, then the set of cooperating BSs can be viewed as one Virtual Antenna Array and if the cardinality of this set is large, then all inter-cell interference can virtually be removed. In reality, backhaul rate and latency limitations and MIMO processing complexity have to be taken into account, and this is the purpose of §4.3 of this thesis.

2.2 Modeling state-of-the-art broadband wireless systems

In our work we try to have as much realistic assumptions as possible corresponding to a state-of-the-art broadband wireless system. In this section we discuss some key assumptions related to this choice, and highlight the differences with the literature.

2.2.1 The fading MIMO relay channel

In §2.1.2 and §2.1.4, the capacity bounds are computed assuming a constant flat-fading channel. Of course, because each wireless link is subject to fading, the relay channel can also be studied in terms of ergodic and outage capacity. In [SEA03] the concept of cooperative diversity is introduced. In [LW04] quasi-static flat fading is assumed and analytical expressions for the outage probability of various half-duplex cooperation protocols are provided. It is shown that most cooperation protocols for the 3-node relay channel provide a diversity order of 2. Papers typically analyze the ergodic capacity ([HZ05][WZ05][YE07]) or the outage capacity ([LW04]), and more recently the Diversity Multiplexing Trade-off (DMT) [YE07].

2.2.1.1 Time variations and tracking of the channel

In our thesis work, we focus mainly on the quasi-static fading channel, i.e. the channel remains constant over a frame, which we define in TDD as two successive slots. However, the channel may change from one frame to the next. Such a model is well suited to low-mobility (e.g. pedestrian) users. For instance, let consider the IEEE802.16e system. Assuming a carrier frequency f_C equal to 3GHz, and denoting by v the maximum relative velocity (expressed in m/s), the maximum Doppler frequency (in Hz) equals $f_D = 10v$ and the channel coherence time $T_C \approx 1/(2f_D)$ is frequently assumed.

At pedestrian velocities of 1m/s to 5 m/s, the channel coherence time thus ranges from 10 ms to 50 ms. The typical frame duration in a TDD system such as IEEE802.16e is 5 ms. Therefore, it is reasonable at these speeds to assume a quasi-static channel. Moreover, at such velocities the channel can be tracked and CSIT can be exploited. In Chapter 3 and Chapter 4 of this thesis, capacity bounds are derived assuming full CSI. In the context of relaying, full CSI means that every node has perfect knowledge of the channel on all links. However, one must pay attention to the delay between the estimation of the channel and the application of the corresponding precoding. Therefore, different degrees of channel knowledge will have to be considered at the transmitter, from full CSI to partial and statistical CSI (see §5.2.1.4). If we now consider an FDD system such as 3GPP LTE, the frame duration is 500 μ s, therefore such system shall be able to track the channel at approximately 10 times higher velocities. When perfect CSIT is assumed, it does not make sense to study the outage performance, because the capacity (or the achievable rates) is known at each frame and the system can adapt the spectral efficiency on a frame-by-frame basis in order to avoid an outage situation. In this case, it makes more sense to compute the average rate over a large number of independent channel realizations.

2.2.1.2 Variations in frequency and space

Very few papers in the literature on the fading relay channel consider a realistic frequency-selective fading channel model. This is probably justified by the difficulty to obtain analytical results with such channel models. In this thesis, because we are interested in broadband wireless systems we have to adopt a channel model with time, frequency and space variations. For homogeneity and ease of comparison between our simulations results, we restrict to a single broadband channel model which is the typical urban model of [B05]. Fortunately, OFDM systems can be modeled as a set of parallel Gaussian channels [RC98][BGP02] and the capacity of a MIMO-OFDM link is the sum over all the subcarriers of the individual MIMO flat-fading channel capacities. Therefore we will not need to assume a broadband signal model in our achievable rate derivations.

In state-of-the-art broadband wireless systems, a transmitter which does not have CSIT will use a set of subcarriers spread over the whole channel bandwidth, in order to benefit from frequency diversity. Therefore, the additional space diversity provided by cooperation protocols typically has a much lower impact than what is often claimed in the literature based on flat fading simulations. Therefore simulations in §5.2.1.1 will show

that the outage performance of a cooperative strategy may be significantly different on a broadband channel (OFDM) compared to a narrowband channel (single-carrier) for a given average SNR, whereas the average achievable rate performance is not much different.

2.2.2 Modeling radio device constraints

2.2.2.1 Transmit power constraints

How to realistically model transmit power constraints is definitely a controversial topic. The most common assumption in the MIMO literature is a sum-power constraint over all the transmit antennas, as opposed to a per-antenna power constraint. The reason for assuming a total transmit power constraint is probably threefold:

- It allows a “fair” comparison between MIMO and SISO systems.
- It simplifies the computation of the MIMO capacity: under perfect CSIT, waterfilling on the MIMO channel eigenmodes is optimum only under total power constraint. Otherwise, it is not possible (to my knowledge) to obtain an analytical expression of the optimum precoder.
- It simplifies benchmarking of coding and signal processing strategies among authors.

The same kind of arguments can be used to justify the adoption of a sum-power constraint on the relay channel: assuming a total source plus relay transmit power allows a “fair” comparison between cooperative relaying strategies, whether a single or multiple devices are allowed to transmit. Another “system-level” argument for adopting a total power constraint is the fact that the same interference will be radiated if a sum-power constraint is assumed. In [HZ05], the authors even assume an average transmit power constraint in time domain (equivalent to an energy constraint) when computing capacity bounds on the TDD relay channel, allowing the source to transmit at different power levels during the first and second slot, as long as its average power is below a certain threshold.

Though we acknowledge the advantages of sum power constraints, we will most often assume (unless explicitly stated) individual transmit power constraint at the source and relay. The motivation for this is that at least in cellular systems, different devices have independent power supply. The argument that a total power constraint allows a “fair comparison” in terms of interference generated is also questionable: at a given transmit

power, a relay located on a lamp pole will generate less interference on remote cells than a Base Station located on a tower. Moreover, in order to be even more realistic the following constraints shall be taken into account:

- Per-antenna power constraints are more realistic than sum-power when devices are allowed to operate at full power, since a typical MIMO transmitter embeds one power amplifier per transmit chain.
- Spectrum Mask constraints. Standards for wideband systems (e.g. [16e05]) typically impose a spectral mask that is almost flat in order to make the interference generated by the system as white as possible. This mask can be modeled by a per sub-carrier transmit power constraint in a MIMO-OFDM system.
- In urban deployments the most stringent constraint can be on the maximum radiated power over all directions. Such a constraint is especially complex because it involves the antenna pattern and the transmit weights.

The impact of the the first two implementation constraints listed above on capacity bounds is addressed in Chapter 5. From this short discussion we can conclude that there is not one but several valid assumptions on power constraints, depending on the goal of the study (e.g. benchmarking coding schemes, predicting performance at system-level, ...). It is therefore of interest to be able to compute capacity bounds under various transmit power constraints.

2.2.2.2 RF impairments

Achieving an accurate modeling of the effect of PA non-linearities, phase noise and other RF impairments in capacity bounds is out-of-scope of this thesis. We account for their existence in our simulations by preventing the average SNR at the input of the A/D converters from exceeding a threshold which we fix at 30dB. This means that even if an MS is very close to the BS or RS and receives an input power which is 60dB above the thermal noise threshold, the achievable rate will be computed assuming only 30dB SNR.

2.2.2.3 Achievable rates and link-to-system interface

As mentioned before, there are several approaches to the study of the MIMO relay channel capacity: a first approach consists in computing expressions for achievable rates, outage or ergodic rates. The problem with this approach is that obtaining such expressions becomes highly involved, especially when the number of nodes increases.

Even in the simple 3-node relay channel, the expression of the relay channel capacity is still unknown after 30 years of investigation. Another approach aims at deriving simple expressions that provide trends under simplifying (e.g. high SNR) assumptions: the DMT analysis belongs to this category. Though DMT is both a powerful and beautiful tool, which facilitates the design of space-time codes, it remains limited in the kind of conclusions it can provide: the fact that a strategy achieves a better DMT trade-off than another strategy does not mean that it provides the best achievable rate under the same conditions. As explained below, achievable rates offer the advantage of allowing throughput prediction, which is a convenient interface towards system-level simulations or resource allocation optimization.

2.2.2.3.1 Degraded capacity

The degraded capacity model is introduced in [CCB95] to predict the spectral efficiency of an OFDM system under a given target error rate with per-subcarrier bit loading, though it was probably used in previous works. Provided that the set of MCS in the system offers a small-enough granularity in terms of spectral efficiency, the throughput can be approximated as:

$$\rho \approx \sum_{i=1}^{N_C} \min \left(\log_2 \left(1 + \frac{\gamma_i}{\Gamma} \right), R_{\max} \right) \quad (2.27)$$

where N_C denotes the number of subcarriers, γ_i denotes the SNR on the i th subcarrier, R_{\max} denotes the maximum spectral efficiency over all the Modulation and Coding Schemes (MCS) of the system, and $\Gamma > 1$ is a factor that captures the degradation w.r.t. the capacity that is due to non-ideal modulation and coding, i.e. finite source alphabet, finite-length coding,etc. The degradation factor Γ is a function of the target error rate, and can be graphically interpreted as the distance (measure in dB of SNR) between the Shannon capacity vs SNR curve and the actual spectral efficiency vs SNR point of operation. Typical values observed for state-of-the-art systems range between 7dB to 3dB, and naturally tend to be in the lower when the coding scheme is powerful (e.g. turbo-code, LDPC) and the decoding is close to ML. Note that in the following we assume for simplification that Γ does not depend on the selected MCS. A typical value for R_{\max} is 5 bits per QAM symbol, which corresponds to 64QAM with code rate 5/6.

The formula (2.27) can be extended to MIMO-OFDM systems. For instance, if N_{SS} spatial streams are multiplexed on each subcarrier, we can write:

$$\rho \approx \sum_{i=1}^{N_c} \sum_{j=1}^{N_{SS}} \min \left(\log_2 \left(1 + \frac{\gamma_{i,j}}{\Gamma} \right), R_{\max} \right) \quad (2.28)$$

where $\gamma_{i,j}$ is the SNR on the j th spatial stream of the i th at the output of the receiver spatial processing. The degraded capacity model will be applied to cooperative links in §5.2.1.3, §5.2.2 and §5.2.3.

2.2.2.3.2 Effective SNR mapping

In [NR98], the notion of effective SNR is introduced to perform error prediction of convolutionally coded systems over frequency-selective channels. The effective SNR γ_{eff} is a function of the MCS, the codeword length, the channel realization and the noise variance. It is defined by:

$$PER_{AWGN}(\gamma_{eff}, i_{MCS}, L) \approx PER(\boldsymbol{\gamma}, i_{MCS}, L) \quad (2.29)$$

where PER_{AWGN} is the PER vs SNR function on the AWGN channel, which depends on the index of the MCS and on the length L of the packet, and $\boldsymbol{\gamma}$ is a vector of SNRs on each state of the fading channel. In [E03], the Exponential Effective SNR Mapping (EESM) is proposed to predict the error rate in OFDM systems with Bit-Interleaved Coded Modulation (BICM). In EESM, the following formula is used:

$$\gamma_{eff} = -\beta \log \left(\frac{1}{N_c} \sum_{i=1}^{N_c} \exp \left(-\frac{\gamma_i}{\beta} \right) \right) \quad (2.30)$$

where β is a parameter that equals 1 for BPSK, 2 for QPSK and shall be optimized for other constellations. A possible criterion to optimize β is the variance of the target SNRs for given target error rate PER_{target} over a large-enough set of N_{trial} independent channel trials:

$$\beta(PER_{target}) = \arg \min_{\beta} \sum_{i=1}^{N_{trial}} \left(\gamma_{eff}(i, \beta, PER_{target}) - \frac{1}{N_{trial}} \sum_{i=1}^{N_{trial}} \gamma_{eff}(i, \beta, PER_{target}) \right)^2 \quad (2.31)$$

For details on EESM, we refer the reader to [BSC04][CSL06][SRS05]. In [BSC04], it is shown that EESM allows an accurate error prediction for turbo-coded systems, up to a few tenths of dBs. In [CSL06], it is shown that EESM can also accurately predict errors

in systems employing HARQ (Chase Combining and/or IR). In [SRS05], the application of EESM to space-time coded systems is discussed. Error prediction is the key link-to-system interface feature, as it allows to predict the throughput and the delay. For instance, under unlimited packet retransmissions, the throughput of a system for a given MCS and channel realization (normalized to have unit variance noise) equals:

$$\rho(i_{MCS}) = R(i_{MCS})(1 - PER(i_{MCS}, \mathbf{H})) \quad (2.32)$$

Simulations in Appendix D show that EESM is accurate-enough for the throughput prediction simulations of this thesis.

2.2.2.3.3 Practical application to throughput prediction

On Figure 9, a practical application of EESM and degraded capacity is illustrated for the IEEE802.16e [16e05] Convolutionally Turbo-Coded (CTC) system. The set of MCS considered here ranges from QPSK with code rate $\frac{1}{2}$ to 64QAM, rate $\frac{5}{6}$, the (uncoded) packet length is fixed to 120 Bytes. The channel model assumed is the urban micro model of [B05]. The average throughput over a large number of independent channel trials is plotted, assuming an Adaptive Modulation and Coding (AMC) strategy that maximizes the throughput under perfect CSI, for a target PER of 5%. The channel codeword is mapped onto subcarriers that are pseudo-randomly interleaved over the whole 10 MHz bandwidth (PUSC). The solid curves (blue and red) represent the AMC throughput with either perfect (ideal) error prediction and with EESM. It can be checked that EESM provides an almost perfect error prediction (the standard deviation observed when optimizing β is around 0.2dB). The dashed curve represents the average degraded capacity with a $\Gamma = 4$ dB degradation factor, and the dotted curve represents the Shannon capacity. It can be observed that over the set of average SNRs for which an MCS matching the PER target can be found, the degraded capacity also provides a fairly accurate throughput estimate.

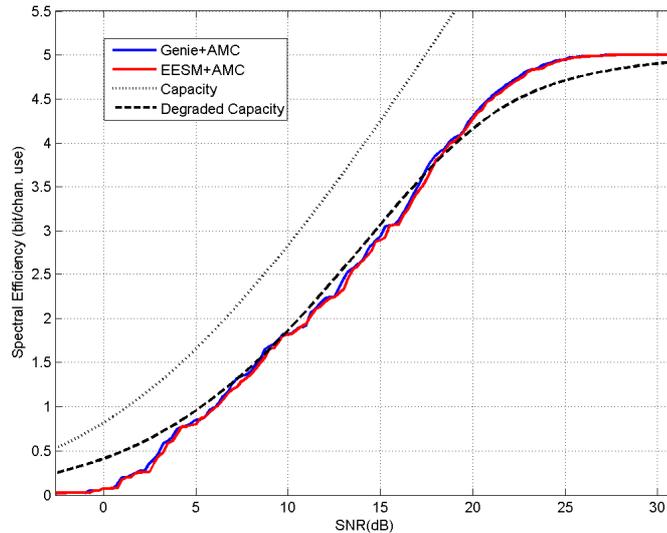


Figure 9: Average Spectral Efficiency in IEEE 802.16e PUSC mode with perfect channel knowledge

2.2.2.3.4 Conclusion on link-to-system interface

Degraded capacity and EESM are two interesting tools to predict the throughput performance of a system. The latter presents the advantage of capturing the exact set of MCS of the system, and achieving a good error prediction accuracy, while the former can be “easily” obtained by inserting a degradation factor and a maximum rate saturation inside achievable rate expressions. Note that more advanced models are being investigated in the literature, such as the MIESM, which uses mutual information under finite alphabet constraint. However, the degraded capacity presents the advantage of having a simpler expression that lends itself easily to convex optimization. We will investigate in this thesis the practical adaptation of the degraded capacity and EESM methodologies for cooperative SISO and MIMO links in §5.2.3.2.

2.2.2.4 From link-level to system-level simulations

The link-to-system interface, whether it is based on EESM or degraded capacity, shall ultimately be used as the input of a system-level simulator. In [VLK07], we publish some preliminary system-level simulation results for cooperative DF and CF protocols. The modeling of phenomena such as the shadowing correlation or the frequency-

selectivity of the co-channel interference can have a significant impact on the conclusions that can be drawn from such simulations. In §5.4 we present a Quasi Monte-Carlo methodology some system-level simulations and discuss the achievable rate performance of DF and CF strategies in cellular deployment topologies.

2.3 Conclusions

In this chapter we have introduced the main system assumptions that will provide the framework for the subsequent chapters. Important assumptions include relay duplexing, transmit power constraints and channel model. The theoretical background on the relay channel is introduced along with capacity bounds and coding strategies for the relay channel (e.g. the cut-set bound, the Decode-and-Forward and Compress-and-Forward strategies). Finally, the connection is made between information-theory and practical system design.

Chapter 3: Capacity bounds for the Gaussian MIMO relay channel - A convex optimization framework

3.1 Introduction and overview of our contribution

In the previous chapter we have overviewed the state-of-the-art on the relay channel. Until recently, the vast majority of studies on cooperative relaying assumed single antenna devices, possibly forming VAAs. However, it is now possible to integrate multiple antennas not only in infrastructure devices (e.g. base stations, fixed relay stations) but also in mobile devices (e.g. handsets), and to exploit MIMO Channel State Information (CSI) not only at the receiving node (CSIR) but also at the transmitting node (CSIT). Point-to-point MIMO with full CSI is now a well investigated topic: the transmit covariance that attains the Gaussian MIMO channel capacity is derived by Telatar in [T99], while the maximization of various other cost functions is performed in [PCL03]. The mature knowledge of the point-to-point MIMO channel has supported the standardization activities in 3GPP-LTE and IEEE802.16m: for a given channel and antenna configuration the capacity can be computed exactly, and every coding and resource allocation strategy can be benchmarked to this reference. Recently, cooperative relaying has been introduced in standardization bodies such as IEEE802.16 j and m and immediately a flurry of coding strategies (most of them probably patented) were proposed by various companies and universities. However, it is very difficult to compare these strategies with each other and it is even unclear how much increase of e.g. spectral efficiency can be theoretically expected by introducing a relay at a given location in a cell, not only at the system-level but even at the link-level. This situation is mainly due to the lack of theoretical results on the MIMO relay channel, even for the simplest three-node topology. Thus, a necessary preliminary step towards a better understanding of the impact of relays in future radio access networks is the extension of the capacity bounds (CSB, AF, DF, CF) to the MIMO case. This topic was still largely unexplored at the beginning of this thesis, especially in the full CSI case, and we therefore decided to focus on it.

We therefore consider in this chapter a single Relay (R) which cooperates with a Source (S) and a Destination (D) to maximize the information rate from S to D. The

number of antennas at S, R and D are respectively denoted N_S , N_R and N_D and each can be greater than 1. Static channels and full CSI are assumed in this chapter (see §2.2.1.1). For reasons explained in §2.1.3 we only briefly address full-duplex relaying and focus more on a “fixed-dynamic” TDD protocol with a relay-receive slot of duration denoted by $t \in [0;1]$ followed by a relay-transmit slot of duration $(1-t)$. In [WZH05], an upper-bound and several lower bounds are derived for the full-duplex MIMO relay channel with full CSI. As explained later in this chapter, the upper-bound in [WZH05] is in general larger than the CSB and besides its numerical evaluation is very computationally complex. In [LVH05], two DF strategies are introduced which are shown to improve the achievable rates bounds in [WZH05]. However, these bounds are still restricted to the full-duplex MIMO relay channel and moreover no generic numerical evaluation procedure is proposed therein. In [YE07][AES05], a Diversity Multiplexing Tradeoff analysis of the full- and half-duplex MIMO relay channel is carried on. The limitation of the DMT tool is that it does not provide actual values for achievable rates and capacity but only trends at high SNR. In [MVA07][HW07], Source and Relay precoders are derived for Linear Relaying (LR) with full MIMO CSI. However, as stated in §2.1.4.3 the problem of LR in the single-relay TDD case is its poor achievable rate performance.

This chapter contributes to the theoretical study of the MIMO relay channel with full CSI as follows:

- The CSB is formulated as a convex optimization problem for both full-duplex and TDD relaying (§3.3). In the full-duplex case, this upper-bound on capacity is tighter than the one proposed in [WZH05]. In the TDD case, the formulation is obtained by exploiting the convexity-preserving property of perspective function [BV04].
- A convex formulation of the achievable rates of DF strategies with either partial or full decoding at the relay is obtained for TDD relaying (§3.5). Two generic procedures are proposed to efficiently compute these upper and lower bounds. For this purpose, various tools are employed from optimization theory and algorithms as well as differentiation techniques. The reader is referred to Appendix A and Appendix B for a review of these tools.
- Sub-optimum source and relay precoder structures are proposed for the DF strategy with full decoding. In this case, either analytical expressions can be derived from

KKT conditions (§3.5.3.1) or at least the problem dimensions can be reduced.
(§3.5.3.2)

Note that parts of this work were published in [SMVC08], which is extended in [SMVC08b] by including the convex formulation in the TDD case, the use of patterned derivatives and a discussion on implementation constraints.

3.2 The Cut-Set Bound with full MIMO CSI

In this section we show that the computation of the CSB can be formulated as a convex optimization problem in the full-duplex and TDD relaying cases. Procedures to solve this problem are then proposed. These procedures will be directly applicable to DF coding strategies in §3.5.

3.3 Formulating the CSB as a convex problem

3.3.1 Full-Duplex Relaying Case

The channels from S to D, S to R and R to D are denoted respectively \mathbf{H}_0 , \mathbf{H}_1 and \mathbf{H}_2 . Moreover, unless explicitly stated otherwise, each node is subject to a maximum transmit power constraint, denoted by P_S and P_R for the source and relay respectively. In full-duplex MIMO relaying, the signals received at the relay and destination can be written as in [WZH05]:

$$\begin{aligned}\mathbf{y}_R &= \mathbf{H}_1 \mathbf{x}_S + \mathbf{n}_R \\ \mathbf{y}_D &= \mathbf{H}_0 \mathbf{x}_S + \mathbf{H}_2 \mathbf{x}_R + \mathbf{n}_D\end{aligned}\tag{3.1}$$

where circularly symmetric complex white Gaussian noise of unit variance is assumed at the relay and destination, i.e. $\mathbf{n}_R \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_R})$ and $\mathbf{n}_D \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_D})$. The capacity C_{FD} of the full-duplex relay channel is upper-bounded by the cut-set bound $C_{CSB,FD}$ whose expression is given by equation (3) in [WZH05]:

$$C_{FD} \leq C_{CSB,FD} = \max_{p(\mathbf{x}_S, \mathbf{x}_R)} \min \left(I(\mathbf{x}_S; \mathbf{y}_D, \mathbf{y}_R | \mathbf{x}_R), I(\mathbf{x}_S, \mathbf{x}_R; \mathbf{y}_D) \right)\tag{3.2}$$

where the maximization is performed over the joint distribution of the source and relay codebooks $p(\mathbf{x}_S, \mathbf{x}_R)$. The authors in [WZH05] show that the optimum $p(\mathbf{x}_S, \mathbf{x}_R)$ is Gaussian and conclude that the optimization of (3.2) must be carried on w.r.t. three matrices $\mathbf{R}_S \triangleq E[\mathbf{x}_S \mathbf{x}_S^H]$, $\mathbf{R}_R \triangleq E[\mathbf{x}_R \mathbf{x}_R^H]$ and the cross-correlation $E[\mathbf{x}_S \mathbf{x}_R^H]$. This optimization seems highly non-trivial and non-convex. Therefore the authors exploit

some matrix inequalities and introduce a scalar parameter ρ that captures the cross-correlation. They finally obtain (see Theorem 3.1 in [WZH05]) an upper-bound which involves a maximization only over \mathbf{R}_S , \mathbf{R}_R and ρ :

$$\begin{aligned}
C_{FD} &\leq C_{CSB,FD} \leq \max_{\rho \in [0;1], \mathbf{R}_S \geq 0, \mathbf{R}_R \geq 0} \min(C_A, C_B) \\
C_A &\triangleq \mathcal{C} \left(\mathbf{R}_S, \sqrt{1-\rho^2} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \right) \\
C_B &\triangleq \inf_{a>0} \mathcal{C} \left(\begin{bmatrix} \mathbf{R}_S & \mathbf{0}_{N_S \times N_R} \\ \mathbf{0}_{N_R \times N_S} & \mathbf{R}_R \end{bmatrix}, \begin{bmatrix} \sqrt{1+\frac{\rho^2}{a}} \mathbf{H}_0 & \sqrt{1+a} \mathbf{H}_2 \end{bmatrix} \right) \\
\text{s.t.} & \quad \text{tr}(\mathbf{R}_S) \leq P_S \quad \text{tr}(\mathbf{R}_R) \leq P_R
\end{aligned} \tag{3.3}$$

Although its derivation is very elegant, this bound unfortunately suffers several restrictions:

- It is in the general case strictly larger than the CSB (e.g. equality with the CSB requires $N_S \leq N_R$)
- Although C_A and C_B are concave in \mathbf{R}_S and \mathbf{R}_R for a fixed ρ , the problem is not convex in $(\mathbf{R}_S, \mathbf{R}_R, \rho)$ and thus the proposed algorithm in [WZH05] includes a non-convex one-dimensional optimization over ρ .
- Its numerical evaluation is computationally intensive. Indeed, it is not possible to obtain a closed-form expression for the derivatives of C_B w.r.t. \mathbf{R}_S and \mathbf{R}_R . Therefore, at each step of the optimization where a gradient shall be computed, we have to evaluate numerically the derivative with respect to each component of these two matrices, which requires a number of evaluations of C_B that is proportional to N_S^2 (resp. N_R^2), and each evaluation of C_B requires by definition to solve a one-dimensional optimization with respect to parameter a .

The above-mentioned limitations can be overcome by considering the joint covariance matrix:

$$\mathbf{R}_{SR} \triangleq \begin{bmatrix} \mathbf{R}_S & E[\mathbf{x}_S \mathbf{x}_R^H] \\ E[\mathbf{x}_R \mathbf{x}_S^H] & \mathbf{R}_R \end{bmatrix} \tag{3.4}$$

Let also define the following matrices:

$$\mathbf{D}_S \triangleq \begin{bmatrix} \mathbf{I}_{N_S} & \mathbf{0}_{N_S \times N_R} \end{bmatrix} \quad \text{and} \quad \mathbf{D}_R \triangleq \begin{bmatrix} \mathbf{0}_{N_R \times N_S} & \mathbf{I}_{N_R} \end{bmatrix} \tag{3.5}$$

From (3.4) and (3.5), the following relationships hold:

$$\mathbf{R}_S = \mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H \quad \text{and} \quad \mathbf{R}_R = \mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H \quad (3.6)$$

Note that if \mathbf{R}_{SR} is PSD, then \mathbf{R}_S and \mathbf{R}_R are PSD too:

$$\mathbf{R}_{SR} \succeq 0 \Rightarrow \mathbf{R}_S \succeq 0 \quad \text{and} \quad \mathbf{R}_R \succeq 0 \quad (3.7)$$

Indeed, for any two vectors $\mathbf{x} \in \mathbb{C}^{N_S}$ and $\mathbf{y} \in \mathbb{C}^{N_R}$, defining $\tilde{\mathbf{x}} \triangleq \mathbf{D}_S^H \mathbf{x}$ and $\tilde{\mathbf{y}} \triangleq \mathbf{D}_R^H \mathbf{y}$, the following holds:

$$\begin{aligned} \mathbf{x}^H \mathbf{R}_S \mathbf{x} &\stackrel{(a)}{=} \mathbf{x}^H \mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H \mathbf{x} = \tilde{\mathbf{x}}^H \mathbf{R}_{SR} \tilde{\mathbf{x}} \stackrel{(b)}{\geq} 0 \\ \mathbf{y}^H \mathbf{R}_R \mathbf{y} &\stackrel{(a)}{=} \mathbf{y}^H \mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H \mathbf{y} = \tilde{\mathbf{y}}^H \mathbf{R}_{SR} \tilde{\mathbf{y}} \stackrel{(b)}{\geq} 0 \end{aligned} \quad (3.8)$$

where (a) comes from equation (3.6) and (b) from the positive semi-definiteness of \mathbf{R}_{SR} . Note that (3.7) is a well-known result from matrix algebra: any principal submatrix of a PSD matrix is also PSD.

The cut-set bound (3.2) can therefore be expressed as:

$$\begin{aligned} C_{CSB,FD} &= \max_{\mathbf{R}_{SR} \succeq 0} \min \left\{ \mathcal{C} \left(\mathbf{R}_{SR}, \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \mathbf{D}_S \right), \mathcal{C} \left(\mathbf{R}_{SR}, \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_2 \end{bmatrix} \right) \right\} \\ \text{s.t.} & \quad \text{tr}(\mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H) \leq P_S \quad \text{and} \quad \text{tr}(\mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H) \leq P_R \end{aligned} \quad (3.9)$$

The objective is concave on the PSD cone, because it is the pointwise minimum of two concave functions[BV04]. The constraints are affine. Therefore the problem is convex. Thus, we can rely on the convex optimization literature (see Appendix B) to solve the problem efficiently.

3.3.2 TDD Relaying Case

Three TDD relaying protocols have been defined in §2.1.3.1. In this section, we consider the more general and more complex Protocol III, from which Protocols I and II can be easily derived. The CSB can be expressed as follows (see equation (77) in [HZ05]):

$$\begin{aligned} C_{CSB,TDD} &= \max_{t \in [0:1], p(\mathbf{x}_S^{(1)}, \mathbf{x}_S^{(2)}, \mathbf{x}_R^{(2)})} \min \{C_A, C_B\} \\ C_A &\triangleq tI(\mathbf{x}_S^{(1)}; \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)} | \mathbf{x}_R^{(1)} = \mathbf{0}) + (1-t)I(\mathbf{x}_S^{(2)}; \mathbf{y}_D^{(2)} | \mathbf{x}_R^{(2)}) \\ C_B &\triangleq tI(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)} | \mathbf{x}_R^{(1)} = \mathbf{0}) + (1-t)I(\mathbf{x}_S^{(2)}, \mathbf{x}_R^{(2)}; \mathbf{y}_D^{(2)}) \end{aligned} \quad (3.10)$$

where the superscript ⁽ⁱ⁾ indicates the slot during which the signal was transmitted or received. Using similar arguments as in the full-duplex case, the cut-set bound for the TDD MIMO relay channel can be expressed as:

$$\begin{aligned}
C_{CSB.TDD} &= \max_{t \in [0:1], \mathbf{R}_S^{(1)} \succeq 0, \mathbf{R}_{SR}^{(2)} \succeq 0} \min \{C_A, C_B\} \\
C_A &\triangleq t \mathcal{C} \left(\mathbf{R}_S^{(1)}, \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \right) + (1-t) \mathcal{C} \left(\mathbf{R}_{SR}^{(2)}, \mathbf{H}_0 \mathbf{D}_S \right) \\
C_B &\triangleq t \mathcal{C} \left(\mathbf{R}_S^{(1)}, \mathbf{H}_0 \right) + (1-t) \mathcal{C} \left(\mathbf{R}_{SR}^{(2)}, [\mathbf{H}_0 \quad \mathbf{H}_2] \right) \\
&\text{s.t.} \\
&\text{tr} \left(\mathbf{R}_S^{(1)} \right) \leq P_S ; \text{tr} \left(\mathbf{D}_S \mathbf{R}_{SR}^{(2)} \mathbf{D}_S^H \right) \leq P_S ; \text{tr} \left(\mathbf{D}_R \mathbf{R}_{SR}^{(2)} \mathbf{D}_R^H \right) \leq P_R
\end{aligned} \tag{3.11}$$

This problem (3.11) is convex in $(\mathbf{R}_S^{(1)}, \mathbf{R}_{SR}^{(2)})$ for a given t , but convexity in $(\mathbf{R}_S^{(1)}, \mathbf{R}_{SR}^{(2)}, t)$ cannot be claimed at this stage. However, let us introduce the following changes of variables into (3.11): $t_1 \triangleq t$, $t_2 \triangleq 1-t$, $\mathbf{Q}_S^{(1)} \triangleq t_1 \mathbf{R}_S^{(1)}$ and $\mathbf{Q}_{SR}^{(2)} \triangleq t_2 \mathbf{R}_{SR}^{(2)}$. The problem now reads:

$$\begin{aligned}
C_{CSB.TDD} &= \max_{t_1 > 0, t_2 > 0, \mathbf{Q}_S^{(1)} \succeq 0, \mathbf{Q}_{SR}^{(2)} \succeq 0} \min \{C_A, C_B\} \\
C_A &\triangleq t_1 \mathcal{C} \left(\mathbf{Q}_S^{(1)} / t_1, \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \right) + t_2 \mathcal{C} \left(\mathbf{Q}_{SR}^{(2)} / t_2, \mathbf{H}_0 \mathbf{D}_S \right) \\
C_B &\triangleq t_1 \mathcal{C} \left(\mathbf{Q}_S^{(1)} / t_1, \mathbf{H}_0 \right) + t_2 \mathcal{C} \left(\mathbf{Q}_{SR}^{(2)} / t_2, [\mathbf{H}_0 \quad \mathbf{H}_2] \right) \\
&\text{s.t.} \\
&\text{tr} \left(\mathbf{Q}_S^{(1)} \right) - t_1 P_S \leq 0 ; \text{tr} \left(\mathbf{D}_S \mathbf{Q}_{SR}^{(2)} \mathbf{D}_S^H \right) - t_2 P_S \leq 0 ; \\
&\text{tr} \left(\mathbf{D}_R \mathbf{Q}_{SR}^{(2)} \mathbf{D}_R^H \right) - t_2 P_R \leq 0 ; t_1 + t_2 - 1 \leq 0
\end{aligned} \tag{3.12}$$

Note that in (3.12) the trivial cases $t=0$ and $t=1$ were excluded from the domain. Indeed, in these two extreme cases it can easily be checked that cooperative relaying degenerates into point-to-point MIMO. The function $g : (\mathbf{X}, t) \mapsto tf(\mathbf{X}/t)$ defined on $\mathcal{S}_+^M \times \mathbb{R}_{++}$ is the perspective (cf. sec 3.2.6 in [BV04]) of the function $f : \mathbf{X} \mapsto \mathcal{C}(\mathbf{X}, \mathbf{H})$ which is concave on \mathcal{S}_+^M , therefore g is concave on $\mathcal{S}_+^M \times \mathbb{R}_{++}$. The problem (3.12) is therefore convex in standard form.

3.4 Computing the CSB

The previous section has shown that the CSB can be expressed as the solution of a convex problem in the full-duplex and TDD cases. Since it does not seem possible to derive a closed-form expression of the solution of (3.9) and (3.12), efficient procedures are sought hereafter to solve them numerically. We focus on solving (3.12) from which the solution of the simpler problem (3.9) will be straightforward.

We start by writing the equivalent epigraph [BV04] form of (3.12) such that closed-form expressions of partial derivatives can be found for the objective and the constraints:

$$\begin{aligned}
C_{CSB,TDD} &= - \min_{(\varepsilon, t_1, t_2, \mathbf{Q}_S^{(1)}, \mathbf{Q}_{SR}^{(2)}) \in \mathcal{X}} -\varepsilon \\
\text{s.t.} & \\
\varepsilon - C_A &\leq 0 ; \quad \varepsilon - C_B \leq 0 & (3.13) \\
\text{tr}(\mathbf{Q}_S^{(1)}) - t_1 P_S &\leq 0 ; \quad \text{tr}(\mathbf{D}_S \mathbf{Q}_{SR}^{(2)} \mathbf{D}_S^H) - t_2 P_S \leq 0 ; \\
\text{tr}(\mathbf{D}_R \mathbf{Q}_{SR}^{(2)} \mathbf{D}_R^H) - t_2 P_R &\leq 0 ; \quad t_1 + t_2 - 1 \leq 0
\end{aligned}$$

where

$$\mathcal{X} \triangleq \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathcal{S}_+^{N_S} \times \mathcal{S}_+^{N_S + N_R} \quad (3.14)$$

The optimization problem in equation (3.13) needs to be carried out with respect to three real-valued variables and two PSD matrices. The computation of closed-form expressions for the partial derivatives and gradients with respect to the time-sharing parameters t_1 and t_2 and to the structured (here Hermitian) matrices $\mathbf{Q}_S^{(1)}$ and $\mathbf{Q}_{SR}^{(2)}$ is detailed in §A.2. From (6.14), let define the following parameterization of matrices $\mathbf{Q}_S^{(1)}$ and $\mathbf{Q}_{SR}^{(2)}$:

$$\mathbf{Q}_S^{(1)} \triangleq \mathcal{F}(\mathbf{r}_S^{(1)}, \mathbf{c}_S^{(1)}, (\mathbf{c}_S^{(1)})^*) \quad (3.15)$$

$$\mathbf{Q}_{SR}^{(2)} \triangleq \mathcal{F}(\mathbf{r}_{SR}^{(2)}, \mathbf{c}_{SR}^{(2)}, (\mathbf{c}_{SR}^{(2)})^*) \quad (3.16)$$

In order to simplify the notations, all the variables in the problem are stacked into the following vector:

$$\mathbf{v} \triangleq \left(\varepsilon, t_1, t_2, (\mathbf{r}_S^{(1)})^T, (\mathbf{c}_S^{(1)})^T, (\mathbf{r}_{SR}^{(2)})^T, (\mathbf{c}_{SR}^{(2)})^T \right)^T \quad (3.17)$$

and the set \mathcal{V} is defined such that

$$\mathbf{v} \in \mathcal{V} \Leftrightarrow (\varepsilon, t_1, t_2, \mathbf{Q}_S^{(1)}, \mathbf{Q}_{SR}^{(2)}) \in \mathcal{X} \quad (3.18)$$

Moreover, the inequality constraints in (3.13) are denoted as $f_j(\mathbf{v}) \leq 0$, $j = 1, \dots, J$ and stacked into a vector-valued function \mathbf{f} as follows:

$$\mathbf{f}(\mathbf{v}) \triangleq (f_1(\mathbf{v}), \dots, f_J(\mathbf{v}))^T \quad (3.19)$$

Using the notations (3.15)-(3.19), the problem can now be solved using the derivatives and gradient expressions of §A.2 and the algorithms of Appendix B. Two alternative ways of solving (3.13) are now presented.

3.4.1 Dual Method

A procedure to solve the dual problem of (3.13) is detailed in §B.3. Here the primal problem is convex and Slater's condition for strong duality holds. Indeed, the following point for instance is strictly feasible:

$$\mathbf{x}_0 \triangleq \left(0, \frac{1}{4}, \frac{1}{4}, \frac{P_S}{8N_S} \mathbf{I}_{N_S}, \frac{\min(P_S, P_R)}{8(N_S + N_R)} \mathbf{I}_{N_S + N_R} \right) \in \tilde{\mathcal{X}} \quad (3.20)$$

which satisfies Slater's condition. The computation of the dual function $g(\boldsymbol{\mu})$ at a given $\boldsymbol{\mu}_0$ requires to minimize the Lagrangian $\mathcal{L}(\mathbf{v}, \boldsymbol{\mu}_0)$ over \mathcal{V} . The gradient of the Lagrangian $\nabla \mathcal{L}(\mathbf{v}, \boldsymbol{\mu}_0)$ can be computed in closed-form from the formulas in Appendix A. However, a straightforward application of gradient descent methods cannot guarantee that the sequence of points belong to \mathcal{V} . As explained hereafter, the projection $\mathcal{P}_{\mathcal{V}}$ on the set \mathcal{V} is simple to implement and we therefore resort to the GPM, which is described in details in §B.1. The projection of t_1 and t_2 onto \mathbb{R}_{++} can be practically handled by restricting their domain to a closed interval $[\eta; +\infty)$ where $\eta > 0$. In this case their projections are respectively $\max(t_1, \eta)$ and $\max(t_2, \eta)$. By selecting a small enough η , the error introduced on the final solution can be made arbitrarily small. The projection of $\mathbf{Q}_S^{(1)}$ and $\mathbf{Q}_{SR}^{(2)}$ onto the PSD cone follows (6.36). Having minimized the Lagrangian, it remains to maximize the dual function. As explained in §B.3, a closed-form expression of a subgradient is easily found and we can therefore resort to the subgradient method. However, as explained in §B.3 step size selection strategies for the subgradient method are empirical and in our simulations the procedure described in the next section for solving the primal problem based on the barrier method turned out to converge significantly faster.

3.4.2 Barrier Method

In order to solve (3.13), a more straightforward approach is to use an interior-point method to solve the primal problem. The barrier method is described in §B.2. Closed-form expressions for the derivatives and gradient of log-barrier functions are obtained from the formulas in §A.1 and §A.2. For instance,

$$\mathcal{D}_{(\mathbf{c}_S^{(1)})} \phi_j \left(\mathbf{v} \left(\varepsilon, t_1, t_2, \mathbf{r}_S^{(1)}, \mathbf{c}_S^{(1)}, \mathbf{r}_{SR}^{(2)}, \mathbf{c}_{SR}^{(2)} \right) \right) = \left(\text{vec} \left(\frac{1}{f_j(\mathbf{v})} \frac{\partial f_j(\mathbf{v})}{\partial \mathbf{Q}_S^{(1)}} \right) \right)^T \mathbf{L}_{u, S^{(1)}} \quad (3.21)$$

where $\mathbf{L}_{u,S^{(1)}}$ is the matrix that maps the $N_S(N_S-1)/2$ independent complex components from the upper-part of $\mathbf{Q}_S^{(1)}$ onto the components of indices $3+N_S+1, \dots, 3+N_S+N_S(N_S-1)/2$ of vector \mathbf{v} . The barrier method can start from any interior point such as the one given by (3.20). In our simulations it converges much faster than the bound in [WZH05] and also significantly faster than the dual method in the section above, and was therefore the preferred method for simulation.

3.5 Decode-and-Forward with full MIMO CSI

The partial and full DF strategies for the Gaussian scalar TDD relay channel have been introduced in §2.1.4.1. In this section, we extend the achievable rate expressions to the Gaussian MIMO relay channel.

3.5.1 Non-Cooperative Decode-and-Forward

We start by treating the simple case of non-cooperative DF (NCDF) as defined in §2.1.4.1. It will be used as a benchmark to highlight the gains due to cooperation in simulations. The achievable rate for NCDF is:

$$R_{NC,DF} \triangleq \max_{t \in [0;1], \mathbf{R}_S^{(1)} \geq 0, \mathbf{R}_R^{(2)} \geq 0} \min \left(t\mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_1), (1-t)\mathcal{C}(\mathbf{R}_R^{(2)}, \mathbf{H}_2) \right) \quad (3.22)$$

s.t. $\text{tr}(\mathbf{R}_S^{(1)}) \leq P_S; \text{tr}(\mathbf{R}_R^{(2)}) \leq P_R$

$$R_{NC,DF} = \max_{t \in [0;1]} \left(t\hat{\mathcal{C}}(\mathbf{H}_1, P_S), (1-t)\hat{\mathcal{C}}(\mathbf{H}_2, P_R) \right) \quad (3.23)$$

$$= \frac{\hat{\mathcal{C}}(\mathbf{H}_1, P_S)\hat{\mathcal{C}}(\mathbf{H}_2, P_R)}{\hat{\mathcal{C}}(\mathbf{H}_1, P_S) + \hat{\mathcal{C}}(\mathbf{H}_2, P_R)}$$

where

$$\hat{\mathcal{C}}(\mathbf{H}, P) \triangleq \max_{\mathbf{R} \geq 0} \mathcal{C}(\mathbf{R}, \mathbf{H}) \text{ s.t. } \text{tr}(\mathbf{R}) \leq P \quad (3.24)$$

Note that problem (3.24) is the MIMO channel capacity with CSIT as solved by Telatar in [T99]. From (3.23), it is clear that the achievable rate of NCDF is upper-bounded by the minimum between the capacity of the first hop link and that of the second hop link:

$$R_{NC,DF} \leq \min \left(\hat{\mathcal{C}}(\mathbf{H}_1, P_S), \hat{\mathcal{C}}(\mathbf{H}_2, P_R) \right) \quad (3.25)$$

3.5.2 Partial Decode-and-Forward

The coding scheme is the same as in the scalar case §2.1.4.1. The source transmits a first message ω_0 at a rate R_0 using a signal $\mathbf{x}_S^{(1)}(\omega_0)$ during the first slot. The relay

decodes $\hat{\omega}_0$ and transmits $\mathbf{x}_R^{(2)}(\hat{\omega}_0)$ during the second slot while S transmits $\mathbf{x}_S^{(2)}(\omega_0, \omega_1) = \mathbf{x}_{S,0}^{(2)}(\omega_0) + \mathbf{x}_{S,1}^{(2)}(\omega_1)$. Because we assume a synchronized scenario, the signals $\mathbf{x}_R^{(2)}(\omega_0)$ and $\mathbf{x}_{S,0}^{(2)}(\omega_0)$ are correlated, whereas ω_1 is mapped onto an independent signal $\mathbf{x}_{S,1}^{(2)}(\omega_1)$ transmitted at rate R_1 using superposition coding. The destination successively decodes $\hat{\omega}_0$ and $\hat{\omega}_1$. The derivation of the achievable rate is a straightforward extension of the proof of Prop. 2 in [HZ05]:

$$\begin{aligned} R_{\text{PDF}} &= \max_{t \in [0;1], p(\mathbf{x}_S^{(1)}, \mathbf{x}_{S,0}^{(2)}, \mathbf{x}_{S,1}^{(2)}, \mathbf{x}_R^{(2)})} R_0 + R_1 \\ &= \max_{t \in [0;1], \mathbf{R}_S^{(1)} \succeq 0, \mathbf{R}_{S,1}^{(2)} \succeq 0, \mathbf{R}_{SR,0}^{(2)} \succeq 0} \min(R_A, R_B) \end{aligned} \quad (3.26)$$

$$R_A \triangleq t\mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_1) + (1-t)\mathcal{C}(\mathbf{R}_{S,1}^{(2)}, \mathbf{H}_0 \mathbf{D}_S) \quad (3.27)$$

$$R_B \triangleq t\mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_0) + (1-t)\mathcal{C}\left(\begin{bmatrix} \mathbf{R}_{S,1}^{(2)} & \mathbf{0}_{N_S \times (N_S + N_R)} \\ \mathbf{0}_{(N_S + N_R) \times N_S} & \mathbf{R}_{SR,0}^{(2)} \end{bmatrix}, [\mathbf{H}_0 \quad \mathbf{H}_0 \quad \mathbf{H}_2]\right) \quad (3.28)$$

where $\mathbf{R}_S^{(1)} \triangleq E[\mathbf{x}_S^{(1)}(\mathbf{x}_S^{(1)})^H]$, $\mathbf{R}_{S,1}^{(2)} \triangleq E[\mathbf{x}_{S,1}^{(2)}(\mathbf{x}_{S,1}^{(2)})^H]$ and $\mathbf{R}_{SR,0}^{(2)} \triangleq E\left[\begin{bmatrix} (\mathbf{x}_{S,0}^{(2)})^T & (\mathbf{x}_R^{(2)})^T \end{bmatrix} \begin{bmatrix} (\mathbf{x}_{S,0}^{(2)})^H & (\mathbf{x}_R^{(2)})^H \end{bmatrix}\right]$ and the transmit power constraints can be written as:

$$\text{tr}(\mathbf{R}_S^{(1)}) \leq P_S \quad ; \quad \text{tr}(\mathbf{D}_R \mathbf{R}_{SR,0}^{(2)} \mathbf{D}_R^H) \leq P_R \quad (3.29)$$

$$\text{tr}(\mathbf{R}_{S,1}^{(2)}) + \text{tr}(\mathbf{D}_S \mathbf{R}_{SR,0}^{(2)} \mathbf{D}_S^H) \leq P_S \quad (3.30)$$

The problem (3.26) with the constraints (3.29) and (3.30) is similar to (3.12) and can be turned into a convex problem in standard form just like (3.13). The Source and Relay precoders during the first and second slot and the time sharing variable which maximize the achievable rate of the partial DF strategy can be computed by exactly the same procedures as the CSB in §3.4.

