ACHIEVABLE RATES FOR GAUSSIAN CHANNELS WITH MULTIPLE RELAYS

Ph.D. Dissertation by

Aitor del Coso Sánchez

Advisor: Dr. Christian Ibars Casas
Tutor: Prof. Miguel Ángel Lagunas Hernández

Department of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)

Barcelona, June 2008

A mis padres, mi hermana y mis hermanos.

Para Marta.

Summary

Multiple-input-multiple-output (MIMO) channels are extensively proposed as a means to overcome the random channel impairments of *frequency-flat* wireless communications. Based upon placing multiple antennas at both the transmitter and receiver sides of the communication, their virtues are twofold. On the one hand, they allow the transmitter: *i*) to concentrate the transmitted power onto a desired *eigen*-direction, or *ii*) to code across antennas to overcome unknown channel fading. On the other hand, they permit the receiver to sample the signal on the space domain. This operation, followed by the coherent combination of samples, increases the signal-to-noise ratio at the input of the detector. In fine, MIMO processing is able to provide large capacity (and reliability) gains within rich-scattered scenarios.

Nevertheless, equipping wireless handsets with multiple antennas is not always possible or worthwhile. Mainly, due to size and cost constraints, respectively. For these cases, the most appropriate manner to exploit multi-antenna processing is by means of *relaying*. This consists of a set of wireless relay nodes assisting the communication between a set of single-antenna sources and a set of single-antenna destinations. With the aid of relays, indeed, MIMO channels can be mimicked in a distributed way. However, the exact channel capacity of single-antenna communications with relays (and how this scheme performs with respect to the equivalent MIMO channel) is a long-standing open problem. To it we have devoted this thesis.

In particular, the present dissertation aims at studying the capacity of Gaussian channels when assisted by multiple, parallel, relays. Two relays are said to be *parallel* if there is no direct link between them, while both have direct link from the source and towards the destination. We focus on three well-known channels: the *point-to-point* channel, the *multi-access* channel and the *broadcast* channel, and study their performance improvement with relays. All over the dissertation, the following assumptions are taken: *i*) full-duplex operation at the relays, *ii*)

transmit and receive channel state information available at all network nodes, and *iii*) *time-invariant*, *memory-less* fading.

Firstly, we analyze the multiple-parallel relay channel, where a single source communicates to a single destination in the presence of N parallel relays. The capacity of the channel is lower bounded by means of the achievable rates with different relaying protocols, *i.e.* *decode-and-forward*, *partial decode-and-forward*, *compress-and-forward* and *linear relaying*. Likewise, a capacity upper bound is provided for comparison, derived using the *max-flow-min-cut* Theorem. Finally, for number of relays growing to infinity, the scaling laws of all achievable rates are presented, as well as the one of the upper bound.

Next, the dissertation focusses on the multi-access channel (MAC) with multiple-parallel relays. The channel consists of multiple users simultaneously communicating to a single destination in the presence of N parallel relay nodes. We bound the capacity region of the channel using, again, the *max-flow-min-cut* Theorem and find achievable rate regions by means of *decode-and-forward*, *linear relaying* and *compress-and-forward*. In addition, we analyze the asymptotic performance of the obtained achievable sum-rates, given the number of users growing without bound. Such a study allows us to grasp the impact of multi-user diversity on access networks with relays.

Finally, the dissertation considers the broadcast channel (BC) with multiple parallel relays. This consists of a single source communicating to multiple receivers in the presence of N parallel relays. For the channel, we derive achievable rate regions considering: *i*) dirty paper encoding at the source, and *ii*) *decode-and-forward*, *linear relaying* and *compress-and-forward*, respectively, at the relays. Moreover, for *linear relaying*, we prove that MAC-BC duality holds. That is, the achievable rate region of the BC is equal to that of the MAC with a sum-power constraint. Using this result, the computation of the channel's weighted sum-rate with *linear relaying* is notably simplified. Likewise, convex resource allocation algorithms can be derived.

Resumen

Los canales múltiple-entrada-múltiple-salida (MIMO) han sido ampliamente propuestos para superar los desvanecimientos aleatorios de canal en comunicaciones inalámbricas no selectivas en frecuencia. Basados en equipar tanto transmisores como receptores con múltiple antenas, sus ventajas son dobles. Por un lado, permiten al transmisor: *i*) concentrar la energía transmitida en una dirección-*propia* determinada, o *ii*) codificar entre antenas con el fin de superar desvanecimientos no conocidos de canal. Por otro lado, facilitan al receptor el muestreo de la señal en el dominio espacial. Esta operación, seguida por la combinación coherente de muestras, aumenta la relación señal a ruido de entrada al receptor. De esta forma, el procesado multi-antena es capaz de incrementar la capacidad (y la fiabilidad) de la transmisión en escenarios con alta dispersión.

Desafortunadamente, no siempre es posible emplazar múltiples antenas en los dispositivos inalámbricos, debido a limitaciones de espacio y/o coste. Para estos casos, la manera más apropiada de explotar el procesado multi-antena es mediante *retransmisión*, consistente en disponer un conjunto de repetidores inalámbricos que asistan la comunicación entre un grupo de transmisores y un grupo de receptores, todos con una única antena. Con la ayuda de los repetidores, por tanto, los canales MIMO se pueden imitar de manera distribuida. Sin embargo, la capacidad exacta de las comunicaciones con repetidores (así como la manera en que este esquema funciona con respeto al MIMO equivalente) es todavía un problema no resuelto. A dicho problema dedicamos esta tesis.

En particular, la presente disertación tiene como objetivo estudiar la capacidad de canales Gaussianos asistidos por múltiples repetidores paralelos. Dos repetidores se dicen paralelos si no existe conexión directa entre ellos, si bien ambos tienen conexión directa con la fuente y el destino de la comunicación. Nos centramos en el análisis de tres canales ampliamente conoci-

dos: el canal *punto-a-punto*, el canal de *múltiple-acceso* y el canal de *broadcast*, y estudiamos su mejora de funcionamiento con repetidores. A lo largo de la tesis, se tomarán las siguientes hipótesis: *i*) operación full-duplex en los repetidores, *ii*) conocimiento de canal tanto en transmisión como en recepción, y *iii*) desvanecimiento sin memoria, e invariante en el tiempo.

En primer lugar, analizamos el canal con múltiples repetidores paralelos, en el cual una única fuente se comunica con un único destino en presencia de N repetidores paralelos. Derivamos límites inferiores de la capacidad del canal por medio de las tasas de transmisión conseguibles con distintos protocolos: *decodificar-y-enviar, decodificar-parcialmente-y-enviar, comprimir-y-enviar*, y *repetición lineal*. Asimismo, con un fin comparativo, proveemos un límite superior, obtenido a través del Teorema de *max-flow-min-cut*. Finalmente, para el número de repetidores tendiendo a infinito, presentamos las leyes de crecimiento de todas las tasas de transmisión, así como la del límite superior.

A continuación, la tesis se centra en el canal de múltiple-acceso (MAC) con múltiples repetidores paralelos. El canal consiste en múltiples usuarios comunicándose simultáneamente con un único destino en presencia de N repetidores paralelos. Derivamos una cota superior de la región de capacidad de dicho canal utilizando, de nuevo, el Teorema de *max-flow-min-cut*, y encontramos regiones de tasas de transmisión conseguibles mediante: *decodificar-y-enviar, comprimir-y-enviar, y repetición lineal*. Asimismo, se analiza el valor asintótico de dichas tasas de transmisión conseguibles, asumiendo el número de usuarios creciendo sin límite. Dicho estudio nos permite intuir el impacto de la diversidad multiusuario en redes de acceso con repetidores.

Finalmente, la disertación considera el canal de broadcast (BC) con múltiples repetidores paralelos. En él, una única fuente se comunica con múltiples destinos en presencia de N repetidores paralelos. Para dicho canal, derivamos tasas de transmisión conseguibles dado: *i*) codificación de canal tipo *dirty paper* en la fuente, *ii*) *decodificar-y-enviar, comprimir-y-enviar*, y *repetición lineal*, respectivamente, en los repetidores. Además, para repetición lineal, demostramos que la dualidad MAC-BC se cumple. Es decir, la región de tasas de transmisión conseguibles en el BC es igual a aquélla del MAC con una limitación de potencia suma. Utilizando este resultado, se derivan algoritmos de asignación óptima de recursos basados en teoría de optimización convexa.

Acknowledgements

Como una mesa, o como un puente, el proyecto vital que aquí termino se ha sustentado en cuatro pilares fundamentales. A los cuatro dedico íntegramente esta tesis y a ellos manifiesto mi más profunda gratitud. Quizá mejor decir cuatro y medio; contando como medio a la maravillosa ciudad de Barcelona, que tanto me ha dado y tan poco me ha pedido. Enumero los pilares de mi mesa:

Mi familia. Sin duda, mi mayor soporte y a la que debo todo: lo que fui, lo que soy y lo que seré. A ella le dedico esta tesis: a mi madre por confiar siempre en mí, y por todo el amor que me da; a mi padre, por ser mi referente en la vida, enseñarme el camino y permitirme tener un modelo que imitar. A mis hermanos, de sangre y políticos, por sostenerme y agruparme, por quererme y dejarme que les quiera, por luchar (entre todos) por la felicidad de todos.

Marta. El amor de mi vida. Sin más. La felicidad por terminar esta etapa es diminuta comparada con la felicidad de tenerte a mi lado. Nunca seré capaz de agradecerte, en tu justa medida, el maravilloso camino en que has convertido mi vida.

Centre Tecnològic de Telecomunicacions de Catalunya. Realizar el doctorado aquí ha sido, sin duda, la elección profesional más acertada que he tomado jamás. Aquí he encontrado un entorno inmejorable de aprendizaje, superación y respeto. Además de un fantástico ambiente. Personalizo mis agradecimientos en dos personas que me han demostrado, de manera sistemática, su apoyo y su confianza: Prof. M.A. Lagunas y Dr. Carles Antón-Haro. En el CTTC he aprendido innumerables cosas, relacionadas unas y no relacionadas otras con su bien conocida excelencia en la investigación.

Dr. Christian Ibars. Mi tutor de tesis. El artífice de que esta tesis tenga el fondo que tiene, se lea como se lee, estudie lo que estudia. *Ha estat tot un plaer aprendre d'ell aquests darrers* *quatre anys; treballar amb ell i crèixer amb ell. Li desitjo fortuna pel futur i tinc l'esperanca que podrem seguir treballant junts, d'una forma o d'altra. També aprofito aquestes línies per felicitar-lo, tant a ell com a la Montse, per la nova família que els espera.*

Aparte de los cuatro sustentos fundamentales de mi doctorado, hay mucha gente a la que debería agredecer todo su apoyo y cariño. Empezando por el grupo humano del CTTC, pasando por el equipo de fútbol y terminando por todos mis viejos y nuevos amigos. Sin embargo, a fin de no convertir esto en una enumeración interminable, resaltaré únicamente las más determinantes. En primer lugar, quiero agredecer a Marc y a Óscar que hayan sido casi hermanos para mí. Han multiplicado mi vida aquí en Barcelona y estoy enormemente agradecido. En segundo lugar, quiero recordar a mis compañeros David Gregoratti, Javier Matamoros, Nizar Zorba y, de nuevo, a Marc Portolés, por enseñarme tanto en las discusiones que hemos tenido juntos. Por mostrarme, sin reservas, sus puntos de vista sobre mi investigación y rebatir mis (casi siempre) desacertados pensamientos. Esta tesis ha mejorado, y de qué manera, gracias a su aporte.