3.5.3 Full Decode-and-Forward

The partial DF strategy requires to implement superposition coding at the source and successive interference cancellation at the destination. In this section, we consider the full DF (FDF) strategy already introduced in §2.1.4.1 for the scalar case, which allows for a reduced implementation complexity at the expense of a lower achievable rate. Remember that FDF is a variant of partial DF in which the relay decodes the whole message and the

source does not superimpose a new message during the second slot (i.e. $R_1 = 0$). In this case, the achievable rate simplifies as:

$$R_{\text{FDF}} = \max_{t \in [0;1], \mathbf{R}_S^{(1)} \geq 0, \mathbf{R}_{SR}^{(2)} \geq 0} \min(tR_A, tR_{B,1} + (1-t)R_{B,2}) \quad (3.31)$$

$$\text{s.t.} \quad R_A \triangleq \mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_1) \quad (3.32)$$

$$R_{B,1} \triangleq \mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_0) \quad (3.33)$$

$$R_{B,2} \triangleq \mathcal{C}(\mathbf{R}_{SR}^{(2)}, [\mathbf{H}_0 \quad \mathbf{H}_2]) \quad (3.34)$$

$$\text{tr}(\mathbf{R}_S^{(1)}) \leq P_S \quad ; \quad \text{tr}(\mathbf{D}_R \mathbf{R}_{SR}^{(2)} \mathbf{D}_R^H) \leq P_R \quad ; \quad \text{tr}(\mathbf{D}_S \mathbf{R}_{SR}^{(2)} \mathbf{D}_S^H) \leq P_S \quad (3.35)$$

For any fixed $t \in [0;1]$, it can be noticed that R_{FDF} is a non-decreasing function of $R_{B,2}$, which only depends on $\mathbf{R}_{SR}^{(2)}$. Therefore the optimization (3.31) can be carried out in two steps:

$$\hat{R}_{B,2} \triangleq \max_{\mathbf{R}_{SR}^{(2)} \geq 0} R_{B,2} \quad (3.36)$$

$$R_{\text{FDF}} = \max_{t \in (0;1], \mathbf{Q}_S^{(1)} \geq 0} \min \left\{ \begin{array}{l} t\mathcal{C}(\mathbf{Q}_S^{(1)} / t, \mathbf{H}_1), \\ t\mathcal{C}(\mathbf{Q}_S^{(1)} / t, \mathbf{H}_0) + (1-t)\hat{R}_{B,2} \end{array} \right\} \quad (3.37)$$

where the trivial case $t = 0$ was again excluded from the domain. As before, problems (3.36) and (3.37) are convex and can be solved using the same procedures as for the CSB and PDF. However, we decide to evaluate sub-optimum precoders at the Source and Relay with a structure that further reduces the optimization complexity.

3.5.3.1 Sub-optimum Source precoder during 1st slot

Let first consider the problem (3.37) in which the source precoder during the first time slot is optimized. Let introduce the SVD of \mathbf{H}_0 and \mathbf{H}_1 :

$$\mathbf{H}_0 = \mathbf{U}_0 \text{diag}(\boldsymbol{\lambda}_0) \mathbf{V}_0^H \quad \text{and} \quad \mathbf{H}_1 = \mathbf{U}_1 \text{diag}(\boldsymbol{\lambda}_1) \mathbf{V}_1^H \quad (3.38)$$

We arbitrarily impose the following structure to the source covariance matrix:

$$\mathbf{R}_S^{(1)} = \underbrace{\mathbf{V}_0 \text{diag}(\mathbf{p}_0) \mathbf{V}_0^H}_{\triangleq \mathbf{R}_0} + \underbrace{\mathbf{V}_1 \text{diag}(\mathbf{p}_1) \mathbf{V}_1^H}_{\triangleq \mathbf{R}_1} \quad (3.39)$$

The structure stems from the intuition that the source shall transmit part of its power on the eigenmodes of the channel to the relay and the rest on the eigenmodes of the channel to the destination (note the similarity with the precoder optimization in [HKE07]). Let L_0

and L_1 denote the number of non-zero singular values of \mathbf{H}_0 and \mathbf{H}_1 . Inserting (3.38) and (3.39) into (3.32) and (3.33) gives:

$$\begin{aligned} R_A &= \log_2 \left| \mathbf{I}_{N_R} + \mathbf{H}_1 \mathbf{R}_0 \mathbf{H}_1^H + \mathbf{H}_1 \mathbf{R}_1 \mathbf{H}_1^H \right| \\ &\stackrel{(a)}{\geq} \log_2 \left| \mathbf{I}_{N_R} + \mathbf{H}_1 \mathbf{R}_1 \mathbf{H}_1^H \right| \stackrel{(b)}{=} \sum_{i=1}^{L_1} \log_2 (1 + \lambda_{1,i}^2 p_{1,i}) \triangleq J_1 \\ R_{B,1} &\stackrel{(a)}{\geq} \log_2 \left| \mathbf{I}_{N_D} + \mathbf{H}_0 \mathbf{R}_0 \mathbf{H}_0^H \right| \stackrel{(b)}{=} \sum_{i=1}^{L_0} \log_2 (1 + \lambda_{0,i}^2 p_{0,i}) \triangleq J_0 \end{aligned} \quad (3.40)$$

Where (a) comes from the Minkowski determinant inequality and (b) comes from (3.39). Inserting these lower bounds on R_A and $R_{B,1}$ into the epigraph form of (3.37) yields the following lower-bound on R_{DF} :

$$R_{DF} \geq \underset{\varepsilon, t \in [0,1], \mathbf{p}_0 \geq \mathbf{0}_{L_0}, \mathbf{p}_1 \geq \mathbf{0}_{L_1}}{-\min} (-\varepsilon) \quad (3.41)$$

$$\text{s.t.} \quad \varepsilon - tJ_1 \leq 0 \quad (3.42)$$

$$\varepsilon - tJ_0 - (1-t)\hat{R}_{B,2} \leq 0 \quad (3.43)$$

$$\mathbf{1}_{L_0}^T \mathbf{p}_0 + \mathbf{1}_{L_1}^T \mathbf{p}_1 - P_S \leq 0 \quad (3.44)$$

Before solving the above-defined problem, it can first be noticed that when the source-relay and source-destination channels are orthogonal, equality occurs in (3.40). If \mathbf{H}_0 and \mathbf{H}_1 have i.i.d. complex Gaussian components, let consider the distribution of the angle ϕ between any two rows \mathbf{h}_0 and \mathbf{h}_1 of respectively \mathbf{H}_0 and \mathbf{H}_1 . For $N_S > 1$ the quantity $\xi \triangleq |\mathbf{h}_1^* \mathbf{h}_0|^2 / \|\mathbf{h}_0\|_2^2 \|\mathbf{h}_1\|_2^2 = \cos^2(\phi)$ is Beta-distributed with parameters 1 and $N_S - 1$ [J06]. This distribution concentrates around 0 as $N_S \rightarrow +\infty$. Therefore, the source precoder that solves (3.41) becomes optimum for (3.37) as N_S grows.

We now derive a procedure for solving (3.41). Introducing the perspective function as in (3.37) allows to turn (3.41) into a convex problem which has a reduced number of dimensions compared to (3.41) leading to a reduction of the computational complexity. Unfortunately, writing the KKT conditions for this problem does not seem to lead to a closed-form expression. However, for a fixed t in (3.41), the optimization w.r.t. $(\varepsilon, \mathbf{p}_1, \mathbf{p}_2)$ is also a convex problem for which the KKT conditions lead to:

$$\hat{p}_{0,i} = (\alpha - 1 / \lambda_{0,i}^2)^+ ; \hat{p}_{1,i} = (\beta - 1 / \lambda_{1,i}^2)^+ \quad (3.45)$$

The solution (3.45) can be obtained by a water-filling algorithm with two water levels $\alpha > 0$ and $\beta > 0$ which are not independent due to the total source power constraint. Define $P_1 \triangleq (\mathbf{1}_{L_1}^T \mathbf{p}_1) / P_S$ the fraction of Source power transmitted on the source-relay

channel eigenmodes, while the rest $1 - P_1$ is transmitted on the eigenmodes of the source-destination channel. Finding α and β amounts to finding the optimum $\hat{P}_1 \in [0; 1]$. Let first assume that both constraints (3.42) and (3.43) are active at the optimum, which yields:

$$tJ_1 = tJ_0 - (1-t)\hat{R}_{B,2} \quad (3.46)$$

It can be checked that $J_1 = 0$ at $P_1 = 0$ and J_1 is non-decreasing with P_1 . Likewise, J_0 is non-increasing with P_1 and equals 0 at $P_1 = 1$. Therefore, the optimum \hat{P}_1 is found by solving (3.46) under the condition that $(1-t)\hat{R}_{B,2} \leq tJ_1$ at $P_1 = 1$. When this constraint is not satisfied, either (3.42) or (3.43) is not active and the solution is trivial (i.e. $\hat{P}_1 = 0$ or $\hat{P}_1 = 1$). In order to solve (3.41), it remains to perform a one-dimensional optimization with respect to the variable t . Fortunately, it can be checked numerically that the solution $\hat{\mathcal{E}}(t)$ of (3.41) for a given t turns out to be a unimodal function of t over the interval $[0; 1]$, i.e. a function that is either strictly increasing or strictly decreasing. Therefore efficient one-dimensional search techniques such as the Golden Section Search (see appendix C.3 in [B99]) can be employed to find the optimum \hat{t} .

To summarize, a sub-optimum approach to source precoder optimization with reduced complexity is proposed in this section. It consists in transmitting a fraction of the source power on the eigenmodes of the channel to the relay and the rest on the eigenmodes of the channel to the destination. The power assignment is provided by a water-filling algorithm with two water levels that are related by the total source power constraint. This precoder tends to become optimum as the number of antennas at the source becomes large.

3.5.3.2 Sub-optimum Source and Relay precoder during 2nd slot

Let us now consider problem (3.36) in which the source and relay precoders are optimized during the second time slot under an individual power constraint. Let us also introduce into (3.36) the SVD of the joint channel $[\mathbf{H}_2 \ \mathbf{H}_0] = \mathbf{U} \text{diag}(\lambda) \mathbf{V}^H$ and perform the following change of variable:

$$\mathbf{R} \triangleq \mathbf{V}^H \mathbf{R}_{SR}^{(2)} \mathbf{V} \quad (3.47)$$

The problem (3.36) can be rewritten as:

$$\begin{aligned} \hat{R}_{B,2} &= \max_{\mathbf{R} \succeq 0} \log_2 \left| \mathbf{I}_{N_D} + \text{diag}(\boldsymbol{\lambda}) \mathbf{R} (\text{diag}(\boldsymbol{\lambda}))^H \right| \\ \text{s.t. } &\text{tr}(\mathbf{D}_S \mathbf{V} \mathbf{R} \mathbf{V}^H \mathbf{D}_S^H) \leq P_S \quad ; \quad \text{tr}(\mathbf{D}_R \mathbf{V} \mathbf{R} \mathbf{V}^H \mathbf{D}_R^H) \leq P_R \end{aligned} \quad (3.48)$$

As in appendix B of [WZH05], we arbitrarily enforce a diagonal structure $\mathbf{R} = \text{diag}(\mathbf{p}_2)$. This turns the matrix optimization problem (3.48) into the following vector optimization:

$$\begin{aligned} \max_{\mathbf{p}_2 \geq 0} & \sum_{i=1}^L \log_2(1 + \lambda_i^2 p_{2,i}) \\ \text{s.t. } & \mathbf{a}^T \mathbf{p}_2 \leq P_S \quad ; \quad \mathbf{b}^T \mathbf{p}_2 \leq P_R \end{aligned} \quad (3.49)$$

where L is the number of non-zero eigenvalues of $[\mathbf{H}_2 \ \mathbf{H}_0]$ and the vectors \mathbf{a} and \mathbf{b} have their components defined by $\mathbf{a}_i \triangleq \sum_{j=1}^{N_S} |V_{N_R+j,i}|^2$ and $\mathbf{b}_i \triangleq \sum_{j=1}^{N_R} |V_{j,i}|^2$. The problem (3.49) can then be solved at a much lower computational cost than the original one.

Note that if we replace the individual power constraints in (3.48) by a sum-power constraint

$$\text{tr}(\mathbf{R}_{SR}^{(2)}) = \text{tr}(\mathbf{R}) \leq P_S + P_R \quad (3.50)$$

then Hadamard determinant inequality can be applied as in [T99]:

$$\left| \mathbf{I}_{N_D} + \text{diag}(\boldsymbol{\lambda}) \mathbf{R} (\text{diag}(\boldsymbol{\lambda}))^H \right| \leq \prod_{i=1}^{N_D} (1 + \lambda_i^2 R_{i,i}) \quad (3.51)$$

with equality if \mathbf{R} is diagonal. Since the constraint (3.50) depends only on the diagonal components of \mathbf{R} , it is clear that the optimum \mathbf{R} in (3.48) is diagonal. In other words, under a total transmit power constraint at the source and relay, precoding with the right singular vectors of the joint channel $[\mathbf{H}_2 \ \mathbf{H}_0]$ is optimal. However, under individual power constraints, this structure is in general suboptimum because the individual power constraints depend on both diagonal and non-diagonal components of \mathbf{R} .

3.6 Simulation Results

In this section, simulation results are presented for the TDD MIMO relay channel. The upper and lower bounds on capacity are evaluated and the sub-optimality of the precoder optimization procedures in §3.5.3.1 and §3.5.3.2 is discussed.

The simulations below assume $N_S = 4$ antennas at the source and $N_R = N_D = 2$ antennas at the relay and destination. Such an antenna configuration is well suited to a cellular downlink scenario. The source and relay are subject to the same power constraint

$P_S = P_R = 1$. The MIMO channel on the S-D, S-R and R-D links is modeled by i.i.d. complex Gaussian components of respective variance γ_0 , γ_1 and γ_2 . Therefore, under the above-defined power constraints, γ_0 , γ_1 and γ_2 represent the average SNR on the S-D, S-R and R-D links. The destination is far from both S and R, with $\gamma_0 = \gamma_2 = 0\text{dB}$. The average SNR γ_1 on the S-R link is varied from 0dB to 30dB.

On Figure 10, the average achievable rates of the partial and full DF strategies are plotted. For comparison purpose, the average capacity of the S-D link is also plotted with (solid line) or without (dashed line) CSIT. In this last case, the source covariance matrix that maximizes the ergodic capacity is [T99] $\mathbf{R}_S = (P_S / N_S) \mathbf{I}_{N_S}$. The average capacity gain provided by CSIT can be decomposed into an array gain and a waterfilling gain [TV05]. Here since $N_S = 2N_D$ there is a 3dB array gain plus a large waterfilling gain because γ_0 is low. It can be observed that single-hop transmission always outperforms non-cooperative DF. This is obvious from inequality (2.20) and the fact that the capacity of the S-D link is larger than that of the R-D link. Both partial and full DF achieve a large rate increase over non-cooperative DF and single-hop transmission, and are less than 0.5 bit (per channel use) away from the cut-set bound for $\gamma_1 \geq 15\text{dB}$. Partial DF outperforms FDF only at low γ_1 , when the rate becomes limited by the S-R link capacity. It can be argued that under the above simulation assumptions, the comparison with non-cooperative DF and single-hop transmission is unfair since the total transmit power is larger for cooperative protocols during the second slot. Therefore, we also plot (dotted curve) the achievable rate of the FDF strategy when the sum-power is constrained to remain lower than P_S during the second slot. The figure shows that even in this case FDF outperforms non-cooperative protocols at $\gamma_1 \geq 5\text{dB}$.

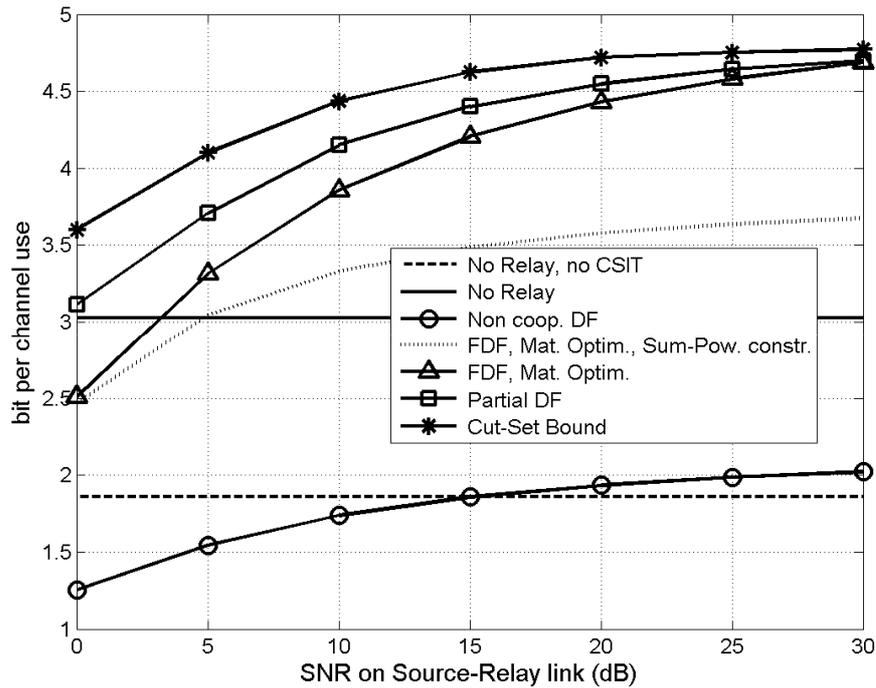


Figure 10: Upper and lower bounds on TDD MIMO relaying channel capacity with Source and Relay precoder optimization in a 4x2x2 antenna configuration.

On Figure 11, various precoder optimization strategies for FDF are compared. The highest rate is achieved by matrix convex optimization of the source and relay precoders during both slots, using one of the procedures described in §3.4. As stated in §3.5.3.1, the vector optimization of the source precoder becomes optimum when N_S is large, but here it can be observed that the incurred loss is already small at $N_S = 4$ (only 0.1 bit). An additional rate penalty occurs at high γ_1 when the sub-optimum source and relay precoder structure of §3.5.3.2 is enforced during the second slot. Overall, the degradation due to sub-optimum precoding is lower than 0.5 bit over the whole SNR range. Finally, the dashed and dotted curves illustrate the large rate loss when precoders are not optimized during the first slot (i.e. $\mathbf{R}_S^{(1)} = (P_S / N_S) \mathbf{I}_{N_S}$) or during both slots ($\mathbf{R}_S^{(1)} = \mathbf{R}_S^{(2)} = (P_S / N_S) \mathbf{I}_{N_S}$, $\mathbf{R}_R^{(2)} = (P_R / N_R) \mathbf{I}_{N_R}$).

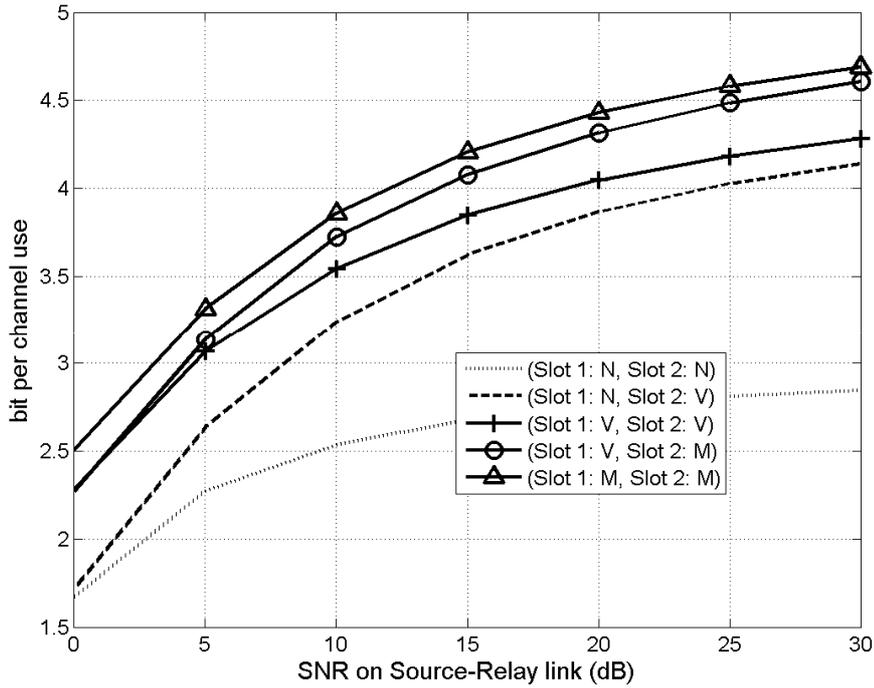


Figure 11: Optimum vs. Sub-optimum precoder optimization for full DF strategy in a 4x2x2 antenna configuration. (Legend: N=No optimization, V=Vector Optimization, M=Matrix Optimization)

On Figure 12, the optimum power fraction \hat{P}_1 and the optimum time-sharing \hat{t} obtained by the sub-optimum procedure of §3.5.3.1 are plotted vs. γ_1 . As expected, when the capacity of the S-R link becomes much larger than that of the other links, most of the source transmit power during the first slot is assigned to the eigenmodes of \mathbf{H}_0 , and most of the time resource is assigned to the second slot in order to maximize R_{DF} .

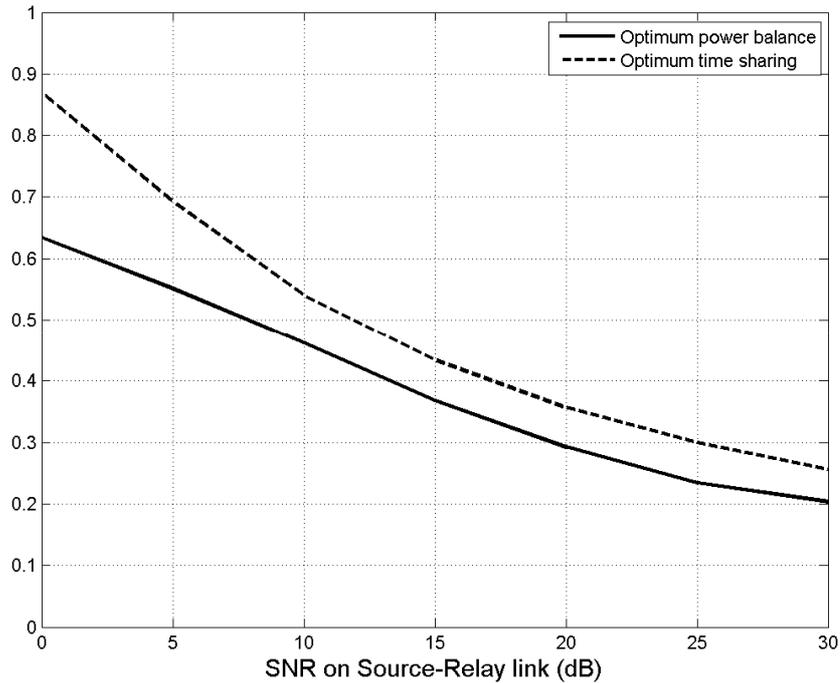


Figure 12: Sub-optimum source precoder during the first slot for full DF strategy: variation of the time-sharing and power balancing vs. SNR on the Source-Relay link

3.7 Conclusions

We presented a generic methodology to maximize capacity upper and lower bounds on the MIMO relay channel with full CSI in the full-duplex and TDD relaying cases. The optimum source and relay transmit covariance matrices and TDD time-sharing parameter can be derived by convex optimization, and the gap between the achievable rate and the capacity can be quantified for various DF strategies. In particular, we verified that for realistic antenna configurations and SNR ranges this gap can actually be quite small. Our optimization procedure illustrates the application of several mathematical tools borrowed from convex optimization theory, nonlinear programming as well as from complex matrix differentiation to the practical problem of precoding for the MIMO relay channel. As explained in Chapter 5, the bounds derived here are not directly applicable to a real system but can be easily modified to provide a good estimate of the throughput envelope that can be achieved. Moreover, they can also serve as a benchmark when studying the effect of practical implementation impairments such as imperfect CSI.

Chapter 4: Distributed Compression for Cooperative MIMO

4.1 Introduction and overview of our contribution

In §2.1.4 we have reviewed coding strategies for the scalar relay channel. We have shown by an analysis of achievable rate expressions and verified by simulations the fact that the CF strategy outperforms DF and LR when the capacity on the link from the relay to the destination becomes large-enough. We have also motivated our focus on TDD relaying due to its simpler practical implementation and we have mentioned the reference [HZ05] in which the achievable rate of a partial CF (PCF) strategy on the Gaussian scalar TDD relay channel is derived. In the PCF strategy of [HZ05], a first message is transmitted during the first slot and is WZ-compressed by the relay while a second message is transmitted by the source directly to the destination during the second slot. In the next section §4.2 of this chapter, we essentially extend these achievable rates to the MIMO case. Our contribution is the following:

- Distributed Gaussian vector compression [GDV04][GDV06] is applied to the specific case of CF relaying (§4.2.1) and the effect of this compression on the achievable rate of CF is analyzed (§4.2.2).
- The achievable rate is maximized with full CSI. A closed-form expression of the optimum WZ coding rates is derived (§4.2.3.1) which differs from the rate-distortion trade-off in [GDV06].
- It is shown in §4.2.3.2 that during the second slot of the TDD protocol an optimum decoding order exists for the messages transmitted by S and R, and this can be used to simplify the optimization of the source and relay covariance matrices. Finally, an iterative procedure is proposed (§4.2.3.3) which jointly optimizes the compression, the transmit covariances and the time resource allocation.
- Simulations are performed in both uplink and downlink cellular scenarios which illustrate the phenomena mentioned above and a comparison with other capacity bounds is performed (§4.2.4).

The first two bullet points are addressed in [SMV07], while the third bullet is included in [SMVC08c].

In §4.2, Compress-and-Forward relaying is applied to TDD in-band MIMO relaying with 3 nodes. In §4.3 we try to apply the Compress-and-Forward strategy to multiple parallel out-of-band multi-antenna relays or equivalently to a coordinated MIMO cellular network, as introduced in §2.1.6. In [C08], the PCF strategy is applied to a set of multiple parallel in-band single-antenna relays. However, the achievable rate region (equation (3.43) in [C08]) is untractable, partly due to the large number of constraints required to define the MAC achievable rate region formed by the multiple parallel relays transmitting to the destination. Del Coso therefore proposes to relax the problem by considering only the MAC sum-rate constraint. As explained in §2.1.6, the problem is simpler for a coordinated cellular network because it can be assumed that the backhaul links have fixed capacity. In [MF07], quantization is proposed for a coordinated cellular uplink with limited backhaul, and it is shown that large spectral efficiency gains can be achieved in spite of the basic source coding at each BS which cannot exploit the correlation between the received signals. Our work in §4.3 relies mainly on the more advanced distributed coding scheme of [DW04] and [SSS08], in which the signals received at each BS are partially decoded and compressed before being processed by a Central Processing Unit. Our contribution essentially consists in a computation of achievable rates when the cooperative BSs are equipped with multiple antennas:

- In §4.3.1 we instantiate the results in [SSS08] for a Gaussian multiple-antenna setting with Gaussian codebook, and formulate the achievable rate as an optimization problem with respect to compression noise covariance matrices. In particular, we show that the problem is simplified under a backhaul sum-rate constraint.
- In §4.3.2 we show that the compression noise distribution which maximizes the achievable rate in the 3-node case corresponds to the Transform Coding approach introduced in [GDV04][GDV06] with the WZ coding rate allocation of [SMV07] that is derived in §4.2.3.1.
- Achievable rates are derived in the case of N-parallel relays in §4.3.3 and an achievable rate region in the multi-user case is derived in §4.3.4.
- Finally, these theoretical results are illustrated by simulations under either per-link or total backhaul rate constraints in §4.3.5.

The above four bullets are the subject of several publications [CS08a][CS08b] and submissions [CS08c][CS08d].

4.2 Compress-and-Forward strategy for the 3-node MIMO relay channel

4.2.1 Signal Model and Achievable Rates

4.2.1.1 Signal Model and Coding strategy

The coding strategy assumed here is an extension to the multiple antenna case of the partial CF described in [HZ05]. The Source, Relay and Destination are equipped with respectively N_S , N_R and N_D antennas. The channels from S to D, S to R and R to D are all assumed static and are denoted respectively \mathbf{H}_0 , \mathbf{H}_1 and \mathbf{H}_2 . During the first slot of duration t , S transmits a first message ω_0 at rate R_0 using the codeword³ $\mathbf{x}_S^{(1)}(\omega_0)$. We assume that $\mathbf{x}_S^{(1)}$ is a proper [NM93] complex Gaussian vector $\mathbf{x}_S^{(1)} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_{N_S}, \mathbf{R}_S^{(1)})$, although this may not be the optimum distribution, as pointed out in section VI.B of [SSS08]. The received signals at R and D read:

$$\mathbf{y}_R^{(1)} = \mathbf{H}_1 \mathbf{x}_S^{(1)} + \mathbf{n}_R^{(1)} \quad (4.1)$$

$$\mathbf{y}_D^{(1)} = \mathbf{H}_0 \mathbf{x}_S^{(1)} + \mathbf{n}_D^{(1)} \quad (4.2)$$

where the superscript (i) indicates that the signals are transmitted or received during the i th slot. The noise at R and D is also assumed (proper) complex white Gaussian of respective covariance $\Sigma_R = \sigma^2 \mathbf{I}_{N_R}$ and $\Sigma_D = \sigma^2 \mathbf{I}_{N_D}$. The compression at R, which will be detailed in §4.2.1.2, consists in mapping $\mathbf{y}_R^{(1)}$ onto the index ω_1 assuming side information $\mathbf{y}_D^{(1)}$ at the destination. The relay forwards ω_1 to D during the second slot of duration $1-t$ by transmitting $\mathbf{x}_R^{(2)}(\omega_1)$ at rate R_1 , while S transmits a new message ω_2 at rate R_2 by means of the codeword $\mathbf{x}_S^{(2)}(\omega_2)$. We denote by P_S and P_R the maximum transmit power at S and R during both slots, i.e. $\text{tr}(\mathbf{R}_S^{(i)}) \leq P_S$ $i=1,2$ and $\text{tr}(\mathbf{R}_R^{(2)}) \leq P_R$. The destination successively decodes $\hat{\omega}_1$ and $\hat{\omega}_2$ (the decoding order will be discussed in §4.2.3.2). The decompression at D can be modeled as a mapping of

³ A complete information-theoretic treatment would not define a codeword as length- N_S random vector but as a length- n sequence of length- N_S vectors drawn i.i.d. from a certain distribution. In this thesis, we are interested in finding the distribution that maximizes the achievable rate, hence the simplified notation used. We rely on [HZ05][GDV06] and [SSS08] for the rate achievability proofs, established for n going to infinity.

$(\mathbf{y}_D^{(1)}, \hat{\omega}_1)$ onto $\hat{\mathbf{y}}_R^{(1)}$, the reconstructed relay observation. After decompression, D decodes $\hat{\omega}_0$ from $(\mathbf{y}_D^{(1)}, \hat{\mathbf{y}}_R^{(1)})$. Finally the achievable rate between S and D is :

$$R_{CF} = tR_0 + (1-t)R_2 \quad (4.3)$$

The rate R_{CF} needs to be maximized w.r.t. the time-sharing parameter, the distribution of the codewords, the compression mapping and the decoding order at the destination. Because the received signals at R and D are Gaussian, recently published results can be applied to design the compression mapping at the relay, as explained in the next section.

4.2.1.2 Vector Source Coding at the Relay

In [HZ05], Høst-Madsen et al. show that the achievable rate of partial CF depends on the variance of a ‘‘compression noise’’ which differs from the quadratic distortion in general. This compression noise was first introduced by Wyner, who derived in [W78] the rate-distortion function for source coding with Gaussian source and Gaussian side information. In [GDV04][GDV06], Gastpar et al. investigate distributed source coding and introduce the Distributed KLT and the Conditional KLT. They show that the rate-distortion coding of a Gaussian vector source with side information at the decoder is achieved by first applying a CKLT and then separately performing WZ coding of each CKLT output at a different rate. The results presented in this section II.B can be viewed as a special case of [GDV06]. A compression noise vector is defined, and its relationship with distortion is clarified. The main difference between our work and [GDV06] will arise in §4.2.3.1 where it will be shown that the code which maximizes the CF achievable rate is not the same code that minimizes the quadratic distortion, which is the design criterion in [GDV06].

Let first define the Conditional Karhunen-Loeve Transform (CKLT) in the CF relaying case. Given the knowledge of $\mathbf{y}_D^{(1)}$, the vector $\mathbf{y}_R^{(1)}$ is Gaussian distributed (see e.g. 4.12.1 in [MS00]) of mean:

$$E[\mathbf{y}_R^{(1)} | \mathbf{y}_D^{(1)}] = \mathbf{R}_{R,D}^{(1)} (\mathbf{R}_D^{(1)})^{-1} \mathbf{y}_D^{(1)} \quad (4.4)$$

and covariance matrix denoted by $\mathbf{R}_{R|D}^{(1)}$ and equal to:

$$\mathbf{R}_{R|D}^{(1)} = \mathbf{R}_R^{(1)} - \mathbf{R}_{R,D}^{(1)} (\mathbf{R}_D^{(1)})^{-1} \mathbf{R}_{D,R}^{(1)} \quad (4.5)$$

where $\mathbf{R}_R^{(1)}$ and $\mathbf{R}_D^{(1)}$ denote the covariance of the received signal at R and D, while $\mathbf{R}_{R,D}^{(1)}$ denotes the cross-correlation between the relay and destination observations:

$$\mathbf{R}_{R,D}^{(1)} \triangleq E \left[\mathbf{y}_R^{(1)} \left(\mathbf{y}_D^{(1)} \right)^H \right] = \mathbf{H}_1 \mathbf{R}_S^{(1)} \mathbf{H}_0^H \triangleq \mathbf{R}_{D,R}^H \quad (4.6)$$

$$\mathbf{R}_D^{(1)} = \mathbf{H}_0 \mathbf{R}_S^{(1)} \mathbf{H}_0^H + \sigma^2 \mathbf{I}_{N_D} \quad (4.7)$$

$$\mathbf{R}_R^{(1)} = \mathbf{H}_1 \mathbf{R}_S^{(1)} \mathbf{H}_1^H + \sigma^2 \mathbf{I}_{N_R} \quad (4.8)$$

After some matrix manipulations, equations (4.5)-(4.8) yield:

$$\mathbf{R}_{R|D}^{(1)} = \mathbf{H}_1 \left(\mathbf{R}_S^{(1)} \right)^{1/2} \left(\mathbf{I}_{N_S} + \mathbf{H}_0 \mathbf{R}_S^{(1)} \mathbf{H}_0^H \right)^{-1} \left(\mathbf{R}_S^{(1)} \right)^{1/2} \mathbf{H}_1^H + \sigma^2 \mathbf{I}_{N_R} \quad (4.9)$$

The CKLT can now be defined as the matrix \mathbf{U}^H such that:

$$\mathbf{R}_{R|D}^{(1)} = \mathbf{U} \text{diag}(\mathbf{s}) \mathbf{U}^H \quad (4.10)$$

The columns of \mathbf{U} are the eigenvectors of the conditional covariance matrix, and the vector \mathbf{s} contains the associated eigenvalues. Note that from (4.9), it is clear that the matrix $\mathbf{R}_{R|D}^{(1)} - \sigma^2 \mathbf{I}_{N_R}$ is positive semi-definite and therefore:

$$s_i \geq \sigma^2 \quad \forall i \in \{1, 2, \dots, N_R\} \quad (4.11)$$

It is shown in Appendix C.2 that the rate-distortion coding of vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the decoder can be modeled by the following relationship:

$$\hat{\mathbf{y}}_R^{(1)} = \mathbf{U} \mathbf{A} \mathbf{U}^H \mathbf{y}_R^{(1)} + \mathbf{U} \mathbf{A} \boldsymbol{\psi} + \mathbf{U} \mathbf{K} \mathbf{y}_D^{(1)} \quad (4.12)$$

where

- $\boldsymbol{\psi}$ is a vector of i.i.d. components ψ_i of variance η_i called the compression noise:

$$\psi_i \sim \mathbb{C}\mathcal{N}(0, \eta_i) \quad (4.13)$$

- $\mathbf{A} \triangleq \text{diag}(\mathbf{a})$ with $a_i \triangleq s_i / (s_i + \eta_i)$ (4.14)

- $\mathbf{K} \triangleq (\mathbf{I} - \mathbf{A}) \mathbf{U}^H \mathbf{R}_{R,D}^{(1)} \left(\mathbf{R}_D^{(1)} \right)^{-1}$ (4.15)

The relationship (4.12) is illustrated on Figure 13:

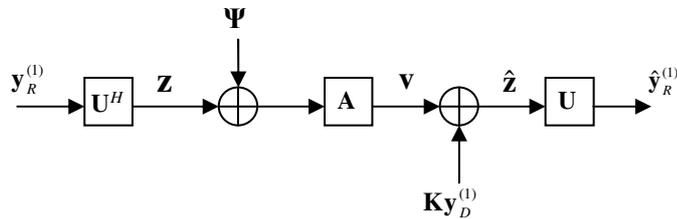


Figure 13: Source coding of Gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the decoder.

The coding scheme consists in applying a CKLT to each source output $\mathbf{y}_R^{(1)}$, followed by independent WZ encoding of each CKLT output sequence z_i , where $\mathbf{z} \triangleq \mathbf{U}^H \mathbf{y}_R^{(1)}$. The destination then performs WZ decoding to obtain \hat{z}_i and applies the inverse CKLT to $\hat{\mathbf{z}}$. Defining r_i as the WZ coding rate for the i th CKLT output z_i , the relationship between this rate and the compression noise is:

$$r_i = \log(1 + s_i / \eta_i) \quad (4.16)$$

Moreover, defining d_i as the squared distortion on the i th component z_i of the transformed vector \mathbf{z} , the following relationship holds between the compression noise and the distortion:

$$d_i = s_i \eta_i / (s_i + \eta_i) \quad (4.17)$$

Inserting (4.17) into (4.16) yields:

$$r_i = \log(s_i / d_i) \quad (4.18)$$

It is shown in Appendix C that the total quadratic distortion is

$$\delta \triangleq E \left[\|\hat{\mathbf{y}}_R^{(1)} - \mathbf{y}_R^{(1)}\|^2 \right] = \sum_{i=1}^{N_R} d_i \quad (4.19)$$

The above results are summarized in the following proposition:

Proposition 4.1:

The source coding of vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the destination can be performed at a rate ρ by applying a Conditional Karhunen-Loève Transform (CKLT) at

the relay followed by separate Wyner-Ziv (WZ) coding of each CKLT output sequence at rate $r_i \geq 0$ such that:

$$\sum_{i=1}^{N_R} r_i \leq \rho \quad (4.20)$$

The following relationship holds between the WZ coding rate, the distortion and the compression noise for each component of the CKLT-transformed source:

$$r_i = \log(1 + s_i / \eta_i) = \log(s_i / d_i) \quad (4.21)$$

In particular, the rate-distortion trade-off $\rho(\delta)$ is achieved by the reverse-waterfilling algorithm:

$$d_i = \begin{cases} \lambda & \text{if } 0 \leq \lambda < s_i \\ s_i & \text{otherwise} \end{cases} \text{ with } \lambda \text{ s.t. } \sum_{i=1}^{N_{\text{eig}}} d_i = \delta \quad (4.22)$$

Proof: See Appendix C.2.

Note that in the scalar case $N_R = 1$, $d_1 = \delta$ and Proposition 4.1 boils down to the rate-distortion coding results of [W78]. The relationship $\eta_i = s_i d_i / (s_i - d_i)$ shows that the compression noise on a component of the transformed vector is approximately equal to the distortion when the latter is small. However, when $r_i \rightarrow 0$ (i.e. no bit is allocated to represent the i th eigenmode) or equivalently $d_i \rightarrow s_i$, then $\eta_i \rightarrow +\infty$. The last part of Proposition 4.1 is the generalization of a well-known result (see e.g. §13.3.3 in [CT91]): the rate-distortion function of parallel Gaussian source is obtained by allocating the distortion according to a reverse-waterfilling algorithm on the eigenvalues of the source covariance matrix. Here, equation (4.22) corresponds to the same algorithm applied to the eigenvalues of the conditional covariance matrix of $\mathbf{y}_R^{(1)}$ given $\mathbf{y}_D^{(1)}$. The term reverse-waterfilling comes from the fact that it can be implemented by progressively increasing the distortion level on every component of the transformed vector until the total distortion δ is reached, under the constraint that the distortion on the i th component cannot exceed the i th eigenvalue of the conditional covariance matrix. As shown by equation (4.5), the latter reduces to the covariance when the product $\mathbf{R}_{R,D}^{(1)} (\mathbf{R}_D^{(1)})^{-1} \mathbf{R}_{D,R}^{(1)}$ goes to zero. This happens for instance when the SNR at the destination is small, then whatever the value of the cross-correlation $\mathbf{R}_{R,D}^{(1)}$, the product $\mathbf{R}_{R,D}^{(1)} (\mathbf{R}_D^{(1)})^{-1} \mathbf{R}_{D,R}^{(1)}$ tends towards zero. This also happens for instance if $\mathbf{R}_S^{(1)} = (P_S / N_S) \mathbf{I}_{N_S}$ and the channels \mathbf{H}_0 and \mathbf{H}_1 are orthogonal.

In these two cases, the CKLT degenerates into a KLT and the side information at the destination cannot be exploited to reduce the rate required at the relay to encode its observation.

4.2.2 Impact of the Compression on the CF achievable rate

In this section, we assume that the compression at the relay is performed as described in §4.2.1.2, and we apply Proposition 4.1 in order to clarify the relationship between the achievable rate of the CF strategy described in §4.2.1.1 and the compression noise defined in §4.2.1.2. The expression of the CF achievable rate is given by the following proposition:

Proposition 4.2:

The partial CF coding strategy defined in §4.2.1.1 achieves a rate given by the solution of the following optimization problem:

$$R_{CF} = \max_{\substack{t \in [0,1], \boldsymbol{\eta} > \mathbf{0} \\ \mathbf{R}_S^{(1)} \succeq \mathbf{0}, \mathbf{R}_S^{(2)} \succeq \mathbf{0}, \mathbf{R}_R^{(2)} \succeq \mathbf{0} \\ \text{tr}(\mathbf{R}_S^{(1)}) \leq P_S, \text{tr}(\mathbf{R}_S^{(2)}) \leq P_S, \text{tr}(\mathbf{R}_R^{(2)}) \leq P_R}} tR_0 + (1-t)R_2 \quad (4.23)$$

where

1. R_0 is equal to the capacity of a virtual MIMO channel:

$$R_0 = \mathcal{C}(\mathbf{R}_S^{(1)}, \tilde{\boldsymbol{\Sigma}}^{-1/2} \tilde{\mathbf{H}}) \quad (4.24)$$

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} \triangleq \begin{bmatrix} \boldsymbol{\Sigma}_D & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_R + \mathbf{U} \text{diag}(\boldsymbol{\eta}) \mathbf{U}^H \end{bmatrix}$$

2. $\boldsymbol{\eta}$ is subject to the following inequality constraint:

$$t \sum_{i=1}^{N_R} \log(1 + s_i / \eta_i) \leq (1-t) R_1 \quad (4.25)$$

3. (R_1, R_2) is constrained to lie within the capacity region of the MAC from (S, R) to D .

Proof:

From the definition of the coding strategy in §4.2.1.1:

$$R_0 = I(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)}, \hat{\mathbf{y}}_R^{(1)}) \quad (4.26)$$

where $\hat{\mathbf{y}}_R^{(1)}$ is given by equation (4.12):

$$\hat{\mathbf{y}}_R^{(1)} = \mathbf{U} \mathbf{A} \mathbf{U}^H \mathbf{y}_R^{(1)} + \mathbf{U} \mathbf{A} \boldsymbol{\psi} + \mathbf{U} \mathbf{K} \mathbf{y}_D^{(1)}$$

Removing $\mathbf{U}\mathbf{K}\mathbf{y}_D^{(1)}$ from $\hat{\mathbf{y}}_R^{(1)}$ does not affect the mutual information (4.26), and from the property (6.68) multiplying the remainder by $\mathbf{U}\mathbf{A}^{-1}\mathbf{U}^H$ does not affect (4.26) either. Therefore,

$$I(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)}, \hat{\mathbf{y}}_R^{(1)}) = I(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)}, \tilde{\mathbf{y}}_R^{(1)}) \quad (4.27)$$

where $\tilde{\mathbf{y}}_R^{(1)}$ is defined as:

$$\tilde{\mathbf{y}}_R^{(1)} \triangleq \mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\Psi} \quad (4.28)$$

From (4.26)-(4.28), the rate R_0 can therefore be expressed as:

$$R_0 = \mathcal{C}(\mathbf{R}_S^{(1)}, \tilde{\boldsymbol{\Sigma}}^{-1/2} \tilde{\mathbf{H}}) \quad (4.29)$$

Finally, equation (4.25) is a direct application of Proposition 4.1. This concludes the proof.

It can be noticed that in the single-antenna case Proposition 4.2 boils down to the CF achievable rates derived in [HZ05]. From the above equations, it can also be observed that R_{CF} meets the MIMO capacity bound when $t \rightarrow 1$ and $\boldsymbol{\eta} \rightarrow \mathbf{0}$. Unfortunately, equation (4.25) shows that achieving these two conditions simultaneously would require $R_1 \rightarrow +\infty$. The trade-off which maximizes R_{CF} is investigated in the next section.

4.2.3 Maximizing the Achievable Rate

The maximization of the CF achievable rate as formulated in Proposition 4.2 is a highly non-trivial non-convex optimization problem. Rather than directly addressing the joint optimization with respect to all the parameters, we start by identifying sub-problems which can be solved by means of convex optimization techniques. First, in §4.2.3.1 a closed-form expression for the optimum WZ coding rates is derived assuming all other parameters constant, then in §4.2.3.2 and §4.2.3.3 the joint optimization of the WZ coding rates and the transmit covariance matrices at the source and relay is addressed.

4.2.3.1 Optimization of the Compression at the relay

Let assume a fixed transmit covariance at S and R during the first and second slot, and a fixed time-sharing parameter t . As a result, R_1 and R_2 are fixed and only R_0 needs to be maximized with respect to the compression noise $\boldsymbol{\eta}$. We show in Appendix C.3 that R_0 can be decomposed as follows:

$$\begin{aligned}
R_0 &= R_{0,d} + R_{0,r} \\
R_{0,d} &\triangleq \mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_0) \quad \text{and} \quad R_{0,r} \triangleq \sum_{i=1}^{N_R} \log \left(\frac{s_i + \eta_i}{\sigma^2 + \eta_i} \right)
\end{aligned} \tag{4.30}$$

The term $R_{0,d}$ is the mutual information between the signal transmitted by the source and the destination observation during the first slot. The term $R_{0,r}$ is the additional information brought by the compressed relay observation reconstructed at the destination. Note that only $R_{0,r}$ depends on $\boldsymbol{\eta}$ and shall be optimized. A contribution to $R_{0,r}$ can be associated to each component of the transformed relay observation. Equation (4.30) shows that this contribution is maximized and tends to $\log(s_i / \sigma^2)$ when the compression noise variance η_i is negligible compared to the thermal noise variance σ^2 . As expected, this contribution vanishes when $\eta_i \rightarrow +\infty$, reflecting the fact that a highly distorted signal component cannot convey information. Also note that the components for which equality holds in (4.11) do not contribute to $R_{0,r}$, although they affect the total distortion \mathcal{D} . The optimization of $R_{0,r}$ with respect to $\boldsymbol{\eta}$ yields the following proposition:

Proposition 4.3:

The Wyner-Ziv coding rates r_i and compression noise η_i which maximize the achievable rate of the partial CF relaying strategy are such that:

$$r_i = \left[\mu + \log \left((s_i - \sigma^2) / \sigma^2 \right) \right]^+ \tag{4.31}$$

$$\eta_i = s_i / (2^{r_i} - 1) \tag{4.32}$$

where μ is a constant such that the following constraint is satisfied:

$$t \sum_{i=1}^{N_R} r_i \leq (1-t) R_1 \tag{4.33}$$

Proof: See Appendix C.3.

In equation (4.31), the ratio $(s_i - \sigma^2) / \sigma^2$ can be interpreted as a useful signal to thermal noise ratio per component, since equality occurs in (4.11) when the i th component contains only noise. From (4.30), it is clear that an eigenmode with a large $(s_i - \sigma^2) / \sigma^2$ ratio will be a large contributor to $R_{0,r}$ and ultimately to the CF achievable rate R_{CF} , provided that the compression noise variance η_i on this eigenmode is low-enough. The

term $\log((s_i - \sigma^2) / \sigma^2)$ in (4.31) can thus be viewed as a rate penalty for the eigenmodes which have a lower potential contribution to R_{CF} . The penalty tends to $-\infty$ when $s_i \rightarrow \sigma^2$, and in this case the corresponding CKLT output will not be encoded. It is interesting to compare equation (4.31) with the reverse-waterfilling algorithm of proposition 4.1 that is obtained when minimizing the total distortion under a rate constraint. In reverse waterfilling, the algorithm tries to spread the total distortion δ uniformly, under the constraint $d_i \leq s_i$. Such a strategy leads to $r_i = \log(s_i / d)$ with d a constant corresponding to the distortion on the eigenmodes which are finally encoded (i.e. $r_i > 0$). If the average number of bits available per vector $R_1(1-t)/t$ is large enough, even the eigenmodes such that $s_i = \sigma^2$ will be encoded, although they cannot convey any information, which shows that reverse-waterfilling is sub-optimum in our problem. The CF achievable rate loss when adopting the reverse-waterfilling algorithm instead of that of proposition 4.3 is illustrated by simulations in [SMV07].

Having optimized the compression at the relay, we now address another sub-problem which is the optimization of source and relay transmit covariance during the second slot, before dealing with the whole problem (4.23).

4.2.3.2 Optimization of the Source and Relay Precoders

Before introducing a joint optimization procedure, the expression of the rates R_1 and R_2 shall be clarified. The partial CF strategy described in §4.2.1.1 considers the simultaneous transmission of the two independent messages ω_1 and ω_2 simultaneously to D during the second slot. Fixed $\mathbf{R}_S^{(2)}$ and $\mathbf{R}_R^{(2)}$, the MAC achievable rate region is defined by the following pentagon (see e.g. ch. 10 in [TV05]):

$$R_1 \leq \mathcal{C}(\mathbf{R}_R^{(2)}, \mathbf{H}_2 / \sigma) \quad (4.34)$$

$$R_2 \leq \mathcal{C}(\mathbf{R}_S^{(2)}, \mathbf{H}_0 / \sigma) \quad (4.35)$$

$$R_1 + R_2 \leq \mathcal{C}\left(\begin{bmatrix} \mathbf{R}_S^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_R^{(2)} \end{bmatrix}, [\mathbf{H}_0 \quad \mathbf{H}_2] / \sigma\right) \quad (4.36)$$

Equality in (4.35) is reached when D decodes ω_1 first, then removes the contribution of $\mathbf{x}_R^{(2)}$ from $\mathbf{y}_D^{(2)}$ so that ω_2 is decoded interference-free. Equality in (4.34) is reached by decoding ω_2 first, and the sum-rate side (4.36) is reached by time-sharing between the two decoding orders. Solving (4.23) under the constraints (4.34)-(4.36) seems a very

difficult non-convex optimization problem. We start by simplifying it by imposing a decoding order at the destination:

Proposition 4.4:

In the single antenna case, the achievable rate of partial Compress-and-Forward is maximum when the relayed message is decoded first.

Proof: See Appendix C.4

In other words, R_{CF} varies along the sum-rate side of the MAC achievable rate pentagon and is optimum only at the corner of this pentagon corresponding to the decoding of ω_1 prior to ω_2 . This is contrary to a statement in a footnote of [HZ05]. Note that we were able to prove this proposition only in the single antenna case, but in the rest of the thesis we conjecture that the proposition remains valid in the multiple antenna case.

Having fixed the decoding order at D, we now turn to the optimization of R_{CF} w.r.t. $\mathbf{R}_S^{(2)}$ and $\mathbf{R}_R^{(2)}$ for a given t . From (4.31) and (4.33), it is clear that given $\mathbf{R}_S^{(1)}$ (i.e. given \mathbf{s}), increasing R_1 allows to increase μ and therefore to increase each rate r_i or equivalently from (4.32) to reduce $\boldsymbol{\eta}$ component-wise. This increases the contribution of R_0 to R_{CF} . In the “full CF” case where only the Relay is allowed to transmit during the 2nd slot (i.e. $R_2 = 0$, $\mathbf{R}_S^{(2)} = \mathbf{0}$), then maximizing R_1 w.r.t. $\mathbf{R}_R^{(2)}$ is a mandatory preliminary step in the maximization of R_{CF} , and this maximization consists in transmit power waterfilling on the eigenmodes of \mathbf{H}_2 as in [T99]. However in general for partial CF, a larger R_{CF} is achieved by letting $R_2 > 0$. From Proposition 4.4, the decoding of ω_2 is interference-free. Therefore, R_2 can be maximized w.r.t. $\mathbf{R}_S^{(2)}$ by waterfilling on the eigenmodes of \mathbf{H}_0 . Likewise, given $\mathbf{R}_S^{(2)}$, R_1 can be maximized w.r.t. $\mathbf{R}_R^{(2)}$ by waterfilling on the eigenmodes of $(\boldsymbol{\Sigma}_D + \mathbf{H}_0 \mathbf{R}_S^{(2)} \mathbf{H}_0^H)^{-1/2} \mathbf{H}_2$. In the following, we decide to optimize R_2 and R_1 successively in the order described above. The intuition behind this simplification is the following: CF is known to outperform DF only when the SNR is much larger on the R-D link than on the S-D link, therefore we can restrict the optimization of CF to this scenario. Thus, we assume that the signal to noise-plus-interference ratio when decoding ω_1 is high. In this case, waterfilling amounts to equal power allocation over all the eigenmodes of \mathbf{H}_2 , and the impact of $\mathbf{R}_S^{(2)}$ on the optimization of R_1 becomes negligible.

4.2.3.3 Iterative Procedure for joint optimization

Having derived in the previous section an optimization procedure for the Source and Relay precoders during the second slot, we now assume that R_1 , R_2 and t are fixed and address the maximization of R_0 w.r.t. the Source precoder during the first slot and the compression at the relay, before addressing the whole problem (4.23).

The following two sub-problems can be identified:

- Fixed \mathbf{U} and $\boldsymbol{\eta}$, the optimization of R_0 w.r.t. $\mathbf{R}_S^{(1)}$ in (4.29) is obtained by transmit power waterfilling on the eigenmodes of $\tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\mathbf{H}}$ as in [T99].
- Fixed $\mathbf{R}_S^{(1)}$, the CKLT is determined and the compression noise $\boldsymbol{\eta}$ which maximizes R_0 is given by Proposition 4.3.

This suggests the use of the non-linear Gauss-Seidel algorithm (see Appendix B.4) to jointly optimize $\mathbf{R}_S^{(1)}$ and $\boldsymbol{\eta}$. Integrating this algorithm with the source and relay precoder optimization of §4.2.3.2, we now propose the following procedure for solving (4.23):

Iterative Procedure for maximizing R_{CF} :

1. Maximize R_2 w.r.t. $\mathbf{R}_S^{(2)}$ by transmit power waterfilling on the eigenmodes of \mathbf{H}_0
2. Maximize R_1 w.r.t. $\mathbf{R}_R^{(2)}$ by transmit power waterfilling on the eigenmodes of $(\boldsymbol{\Sigma}_D + \mathbf{H}_0\mathbf{R}_S^{(2)}\mathbf{H}_0^H)^{-1/2}\mathbf{H}_2$
3. Outer loop: Maximize R_{CF} w.r.t. $t \in [0;1]$
 Inner-loop: Maximize R_0 w.r.t. $\mathbf{R}_S^{(1)}$ and $\boldsymbol{\eta}$ by iterating between steps a. and b. ($\tilde{\boldsymbol{\Sigma}}$ is initialized to $\sigma^2\mathbf{I}_{N_R+N_D}$):
 - (a) Maximize R_{CF} w.r.t. $\mathbf{R}_S^{(1)}$ by transmit power waterfilling on the eigenmodes of $\tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\mathbf{H}}$.
 - (b) Maximize R_{CF} w.r.t. $\boldsymbol{\eta}$ by optimum Wyner-Ziv coding rates allocation (Proposition 4.3).

The outer loop is a one-dimensional maximization of $R_{CF} = tR_0 + (1-t)R_2$ with respect to $t \in [0;1]$. It can practically be performed by uniformly quantizing this interval, and the quantization step determines the time accuracy of the final solution. Unfortunately the convergence of the Gauss-Seidel algorithm in the inner-loop cannot be guaranteed. Indeed, the conditions for convergence given in §2.7 of [B99] cannot be

verified, because \mathbf{U} depends on $\mathbf{R}_S^{(1)}$ and therefore the assumption that $\tilde{\Sigma}$ is fixed in the optimization step 3.a is an approximation. Thus the joint optimization procedure cannot be claimed optimal. The convergence of the inner-loop will be assessed in the next section for realistic SNR values, and the sub-optimality of the whole procedure will be evaluated by a comparison with the cut-set bound and the achievable rate of other relaying strategies.

4.2.4 Simulation Results

In this section we analyze by simulations the achievable rate performance of partial CF, and perform comparisons with other relaying strategies for the TDD MIMO relay channel with full CSI. In the comparison, we will consider non-cooperative DF and partial DF strategies for which achievable rates are computed in Chapter 3, but not linear relaying for reasons explained in §2.1.4.3.

4.2.4.1 Downlink Mobile Relaying scenario

We consider a downlink TDD mobile relaying scenario where a Base Station (S) equipped with $N_S = 4$ antennas per sector transmits to a dual antenna Mobile Station (D) which is assisted by another dual antenna MS (R) in its neighborhood. The average SNRs on the S-D, S-R and R-D links are denoted respectively γ_0 , γ_1 and γ_2 . In the simulations we assume that $\gamma_0 = \gamma_1$ varies between 0dB and 20dB and γ_2 is fixed to 30dB (i.e. R and D are close to each other). The MIMO fading on each link is modeled by i.i.d. complex Gaussian components. On Figure 14, the average CF achievable rate obtained from the iterative procedure of §4.2.3.3 is plotted (solid black curves) for both the partial and full CF strategies. The dashed curves represent the MIMO channel capacity of the S-D link, the achievable rate with the non-cooperative DF and with the partial DF strategies. On the figure, the three curves associated to full CF, partial DF and the S-D link capacity (i.e. no relaying) almost overlap and it can be observed that only the partial CF strategy yields a significant rate increase over the S-D link capacity, whereas non-cooperative DF relaying achieves a rate much lower than the S-D link capacity. Further simulation results (not plotted here) show that at $\gamma_2 = 30\text{dB}$, partial DF starts to outperform partial CF only when γ_1 is at least 10dB higher than γ_0 , i.e. when the capacity on the S-R link becomes much higher than on the S-D link. On the figure, the dotted curves represent the cut-set

bound and the capacity of the Virtual MIMO channel from S to (R,D), obtained by waterfilling the source transmit power on the eigenmodes of $\tilde{\mathbf{H}}$. The cut-set bound is much lower than the VMIMO bound, which shows that although γ_2 is high, the capacity of the R-D link cannot be assumed infinite. Partial CF achieves a rate very close to the cut-set bound and is therefore almost capacity-achieving, although in our simulations we stopped the inner-loop after only two iterations. Thus, partial CF seems well-suited to this downlink mobile cooperation scenario, and improving the optimization procedure could in this case only yield a marginal rate increase.

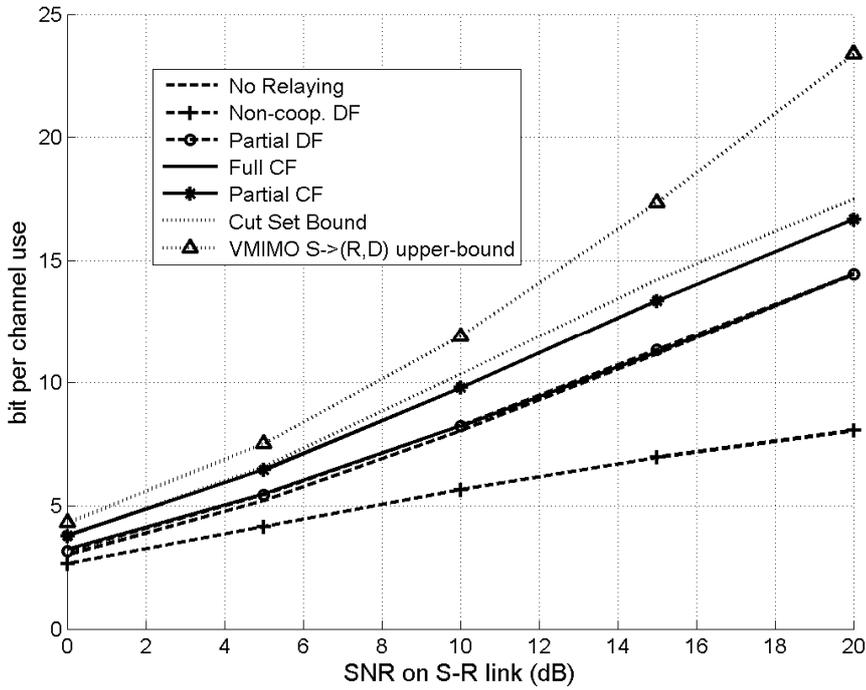


Figure 14: Average CF achievable rate and comparison with DF strategies and capacity upper-bounds

On Figure 15 we compare the achievable rate obtained by applying the complete optimization procedure of §4.2.3.3 with a simpler optimization in which we fix $\mathbf{R}_S^{(1)} = \mathbf{R}_S^{(2)} = (P_S / N_S) \mathbf{I}_{N_S}$ and $\mathbf{R}_R^{(2)} = (P_R / N_R) \mathbf{I}_{N_R}$ and only optimize t and $\boldsymbol{\eta}$. As explained in [TV05], optimizing the source transmit covariance increases the S-D link capacity by a 3dB power gain (because $N_S = 2N_D$) plus a waterfilling gain that becomes negligible at high SNR. The SNR gain is only about 1dB for the cut-set bound and for partial CF. Intuitively, this smaller gain is justified by the fact that the number of transmit

antennas is equal to the number of receive antennas on the 4×4 virtual MIMO channel $\tilde{\mathbf{H}}$ and therefore the rate R_0 , which is the main component of R_{CF} , does not benefit from any power gain.

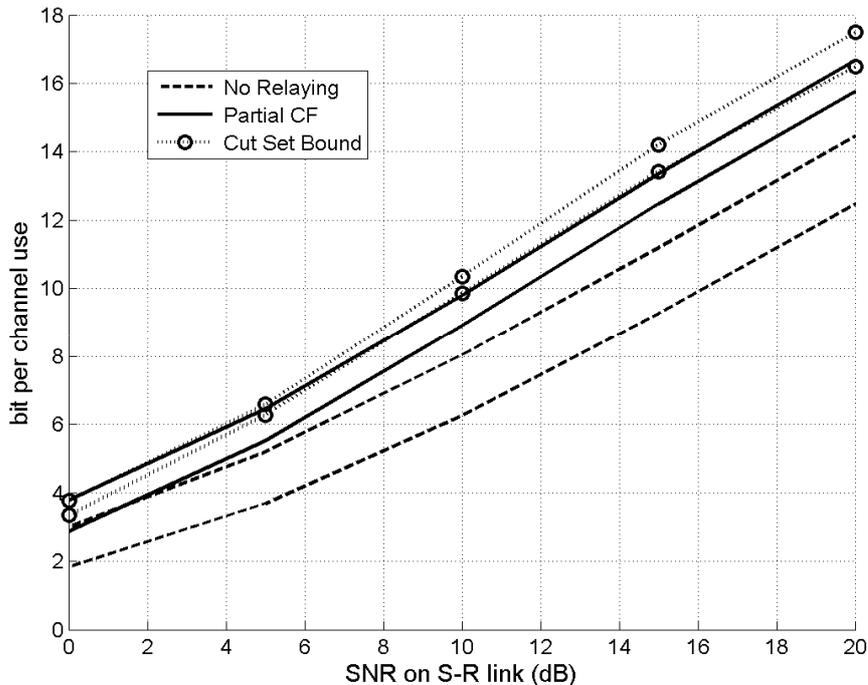


Figure 15: Capacity bounds with and without transmit covariance optimization.

4.2.4.2 Uplink Fixed Relaying scenario

We now consider a cellular uplink scenario with fixed relaying in which the mobile is equipped with $N_s = 2$ antennas, and the BS and RS are equipped with $N_R = N_D = 4$ antennas. The same i.i.d. Rayleigh MIMO channel model is assumed as in the previous section. We assume a high SNR between the RS and BS $\gamma_2 = 20\text{dB}$. The MS is far from the BS ($\gamma_0 = 0\text{dB}$) and we plot on capacity bounds as a function of the SNR γ_1 on the S-R link. Note that we do not optimize transmit covariance matrices as this optimization is not expected to provide a significant gain in this antenna configuration.

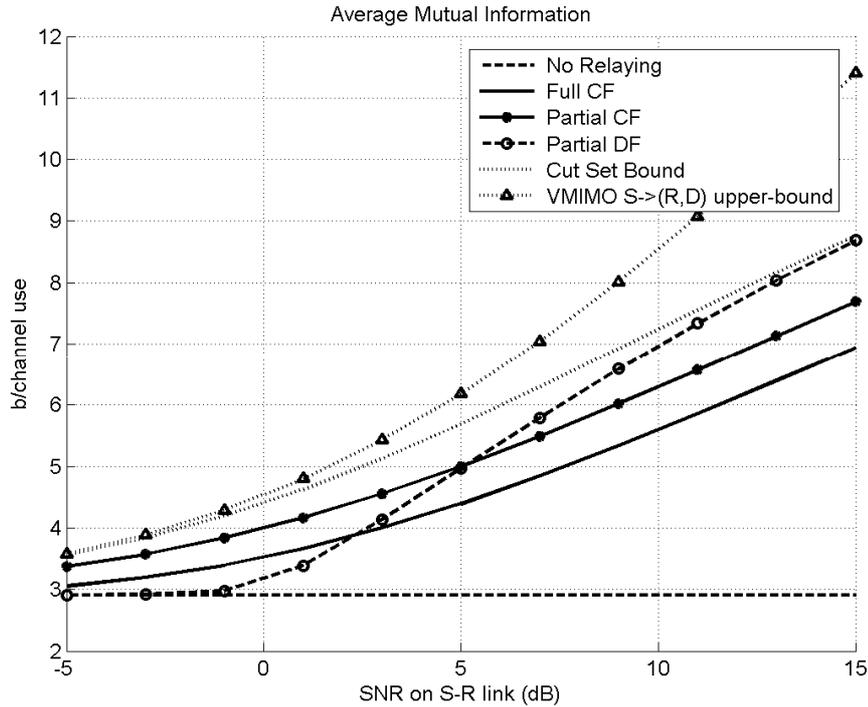


Figure 16: Comparison of CF with other capacity bounds in a cellular uplink scenario with fixed relaying ($\gamma_0 = 0\text{dB}$, $\gamma_2 = 20\text{dB}$, $N_S = 2$, $N_R = N_D = 4$).

It can be observed on the figure that partial CF outperforms partial DF at low SNR on the S-R link. However, as the MS gets closer to the RS the partial DF strategy shall be selected. Implementing a link adaptation between these two strategies allows to always operate at less than 1 bit per channel use from the cut-set bound, and hence from the capacity. Note that full CF never yields a significant gain compared to direct link or partial DF. This shows that compressing the relay observation demands too much capacity on the R-D link to create a VAA.

4.2.5 Conclusions

We have derived in this section achievable rates for CF in the MIMO case with full CSI. These rates can be computed by an iterative procedure which optimizes the WZ compression at the relay, the transmit covariance matrices at the Source and Relay and TDD time-sharing parameter. Simulations show that partial CF outperforms partial DF and is almost capacity-achieving in scenarios where the capacity of the R-D link is sufficiently high. This condition may occur either in cellular uplink where a fixed relay

benefits from a strong link to the BS or in downlink mobile relaying scenarios provided the cooperating mobiles are very close to each other. In this second case, the optimization of the source transmit covariance matrix makes sense especially if the number of transmit antennas at the BS is larger than the total number of antennas at the relay and destination. However, in in-band relaying scenarios, the achievable rates of CF remain far away from the capacity of a Virtual MIMO channel, because the number of bits to compress the observation at the relay is large and conveying these samples to the destination takes a significant part of the TDD frame. In the next sections of this chapter we will consider distributed compression for out-of-band uplink relaying and we will see that CF becomes especially attractive if the backhaul rate is large.

4.3 Extension to multiple out-of-band relays or cooperative base stations

4.3.1 Distributed Compression strategy

4.3.1.1 Coding Strategies

We consider a source S equipped with N_S antennas transmitting data to $M + 1$ base stations BS_0, BS_1, \dots, BS_M each equipped with N_i $i = 0, \dots, M$ antennas. Without loss of generality, we assume that BS_0 is the CPU and that each BS is connected to BS_0 through a backhaul of fixed rate ρ_i $i = 1, \dots, M$. Note that the user-to-BS assignment is assumed given and its optimization is out-of-the-scope of this thesis (see e.g. [KM07]). The received signal at each BS is

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_S + \mathbf{n}_i \quad i = 0, \dots, M \quad (4.37)$$

where \mathbf{H}_i is the (static) channel matrix from S to each BS and $\mathbf{n}_i \sim \sigma^2 \mathbf{I}_{N_i}$ for $i = 0, \dots, M$.

In Theorem 1 of [SSS08], a distributed coding strategy is considered in which the source maps a message ω_{CF} onto $\mathbf{x}_S(\omega_{CF})$ that is decoded at the CPU. Each BS, upon receiving \mathbf{y}_i maps it onto an auxiliary variable $\hat{\mathbf{y}}_i(s_i)$ where s_i is the Wyner-Ziv bin index and ω_{CF} is decoded once all the decoded WZ bin indices \hat{s}_i are available. Though Theorem 1 applies to discrete channels, the extension to Gaussian channels and

continuous variables holds and is given in section VI of [SSS08]. This scheme achieves the following rate:

Proposition 4.5 (Theorem 1 of [SSS08] - Distributed Compress-and-Forward with Joint Decoding (DCF-JD))

The following rate is achievable by distributed compression:

$$R_{DCF,JD} = \max_{p(\mathbf{x}_S, \{\mathbf{y}_i\}_0^M, \{\hat{\mathbf{y}}_i\}_1^M)} I(\mathbf{x}_S; \mathbf{y}_0, \{\hat{\mathbf{y}}_i\}_1^M) \quad (4.38)$$

$$s.t. \quad \forall \mathcal{G} \subseteq \{1, \dots, M\} : I(\hat{\mathbf{y}}_{\mathcal{G}}; \mathbf{y}_{\mathcal{G}} | \hat{\mathbf{y}}_{\{1, \dots, M\} \setminus \mathcal{G}}) < \sum_{i \in \mathcal{G}} \rho_i \quad (4.39)$$

$$and \quad p(\mathbf{x}_S, \{\mathbf{y}_i\}_0^M, \{\hat{\mathbf{y}}_i\}_1^M) = p(\mathbf{x}_S) p(\{\mathbf{y}_i\}_0^M | \mathbf{x}_S) \prod_{i=1}^M p(\hat{\mathbf{y}}_i | \mathbf{y}_i) \quad (4.40)$$

The coding strategy used to prove Theorem 1 makes use of random coding and binning. Decoding is based on strong typicality. The error analysis relies on the generalized Markov Lemma [HK80] to show that the decoding error probability goes down to zero when the codeword length goes to infinity. This lemma requires that the following Markov chain holds:

$$\{\mathbf{x}_S, \hat{\mathbf{y}}_{\{1, \dots, M\} \setminus i}, \mathbf{y}_{\{0, \dots, M\} \setminus i}\} \rightarrow \mathbf{y}_i \rightarrow \hat{\mathbf{y}}_i \quad (4.41)$$

to guarantee that the auxiliary variables are jointly strongly typical with the BS observations and with the source with a probability close to one as the codeword length goes to infinity, which guarantees an error-free decoding. This Markov chain relationship is reflected by (4.40). Note that (4.41) is an extension of (6.69) to source coding with multiple sources.

Note that prior to [SSS08] other strategies were proposed for distributed WZ coding which can be applied to the BS cooperation case as well. In [DW04], parallel Wyner-Ziv coding with sequential decoding is performed. A permutation π of the set $\{1, \dots, M\}$ gives the decoding order at the CPU, the observation $\mathbf{y}_{\pi(1)}$ is WZ-encoded assuming side information \mathbf{y}_0 at the CPU. The observation $\mathbf{y}_{\pi(2)}$ is WZ-encoded assuming side information $(\mathbf{y}_0, \hat{\mathbf{y}}_{\pi(1)})$ and so on. With this strategy, the achievable rate is given by Proposition 4.6:

Proposition 4.6 ([DW04] - Distributed Compress-and-Forward with sequential decoding (DCF-SD))

$$R_{DCF,SD} = \max_{\pi, p(\mathbf{x}_S, \{\hat{\mathbf{y}}_i\}_1^M)} I(\mathbf{x}_S; \mathbf{y}_0, \{\hat{\mathbf{y}}_i\}_1^M) \quad (4.42)$$

$$\text{s.t.} \quad I(\mathbf{y}_{\pi(i)}; \hat{\mathbf{y}}_{\pi(i)} \mid \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}) < \rho_j \quad (4.43)$$

where the distribution is such that the following Markov chain holds:

$$\left(\mathbf{x}_S, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}, \{\mathbf{y}_j\}_{\{0, \dots, M\} \setminus \pi(i)} \right) \rightarrow \mathbf{y}_{\pi(i)} \rightarrow \hat{\mathbf{y}}_{\pi(i)} \quad (4.44)$$

In this chapter we will compute achievable rates for the two coding/decoding strategies of Propositions 4.5 and 4.6. Note that contrary to [W78][GDV06] and [DW04], [SSS08] does not compute a rate-distortion performance but only performs an error analysis where the ultimate goal is to decode the source data. Moreover at least two possible improvements of the achievable rate are proposed in [SSS08]:

- Proposition 4.5 assumes that all bin indices are decoded correctly at the destination. However, this constraint is actually not really required and an error in the decoding of auxiliary variable $\hat{\mathbf{y}}_i$ at the CPU could be tolerated as long as the original source message ω_{CF} is correctly decoded. This leads to Corollary 1 in [SSS08].
- Partial decoding at each BS can be considered. In this more complex coding strategy, the source splits the data into $M + 1$ messages $(\omega_0, \omega_1, \dots, \omega_M, \omega_{CF})$ where ω_i is decoded by the i th BS and ω_{CF} is only decoded at the CPU. Each BS compresses its observation given its decoded message and forwards both the decoded message and the compressed observation to the CPU.

Unfortunately, for time reasons we were not able in this thesis to address these two potential improvements. We will therefore restrict ourselves in the following to the achievable rates of Proposition 4.5 and 4.6 and attempt to compute them numerically in the Gaussian MIMO case.

4.3.1.2 Upper and Lower Bounds on the Gaussian MIMO channel

In this section, we start by defining a codebook which makes the optimization problems of Propositions 4.5 and 4.6 numerically tractable. Thereafter, we instantiate Propositions 4.5 and 4.6 for this codebook which provides us achievable rate expressions. Finally, we present some simple upper-bounds on the achievable rate which will be useful as benchmarks in our simulations.

4.3.1.2.1 Codebook definition

We constrain \mathbf{x}_s and $\hat{\mathbf{y}}_i$ for $i=1,\dots,M$ to be proper [NM93] complex Gaussian vectors. Note that as mentioned in [SSS08] this distribution may not be optimal. From (4.37) and Lemma 3 in [NM93], the vector \mathbf{y}_i is also proper for $i=0,\dots,M$. Moreover, we assume that \mathbf{y}_i and $\hat{\mathbf{y}}_i$ are jointly proper such that there exists a constant matrix \mathbf{M}_i and a vector $\boldsymbol{\varphi}_i$ such that:

$$\hat{\mathbf{y}}_i = \mathbf{M}_i \mathbf{y}_i + \boldsymbol{\varphi}_i \quad i=1,\dots,M \quad (4.45)$$

and we further assume that \mathbf{M} is non-singular, which means that we do not add useless dimensions to the auxiliary random variable $\hat{\mathbf{y}}_i$. Let define the following vectors:

$$\hat{\mathbf{y}}_i' \triangleq \mathbf{y}_i + \boldsymbol{\varphi}_i' \quad (4.46)$$

$$\boldsymbol{\varphi}_i' \triangleq \mathbf{M}_i^{-1} \boldsymbol{\varphi}_i \quad (4.47)$$

We have:

$$\begin{aligned} I(\mathbf{y}_i; \hat{\mathbf{y}}_i) &\triangleq H(\hat{\mathbf{y}}_i) - H(\hat{\mathbf{y}}_i | \mathbf{y}_i) \\ &= H(\mathbf{M}_i \mathbf{y}_i + \boldsymbol{\varphi}_i) - H(\boldsymbol{\varphi}_i) \\ &= H(\mathbf{M}_i \mathbf{y}_i') - H(\mathbf{M}_i \boldsymbol{\varphi}_i') \\ &\stackrel{(a)}{=} H(\hat{\mathbf{y}}_i') - H(\boldsymbol{\varphi}_i') \\ &= I(\mathbf{y}_i; \hat{\mathbf{y}}_i') \end{aligned} \quad (4.48)$$

where (a) comes from (6.68). Likewise, it can be checked that:

$$I(\mathbf{x}; \mathbf{y}_0, \hat{\mathbf{y}}_i) = I(\mathbf{x}; \mathbf{y}_0, \hat{\mathbf{y}}_i') \quad (4.49)$$

Renaming $\hat{\mathbf{y}}_i'$ and $\boldsymbol{\varphi}_i'$ as respectively $\hat{\mathbf{y}}_i$ and $\boldsymbol{\varphi}_i$, it can be concluded from (4.48) and (4.49) that when the compression codebook is constrained to be proper complex Gaussian distributed and when the auxiliary variables are constrained to be jointly Gaussian with

the observations then the achievable rates of Propositions 4.5 and 4.6 can be computed with the following codebook:

$$\hat{\mathbf{y}}_i = \mathbf{y}_i + \boldsymbol{\Phi}_i \quad (4.50)$$

where the auxilliary variables are equal to the observations plus an additive compression noise. Notice that the Gaussian codebook distribution that we selected satisfies the Markov chain relationship (4.44).

4.3.1.2.2 Achievable rate expressions

With the codebook defined in §4.3.1.2.1, we now compute achievable rates from Propositions 4.5 and 4.6:

Proposition 4.7

The following rate is achievable by distributed WZ coding with sequential decoding:

$$R_{DWZ,SD} = \max_{\pi(\cdot)} \mathcal{C} \left(\mathbf{R}_S, \begin{pmatrix} (1/\sigma) \mathbf{H}_0 & \mathbf{0} \\ \mathbf{0} & \text{diag} \left\{ (\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_i)^{-1/2} \mathbf{H}_i \right\}_1^M \end{pmatrix} \right) \quad (4.51)$$

$$s.t. \quad \log \left| \mathbf{I}_{N_S} + \boldsymbol{\Phi}_{\pi(i)}^{-1} \mathbf{R}_{\mathbf{y}_{\pi(i)} | \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}} \right| \leq \rho_{\pi(i)} \quad i = 1, \dots, M \quad (4.52)$$

where

$$\mathbf{R}_{\mathbf{y}_{\pi(i)} | \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}} = \mathbf{H}_{\pi(i)} \left(\mathbf{I} + \mathbf{R}_S \left(\mathbf{H}_0^H \mathbf{H}_0 + \sum_{j=1}^{i-1} \mathbf{H}_{\pi(j)}^H (\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_{\pi(j)})^{-1} \mathbf{H}_{\pi(j)} \right) \right)^{-1} \mathbf{R}_S \mathbf{H}_{\pi(i)}^H + \sigma^2 \mathbf{I} \quad (4.53)$$

$$\boldsymbol{\Phi}_{\pi(i)} = \mathbf{U}_{\pi(i)} \text{diag} \left(\boldsymbol{\eta}_{\pi(i)} \right) \mathbf{U}_{\pi(i)}^H \quad (4.54)$$

with $\mathbf{U}_{\pi(i)}$ defined by

$$\mathbf{R}_{\mathbf{y}_{\pi(i)} | \mathbf{y}_0, \hat{\mathbf{y}}_{\pi(1)}, \dots, \hat{\mathbf{y}}_{\pi(i-1)}} = \mathbf{U}_{\pi(i)} \text{diag} \left(\mathbf{s}_{\pi(i)} \right) \mathbf{U}_{\pi(i)}^H \quad (4.55)$$

and $\{\boldsymbol{\eta}_i\}_1^M$ is computed for a given permutation π by applying the optimum WZ coding rate allocation of Proposition 4.3.

Proof:

From Appendix C.2 the compression noise covariance at the i th step of the sequential WZ coding is given by $\boldsymbol{\Phi}_{\pi(i)} = \mathbf{U}_{\pi(i)} \boldsymbol{\eta}_{\pi(i)} \mathbf{U}_{\pi(i)}^H$. Therefore, the achievable rate (4.42) can be

computed by forming a virtual MIMO channel where the observation at the i th remote BS is subject to both thermal noise and compression noise, which gives equation (4.51). The expression (4.53) of the conditional covariance matrix is obtained from the definition of the conditional covariance (4.5) after a few matrix manipulations (including the application of the inversion lemma).

Note that in Proposition 4.7 we do not attempt to optimize the source covariance \mathbf{R}_S . Apart from the fact that such an optimization would further complicate the problem, we also expect that it will not significantly increase the achievable rate in the uplink of a coordinated network as the total number of receive antennas becomes much larger than N_S . One can also notice that the complexity of the optimization problem (4.51) for a given permutation π is quite low because the achievable rate (4.51) is just a sum of contributions given by equation (4.30) and obtained by applying M times the algorithm of Proposition 4.3. However, the exhaustive search of the best permutation makes the overall complexity proportional to $M!$. This makes the problem prohibitively complex when the number of cooperating BSs becomes larger than 4-5. We expect that the same complexity issue will occur with the coding strategy of Proposition 4.5, since the number of constraints in (4.39) is proportional to the number of subsets \mathcal{G} in $\{1, \dots, M\}$. In an attempt to simplify the problem, we replace the individual backhaul link rate constraints by a backhaul sum-rate constraint. This leads to the following achievable rate:

Proposition 4.8

Under a backhaul sum-rate constraint ρ , the following rate is achievable by DCF with either Joint or Sequential Decoding:

$$R_{DWZ} = \max_{\{\Phi_i\}_1^M \succeq 0} \mathcal{C} \left(\mathbf{R}_S, \begin{pmatrix} (1/\sigma)\mathbf{H}_0 & \mathbf{0} \\ \mathbf{0} & \text{diag}\left\{(\sigma^2\mathbf{I} + \Phi_i)^{-1/2} \mathbf{H}_i\right\}_1^M \end{pmatrix} \right) \quad (4.56)$$

$$\text{s.t.} \quad \log \left| \mathbf{I} + \text{diag}(\Phi_1^{-1}, \dots, \Phi_M^{-1}) \mathbf{R}_{\{y_i\}_1^M | y_0} \right| \leq \rho \quad (4.57)$$

where $\mathbf{R}_{\{y_i\}_1^M | y_0}$ is the conditional covariance of all the remote BS observations given the observation of the decoding BS and is given by:

$$\mathbf{R}_{\{y_i\}_1^M | y_0} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \left(\mathbf{I} + \frac{\mathbf{R}_S}{\sigma^2} \mathbf{H}_0^H \mathbf{H}_0 \right)^{-1} \mathbf{R}_S \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix}^H + \sigma^2 \mathbf{I} \quad (4.58)$$

Proof: See Appendix C.5

A somewhat surprising corollary of Proposition 4.8 is that with DCF-SD under a backhaul sum-rate constraint, the order in which compressed observations are decoded does not matter, which removes the need for a search over all possible permutations. Unfortunately, the optimum compression noise in problem (4.56) may not satisfy (4.54) and therefore the DCF-SD under a backhaul sum-rate constraint may not be implementable by transform coding contrary to DCF-SD under a per-link constraint.

In this section we have expressed achievable rates with Gaussian codebooks for the DCF-JD and DCF-SD coding strategies under either per-link rate or sum-rate constraint for the backhaul. Under a per-link backhaul rate constraint, the achievable rates of the DCF-SD strategy are easily obtained for small sets of cooperative BSs by applying the same Transform Coding approach as for the 3-node relay channel in a sequential way. However, the complexity of the optimization of the achievable rate for DCF-SD and DCF-JD becomes prohibitive as the size of the set of cooperating BSs grows. Under a backhaul sum-rate constraint, this complexity issue can be alleviated and an achievable rate for both strategies can be obtained by solving a constrained optimization problem over the set of compression noise covariance matrices. Before attempting to solve this optimization problem, we derive some upper-bounds which provide some insight on the achievable rate performance of the DCF strategies.

4.3.1.3 Upper-bounds on achievable rate

A first obvious upper-bound is the capacity of the virtual MIMO channel without any compression noise. It can be viewed as the achievable rate limit when the backhaul rate grows to infinity:

Upper-Bound 1 (Virtual MIMO Channel capacity):

$$R_{DCF} \leq \mathcal{C} \left(\mathbf{R}_S, \frac{1}{\sigma} \begin{bmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \right) \quad (4.59)$$

A second upper-bound is obtained by applying the cut-set bound to the network including the backhaul links. Considering the MAC cut between $\{\mathbf{S}, \{\text{BS}_i\}_1^M\}$ and BS_0 readily gives the following upper-bound:

Upper-Bound 2 (Cut-Set Bound):

$$R_{DCF} \leq \mathcal{C} \left(\mathbf{R}_S; \frac{\mathbf{H}_0}{\sigma} \right) + \rho \quad (4.60)$$

The access rate of the distributed CF scheme cannot exceed the access rate to the CPU plus the backhaul sum-rate. This second upper-bound turns out to be easily achieved in the multi-user case when the coordinated network is backhaul-limited, i.e. when the backhaul rate is not large enough to let the distributed CF scheme achieve a significant fraction of the VMIMO channel capacity (that is obtained with infinite backhaul rate), as illustrated in [CS08d].

4.3.2 Two cooperative BS case

When $M = 1$, the backhaul sum-rate and per-link rate constraints are equivalent and Proposition 4.8 simplifies as:

$$R_{DWZ} = \max_{\Phi \succeq 0} \mathcal{C} \left(\mathbf{R}_S, \begin{pmatrix} (1/\sigma)\mathbf{H}_0 & \mathbf{0} \\ \mathbf{0} & (\sigma^2\mathbf{I} + \Phi)^{-1/2} \mathbf{H}_1 \end{pmatrix} \right) \quad (4.61)$$

$$\text{s.t.} \quad \log \left| \mathbf{I} + \Phi^{-1} \mathbf{R}_{y_1|y_0} \right| \leq \rho \quad (4.62)$$

$$\mathbf{R}_{y_1|y_0} = \mathbf{H}_1 \left(\mathbf{I} + \frac{\mathbf{R}_S}{\sigma^2} \mathbf{H}_0^H \mathbf{H}_0 \right)^{-1} \mathbf{R}_S \mathbf{H}_1^H + \sigma^2 \mathbf{I} \quad (4.63)$$

This problem is solved in Theorem 2 of [CS08d] to which we refer for the details of the solution. In order to turn (4.61) into a concave function, we introduced the change of variable $\mathbf{A} \triangleq \Phi^{-1}$. Unfortunately, the constraint (4.62) turns out to be concave in \mathbf{A} and therefore the problem is not convex. The proof of Theorem 2 in [CS08d] is made of two steps: *i*) solving the KKT conditions provides a necessary (but not sufficient) condition for the compression noise to be optimum *ii*) checking the general sufficiency condition

(see Appendix B.3) for optimality guarantees that the solution given by the KKT is the optimum. Finally, the solution is

$$\hat{\Phi} = \mathbf{U} \text{diag}(\boldsymbol{\alpha}^{-1}) \mathbf{U}^H \quad (4.64)$$

with

$$\alpha_i = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma^2} - \frac{1}{s_i} \right) - \frac{1}{\sigma^2} \right]^+ \quad (4.65)$$

and λ is such that the backhaul rate constraint (4.62) is satisfied. Equations (4.64)-(4.65) correspond exactly to the optimum WZ coding rate allocation of Proposition 4.3. Theorem 2 in [CS08d] is therefore a proof that the transform coding approach and the Wyner-Ziv coding rate allocation of Proposition 4.3 correspond to the optimum Gaussian compression codebook.

4.3.3 Multiple (more than two) cooperative BS case

The Problem of proposition 4.8 in the general case $M > 1$ is addressed in [CS08c]. As explained in the previous section, this problem is not convex. Moreover, the KKT conditions do not seem to lead to a closed-form solution. We therefore decided to solve the dual problem (see Appendix B.3), and claim in [CS08c][CS08d] that the duality gap is zero. Introducing the same change of variable $\mathbf{A}_i \triangleq \Phi_i^{-1}$ as in the previous section, the Lagrangian can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_M, \lambda) = & -\log \left| \mathbf{I} + \frac{\mathbf{R}_S}{\sigma^2} \mathbf{H}_0^H \mathbf{H}_0 + \mathbf{R}_S \sum_{i=1}^M \mathbf{H}_i^H (\mathbf{A}_i \sigma^2 + \mathbf{I})^{-1} \mathbf{A}_i \mathbf{H}_i \right| \\ & + \lambda \left(\log \left| \mathbf{I} + \text{diag}(\{\mathbf{A}_i\}_1^M) \mathbf{R}_{\{y_i\}_1^M | y_0} \right| - \rho \right) \end{aligned} \quad (4.66)$$

The minimization of the Lagrangian can be carried on by the Gauss-Seidel algorithm (see Appendix B.4) because it is defined on the domain \mathcal{S}_+^M which is the Cartesian product of the domain of each variable. The Gauss-Seidel algorithm is especially well suited to the minimization of (4.66) because the problem

$$\hat{\mathbf{A}}_i = \arg \max_{\mathbf{A}_i \succeq 0} \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_{i-1}, \mathbf{A}_i, \mathbf{A}_{i+1}, \dots, \mathbf{A}_M) \quad (4.67)$$

has a closed-form solution which is obtained from KKT conditions and is given by Theorem 2 of [CS08c]:

$$\hat{\mathbf{A}}_i = \mathbf{U}_i \text{diag}(\mathbf{a}) \mathbf{U}_i^H \quad (4.68)$$

where \mathbf{U}_i is defined from the eigen-decomposition

$$\mathbf{R}_{y_i|y_0,\hat{y}_{\{1,\dots,M\}^c}} = \mathbf{U}_i \text{diag}(\mathbf{s}_i) \mathbf{U}_i^H \quad (4.69)$$

and

$$\alpha_i = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma^2} - \frac{1}{s_{i,j}} \right) - \frac{1}{\sigma^2} \right]^+ \quad j = 1, \dots, N_i \quad (4.70)$$

The maximization of the dual function can be performed by a subgradient method (see Appendix B.3)

4.3.4 Multi-user case: sum-rate and achievable rate region

In the previous section §4.3.3, we have computed an achievable rate for distributed compression of a single user equipped with multiple antennas by several cooperating BSs also equipped with multiple antennas. However, in state-of-the-art systems, even a single BS is capable to spatially multiplex several users, and from a system spectral efficiency standpoint, it would be a waste of resource to dedicate several BSs to a single user. It is therefore of more practical interest to investigate the achievable rate of distributed compression in the multi-user case. In [CS08b][CS08c][CS08d], we derive an achievable rate region for distributed compression by solving a weighted sum-rate optimization problem. The procedure is quite similar to that of the previous section, except that the minimization of the Lagrangian cannot be solved as easily as in the single-user case. For brevity, we do not further discuss this derivation and refer the reader to the above-mentioned papers for details. Instead, we restrict our multi-user simulations to the sum-rate side of the achievable rate region, as it already allows interesting observations to be made. Notice that the sum-rate can be easily obtained as a special case of the single-user achievable rate, defining $\mathbf{H}_i \triangleq [\mathbf{H}_{i,1} \quad \mathbf{H}_{i,2} \quad \dots \quad \mathbf{H}_{i,K}]$ $i = 0 \dots M$ where K is the number of users and $\mathbf{H}_{i,j}$ is the channel matrix from the j th user to the i th BS.

4.3.5 Simulation results

4.3.5.1 Diamond out-of-band uplink relaying configuration

We want to study by simulations how the achievable rate of distributed CF for out-of-band uplink relaying is affected by the backhaul rate in the single-user and multi-user cases. We want to look at the influence of several parameters, such as the deployment

topology, the number of antennas, the number of users and the type of compression used. For this purpose, we consider the “diamond” topology of Figure 17, in which the MS is only connected to two parallel RSs but not to the CPU. For simplicity, the same average SNR γ is assumed on the two “access links” MS- RS_1 and MS- RS_2 links (note that in this case the macro-diversity gain is maximum). This scenario is well suited to rural deployment, in which RSs are deployed for coverage extension beyond the range of the BS. Let also assume that each RS is connected to the BS by a backhaul link of rate $\rho/2$. Note that if the backhaul link is a radio link, then at high SNR we can assume that ρ scales linearly with the bandwidth allocated to this link. Therefore, increasing ρ amounts to increasing the backhaul link bandwidth.