In third place, many thanks are due to Dr. Osvaldo Simeone for his kind support during my stay at the NJIT, and thereafter. I have learned tons of useful knowledge from him and I have enjoyed thousand of intelligent discussions with him (obviously, "intelligent" on his side). He is a superb teacher, and an impressively speddy reviewer/feedback-provider. Thanks!. Likewise, I'm in debt with Prof. U. Spagnolini, from PoliMi, for making my short-stay in the beautiful city of Milano so profitable.

Finalement, je voudrais bien remercier Sebastien Simoens de son appui pendant mon merveilleux stage à Motorola Labs, à Paris. Là-bas et à côté de lui, j'ai énormememt appris et je me suis beaucoup amusé. En mettant en contrast très vivement ses idées avec les miennes, il a été capable de obtenir le mieux de moi-même. J'ai beaucoup plu de travailler en équipe avec lui.

I end this never-ending acknowledgements sheet thanking Prof. Ana I. Pérez-Neira, Prof. Erik. G. Larsson, Dr. Mischa Dohler, Dr. Osvaldo Simeone, Dr. A. Pascual-Iserte, Dr. M. Navarro and Dr. N. Zorba for taking part of my Ph.D. defense committee. They have provided priceless feedback to improve the quality of this dissertation.

Finiquito citando[1] un poema de Kipling que, para mí, retrata perfectamente los dos principios irrenunciables que me han acompañado en esta aventura: sinceridad conmigo mismo, y honestidad con los demás. Ahí va el poema (titulado *If*):

Si puedes mantener la cabeza sobre los hombros
cuando otros la pierden y te cargan su culpa.
Si confías en tí mismo aún cuando todos de tí dudan,
pero aún así tomas en cuenta sus dudas.
Si puedes soñar y no hacer de tus sueños tu guía,
si puedes pensar sin hacer de tus pensamientos tu meta.
Si triunfo y derrota se cruzan en tu camino
y tratas de igual manera a ambos impostores.
Si puedes hacer un montón con todas tus victorias.
Si puedes arrojarlas al capricho del azar, y perder,
y remontarte de nuevo a tus comienzos
sin que salga de tus labios una queja.
Si logras que tus nervios y el corazón sean tu fiel compañero
y resistir aunque tus fuerzas se vean menguadas,
con la única ayuda de la volutad que dice: Adelante!
Si ante la multitud das a la virtud abrigo,
si áun marchando con reyes guardas tu sencillez.
Si no pueden herirte ni amigos ni enemigos,
si todos te reclaman y ninguno te precisa.
Si puedes rellenar un implacable minuto
con sesenta segundos de combate bravío,
tuya es la tierra y sus codiciados frutos
y lo que es más: serás un hombre, hijo mío!

Aitor del Coso

Castelldefels, 26 de junio de 2008.

[1]He de reconocer que citar un autor que apenas he leído habla muy mal de mi persona.

Contents

List of Figures

Glossary

Tabulated below are the most important acronyms used in this dissertation.

A&F	Amplify-and-forward
BC	Broadcast channel
C&F	Compress-and-forward
D&F	Decode-and-forward
D-WZ	Distributed Wyner-Ziv
FD	Full-duplex
GP	Gradient projection
HD	Half-duplex
KKT	Karush-Kuhn-Tucker
LoS	Line of sight
LR	Linear relaying
MAC	Multi-access channel
MIMO	Multiple-input-multiple-output
MPRC	Multiple-parallel relay channel
MPR-BC	Broadcast channel with multiple-parallel relays
MPR-MAC	Multi-access channel with multiple-parallel relays
OFDM	Orthogonal frequency division multiplexing
PD&F	Partial decode-and-forward
SNR	Signal-to-noise ratio

Notation

Tabulated below are the most important notations and symbols used in this dissertation.

a	Scalar notation
a^*	Conjugate of a
$\boldsymbol{a}$	Vector notation
$\boldsymbol{a}_{1:N}$	$[a_1, \cdots, a_N]$
$\boldsymbol{a}_{\mathcal{G}}$	$\{a_i : \ a_i \in \mathcal{G}\}$
$\boldsymbol{A}$	Matrix notation
$[\boldsymbol{A}]_{r,c}$	The element in row r and column c of matrix $\boldsymbol{A}$
$\boldsymbol{A}^T$	Transpose of $\boldsymbol{A}$
$\boldsymbol{A}^\dagger$	Conjugate transpose of $\boldsymbol{A}$
$\boldsymbol{A}^\sharp$	Conjugate transpose of $\boldsymbol{A}$ with respect to the opposite diagonal
	e.g., $\boldsymbol{A} = [a_1, \ a_2 \ ; \ a_3, \ a_4]$ then $\boldsymbol{A}^\sharp = [a_4^*, \ a_2^* \ ; \ a_3^*, \ a_1^*]$.
$\mathbb{C}$	The complex numbers
$\mathbb{C}_{\text{SLT}}^{n \times n}$	The set of complex, strictly lower triangular, square matrices of dimension n
$\mathbb{C}_{\text{SUT}}^{n \times n}$	The set of complex, strictly upper triangular, square matrices of dimension n
$\mathcal{C}(x)$	$\log_2(1+x)$
$\mathcal{CN}(\eta, \sigma^2)$	Complex Gaussian distribution with mean η and variance σ^2
c.d.f.	Cumulative density function
$\mathrm{coh}(\cdot)$	Convex hull
$\det(\cdot)$	Determinant
$\mathrm{diag}(\boldsymbol{a})$	Diagonal matrix whose diagonal elements are the elements of vector $\boldsymbol{a}$
$\mathrm{E}\{\cdot\}$	Expectation

$H(x)$	Differential Entropy: $-\int p(x) \log_2 p(x)\, dx$
i.i.d	Independently, identically distributed
$I(x;y)$	Differential mutual information: $\int p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} dxdy$
$\log(\cdot)$	Natural logarithm
$\log_2(\cdot)$	Logarithm in base 2
$\max$	Operator that takes the maximum of functions or vectors
$\min$	Operator that takes the minimum of functions or vectors
p.d.f.	Probability density function
$\Pr\{\cdot\}$	Probability
$p_x(x)$	Probability density function of random variable x
$p(x)$	Probability density function of random variable x
$\mathbb{R}$	The real numbers
s.t.	Subject to
$\mathrm{tr}\{\cdot\}$	Trace
w.l.o.g	Without loss of generality
$W_0(\cdot)$	Branch zero of the Lambert W function
$\mathcal{X}$	Set notation
$[x]^+$	$\max\{0, x\}$
$\mathbf{1}_n$	$n \times 1$ vector whose elements are all 1

Chapter 1

Introduction

1.1 Motivation

Multiple-antenna processing is widely used to boost performance of telecommunications systems: from remote sensing, radar and radio-astronomy to wireless communications and media broadcasting. Serve as example the stunning *Very Large Array* (VLA) that the National Radio Astronomy Observatory (NRAO) owns in New Mexico to explore the confines of the universe.

Virtues of deploying multiple antennas are many and wide. First, it allows the transmitter to concentrate the transmitted power onto a desired spatial direction; thus, augmenting the received signal-to-noise ratio at the intended receiver and drastically reducing the interference onto non-desired ones. This processing is referred to as *beamforming*, for which channel knowledge is required on transmission [2, Chapter 5]. With line-of-sight (LoS) propagation, it consists of conjugating the steering vector of the desired direction. In contrast, with non-LoS and rich-scattering, the beamforming consists of transmitting along the eigen-vectors of the channel matrix [3]. Unfortunately, the source node is not always aware of the channel values.

For those cases, multi-antenna processing takes the form of space-time coding [4–6], consisting of coding the transmitted message along different antennas and channel uses. Examples of these codes are space-time (orthogonal) block codes and space-time trellis codes [7]. With them, deep fading on several antennas can be overcome by good channels on others. Thus, *extra* diversity is introduced into the system (referred to as spatial diversity), which makes communication more reliable.

On the other side, having multiple antennas at the receiver allows to coherently combine the received samples, increasing the signal-to-noise ratio at the input of the detector [7, Chapter 5]. This is referred to as *coherent detection*, and receive channel knowledge is required. Likewise, multiple receive antennas can be used for parameter estimation such as *e.g.,* direction of arrival. In fact, with multiple antennas, the receiver samples the signal onto two dimensions: time and space. Accordingly, the number of available samples increases, thus boosting the performance of detectors and/or estimators [8]. The outstanding feature of this technology is that gains come without sacrificing additional degrees of freedom as time or frequency.

When applied to wireless communications, multi-antenna processing augments the channel capacity in two manners: on the one hand, with transmit and receive channel knowledge, beamforming and coherent detection increases the available power at the receiver node, thus allowing for higher data rates. On rich-scattering scenarios, this is translated into the parallelization of the communication onto its eigen-channels [3]. On the other hand, without channel knowledge at the transmitter, space-time coding overcomes fading by stabilizing the received power at the destination. Thus, it increases the set of rates that can be reliably transmitted. Unfortunately, for this case, the more reliability required, the lower the transmitted rate can be, and *viceversa*. This concept is known as the diversity-multiplexing tradeoff [9].

Nonetheless, deploying multiple antennas at the wireless handsets (*e.g.* mobile telephones, GPS receivers, WLAN cards) is not always possible due to size constraints. Likewise, in some cases, their deployment is not worthwhile regarding a cost-performance criterion. For these cases, the most suitable manner to exploit spatial diversity is by means of *relaying* [10–12]. Relay channels consist of a set of wireless relay nodes assisting the communication between a set of single-antenna sources and a set of single-antenna destinations. With such a setup, multi-antenna processing can be thus performed distributedly [13, 14]. As an example, consider the network in Fig. 1.1 where a user S wants to transmit a message to user D, and makes use of

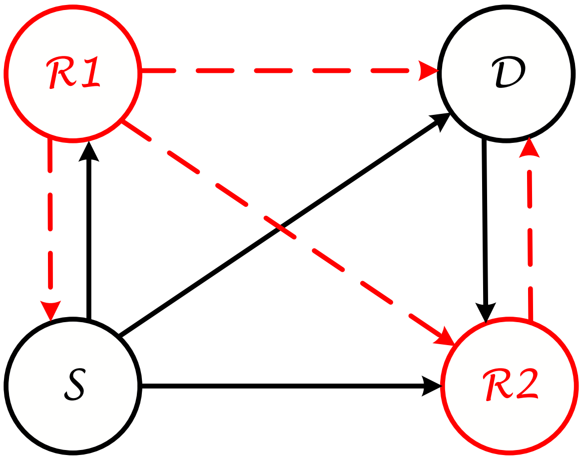

Figure 1.1: Relay channel.

relay nodes R_1 and R_2. The node R_1 helps the source to transmit the message by means of *e.g.* distributed beamforming. On the other hand, relay R_2 helps node D to receive it using *e.g.,* distributed coherent detection. For such a communication with relays, a fundamental query rapidly comes up: how close are the capacity and robustness of this setup to those of the equivalent MIMO channel?. Or, in other words, can MIMO communications be mimicked using distributed antennas?. Answering this question is the main aim and motivation of this dissertation.