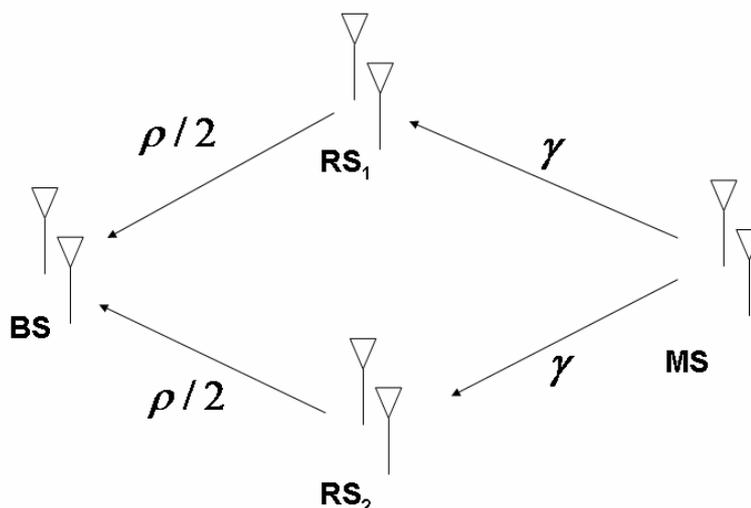


Figure 17: Symmetric Diamond out-of-band relay deployment topology with 2 parallel relays

We assume that each MS has the same number of antennas N_S and each RS has the same number of antennas N_R . At low SNR γ on the access link the ergodic capacity can be approximated as [TV05]:

$$C_{S-R} \approx N_R \gamma \log_2 e \quad (4.71)$$

When the backhaul rate ρ goes to infinity, the achievable rate of NCDF is limited by (4.71) and even if the best relay is selected, the achievable rate of all DF strategies will scale as N_R . In contrast, the Cut-Set Bound converges to the capacity of a VMIMO channel and therefore we expect that the following ergodic rate could be approached by distributed compression:

$$C_{CSB} \approx 2N_R \gamma \log_2 e \quad (4.72)$$

This phenomenon is illustrated on Figure 18, where the average achievable rate is plotted as a function of the average SNR on the MS-RS link in the same diamond configuration as Figure 17. At 0dB SNR, the virtual MIMO channel capacity is 2.8 bit/channel use, versus 2 bits for DF with best relay selection. This represents a 40% potential achievable rate increase. It would also be possible to increase the rate by broadcasting one message to each relay. However, intuitively if the SNR on both links are similar, it means that each message is assigned half the power and since the capacity at low SNR is proportional to γ we cannot expect a significant sum-rate increase.

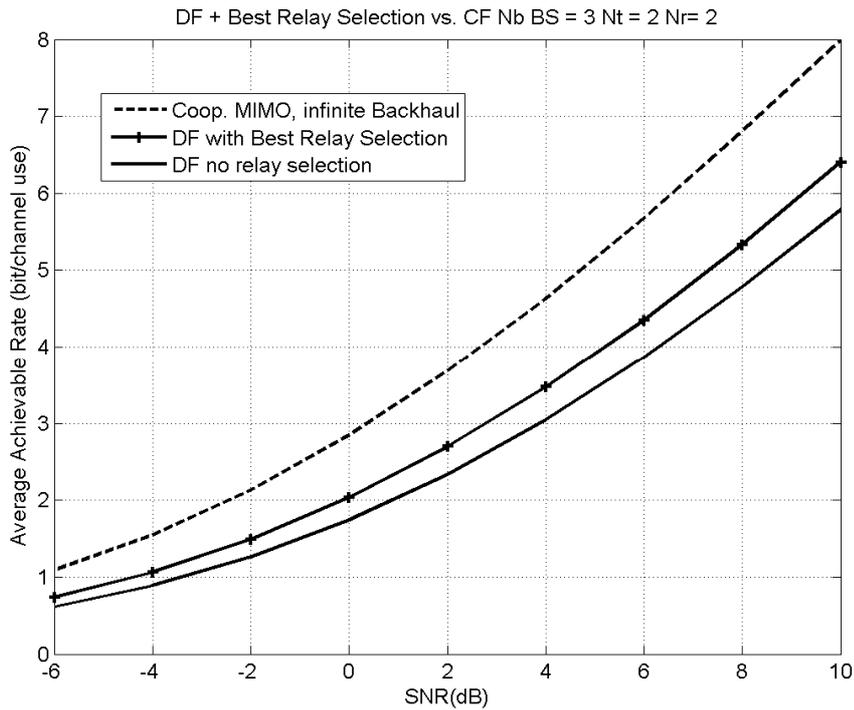


Figure 18: Average achievable rate of the DF protocol and Channel capacity in a symmetric diamond Relay channel at low SNR on the MS-RS links and high backhaul rate, single user, $N_S = 2$, $N_R = 2$.

Let now consider the DCF-SD strategy of Proposition 4.7. Each relay compresses its observation and forwards it to the BS for decompression followed by joint MIMO receiver processing. The compression at each RS assumes that a certain amount of side information is available at the BS, which depends on the decompression order at the BS. In this diamond configuration, there are only two possible orders and we select for each channel realization the order which provides the highest achievable rate. On Figure 19, we compare DCF-SD with a lower complexity Quantize-and-Forward (QF) strategy in which uniform scalar quantization is applied on a per-antenna basis. In this case, the achievable rate can be estimated by roughly approximating the quantization noise by AWGN. When ρ grows to infinity, both strategies achieve the cooperative MIMO capacity. Figure 19 addresses the following question: **in a real network what is the minimum required backhaul capacity to achieve most (here 90%) of the cooperative uplink MIMO capacity?**

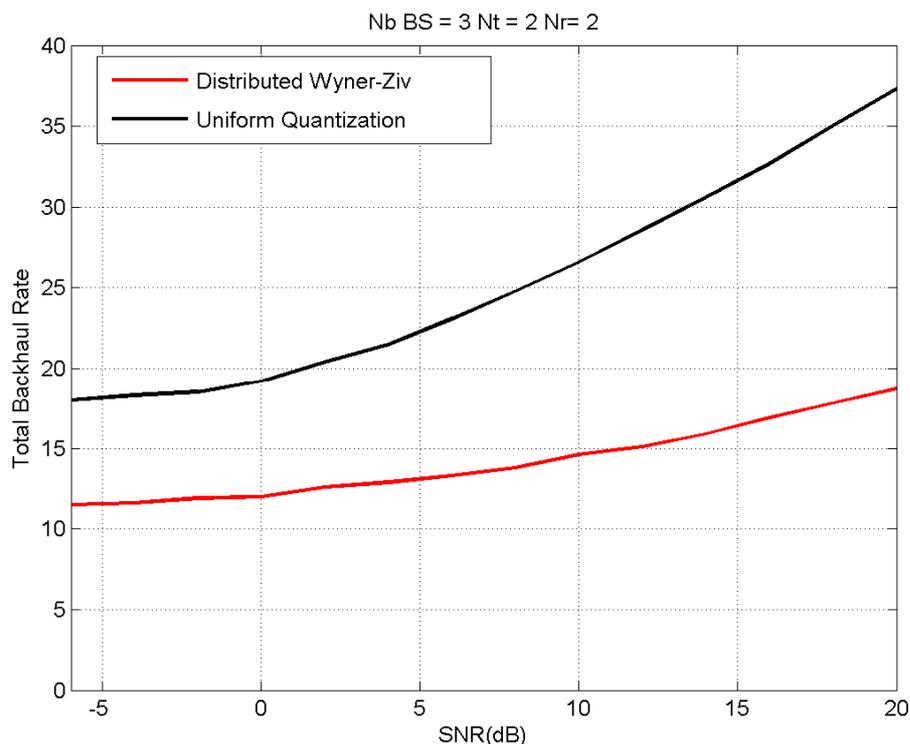


Figure 19: Required backhaul capacity to achieve 90% of the uplink capacity as a function of the SNR on the MS-RS link in a symmetric diamond topology, single user, $N_S = 2$, $N_R = 2$.

The figure shows that at $\gamma = 10\text{dB}$, with DCF-SD approximately $\rho = 15$ bits of backhaul rate are required per channel use in order to achieve 90% of the cooperative

MIMO capacity which from Figure 18 equals 8 bit/channel use at that value of SNR. At $\gamma = 0\text{dB}$, a total of 12 bits are required for a cooperative MIMO capacity which equals 3 bits/channel use. From these preliminary results we get the rough indication that the backhaul sum-rate shall be equal to as much as four times the access rate in order to achieve most of the cooperative MIMO capacity by distributed compression techniques. Also note that the distributed WZ compression of the DCF-SD scheme greatly reduces the backhaul rate requirements compared to sub-optimum source coding schemes such as uniform quantization. Indeed, at high SNR, the number of bits required by uniform quantization to maintain the distortion to thermal noise ratio at a certain level grows logarithmically with the SNR (i.e. linearly with the SNR in dB). However, DCF-SD benefits from the fact that the conditional covariance of the observation at one relay given the observation at the other relay decreases as the SNR increases, which is reflected in the slope of the required backhaul rate vs SNR curves on Figure 19 which is steeper for uniform quantization than for DCF-SD.

On Figure 20, one can see that at $\gamma = 10\text{dB}$ the ratio between the access capacity and the required backhaul rate remains approximately equal to 1/2 with DCF-SD when the number of antennas at the RS is doubled, whereas it falls down to 1/3 or 1/5 with uniform quantization. This shows that when N_R grows while N_S remains constant, it becomes necessary to apply a linear transform such as the KLT or CKLT at the RS (as in DCF-SD) instead of low-complexity QF, in order to have a backhaul load that scales as the number of degrees of freedom in the signal and not as the number of antennas at the RS.

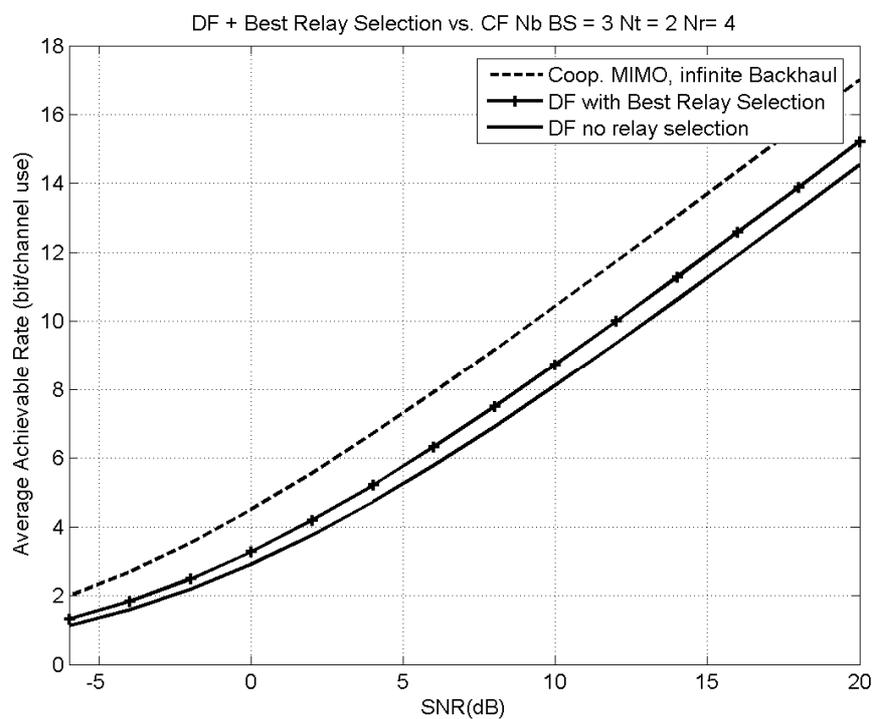
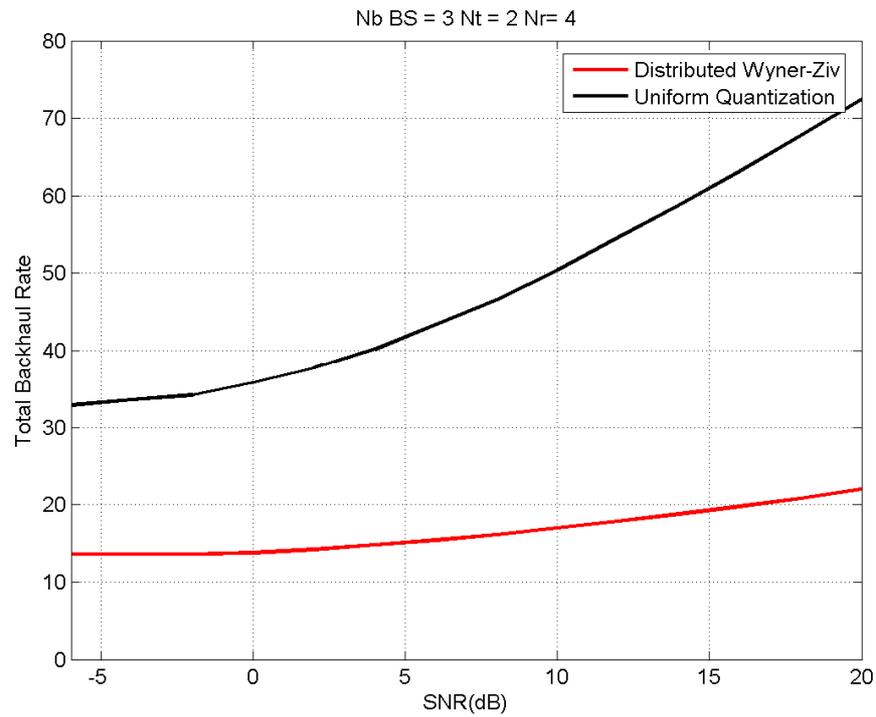


Figure 20: Same as Figure 18 and Figure 19 in a $N_S = 2$ $N_R = 4$ antenna configuration, two-users.

On Figure 21, the same simulation is run in a two-user scenario with $N_S = 2$ $N_R = 2$ antenna configuration. The achievable sum-rate is plotted as a function of the backhaul sum-rate. In this case, at high SNR on the MS-RS link the achievable sum-rate of DF protocols scales as $\min(2N_S, N_R)$ whereas the capacity of the virtual MIMO channel scales as $\min(2N_S, 2N_R)$. Thus DCF-SD compression can also provide large capacity gains at high SNR. However, this situation will only occur for a large density of RSs. It can be noticed that also in the multi-user case the ratio between the access sum-rate and the required backhaul sum-rate remains approximately equal to 1/2 at 10 dB SNR.

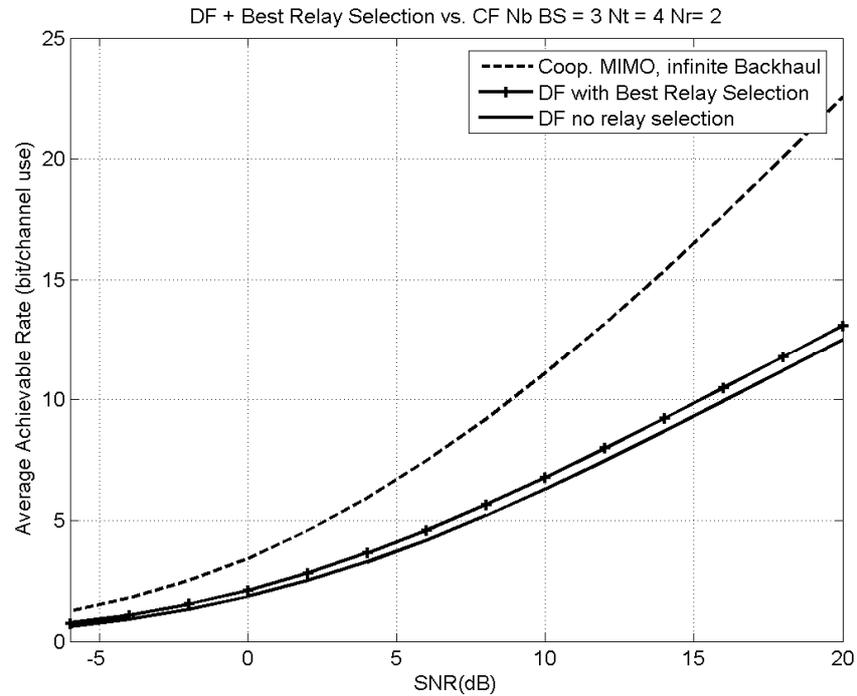
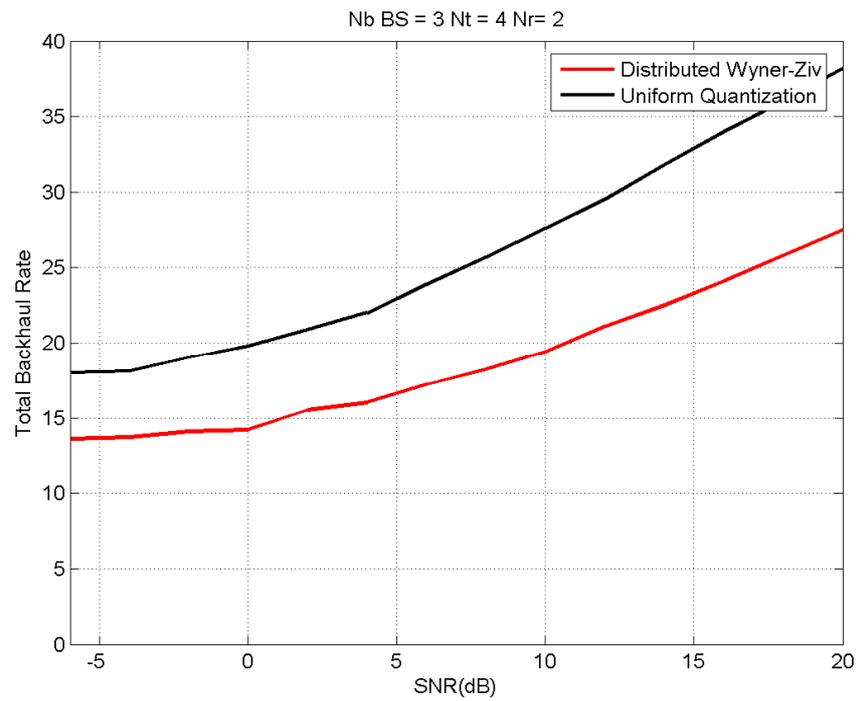


Figure 21: Same as Figure 18 and Figure 19 in a $N_S = 2$ $N_R = 2$ antenna configuration, two users.

4.3.5.2 Coordinated Network configuration

In the previous section, we illustrated the DCF-SD strategy in a diamond topology in order to analyze the backhaul requirements of distributed compression. However, we stuck to the coding strategy of Proposition 4.7. Many interesting simulation results concerning the other capacity bounds of this chapter are presented in [CS08a][CS08b][CS08c] and [CS08d]. For brevity, we refer the interested reader to these references.

4.3.5.3 Conclusions from simulations

The simulations in this section have shown that when some spare capacity is available on the cellular backhaul, it can be used to improve the uplink throughput of remote users by creating a Virtual Antenna Array. Simulations show that with distributed compression the required backhaul sum-rate must be approximately equal to four times the access sum-rate in order to achieve most of the virtual MIMO channel capacity over a wide range of SNRs. Distributed compression can thus allow to increase cell-edge throughput in lightly-loaded networks. In addition, the backhaul can also be used to spatially multiplex several users at high SNR. This might become especially relevant in future backhauls based on high capacity optical fiber.

Chapter 5: From Capacity bounds to practical implementation

5.1 Introduction and overview of our contribution

In the previous chapters, we have computed capacity bounds for the MIMO relay channel with an emphasis on the DF and CF strategies for which we derived achievable rates in the full CSI case. In this chapter, we review various issues which arise when we consider a practical implementation of these strategies in a state-of-the-art broadband wireless access network such as IEEE802.16 [16j07][16m06]. Our contribution is the following:

- In §5.2 we review the implementation of cooperative DF relaying.
 - First, we show that the capacity bounds that we derived in the previous chapters can be extended to model MIMO-OFDM transmission, various transmit power constraints and constraints related to the finite set of Modulation and Coding Schemes.
 - In §5.2.1.4 we study the effect of imperfect CSI. We propose some modifications to the achievable rate optimization problem to handle the case of statistical CSI and we verify that quantized precoder codebooks can also be applied to cooperative relaying.
 - We conduct a detailed study of two practical implementations of cooperative DF Protocol I based on the convolutionally turbo-coded mode of IEEE802.16e.
 - The first implementation is a cooperative Incremental Redundancy strategy. We derive the parameters of an EESM error predictor for cooperative IR and compute its throughput performance under a target error rate. We verify that the throughput envelope can be well approximated by the degraded achievable rate which is obtained by simple modifications of the information-theoretic formulas of previous chapters.
 - However, the peak rate of cooperative IR may be limited if the set of MCS does not allow very high spectral efficiencies per symbol. In

such situations, we show that a superposition coding strategy during the first slot of the TDD protocol can overcome the peak rate saturation problem.

- In §5.3, we review some implementation constraints for the CF strategy. We show that as for DF, the capacity bounds can be extended to handle practical constraints such as MIMO-OFDM transmission. We also briefly describe how practical Wyner-Ziv coding can be realized and what performance can be expected.
- In §5.4 we conduct some system-level simulations to check whether the performance observed from link-level simulations can match practical deployment scenarios. We present a Quasi Monte-Carlo system simulation framework and review the main parameters, before running some simulations in single-cell and multi-cell downlink scenarios to assess how cooperative DF strategies can increase the cellular throughput.
 - In the single-cell scenario we illustrate the effect of shadowing and relay density. We show that cooperative partial DF Protocol III is the most efficient and allows a large increase of achievable rate in the vicinity of the RS and at cell edge. When full CSI is available, even larger gains are achievable by cooperative beamforming (Protocols II and III), as predicted by link-level simulations.
 - In the multi-cell scenario, we model additional effects such as inter-sector and inter-cell interference. We show that a careful positioning of RSs in the deployment is required if RSs cannot handle a connection to multiple BSs. We study the potential gains of non-orthogonal resource allocation with a spatial reuse of the relay time-frequency slot and show that it allows a large increase of spectral efficiency. Moreover, spatial reuse is possible with Protocol I but cannot be directly implemented with Protocol III. Therefore, Protocol I can be preferred in many cases at the system-level although it is outperformed by Protocol III at the link level.

5.2 Practical Implementation of Decode-and-Forward strategies

5.2.1 Constrained Capacity Bounds

5.2.1.1 MIMO-OFDM transmission

If the system is wideband and employs Orthogonal Frequency Division Multiplexing (OFDM), the equations of Chapter 3 must be modified to account for the parallel transmission on multiple sub-carriers. A term $\mathcal{C}(\mathbf{R}, \mathbf{H})$ shall be replaced by a sum $\sum_{i=1}^{N_c} \mathcal{C}(\mathbf{R}_i, \mathbf{H}_i)$ where N_c is the number of sub-carriers, and \mathbf{R}_i and \mathbf{H}_i are the signal covariance and MIMO channel matrices on the i th sub-carrier. Likewise a transmit power constraint shall now read $\sum_{i=1}^{N_c} \text{tr}(\mathbf{R}_i) \leq P$. A spectral mask constraint can be translated into a per-subcarrier power constraint. For instance, assuming a perfectly flat mask gives: $\text{tr}(\mathbf{R}_i) \leq P / N_c$. Per-antenna and per-subcarrier power constraints can also be applied simultaneously. Note that all these changes do not affect the convexity of the optimization problems and therefore the convex optimization procedures presented in Chapter 3 remain applicable.

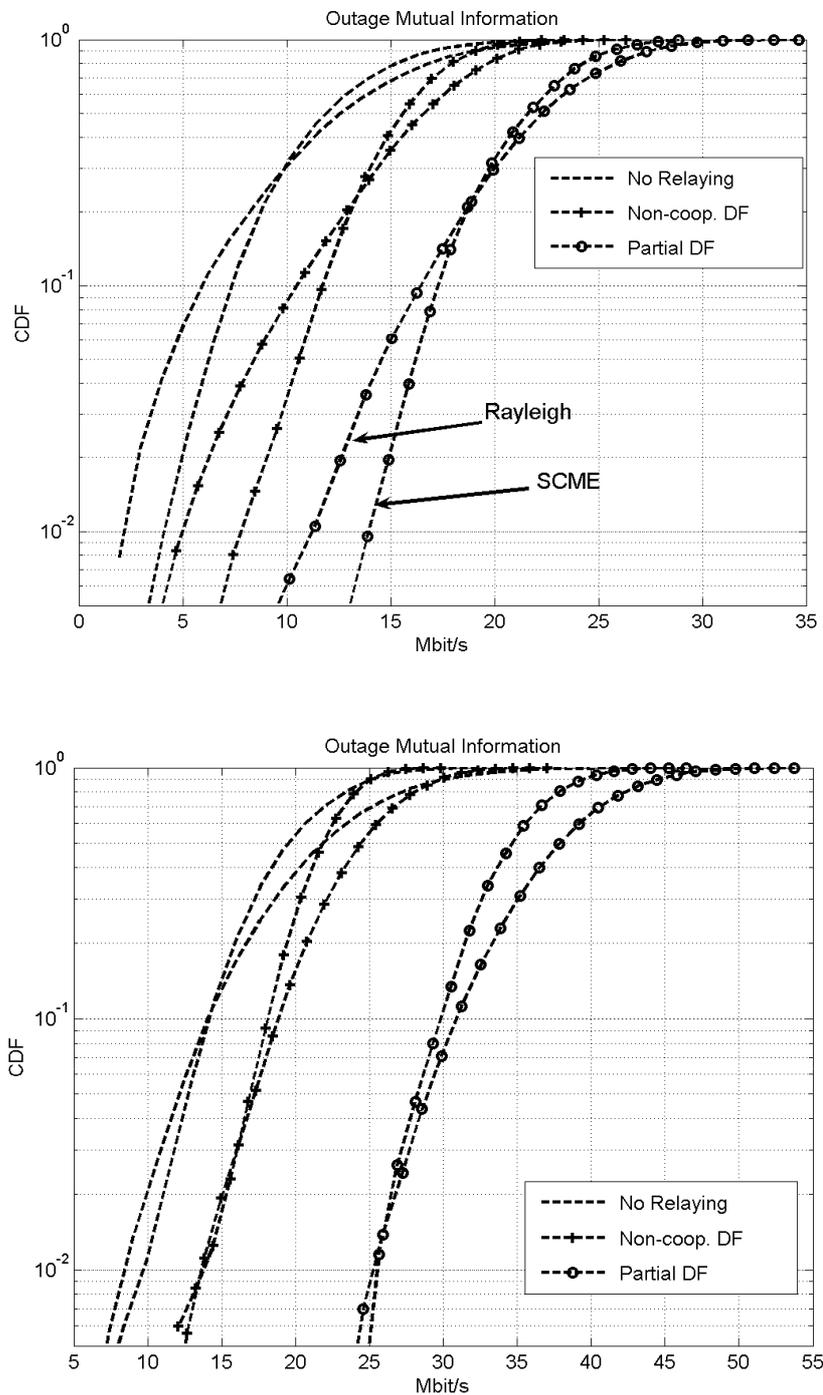


Figure 22: CDF of the achievable rate for DF protocols in a $N_S = N_R = 2$ configuration at $(\gamma_0 = 5\text{dB}, \gamma_1 = 20\text{dB}, \gamma_2 = 10\text{dB})$, for a broadband channel (SCME typical urban model) and for a single-carrier Rayleigh i.i.d. model. Top: $N_D = 1$, Bottom: $N_D = 2$

On Figure 22, the impact of frequency diversity on the outage performance of cooperative DF strategies is illustrated. The CDF of the achievable rate is plotted for two different systems. The first one is a single-carrier system for which the MIMO channel is modeled as Rayleigh i.i.d. the second is a broadband OFDM system with 10 MHz channel bandwidth for which the SCME typical urban channel model is assumed. The average SNR is the same for both systems and the transmit covariance matrices are not optimized (i.e. isotropic transmission). In the $2 \times 2 \times 1$ antenna configuration (top figure), it can be observed that the frequency diversity provides a significant rate increase at low outage probability (less than 5%). However, in the $2 \times 2 \times 2$ configuration, it looks like there is enough space diversity and the benefits of additional frequency diversity cannot be observed. Other simulation results could confirm the fact that very often in scenarios of practical interest the conclusions drawn from an analysis of cooperative coding strategies in the single-carrier MIMO relay channel are directly applicable to the broadband MIMO-OFDM case. This motivates our interest (and the interest of the research community) in the single-carrier case since it is easier to analyze and less complex to simulate.

5.2.1.2 Transmit Power Constraints

As mentioned in the previous section, a spectral mask constraint can be easily modeled while preserving the convexity of the achievable rate optimization problem. Likewise, the sum-power constraints $\text{tr}(\mathbf{R}) \leq P$ on transmit covariance matrices can be replaced by per-antenna power constraints $R_{i,i} \leq P/N$ where P is the total device power and N is the number of antennas. Again, this change in the constraints preserves the convexity of the optimization problems.

5.2.1.3 Modulation and coding constraints

A state-of-the-art broadband wireless system (e.g. [16e05]) typically encodes finite-length packets with a turbo-code or an LDPC, and maps the output onto finite alphabet symbols (e.g. QAM). In this case, the precoders and rate allocation algorithms derived in Chapter 3 are not directly applicable, since they are designed assuming Gaussian i.i.d. codewords of infinite length. This issue is well-known in OFDM systems and solutions such as bit and power loading have been proposed as realistic alternatives to waterfilling [CCB95]. An interesting property of the degraded capacity formula (2.27) is that it remains convex

in $\boldsymbol{\gamma}$, the vector of SNRs per subcarrier and spatial stream. The SNR degradation factor Γ and the maximum rate saturation can thus be introduced in the achievable rate expressions of Chapter 3 without affecting the convexity of the optimization problem. However, note that the rate saturation shall be preferably introduced as an additional inequality constraint in order to keep the objective and constraints differentiable. For instance, the equation (3.27) in the expression of the PDF achievable rate can be replaced by four inequality constraints:

$$\begin{aligned}
R_A &\leq t_1 \mathcal{C}\left(\frac{\mathbf{Q}_S^{(1)}}{t_1}, \frac{1}{\sqrt{\Gamma}} \mathbf{H}_1\right) + t_2 \mathcal{C}\left(\frac{\mathbf{Q}_{S,1}^{(2)}}{t_2}, \frac{1}{\sqrt{\Gamma}} \mathbf{H}_0 \mathbf{D}_S\right) \\
R_A &\leq t_1 \min(N_S, N_D) R_{\max} + t_2 \mathcal{C}\left(\frac{\mathbf{Q}_{S,1}^{(2)}}{t_2}, \frac{1}{\sqrt{\Gamma}} \mathbf{H}_0 \mathbf{D}_S\right) \\
R_A &\leq t_1 \mathcal{C}\left(\frac{\mathbf{Q}_S^{(1)}}{t_1}, \frac{1}{\sqrt{\Gamma}} \mathbf{H}_1\right) + t_2 \min(N_S, N_D) R_{\max} \\
R_A &\leq t_1 \min(N_S, N_D) R_{\max} + t_2 \min(N_S, N_D) R_{\max}
\end{aligned} \tag{5.1}$$

where R_{\max} is the maximum spectral efficiency in the set of MCS of the system. For instance, if the largest constellation is 64QAM and the highest code rate is 5/6 then $R_{\max} = 5$ bits (per QAM symbol). Including rate saturation constraints into the optimization starts to make the latter really complex. However, simulations show that the effect of rate saturation on the achievable rate starts to become marginal when the set of MCS includes very high order constellations combined with high code rates. For instance on Figure 23 we plot for each coding strategy the achievable rate under four increasing levels of degradation:

1. $\Gamma = 0\text{dB}$, $R_{\max} = +\infty$ (ideal reference curve)
2. $\Gamma = 4\text{dB}$, $R_{\max} = 10$ b (SNR degradation but almost no rate saturation)
3. $\Gamma = 4\text{dB}$, $R_{\max} = 6$ b (SNR degradation and moderate rate saturation)
4. $\Gamma = 4\text{dB}$, $R_{\max} = 4$ b (SNR degradation and severe rate saturation)

It can be observed that the SNR degradation accounts for most of the achievable rate decrease whereas the effect of rate saturation becomes marginal when high order constellations can be combined with high code rates. Indeed, increasing the maximum spectral efficiency per symbol from $R_{\max} = 6$ (e.g. 256QAM rate $\frac{3}{4}$) to $R_{\max} = 10$ yields an achievable rate gain of less than 0.3 bit at high SNR. Therefore initially in our system-

level simulations we only modeled the SNR degradation and not the rate saturation phenomenon. However, in §5.4.4 both phenomena were accounted for.

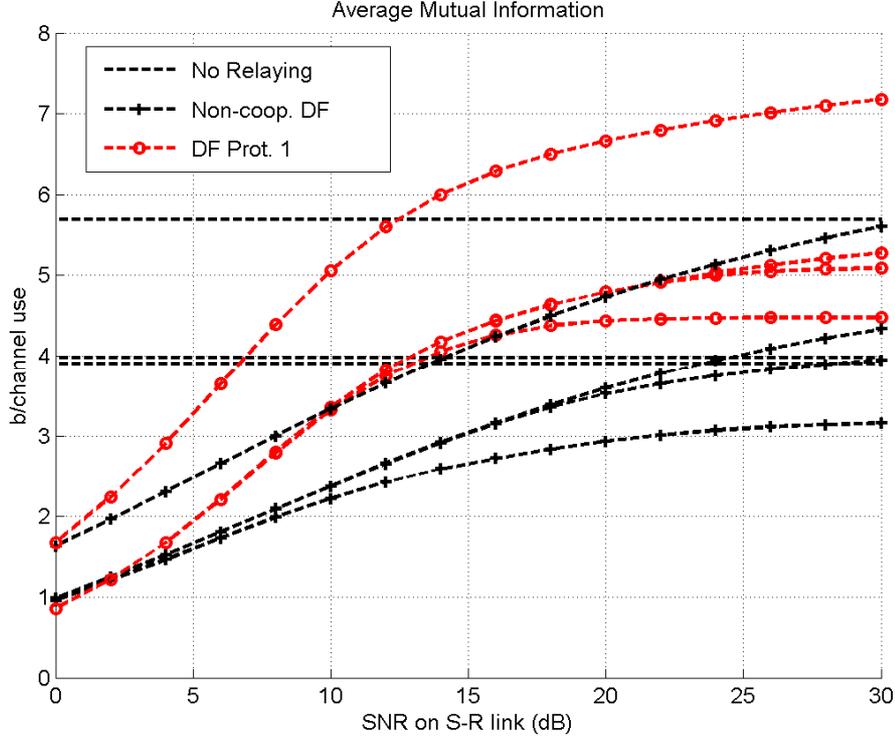


Figure 23: Impact of achievable rate degradation parameters (SNR degradation Γ and maximum rate per QAM symbol R_{\max}) on the achievable rate of cooperative DF protocols in the multiple-antenna case ($N_S = N_R = N_D = 2$, $\gamma_0 = 10\text{dB}$, $\gamma_2 = 15\text{dB}$) for 4 increasing levels of degradation ($\Gamma = 0\text{dB}$, $R_{\max} = +\infty$), ($\Gamma = 4\text{dB}$, $R_{\max} = 10$), ($\Gamma = 4\text{dB}$, $R_{\max} = 6$), ($\Gamma = 4\text{dB}$, $R_{\max} = 4$)

Finally, note that another more recent approach includes the finite alphabet assumption in the information-theoretic optimization [LTV05]. The problem is that it does not model the finite codeword length effect and moreover it requires having an a priori knowledge of the constellation on each subcarrier and spatial stream, which is a chicken and egg problem.

5.2.1.4 Imperfect CSI

State-of-the-art TDD systems often rely on UL/DL channel reciprocity to obtain CSIT without having to explicitly feed back the estimated channel coefficients. In this

case, CSIT imperfection⁴ comes from the estimation noise and from the variations of the channel between the moment it is estimated in one direction and the moment it is applied in the other direction. However, in cooperative relaying it is hard if not impossible to avoid explicit signaling because for instance the source cannot estimate the relay-destination channel by means of reciprocity. Moreover, even if the channel is very slowly varying (e.g. pedestrian speed), feeding each MIMO channel estimate back to the source typically requires a prohibitively large number of bits to achieve a quantization noise variance which is of the same order of magnitude as the channel estimation MSE.

One possible solution to overcome this problem is to feed back the channel covariance matrix instead of the channel time or frequency response. The rate at which the covariance needs to be fed back is low. Indeed, the channel covariance is assumed to remain almost constant over time frame much larger than the channel coherence time. Moreover, for a MIMO-OFDM system, a single frequency-domain channel covariance matrix needs to be fed back for all subcarriers. In this case, the CSI and therefore the capacity and the achievable rate of the various relaying strategies shall be treated as random variables. The precoders and resource allocation can be designed to maximize an objective which is a function of the CSI distribution (e.g. outage capacity, average capacity). This solution is investigated in §5.2.1.4.1 below. Another solution is to feed back the precoder instead of the CSI, where the precoder is selected from a small set (i.e. a set of small cardinality) of structured matrices. Thus the precoder can be quantized on a small number of bits (e.g. 4 bits) and fed back at intervals much lower than the coherence time while keeping the feedback load at an acceptable level. In MIMO-OFDM systems the channel correlation in frequency-domain can be exploited to reduce the feedback load. Quantized precoders are investigated in §5.2.1.4.2 below.

5.2.1.4.1 Statistical and Hybrid CSI

5.2.1.4.1.1 Background and prior art on statistical CSI exploitation

Statistical CSI is understood in this thesis as the knowledge of the average SNR and of the channel covariance matrix. The knowledge of the statistics of a potential interferer is not a topic addressed here. The estimation of the average SNR is a well-known topic and

⁴ Note that in the presence of interference (not treated here) the channel is not reciprocal because interference is different at each node.

without spending more time on it we will assume that it can be perfectly estimated. In [JVG01], the MIMO channel is modeled as a product:

$$\mathbf{H} = \mathbf{W}\mathbf{R}_T^{1/2} \quad (5.2)$$

where \mathbf{W} is i.i.d. Rayleigh fading and \mathbf{R}_T is the transmit correlation matrix. The rows of the channel matrix are assumed uncorrelated but the columns are correlated (i.e. the transmit antennas are correlated but the receive antennas are not). A row vector \mathbf{h}_i for $i = 1, \dots, N_R$ has a covariance matrix \mathbf{R}_T / N_R . In a MIMO-OFDM system, this channel covariance matrix can be estimated by averaging over all the subcarriers. Then for each received packet, a simple AR model can be implemented in time-domain as follows:

$$\hat{\mathbf{R}}(k) = (1 - \alpha)\hat{\mathbf{R}}(k-1) + \alpha \left(\frac{1}{N_C} \sum_{i=1}^{N_C} \hat{\mathbf{H}}_i^H(k) \hat{\mathbf{H}}_i(k) \right) \quad (5.3)$$

where k is the index of the OFDM symbol. It is shown in [JVG01] that the ergodic capacity of a point-to-point MIMO channel is maximized by precoding with the eigenvectors of \mathbf{R}_T . Moreover, the power loading vector shall be ordered like the eigenvalues, i.e. more power shall be allocated to the largest eigenvalues. In [JB04], a more complex channel model is considered:

$$\mathbf{H} = \mathbf{R}_R^{1/2} \mathbf{W} \mathbf{R}_T^{1/2} \quad (5.4)$$

and it is assumed that both the transmit and receive correlation matrices \mathbf{R}_T and \mathbf{R}_R are known. In this case, the authors show that the optimum transmit directions are the eigenvectors of \mathbf{R}_T but the optimum power allocation depends on \mathbf{R}_R . In both [JVG01] and [JB04], the optimum power allocation is obtained by numerical (convex) optimization, but no closed-form expression is provided. Both papers also verify that at low SNR or when one eigenvalue is much larger than the other ones then the optimum scheme degenerates into beamforming.

5.2.1.4.1.2 Application to the relay channel

As far as we know, the MIMO relay channel capacity bounds have not yet been extended to the case of covariance feedback. Let consider for instance the DF strategy with Protocol I. If the objective is the average achievable rate, the optimization can now be formulated as follows:

$$\begin{aligned}
R_{FDF,I} &= \max_{t_1 > 0, t_2 > 0, \mathbf{Q}_S^{(1)} \succeq 0, \mathbf{Q}_R^{(2)} \succeq 0} \min \{R_A, R_B\} \\
R_A &= E_{\mathbf{H}_1} \left[t_1 C(\mathbf{Q}_S^{(1)} / t_1, \mathbf{H}_1) \right] \\
R_B &= E_{\mathbf{H}_0, \mathbf{H}_2} \left[t_1 C(\mathbf{Q}_S^{(1)} / t_1, \mathbf{H}_0) + t_2 C(\mathbf{Q}_R^{(2)} / t_2, \mathbf{H}_2) \right] \\
t_1 + t_2 &\leq 1; \quad \text{tr}(\mathbf{Q}_S^{(1)}) - t_1 P_S \leq 0; \quad \text{tr}(\mathbf{Q}_R^{(2)}) - t_2 P_R \leq 0
\end{aligned} \tag{5.5}$$

The expectation is a convexity-preserving operation [BV04], and therefore the problem (5.5) is convex. It cannot (probably) be solved in closed-form but numerical optimization can still be performed as in Chapter 3. The partial derivatives can be computed by taking the expectation of the closed-form expressions derived in Appendix A:

$$\frac{\partial E_{\mathbf{H}} [tC(\mathbf{Q}/t, \mathbf{H})]}{\partial \mathbf{Q}} = E_{\mathbf{H}} \left[\frac{\partial tC(\mathbf{Q}/t, \mathbf{H})}{\partial \mathbf{Q}} \right] \tag{5.6}$$

Given the lack of a closed-form expression for the partial derivatives, the computational complexity of the numerical optimization (5.5) will be very high. Various approaches of much lower complexity can be considered at the expense of a lower rate. One possible approach for reducing the numerical complexity can be as in §3.5.3.1 to impose a structure to the source precoder during the first slot, while the relay precoder during the second slot can be optimized by applying [BV04] to the point-to-point MIMO link between R and D. A precoder structure which makes sense intuitively is

$$\mathbf{R}_S^{(1)} = \mathbf{U}_0 \text{diag}(\mathbf{p}_0) \mathbf{U}_0^H + \mathbf{U}_1 \text{diag}(\mathbf{p}_1) \mathbf{U}_1^H \tag{5.7}$$

where the columns of \mathbf{U}_0 and \mathbf{U}_1 are the eigenvectors of $\mathbf{R}_{T,0}$ and $\mathbf{R}_{T,1}$. A further simplification can then be achieved by limiting the rank of the precoder (5.7) to the eigenvalues which are greater than a certain threshold, and by allocating the transmit power uniformly over the set of eigenmodes. The problem then amounts to finding the optimum fraction of source transmit power that shall be allocated to the eigenmodes of the Source-Relay channel.

5.2.1.4.1.3 The case of Hybrid CSI

Hybrid CSI is defined here as in [GHS06], by the fact that a given node may have perfect CSI for some channels but only statistical CSI for some other channels. The hybrid CSI case is especially relevant to cellular relaying with fixed relays. For instance in the downlink, the BS-RS channel \mathbf{H}_1 is varying very slowly and can therefore be easily tracked, whereas the RS-MS and BS-MS channels \mathbf{H}_2 and \mathbf{H}_0 may not be

trackable if the MS is moving too fast. With hybrid CSI, the full-fledge optimization remains highly complex, but the sub-optimum precoder structure can be improved by replacing the matrix \mathbf{U}_1 in (5.7) by the matrix \mathbf{V}_1 of the right singular vectors of \mathbf{H}_1 .

5.2.1.4.2 Quantized CSI and Quantized Precoders

5.2.1.4.2.1 Background on Quantized Precoders

When it is possible to track the channel but the feedback load would be prohibitive, an efficient technique consists in selecting the “best” MIMO precoder from a set of reduced cardinal N (e.g. $N = 16$). Only the index of the precoder needs to be fed back, which requires only $\log_2(N)$ bits. In [LHS03], the authors restrict to beamforming in a point-to-point MIMO link (here the S-D link), i.e. a single spatial stream is sent even if both the source and destination have multiple antennas. Denoting by \mathbf{z} the receiver weights (which are normalized such that $\|\mathbf{z}\|_2 = 1$) and by \mathbf{w} the transmitter weights, it is straightforward that the maximization of the capacity is equivalent to the maximization of the SNR at the output of the receiver, which is proportional to $|\mathbf{z}^H \mathbf{H} \mathbf{w}|^2$. The optimal receiver weights are such that $|\mathbf{z}^H \mathbf{H} \mathbf{w}| = \|\mathbf{H} \mathbf{w}\|_2$ which corresponds to the Maximum Ratio Combining solution. The optimal transmit weights (assuming an MRC receiver) are given by the solution of the following problem:

$$\hat{\mathbf{w}} = \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmax}} \|\mathbf{H} \mathbf{w}\|_2 \quad (5.8)$$

If \mathcal{W} is the set of unit vectors of length N_s , then the well-known solution is the dominant right singular vector of \mathbf{H} , which corresponds to Maximum Ratio Transmission (MRT). Quantized precoding requires to find the best precoder codebook \mathbf{W} of a given size N , defined as the $N_s \times N$ matrix whose columns are the transmit weights vectors \mathbf{w}_i $i = 1, \dots, N$. Given a precoder codebook, the transmit weights for a given channel realization can be found by exhaustive search:

$$\hat{\mathbf{w}} = \underset{1 \leq i \leq N}{\operatorname{argmax}} \|\mathbf{H} \mathbf{w}_i\|_2 \quad (5.9)$$

The authors in [LHS03] design quantized precoders for uncorrelated fading channels, i.e. the components of \mathbf{H} are assumed i.i.d. complex Gaussian. They show that for this specific channel distribution, the precoder codebook which minimizes the average SNR loss with respect to the MRT can be obtained by maximizing the minimum angle between all pairs of weight vectors:

$$\delta(\mathbf{W}) = \min_{1 \leq k < l \leq N} \sqrt{1 - |\mathbf{w}_k^H \mathbf{w}_l|^2} \quad (5.10)$$

This criterion is called the Grassmannian Beamforming Criterion, because it is analog to the Grassmannian Line Packing problem. Quantized Equal Gain Transmitters (QEGT) and Quantized Maximum Ratio Transmitters (QMRT) can be designed by maximizing $\delta(\mathbf{W})$ over the set of complex matrices with respectively unit modulus coefficients and unit vector columns. The authors in [LHS03] state that the achievable rate difference between QMRT and QEGT is minor.

Quantized precoder codebooks have been extended to handle the multiple-stream case and the correlated MIMO channel case. For instance, in [LH05] multiple spatial streams can be sent and the design criterion can be either the minimization of the error probability (assuming an ML decoder) or the maximization of the capacity. Different design criteria imply the use of different definitions for the distance $\delta(\mathbf{W})$. Moreover, quantized precoder codebooks are also specified in 3GPP-LTE and IEEE802.16m standards for single-user and multi-user MIMO-OFDM systems. In order to reduce the feedback load, a single precoder is assigned per time-frequency resource block over which the channel is assumed to remain (roughly) constant.

5.2.1.4.2.2 Application to cooperative relaying

Because cooperation and limited feedback are two hot research topics for NG-BWA systems, the application of quantized precoders to cooperative transmission has recently drawn attention. However, as we are writing this thesis there are still very few papers on the topic, and they are limited to non-cooperative relaying [YWK07] and the AF strategy [ZAL07].

In this section we investigate the application of Grassmannian beamforming to cooperative DF Protocols II and III, in which both S and R have to cooperatively beamform to D during the second slot. The source covariance during the first slot is set to $(P_S / N_S) \mathbf{I}_{N_S}$. The joint source-relay transmit precoder is constrained to be a QEGT and designed based on the Grassmannian beamforming criterion (5.10). On Figure 24, we evaluate the achievable rate performance of quantized precoding compared to the optimal transmit covariance given by convex matrix optimization and to the lower complexity approach of §3.5.3.2. It can be checked that a 2-bit codebook outperforms the vector optimization approach and achieves all the cooperation gain provided by the matrix

optimization approach. On Figure 25, a 2x2x1 antenna configuration is considered. In this case, the vector optimization approaches the optimal precoder. It is well-known that for larger antenna arrays a larger codebook size is required. However, here with 5-bit codebook the performance of the vector optimization approach can be approached closely, still increasing the codebook size from 16 to 32 yields a very small gain and it seems useless to further increase it.

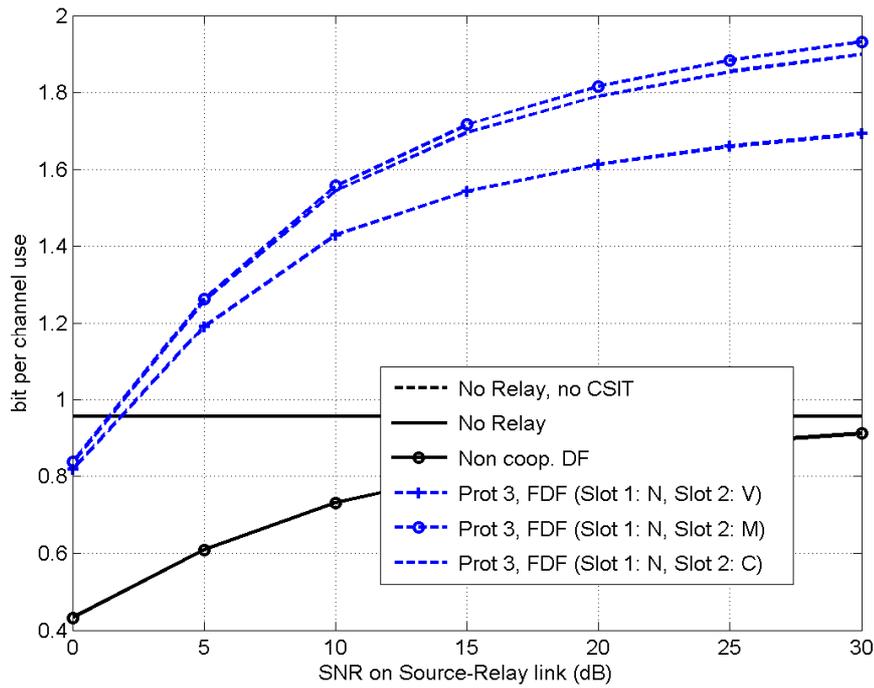


Figure 24: Impact of using a 2-bit precoder codebook in cooperative FDF Protocol III, in a 1x1x1 antenna configuration with $\gamma_0 = \gamma_2 = 0dB$ (Legend: N=No optimization, V=Vector Optimization, M=Matrix Optimization, C=Quantized Precoder Codebook)

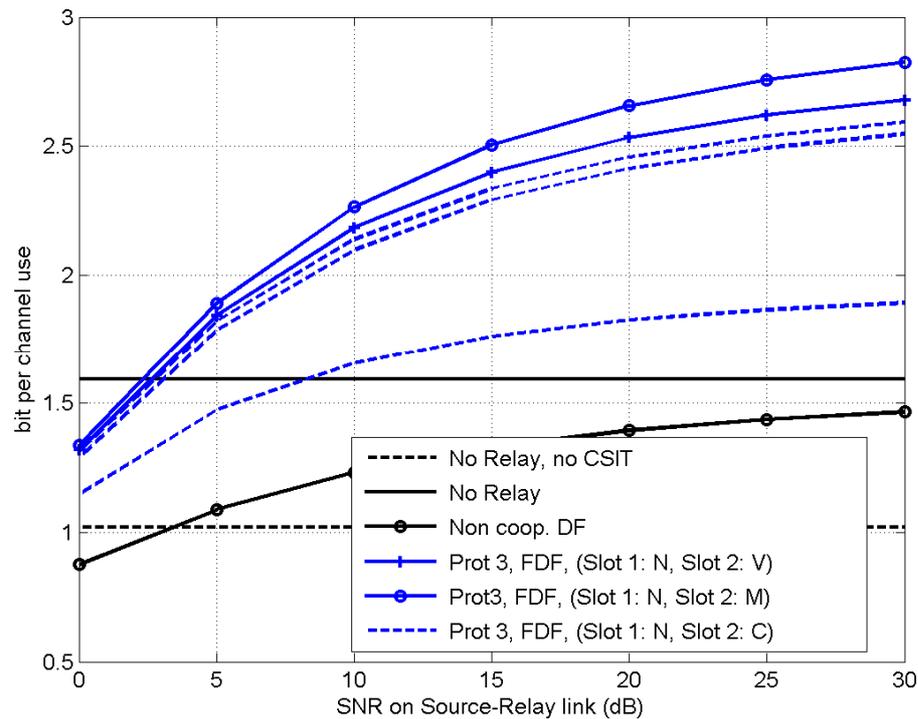


Figure 25: Same as Figure 24 in a 2x2x1 antenna configuration with 2, 4, 5 bit precoder codebook size

This very preliminary performance evaluation tends to show that quantized precoding can be applied successfully to enable cooperative relaying with limited feedback load. However, a more in-depth analysis will be needed to design codebooks which are specific to cooperative relaying, by taking into account the distributed channel characteristics, device and antenna power constraints as well as achievable rate expressions.

5.2.2 An implementation based on cooperative IR

System simulations performed in §5.4 show that cooperative (F)DF Protocol I (i.e. with orthogonal Source and Relay transmissions) is the most interesting at the system-level when at least two relays are deployed per BS sector. We now investigate how to implement such a coding strategy in a real system. Due to time limitations our simulations are limited to the single-antenna case, but as commented at the end of this section we expect that the conclusions drawn from the single-antenna analysis can be extended to the multi-antenna case.

5.2.2.1 Description of the IEEE802.16e convolutional turbo-code

Our coding strategy is based on the Convolutional Turbo Codes (CTC) of IEEE802.16e [16e05]. The CTC consists in two duo-binary constituent encoders which are Circular Recursive Systematic Convolutional codes. Figure 26 depicts the CTC encoder. The data block is split into two sub-blocks of bits A and B that are fed to the first constituent encoder shown in Figure 27 to generate parity bits Y_1 and W_1 . Afterwards, an interleaved version of A and B is encoded in order to generate additional parity bits (Y_2 and W_2).

The CTC specification in [16e05] allows a set of different coded block sizes to be generated. The larger the block size, the better the BLER performance of the turbo-code. We picked a $N=120B=960$ data bits block size for our simulations. The mother code size is $3N$.

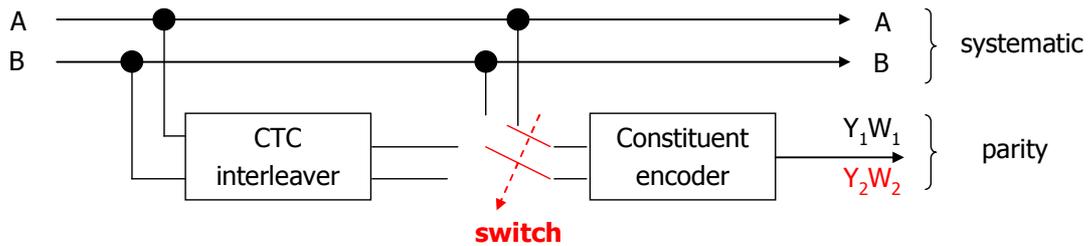


Figure 26 CTC encoder

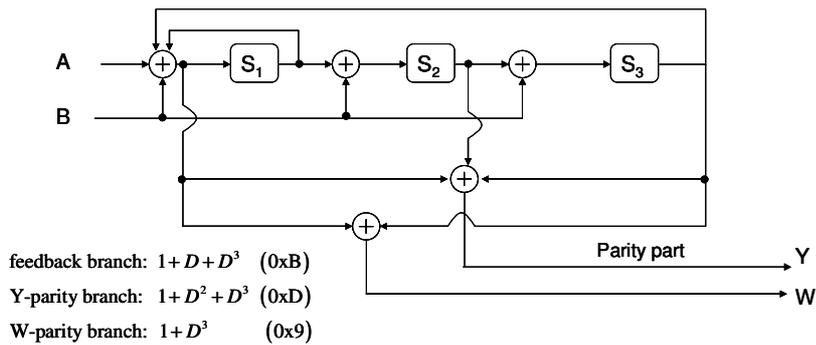


Figure 27 CTC constituent encoder

Note that in our study we do not consider the optional HARQ-IR mode of 802.16e which is defined in sections 8.4.9.2.3.4 and 8.4.9.2.3.5 in [16e05]. Instead we consider that the Source (BS) always transmits the first P coded bits with $P > N$ and N the number of data bits, such that the code rate is N/P . The first N bits of the codeword are systematic bits and the next $P-N$ bits are parity bits, as represented on Figure 29.

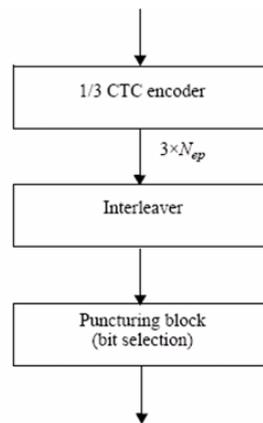


Figure 28 Block diagram of subpacket generation for CTC

We designed a link simulator which implements IEEE802.16e CTC Modulation and Coding Schemes. According to the standard, the following code rates can be obtained by puncturing the rate 1/3 mother code: $\frac{1}{2}$, $\frac{3}{4}$, $\frac{2}{3}$ and $\frac{5}{6}$. An MCS consists in the association of a code rate with a constellation QPSK, 16QAM or 64QAM. The MCS with the highest spectral efficiency is 64QAM rate $\frac{5}{6}$ with 5 bits per QAM symbol.

5.2.2.2 Cooperative IR strategies

The mother code is represented on Figure 29. The first N bits are systematic bits.

Several strategies are considered as illustrated on Figure 29:

- **Direct Transmission** (i.e. No Relaying) from Source to Destination.
- **(Cooperative) HARQ-no IR**
 - In this case, the Source and Relay respectively transmit the first N/R_S and N/R_R coded bits where R_S is the source code rate and R_R is the relay code rate
 - Note that if $R_S=R_R$ then the coded bits are just repeated
 - If cooperative DF is assumed, the Destination performs reliability (LLR) combining of the two received signals.
- **Cooperative IR v1**
 - The Source transmits the first N/R_S coded bits. The Relay forwards the last N/R_R coded bits. Some of the bits may be repeated if $N/R_S+N/R_R>3N$
 - Note that in this strategy the Relay does not transmit systematic bits
- **Cooperative IR v2**

- The Source transmits the first N/R_S coded bits. The Relay forwards the first N coded bits (the systematic bits) and the last $N/R_R - N$ coded bits.
- Note that in this strategy the Relay repeats the systematic bits

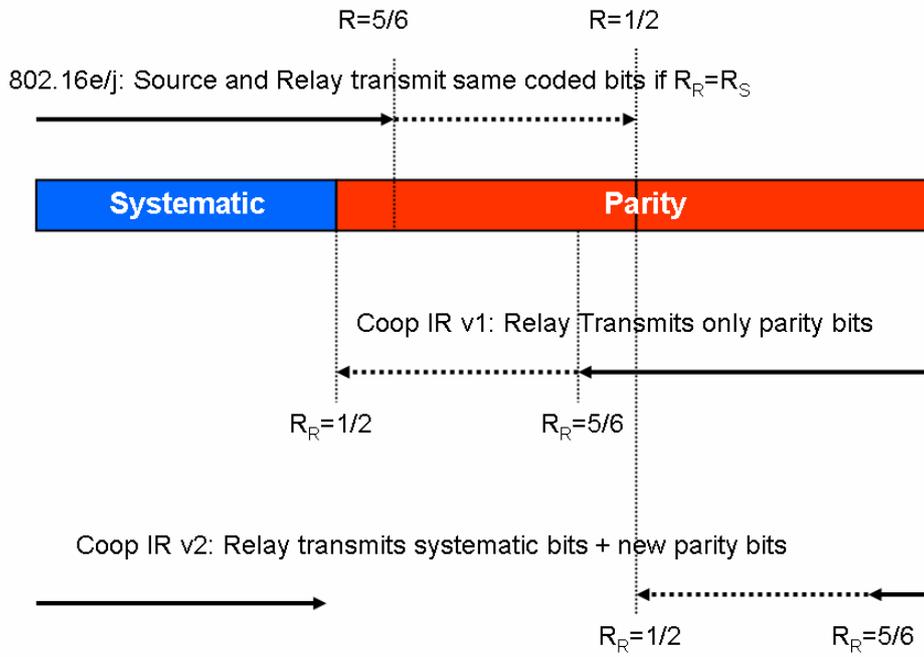


Figure 29: Sequence of coded bits transmitted by the relay for different cooperative IR strategies (the bits corresponding to the solid line are always transmitted by the relay, the bits corresponding to the dashed line are optionally transmitted depending on the relay code rate)

5.2.2.3 Cooperative IR throughput performance

The simulations in this section are based on the EESM error prediction model detailed in Appendix D. On Figure 30, the average throughput per QAM symbol of the different cooperative IR strategies is plotted versus the average SINR γ_2 on the R-D link, assuming a $\gamma_1 = 30$ dB SINR on the S-R link and either $\gamma_0 = 5$ dB or 10dB on the S-D link. This simulation scenario thus corresponds to downlink with a fixed RS having an excellent link with the BS. The MCS on the 1st hop is selected for a target BLER of 1% at the relay and the MCS at the relay is based on a 5% target BLER at the final destination, after potential combining that is strategy-dependent.

Figure 30 shows that a large gain of up to 5dB can be achieved by adopting cooperative IR v1 strategy. **The cooperative IR strategy that avoids as much as possible repeating**

coded bits clearly outperforms other strategies and yields a large performance increase of 5dB, which translates into a 2x throughput gain at low SINR when the SINR on the Source-Destination link equals 10dB. This gain reduces to about 2dB when the SINR on the S-D link is 5dB, and vanishes at lower SINRs. In the next section, we introduce another coding strategy for cooperative DF Protocol I and we perform a comparison of throughput and degraded capacity to analyze the results.

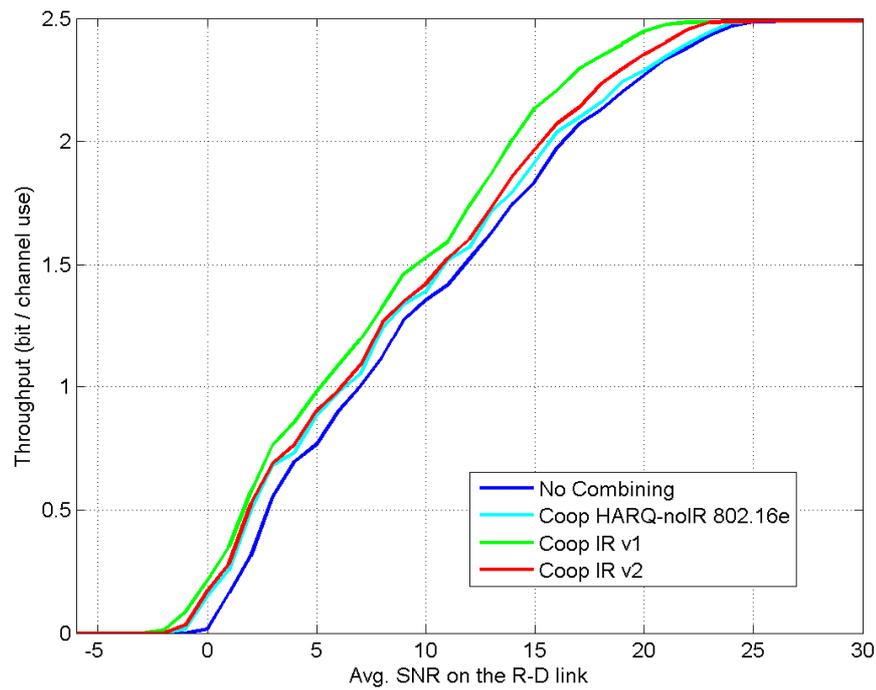
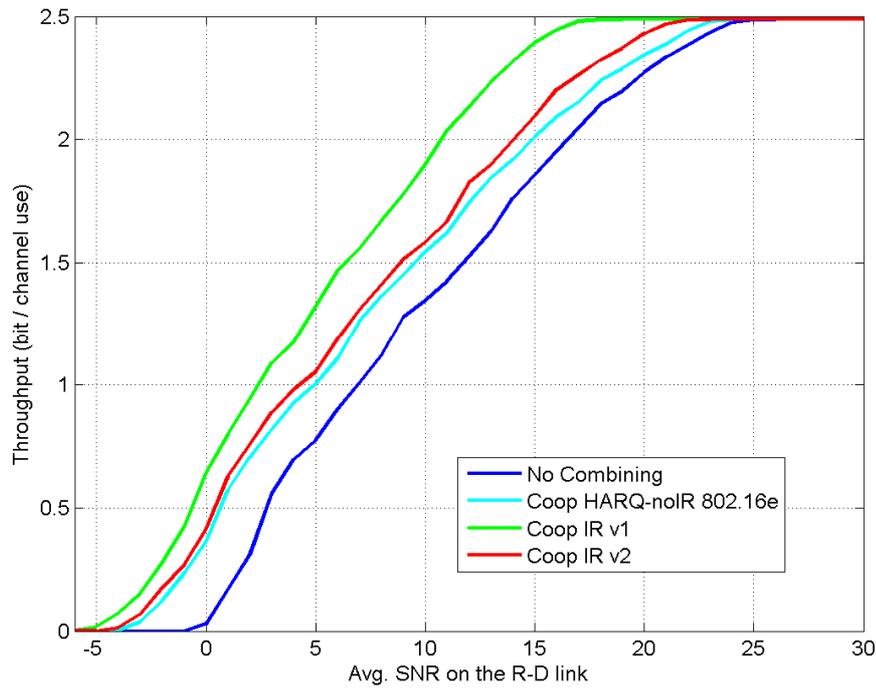


Figure 30: Average Throughput vs SINR γ_2 on R-D link of cooperative IR strategies, assuming $\gamma_1 = 30\text{dB}$ (S-R link). Top: $\gamma_0 = 10\text{dB}$ (S-D link) Bottom: $\gamma_0 = 5\text{dB}$ (S-D link)

5.2.2.4 Degraded capacity model for cooperative IR

In the previous section, we have computed the throughput performance of cooperative IR by applying the EESM. We now check whether the throughput envelope of cooperative IR can be predicted by the degraded capacity model. A straightforward extension of equation (2.15) gives:

$$\rho_{DF,P1} \approx \max_t \min(t\rho_{SR}, t\rho_{SD} + (1-t)\rho_{RD}) \quad (5.11)$$

where ρ_{SD} , ρ_{SR} and ρ_{RD} are given by equation (2.27)

$$\rho_{SD} \approx \frac{1}{N} \sum_{i=1}^N \min(\log_2(1 + \gamma_{SD,i} / \Gamma), R_{\max}) \quad (5.12)$$

where $\gamma_{SD,i}$ is the SINR on the i th subcarrier of the S-D link. We have seen in §2.2.2.3.3 that for IEEE802.16e CTC the degradation factor and maximum rate can be set to $\Gamma \approx 4\text{dB}$ and $R_{\max} = 5$.

On Figure 31, the throughput performance of cooperative IR v1 is plotted in green in the same simulation conditions as on Figure 30, top, and compared to the degraded achievable rate of equation (5.11) that is plotted in dotted black. The following can be observed:

- At low SNR on the R-D link, the degraded achievable rate is equal to the throughput of the single-hop transmission. Indeed cooperative DF Protocol I theoretically degenerates into single-hop transmission when the S-D link has higher capacity than either the S-R link or the R-D link. However in our simulations the MCS during 1st slot was selected solely based on the predicted S-R link quality. Therefore, since the S-R link is good, the source transmits using the least robust MCS and the destination cannot decode the received signal, although it contains some mutual information. This phenomenon does not correspond to an actual throughput loss in a real system: an AMC algorithm will select the best between single-hop and multi-hop transmission.
- **At high SNR on the R-D link, the peak throughput only reaches 2.5 bit/symbol compared to an expected 3 bit/symbol predicted by the degraded capacity model.** Contrary to the previous effect this 20% peak throughput loss corresponds to a fundamental limitation of the cooperative IR implementation: for a given target error rate, at low SNR, combining the coded bit LLRs received during the 1st slot with the LLRs of the signal received during the second slot

prior to decoding allows to select an MCS with a higher spectral efficiency during the second slot and to increase the throughput. However, **if at high SNR on the R-D link the MCS with the largest spectral efficiency is already used by the relay, then LLR combining cannot further increase the throughput.**

Based on the above observations, we propose the following formula for the degraded achievable rate of the cooperative IR Protocol I strategy:

$$\rho_{IR,P1} \approx \max \left(\rho_{SD}, \max_t \min \left(t \rho_{SR}, \min \left(t \rho_{SD} + (1-t) \rho_{RD}, R_{\max} \right) \right) \right) \quad (5.13)$$

This time, Figure 31 shows that the formula (5.13) matches very well the actual throughput envelope.

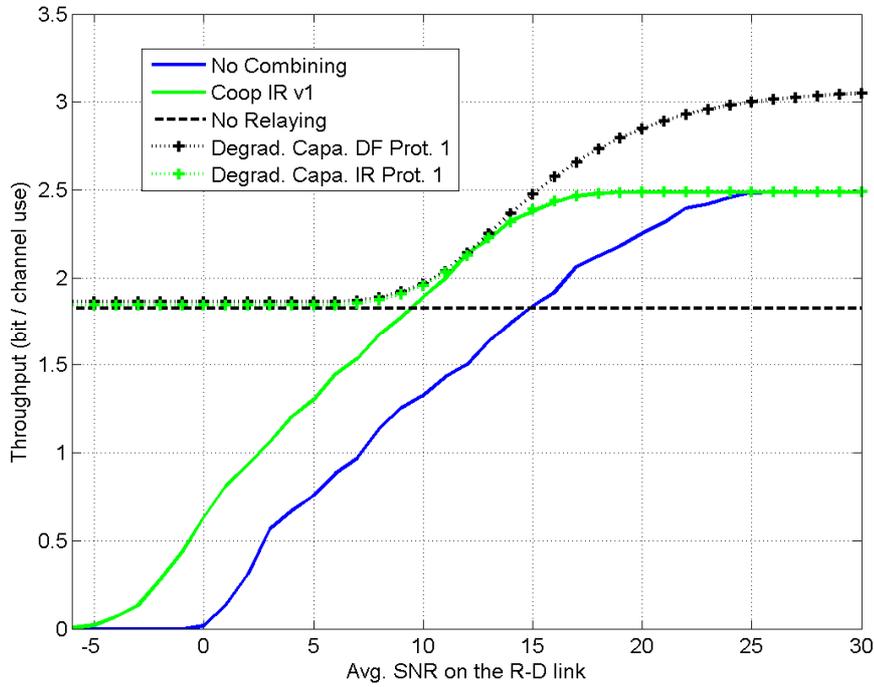


Figure 31: Comparison of cooperative IR throughput and degraded achievable rate assuming $\gamma_1 = 30\text{dB}$ (S-R link) and $\gamma_0 = 10\text{dB}$ (S-D link).

From expression (5.13) we can easily check that the peak throughput of cooperative IR cannot exceed $\rho_{\max} = (R_{\max})^2 / (2R_{\max}) = R_{\max} / 2 = 2.5$. This also corresponds to the peak throughput of non-cooperative DF. In other words, if rate saturation occurs at high SNR, the throughput of cooperative IR cannot exceed that of non-cooperative DF. We present a solution to overcome this issue in §5.2.3.

5.2.2.5 Conclusions on cooperative IR

The study conducted in this section has shown that cooperative IR allows a practical implementation of cooperative DF Protocol I. Therefore a significant throughput increase can be obtained in downlink scenarios with fixed relays, compared to non-cooperative DF. The best cooperative IR strategy consists in avoiding the retransmission of coded bits, which is consistent with the information theoretic statement that the codewords transmitted during the first and second slot shall be uncorrelated in order to maximize the mutual information. The throughput of cooperative IR for a given set of MCS can be accurately predicted by EESM and the throughput envelope can also be predicted by the degraded capacity model, provided that the granularity of the set of MCS is small enough. An interesting observation is that, at least in the single-antenna case, rate saturation can occur at high SNR even if the set of MCS allows a spectral efficiency as high as 5 bits per QAM symbol. In this case, cooperative IR cannot achieve a peak throughput increase compared to non-cooperative DF.

5.2.3 An implementation based on superposition coding

In this section we present a practical implementation of cooperative DF Protocol I with the superposition coding at the source during the first slot. This strategy has been shown in §2.1.4.1.1 to achieve a lower rate than cooperative DF Protocol I without superposition coding, so one may wonder why we want to study its implementation. It turns out that the comparison of §2.1.4.1.1 is not always valid when the rate saturation (see §5.2.1.3) phenomenon is accounted for. On Figure 23, the impact of rate saturation on cooperative DF Protocol I is illustrated. It can be observed that above a certain threshold on the average SNR γ_1 on the S-R link, the achievable rate saturates. Therefore at high γ_1 an idea that arises is to exploit the link margin on the S-R link by allocating the excess power to a signal that is superimposed onto the message transmitted to the relay. This superimposed signal can convey a message transmitted directly to the destination.

5.2.3.1 Coding strategy

We propose the following practical implementation of cooperative DF Protocol I with superposition coding:

1. If the target PER on the BS-RS link can be achieved by the MCS with the highest nominal rate, then superposition coding can be considered. The source S

will send a signal $\mathbf{x}_s = \sqrt{\delta P_s} \mathbf{x}_1(\omega_1) + \sqrt{(1-\delta)P_s} \mathbf{x}_2(\omega_2)$ where ω_1 is the 1st layer message, ω_2 is the 2nd layer message, and δ is the fraction of the transmit power allocated to the 1st layer.

2. The 2nd layer message is always mapped onto the MCS with the highest rate. The fraction δ is increased under a target PER constraint for the 2nd layer message after successive decoding at the RS.
3. The 1st layer message is always mapped onto the MCS with the highest rate such that the target PER at the destination D is met.
4. The 2nd layer message decoded by R is mapped onto an MCS and forwarded to D during the 2nd hop slot such that the PER at D is below the target PER. Note that in this thesis we consider a simple strategy in which there is no successive decoding and combining at D, which would further (slightly) increase the throughput at the expense of higher decoding complexity.

5.2.3.2 Simulation Results

On Figure 32, we assume a 10dB SNR on the S-D link as on Figure 30 and compare the throughput performance of cooperative IR v1 and cooperative DF with superposition coding. As expected, cooperative DF with superposition coding outperforms cooperative IR **at high SNR on the R-D link (above 15dB) and yields an additional +20% throughput**. Of course, the peak throughput increases with the capacity of the S-D link, but when the latter increases, the relative improvement of relaying compared to 1 hop transmission reduces.

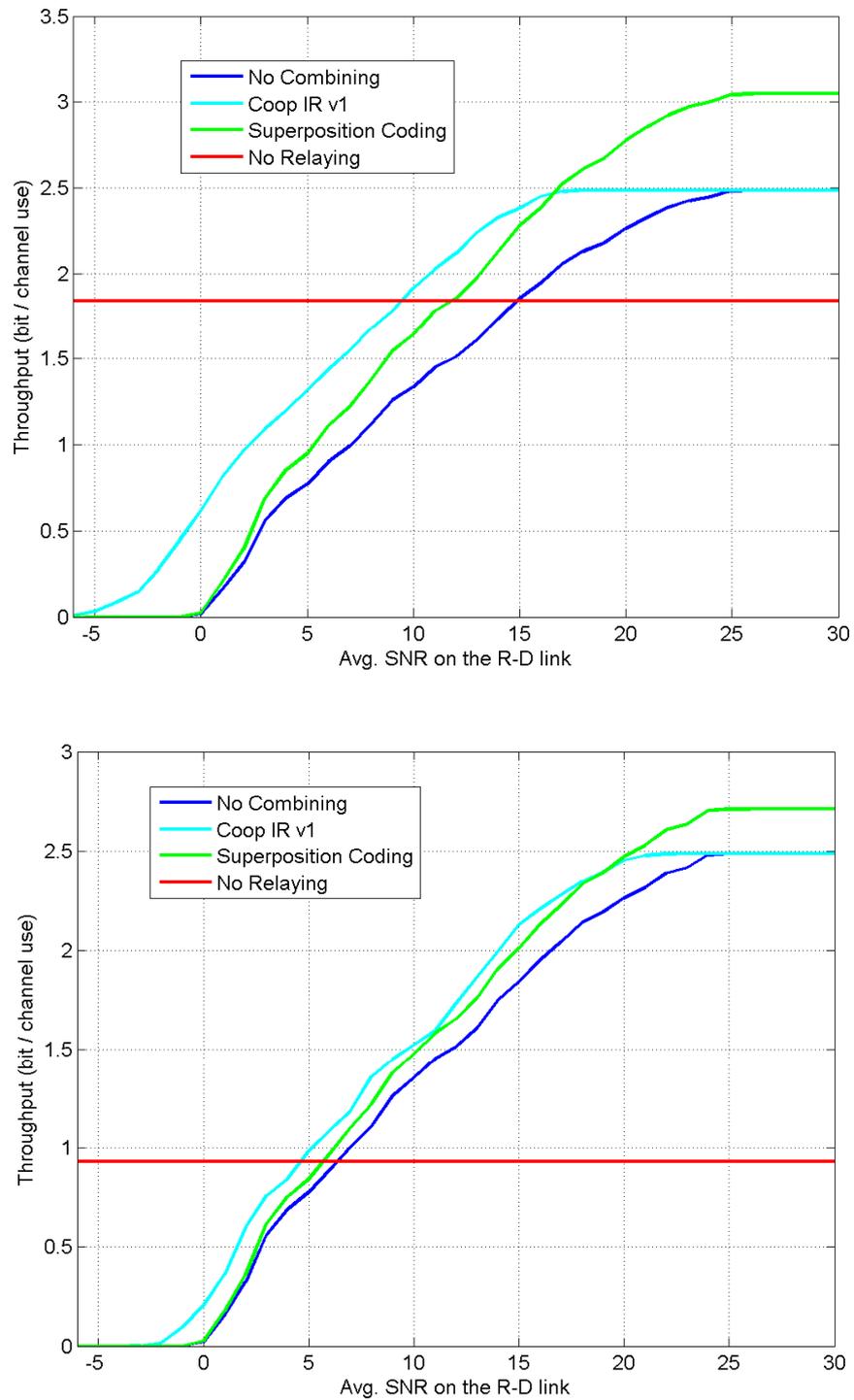


Figure 32: Average Throughput performance of cooperative DF with superposition coding and with cooperative IR v1 strategy, assuming $\gamma_1 = 30\text{dB}$ (S-R link). Top $\gamma_0 = 10\text{dB}$ (S-D link) Bottom: $\gamma_0 = 5\text{dB}$ (S-D link).

On Figure 33, it can be observed that the degraded achievable rate expression computed in equation (5.11) matches very well the average throughput envelope of the maximum between single-hop throughput, cooperative IR and cooperative DF with superposition coding. This shows that the superposition coding allows to alleviate the rate saturation problem at high SNR.

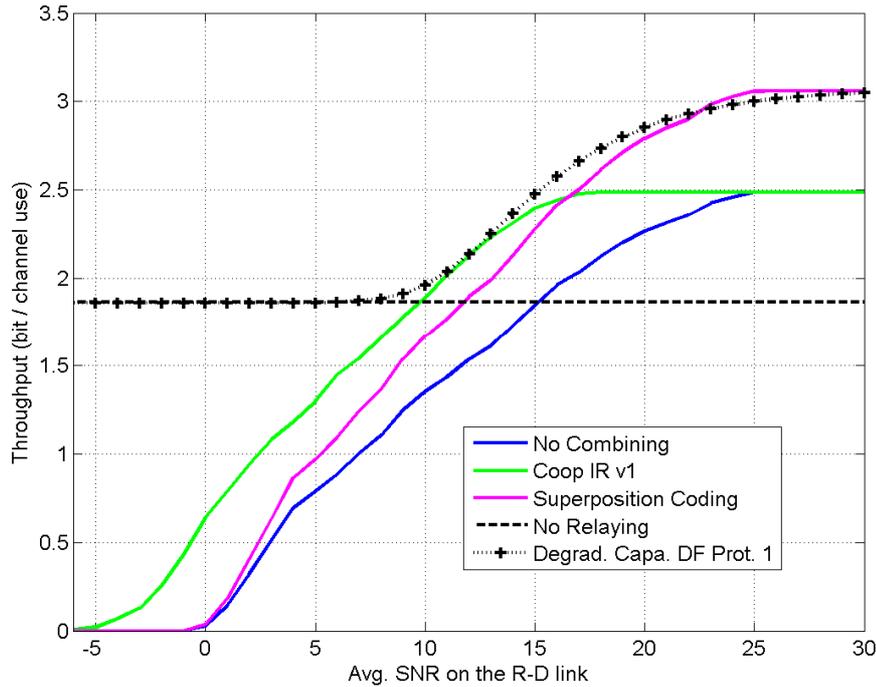


Figure 33: Comparison of average throughput and average degraded achievable rate (Same assumptions as Figure 32, top)

5.2.4 Conclusions and perspectives

The achievable rate gains predicted by information theory for cooperative DF can be realized in practical systems by combining techniques such as Incremental Redundancy and Superposition Coding. We studied in this section the case of a single-antenna system based on the IEEE802.16e standard. Its throughput performance was analyzed by both the EESM and the degraded achievable rate models and both were shown to be applicable when applied cautiously. Throughput simulations showed that cooperative IR is subject to a rate saturation problem at high SNR which can be overcome by combining it with superposition coding. The extension of these practical coding

techniques to the multiple-antenna case should be (relatively) straightforward. For instance soft-output sphere detectors can provide the LLRs required for reliability combining. However, it would be interesting to study for instance whether or not the rate saturation problem which we identified in the single-antenna case remains a serious issue in the multiple-antenna case.

5.3 Practical implementation of Compress-and-Forward strategies

In the previous section, we have investigated the practical implementation of DF and we have validated the fact that the achievable rate performance predicted by information theory can be approached in a real system. We now adopt a similar approach for CF. However for the sake of brevity we will limit ourselves to key points that are specific to CF.

5.3.1 Constrained Capacity Bounds

5.3.1.1 MIMO-OFDM transmission

The MIMO results presented in Chapter 4 can be easily extended to MIMO-OFDM by considering block-diagonal frequency-domain channel matrices. For instance the channel matrix on the S-D link can be written as:

$$\mathbf{H}_0 \triangleq \text{diag}(\mathbf{H}_{0,i})_{i=1,\dots,N_C} \quad (5.14)$$

where $\mathbf{H}_{0,i}$ is the $N_D \times N_S$ channel matrix on the i th subcarrier out of a total of N_C subcarriers. From parallel channel arguments (see e.g. 10.4 in [CT91]), the signal transmitted by S on different subcarriers can be assumed uncorrelated and therefore the conditional covariance will also be block-diagonal:

$$\mathbf{R}_{R/D}^{(1)} = \text{diag}(\mathbf{R}_{i,R/D}^{(1)})_{i=1,\dots,N_C} \quad (5.15)$$

The CKLT can therefore also be expressed as a block-diagonal matrix:

$$\mathbf{R}_{R/D}^{(1)} = \mathbf{U} \text{diag}(\mathbf{s}) \mathbf{U}^H \quad \text{with } \mathbf{U} = \text{diag}(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_{N_C}) \quad (5.16)$$

Therefore a per-subcarrier CKLT \mathbf{U}_i^H can be defined. It must be applied to the vector of length N_R formed by stacking the i th Discrete Fourier Transform output for each antenna at the relay. Achievable rates are now obtained by summing independent contributions over the N_C subcarriers. The optimum WZ coding rates can still be computed from Proposition 4.3, with the summation index i now ranging from 1 to $N_C N_R$, reflecting the

fact that the total rate R_1 available at the relay to quantize its observation shall be shared between all the spatial eigenmodes over all the subcarriers.

5.3.1.2 Practical Wyner-Ziv coding

The CF achievable rates are derived in Chapter 4 assuming an ideal WZ coding applied to each CKLT output. Practical WZ coding implementations are reviewed in [XLC04]. The WZ encoder consists in a quantizer followed by a Slepian-Wolf (SW) encoder. In order to approach the rate-distortion trade-off, the quantizer shall be rate-distortion⁵ approaching and the channel code used for SW coding shall be capacity-approaching. In [LSX05][XLC04], practical WZ coding of a Gaussian source operating at less than 0.5dB from the rate-distortion curve is obtained by combining LDPC-based SW Coding with trellis-coded quantization. Therefore, quasi-ideal rate-distortion coding is a realistic assumption from an implementation standpoint. If an implementation is considered which performs at more than say 1dB from the rate-distortion trade-off, it could be interesting to study whether a model similar to degraded capacity can be found. Intuitively, one possibility could be to introduce a compression noise degradation factor to model the non-ideal WZ coding.

5.3.1.3 Source coding without side information

The rate-distortion trade-off for gaussian vectors without side information is given in section 13.3.3 of [CT91]. The relay shall apply a KLT followed by rate-distortion (without side-information) encoding of each output. In this case, the rate allocation of Proposition 4.3 cannot be applied, but the total distortion can still be minimized by reverse-waterfilling. An even simpler source coding which does not even require CSI consists in applying per antenna quantization without any linear transform. In this case, an OFDM signal can be quantized directly in time-domain, which is not possible for techniques exploiting CSI. Overall these source-coding strategies result in a simpler implementation but may lead to a large reduction of the achievable rate as suggested by simulations in §4.3.5.1.

⁵ Here, for the quantizer we are referring to the rate-distortion trade-off without side information.

5.4 Deployment aspects

Up to now in this thesis we have studied the relay channel at the link level. However, the key question that equipment manufacturers and operators would like to answer can be summarized as follows: what is the benefit of deploying fixed Relay Stations as a complement to cellular Base Stations? The benefit shall be ultimately expressed in the framework of a business model. In [FIR08] we attempted to bridge together technical results obtained from system-level simulations with a network cost model in order to address this very complex question. In order to keep the focus of this thesis on cooperative MIMO coding strategies we will not enter into such considerations in this chapter, but only provide a few simulation results which illustrate connections between the link-level simulation results and the system-level performance. We will start by describing a simple simulation methodology to study the effect of deployment topology on the throughput, before analyzing results in single-cell and multi-cell environments.

5.4.1 Single-cell simulation methodology

We start by considering a single-cell deployment as represented on Figure 34.

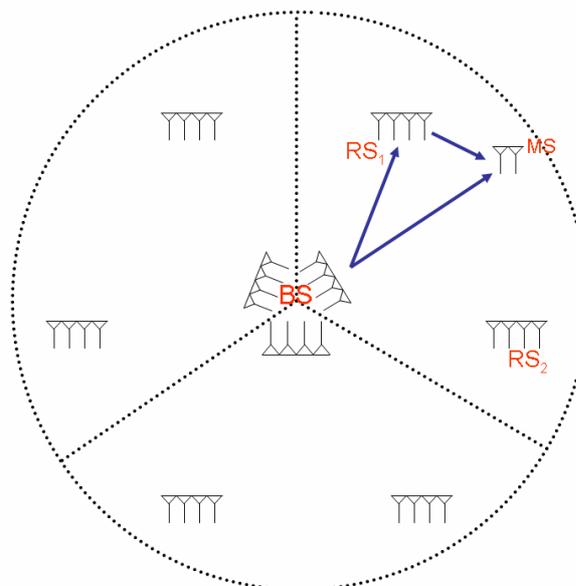


Figure 34: Single-cell deployment topology in a downlink with 4x4x2 antenna configuration

5.4.1.1 Simulation Parameters

We consider a 3-sector BS equipped with 2 to 4 antennas per sector. The RSs are not sectorized, and have 2 to 4 omnidirectional (in azimuth) antennas of 3dB gain. We position one or two RSs per sector, at a distance equal to 80% of the cell range. The BS-RS link is assumed LOS, but the BS-MS and RS-MS are NLOS. An OFDM system is assumed with N_C data sub-carriers. The size of the FFT is N_{FFT} , so that the bandwidth is approximately $B \approx N_C / (N_{FFT} T)$ where $1/T$ is the sampling rate. A cyclic prefix of N_{CP} samples is accounted for in the throughput calculations. The total OFDM symbol duration is $T_S = (N_{FFT} + N_{CP})T$. The path loss, shadowing and fast fading are given by the SCME typical urban model [B05]. Moreover we introduce a shadowing correlation between two links originating from different sites (e.g. two different BSs) but ending at the same MS. Likewise, in the uplink we assume shadowing correlation between the links originating from a given MS. This correlation coefficient is set to 0.5. This reduces the benefits of macro-diversity.

These parameters are summarized in the following table:

Parameter	Value
Cell Radius (m)	Defined by cell-edge coverage probability
BS to RS distance	$d_{BS-RS} = 0.8 r_{cell}$
Antenna Gain (dBi)	$G_{RS} = 3 ; G_{MS} = 1 ; G_{BS} = 15 - 12 \left(\frac{\theta}{\theta_{3dB}} \right)^2$ $\theta_{3dB} = 90^\circ$
Number of Antennas	$N_{BS} = 2$ to 8 per sector; $N_{RS} = 2$ to 4; $N_{MS} = 1$ to 2
Path Loss Exponent	$n_{NLOS} = 4.05$; $n_{LOS} = 2.6$
Shadowing (std. dev. dB)	$\sigma_{S,NLOS} = 10$; $\sigma_{S,LOS} = 4$
Transmit Power (dBm)	$P_{BS} = 40$; $P_{RS} = 36$; $P_{MS} = 23$
Cyclic Prefix Overhead	$N_{CP} / N_{FFT} = 1/8$
Data Carrier Ratio	$N_C / N_{FFT} = 3/4$
Channel Bandwidth (MHz)	$B = 10-20$
Noise PSD (dBm/MHz)	$N_0 = -113.9$
Noise Figure (dB)	$F_{BS} = 4$; $F_{RS} = 4$; $F_{SS} = 7$
Carrier Frequency (MHz)	$F_c = 2500$
Signal to RF Impairments Ratio (dB)	$\left(\frac{C}{I} \right)_{RF} = 30$

Table 1: System Simulation Parameters

Note that:

- For the sake of brevity we limit our system simulations to the downlink. Therefore cooperative DF will be the preferred coding strategy whereas cooperative CF will not be considered. Anyway we verified (not shown in this thesis manuscript) that cooperative CF does not provide large gains at the system-level in an in-band relaying scenario, and is better suited to out-of-band relaying and BS cooperation.
- In this single-cell scenario we neglect the impact of co-sector interference (i.e. we assume each sector operates on a different channel)
- The cell radius is defined at 75% coverage probability, i.e. such that 75% of the users experience an average SNR greater than -6dB at cell edge. This last value

takes into account the fact that the BS will broadcast the beacon in a robust Modulation and Coding Scheme, using Cyclic Delay Diversity, which should allow correct reception at low SNR.

- We locate the RS at 0.8 times the cell radius. At this distance, the SNR on the BS-RS link is limited only by the RF impairments (i.e. to about 30 dB) because of the LOS propagation. Preliminary simulation results (not shown in this thesis for brevity) suggested that this BS-RS distance is the best if the relays are deployed in order to increase the throughput within the range of the BS. However, if we wanted to study a scenario in which the RSs are deployed to extend the range of the cell beyond the BS coverage, then the best BS-RS distance would be larger than (e.g. 1.2 times) the cell radius. With the selected BS-RS distance and the adopted path loss model, we have an average SNR_{RS-MS} 6dB larger than SNR_{BS-MS} at cell edge on the BS-RS axis.
- Although we allow up to 2 relays per sector, we do not consider coding strategies for the diamond topology but only for the 3-node relay channel.

5.4.1.2 Quasi Monte-Carlo simulation methodology

Very (too?) often in the literature on relaying the emphasis is put on the spatial diversity provided by the independence of small-scale fading on the BS-MS and RS-MS links. However, in a broadband system, shadowing plays a more important role than small-scale fading. Indeed, in the SCME channel model the log-normal shadowing standard deviation is as high as 10 dB for NLOS links. In our simulations we either compute the average achievable rate of the CDF of the achievable rate over a large number of realizations of both small-scale and large-scale (a.k.a. shadowing) fading. Such simulations can be very computationally intensive if Monte-Carlo (MC) simulation methodology is used. Therefore, we investigated the application of Quasi Monte-Carlo (QMC) [N92] methodology to our simulations. QMC is a generic tool that to our knowledge is not commonly used in communications but is well known in other domains such as financial mathematics. For illustration purpose, let apply the principle of QMC to the generation of log-normal shadowing in the context of our simulations. In the MC methodology, for a given location in the cell, N_{trial} i.i.d. real Gaussian vectors $\mathbf{s}_i \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_3)$ are generated to model the shadowing on the 3 links. The components of \mathbf{s}_i can then be correlated (to model site-correlation) but different trials are always

uncorrelated $E[\mathbf{s}_i \mathbf{s}_j^H] = \mathbf{0}$. The first and second order moments of the performance function that we want to monitor (here the achievable rate) are computed:

$$m_{\mathcal{F}} \approx \frac{1}{N_{\text{trial}}} \sum_{i=1}^{N_{\text{trial}}} \mathcal{F}(\mathbf{s}_i) \quad v_{\mathcal{F}} \approx \frac{1}{N_{\text{trial}}} \sum_{i=1}^{N_{\text{trial}}} (\mathcal{F}(\mathbf{s}_i) - m_{\mathcal{F}})^2 \quad (5.17)$$

In contrast, in QMC the i th trial is correlated to the previous trials $1, \dots, i-1$ in order to accelerate the convergence of the estimators (5.17) of the first and second order moments. The QMC relies on so-called Low-Discrepancy Sequences (LDS). An LDS is a sequence of deterministic vectors which is designed to quickly “explore” the domain of a random vector by avoiding the generation of points \mathbf{s}_i and \mathbf{s}_j , $j \neq i$ which are too close to each other (w.r.t. Euclidean distance). The LDS must preserve the randomness properties of the sequence, i.e. the estimated cross-correlation of the components of \mathbf{s}_i must be close to zero and the estimators (5.17) must become unbiased after a sufficiently large number of trials, as if a classical pseudo-random generator was used. In order to clarify this very empirical definition of LDS, we illustrate it in the two-dimensional case with $N_{\text{trial}} = 10000$ on Figure 35. It can be observed that the pseudo-random sequence contains “clusters” of points which are very close to each other and for which the value of the function $\mathcal{F}(\mathbf{s}_i)$ is approximately constant if the function is continuous. In contrast the LDS avoids the occurrence of such clusters of points. From Figure 35, it is obvious that LDS can be used to generate random uniformly-distributed user locations in a system simulator such that at all the possible locations of the deployment are explored in the minimum possible number of trials. Of course, the uniform LDS can be transformed into a Gaussian distribution or any other useful distribution of interest and therefore LDSs can also be used to generate the shadowing process or even the fast-fading process. In our simulations we used Sobol’s [S77] sequences but many other LDSs have been proposed in the literature. For the sake of brevity, we will not enter into a detailed discussion on LDSs and QMC. However, we insist on the fact that **when the number of dimensions of the random vector is small** - which is the case in 3-node shadowing simulation - **the QMC methodology leads to a tremendous simulation complexity reduction. In our simulations, we observed a complexity reduction by a factor of 10** for the same accuracy of the achievable rate estimates.