Indistinctively, the relay nodes can be either infrastructure nodes, placed by the service provider in order to enhance coverage and rate [15], or a set of users that cooperate with the source and destination, while having own data to transmit [16]. Nonetheless, in order to apply relaying to current wireless networks, still a number of practical issues need to be addressed. Those issues are out of the scope of this dissertation, which is focused on the theoretical performance only. However, for completeness, we briefly highlight some of them:

1. *How to synchronize spatially separated antennas*: beamforming and space-time coding require symbol synchronization among the transmit antennas. When building them up using physically separated relays, this leads to a distributed synchronization problem that implies serious difficulties in practice. Indeed, for single carrier communications with high-data rates, synchronization becomes unfeasible. However, the problem can be softened resorting to OFDM modulation, due to the higher length of the multicarrier symbol.

2. *How to implement distributed channel state information acquisition:* when decoding a space-time signal, the receiver must be aware of the channel with all the transmit antennas. In relay-based transmissions, this forces the receiver to know *a priori* the set of relays involved in the communication. This is an unlike scenario for relaying, which generally takes an opportunistic approach: relays transmit or do not transmit based on local decisions, without centralized control [17]. To solve the problem, and decode without channel knowledge, differential space-time codes can be utilized [18]. However, those have a 3 dB penalization.

3. *How to merge routing and PHY layer forwarding*: the optimum selection of nodes performing layer 3 routing, and the selection of nodes performing PHY layer relaying is an involved cross-layer research topic.

4. *How to provide incentives for cooperation among selfish nodes*: incentive mechanisms must be put in place in order to make egoistic nodes relay each others' messages, and to ensure a fair, reciprocal exchange of communication resources.

1.2 Previous Work

The study of relay channels dates back to the 70s, when Cover, El Gamal and Van der Meulen introduced the three-terminal network [10, 19, 20]. This is composed of one source, one relay and one destination, all three with a single antenna. For the network, the authors presented several relaying techniques (namely, *decode-and-forward D&F* and *compress-and-forward C&F*) and compared their achievable rates with the *max-flow-min-cut* capacity bound [21]. Years later, other relaying protocols such as *partial decoding PD*, *amplify-and-forward A&F* and *linear relaying LR*, helped to widen the analysis [22–24]. Among all, *D&F* and *C&F* turned out to be asymptotically optimal for the relay infinitely close to source and destination, respectively, while *PD* was shown to outperform all other techniques for the half-duplex relay [22]. The capacity of the channel, however, remained an open problem.

The extension of the results in [10] to multiple relays also attracted a lot of attention [11,25–34]. Two main approaches have been followed: the *multi-level relay* model[1] [25] and the *multiple-*

[1] *Multi-level relaying* has been referred to as *Information-theoretic multi-hopping* in [27]

parallel relay model [26]. The former considers every relay node receiving information from the source and/or a subset of relays and forwarding it to the destination and/or another subset of relays. Every subset of relays is referred to as a *level-set*, across which information *flows* in the network [28]. Intuitively, it brings down classical *multi-hop* transmission to the physical-layer. The capacity of the multi-level channel is also unknown. However, an achievable rate with *decode-and-forward* is derived by Xie and Kumar in [25, Theorem 1], and with *compress-and-forward* and superposition coding schemes by Kramer *et. al.* in [11, Theorems 3 and 4]. Moreover, as for the single relay case, *decode-and-forward* is capacity-achieving for the degraded *multi-level* relay channel [25, Theorem 3.2].

On the other hand, the *multiple-parallel* relay channel arranges all the relay nodes into a unique set. That is, relays receive information only from the source, and retransmit directly to the destination. This channel was pioneered by Schein and Gallager in [26, 35] and accounts for scenarios where relays have directional antennas to the destination (as *e.g.*, cellular uplink channels with relays) or where delay constraints do not allow multi-hop transmission. The capacity of the channel is unknown, and few achievable rates have been reported so far: for N relays with orthogonal components, *decode-and-forward* and *amplify-and-forward* were studied by Maric and Yates in [29] and [30], respectively. Moreover, *decode-and-forward* with half-duplex relays is considered in [31]. Finally, the *multiple-parallel* relay channel capacity, although unknown, is shown by Gastpar to scale as the logarithm of the number of relays [32], given Rayleigh fading. A variant of the channel, where relays are independently connected to the destination via lossless links of limited capacity, is considered by Sanderovich *et. al.* in [36–38]. For such a network, authors present achievable rates with *compress-and-forward*, assuming ergodic and block-fading, respectively, and considering distributed Wyner-Ziv compression at the relays [39]. We build upon this result to derive the multi-relay *compress-and-forward* results of the dissertation.

The application of all previous results to schemes with multiple sources (MAC) and multiple destinations (BC) recently witnessed an immense growth of interest [11, 40–44]. First, the MAC assisted by a relay is presented in [11, 40] where, assuming *time-invariant* fading and channel state information (CSI) at the sources, the rate regions with *D&F* and Wyner-Ziv *C&F* are derived. Additionally, *A&F* is analyzed in [41] for the MAC without CSI at the sources, and shown to be optimal at the high multiplexing gain regime. On the other hand, the BC with relay is first studied by Liang *et al.* in [42]. A network with a single source and two receivers, one acting as relay of the other, is considered and rate regions derived using *D&F* and *PD*. In parallel, *A&F* for the two-hop BC (*i.e.*, no connectivity between source and destinations) is studied by Jafar *et al.* in [43]. In that work, the optimal power allocation on relays is presented, as well as the celebrated MAC-BC duality. In this dissertation, we extend this duality result to the BC with direct link between the transmitter and the receivers.

1.3 Organization of the Dissertation

This dissertation aims at studying the capacity of Gaussian channels when assisted by multiple, parallel, relays. Two relays are said to be *parallel* if there is no direct link between them, while both have direct link from the source and towards the destination. In particular, our analysis focuses on three well-known channels: the *point-to-point* channel, the *multi-access* channel and the *broadcast* channel. For them, we present achievable rates and capacity outer regions. All over the dissertation, we assume transmit and receive channel state information available at all network nodes, and *time-invariant*, *memory-less* fading. The research contributions of this dissertation are presented in Chapters 3-5. Additionally, Chapter 2 introduces the necessary mathematical background to follow the derivations, and Chapter 6 summarizes conclusions and outlines future work.

Research results are organized as follows:

Chapter 3. This part of the dissertation analyzes the multiple-parallel relay channel, where a single source communicates to a single destination in the presence of N parallel relays. All network nodes have a single antenna. The capacity of the channel is upper bounded using the *max-flow-min-cut* Theorem, and lower bounded by means of the achievable rates with *decode-and-forward*, *partial decode-and-forward*, *compress-and-forward* and *linear relaying*. Additionally, for $N \to \infty$, the asymptotic performance of all achievable rates is presented. Finally, results are extended to relay nodes under a half-duplex constraint.

The contributions of this chapter are published in part on:

- **A. del Coso** and C. Ibars, ”Achievable rates for the AWGN channel with multiple parallel relays", *submitted to IEEE Trans. on Wireless Communications*, Feb., 2008.

- **A. del Coso** and C. Ibars, "Distributed antenna channels with regenerative relaying: relay selection and asymptotic capacity" , *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, Nov.
- **A. del Coso** and S. Simoens, "Distributed compression for the uplink channel of a coordinated cellular network with a backhaul constraint", *in Proc. 9th IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC2008)*, Recife, Brazil, Jul. 2008
- **A. del Coso**, and C. Ibars, "Partial decoding for synchronous and asynchronous Gaussian multiple relay channels", *in Proc. IEEE International Conference in Communications (ICC07)*, Glasgow, UK, Jun. 2007
- **A. del Coso**, C. Ibars, "Achievable rate for Gaussian multiple relay channels with linear relaying functions", *in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, Hawai'i,USA, Apr., 2007
- **A. del Coso**, C. Ibars, "Capacity of Decode-and-forward Cooperative Links with full channel state information", *in Proc. 39th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2005

Furthermore, while developing the material of this chapter, I co-authored the partial extension of previous results to the MIMO relay channel. Such an extension belongs to the Ph.D. dissertation of Mr. Sebastien Simoens:

- S. Simoens, O. Muñoz, J. Vidal and **A. del Coso**, "Capacity bounds for Gaussian MIMO relay channel with channel state information", *in Proc. 9th IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC2008)*, Recife, Brazil, Jul. 2008.
- S. Simoens, O. Muñoz, J. Vidal and **A. del Coso**, "Partial compress-and-forward cooperative relaying with full MIMO channel state information", *submitted to IEEE Globecom 2008*, New Orleans, Nov. 2008

Chapter 4. This part of the dissertation focusses on the multi-access channel with multiple-parallel relays. It consists of multiple users simultaneously communicating to a single destination in the presence of N parallel relay nodes. We bound the capacity region of the channel using, again, the *max-flow-min-cut* Theorem and find achievable rate regions by means of *decode-and-forward*, *linear relaying* and *compress-and-forward*. We omit partial decoding as it is shown in Chapter 3 to perform exactly as D&F for large number of relays. Additionally, we study the asymptotic achievable sum-rates of the channel when the number of users grows to infinity, thus estimating the impact of multi-user diversity on relay-based access networks.

The research results of this chapter are published in part on:

- **A. del Coso** and C. Ibars, ”Linear relaying for the Gaussian multiple access and broadcast channels”, *submitted to IEEE Trans. on Wireless Communications*, Mar., 2008.
- **A. del Coso** and S. Simoens, ”Distributed compression for the uplink of a backhaul-constrained coordinated cellular network”, *submitted to IEEE Trans. on Wireless Communications*, Jun. 2008
- **A. del Coso**, and C. Ibars, ”Multi-access channels with multiple decode-and-forward relays: rate region and asymptotic sum-rate”, *in Proc. 9th IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC2008)*, Recife, Brazil, Jul. 2008
- **A. del Coso** and C. Ibars, ”The amplify-based multiple-relay multiple-access channel: capacity region and MAC-BC duality”, *in Proc. IEEE Information Theory Workshop (ITW2007)*, Bergen, Norway, Jul., 2007
- Z. Ahmad, **A. del Coso**, and C. Ibars, ”TDMA network design using decode-and-forward relays with finite set modulation”, *in Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2008)*, Cannes, France, Sep. 2008

Chapter 5. In this final part, we consider the broadcast channel with multiple parallel relays. This consists of a single source communicating to multiple receivers in the presence of N parallel relays. We derive achievable rate regions for the channel considering: *i*) *decode-and-forward*, *linear relaying* and *compress-and-forward*, respectively, at the relays and *ii*) dirty paper encoding at the source. Interestingly, we demonstrate that, for linear relaying, MAC-BC duality holds, which simplifies the computation of the BC weighted sum-rate.

The results of this Chapter are published in part on two publications mentioned for Chapter 4:

- **A. del Coso** and C. Ibars, ”Linear relaying for the Gaussian multiple access and broadcast channels”, *submitted to IEEE Trans. on Wireless Communications*, Mar., 2008.
- **A. del Coso** and C. Ibars, ”The amplify-based multiple-relay multiple-access channel: capacity region and MAC-BC duality”, *in Proc. IEEE Information Theory Workshop (ITW2007)*, Bergen, Norway, Jul., 2007

Research results not included in the Dissertation. At the very beginning of my graduate studies, I chose *wireless cooperative networks* to be the major subject of my research. In particular, I thought of *e.g.,* outage analysis for cooperative wireless sensor networks (WSN), diversity-multiplexing trade-off in cooperative multicasting, OFDM relay networks, distributed receiver architectures, code design, etc.

However, once I became more and more involved with cooperative communications (and after attending a masterly class from Pr. Abbas El Gamal) I realized that Information Theory needed to be the mandatory minor to be pursued. At the end, the minor has transmuted into the major. Nonetheless, during the transmutation, I got time to research on cooperative sensor networks. In particular, under the kind support of Prof. U. Spagnolini, I derived the diversity order of virtual MIMO channels in cluster-based sensor networks. Additionally, I studied resource allocation strategies for those networks under energy constrains. Results are published in:

- **A. del Coso**, U. Spagnolini and C. Ibars ”Cooperative distributed MIMO channels in wireless sensor networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 25, no. 2, pp.402-414, Feb. 2007
- **A. del Coso**, S. Savazzi, U. Spagnolini and C. Ibars, ”Virtual MIMO channels in cooperative multi-hop wireless sensor networks”, *in Proc. 40th Annual Conference in Information Sciences and Systems (CISS)*, Princeton, USA, Mar. 2006.

- **A. del Coso** and C. Ibars, ”Spatial diversity of cluster-based cooperative wireless sensor networks”, *in Proc. MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks*, Berkeley, USA, Apr. 2006

Finally, while on intern in New Jersey, I analyzed cooperative multicast networks under the guidance of Prof. Y. Bar-ness and Dr. O. Simeone. The results derived therein are published in part on:

- **A. del Coso**, O. Simeone, Y. Bar-ness and C. Ibars, ”Space-time coded cooperative multicasting with maximal ratio combining and incremental redundancy”, *in Proc. IEEE International Conference in Communications (ICC07)*, Glasgow, UK, Jun. 2007
- **A. del Coso**, O. Simeone, Y. Bar-ness and C. Ibars, ”Outage capacity of two phase space-time coded cooperative multicasting”, *in Proc. 40th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2006

Chapter 2

Background

2.1 Channel Capacity

Claude E. Shannon presented his Mathematical Theory of Communications [1] in 1948 and launched digital communications as understood nowadays. Pioneering modern times, Shannon addressed the fundamental problem of communication: *how to reproduce at one point either exactly or approximately a message selected at another point, regardless of its meaning* [1, p. 379]. His theory involved a big bang on communications, whose direct and indirect contributions (incredibly huge and wide) have boosted telecommunications, signal processing and network design. In particular, the two fundamental starting points of digital communications are the result of Shannon's imagination:

1. The differential entropy as the measure of the uncertainty of a continuous random variable

$$H\left(x\right) = -\int_{-\infty}^{\infty} p\left(x\right) \log_2 p\left(x\right) dx. \tag{2.1}$$

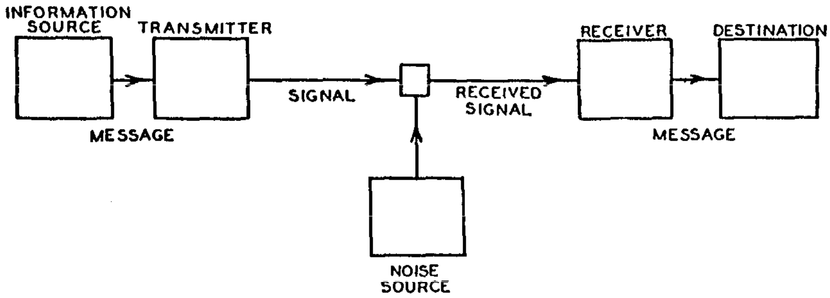

Figure 2.1: Schematic diagram of a communication system. *Source: C.E. Shannon* [1].

With such a selection, Shannon aimed at finding a continuous measure on $p(x)$ with maximum uncertainty for the uniform distribution (subject to a give support set). However, it is the relationship between the entropy and the number of "typical" sequences generated from independent realizations of this variable, what makes the measure so useful [21, Theorem 9.2.1].

2. The schematic diagram of a general communication system (see Fig. 2.1). This consists essentially in five parts [1, Section 1]: *i*) an information source, *which produces a message or sequence of messages to be communicated to the receiving terminal*, *ii*) a transmitter, which *operates on the message to produce a sequence of signals suitable for transmission over the channel*. We refer to this operation as encoding. *iii*) The channel, which *is merely the medium used to transmit the signal from transmitter to receiver*. *iv*) The receiver, *which performs the inverse operation of that done by the transmitter, decoding the message from the signal*, and *v*) the destination, to *whom the message is intended*.

Two main reasons brought Shannon to these two concepts. On the one hand, he aimed at validating symbolic communications. That is, the process of mapping the messages to be transmitted onto symbols, which are random under the receiver glance and has to be deciphered by him (as for *e.g.* telegraphy or Morse). On the other hand, by measuring the uncertainty of random variables, he responded to the necessity of quantifying the unawareness at the receiver of the random symbol transmitted by the source. Merging both concepts, Shannon put the foundations of digital communications, introducing the concept of *reliable transmission*: how many messages can be transmitted from a source towards a destination without errors.

In the following, we revisit three important results derived by, or due to, Shannon's work.

2.1.1 Gaussian Channel

Consider a single source communicating to a single destination. Assume that, for every signal x transmitted by the transmitter, the receiver receives:

$$y = h \cdot x + z, \tag{2.2}$$

where $z \sim \mathcal{CN}(0, N_o)$ is additive white Gaussian noise, and h is a non-random complex scalar. This is referred to as Gaussian channel with *time-invariant* fading.

Let the source randomly select message $\omega \in \mathcal{W} = \left\{1, \cdots, 2^{nR}\right\}$ for transmission, given $p_{\mathcal{W}}(\omega) = \frac{1}{2^{nR}}, \forall\omega$. As stated by Shannon, the message is then passed from the source to the transmitter, which encodes it using the mapping:

$$f : \mathcal{W} \to \mathcal{X}^n, \tag{2.3}$$

where $\mathcal{X}$ is the transmitter signal space and n the number of symbols within the transmitter signal sequence. In general, the mapping is constrained to satisfy:

$$\frac{1}{n}\sum_{t=1}^{n} |x^t(\omega)|^2 \leq P, \ \ \forall\omega. \tag{2.4}$$

The signal sequence $\boldsymbol{x}^n = f(\omega)$ is then plugged onto the channel, and received at the destination following (2.2). Finally, the receiver attempts to recover the message using the de-mapping:

$$g : \mathcal{Y}^n \to \mathcal{W}, \tag{2.5}$$

and an error occurs whenever $g(\boldsymbol{y}^n) \neq \omega$.

The mapping and de-mapping functions, f and g, together with the message set $\mathcal{W} = \left\{1, \cdots, 2^{nR}\right\}$ and the signal space $\mathcal{X}$, are compactly referred to as $\left(n, 2^{nR}\right)$ channel code. A transmission rate R [bits/symbol] is said to be achievable if there exists a sequence of channel codes $\left(n, 2^{nR}\right)$ for which

$$\lim_{n\to\infty} \frac{1}{2^{nR}} \sum_{\omega} \Pr\left\{g(\boldsymbol{y}^n) \neq \omega | \omega \text{ was sent}\right\} = 0. \tag{2.6}$$

That is, if there are no errors (on average). This is what we refer to as *reliable communication*. Finally, the capacity of the channel is the supremum of all rates that are achievable.

Proposition 2.1 *[1, Theorem 17] The capacity of the Gaussian channel with time-invariant fading, given a source power constraint P and noise power N_o is*

$$\begin{aligned} \mathcal{C} &= \max_{p(x): \int_{-\infty}^{\infty} |x|^2 p(x) dx = P} I(x; y) \\ &= \mathcal{C}\left(|h|^2 \frac{P}{N_o}\right). \end{aligned} \tag{2.7}$$

As thoroughly explained in [21, Section 10.1], the capacity is achievable using a random codebook at the transmitter, with codewords of length $n \to \infty$, generated *i.i.d.* from $x \sim \mathcal{CN}(0, P)$. At the receiver, joint typical decoding is capacity-achieving.

2.1.2 Gaussian Multi-access Channel

Consider now two independent sources communicating simultaneously to a single destination. Assume that, for every signal x_1 and x_2 transmitted by user 1 and user 2, respectively, the receiver receives:

$$y = h_1 \cdot x_1 + h_2 \cdot x_2 + z, \tag{2.8}$$

where $z \sim \mathcal{CN}(0, N_o)$, and h_1, h_2 are non-random complex scalars. This is referred to as Gaussian multi-access channel (MAC) with *time-invariant* fading.

The sources select messages $\omega_1 \in \mathcal{W}_1 = \left\{1, \cdots, 2^{nR_1}\right\}$ and $\omega_2 \in \mathcal{W}_2 = \left\{1, \cdots, 2^{nR_2}\right\}$ for transmission, respectively. Then, they pass them to their transmitters, who put the messages onto the channel as explained in the previous subsection. However, in this case, the transmitters use a MAC channel code $\left(n, 2^{nR_1}, 2^{nR_2}\right)$, which is defined by: *i*) the two independent sets of messages $\mathcal{W}_1$ and $\mathcal{W}_2$; *ii*) two signal spaces $\mathcal{X}_1$ and $\mathcal{X}_2$, *iii*) two transmit mappings f_1 and f_2, one per each user:

$$f_u : \mathcal{W}_u \to \mathcal{X}_u^n, \ \ u = 1, 2, \tag{2.9}$$

where each mapping satisfies the per-user power constraint:

$$\frac{1}{n} \sum_{t=1}^{n} |x_u^t(\omega_u)|^2 \leq P_u, \ \ \forall \omega_u, \ \ u = 1, 2. \tag{2.10}$$

And, *iv*) a receiver demapping g

$$g : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2. \tag{2.11}$$

In the channel, an error occurs whenever $g\left(\boldsymbol{y}^n\right) \neq (\omega_1, \omega_2)$. A transmission rate duple (R_1, R_2) [bits/symbol] is said to be achievable if there exists a sequence of codes $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ for which:

$$\lim_{n\to\infty} \frac{1}{2^{nR_1}2^{nR_2}} \sum_{\omega_2,\omega_2} \Pr\left\{g\left(\boldsymbol{y}_d^n\right) \neq (\omega_1, \omega_2)|(\omega_1, \omega_2) \text{ was sent}\right\} = 0. \tag{2.12}$$

That is, if the communication is reliable for both users. The capacity region is the closure of all rate duples that are achievable.