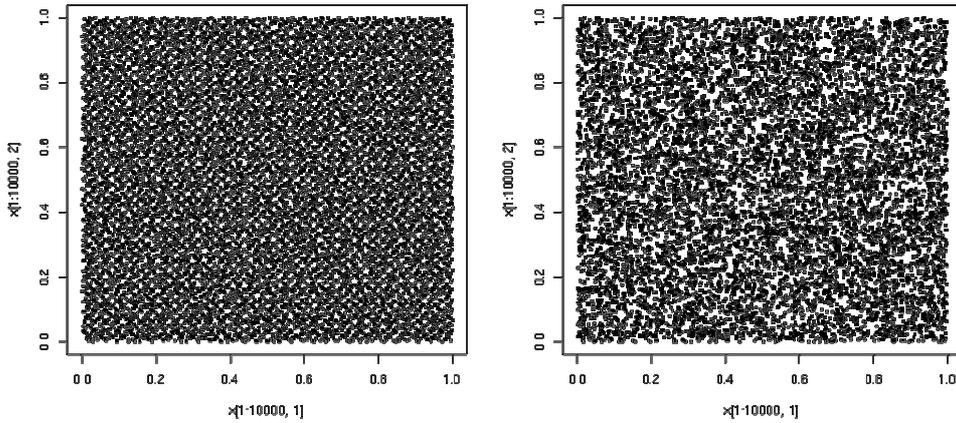


Figure 35: Generation of 10000 points uniformly distributed in $[0;1] \times [0;1]$ (left: Low Discrepancy Sequence, right: Matlab Pseudo-Random generator)

5.4.2 Simulation results with 1 relay per sector

On Figure 36, we plot the average rate vs. the MS location in the cell for direct link (left) and non-cooperative DF. The peak rate around the BS is large and users can be served at large spectral efficiency (up to 12 b/s/Hz thanks to the 2 antennas) but most of the cell can only be served at an average spectral efficiency lower than 2 b/s/Hz. The weakest coverage is at inter-sector border, because we did not consider sector cooperation. The addition of one RS per sector creates “hot-zones” around which large spectral efficiency is achievable (up to 7 b/s/Hz) but due to the fact that the RS has lower power than the BS and is equipped with omni-directional antennas, the range of these hot-spots is limited.

On Figure 37 and Figure 38, we plot the average achievable rate with cooperative DF and the cooperation gain, defined as the ratio of the achievable rate using cooperative techniques to the achievable rate using non-cooperative techniques. It can be observed that the coverage is improved on the BS-RS axis, but the coverage of the inter-sector border remains insufficient. Cooperation yields largest gains at cell edge on the inter-sector border (this is where the macro-diversity yields largest gains), but also around each RS. Indeed, when the MS is close to the RS, the duration of the 1st slot is not negligible compared to that of the 2nd slot, which is in favour of cooperative DF Protocols I and III, compared to Protocol II as shown on Figure 38. Note that even when the MS is far from

the BS and the RS, Protocol II cannot outperform Protocol I, mainly because CSIT is not available in this simulation. It can also be checked that the deployment of only 1 relay per sector is not sufficient to guarantee homogeneous coverage.

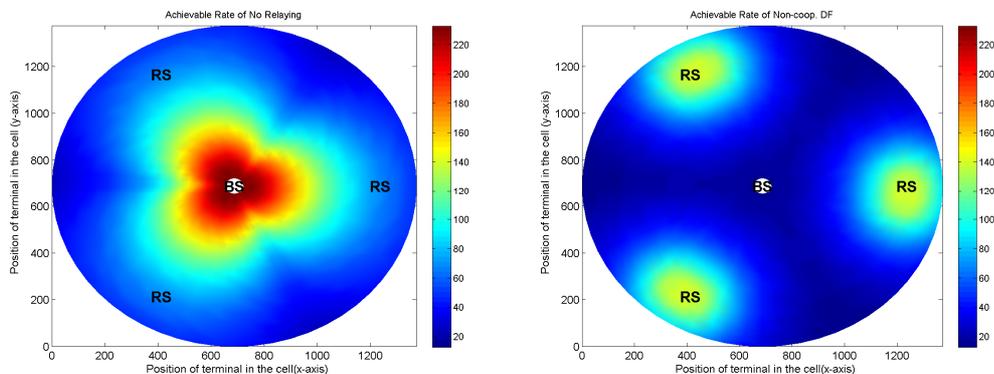


Figure 36: Average capacity of Direct Link (left) and 2-hop non-cooperative DF (right) vs. MS position in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$.

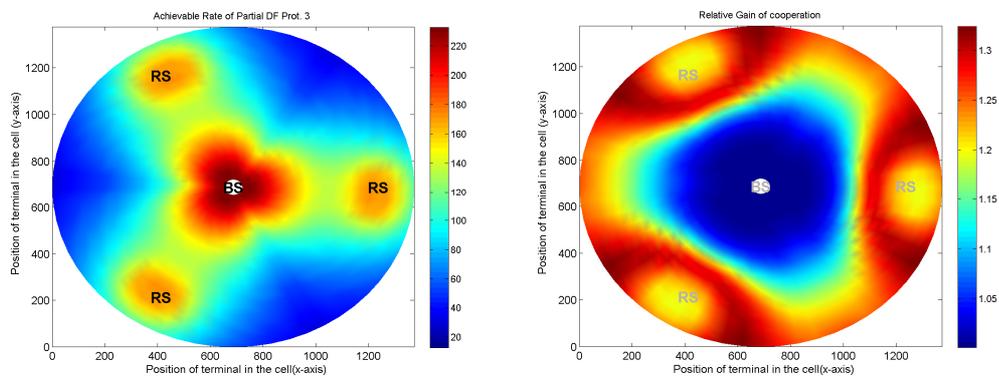


Figure 37: Average rate of cooperative partial DF Protocol III (left) and cooperation gain vs. MS position in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$.

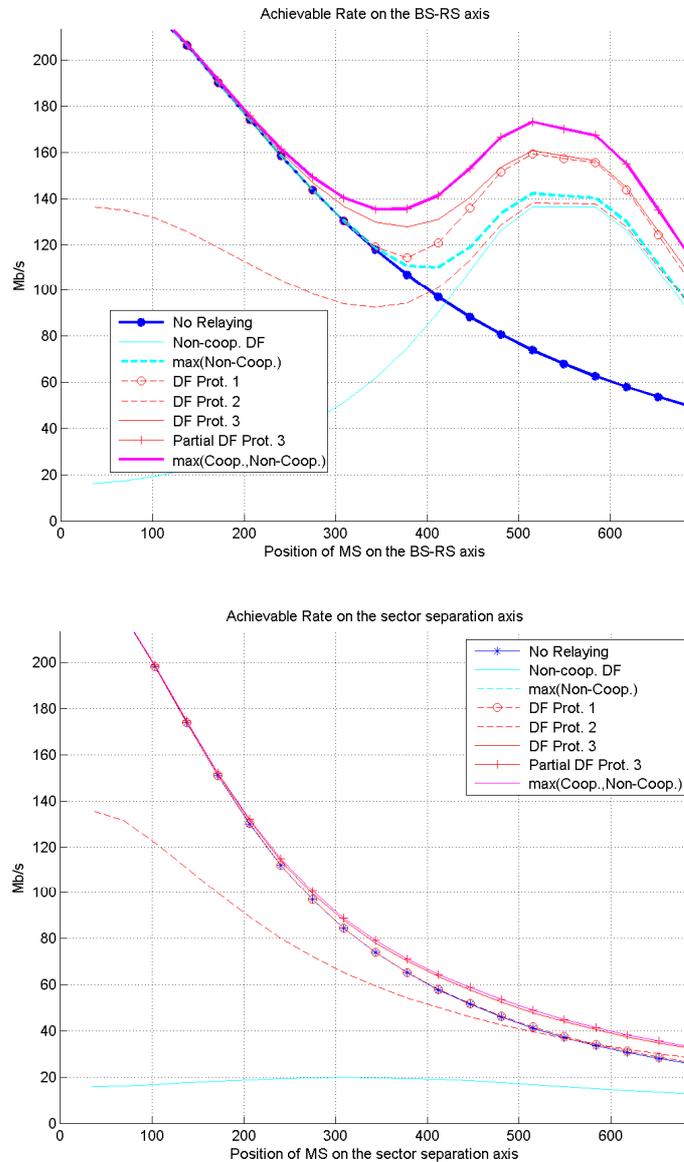


Figure 38: Average capacity of various strategies vs. MS position on the BS-RS axis (top) and inter-sector separation axis (bottom) in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with CSIR only, $B=20\text{MHz}$.

5.4.3 Simulation results with 2 relays per sector

The deployment of 2 RS per sector significantly improves the coverage. We illustrate this in the downlink in the following.

5.4.3.1 Cooperative DF vs. non-Cooperative DF

With two RS per sector, the average spectral efficiency at cell edge on the sector separation axis is improved by a factor of 2.2, when using NCDF, and by a factor of 2.8 when using PDF Protocol III. It can be observed that the cooperation gain is around 1.2 in most of the cell (except around the BS). In terms of average spectral efficiency over the cell, assuming a uniform user distribution, direct link yields 4 b/s/Hz, while relaying yields 5 b/s/Hz, and cooperative relaying (PDF Protocol III) 6 b/s/Hz, i.e. a 20% increase.

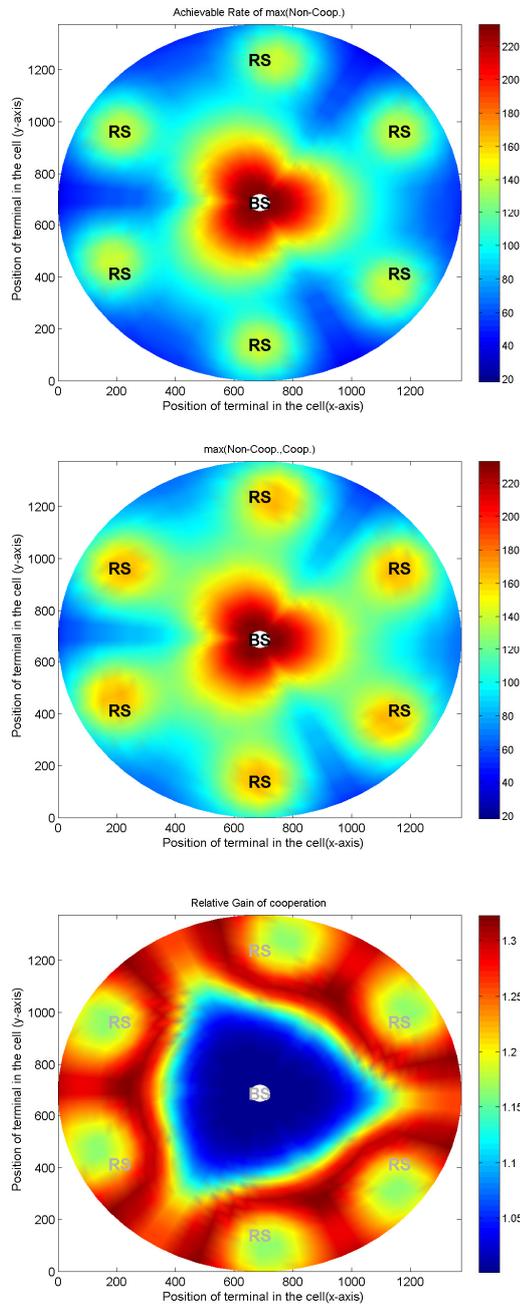


Figure 39: Coverage Improvement by Cooperative DF. Top: Average achievable rate without cooperation vs. MS position; Center: Average achievable rate with Cooperation vs. MS position; Bottom: Cooperation gain vs. MS position, $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, CSIR only, $B=20\text{MHz}$.

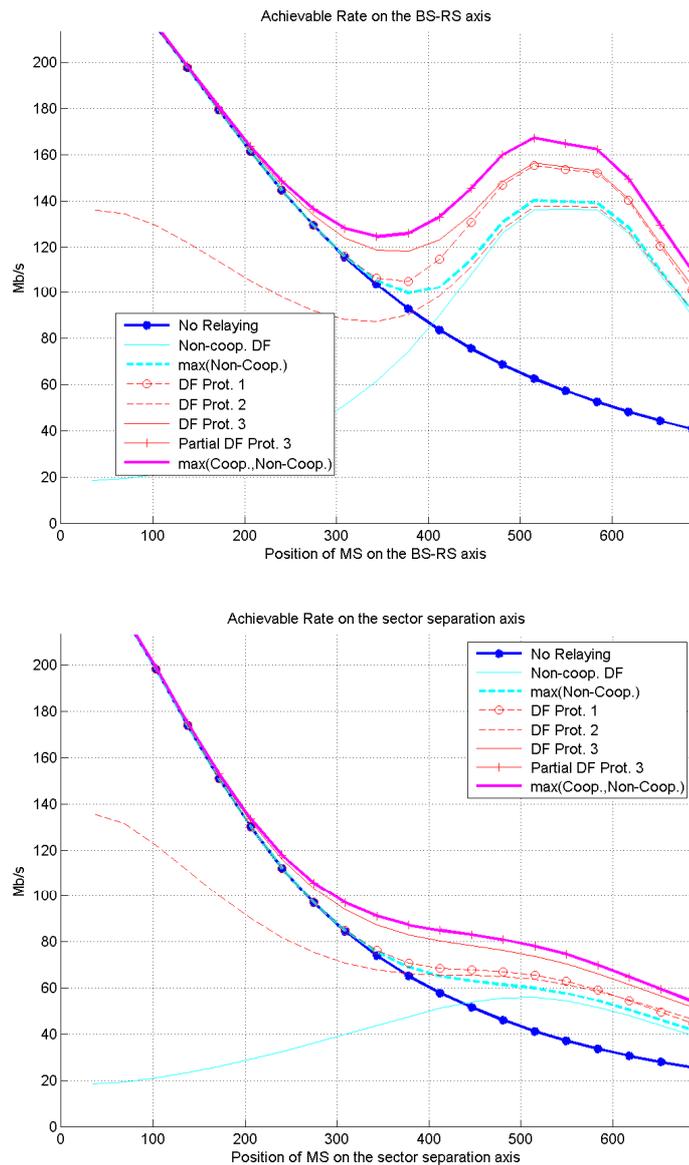


Figure 40: Average achievable rate of Cooperative DF vs. MS position on BS-RS axis (top) and on sector separation axis (bottom) in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, CSIR only, $B=20$ MHz.

The impact of a large number of antennas at the BS and RS on the downlink single-user average capacity with CSIR only is limited, as illustrated on Figure 41 where the average capacity is plotted for a configuration with 2 antennas at BS, RS and MS. The only effect of having 4 antennas at BS and RS is a higher capacity in the immediate vicinity of the RS, due to the larger capacity of the BS-RS link.

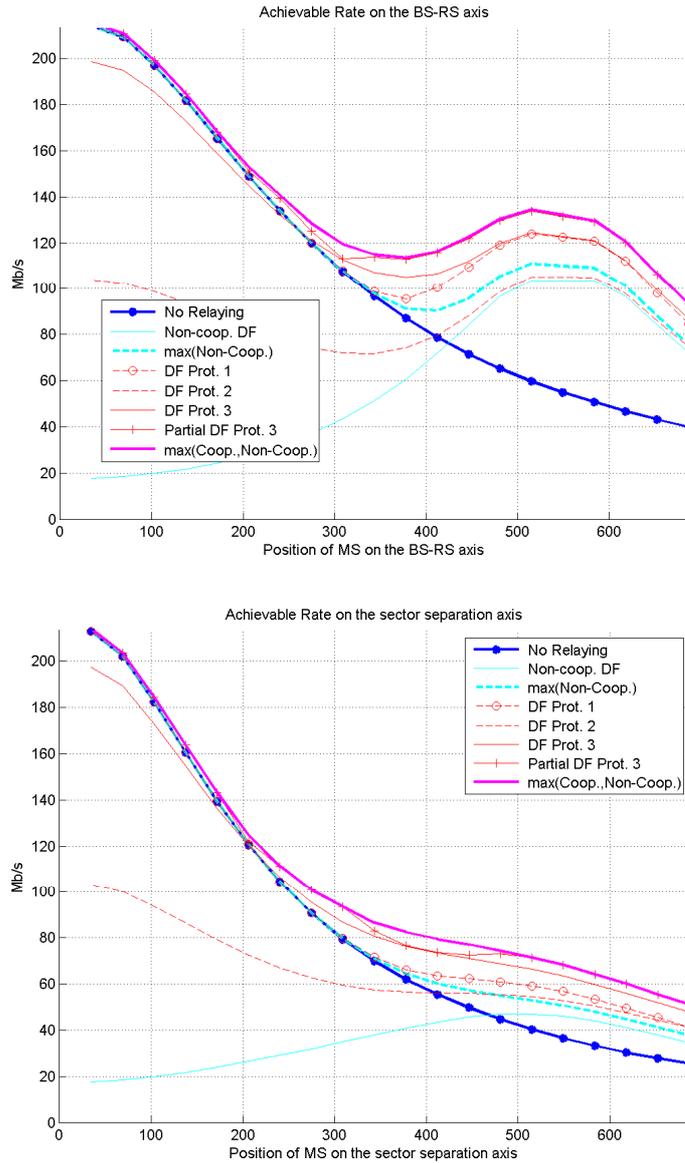


Figure 41: Average capacity on BS-RS axis (left) and on sector separation axis (right) in a $N_{BS} = N_{RS} = N_{MS} = 2$ antenna configuration, no CSIT, $B=20\text{MHz}$, 2 relays/sector.

5.4.3.2 Impact of Shadowing

5.4.3.2.1 What if the relay is in NLOS with the base station?

An important assumption which is made throughout this thesis is that the BS-RS link benefits from a high SNR. This will be true if the RS is in LOS. However, especially in urban environment, it may happen that the only available relay sites are in NLOS with the BS. From the SCME path loss model and in our simulation scenario if the RS is located at

0.8 times the cell radius, then $SNR_{BS-RS} \approx 10dB$ if only the distance-dependent path loss is taken into account. This is a priori not enough to significantly increase the capacity in the cell. However, if we assume that shadowing is essentially due to fixed components of the environment such as buildings, then the shadowing shall remain almost constant during the relay lifetime. In this case, statistically it should be feasible to find relay locations where the path loss including shadowing $L(d)+S$ is for instance equal to $L(d)+\sigma_s$. Since $\sigma_s = 10dB$ in NLOS, this would lead to $SNR_{BS-RS} \approx 20dB$, which is enough for efficient relay operation. Moreover, since the NLOS multipath channel is more spatially rich a large capacity could be achieved in 4x4 antenna configurations.

5.4.3.2.2 Impact of shadowing correlation from MS to RS and BS

On Figure 42, we illustrate the impact of the shadowing correlation on the links from a given MS to the various RS and BS, by plotting the average capacity guaranteed with 90% coverage probability, as a function of the MS location on the BS-RS axis. When the correlation is low (uncorrelated shadowing on top figure), then a MS experiencing severe shadowing on the BS-MS link can switch to 2-hop forwarding if its link to one of the RS benefits from better shadowing conditions. On the contrary, if the shadowing is highly correlated on all links (e.g. correlation of 0.5 on the right figure), then its effects cannot be mitigated by relaying. However, notice that cooperative relaying still improves the rate w.r.t. non-cooperative relaying. When averaged over the cell, the average spectral efficiency at 90% coverage probability is 1.7b/s/Hz for direct link, vs 2.5b/s/Hz Mb/s with non-cooperative DF and 3 b/s/Hz with cooperative DF.

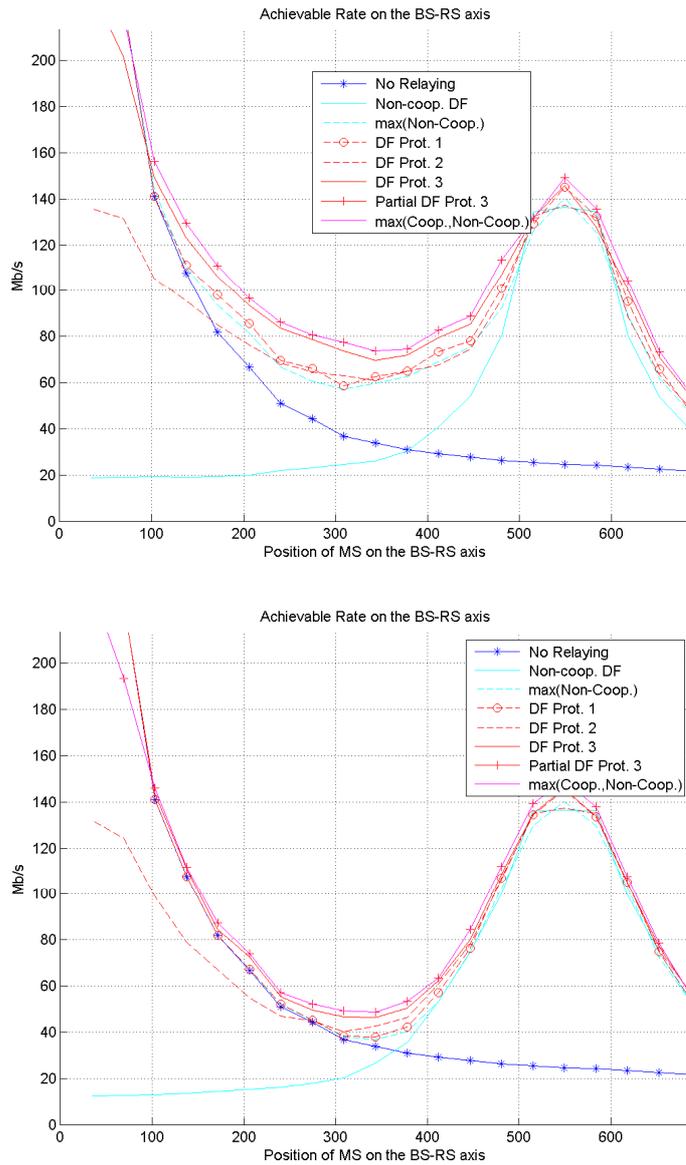


Figure 42: Impact of shadowing correlation on the achievable rate of cooperative DF (on the BS-RS axis at 90% coverage probability). Top: No Correlation; Bottom: Correlation of 0.5. $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration.

5.4.3.3 Impact of CSIT

We now optimize the transmit covariance for DF Protocols II and III with CSIT, as described in Chapter 3. The gain is high only when $SNR_{BS-MS} \approx SNR_{RS-MS}$ and both are low. This happens in areas far from the BS and RS. The largest cooperation gain is thus achieved on the inter-sector axis. On Figure 43, we see that even in the two-relay case,

non-cooperative relaying cannot significantly improve coverage if the MS is on the sector separation axis. However, if Protocol II -or even better Protocol III- is used, a large capacity improvement can be achieved. For instance in the 4x4x2 antenna configuration of Figure 43, there is a 50% achievable rate increase for Protocol III with full CSI (w.r.t. CSIR only) at a 300m distance from the BS and almost a 100% increase at cell edge.

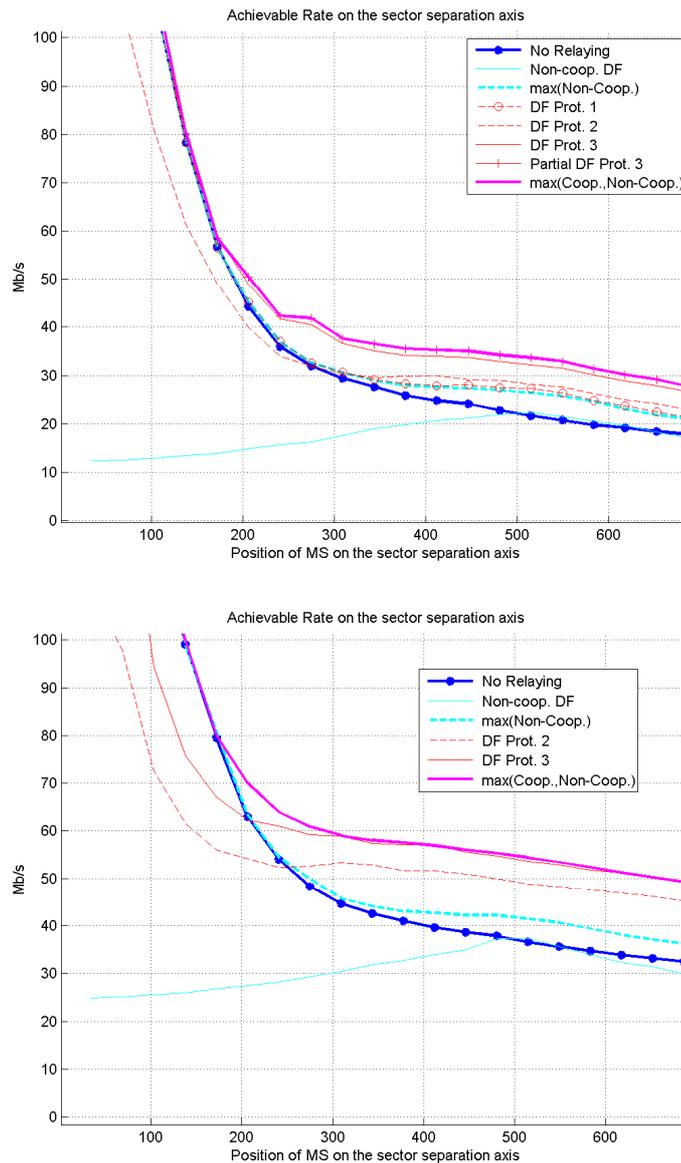


Figure 43: Effect of full CSI on cooperative DF Protocols. Average achievable rate on the sector separation axis at 90% coverage probability in a $N_{BS} = N_{RS} = 4$, $N_{MS} = 2$ antenna configuration, with 2 relays per sector. Shadowing Correlation = 0.5. Top: CSIR only; Bottom: full CSI

5.4.3.4 Conclusions from single-cell simulations

In this section, we tried to understand the effect of deployment topology and macroscopic propagation on cooperative relaying. We performed simulations in an urban micro-cell scenario, with 3-sectors at the BS and a variable number of relays per sector in LOS with the BS and located close to cell edge. When a single RS is deployed per sector, with our simulation assumptions the relay cannot significantly improve the coverage at cell edge in the sector separation area. Therefore, we consider the deployment of two relays per sector to provide a more homogeneous high rate coverage. Our simulations show that in this case the single-user spectral efficiency, averaged over the cell area, is increased by 25% thanks to non-cooperative relaying. Cooperation yields another 20% increase w.r.t. non-cooperative relaying, Protocol III outperforming other protocols. We discuss the LOS assumption between BS and RS, and conjecture that a pre-selection of relay sites shall allow a good enough link between the BS and RS. We briefly study the effect of shadowing correlation from the MS to the BS and RSs, and verify that a high shadowing correlation reduces the gain brought by relaying in poor coverage areas, but does not seem to affect the cooperative vs. non-cooperative relaying comparison, which is more related to microscopic propagation effects. Finally, we show that the exploitation of full CSI at the transmitter side benefits to all strategies (direct link, non-cooperative and cooperative DF relaying) and results in large capacity gains. Before drawing conclusions at the system-level, we will have to consider multi-cell scenarios in order to account for co-channel interference, which may change our conclusions.

5.4.4 Simulations in a multi-cell scenario

We now make our system-level simulations a bit more realistic compared to the single-cell simulations of §5.4.1. The basic system assumptions of table 1 are kept, and the antenna configuration is 2x2x2. Note that:

- We consider a multi-sector multi-cell deployment. An MS at cell edge may thus associate with a neighbouring BS that offers the best 2-hop throughput (i.e. association and routing are performed)

- The inter-cell and inter-sector interference are modeled. We consider two possible frequency reuse scenarios: 1/3/3 and 1/3/1, where the first figure denotes the number of cells in a reuse pattern, the second denotes the number of sectors and the last denotes the number of channels. In the 1/3/3 case, each sector operates on a different channel, thus there is no inter-sector interference and only inter-cell interference. In the 1/3/1 case, the inter-sector interference is the dominant source of interference.

Note that our system-level simulator still relies on some simplifying assumptions:

- Scheduling assumptions

Full-buffer traffic is assumed. Moreover, we restrict the study to DL and assume a synchronized TDD/TDMA/OFDMA system, therefore the DL interference is modeled by assuming that all other BSs on the same channel are interfering all the time. We model the interference from RSs by assuming that all the RSs and BSs of neighbouring sectors and cells transmit continuously, which is a worst-case assumption. Aggregate cell throughput is computed by averaging the single-user throughput assuming a uniform MS distribution, neglecting the MAC overhead and assuming that each DL connection is granted the same time-frequency resource (this resource includes the 1st and 2nd hop slots). The effect of granularity in the time-frequency resource allocation (due to the limited number of OFDM symbols and frequency subchannels) is neglected.

- Degraded achievable rate link-to-system interface with rate saturation

A degraded-capacity model is assumed with a 4dB degradation and a maximum rate of 5 data bits per QAM symbol. Furthermore, the minimum average SNR requirement is computed assuming a 3dB cyclic combining gain at the BS, an MRC gain of $10\log(N_R)$ dB and a 0dB SNR requirement for the most robust MCS. In the following a 2x2x2 antenna configuration is assumed, leading to a minimum SNR requirement of -6dB, below which the achievable rate is zero. The peak⁶ spectral efficiency in b/s/Hz/cell is computed assuming the 10 MHz WiMax PHY parameters for the PHY overhead (1/8 relative Cyclic Prefix duration, 720 useful data subcarriers per OFDM symbol, 23/25 oversampling factor). Finally, full CSI is assumed.

- The frequency-selectivity of the interference is not modeled.

⁶ Peak spectral efficiency means that a single-user with full-buffer traffic is served.

- Fractional frequency reuse is not modeled. Therefore, 1/3/1 deployment is severely affected by interference. Fractional frequency reuse could alleviate this interference.

5.4.4.1 Simulations with 1/3/3 frequency reuse and no relay-slot reuse

5.4.4.1.1 Multi-cell 1/3/3 deployment without relays

On Figure 44, simulation results are plotted for a given multi-cell deployment without relays and a single shadowing realization⁷. On Figure 45 the CDF of the spectral efficiency over a large number of shadowing realizations is plotted. As expected, half of the users are served at a spectral efficiency lower than 2 b/s/Hz/cell. In the multi-cell deployment scenario, MSs can associate with the BS offering the best SINR, which may not be the closest one due to the large shadowing caused by the NLOS propagation on the BS-MS link. Therefore, it can be checked by comparing the SNR CDFs on Figure 45 and Figure 46 that the multi-cell deployment achieves a 95% coverage at -6dB average SINR whereas an isolated cell would achieve only 80% coverage.

⁷ Note that the spatial correlation of the shadowing can be observed on this plot, because it represents a single trial of the shadowing.

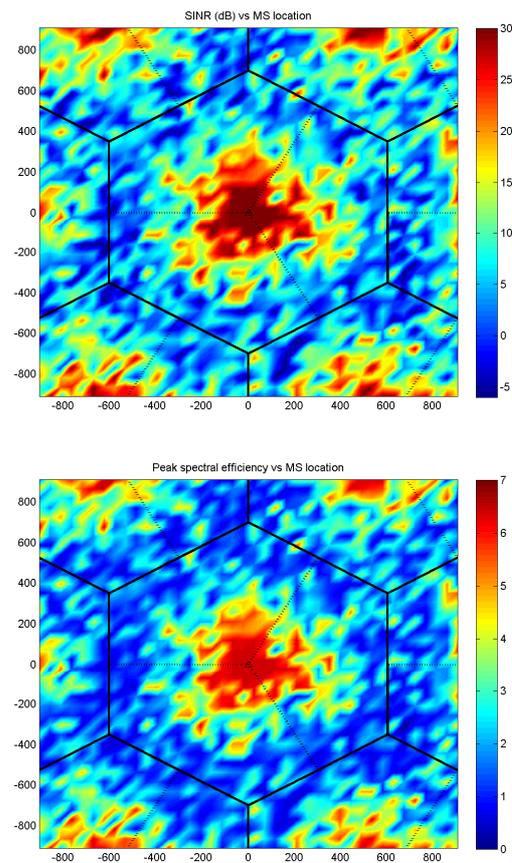


Figure 44: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/3 multi-cell deployment without relays

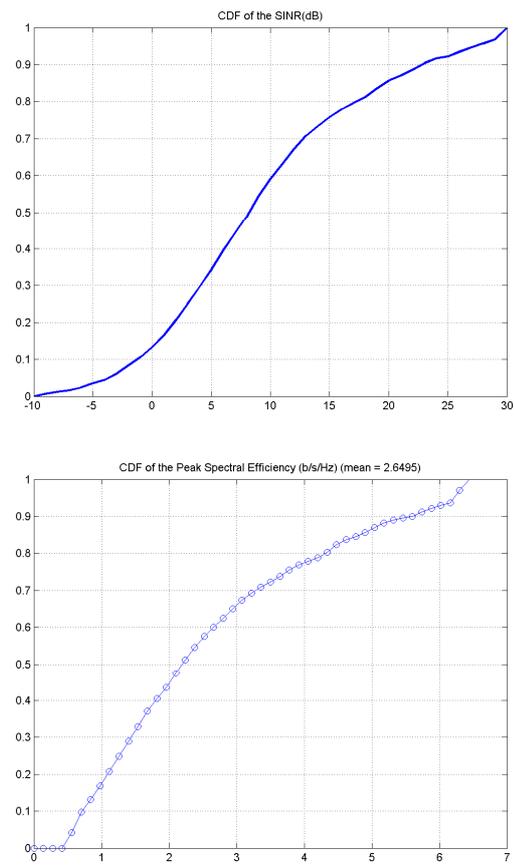


Figure 45: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 multi-cell deployment without relays

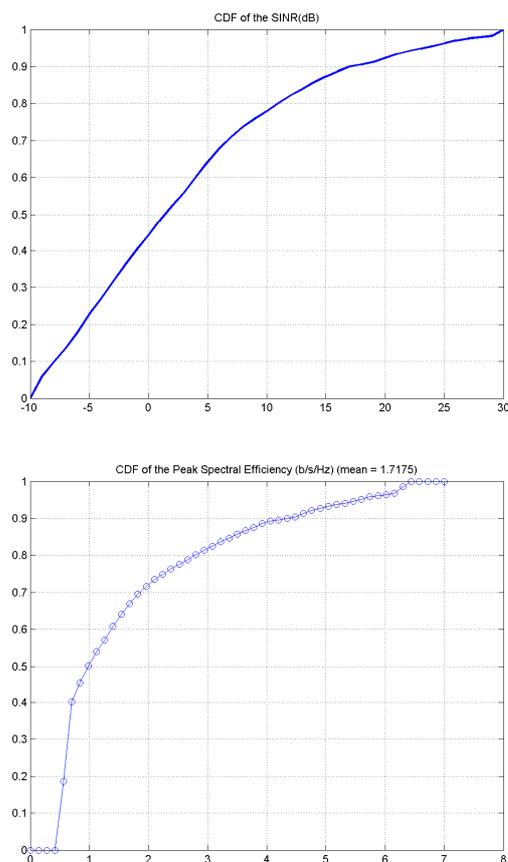


Figure 46: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 single-cell deployment without relays

5.4.4.1.2 Multi-cell 1/3/3 deployment with relays

We now assume the same BS deployment with 2 RSs per sector. Moreover, we assume that the RSs are in LOS with their main BS, but in NLOS with the co-channel neighboring BSs. This is an important assumption, because in this case the DL co-channel interference from neighbouring BSs is negligible and there is no need for a RS to associate to multiple BSs. Otherwise, it can be checked that the SINR on the BS to RS link is low (around 6 dB) and relaying gain vanishes.

On Figure 47 (top), the SINR to the closest BS or RS is plotted. As expected, hotspots are created around RSs. The peak spectral efficiency achieved by cooperative relaying with slow link adaptation between Protocols I and III is also plotted on this figure (bottom). On Figure 48, the CDF of the average achievable rate is plotted for each

strategy (the black curve represents the slow link adaptation between Protocol I and Protocol III). It can be observed that **relaying increases the throughput especially in poor coverage areas**: at 90% coverage probability, direct transmission achieves 0.7 b/s/Hz/cell, non-cooperative relaying achieves 1.2 b/s/Hz/cell and cooperative relaying achieves 1.5 b/s/Hz/cell. On average, direct transmission achieves 2.5 b/s/Hz/cell, non-cooperative relaying achieves 2.8 b/s/Hz/cell and cooperative relaying achieves 3.3 b/s/Hz/cell. **Here, thanks to cooperative beamforming (FDF Protocol III with CSIT) the throughput at cell edge is more than doubled. Non-cooperative relaying increases the average cell throughput by 15% and cooperative relaying increases it by 30%.**

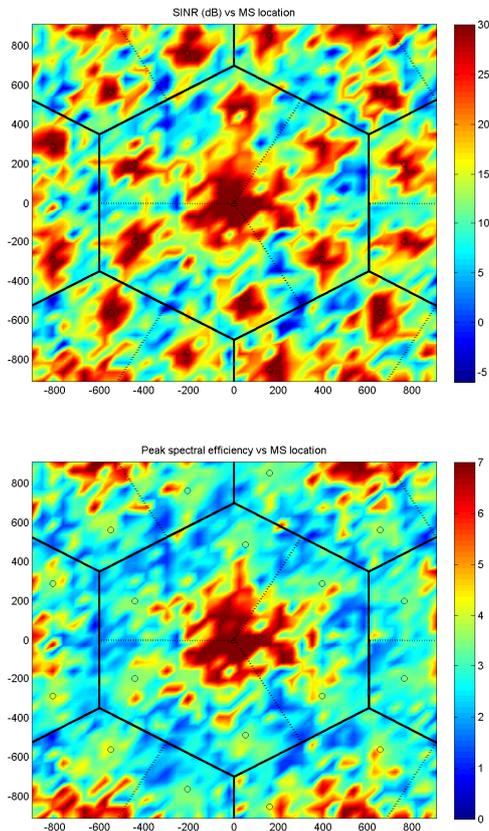


Figure 47: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/3 multi-cell deployment with 2 relays per sector

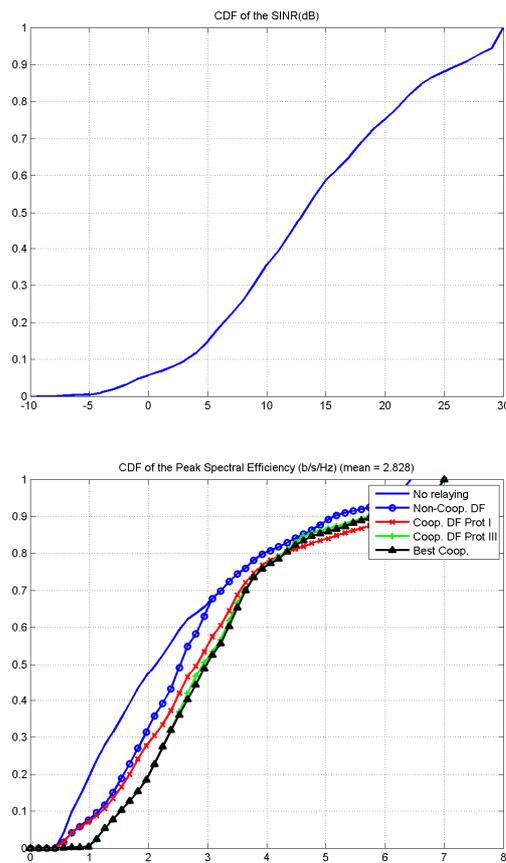


Figure 48: CDF of the SINR (top) and peak spectral efficiency (bottom) over all user locations in a 1/3/3 multi-cell deployment with 2 relays per sector

5.4.4.2 Multi-cell 1/3/3 deployment with relays and re-use of the relay slot

We now remove the constraint that two RSs transmit on orthogonal resource. In this case they interfere with each other, but two users can be scheduled simultaneously. We assume that the user density is large enough to always find a pair of users in the same area such that they request a relay slot of similar duration. In this case, the time-sharing variable can be optimized as described in section 2.2.2 of [FIR07b] for the two relay case. The resulting CDF is plotted on Figure 49. Several observations can be made:

- The interference between the two RSs does not have a big impact on the SINR distribution. This is due to the fact that in the considered scenario the relay sub-cell footprints do not overlap and also to the fact that NLOS propagation is always assumed on the RS-MS link. Clearly, reuse of the relay slot would not be possible for two MSs in LOS with the same two RSs.

- Reusing the relay slot yields a large average spectral efficiency increase: 2.4 b/s/Hz/cell for direct transmission, 3.1 b/s/Hz/cell for non-cooperative relaying and 3.6 b/s/Hz/cell for cooperative relaying. **Non-cooperative relaying increases the average cell throughput by 30% and cooperative relaying increases it by 50%.**
- **When relay slot is reused, Protocol I almost always outperforms Protocol III,** except for users at cell edge and the cooperation gain is achieved mainly for users around the RSs.

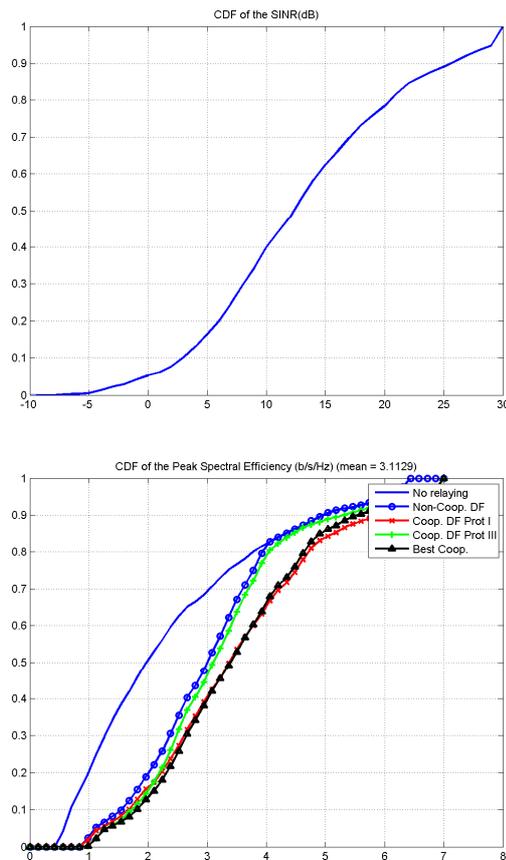


Figure 49: CDF of the SINR (top) and of the peak spectral efficiency (bottom) versus user location in a 1/3/3 single-cell deployment with 2 relays per sector

5.4.4.3 Multi-cell 1/3/1 deployment without fractional frequency reuse

In a 1/3/1 scenario, due to the high interference (fractional frequency reuse is not implemented), the SINR at the RS is only around 10dB, but this is still enough to achieve gains by relaying. The average cell throughput is therefore much lower than for the 1/3/3 scenario (Figure 50 and Figure 51). However, in terms of spectral efficiency the 1/3/1

scenario outperforms 1/3/3. The cell spectral efficiency plots still show a large gain for relaying (3.6 b/s/Hz/cell to 4.7 b/s/Hz/cell) and cooperation (4.7 b/s/Hz/cell to 5.9 b/s/Hz/cell).

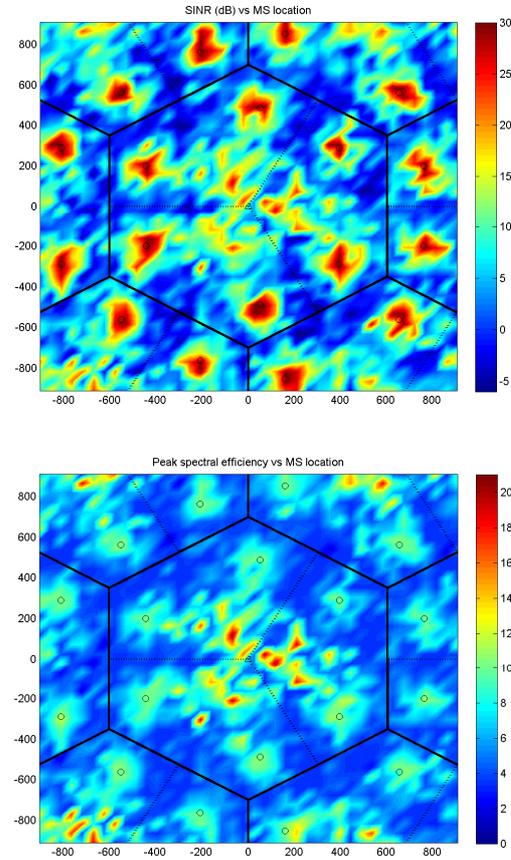


Figure 50: Average SINR (top) and peak spectral efficiency (bottom) versus user location in a 1/3/1 multi-cell deployment with 2 relays per sector

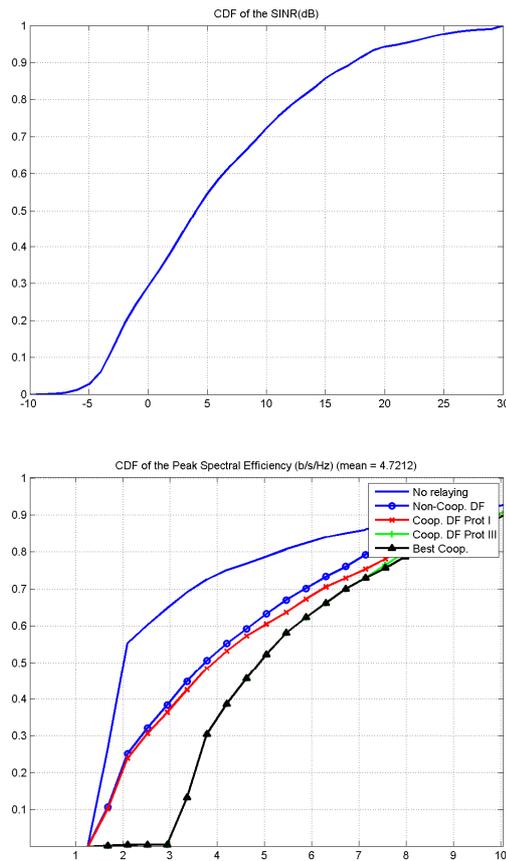


Figure 51: CDF of the SINR (top) and of the peak spectral efficiency (bottom) versus user location in a 1/3/1 single-cell deployment with 2 relays per sector

5.4.4.4 Summary and conclusions

The following conclusions can be drawn from the above multi-cell system-level simulations:

- If the relays do not have the capability to associate to multiple BSs, then they should be carefully deployed, avoiding that a RS be in LOS from two co-channel BSs.
- In both 1/3/1 and 1/3/3 scenarios a significant spectral efficiency gain is achieved by relaying and cooperation.
- Protocol III with CSIT allows a large increase of the spectral efficiency for cell-edge users (2x increase at 90% coverage)
- **Reusing the relay slot is a strategy that leads to a large increase of the spectral efficiency, resulting in a +50% cell capacity gain (30% due to relaying and 20% due to cooperation).** In this case Protocol I outperforms Protocol III for users around

the relay. **Therefore relaying protocol I is an attractive solution at the system level.**

5.5 Conclusions

A large number of issues arise when considering the practical implementation of cooperative coding strategies in future broadband wireless systems. In this chapter, we have reviewed some of them. Sometimes we only scrapped the surface and a more in-depth work would definitely be needed. However, the general observation that we can make is that the information-theoretic study is not disconnected from the real implementation but on the contrary can provide very useful tools to analyze and predict the performance of cooperative coding strategies in a real implementation.