Proposition 2.2 *[21, Section 14.3.6] The capacity region of the Gaussian MAC with time-invariant fading, given a source power constraints P_1 and P_2, and noise power N_o is*

$$\begin{aligned}\mathcal{C} &= \text{coh}\left(\bigcup_{p_{x_1},p_{x_2}:\int_{-\infty}^{\infty}|x|^2 p_{x_u}(x)dx\leq P_u,\ u=1,2.} \left\{R_{1,2}: \begin{array}{c} R_1 \leq I\left(x_1; y|x_2\right) \\ R_2 \leq I\left(x_2; y|x_1\right) \\ R_1 + R_2 \leq I\left(x_1, x_2; y\right)\end{array}\right\}\right) \\ &= \left\{R_{1,2}: \begin{array}{c} R_1 \leq \mathcal{C}\left(|h_1|^2 \frac{P_1}{N_o}\right) \\ R_2 \leq \mathcal{C}\left(|h_2|^2 \frac{P_2}{N_o}\right) \\ R_1 + R_2 \leq \mathcal{C}\left(|h_1|^2 \frac{P_1}{N_o} + |h_2|^2 \frac{P_2}{N_o}\right)\end{array}\right\}\end{aligned} \tag{2.13}$$

This proposition points out one of the most interesting features of the MAC: while for the Gaussian channel we search for the distribution $p(x)$ that maximizes the mutual information (2.7), for the MAC we consider the union over all possible distributions p_{x_1}, p_{x_2} that satisfy the power constraint. This is common for all multi-access networks as *e.g.*, with multiple antennas [45], with memory [46], etc. Nevertheless, here, all distributions are outperformed by the Gaussian distribution. Indeed, as noted in [21, Chapter 14], any point of the region is achievable using random codebooks at the transmitters, generated *i.i.d.* from $x_u \sim \mathcal{CN}(0, P_u)$, $u = 1, 2$, respectively. At the receiver, joint typical decoding is again optimal.

2.1.3 Gaussian Broadcast Channel

Let a single source communicate simultaneously with two destinations, transmitting different messages to each other. Assume that, for every signal x transmitted by the transmitter, the two receivers receive:

$$\begin{aligned} y_1 &= h_1 \cdot x + z_1, \\ y_2 &= h_2 \cdot x + z_2, \end{aligned} \tag{2.14}$$

where $z_u \sim \mathcal{CN}(0, N_o)$ and h_u is non-random complex scalar, for $u = 1, 2$. This is referred to as Gaussian broadcast channel (BC) with *time-invariant* fading. Consider hereafter, *without loss of generality*, that $|h_1| > |h_2|$.

The source selects two messages $\omega_1 \in \mathcal{W}_1 = \left\{1, \cdots, 2^{nR_1}\right\}$ and $\omega_2 \in \mathcal{W}_2 = \left\{1, \cdots, 2^{nR_2}\right\}$, for transmission to destination 1 and destination 2, respectively. The transmitter sends them over the channel using a BC channel code $\left(n, 2^{nR_1}, 2^{nR_2}\right)$, which is defined by: *i*) the two independent sets of messages $\mathcal{W}_1$ and $\mathcal{W}_2$; *ii*) a unique signal space $\mathcal{X}$; *iii*) a unique transmit mapping:

$$f : \mathcal{W}_1 \times \mathcal{W}_2 \rightarrow \mathcal{X}^n, \tag{2.15}$$

that satisfies the source power constraint:

$$\frac{1}{n} \sum_{t=1}^{n} \left|x^t(\omega_1, \omega_2)\right|^2 \leq P, \ \ \forall \omega_1, \omega_2. \tag{2.16}$$

And, *iv*) two receiver demappings g_1 and g_2, one per each destination

$$g_u : \mathcal{Y}_u^n \rightarrow \mathcal{W}_u, \ u = 1, 2. \tag{2.17}$$

In the channel, an error occurs whenever $g_1(\boldsymbol{y}_1^n) \neq \omega_1$ or $g_2(\boldsymbol{y}_2^n) \neq \omega_2$. A transmission rate duple (R_1, R_2) [bits/symbol] is said to be achievable if there exists a sequence of codes $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ for which no errors occur:

$$\lim_{n\to\infty} \frac{1}{2^{nR_1}2^{nR_2}} \sum_{\omega_2, \omega_2} \Pr\left\{ \bigcup_{u=1,2} g_u(\boldsymbol{y}_u^n) \neq \omega_u | (\omega_1, \omega_2) \text{ was sent} \right\} = 0. \tag{2.18}$$

The capacity region is then the closure of all achievable rate duples.

Proposition 2.3 *[21, Example 14.6.6] The capacity region of the Gaussian BC with time-invariant fading, source power constraint P and noise power N_o at both receivers is*

$$\mathcal{C} = \bigcup_{0 \leq \alpha \leq 1} \left\{ R_{1,2} : \begin{array}{c} R_1 \leq \mathcal{C}\left(|h_1|^2 \frac{\alpha \cdot P}{N_o}\right) \\ R_2 \leq \mathcal{C}\left(|h_2|^2 \frac{(1-\alpha)\cdot P}{N_o + |h_2|^2 \alpha \cdot P}\right) \end{array} \right\} \tag{2.19}$$

There are two different coding schemes that are able to achieve any point of the capacity region. The first one assumes all the processing efforts at the receivers, while the second one centralizes the processing at the transmitter:

1. *Superposition coding at the transmitter plus successive interference cancellation (SIC) at the receivers* [21, Section 14.1.3]: the transmitter generates two random codebooks, with codewords of length $n \to \infty$, generated *i.i.d.* from $x_1 \sim \mathcal{CN}(0, \alpha P)$ and $x_2 \sim \mathcal{CN}(0, (1-\alpha)P)$, respectively. It maps message ω_1 using the first codebook and ω_2 using the second codebook. Finally, the source transmits the superposition of both mappings. Receiver 2 decodes message ω_2 using joint typicality, considering the codeword intended to user 1 as interference. In contrast, receiver 1 first decodes ω_2; then, it removes the ω_2 contribution onto the received signal, and decodes ω_1 without interference. This is referred to as SIC.

2. *Dirty Paper coding at the transmitter and independent joint typical decoding at the receivers* [47]: The source encodes ω_2 as if destination 2 were alone in the network, *i.e.*, using random coding. Next, the source encodes the message ω_1 utilizing the (non-causal) knowledge of the interference that the signal intended to user 2 is going to make on user 1. It was shown by Costa in [47] that such a coding allows the first user to decode ω_1, using joint typical decoding, as if it had not interference from the second user. In contrast, the second receiver decodes ω_2 suffering the interference of the signal intended to the first user.

2.2 Constrained Optimization

In a wide range of communication problems, designers aim at maximizing a network cost function given a constraint on the network resources. This is referred to as constrained optimization [48, Chapter 3], and is of capital importance on the research, design and development of wireless communications. In this section, we review its main results, which are used along the dissertation.

Consider an optimization problem of the form

$$\begin{aligned}\max_{x_1,\cdots,x_N} \quad & f\left(\boldsymbol{x}_{1:N}\right) \\ \text{s.t.} \quad & g_i\left(\boldsymbol{x}_{1:N}\right) \leq R_i, \quad i = 1, \cdots, m,\end{aligned} \tag{2.20}$$

where $f(\cdot)$ and $g_i(\cdot)$ are continuously differentiable functions, not necessarily convex or concave. The union of all points $\boldsymbol{x}_{1:N}$ where the constraints are satisfied, *i.e.* $g_i(\boldsymbol{x}_{1:N}) \leq R_i$, $i = 1, \cdots, m$, is denoted by *feasible set* $\mathcal{X}$. Every point of the feasible set is referred to as *feasible point*.

The problem consists in finding the maximum value of the function $f(\cdot)$ within the feasible set $\mathcal{X}$. Usually, many local maxima are found in the function. The objective of the search is to find the greater of these, which is called the global maximum. It can occur that the function is not uniquely attained: that is, the global maximum is achieved at more than one feasible point. As an example, the function x^2 takes maximum value within the feasible set $\mathcal{X} = [-1, 1]$ at the points $x = 1$ and $x = -1$. This will have no implications on the theory.

In order to solve such a constrained maximization, a parameterized function, which is always an upper bound on $f(\cdot)$ within the feasible set, is defined: the Lagrangian function. It is constructed using Lagrange multipliers $\lambda_1, \cdots, \lambda_m \geq 0$ and reads as follows:

$$\mathcal{L}\left(\boldsymbol{x}_{1:N}, \boldsymbol{\lambda}_{1:m}\right) = f\left(\boldsymbol{x}_{1:N}\right) - \sum_{i=1}^{m} \lambda_i \cdot \left(g_i\left(\boldsymbol{x}_{1:N}\right) - R_i\right). \tag{2.21}$$

As mentioned, it satisfies

$$\mathcal{L}\left(\boldsymbol{x}_{1:N}, \boldsymbol{\lambda}_{1:m}\right) \geq f\left(\boldsymbol{x}_{1:N}\right), \quad \forall \boldsymbol{x}_{1:N} \in \mathcal{X}. \tag{2.22}$$

since for all the feasible points: $g_i(\boldsymbol{x}_{1:N}) \leq R_i$, $i = 1, \cdots, m$. This is indeed the *reason-of-existence* for the Lagrangian: it is born to upper bound the function at hand, allowing thus to upper bound its maximum value and, in cases, to obtain it. Consider the non-convex maximization: $f(x) = x^3 - e^x$ constrained to $x^2 \leq 25$. The function and its associated Lagrangian (parameterized for $\lambda = 1$ and $\lambda = 1.2$) is plotted in Fig. 2.2 with respect to the feasible set. The relationship (2.22) is clearly shown: the Lagrangian upper bounds the function.

Making use of the Lagrangian, many results regarding the maximum value of the function can be derived [48]. Here, we present the two most important. First, the necessary condition that the maximum of (2.20) must satisfy: the Karush-Kuhn-Tucker (KKT) conditions. Later,

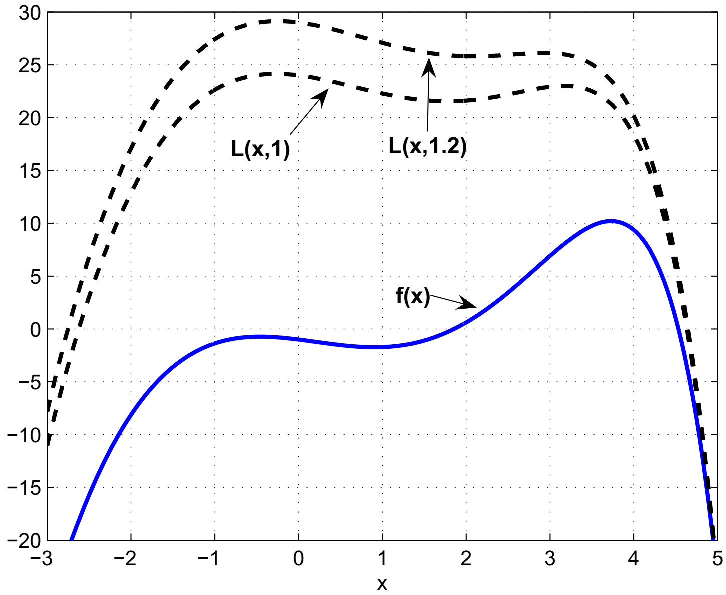

Figure 2.2: Lagrangian function.

the general sufficiency condition for a feasible point to attain the maximum. We will show that, for convex problems, both are equivalent and thus KKTs are necessary and sufficient for optimality.