General conclusions and possible future work

We now draw some general conclusions on the results obtained within this thesis and propose some research directions to complete or extend our work. We do not attempt to summarize again our achievements chapter by chapter, as this was already done in §1.2, but rather highlight some key take-away messages.

- Although the capacity of the three-node Gaussian relay channel remains unknown, simulations show that a combination of partial DF and CF yields an achievable rate envelope which is only a few tenths of dBs below the cut-set upper-bound on capacity. Thus we do not see a big incentive in performing research on even more advanced coding strategies that might approach even closer to the cut-set bound. This observation is in agreement with the current research trend in the IT community to focus on more complex topologies and traffic, especially on relaying with multiple hops or multiple parallel relays, the multiple access and broadcast relay channels and multi-way relaying. Another research path which we believe deserves interest is interference-aware relaying (see e.g. [ZKL08]). Indeed, on the one hand our network simulations in Chapter 5 show that the relaying strategy shall not be designed or selected in isolation of the rest of the network but on the other hand trying to design a coding strategy for a large topology often leads to prohibitive complexity, thus interference-aware relaying can be viewed as a trade-off between these two requirements.
- We decided to focus on three-node MIMO TDD relaying with full CSI because it provides useful capacity bounds for the throughput prediction of future BWA networks. We show in Chapter 3 and Chapter 4 that the cut-set bound and the achievable rates of the partial DF and CF strategies can be computed efficiently by convex optimization. An outcome of the optimization process is the optimum resource allocation (here the optimum time resource allocation) and the optimum transmit precoders at the source and relay.

- Our link-level simulations in Chapter 3 show that in a typical cellular downlink scenario partial DF can yield about 50% achievable rate increase compared to conventional point-to-point single-hop and multi-hop transmission. This gain reduces to about 20% if the sum-power of the source and relay is normalized.
- Our link-level simulations in Chapter 4 show that in a typical cellular uplink scenario partial CF can outperform partial DF by up to 40% at low SNR on the source-relay and source-destination links. However, partial CF requires a very high rate on the relay-destination link to become efficient. Therefore we believe that it is better suited to out-of-band relaying and to BS cooperation.
- The rate gains we observed in our link-level simulations suggest that cooperative relaying is an attractive solution to increase the spectral efficiency of future BWA networks. We hope that our work will support the design of practical precoders and coding strategies for cooperative MIMO relaying, like the knowledge of capacity bounds for point-to-point MIMO supported the design of currently standardized single-user MIMO coding schemes. However, we believe that more work is needed towards practical implementation:
 - A valuable research topic in this direction is the quantization and signaling of CSI for cooperative MIMO-OFDMA links, which can be combined with the exploitation of statistical CSI. Another direction is the extension of precoding with CSI to the multi-user MIMO relaying case, because the emerging standards IEEE802.16m and 3GPP LTE+ are likely to support multi-user MIMO.
 - We verify in Chapter 5 that by introducing some “degradation” parameters into our capacity bounds we are able to predict the throughput of a real system with a good accuracy (i.e. within about a dB). This topic would require more work on some specific aspects, for instance on the effect of rate saturation on cooperative MIMO relaying.
- Our network simulations in Chapter 5 essentially show that macroscopic propagation and interference play a key role in extending link-level results to the system-level:

- A first outcome is that fixed RSs shall either avoid or remove co-channel interference in order to achieve the high SNR on the BS-RS link required for DF and CF to be efficient in the downlink and uplink respectively.
 - Simulations in a downlink noise-limited deployment with two relays per sector show that relaying increases the cell average spectral efficiency by about 25% and cooperation (partial DF) adds another 20% gain on top of this. Moreover, with full CSI and optimum precoders at the source and relay the achievable rate is almost doubled at cell edge with partial DF. These results represent an a posteriori motivation for our research on MIMO precoding for the relay channel. Note that the figures aim at providing an order of magnitude for the potential gains to be expected from cooperation, but of course they are directly dependent on our choice of simulation parameters, which we tried to chose as realistic as possible.
 - Another observation is that spatial reuse has a strong impact on the performance of cooperative protocols. Indeed, simulations in Chapter 3 and Chapter 4 show that partial DF and CF strategies achieve the highest rate at the link-level. However, both are based on the TDD protocol III defined in §2.1.3.1 which involves simultaneous transmission by the Source and Relay during the second slot. This protocol does not allow multiple relay transmission, contrary to protocol I, and therefore the rate gain vanishes as soon as two relays (or more) are deployed per sector. Thus at the system-level protocol I can be preferred to protocol III.
- In Chapter 4, we derive some achievable rates for distributed compression applied to a coordinated uplink with multiple antenna network devices. We show how spare cellular backhaul capacity can be exploited to increase the wireless access capacity and again we provide by simulations some orders of magnitude on the required ratio of backhaul to access capacity in order to optimally exploit the potential of distributed compression. We believe that there is a high potential in distributed compression for next generation wireless networks but again a significant amount of work needs to be carried on before it becomes applicable to a real system. Among the many issues not addressed in this thesis are the selection of set of MSs and a set of

receiving BSs for each given time-frequency resource element, the problem of CSI signaling in a MIMO-OFDMA system, the backhaul latency issues, ...etc.

Overall, relaying and cooperation deserve well the interest that they have been raising over the last few years. However, by breaking the point-to-point paradigm they also create many challenges, the surface of which we sometimes only scratched in this thesis. Finally, we believe that in future works the topic of cooperation shall be addressed jointly with the problem of interference, which is a complex-enough problem to give headaches to PhD students for years ahead.

Appendix A Differentiation with respect to complex structured matrices

In this thesis, we need to differentiate real-valued functions of complex positive semi-definite matrices. Although the derivation with respect to complex matrices is well-known (e.g. [PP08]), the fact that the matrices with which we are dealing have a special structure requires the differentiation to be handled carefully. We start by reviewing classical results on complex differentiation when the matrix does not have a special structure. We then review a recently published methodology for differentiation with respect to structured matrices. Finally, we propose an alternative way to handle the case of structured matrices and discuss its advantages and drawbacks.

A.1 Differentiation with respect to unstructured matrices

The functions we are dealing with in this thesis are non-analytical and therefore we will resort to the generalized complex derivative and conjugate complex derivative, defined respectively as

$$\begin{aligned} \left(\frac{\partial f}{\partial \mathbf{X}}\right)_{i,j} &\triangleq \frac{\partial f}{\partial X_{i,j}} \triangleq \frac{1}{2} \left(\frac{\partial f}{\partial \Re(X_{i,j})} - i \frac{\partial f}{\partial \Im(X_{i,j})} \right) \\ \left(\frac{\partial f}{\partial \mathbf{X}^*}\right)_{i,j} &\triangleq \frac{\partial f}{\partial X_{i,j}^*} \triangleq \frac{1}{2} \left(\frac{\partial f}{\partial \Re(X_{i,j})} + i \frac{\partial f}{\partial \Im(X_{i,j})} \right) \end{aligned} \quad (6.1)$$

And the differential reads:

$$df = \text{tr} \left(\left(\frac{\partial f}{\partial \mathbf{X}} \right)^T d\mathbf{X} + \left(\frac{\partial f}{\partial \mathbf{X}^*} \right)^T d\mathbf{X}^* \right) \quad (6.2)$$

If the function is real-valued, then

$$\frac{\partial f}{\partial \mathbf{X}^*} = \left(\frac{\partial f}{\partial \mathbf{X}} \right)^* \quad (6.3)$$

and the differential (6.2) simplifies as:

$$df = 2 \text{Tr} \left(\Re \left(\left(\frac{\partial f}{\partial \mathbf{X}^*} \right)^H d\mathbf{X} \right) \right) = \langle \nabla f, d\mathbf{X} \rangle \quad (6.4)$$

where the gradient is defined as:

$$\nabla f \triangleq 2 \frac{\partial f}{\partial \mathbf{X}^*} \quad (6.5)$$

In this thesis, we will need the following partial derivatives:

$$\frac{\partial \mathcal{C}(\mathbf{X}, \mathbf{H})}{\partial \mathbf{X}} = \left(\mathbf{H}^H (\mathbf{I}_N + \mathbf{H}\mathbf{X}\mathbf{H}^H)^{-1} \mathbf{H} \right)^T \quad \frac{\partial \mathcal{C}(\mathbf{X}, \mathbf{H})}{\partial \mathbf{X}^*} = \mathbf{0}_N \quad (6.6)$$

$$\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} = (\mathbf{B}\mathbf{A})^T \quad \frac{\partial \text{tr}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}^*} = \mathbf{0}_N \quad (6.7)$$

$$\frac{\partial (t\mathcal{C}(\mathbf{X}/t, \mathbf{H}))}{\partial t} = \mathcal{C}(\mathbf{X}/t) - \frac{1}{t} \text{tr} \left(\mathbf{H} \left(\mathbf{I}_M + \frac{1}{t} \mathbf{X}\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{X}\mathbf{H}^H \right) \quad (6.8)$$

In [HG07], a notation is introduced which simplifies the expression of the differential and of the chain rule, especially for matrix valued functions of matrices. We use the notation of [HG07] throughout this thesis, which is defined from the differential expression:

$$df = (\mathcal{D}_{\mathbf{X}} f) \text{dvec}(\mathbf{X}) + (\mathcal{D}_{\mathbf{X}^*} f) \text{dvec}(\mathbf{X}^*) \quad (6.9)$$

With the notations of [HG07], the generalized complex derivative $\mathcal{D}_{\mathbf{X}} f$ and conjugate complex derivative $\mathcal{D}_{\mathbf{X}^*} f$ of a real-valued function are now row-vectors:

$$\mathcal{D}_{\mathbf{X}} f \triangleq \left(\text{vec} \left(\frac{\partial f}{\partial \mathbf{X}} \right) \right)^T \quad \mathcal{D}_{\mathbf{X}^*} f \triangleq \left(\text{vec} \left(\frac{\partial f}{\partial \mathbf{X}^*} \right) \right)^T \quad (6.10)$$

The gradient is

$$\nabla f \triangleq 2(\mathcal{D}_{\mathbf{X}^*} f)^T \quad (6.11)$$

Let define $h(\mathbf{X}, \mathbf{X}^*) = g(\mathbf{U}(\mathbf{X}, \mathbf{X}^*), \mathbf{U}^*(\mathbf{X}, \mathbf{X}^*))$. The chain rule then reads:

$$\mathcal{D}_{\mathbf{X}} h = (\mathcal{D}_{\mathbf{U}} g)(\mathcal{D}_{\mathbf{X}} \mathbf{U}) + (\mathcal{D}_{\mathbf{U}^*} g)(\mathcal{D}_{\mathbf{X}} \mathbf{U}^*) \quad (6.12)$$

$$\mathcal{D}_{\mathbf{X}^*} h = (\mathcal{D}_{\mathbf{U}} g)(\mathcal{D}_{\mathbf{X}^*} \mathbf{U}) + (\mathcal{D}_{\mathbf{U}^*} g)(\mathcal{D}_{\mathbf{X}^*} \mathbf{U}^*) \quad (6.13)$$

A.2 Differentiation with respect to structured complex matrices

In the previous section, it was assumed that all matrix components could vary independently. However, in this thesis we are dealing only with PSD matrices, which are by definition Hermitian-symmetric. In [PP08], the case of real structured matrices is addressed. The structured matrix is expressed as a function of an unstructured one, and the chain rule is used to find the derivative with respect to the unstructured matrix. In [HP08], the case of complex structured matrices is addressed. The authors call them

“patterned matrices”. Let consider for example Hermitian-symmetric matrices. Using similar notations as example 5 of [HP08], an $N \times N$ Hermitian-symmetric matrix \mathbf{X} can be generated by the following so-called “pattern producing” function:

$$\mathcal{F}: \mathbb{R}^N \times \mathbb{C}^{N(N-1)/2} \times \mathbb{C}^{N(N-1)/2} \rightarrow \mathbb{C}^{N \times N} \quad (6.14)$$

$$(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) \mapsto \text{mat}(\mathbf{L}_d \mathbf{r} + \mathbf{L}_l \mathbf{c} + \mathbf{L}_u \mathbf{c}^*) \triangleq \mathbf{X}$$

where \mathbf{L}_d is an $N^2 \times N$ matrix that maps the N independent components of the real vector \mathbf{r} onto the diagonal of \mathbf{X} , \mathbf{L}_l (resp. \mathbf{L}_u) is an $N^2 \times (N(N-1)/2)$ matrix that maps the $N(N-1)/2$ independent components of \mathbf{c} (resp. \mathbf{c}^*) onto the lower-triangular (resp. upper-triangular) part of \mathbf{X} . The derivatives of \mathcal{F} read:

$$\mathcal{D}_r \mathcal{F} = \mathbf{L}_d ; \mathcal{D}_c \mathcal{F} = \mathbf{L}_l ; \mathcal{D}_{c^*} \mathcal{F} = \mathbf{L}_u \quad (6.15)$$

The generalized complex derivatives with respect to \mathbf{r} , \mathbf{c} and \mathbf{c}^* of a function of a Hermitian-symmetric matrix can now be computed using equations (15), (16) and (17) of [HP08]:

$$\mathcal{D}_r f(\mathbf{X}) = \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_r \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) + \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_r \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) \quad (6.16)$$

$$\mathcal{D}_c f(\mathbf{X}) = \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_c \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) + \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_c \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) \quad (6.17)$$

$$\mathcal{D}_{c^*} f(\mathbf{X}) = \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_{c^*} \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) + \mathcal{D}_{\tilde{\mathbf{X}}} f(\tilde{\mathbf{X}}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \mathcal{D}_{c^*} \mathcal{F}(\mathbf{r}, \mathbf{c}, \mathbf{c}^*) \quad (6.18)$$

where $\tilde{\mathbf{X}}$ is the extension of \mathbf{X} to the set of unpatterned matrices. One important aspect that is mentioned in [HP08] and not in [PP08] is the problem of dimension. In the chain rule, each function must be differentiable, which requires that each variable can be changed independently of the other. In the case of Hermitian-symmetric matrices, the variables in vectors \mathbf{r} and \mathbf{c} shall be independent and the number of independent real variables (one per real variables and two per complex variable) shall be equal to the real dimension of the set of patterned matrices, e.g. $N + N(N-1)$. This condition is verified with the pattern-producing function (6.14), but many other parameterizations are possible which do not verify this condition on dimension.

Applying the above methodology to the function $\mathbf{X} \mapsto \mathcal{C}(\mathbf{X}, \mathbf{H})$ gives

$$\mathcal{D}_{\tilde{\mathbf{X}}} \mathcal{C}(\tilde{\mathbf{X}}, \mathbf{H}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \triangleq \left(\text{vec} \left(\frac{\partial \mathcal{C}(\tilde{\mathbf{X}}, \mathbf{H})}{\partial \tilde{\mathbf{X}}} \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} \right) \right)^T = \left(\text{vec} \left(\left(\mathbf{H}^H (\mathbf{I} + \mathbf{H} \mathbf{X} \mathbf{H}^H)^{-1} \mathbf{H} \right)^T \right) \right)^T \quad (6.19)$$

and

$$\mathcal{D}_{\tilde{\mathbf{X}}^*} \mathcal{C}(\tilde{\mathbf{X}}, \mathbf{H}) \Big|_{\tilde{\mathbf{X}}=\mathbf{X}} = \mathbf{0}_{1 \times N^2} \quad (6.20)$$

Inserting (6.19),(6.20) and (6.15) into (6.16) yields:

$$\mathcal{D}_r \mathcal{C}(\mathbf{X}, \mathbf{H}) = \left(\text{vec} \left(\left(\mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{X}\mathbf{H}^H)^{-1} \mathbf{H} \right)^T \right) \right)^T \mathbf{L}_d \quad (6.21)$$

Likewise,

$$\mathcal{D}_c \mathcal{C}(\mathbf{X}, \mathbf{H}) = \left(\text{vec} \left(\left(\mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{X}\mathbf{H}^H)^{-1} \mathbf{H} \right)^T \right) \right)^T \mathbf{L}_u \quad (6.22)$$

Finally, the gradient of \mathcal{C} is given by Theorem 2 of [HP08]:

$$\nabla \mathcal{C} = \left[\mathcal{D}_r \mathcal{C}(\mathbf{X}, \mathbf{H}) \quad 2\mathcal{D}_c \mathcal{C}(\mathbf{X}, \mathbf{H}) \right]^T \quad (6.23)$$

Note that because of the mapping \mathcal{F} defined by (6.14) is linear, it preserves convexity. However, in this thesis, the matrices with which we are dealing are not only Hermitian but also PSD. In [HP08] the Cholesky decomposition $\mathbf{Q} = \mathbf{L}\mathbf{L}^H$ is proposed to parameterize PSD matrices. Unfortunately in this case the mapping is non-linear and a function which is convex in \mathbf{Q} may be non-convex in \mathbf{L} . This is the reason why in the algorithms proposed in this thesis the Hermitian symmetry is guaranteed by the pattern-producing function, but the positive semi-definiteness is enforced by either gradient projection or a barrier function.

A.3 An alternative way to differentiate with respect to structured matrices

In the previous section, the approach of [HP08] to patterned complex matrix derivatives was presented. This approach presents the interest of expressing the gradient as a function of the minimum set of independent variables. Therefore, the numerical complexity of gradient descent algorithms is minimized. In this section, we propose an alternative way to compute the gradient for structured matrices, which is instantiated here for Hermitian matrices.

Let now assume that the matrix \mathbf{X} is an $N \times N$ unstructured complex matrix with $2N^2$ real dimensions. Let define

$$\mathbf{U}(\mathbf{X}, \mathbf{X}^*) \triangleq \frac{\mathbf{X} + (\mathbf{X}^*)^T}{2} = \frac{\mathbf{X} + \mathbf{X}^H}{2} \quad (6.24)$$

This function generates a Hermitian matrix which is equal to \mathbf{X} only if the latter is Hermitian. Let now define

$$g : \mathbf{X} \mapsto \mathcal{C}(\mathbf{U}(\mathbf{X}, \mathbf{X}^*), \mathbf{H}) \quad (6.25)$$

Applying the chain rule as defined in [HG07]:

$$\mathcal{D}_{\mathbf{X}} g = \mathcal{D}_{\mathbf{U}} g \mathcal{D}_{\mathbf{X}} \mathbf{U} + \underbrace{\mathcal{D}_{\mathbf{U}^*} g}_{=0_{N^2}} \mathcal{D}_{\mathbf{X}} \mathbf{U}^* = \mathcal{D}_{\mathbf{U}} g \mathcal{D}_{\mathbf{X}} \mathbf{U} \quad (6.26)$$

$$\mathcal{D}_{\mathbf{X}^*} g = \mathcal{D}_{\mathbf{U}} g \mathcal{D}_{\mathbf{X}^*} \mathbf{U} + \underbrace{\mathcal{D}_{\mathbf{U}^*} g}_{=0_{N^2}} \mathcal{D}_{\mathbf{X}^*} \mathbf{U}^* = \mathcal{D}_{\mathbf{U}} g \mathcal{D}_{\mathbf{X}^*} \mathbf{U} \quad (6.27)$$

From (6.24) we have:

$$\mathcal{D}_{\mathbf{X}} \mathbf{U} = \mathcal{D}_{\mathbf{X}^*} \mathbf{U} = \frac{1}{2} \mathbf{I}_{N^2} \quad (6.28)$$

Introducing (6.28) into (6.27) and (6.26) and switching back to convential notation gives:

$$\frac{\partial g}{\partial \mathbf{X}} = \frac{1}{2} \left(\mathbf{H}^H \left(\mathbf{I} + \mathbf{H} \left(\frac{\mathbf{X} + \mathbf{X}^H}{2} \right) \mathbf{H}^H \right)^{-1} \mathbf{H} \right)^T \quad (6.29)$$

$$\frac{\partial g}{\partial \mathbf{X}^*} = \frac{1}{2} \mathbf{H}^H \left(\mathbf{I} + \mathbf{H} \left(\frac{\mathbf{X} + \mathbf{X}^H}{2} \right) \mathbf{H}^H \right)^{-1} \mathbf{H} \quad (6.30)$$

Note that g is not convex on the set of unstructured matrices. However, nothing prevents us from starting a gradient descent from an initial point $\mathbf{X}_0 \in \mathcal{S}_+^N$. In this case, the gradient equals:

$$\nabla g|_{\mathbf{x}_0} = 2 \frac{\partial g}{\partial \mathbf{X}^*} \Big|_{\mathbf{x}_0} = \mathbf{H}^H \left(\mathbf{I} + \mathbf{H} \mathbf{X}_0 \mathbf{H}^H \right)^{-1} \mathbf{H} \quad (6.31)$$

As long as it is checked that the sequence of points in the descent lies in \mathcal{S}_+^N , the gradient expression (6.31) remains valid. Ultimately, the minimization leads to the same optimum point as the patterned derivative approach of [HP08].

The main interest in the approach that we introduce here is that the gradient is directly obtained from the unconstrained case as $(\partial f / \partial \mathbf{X})^*$.

Appendix B Numerical Optimization Algorithms

In this appendix, we review the numerical optimization algorithms which are used in this thesis. The two references on which we rely are [BV04] and [B99]. The former provides an in-depth analysis of convex sets, convex optimization theory and algorithms for convex optimization. The latter also addresses numerical algorithms for non-convex problems. We illustrate these algorithms in the context of the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{v} \in \mathcal{V}} o(\mathbf{v}) \\ \text{s.t. } \mathbf{f}(\mathbf{v}) \leq \mathbf{0}_{J \times 1} \end{aligned} \quad (6.32)$$

where o is a real-valued objective function, \mathbf{v} the variable (a vector or a matrix), \mathcal{V} is the domain of the problem and the inequality constraints are denoted as $f_j(\mathbf{v}) \leq 0$, $j = 1, \dots, J$ and stacked into a vector-valued function \mathbf{f} as follows:

$$\mathbf{f}(\mathbf{v}) \triangleq (f_1(\mathbf{v}), \dots, f_J(\mathbf{v}))^T \quad (6.33)$$

The subset of \mathcal{V} on which the constraints are satisfied is called the feasible set. The subset $\tilde{\mathcal{V}}$ of \mathcal{V} over which all the constraints are inactive, i.e. $f_j(\mathbf{v}) < 0 \forall j \in \{1, \dots, J\}$ is called the interior set of \mathcal{V} . If the domain, the objective and the inequality constraints are convex, then the problem is convex in standard form and any local optimum is a global optimum. For differentiable convex problems, the unique optimum point can sometimes be found by solving the necessary first order conditions for optimality (better known as Karush-Kuhn-Tucker (KKT) conditions). However, most often KKT conditions do not bring a closed-form solution and one has to resort to numerical optimization.

B.1 Gradient projection

The Gradient Projection Method (GPM) is described in section 2.3 of [B99] and used in §3.4.1 of this thesis. Let us assume in this section that the inequality constraints have been included in the definition of the domain, i.e. the domain is the feasible set. In order to introduce the GPM, let first consider a steepest descent. At the k th step the candidate next points are $\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} - s^{(k)} \nabla o|_{\mathbf{v}^{(k)}}$ it may happen that for some step sizes the candidate next point does not belong to \mathcal{V} . For instance adding a Hermitian matrix to a PSD matrix does not guarantee that the resulting matrix is PSD. The Gradient Projection Method (GPM) guarantees that the direction and step size lead to a feasible point. The GPM is an iterative algorithm which computes at step k the following points:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \alpha^{(k)} (\bar{\mathbf{v}}^{(k)} - \mathbf{v}^{(k)}) \quad (6.34)$$

$$\bar{\mathbf{v}}^{(k)} = \mathcal{P}_{\mathcal{V}} \left(\mathbf{v}^{(k)} - s^{(k)} \nabla L|_{\mathbf{v}^{(k)}} \right) \quad (6.35)$$

where $s^{(k)}$ is a positive scalar, $\alpha^{(k)} \in (0;1]$ is the step size and $\mathcal{P}_{\mathcal{V}}$ denotes the projection on \mathcal{V} . The GPM presents a practical interest when the projection operator turns out to be simple. For instance in [YB03], the GPM is used because the feasible set is the set of PSD matrices of unit Frobenius norm and the projection on such a set can be performed at a relatively low computational cost. In our thesis, the GPM is used to minimize a Lagrangian with respect to PSD matrices. Therefore a projection from the set of Hermitian matrices onto the PSD cone is needed. Let consider the eigenvalue decomposition of an $N \times N$ Hermitian matrix: $\mathbf{M} = \mathbf{U} \text{diag}(\boldsymbol{\lambda}) \mathbf{U}^H$. The projection of \mathbf{M} onto \mathcal{S}_+^N is [YB03]:

$$\mathcal{P}_{\mathcal{S}_+^N}(\mathbf{M}) = \mathbf{U} \text{diag}(\boldsymbol{\lambda}^+) \mathbf{U}^H \quad (6.36)$$

There exist lots of variants of the GPM and we refer the reader to [B99] for a detailed review. One important parameter which is left to the implementer is the choice of the step size selection strategy. The one we picked in this thesis is the Armijo rule along the feasible direction. Fixed a constant $s^{(k)} = s$, the Armijo rule gives a procedure to select $\alpha^{(k)}$ at each iteration. Fixed two scalars $\sigma \in (0,1)$ and $\beta \in (0,1)$, then $\alpha^{(k)} = \beta^{m_k}$ where m_k is the first non-negative integer such that

$$o(\mathbf{v}^{(k)}) - o\left(\mathbf{v}^{(k)} + \beta^m (\bar{\mathbf{v}}^{(k)} - \mathbf{v}^{(k)})\right) \geq -\sigma \beta^m \left\langle \nabla o(\mathbf{v}^{(k)}), \bar{\mathbf{v}}^{(k)} - \mathbf{v}^{(k)} \right\rangle \quad (6.37)$$

The choice of parameters β and σ is empirical, but the convergence to a stationary point (the optimum if the problem is convex) is proven for various step size selection strategies in [B99], including the Armijo rule.

B.2 The barrier method: an interior point algorithm

The barrier method is described in section 11.3.1. of [BV04] and is used in section 3.4.2. of this thesis. Given two fixed real parameters $\alpha > 0$ and $\beta > 1$, the barrier method is an iterative procedure which solves at the i th iteration the following unconstrained minimization problem:

$$\min_{\mathbf{v} \in \mathcal{V}} \left\{ o(\mathbf{v}) + \frac{1}{m(i)} \sum_{j=1}^J \phi_j(\mathbf{v}) \right\} \quad (6.38)$$

$$\phi_j(\mathbf{v}) \triangleq -\log(-f_j(\mathbf{v})) \quad (6.39)$$

where $m(1) = \alpha$ and $m(i) = \beta m(i-1)$. The function ϕ_j is the logarithmic barrier associated to the j th inequality constraint. This function tends to $+\infty$ when $f_j(\mathbf{v}) \rightarrow 0$. It is important to define barrier functions for all the constraints defining the feasible set (and not only the inequality constraints). For instance, as pointed out in Theorem 5.1 of [T01], $-\log|\mathbf{X}|$ is a barrier function for the positive definiteness constraint $\mathbf{X} \succ 0$. Once all the barriers are defined, the optimization can be carried on without the need for any projection as long as the descent starts from an initial point \mathbf{v}_0 in $\tilde{\mathcal{V}}$, i.e. an interior point. In this thesis we used a steepest descent with backtracking line search [BV04] to provide the step size. For sufficiently small step size, the candidate next point is guaranteed to lie within the interior set and the convergence to the optimum is proven in [BV04].

B.3 Solving the dual problem

The dual problem associated to (6.32) is:

$$\max_{\boldsymbol{\mu} \geq 0} g(\boldsymbol{\mu}) \quad \text{where} \quad g(\boldsymbol{\mu}) \triangleq \inf_{\mathbf{v} \in \mathcal{Y}} \left(\underbrace{o(\mathbf{v}) + \boldsymbol{\mu}^T \mathbf{f}(\mathbf{v})}_{\triangleq \mathcal{L}(\mathbf{v}, \boldsymbol{\mu})} \right) \quad (6.40)$$

where $\mathcal{L}(\mathbf{v}, \boldsymbol{\mu})$ denotes the Lagrangian and $g(\boldsymbol{\mu})$ is the dual function. The dual problem is always convex, even when the primal is not. The difference between the solution of the primal problem and the solution of the dual is called the duality gap. This quantity is always non-negative. The dual problem may be easier to solve than the primal, especially when the latter is non-convex, but in this case the duality gap must be quantified. Often when the primal is convex the duality gap is zero and it is said that strong duality holds. Proving strong duality can be established by simply proving that the interior set is non-empty. This last condition is called Slater's condition (see section 5.2.3 in [BV04]). An example application is provided in section 3.4.1 of this thesis. If the primal problem is non-convex, then proving that the duality gap is zero is more difficult, but not impossible. One possibility is to show that the following general sufficiency condition is satisfied:

General Sufficiency condition (Proposition 3.3.4 in [B99]):

Let $\hat{\mathbf{v}}$ and $\hat{\boldsymbol{\mu}}$ two vectors such that $\hat{\mathbf{v}}$ is a minimizer of the Lagrangian function $\mathcal{L}(\mathbf{v}, \hat{\boldsymbol{\mu}})$ and $\boldsymbol{\mu} \geq 0$ with $\mu_j = 0$ for all j belonging to the set of non-active constraints at $\hat{\mathbf{v}}$. Then $\hat{\mathbf{v}}$ is a global minimum of the problem.

An example application of the general sufficiency condition is given in [CS08a][CS08c][CS08d].

From (6.40), it can be observed that the dual problem can be decomposed into two optimization problems:

- **Minimization of the Lagrangian.** The computation of the dual function at a given point $\boldsymbol{\mu}_0$ requires to minimize the Lagrangian $\mathcal{L}(\mathbf{v}, \boldsymbol{\mu}_0)$ with respect to $\mathbf{v} \in \mathcal{V}$.
- **Maximization of the dual function.** The dual function $g(\boldsymbol{\mu})$ needs to be maximized on \mathbb{R}_+^J .

If the Lagrangian is differentiable, the first optimization problem can be solved by e.g. a classical gradient descent method or by the GPM depending on the set \mathcal{V} . The second optimization may be less straightforward. Indeed, from definition (6.40) it is in general difficult to derive a closed-form expression of the gradient ∇g (provided it exists). However, as shown below, a closed-form expression of a subgradient can be found (see also sec. 6.3 of [B99]). Since the dual function is concave in $\boldsymbol{\mu}$, a vector \mathbf{h} is a subgradient of g at $\boldsymbol{\mu}_0$ if for all $\boldsymbol{\mu}_1$:

$$g(\boldsymbol{\mu}_1) \leq g(\boldsymbol{\mu}_0) + \mathbf{h}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \quad (6.41)$$

Let $\hat{\mathbf{v}}_0$ and $\hat{\mathbf{v}}_1$ be minimizers of the Lagrangian at respectively $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. Then

$$\begin{aligned} g(\boldsymbol{\mu}_1) &\triangleq \mathcal{L}(\hat{\mathbf{v}}_1, \boldsymbol{\mu}_1) \leq \mathcal{L}(\hat{\mathbf{v}}_0, \boldsymbol{\mu}_1) \\ \Rightarrow g(\boldsymbol{\mu}_1) - g(\boldsymbol{\mu}_0) &\leq \mathcal{L}(\hat{\mathbf{v}}_0, \boldsymbol{\mu}_1) - \mathcal{L}(\hat{\mathbf{v}}_0, \boldsymbol{\mu}_0) = \mathbf{f}(\hat{\mathbf{v}}_0)^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \end{aligned} \quad (6.42)$$

From (6.41) and (6.42) it can be concluded that $\mathbf{f}(\hat{\mathbf{v}}_0)$ is a subgradient of g at $\boldsymbol{\mu}_0$, and the dual can be solved by subgradient methods. The subgradient method generates a sequence of dual-feasible points according to the following iteration:

$$\boldsymbol{\mu}^{(k+1)} = \mathcal{P}_{\mathcal{M}}(\boldsymbol{\mu}^{(k)} + s^{(k)} \mathbf{h}) \quad (6.43)$$

where \mathbf{h} is the subgradient, $s^{(k)}$ is a positive scalar step size and $\mathcal{P}_{\mathcal{M}}$ is the projection on the set \mathcal{M} of dual-feasible points. Proposition 6.3.1 in [B99] states that for sufficiently small step size, the distance to the optimum $\hat{\boldsymbol{\mu}}$ is reduced at each iteration. Unfortunately, the practical step size selection strategies are quite empirical, as explained in .sec. 6.3 of [B99]. For instance, in this thesis we computed the step size as:

$$s^{(k)} = \alpha^{(k)} \frac{(\tilde{g} - g(\boldsymbol{\mu}^{(k)}))}{\|\mathbf{h}^{(k)}\|^2} \quad (6.44)$$

where $\alpha^{(k)} = (1+m)/(k+m)$ with m a fixed positive integer and \tilde{g} an upper-bound on the optimum dual. Any primal-feasible solution of the first problem is an upper-bound, but not all solutions are primal-feasible, and therefore \tilde{g} may be updated infrequently. In §3.4, the CSB is computed either by solving the primal problem with an interior point method or by solving the dual problem using the GPM to minimize the Lagrangian and the subgradient method to maximize the dual function. We observed that in our simulations solving the primal problem was faster, and this seems to be due to a slow convergence of the subgradient method.

B.4 The non-linear Gauss-Seidel method

This algorithm is classified in section 2.7 of [B99] as a Block Coordinate Descent method. It applies to problems of the form

$$\min_{\mathbf{v} \in \mathcal{V}} o(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \quad (6.45)$$

where $\mathbf{v} \triangleq (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ and \mathcal{V} is a Cartesian product of convex sets: $\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 \times \dots \times \mathcal{V}_n$.

It is an iterative algorithm which optimizes each variable one after the other in a cyclic order:

$$\mathbf{v}_i^{(k+1)} = \arg \min_{\xi \in \mathcal{V}_i} o(\mathbf{v}_1^{(k+1)}, \dots, \mathbf{v}_{i-1}^{(k+1)}, \xi, \mathbf{v}_{i+1}^{(k)}, \dots, \mathbf{v}_n^{(k)}) \quad (6.46)$$

If the optimization with respect to each independent variable \mathbf{v}_i has a unique solution, and if the problem (6.45) has a unique solution (e.g. if it is convex) then the Gauss-Seidel algorithm converges to the optimum. In case the problem is non-convex, there is a unique minimum to the problem (6.45) and it is provided by the Gauss-Seidel algorithm if some contraction conditions are verified for the mapping $T(\mathbf{v}) = \mathbf{v} - \gamma \nabla o$ where γ is a positive scalar. These conditions are mentioned in [PC06] which refers to [BT89] for details. It is interesting to apply the Gauss-Seidel algorithm when the optimization w.r.t. each variable is easy to solve. One well-known example application is Yu's iterative waterfilling algorithm for the MIMO MAC [YRBC04]. Example applications of the Gauss-Seidel algorithm in this thesis can be found in §4.2.3.3 and §4.3.3.

Appendix C Proofs of Propositions

C.1 Proof of Proposition 2.1

During the first slot, S transmits $\omega_{d,1}$ and ω_r via superposition coding as follows:

$$\mathbf{x}_S^{(1)} = \sqrt{\alpha^{(1)} P_S} \mathbf{u}^{(1)}(\omega_{d,1}) + \sqrt{(1-\alpha^{(1)}) P_S} \mathbf{v}^{(1)}(\omega_r) \quad (6.47)$$

where $\alpha^{(1)} \in [0;1]$ is the fraction of source transmit power allocated to the transmission of the direct message during the first slot.

The relay first decodes $\omega_{d,1}$ from $\mathbf{y}_R^{(1)}$ and removes the contribution depending on this message from its observation before decoding ω_r . The rates $R_{d,1}$ and R_r are therefore constrained by:

$$R_{d,1} \leq t \log \left(1 + \frac{\alpha^{(1)} \gamma_1}{1 + (1-\alpha^{(1)}) \gamma_1} \right) \quad (6.48)$$

$$R_r \leq t \log (1 + (1-\alpha^{(1)}) \gamma_1) \quad (6.49)$$

Moreover, D can decode $\omega_{d,1}$ if

$$R_{d,1} \leq t \log \left(1 + \frac{\alpha^{(1)} \gamma_0}{1 + (1-\alpha^{(1)}) \gamma_0} \right) \quad (6.50)$$

During the second slot, the relayed message ω_r is used by S and R to cooperate while S sends the second direct message $\omega_{d,2}$ via superposition coding:

$$\mathbf{x}_S^{(2)} = \sqrt{\alpha^{(2)} P_S} \mathbf{u}^{(2)}(\omega_{d,2}) + \sqrt{(1-\alpha^{(2)}) P_S} \frac{\mathbf{h}_0^*}{|\mathbf{h}_0|} \mathbf{v}^{(2)}(\omega_r) \quad (6.51)$$

$$\mathbf{x}_R^{(2)} = \sqrt{P_R} \frac{\mathbf{h}_2^*}{|\mathbf{h}_2|} \mathbf{v}^{(2)}(\omega_r) \quad (6.52)$$

where $\alpha^{(2)} \in [0;1]$ is the fraction of source transmit power allocated to the transmission of the direct message during the second slot.

The destination starts by decoding ω_r from $\mathbf{y}_D^{(1)}$ and $\mathbf{y}_D^{(2)}$, and removes its contribution from the observation before decoding $\omega_{d,2}$ which imposes:

$$R_r \leq t \log (1 + (1-\alpha^{(1)}) \gamma_0) + (1-t) \log \left(1 + \frac{(\sqrt{(1-\alpha^{(2)}) \gamma_0} + \sqrt{\gamma_2})^2}{1 + \alpha^{(2)} \gamma_0} \right) \quad (6.53)$$

$$R_{d,2} \leq (1-t) \log (1 + \alpha^{(2)} \gamma_0) \quad (6.54)$$

Note that:

$$\begin{aligned}
\gamma_1 \geq \gamma_0 &\Rightarrow \alpha^{(1)} \gamma_1 + \alpha^{(1)} (1 - \alpha^{(1)}) \gamma_0 \gamma_1 \geq \alpha^{(1)} \gamma_0 + \alpha^{(1)} (1 - \alpha^{(1)}) \gamma_0 \gamma_1 \\
&\Rightarrow \alpha^{(1)} \gamma_1 (1 + (1 - \alpha^{(1)}) \gamma_0) \geq \alpha^{(1)} \gamma_0 ((1 + (1 - \alpha^{(1)}) \gamma_1)) \\
&\Rightarrow \frac{\alpha^{(1)} \gamma_1}{1 + (1 - \alpha^{(1)}) \gamma_1} \geq \frac{\alpha^{(1)} \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0} \\
&\Rightarrow t \log \left(1 + \frac{\alpha^{(1)} \gamma_1}{1 + (1 - \alpha^{(1)}) \gamma_1} \right) \geq t \log \left(1 + \frac{\alpha^{(1)} \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0} \right)
\end{aligned} \tag{6.55}$$

Therefore if $\gamma_1 \geq \gamma_0$, then (6.50) \Rightarrow (6.48) and (6.48)-(6.54) give:

$$\begin{aligned}
R_{PSC,TDD} &\triangleq R_r + R_{d,1} + R_{d,2} \\
&= \max_{t, \alpha^{(1)}, \alpha^{(2)}} \min \left(\underbrace{t \log(1 + (1 - \alpha^{(1)}) \gamma_1) + t \log \left(1 + \frac{\alpha^{(1)} \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0} \right)}_{\triangleq A} + (1 - t) \log(1 + \alpha^{(2)} \gamma_0), \right. \\
&\quad \left. \underbrace{t \log(1 + (1 - \alpha^{(1)}) \gamma_0)}_{\triangleq B} + (1 - t) \log \left(1 + \frac{(\sqrt{(1 - \alpha^{(2)}) \gamma_0} + \sqrt{\gamma_2})^2}{1 + \alpha^{(2)} \gamma_0} \right)}_{\triangleq C} \right. \\
&\quad \left. + \underbrace{t \log \left(1 + \frac{\alpha^{(1)} \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0} \right)}_{\triangleq D} + \underbrace{(1 - t) \log(1 + \alpha^{(2)} \gamma_0)}_{\triangleq E} \right)
\end{aligned} \tag{6.56}$$

The above expression can be simplified as follows:

$$B + D = t \log(1 + \gamma_0) \tag{6.57}$$

$$\begin{aligned}
C + E &= (1 - t) \log \left(\frac{1 + \alpha^{(2)} \gamma_0 + (1 - \alpha^{(2)}) \gamma_0 + \gamma_2 + 2\sqrt{(1 - \alpha^{(2)}) \gamma_0 \gamma_2}}{1 + \alpha^{(2)} \gamma_0} \right) + (1 - t) \log(1 + \alpha^{(2)} \gamma_0) \\
&= (1 - t) \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{(1 - \alpha^{(2)}) \gamma_0 \gamma_2})
\end{aligned} \tag{6.58}$$

Plugging (6.57) and (6.58) into (6.56) gives:

$$\begin{aligned}
R_{PSC,TDD} &= \max_{t, \alpha^{(1)}, \alpha^{(2)}} \min \left(\underbrace{t \log(1 + (1 - \alpha^{(1)}) \gamma_1) + t \log \left(1 + \frac{\alpha^{(1)} \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0} \right)}_{\triangleq A} + (1 - t) \log(1 + \alpha^{(2)} \gamma_0), \right. \\
&\quad \left. t \log(1 + \gamma_0) + (1 - t) \log(1 + \gamma_0 + \gamma_2 + 2\sqrt{(1 - \alpha^{(2)}) \gamma_0 \gamma_2}) \right)
\end{aligned} \tag{6.59}$$

Comparing the expression (6.59) with the expression of $R_{PDF,P3}$ given by (2.12), the superposition coding increases the achievable rate if and only if the following condition is satisfied:

$$\begin{aligned}
A \geq t \log(1 + \gamma_1) &\Leftrightarrow t \log(1 + (1 - \alpha^{(1)}) \gamma_1) + t \log\left(\frac{1 + \gamma_0}{1 + (1 - \alpha^{(1)}) \gamma_0}\right) - t \log(1 + \gamma_1) \geq 0 \\
&\Leftrightarrow \log\left(\frac{(1 + (1 - \alpha^{(1)}) \gamma_1)(1 + \gamma_0)}{(1 + \gamma_1)(1 + (1 - \alpha^{(1)}) \gamma_0)}\right) \geq 0 \\
&\Leftrightarrow \left(1 - \frac{\alpha^{(1)} \gamma_1}{1 + \gamma_1}\right) \Big/ \left(1 - \frac{\alpha^{(1)} \gamma_0}{1 + \gamma_0}\right) \geq 1 \\
&\Leftrightarrow \frac{\gamma_1}{1 + \gamma_1} \leq \frac{\gamma_0}{1 + \gamma_0} \\
&\Leftrightarrow \gamma_1 \leq \gamma_0
\end{aligned} \tag{6.60}$$

Therefore, we conclude that

$$R_{PSC,TD} \leq R_{PDF,P3} \tag{6.61}$$

C.2 Proof of Proposition 4.1

The proof can be obtained as a special case of Theorem 3 and corollary 4 in [GDV06]. The latter considers a Gaussian source vector \mathbf{x} which is split into two (correlated) parts \mathbf{x}_1 and \mathbf{x}_2 and shall be reconstructed from a compressed version of \mathbf{x}_1 and a noisy observation of \mathbf{x}_2 . Our problem is slightly different as we are only interested in reconstructing \mathbf{x}_1 and not the whole vector \mathbf{x} , thus we do not take into account the distortion on \mathbf{x}_2 . As in [W78], let first consider the rate-distortion coding of the Gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at both the encoder and the decoder. It can be realized by the distribution $f(\hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)})$ which is generated on Figure 52.