Proposition 2.4 *(Karush-Kuhn-Tucker Necessary Conditions [48, Proposition 3.3.1]) Let $\boldsymbol{x}^*_{1:N}$ be a local maximum of (2.20). Then, there exist a unique Lagrange multiplier vector $\boldsymbol{\lambda}^*_{1:m}$ for which the KKT conditions:*

$$\begin{aligned} &\frac{\partial \mathcal{L}\left(\boldsymbol{x}^*_{1:N}, \boldsymbol{\lambda}^*_{1:m}\right)}{\partial x_i} = 0, \;\; i = 1, \cdots, N \\ &\lambda^*_j \geq 0, \;\; j = 1, \cdots, m \\ &\lambda^*_j \cdot \left(g_j\left(\boldsymbol{x}^*_{1:N}\right) - R_j\right) = 0 \;\; j = 1, \cdots, m. \end{aligned} \tag{2.23}$$

hold.

Being the condition necessary, it is not sufficient. It can occur, thus, that a feasible point satisfies the KKTs and it is not a local maximum of the function. Therefore, it is not the global either. In order to be it, it must also satisfy the sufficient condition for optimality.

Proposition 2.5 *(General Sufficiency Condition [48, Proposition 3.3.4]) Consider the problem (2.20). Let $\boldsymbol{x}^*_{1:N}$ be a feasible point which, together with a Lagrange multiplier vector $\boldsymbol{\lambda}^*_{1:m} \geq 0$,* *satisfies*

$$\lambda_j^* \cdot \left(g_j\left(\boldsymbol{x}_{1:N}^*\right) - R_j\right) = 0, \;\; j = 1, \cdots, m. \tag{2.24}$$

and maximizes the Lagrangian $\mathcal{L}\left(\boldsymbol{x}_{1:N}, \boldsymbol{\lambda}_{1:m}^*\right)$ *over the feasible set:*

$$\boldsymbol{x}_{1:N}^* \in \arg\max_{\boldsymbol{x}_{1:N} \in \mathcal{X}} \mathcal{L}\left(\boldsymbol{x}_{1:N}, \boldsymbol{\lambda}_{1:m}^*\right). \tag{2.25}$$

Then, $\boldsymbol{x}_{1:N}^*$ *is the global maximum of the problem.*

Mixing both Propositions, a generic strategy to solve constrained optimizations is as follows: first, we search for all feasible points that satisfy the KKT conditions. That is, find all points for which (2.23) has a unique solution on $\boldsymbol{\lambda}_{1:m}$. Among all, we next find out one for which the general sufficiency condition holds. This will be the global maximum of the problem. However, many and wide drawbacks can be drawn on this procedure: *i*) there can be ∞ points satisfying the KKTs, making the computation tedious. *ii*) Usually, these points are obtained resorting to iterative algorithms, making the computational time excessively long. *iii*) The general sufficiency condition is not easy to demonstrate when the Lagrangian is not a concave function on $\boldsymbol{x}_{1:N}$.

Fortunately, when the function $f(\cdot)$ is concave, and the constraints $g_i(\cdot)$, $i = 1, \cdots, m$ are convex, both necessary and sufficient conditions become identical. In fact, for this case the Lagrangian in (2.21) is concave with respect to $\boldsymbol{x}_{1:N}$. Thereby, it takes its maximum (2.25) at the point where:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{x}_{1:N}^*, \boldsymbol{\lambda}_{1:m}^*\right)}{\partial x_i} = 0, \;\; i = 1, \cdots, N, \tag{2.26}$$

which is identical to the KKT condition in Proposition 2.4. Hence, KKTs become sufficient also. This simplifies notably the search, which is reduced to solving (on $\boldsymbol{x}_{1:N}$ and $\boldsymbol{\lambda}_{1:N}$) the system of equations defined by the KKT conditions (2.23). Every solution is a global maximum due to the sufficiency condition. Notice that the KKTs are defined by means of $N+m$ equations with $N+m$ variables. Thus, the problem is solvable. However, the solution is not necessarily unique.

2.2.1 Iterative Algorithms

Many times in optimization, the KKT conditions (2.23) cannot be solved in closed-form. That is, the system of $N+m$ equations with $N+m$ variables is solvable but an analytical expression for the solution cannot be given. It has to be obtained, then, resorting to iterative algorithms.

Here, we present two well-known algorithms to solve constrained optimizations iteratively. Additionally, we provide an extension for the case where the cost function is not differentiable.

Block-coordinate Ascent Algorithm

Let us consider a special case of problem (2.20):

$$\begin{aligned} \max_{x_1,\cdots,x_N} \quad & f\left(\boldsymbol{x}_{1:N}\right) \\ \text{s.t.} \quad & g_i\left(x_i\right) \leq R_i,\ i=1,\cdots,N \end{aligned} \tag{2.27}$$

where $g_i(\cdot)$ are convex functions on x_i. No assumption is taken on the concavity of $f(\cdot)$. Besides the convexity of $g_i(\cdot)$, the main difference with the original problem is that, now, the optimization variables are decoupled: each constraint only involves one variable. In other words, being the convex set $\mathcal{X}_i = \{x_i \mid g_i(x_i) \leq R_i\}$, then the feasible set $\mathcal{X}$ is the cartesian product of convex sets:

$$\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N. \tag{2.28}$$

The algorithm consists of iteratively optimizing the function with respect to one x_i while keeping the other variables fixed:

$$x_i^{t+1} \in \arg\max_{x_i \in \mathcal{X}_i} f\left(x_1^{t+1},\cdots,x_{i-1}^{t+1},x_i,x_{i+1}^{t},\cdots,x_N^{t}\right),\ \ i=1,\cdots,N, \tag{2.29}$$

with t the iteration index. Intuitively, it performs a cyclic best-effort among variables. The algorithm is also known as Non-Linear *Gauss-Seidel* algorithm, and makes sense whenever the optimization (2.29) is fairly more easy to solve than (2.27) [49, Section 3.3.5]. Clearly, each iteration increases the cost function. Two convergence results and a corollary can be drawn for it.

Proposition 2.6 *[48, Proposition 2.7.1] Consider (2.27) with $f(\cdot)$ continuously differentiable. Furthermore, let*

$$\max_{\xi \in \mathcal{X}_i} f\left(x_1, \cdots, x_{i-1}, \xi, x_{i+1}, \cdots, x_N\right), \quad i = 1, \cdots, N \tag{2.30}$$

exist and be uniquely attained. Then, the limit point of sequence $\boldsymbol{x}_{1:N}^t$ is a stationary point, i.e., $\boldsymbol{x}_{1:N}^t \rightarrow \boldsymbol{x}_{1:N}^*$ with $\nabla f\left(\boldsymbol{x}_{1:N}^*\right)^\dagger \cdot \left(\boldsymbol{x}_{1:N} - \boldsymbol{x}_{1:N}^*\right) \leq 0$, $\forall \boldsymbol{x}_{1:N} \in \mathcal{X}$.

Remark 2.1 *Notice that all local maxima are stationary points, but not all stationary points are local maxima [48, Proposition 2.1.2]. Therefore, the sequence is guaranteed to converge, but not necessarily to a local maximum.*

Corollary 2.1 *[49, Proposition 3.9] Whenever $f(\cdot)$ is also concave, the limit point of sequence $\boldsymbol{x}_{1:N}^t$ converges to the global maximum of the problem.*

Proof: As mentioned in [48, Proposition 2.1.2], for concave cost function $f(\cdot)$, stationarity is not only necessary but also sufficient for global-optimality. Thus, the limit point satisfies Proposition 2.5. ■

Proposition 2.7 *[49, Proposition 3.10] Consider the assumptions in Prop. 2.6, with $f(\cdot)$ not necessarily concave. Furthermore, let the mapping $R\left(\boldsymbol{x}_{1:N}\right) = \boldsymbol{x}_{1:N} + \gamma \cdot \nabla f\left(\boldsymbol{x}_{1:N}\right)$ be a block-contraction*[1] *for some γ. Then, the limit point of sequence $\boldsymbol{x}_{1:N}^t$ converges to a global maximum of the problem.*

Gradient Projection Algorithm

For the previous algorithm, two strong assumptions were taken: *i*) the feasible set needed to be defined as the cartesian product of convex sets; *ii*) the maximization with respect to one variable, keeping the rest fixed, needed to be uniquely attained. Those two assumptions restrict the applicability of the algorithm. We thus propose new alternatives to iteratively solve optimization problems not satisfying *i*) and/or *ii*).

[1] See [49, Section 3.1.2] for the definition of block-contraction, and sufficient conditions to hold.

Consider now the more general problem (2.20), and let $\mathcal{X}$ be the feasible set, closed but not necessarily convex. The *Gradient Projection* method searches for the maximum iterating as follows [48, Section 2.3]:

$$\boldsymbol{x}_{1:N}^{t+1} = \boldsymbol{x}_{1:N}^{t} + \gamma_t \left(\bar{\boldsymbol{x}}_{1:N} - \boldsymbol{x}_{1:N}^{t} \right), \tag{2.31}$$

where t is the iteration index, $\gamma_t \in (0, 1]$ the step-size, and:

$$\bar{\boldsymbol{x}}_{1:N} = \left[\boldsymbol{x}_{1:N}^{t} + s_t \cdot \nabla f \left(\boldsymbol{x}_{1:N}^{t} \right) \right]^{\perp}. \tag{2.32}$$

Here, $s_t \geq 0$ is a scalar, $\nabla f(\cdot)$ the gradient of $f(\cdot)$ and $[\cdot]^{\perp}$ the projection onto the feasible set $\mathcal{X}$. Two important issues must be noted here: *i*) the projection of a vector $\boldsymbol{x}_{1:N}$ onto the feasible set $\mathcal{X}$ consists of finding a vector $\boldsymbol{a}_{1:N}^{*} \in \mathcal{X}$ such that:

$$\boldsymbol{a}_{1:N}^{*} \in \arg \min_{\boldsymbol{a}_{1:N} \in \mathcal{X}} ||\boldsymbol{a}_{1:N} - \boldsymbol{x}_{1:N}|| \Rightarrow \left[\boldsymbol{x}_{1:N} \right]^{\perp} = \boldsymbol{a}_{1:N}^{*}, \tag{2.33}$$

given a preselected norm. *ii*) The algorithm is not, in general, amenable for parallel implementation. Although the computation of $\boldsymbol{x}_{1:N}^{t} + s_t \cdot \nabla f(\boldsymbol{x}_{1:N}^{t})$ can be easily parallelized, the computation of the projection requires the solution of a non-trivial optimization involving all components of $\boldsymbol{x}_{1:N}$.