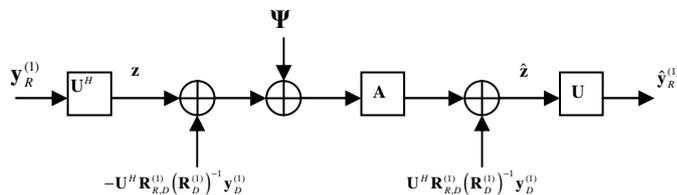


Figure 52: Rate-distortion coding of gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the encoder and decoder

Because the CKLT is a unitary transform, it preserves the quadratic distortion:

$$\mathcal{D} \triangleq E \left[\left\| \mathbf{y}_R^{(1)} - \hat{\mathbf{y}}_R^{(1)} \right\|_2^2 \middle| \mathbf{y}_D^{(1)} \right] = E \left[\left\| \hat{\mathbf{z}} - \mathbf{z} \right\|_2^2 \middle| \mathbf{y}_D^{(1)} \right] \quad (6.62)$$

Furthermore, a fundamental property of the CKLT is that the transformed vector $\mathbf{z} \triangleq \mathbf{U}^H \mathbf{y}_R^{(1)}$ has conditionally uncorrelated components given the side information $\mathbf{y}_D^{(1)}$. Therefore, rate-distortion encoding can be performed separately on each component of the transformed vector using the same scheme as in section 3 of [W78]: first, the conditional expectation $E[\mathbf{z} | \mathbf{y}_D^{(1)}] = \mathbf{U}^H \mathbf{R}_{R,D}^{(1)} (\mathbf{R}_D^{(1)})^{-1} \mathbf{y}_D^{(1)}$ is removed, then rate-distortion coding of independent Gaussian variables is performed (section 13.3.3 in [CT91]), and finally the conditional expectation is added back at the destination to obtain the reconstructed signal $\hat{\mathbf{z}}$ which is transformed into $\hat{\mathbf{y}}_R^{(1)}$ by an inverse CKLT. The rate-distortion function with side information at both the encoder and decoder is found by minimizing the information rate distortion function with respect to the distribution $f(\hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)})$ for a given sum-distortion \mathcal{D} :

$$r_{R|D}(\mathcal{D}) = \min_{f(\hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)})} I(\mathbf{y}_R^{(1)}; \hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_D^{(1)}) \quad (6.63)$$

Note that the two schemes of Figure 13 and Figure 52 result in the same input-output relationship. From equations (4.12)-(4.15), it is clear that the distribution is determined by the choice of a compression noise vector $\boldsymbol{\eta}$. Let now define the vector \mathbf{d} as the squared distortion per component of the transformed vector \mathbf{z} :

$$d_i \triangleq E \left[\left| \hat{z}_i - z_i \right|^2 \middle| \mathbf{y}_D^{(1)} \right] \quad (6.64)$$

The following lines show the relationship between the compression noise $\boldsymbol{\eta}$ and the component-wise distortion \mathbf{d} . From (4.12) and (4.4) we can write:

$$\hat{\mathbf{z}} - \mathbf{z} \triangleq \mathbf{U}^H (\hat{\mathbf{y}}_R^{(1)} - \mathbf{y}_R^{(1)}) = (\mathbf{A} - \mathbf{I}) \mathbf{U}^H (\mathbf{y}_R^{(1)} - E[\mathbf{y}_R^{(1)} | \mathbf{y}_D^{(1)}]) + \mathbf{A} \boldsymbol{\psi} \quad (6.65)$$

Then by definition of conditional covariance and since the matrix \mathbf{A} is diagonal, inserting (6.65) into (6.64) gives:

$$d_i = (a_i - 1)^2 s_i + a_i^2 \eta_i \quad (6.66)$$

Finally, replacing a_i in (6.66) by its definition given by (4.14) leads to equation (4.17): $d_i = s_i \eta_i / (s_i + \eta_i)$. Therefore, the distribution $f(\hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)})$ is equivalently determined by the choice of either $\boldsymbol{\eta}$ or \mathbf{d} . It will be shown in section 4.2.3.1 that the distribution which minimizes the sum-distortion in (6.63) may not be the best for our CF relaying problem. We therefore compute the rate required to achieve a component-wise distortion \mathbf{d} :

$$\begin{aligned} r_{R|D}^{(1)}(\mathbf{d}) &= I(\mathbf{y}_R^{(1)}; \hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_D^{(1)}) \\ &\stackrel{(a)}{=} I(\mathbf{U}^H \mathbf{y}_R^{(1)}; \mathbf{U}^H \hat{\mathbf{y}}_R^{(1)} | \mathbf{y}_D^{(1)}) \\ &\triangleq I(\mathbf{z}; \hat{\mathbf{z}} | \mathbf{y}_D^{(1)}) \\ &= H(\hat{\mathbf{z}} | \mathbf{y}_D^{(1)}) - H(\hat{\mathbf{z}} | \mathbf{y}_D^{(1)}, \mathbf{z}) \\ &\stackrel{(b)}{=} H(\mathbf{A}\mathbf{z} + \mathbf{A}\boldsymbol{\psi} + \mathbf{K}\mathbf{y}_D^{(1)} | \mathbf{y}_D^{(1)}) - H(\mathbf{A}\mathbf{z} + \mathbf{A}\boldsymbol{\psi} + \mathbf{K}\mathbf{y}_D^{(1)} | \mathbf{z}, \mathbf{y}_D^{(1)}) \\ &\stackrel{(c)}{=} H(\mathbf{z} + \boldsymbol{\psi} | \mathbf{y}_D^{(1)}) - H(\boldsymbol{\psi}) \\ &\stackrel{(d)}{=} \sum_{i=1}^{N_R} \log \left(s_i + \frac{s_i d_i}{s_i - d_i} \right) - \sum_{i=1}^{N_R} \log \left(\frac{s_i d_i}{s_i - d_i} \right) \\ &= \sum_{i=1}^{N_R} \log(s_i / d_i) \triangleq \sum_{i=1}^{N_R} r_i \end{aligned} \quad (6.67)$$

where

(a) follows from equation (13) in [NM93], which gives the entropy of the product of a proper complex Gaussian vector \mathbf{x} by a non-singular matrix \mathbf{M} :

$$H(\mathbf{M}\mathbf{x}) = H(\mathbf{x}) + 2 \log |\det(\mathbf{M})| \quad (6.68).$$

(b) is straightforward from equation (4.12)

(c) stems from (6.68) and the fact that the entropy of a known variable is zero.

(d) from the fact that the components of \mathbf{z} are conditionally independent given $\mathbf{y}_D^{(1)}$ by definition of the CKLT.

As in [W78], let denote by $r^*(\mathbf{d})$ the rate with side information at the decoder only. The equality between $r^*(\mathbf{d})$ and $r_{R|D}(\mathbf{d})$ follows from section 3 of [W78] in which the equivalence between Fig. 1 and Fig. 2 and the fact that the following Markov chain holds:

$$\mathbf{y}_D \rightarrow \mathbf{y}_R \rightarrow \mathbf{v} \quad (6.69)$$

(where \mathbf{v} is defined on Figure 13) lead to:

$$I(\mathbf{y}_R; \mathbf{v} | \mathbf{y}_D) = I(\mathbf{y}_R; \hat{\mathbf{y}}_R | \mathbf{y}_D) \quad (6.70)$$

Finally, the rate-distortion function $r^*(\delta)$ is obtained by minimizing the rate under total squared distortion δ . This constrained problem is convex and the solution is given by the well-known (section 13.3.3 in [CT91]) reverse waterfilling algorithm.

C.3 Proof of Proposition 4.3

Applying the chain rule for mutual information to (4.27) gives:

$$R_0 = I(\mathbf{x}_S^{(1)}; \tilde{\mathbf{y}}_R^{(1)}, \mathbf{y}_D^{(1)}) = \underbrace{I(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)})}_{\triangleq R_{0,d}} + \underbrace{I(\mathbf{x}_S^{(1)}; \tilde{\mathbf{y}}_R^{(1)} | \mathbf{y}_D^{(1)})}_{\triangleq R_{0,r}} \quad (6.71)$$

The first term $R_{0,d}$ is equal to $\mathcal{C}(\mathbf{R}_S^{(1)}, \mathbf{H}_0)$. The second term $R_{0,r}$ can be computed as follows:

$$\begin{aligned} R_{0,r} &= H(\mathbf{y}_R^{(1)} + \mathbf{U}\Psi | \mathbf{y}_D^{(1)}) - H(\mathbf{y}_R^{(1)} + \mathbf{U}\Psi | \mathbf{x}_S^{(1)}, \mathbf{y}_D^{(1)}) \\ &\stackrel{(a)}{=} H(\mathbf{y}_R^{(1)} + \mathbf{U}\Psi | \mathbf{y}_D^{(1)}) - H(\mathbf{y}_R^{(1)} + \mathbf{U}\Psi | \mathbf{x}_S^{(1)}) \\ &= \log |\mathbf{R}_{R|D}^{(1)} + \mathbf{U} \text{diag}(\boldsymbol{\eta}) \mathbf{U}^H| - \log |\mathbf{R}_{R|S}^{(1)} + \mathbf{U} \text{diag}(\boldsymbol{\eta}) \mathbf{U}^H| \\ &\stackrel{(b)}{=} \log |\mathbf{U} \text{diag}(\mathbf{s} + \boldsymbol{\eta}) \mathbf{U}^H| - \log |\mathbf{U}(\sigma^2 \mathbf{I}_{N_R} + \text{diag}(\boldsymbol{\eta})) \mathbf{U}^H| \\ &= \sum_{i=1}^{N_R} \log \left(\frac{s_i + \eta_i}{\sigma^2 + \eta_i} \right) \end{aligned} \quad (6.72)$$

where (a) comes from the fact that $\mathbf{y}_D^{(1)}$ is a noisy version of $\mathbf{H}_0 \mathbf{x}_S^{(1)}$ and (b) from the fact that white thermal noise was assumed in our signal model.

The maximization of $R_{0,r}$ w.r.t. $\boldsymbol{\eta}$ can now be performed. From (4.16) we have $\eta_i = s_i / (2^r - 1)$, which can be inserted into (6.72), resulting in the following equivalent problem:

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{i=1}^{N_R} \log \left(\frac{s_i 2^{r_i}}{\sigma^2 (2^{r_i} - 1) + s_i} \right) \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^{N_R} r_i \leq R_1 \\ r_i \geq 0 \text{ for } 1 \leq i \leq N_R \end{cases} \end{aligned} \quad (6.73)$$

It can be checked that this objective is concave in \mathbf{r} , and since the inequality constraints are affine the problem is convex in standard form. The KKT conditions yield after a few simple calculations the solution (4.31).

C.4 Proof of Proposition 4.4

Let parameterize the sum-rate side of the achievable rate region (4.36), denoting by $\alpha \in [0;1]$ the fraction of the time during which ω_1 is decoded first, assuming that the rest of the time ω_2 is decoded first. In the single-antenna case the channel matrices are complex scalar denoted by H_0 , H_1 and H_2 which are normalized in this proof such that $\sigma^2 = 1$. The achievable rates read as:

$$R_1 = \alpha \log \left(1 + \frac{|H_2|^2 P_R}{1 + |H_0|^2 P_S} \right) + (1 - \alpha) \log \left(1 + |H_2|^2 P_R \right) \quad (6.74)$$

$$R_2 = \alpha \log \left(1 + |H_0|^2 P_S \right) + (1 - \alpha) \log \left(1 + \frac{|H_0|^2 P_S}{1 + |H_2|^2 P_R} \right) \quad (6.75)$$

$$R_0 = \log \left(1 + |H_0|^2 P_S + \frac{|H_1|^2 P_S}{1 + \eta} \right) \quad (6.76)$$

$$\text{where } \eta = \left(1 + |H_1|^2 P_S + |H_0|^2 P_S \right) \left(2^{\frac{(1-t)R_1}{t}} - 1 \right)^{-1} \left(1 + |H_0|^2 P_S \right)^{-1} \quad (6.77)$$

From (4.3) and (6.74)-(6.77), it can be shown (after tedious calculations) that:

$$\frac{\partial R_{CF}}{\partial \alpha} = -2^{\frac{(1-t)R_1}{t}} \left(2^{\frac{(1-t)R_1}{t}} - 1 \right)^{-1} (1 + \eta)^{-1} (1 - t) \frac{\partial R_1}{\partial \alpha} \geq 0 \quad (6.78)$$

Since from (6.74) we have $\partial R_1 / \partial \alpha \leq 0$, therefore $\partial R_{CF} / \partial \alpha \geq 0$. In general, inequality (6.78) is strict and $\gamma = 1$ is optimum, Q.E.D.

C.5 Proof of Proposition 4.8

Equations (4.56) and (4.58) are straightforward. We show that (4.57) is equivalent to (4.39) in Proposition 4.5 under a sum-rate constraint. Indeed, we have:

$$\begin{aligned} \forall \mathcal{G} \subseteq \{1, \dots, M\} : I(\hat{\mathbf{y}}_{\mathcal{G}}; \mathbf{y}_{\mathcal{G}} | \mathbf{y}_0, \hat{\mathbf{y}}_{\{1, \dots, M\} \setminus \mathcal{G}}) &\leq I(\{\hat{\mathbf{y}}_i\}_1^M; \{\hat{\mathbf{y}}_i\}_1^M | \mathbf{y}_0, \hat{\mathbf{y}}_{\{1, \dots, M\} \setminus \mathcal{G}}) \\ &\leq I(\{\hat{\mathbf{y}}_i\}_1^M; \{\hat{\mathbf{y}}_i\}_1^M | \mathbf{y}_0) \end{aligned} \quad (6.79)$$

Since equality holds in (6.79) when $\mathcal{G} = \{1, \dots, M\}$, the constraint (4.39) is equivalent to (6.79) under a backhaul sum-rate constraint. Therefore the rate of Proposition 4.8 is achievable by the DCF-JD strategy of Proposition 4.5. It remains to check whether it is also achievable by DCF-SD strategy of Proposition 4.6. For a given permutation π , the minimum rate required to compress the i th BS observation is equal to $I(\mathbf{y}_{\pi(i)}; \hat{\mathbf{y}}_{\pi(i)} | \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1})$. Therefore the minimum sum-rate is:

$$\begin{aligned} \sum_{i=1}^M I(\mathbf{y}_{\pi(i)}; \hat{\mathbf{y}}_{\pi(i)} | \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}) &\stackrel{(a)}{=} \sum_{i=1}^M I(\{\mathbf{y}_j\}_1^M; \hat{\mathbf{y}}_{\pi(i)} | \mathbf{y}_0, \{\hat{\mathbf{y}}_{\pi(j)}\}_1^{i-1}) \\ &\stackrel{(b)}{=} I(\{\mathbf{y}_i\}_1^M; \{\hat{\mathbf{y}}_{\pi(i)}\}_1^M | \mathbf{y}_0) \end{aligned} \quad (6.80)$$

where (a) comes from the Markov chain relationship (4.44) and (b) from the Chain rule. Equation (6.80) is equivalent to (4.57) which concludes the proof.

Appendix D EESM model for cooperative links

In this annex we provide details on error prediction for the cooperative DF strategies of §5.2.2 (cooperative IR) and §5.2.3 (superposition coding). For cooperative IR, we need to predict the error rate for an equivalent code rate formed by the combined transmission by the source and relay. Therefore, we compute by simulations EESM parameters for code rates in the range 1/3 to 5/6 with a granularity thin enough to allow the prediction for any code rate by linear interpolation of the tabulated code rates. The assumptions are the following:

- Data block size of 120B
- PUSC subchannel to subcarrier mapping
- Code rates: 1/3, 2/5, 1/2, 3/4, 5/6
- Channel: 40 independent snapshots of SCME typical urban channel [B05].
- EESM beta factor optimized to minimize the standard deviation of $\gamma_{eff}(\beta)$ at target BLER of 5%
- No channel estimation error

Link simulation results plotted on Figure 53 illustrate the BLER prediction performance of EESM in the non-cooperative case. The code rate determines the color of the curve. Solid line curve represents the AWGN channel performance, whereas the performance on actual channel snapshots is represented by clouds of points at values of BLER logarithmically spaced between 100% and 1%.

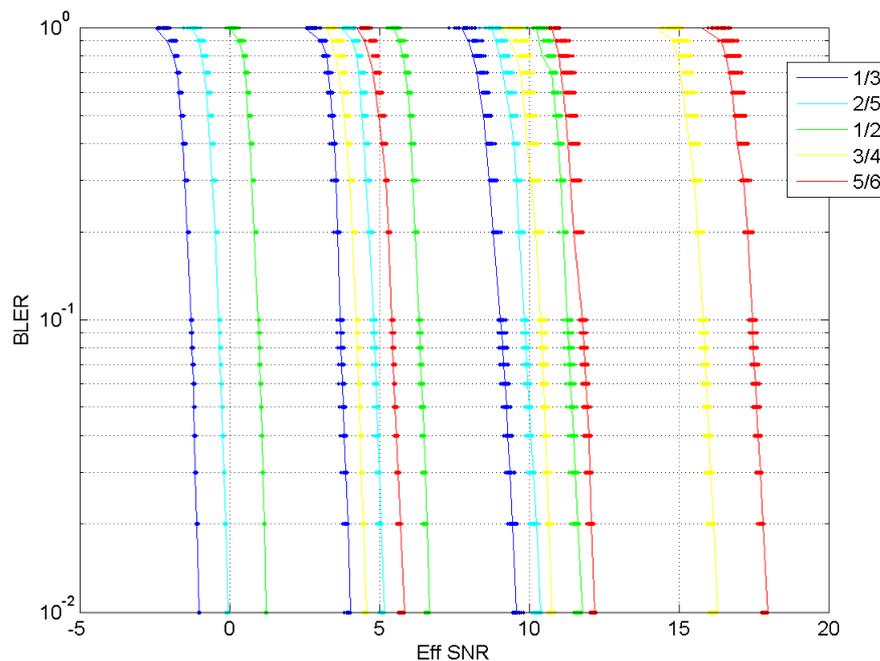


Figure 53: BLER vs EESM SNR_{eff} curves on AWGN and Actual Channel snapshots for code rates 1/3 to 5/6 and constellations QPSK, 16QAM and 64QAM.

On Table 2, the error prediction performance of EESM corresponding to Figure 53 is summarized. As mentioned before, a look-up table is generated for each MCS. For a given channel snapshot, the prediction error (measured in dB) is the difference between the actual SNR required to meet the target PER and the predicted SNR. The Root Mean Squared Error of the predictor which is provided in Table 2 gives an idea of the SNR margin that is to be taken by the MCS selection algorithm to avoid a bad MCS selection. Typically, twice the RMSE is enough to minimize the bad MCS selection event. It means that in the worse case, the throughput performance is degraded by twice the RMSE. However, the impact on the average throughput is lower than the worst case degradation. The RMSE figures of Table 2 confirm that EESM can predict the throughput envelope with an accuracy better than 0.2 dB.

Modulation	Code Rate	EESM Beta	EESM RMSE (dB)
QPSK	1/3	1.58	0.02
QPSK	2/5	1.58	0.02
QPSK	1/2	1.58	0.02
QPSK	3/4	1.74	0.06
QPSK	5/6	1.74	0.08
16QAM	1/3	3.98	0.05
16QAM	2/5	4.57	0.04
16QAM	1/2	5.01	0.06
16QAM	3/4	7.94	0.10
16QAM	5/6	8.32	0.13
64QAM	1/3	19.95	0.08
64QAM	2/5	15.85	0.07
64QAM	1/2	16.60	0.08
64QAM	3/4	28.84	0.12
64QAM	5/6	33.11	0.12

Table 2: Error Prediction performance of EESM

We now evaluate the error prediction accuracy of EESM for cooperative links. The following techniques used in cooperative coding strategies require specific study:

- Reliability combining of packets with partial repetition of coded bits, different constellation and large SINR difference
- Superposition coding

The EESM modeling of superposition coding is studied in [FIR07], and it is concluded that the EESM can be computed by assuming a Gaussian-distributed interferer and modifying the SINRs for each sub-carrier accordingly. We will focus on the first bullet point. In [CSL06], a method to compute the exponential effective SNR is proposed for HARQ involving incremental redundancy with partial retransmission of coded bits, and possibly a change of constellation by means of a demapping penalty [RGC02]. The method originally proposed in [CSL06] computes the effective SNR by considering the channel SINR for all the coded bits over all the retransmissions. When a coded bit is

repeated, the SINRs add-up. The beta parameter must be selected based on the equivalent code rate. The problem with this method is that an SINR must be stored for each coded bit. Therefore, [CSL06] proposes a recursive formula that replaces the SINR for the coded bits of all previous transmissions by a single effective SNR value. Such simplification completely removes the need for storage, but reduces the prediction accuracy. Finally, binning is proposed as a trade-off between the recursive and the full-accuracy methods. We want to check if the accuracy of this recursive method is enough for our objective.

For EESM, the recursive formula for computing the effective SNR at the k th transmission attempt is the following:

$$\gamma_{eff}^1 = -\beta \log \left(\frac{1}{|U_1|} \sum_{i \in U_1} \exp \left(-\frac{\gamma_{i,1}}{\beta} \right) \right) \quad (6.81)$$

$$\gamma_{eff}^k = -\beta \log \left(\frac{1}{|U_k|} \left(\sum_{i \in U_{k-1}} \exp \left(-\frac{\gamma_{eff}^{k-1} + I_{i,k} \gamma_{i,k}}{\beta} \right) + \sum_{i \in (U_k \setminus U_{k-1})} \exp \left(-\frac{\gamma_{i,k}}{\beta} \right) \right) \right) \quad (6.82)$$

where U_k is the set of received coded bits at the k th transmission attempt and $I_{i,k} = 1$ if the i th coded bit of the k th transmission is a repetition of a previously transmitted coded bit.

Formula (6.82) assumes that the same constellation was used in all transmissions. If the constellation is changed, then a reference constellation can be chosen, for instance QPSK, and the β shall be taken for the equivalent code rate of the reference constellation. A demapping penalty is added to γ_{eff}^{k-1} and $\gamma_{i,k}$ that corresponds to the difference in dB between the BLER vs. SNR curves of the M-QAM and reference QPSK. In the following we assumed demapping penalties of 5dB and 10dB respectively for 16QAM and 64QAM:

$$\begin{aligned} P(QPSK, 16QAM) &\approx 5dB \\ P(QPSK, 64QAM) &\approx 10dB \\ P(16QAM, 64QAM) &\approx 5dB \end{aligned} \quad (6.83)$$

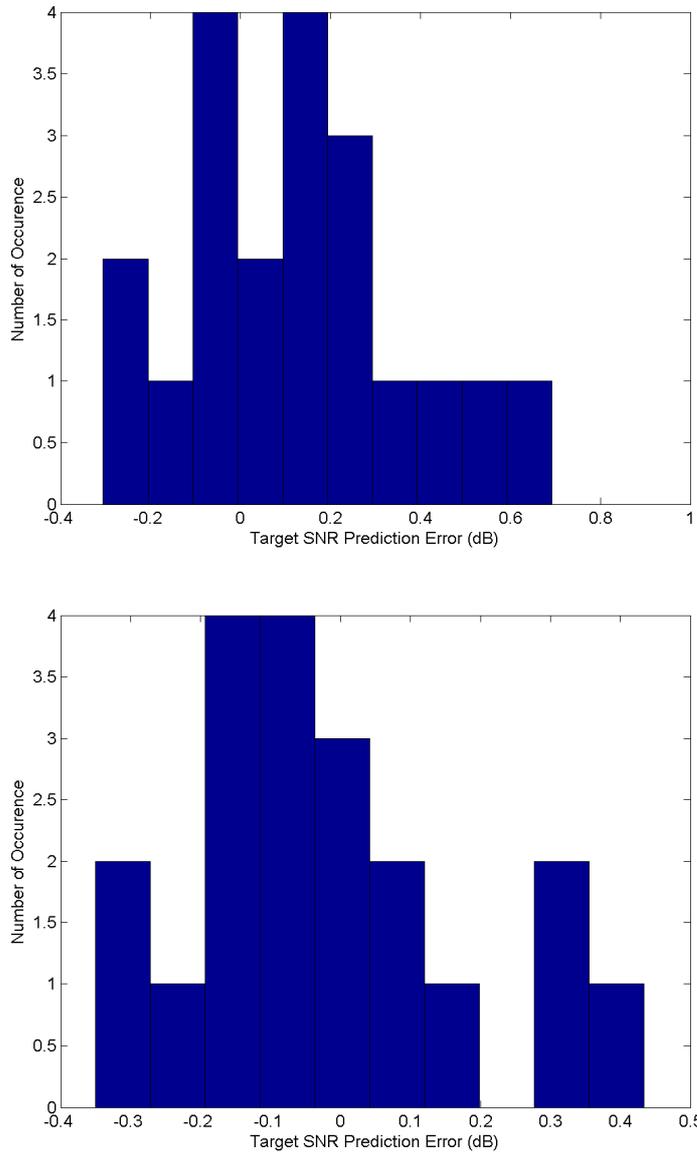


Figure 54: Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. S-R: 64QAM, R=5/6 R-D: 64QAM, R=1/2, Cooperative IR v1.

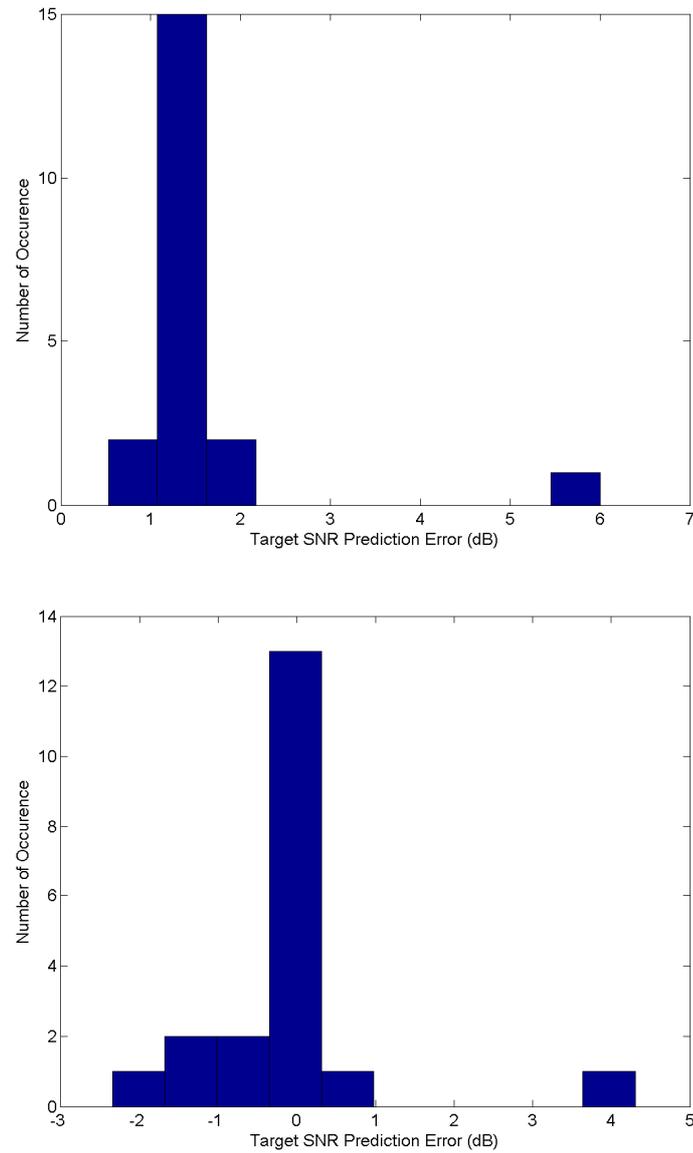


Figure 55: Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. S-R: 16QAM, R=3/4 R-D: 16QAM, R=3/4, Cooperative IR v1.

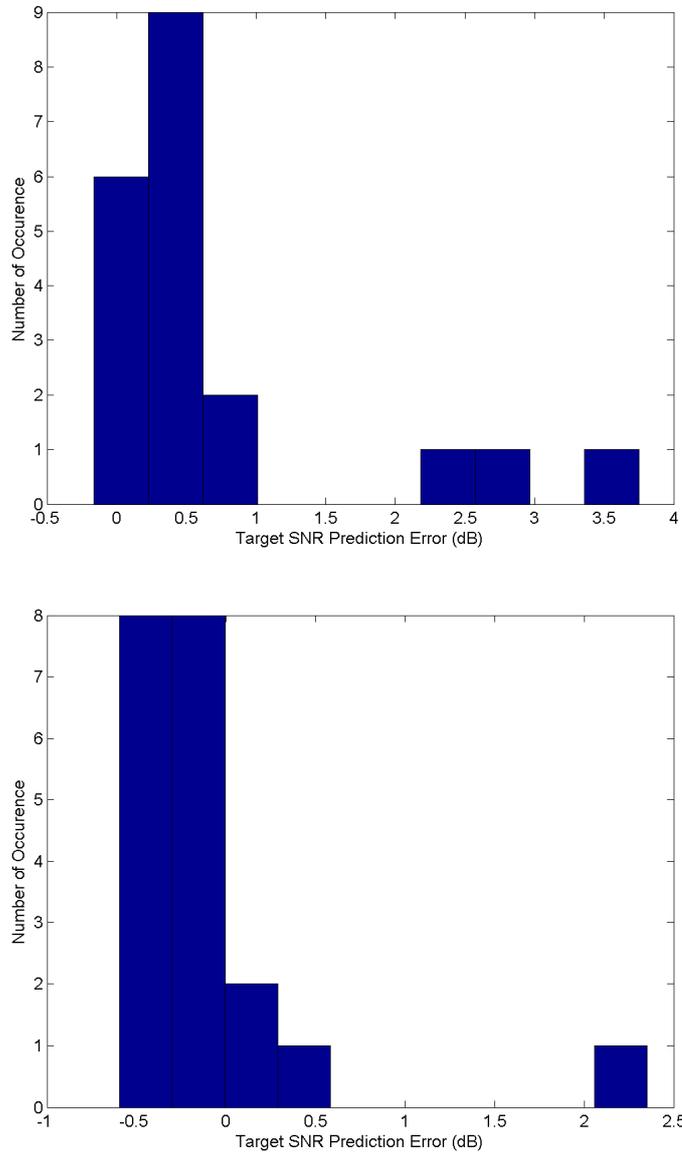


Figure 56 Distribution of target SNR prediction errors for recursive (top) and non-recursive (bottom) EESM. BS-RS: 16QAM, $R=3/4$ RS-MS: 64QAM, $R=1/2$, Cooperative IR v1.

From the above simulation results, we conclude that we should select the non-recursive formula to compute the EESM for our simulations of cooperative IR, in order to benefit from the best error prediction accuracy. Note that whatever the formula chosen, in some cases a large prediction error may occur. However, these events -which occur at low SNR- correspond to a positive target SNR prediction error, which will only result in a conservative selection of the MCS, but not to a fatal misadaptation. Therefore the impact on throughput should be low (a few percent).

References

[11n08]	IEEE P802.11n/D5.00, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 4: Enhancements for Higher Throughput. Jul 2008.
[16e05]	IEEE Std 802.16e-2005 "Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems- Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands". Feb 2006.
[16j07]	IEEE 802.16j Task Group. "Baseline Document for Draft Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Multihop Relay Specification" Document nb. 802.16j-06/026r4. June 2007.
[16m06]	IEEE 802.16m Task Group "Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Advanced Air Interface" PAR approved by the IEEE-SA Standards Board on 6 Dec. 2006.
[ACH07]	J.G. Andrews, Wan Choi, and R.W. Heath Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," <i>IEEE Wireless Communications</i> , vol. 14, no. 6, Dec. 2007.
[AES05]	K.Azarian, H.El Gamal, P.Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels", <i>IEEE Trans. on Information Theory</i> , vol. 51, no.12, Dec. 2005.
[AV06]	A. Agustín and J. Vidal, "TDMA cooperation with spatial reuse of the relay slot and interfering power distribution information," <i>IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)</i> , May 2006.
[AV07]	A. Agustin, J. Vidal "Radio Resource Optimization for the Half-Duplex Relay-Assisted Multiple Access Channel" <i>IEEE Signal Processing Advances for Wireless Communications (SPAWC)</i> , Jun. 2007
[B05]	D.S. Baum et. al., "An Interim Channel Model for Beyond-3G Systems" in <i>Proc. IEEE VTC Spring</i> , Vol 5, May 2005.
[B99]	D. P. Bertsekas "Nonlinear Programming". <i>Athena Scientific</i> . 2 nd edition, 1999.
[BGP02]	H. Bolcskei, D. Gesbert, A. Paulraj, "On the capacity of wireless systems employing OFDM-based spatial multiplexing", <i>IEEE Trans.Communications</i> , February 2002.
[BSC04]	Y.W. Blankenship, P.J. Sartori, B.K. Classon, V. Desai, K.L. Baum "Link error prediction methods for multicarrier systems" <i>IEEE Vehicular Technology Conference (VTC2004-Fall)</i> , Sep 2004.
[BT89]	D. P. Bertsekas and J. N. Tsitsiklis, "Parallel and Distributed Computation: Numerical Methods". <i>Englewood Cliffs, NJ: Prentice-Hall</i> , 1989.
[BV04]	S. Boyd and L. Vandenberghe "Convex Optimization" <i>Cambridge University Press</i> , 2004. Avail. http://www.stanford.edu/~boyd/cvxbook/
[C08]	A. Del Coso "Achievable Rates for Gaussian Channels with Multiple Relays" <i>PhD Thesis</i> , Jun. 2008.
[CCB95]	P.S. Chow, J.M. Cioffi, J.A.C Bingham "A Practical Discrete Multitone Loading Algorithm for Data Transmission over Spectrally Shaped Channels" <i>IEEE Trans. on Communications</i> , Vol 43, No 2/3/4 Feb/mar/Apr 1995.
[CEG79]	T. M. Cover, A.A. El Gamal, "Capacity theorems for the relay channel", <i>IEEE Trans. on</i>

	<i>Information Theory</i> , vol. IT-25, No 5, Sep 1979.
[CS08a]	A. Del Coso, S. Simoens “Distributed Compression for the uplink of a coordinated cellular network with a backhaul constraint” <i>Proceedings of IEEE SPAWC 2008</i> , Jul. 2008.
[CS08b]	A. Del Coso, S. Simoens “Uplink Rate Region of a Coordinated Cellular Network with Distributed Compression” <i>Proceedings of the IEEE Int. Sympos. Information Theory (ISIT)</i> , Jul. 2008.
[CS08c]	A. Del Coso, S. Simoens “Distributed Compression for MIMO Coordinated Networks with a Backhaul Constraint” <i>Submitted to IEEE Trans. on Wireless Communications, Sep. 2008. Accepted for publication in 2009.</i>
[CS08d]	A. del Coso and S. Simoens, “Distributed compression for the uplink of a backhaul-constrained coordinated cellular network,” <i>submitted to IEEE Trans. on Signal Processing</i> , 2008. arXiv:0802.0776
[CSL06]	B. Classon, P. Sartori, R. Love, Y. Sun “Effective OFDM-HARQ System Evaluation using a Recursive EESM Link Error Prediction” <i>Proceedings of IEEE WCNC 2006</i> . Las Vegas, NV, USA. Apr. 2006.
[CT91]	T. Cover, J.A. Thomas “Elements of Information Theory” Wiley-Interscience, 1991
[DDA02]	M. Dohler, B. Rassool and A. Aghvami “Link Capacity Analysis for Virtual Antenna Arrays” <i>Proceedings of IEEE Vehicular Technology Conference (VTC-Fall)</i> , 2002.
[DW04]	S.C.Draper and G.W. Wornell “Side Information aware coding strategies for sensor networks” <i>IEEE Journal on Selected Areas in Communications (JSAC)</i> , vol. 22, no 6, Aug. 2004.
[E03]	Ericsson, “System-level evaluation of ofdm - further considerations” Tech. Rep., 3GPP TSGRAN WG1, Nov. 2003.
[FIR07]	IST Fireworks Project Deliverable 4D3 “Detailed Description of Coding and Modulation Techniques for L1 relaying and Cooperative Communication” Dec 2007. http://fireworks.intranet.gr
[FIR07b]	IST Fireworks Project Deliverable 2D2 “Advanced radio resource management algorithms for relay-based networks” Jul.2007. http://fireworks.intranet.gr
[FIR07c]	IST Fireworks Project Deliverable 2D1 “Cellular deployment concepts for relay-based systems” Jan. 2007. http://fireworks.intranet.gr
[FIR08]	IST Fireworks Project Deliverable 1D3 “FIREWORKS Business Models” Mar. 2008. http://fireworks.intranet.gr
[FKV06]	G.J. Foschini, K. Karakayali and R.A. Valenzuela, “Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency,” <i>IEE Proceedings Communications</i> , vol. 153, no. 4, pp. 548–555, Aug. 2006.
[GDV04]	M. Gastpar, P.L. Dragotti and M. Vetterli “On compression using the distributed Karhunen Loeve Transform” <i>Proceedings of IEEE ICASSP</i> , May 2004
[GDV06]	M. Gastpar, P.L. Dragotti and M. Vetterli “The Distributed Karhunen-Loève Transform” <i>IEEE Trans. on Information Theory</i> , Vol 52, No 12, Dec. 2006.
[GHS06]	D. Gesbert, A. Hjørungnes, S. Skjevling “Cooperative Spatial Multiplexing with Hybrid Channel Knowledge” <i>Proceedings of the International Zurich Seminar on Broadband Communications (IZS)</i> , 2006.
[GMZ06]	A. El Gamal, M. Mohseni, S. Zahedi “Bounds on Capacity and Minimum Energy-Per-Bit for AWGN Relay Channels”. <i>IEEE Trans. on Information Theory</i> , Vol 52, No 4, Apr 2006.
[HG07]	A. Hjørungnes, D. Gesbert “Complex-Valued Matrix Differentiation: Techniques and Key

	Results” <i>IEEE Trans. on Signal Processing</i> , Vol. 55, No 6, part I, pp. 2740-2746, Jun. 2007.
[HK80]	T.S. Han, K. Kobayashi “A unified achievable rate region for a general class of multiterminal source coding systems” <i>IEEE Trans. On Information Theory</i> , vol. IT-26, no 3, May 1980.
[HKE07]	I. Hammerstrom, M. Kuhn, C. Esli, Z. Jihan, A. Wittneben, G. Bauch “MIMO two-way relaying with transmit CSI at the relay” <i>Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)</i> , Jun. 2007.
[HP08]	A. Hjørungnes, D. Palomar “Patterned Complex-Valued Matrix Derivatives” In <i>Proc. for Fifth IEEE Workshop on Sensor Array and Multi-Channel Signal Processing, SAM 2008</i> , (Darmstadt, Germany), Jul. 2008. Avail. at http://brage.unik.no/personer/arehj/
[HW07]	I. Hammerstrom, A. Wittneben “Power Allocation Schemes for Amplify-and-Forward MIMO-OFDM Relay Links”, <i>IEEE Trans. on Wireless Communications</i> , Vol. 6, No. 8, pp. 2798-2802, Aug. 2007.
[HZ05]	A. Høst-Madsen, J. Zhang “Capacity Bounds and Power Allocation for Wireless Relay channels” <i>IEEE Transactions on Information Theory</i> , Vol 51, N°6, pp. 2020-2040, June 2005
[J06]	N. Jindal, “MIMO Broadcast Channels With Finite-Rate Feedback” <i>IEEE Trans. On Information Theory</i> , Vol. 52, No. 11, Nov. 2006.
[JB04]	E.A. Jorswieck, H. Boche “Channel Capacity and Capacity-Range of Beamforming in MIMO Wireless Systems Under Correlated Fading with covariance Feedback” <i>IEEE Trans. On Wireless Commun.</i> , Vol. 3, No. 5, Sep. 2004.
[JVG01]	S.A. Jafar, S. Vishwanath, A. Goldsmith “Channel Capacity and Beamforming for Multiple Transmit and Receive Antennas with Covariance Feedback” <i>IEEE Int. Conf. Commun. (ICC)</i> , 2001.
[K04]	G. Kramer “Models and Theory for Relay Channels with Receive Constraints” Proceedings of the 42nd Allerton Conf. on Commun., Control, and Comp., Sept. 29–Oct. 1, 2004
[KVF06]	K. Karakayali, G.J. Foschini, R.A. Valenzuela, and R.D. Yates, “On the maximum common rate achievable in a coordinated network,” in Proc. IEEE International Conference on Communications (ICC), Turkey, Jun. 2006.
[KGG05]	G. Kramer, M. Gastpar, P. Gupta, “Cooperative Strategies and Capacity Theorems for Relay Networks”, <i>IEEE Trans. on Information Theory</i> , Vol 51. No 9. Sep 2005.
[KM07]	M. Kamoun and L. Mazet, “Base-station selection in cooperative single frequency cellular network,” in Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications, Helsinki, Finland, Jun. 2007.
[LHS03]	D. J. Love, R.W. Heath Jr, T. Strohmer “Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems” <i>IEEE Trans on Information Theory</i> , Vol 49, No 10, Oct 2003.
[LH05]	D.J. Love, R.W. Heath Jr. “Multimode Precoding for MIMO Wireless Systems” <i>IEEE Trans. on Signal Processing</i> , Vol. 53, No. 10, Oct. 2005.
[LSX05]	Z. Liu, V. Stankovic and Z. Xiong “Wyner-Ziv coding for the half-duplex relay channel” <i>Proceedings of IEEE ICASSP</i> , Mar. 2005
[LTV05]	A. Lozano, A.M. Tulino, S. Verdu, “Mercury/waterfilling: optimum power allocation with arbitrary input constellations” <i>IEEE Int. Symposium on Information Theory (ISIT)</i> . Sep 2005.

[LVH05]	C.K. Lo, S. Vishwanath, R.W. Heath Jr. "Rate Bounds for MIMO Relay Channels Using Precoding" <i>Proc. IEEE Globecom'05</i> . Nov 2005.
[LW04]	J. N. Laneman, D. N. C. Tse and G. W. Wornell. "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior" <i>IEEE Trans. On Information Theory</i> , Vol. 50, No 12, Dec 2004.
[MF07]	P. Marsch and G. Fettweis, "A framework for optimizing the uplink performance of distributed antenna systems under a constrained backhaul," in Proc. IEEE International Conference on Communications (ICC), Glasgow, UK, Jun. 2007.
[MS00]	T.K. Moon, W.C. Stirling "Mathematical Methods and Algorithms for Signal Processing" Prentice Hall, 2000.
[MVA07]	O. Muñoz, J. Vidal, A. Agustin, "Linear Transceiver Design in Non-Regenerative Relays with Channel State Information", in <i>IEEE trans. on Signal Processing</i> , Vol. 55, No. 6, Jun. 2007
[N92]	H. Niederreiter "Random Number Generation and Quasi-Monte Carlo Methods" CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.
[NBK04]	Nabar, R.U.; Bolcskei, H.; Kneubuhler, F.W. "Fading relay channels: performance limits and space-time signal design". <i>IEEE Journal on Selected Areas in Communications (JSAC)</i> , Vol. 22, No 6, Aug. 2004
[NHH04]	Nosratinia, T. E. Hunter, A. Hedayat, "Cooperative Communication in Wireless Networks", <i>IEEE Communications Magazine</i> , Oct 2004
[NM93]	F.D. Neeser and J. Massey "Proper Complex Random Processes with Applications to Information Theory" <i>IEEE Trans. On Inform. Theory</i> , Vol. 39, No 4, Jul. 1993.
[NR98]	S. Nanda and K. Rege, "Frame error rates for convolutional codes on fading channels and the concept of effective e_b/n_0 ," <i>IEEE Trans. on Veh. Technol.</i> , vol. 47, no. 4, Nov. 1998.
[PC06]	D.P. Palomar, M. Chiang "A Tutorial on Decomposition Methods for Network Utility Maximization" <i>IEEE Journal on Selected Areas in Communications (JSAC)</i> , Vol. 24, No 8, Aug. 2006
[PCL03]	D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization," <i>IEEE Trans. on Signal Processing</i> , vol. 51, No. 9, Sep. 2003.
[PP08]	K. B. Petersen, M.S. Pedersen "The Matrix Cookbook" Version: February 16, 2008, http://matrixcookbook.com
[RC98]	G.G. Raleigh, J.M. Cioffi "Spatio-Temporal Coding for Wireless communication", <i>IEEE Transactions on Communications</i> , Vol 46, No 3, March 1998.
[RGC02]	R. Ratasuk, A. Ghosh, B. Classon "Quasi-Static Method for Predicting Link-Level Performance" Proceedings of IEEE VTC Spring 2002 Conference. Birmingham, AL. U.S.A.
[RW07]	B. Rankov, A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels", <i>IEEE Journal on Selected Areas in Communications (JSAC)</i> , Vol. 25, No. 2, Feb. 2007.
[S77]	I. Sobol, "Uniformly Distributed Sequences with an Additional Uniform Property", <i>USSR Computational Mathematics and Mathematical Physics</i> , Volume 16, 1977, pages 236-242.
[SEA03]	A. Sendranis, E. Erkip, B. Aazhang "User cooperation diversity. Part I. System description" <i>IEE Trans. on Communications</i> , Vol 51, No 11, Nov. 2003.
[SMV07]	S. Simoens, O. Muñoz, J. Vidal "Achievable Rates of Compress-and-Forward Cooperative Relaying on Gaussian Vector Channels", <i>Proceedings of IEEE Int. Conf. Communications</i>

	(ICC), Jun. 2007.
[SMVC08]	S. Simoens, O. Muñoz, J. Vidal, A. Del. Coso “Capacity Bounds for Gaussian MIMO Relay Channel with full Channel State Information” in <i>Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)</i> , Recife, Brazil, Jul. 2008. Avail. at http://sebastien.simoens.free.fr/publications_simoens.html
[SMVC08b]	S. Simoens, O. Muñoz, J. Vidal, A. Del Coso “On the Gaussian MIMO Relay Channel with full Channel State Information” <i>Submitted to IEEE Trans. on Signal Processing in Sep. 2008. Accepted for publication in 2009.</i> http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4813244&isnumber=4359509.
[SMVC08c]	S. Simoens, O. Muñoz, J. Vidal, A. Del Coso “On Compress-and-Forward Cooperative Relaying in MIMO systems” <i>Submitted to IEEE Trans. on Signal Processing in Nov. 2008. Accepted for publication in 2009.</i>
[SRS05]	S. Simoens, S. Rouquette-Léveil, P. Sartori, Y. Blankenship, B. Classon “Error Prediction for Adaptive Modulation and Coding in Multiple Antenna OFDM Systems” <i>EURASIP Signal Processing Journal, special issue on cross-layer optimization</i> , Aug. 2006
[SSS08]	A. Sanderovich, S. Shamaï, Y. Steinberg, G. Kramer “Communication via Decentralized Processing” <i>IEEE Trans. Inform. Theory</i> , Vol. 54, No 7, Jul. 2008.
[T01]	M.J. Todd “Semidefinite Optimization” in <i>Acta Numerica</i> 10, pp 515-560. Cambridge Univ. Press. 2001.
[T05]	Wen Tong et al. “Duplex and Multiplex Configurations for OFDMA In-Band Relay” <i>IEEE 802.16 MMR study group</i> . Doc nb. C802.16mmr-05/011. Sep. 2005.
[T99]	E. Telatar, “Capacity of multi-antenna gaussian channels” <i>European Trans. Telecommun.</i> , vol. 10, No. 6, Nov/Dec 1999.
[TV05]	D. Tse, P. Viswanath “Fundamentals of Wireless Communications” <i>Cambridge University Press</i> , 2005.
[V71]	E. C. van der Meulen, “Three-terminal communication channels,” <i>Adv. Appl. Prob.</i> , vol. 3, pp. 120-154, 1971.
[VLK07]	S. Valentin, H. S. Lichte, H. Karl, G. Vivier, S. Simoens, J. Vidal, and A. Agustin “Cooperative wireless networking beyond store-and-forward: Perspectives in PHY and MAC design”, <i>Wireless Personal Communications</i> , Nov. 2007
[W78]	A.D. Wyner “The rate-distortion function for source coding with side information at the decoder –II: General Sources” <i>Information and Control</i> , Vol 38, pp 60-80, 1978
[WZ76]	A. D. Wyner and J. Ziv. “The rate-distortion function for source coding with side information at the decoder,” <i>IEEE Trans. on Information Theory</i> , Vol. 22. pp. 1-11. Jan. 1976.
[WZH05]	B. Wang, J. Zhang, A. Høst-Madsen “On the Capacity of MIMO Relay Channels” <i>IEEE Trans. Information Theory</i> , Vol. 51, No 1, Jan 2005.
[XLC04]	Z. Xiong, A. D. Liveris and S. Cheng “Distributed Source Coding for Sensor Networks” <i>IEEE Signal Processing Magazine</i> , Vol 21, No 5, Sep. 2004.
[XS07]	Feng Xue; S. Sandhu “Cooperation in a Half-Duplex Gaussian Diamond Relay Channel”, <i>IEEE Trans. on Information Theory</i> , Vol. 53, No. 10, Oct. 2007.
[YB03]	S. Ye and R. Blum “Optimized Signaling for MIMO Interference Systems with Feedback” <i>IEEE Trans. Signal Processing</i> . Vol 51, No 11, Nov 2003.
[YE07]	M. Yuksel, E. Erkip “Multiple-Antenna Cooperative Wireless Systems: A Diversity–Multiplexing Tradeoff Perspective” <i>IEEE Trans. Inform. Theory</i> , Vol. 53, No 10, Oct 2007.

[YRBC04]	W. Yu, W. Rhee, S. Boyd, J.M. Cioffi “Iterative Water-Filling for Gaussian Vector Multiple-Access Channels” <i>IEEE Trans. Inform. Theory</i> , Vol. 50, No 1, Jan 2004.
[YWK07]	Byung K. Yi, Shu Wang and Soon Y. Kwon “On MIMO Relay with Finite-Rate Feedback and Imperfect Channel Estimation” <i>IEEE Global Conf. on Communications (Globecom)</i> , Nov. 2007.
[ZAL07]	Yi Zhao, Raviraj Adve and Teng Joon Lim “Beamforming with Limited Feedback in Amplify-and-Forward Cooperative Networks” <i>IEEE Global Conf. on Communications (Globecom)</i> , Nov. 2007.
[ZKL08]	A. Zaidi, S. Kotagiri, J.N. Laneman., L. Vandendorpe “Cooperative Relaying with State Available Non-Causally at the Relay”, <i>IEEE Trans. Inf. Theory</i> , 2008