The selection of the step size γ_t and the scalar s_t is generally crucial to obtain good convergence properties for the algorithm. Some of the commonly used rules to choose them are brilliantly revisited by Bertsekas in [48, Section 2.3.1], *e.g.,* limited maximization rule, Armijo's rule along the feasible direction, Armijo's rule along the feasible arc and diminishing step-size. We refer the reader to that textbook for in-depth discussion.

Proposition 2.8 *[48, Proposition 2.3.1] Let $\boldsymbol{x}_{1:N}^{t}$ be a sequence generated by the gradient projection method, with γ_t and s_t chosen by the limited maximization rule or the Armijo rule along the feasible direction. Then, every limit point of $\boldsymbol{x}_{1:N}^{t}$ is a stationary point.*

Corollary 2.2 *Whenever $f(\cdot)$ is concave, and the feasible set $\mathcal{X}$ convex, the limit point of sequence $\boldsymbol{x}_{1:N}^{t}$ converges to the global maximum of the problem.*

One important property of the algorithm is that, when the feasible set is defined as the cartesian product of sets, *i.e.* $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$, the algorithm can be parallelized without loosing convergence properties [49, Section 3.3.4]. That is, we can define it as:

$$x_n^{t+1} = x_n^t + \gamma_t \left(\bar{x}_n - x_n^t \right), \quad n = 1, \cdots, N \tag{2.34}$$

and:

$$\bar{x}_n = \left[x_n^t + s_t \cdot \nabla_{x_n} f\left(\boldsymbol{x}_{1:N}^t \right) \right]^{\perp}, \quad n = 1, \cdots, N. \tag{2.35}$$

where $\nabla_{x_n} f(\cdot)$ is the gradient with respect to x_n, and the projection is independently carried out onto each set $\mathcal{X}_n$.

Subgradient Methods

In the previous analysis and algorithms, we assumed that the objective function is continuously differentiable. In fact, we considered the gradient known and computable in closed-form. However, this might not be the case for all optimizations in communications; thus, here we propose an alternative: the *subgradient search*.

It iterates exactly as the gradient method. However, instead of computing the gradient, a subgradient is utilized. A subgradient is defined as a function $\boldsymbol{h}(\boldsymbol{x}_{1:N})$ that satisfies:

$$f(\boldsymbol{a}_{1:N}) - f(\boldsymbol{x}_{1:N}) \geq \boldsymbol{h}(\boldsymbol{x}_{1:N})^{\dagger} (\boldsymbol{a}_{1:N} - \boldsymbol{x}_{1:N}), \quad \forall \boldsymbol{a}_{1:N} \neq \boldsymbol{x}_{1:N}. \tag{2.36}$$

With it, the algorithm iterates:

$$\boldsymbol{x}_{1:N}^{t+1} = \boldsymbol{x}_{1:N}^t + \gamma_t \left(\bar{\boldsymbol{x}}_{1:N} - \boldsymbol{x}_{1:N}^t \right), \tag{2.37}$$

where t is the iteration index, $\gamma_t \in (0, 1]$ the step-size, and:

$$\bar{\boldsymbol{x}}_{1:N} = \left[\boldsymbol{x}_{1:N}^t + s_t \cdot \boldsymbol{h}\left(\boldsymbol{x}_{1:N}^t \right) \right]^{\perp}. \tag{2.38}$$

In contrast with the gradient method, the new iterate may not improve the cost function. However, as discussed in [49, Section 6.3.1], diminishing step-size rules guarantees convergence to a stationary point of the cost function.

2.2.2 The Dual Problem

The algorithms introduced in the previous subsection are of wide use in network optimization for *e.g.* covariance optimization for MIMO multi-user networks [50] and distortion design for multi-source compression [51]. However, there are many other problems where the applicability is limited. On one hand, the block-coordinate approach requires the feasible set to be the cartesian product of convex sets, which is not common. On the other hand, the gradient projection is built up upon the projection of vectors onto feasible sets. This can be computationally intensive[2], and in some cases, unfeasible. Hence, new approaches are needed. Here, we present a powerful one: *dual decomposition*.

Consider the optimization problem

$$\begin{aligned} p^* \quad = \quad & \max_{x_1,\cdots,x_N} \quad f\left(\boldsymbol{x}_{1:N}\right) \\ & \text{s.t.} \quad\quad g\left(\boldsymbol{x}_{1:N}\right) \leq R \end{aligned} \tag{2.39}$$

For it, the dual problem is defined as follows. First, the Lagrangian function is constructed on $\lambda \geq 0$ as in (2.21):

$$\mathcal{L}\left(\boldsymbol{x}_{1:N}, \lambda\right) = f\left(\boldsymbol{x}_{1:N}\right) - \lambda \cdot \left(g\left(\boldsymbol{x}_{1:N}\right) - R\right). \tag{2.40}$$

Next, the dual function is obtained as the maximum of each parameterized function

$$q\left(\lambda\right) = \max_{x_1,\cdots,x_N} \mathcal{L}\left(\boldsymbol{x}_{1:N}, \lambda\right). \tag{2.41}$$

Function $q\left(\lambda\right)$ has two main, and fundamental, properties: *i*) it is convex on λ, since it is the point-wise maximum of a family of affine functions [52, Section 5.1.2]. This holds even for non-convex problems. *ii*) $q\left(\lambda\right) \geq p^*, \forall \lambda \geq 0$. To demonstrate such an inequality, we can easily infer from (2.22) that:

$$\mathcal{L}\left(\boldsymbol{x}_{1:N}, \lambda\right) \geq f\left(\boldsymbol{x}_{1:N}\right), \;\; \forall \lambda \geq 0, \forall \boldsymbol{x}_{1:N} \in \mathcal{X} \Rightarrow \; \max \mathcal{L}\left(\boldsymbol{x}_{1:N}, \lambda\right) \geq \max f\left(\boldsymbol{x}_{1:N}\right) \tag{2.42}$$

Finally, the dual problem is defined as

$$q^* = \min_{\lambda \geq 0} q\left(\lambda\right). \tag{2.43}$$

The dual analysis has a nice interpretation, that can be grasped from Fig. 2.2. First, the maximum of each parameterized function is found, aiming at upper bounding the global maximum of f. Later, the minimum of all upper bounds is selected as the tightest upper bound on the sought maximum.

[2] Recall that the projection is defined by means of a minimization.

In order to estimate how good the bound is, the duality gap is used; it is defined as $q^* - p^* \geq 0$. An optimization is said to have zero-duality gap if and only if $q^* - p^* = 0$. For those cases, solving the dual problem solves the primal too.

Proposition 2.9 *(Strong Duality Theorem for Convex Problems [48, Proposition 5.3.1]) Consider the problem (2.39), with: $f(\cdot)$ concave, $g(\cdot)$ convex, non-empty feasible set $\mathcal{X}$, and finite optimal value f^*. Furthermore, let be there at least one $\boldsymbol{x}_{1:N} \in \mathcal{X}$ such that $g(\boldsymbol{x}_{1:N}) < R$. For such a problem, the duality gap is zero.*

With this result, it is demonstrated that convex problems has zero-duality gap. But, what about non-convex problems?. In general, no claim can be made. However, there is a specific family of non-convex problems for which $p^* = q^*$; the one who satisfies the *time-sharing* property.

Definition 2.1 *(Time-Sharing Property [53, Definition 1]) Let $\boldsymbol{x}^*_{1:N}$ and $\boldsymbol{y}^*_{1:N}$ be the optimal solutions for problem (2.39) with $R = R_x$ and $R = R_y$, respectively. An optimization problem of the form of (2.39) is said to satisfy the time-sharing property if for any R_x, R_y and any $0 \leq \nu \leq 1$, there is always a feasible point $\boldsymbol{z}_{1:N}$ such that $g(\boldsymbol{z}_{1:N}) \leq \nu R_x + (1-\nu) R_y$ and $f(\boldsymbol{z}_{1:N}) \geq \nu f(\boldsymbol{x}_{1:N}) + (1-\nu) f(\boldsymbol{y}_{1:N})$. That is, if the maximum of (2.39) is concave with respect to R.*

Proposition 2.10 *(Strong Duality Theorem for Non-Convex Problems [53, Theorem 1]) Consider the problem (2.39), with $f(\cdot)$ not necessarily concave and $g(\cdot)$ not necessarily convex. If the optimization satisfies the time-sharing property, the duality gap is zero.*

Therefore, convex problems (and a family of non-convex ones) can be solved resorting to dual decomposition. However, is this really worthwhile? In most of the cases, it is. The main reason is that the dual function decouples the constraint. In other words, the involved constrained maximization (2.39) turns into a simple non-constrained one in (2.41). Clearly, the latter can be straightforwardly solved resorting to iterative algorithms as *e.g.*, unconstrained *block-coordinate* algorithm or unconstrained gradient method [48, Chapter 1].

However, it is still necessary to find the minimum of the dual function $q(\lambda)$. As mentioned, the function is convex. Therefore, a gradient method can be utilized to efficiently search for the minimum with guaranteed convergence. Unfortunately, the dual function is not continuously differentiable in all cases, and for the cases where it is, the gradient may not be known. We can thus utilize the subgradient search, which consists of following search direction $-h(\lambda)$ such that[3]:

$$\frac{q(\lambda') - q(\lambda)}{\lambda' - \lambda} \geq h(\lambda), \quad \forall \lambda' \neq \lambda. \tag{2.44}$$

Let us now define

$$\boldsymbol{x}_{1:N}(\lambda) \in \arg \max_{x_1, \cdots, x_N} \mathcal{L}(\boldsymbol{x}_{1:N}, \lambda) \tag{2.45}$$

$$\boldsymbol{x}_{1:N}(\lambda') \in \arg \max_{x_1, \cdots, x_N} \mathcal{L}(\boldsymbol{x}_{1:N}, \lambda'). \tag{2.46}$$

Hence, it is clear that

$$\begin{aligned} q(\lambda) &= f(\boldsymbol{x}_{1:N}(\lambda)) - \lambda \cdot (g(\boldsymbol{x}_{1:N}(\lambda)) - R), \quad (2.47) \\ q(\lambda') &= f(\boldsymbol{x}_{1:N}(\lambda')) - \lambda' \cdot (g(\boldsymbol{x}_{1:N}(\lambda')) - R) \\ &\geq f(\boldsymbol{x}_{1:N}(\lambda)) - \lambda' \cdot (g(\boldsymbol{x}_{1:N}(\lambda)) - R) \end{aligned}$$

where inequality follows from the fact that there is no greater value than the global maximum. Therefore, it is possible to bound

$$q(\lambda') - q(\lambda) \geq -(\lambda' - \lambda)(g(\boldsymbol{x}_{1:N}(\lambda)) - R). \tag{2.48}$$

As a result, the following choice satisfies the subgradient condition (2.44):

$$h(\lambda) = R - g(\boldsymbol{x}_{1:N}(\lambda)). \tag{2.49}$$

2.3 Stochastic Convergence

A sequence of random variables $\{x_n\}_{n=1}^{\infty}$ is said to converge to a random variable x if, at the limit, both the sequence and x behave equally. Such a convergence is denoted by:

$$x_n \to x. \tag{2.50}$$

[3]When searching for the minimum, we must follow the opposite direction of the gradient.

However, what does ”behave equally” mean for random variables? It can mean different things, and each meaning will define a type of convergence. Before reviewing them, let us state that a non-random sequence $\{s_n\}$ converges to s if and only if:

$$\forall \epsilon > 0, \ \ \exists N \ \text{ such that } \ |s_n - s| < \epsilon, \ \ \forall n \geq N. \tag{2.51}$$

With this in mind, four kinds of convergence can be defined for random sequences [54, pp. 40-42]:

- *Almost sure convergence*, denoted by $x_n \overset{a.s.}{\to} x$, and satisfied whenever Pr $\{\mathcal{O}\} = 1$, where $\mathcal{O}$ is the event

  $$\mathcal{O} = \{\forall \epsilon > 0, \ \ \exists N \ \text{ such that } \ |x_n - x| < \epsilon, \ \ \forall n \geq N\}. \tag{2.52}$$

  This is the strongest type of convergence, forcing to be 1 the probability of having a sampled sequence for which (2.51) holds. Nonetheless, notice that the complementary event $\mathcal{O}^c$ can occur (that is, having a sampled sequence which does not converge), but with probability zero.

- *Convergence in mean-square*, denoted by $x_n \overset{m.s.}{\to} x$, and satisfied whenever

  $$\lim_{n\to\infty} \mathrm{E}\left\{|x_n - x|^2\right\} = 0. \tag{2.53}$$

  The mean is taken over the probability density function of both x_n and x.

- *Convergence in probability*, denoted by $x_n \overset{P}{\to} x$, and satisfied whenever

  $$\lim_{n\to\infty} \Pr\{|x_n - x| < \epsilon\} = 1, \ \ \forall \epsilon > 0. \tag{2.54}$$

  Clearly, such a convergence is guaranteed whenever almost sure converge holds. However, the converse is not true.

- *Convergence in law*, denoted by $x_n \overset{L}{\to} x$. It holds whenever the *c.d.f.* of x, and that of x_n satisfy, for all y belonging to the support set:

  $$\lim_{n\to\infty} F_{x_n}(y) = F_x(y). \tag{2.55}$$

Among all, the convergence in law is the weakest one, holding whenever all others are satisfied. On the other hand, convergence in mean square and almost sure convergence are the strongest ones, making all others to hold too. Nevertheless, it can occur that one is satisfied while the other is not. Finally, as mentioned, convergence in probability is satisfied if almost sure and/or in mean square hold, and implies convergence in law.

Convergence in probability will be the one used to present the convergence results of this thesis. Let us introduce four lemmas regarding it.

Lemma 2.1 *Let $x_n > 0\ \forall n$ be a sequence of positive random variables, and $0 < g(n) < \infty$, $\forall n$ be a non-random sequence. Then:*

$$x_n - g(n) \xrightarrow{P} 0 \Longrightarrow \log_2(x_n) - \log_2(g(n)) \xrightarrow{P} 0. \tag{2.56}$$

Proof: Consider $x_n - g(n) \xrightarrow{P} 0$; from definition (2.54) that means:

$$\lim_{n\to\infty} \Pr\{|x_n - g(n)| < \epsilon\} = 1, \ \ \forall \epsilon > 0. \tag{2.57}$$

Such a probability can be rewritten as:

$$\Pr\{|x_n - g(n)| < \epsilon\} = \Pr\{x_n < g(n) + \epsilon\} - \Pr\{x_n \leq g(n) - \epsilon\} \tag{2.58}$$

Clearly, being the logarithm a continuous, non-decreasing function, the first term can be expressed as:

$$\Pr\{x_n \leq g(n) + \epsilon\} = \Pr\{\log_2(x_n) \leq \log_2(g(n) + \epsilon)\} \tag{2.59}$$

On the other hand, recalling that $x_n > 0$, we can compute the second term as

$$\Pr\{x_n < g(n) - \epsilon\} = \begin{cases} 0 & \epsilon \geq g(n) \\ \Pr\{\log_2(x_n) < \log_2(g(n) - \epsilon)\} & \epsilon < g(n) \end{cases} \tag{2.60}$$

Defining $\log_2(0) \triangleq -\infty$, we can compactly write the probability above as:

$$\Pr\{x_n \leq g(n) - \epsilon\} = \Pr\left\{\log_2(x_n) \leq \log_2\left([g(n) - \epsilon]^+\right)\right\}. \tag{2.61}$$

Plugging both (2.59) and (2.61) into (2.58), makes:

$$\begin{aligned} \Pr\{|x_n - g(n)| < \epsilon\} &= \Pr\left\{\log_2\left([g(n) - \epsilon]^+\right) < \log_2(x_n) < \log_2(g(n) + \epsilon)\right\} \\ &\leq \Pr\{|\log_2(x_n) - \log_2(g(n))| < \delta(\epsilon, n)\} \\ &\leq \Pr\left\{|\log_2(x_n) - \log_2(g(n))| < \max_i \delta(\epsilon, i)\right\} \end{aligned} \tag{2.62}$$

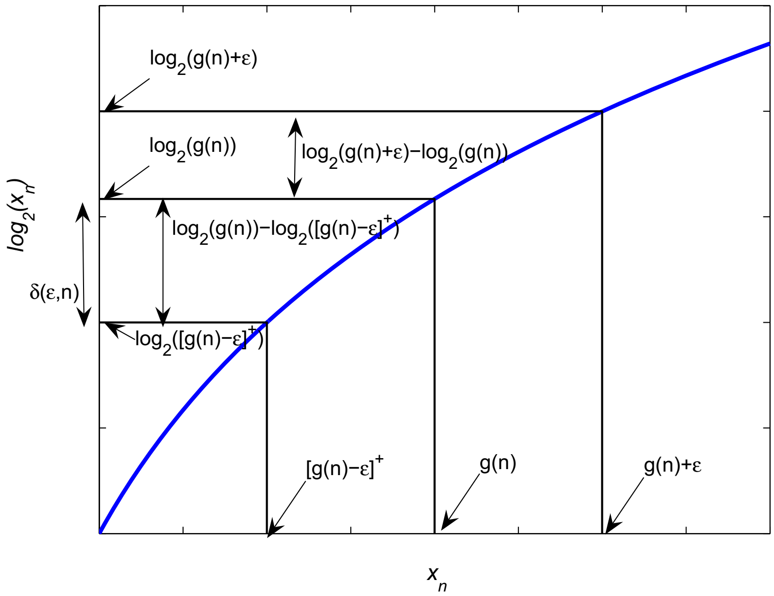

Figure 2.3: Convergence in probability for the $\log_2$ function.

where we have defined

$$\begin{aligned}\delta(\epsilon, n) &= \max\left\{\log_2(g(n)+\epsilon) - \log_2(g(n)), \log_2(g(n)) - \log_2\left([g(n)-\epsilon]^+\right)\right\} \\ &= \max\left\{\log_2\left(1+\frac{\epsilon}{g(n)}\right), -\log_2\left(\left[1-\frac{\epsilon}{g(n)}\right]^+\right)\right\}\end{aligned} \tag{2.63}$$

The first inequality of (2.62) follows directly from Fig. 2.3. The second one is straightforward, noticing that augmenting the interval increases the probability. Now, making use of (2.62) and denoting $\epsilon' = \max_i \delta(\epsilon, i)$, given (2.57) the following holds:

$$\lim_{n\to\infty} \Pr\{|\log_2(x_n) - \log_2(g(n))| < \epsilon'\} = 1 \ \ \forall\epsilon > 0. \tag{2.64}$$

However, this does not demonstrate convergence in probability unless the range of ϵ' is $(0, \infty)$. Let then study the possible values of $\epsilon' = \max_i \delta(\epsilon, i)$. First, since $0 < g(n) < \infty$, it is clear that:

$$\lim_{\epsilon\to 0} \delta(\epsilon, n) = 0 \ \ \forall n. \Longrightarrow \lim_{\epsilon\to 0} \epsilon' = 0 \tag{2.65}$$

Likewise, it is possible to derive that

$$\lim_{\epsilon\to g(n)} \delta(\epsilon, n) = \infty \ \ \forall n. \Longrightarrow \lim_{\epsilon\to \min_n g(n)} \epsilon' = \infty. \tag{2.66}$$

Finally, being $\delta(\epsilon, n)$ a continuous function on $\epsilon \in (0, \infty)$, so it is $\max_i \delta(\epsilon, i)$. Accordingly, $\epsilon' = \max_i \delta(\epsilon, i)$ will take all possible values between $(0, \infty)$ when ϵ ranges from

$(0, \min_n g(n)]$. Therefore, (2.64) holds $\forall \epsilon' > 0$, which demonstrates convergence and concludes the proof. ■

Lemma 2.2 *Let x_n, y_n be two independent sequences of random variables, and $g(n)$, $f(n)$ be two non-random sequences. Then:*

$$\left.\begin{array}{l} x_n - g(n) \xrightarrow{P} 0 \\ y_n - f(n) \xrightarrow{P} 0 \end{array}\right\} \Longrightarrow \min\{x_n, y_n\} - \min\{g(n), f(n)\} \xrightarrow{P} 0. \tag{2.67}$$

Remark 2.2 *The same holds when taking the* max *instead of the* min.

Proof: We follow similar steps as those in Lemma 2.1. However, in this case, we have a two dimensional mapping $\mathcal{T}(x, y) = \min\{x, y\}$ instead of a one-dimensional mapping $\mathcal{T}(x) = \log_2(x)$. First of all, being x_n and y_n independent, and given the individual convergence of both random sequences we can derive that:

$$\begin{aligned} &\lim_{n\to\infty} \Pr\{|x_n - g(n)| < \epsilon \,\cap\, |y_n - f(n)| < \epsilon\} = \\ &\qquad \lim_{n\to\infty} \Pr\{|x_n - g(n)| < \epsilon\} \cdot \Pr\{|y_n - f(n)| < \epsilon\} = 1, \ \forall \epsilon > 0. \end{aligned} \tag{2.68}$$

Let us now define the sets $\mathcal{X}(n) = (g(n) - \epsilon, g(n) + \epsilon)$ and $\mathcal{Y}(n) = (f(n) - \epsilon, f(n) + \epsilon)$, so that:

$$\Pr\{|x_n - g(n)| < \epsilon \,\cap\, |y_n - f(n)| < \epsilon\} = \Pr\{x_n \in \mathcal{X}(n) \,\cap\, y_n \in \mathcal{Y}(n)\}. \tag{2.69}$$

Notice that, within the cartesian product $\mathcal{X}(n) \times \mathcal{Y}(n)$, the mapping $\mathcal{T}(x, y) = \min\{x, y\}$ takes supremum and infimum values:

$$\begin{aligned} \sup_{(x_n, y_n)\in\mathcal{X}(n)\times\mathcal{Y}(n)} \min\{x_n, y_n\} &= \min\{g(n), f(n)\} + \epsilon. \\ \inf_{(x_n, y_n)\in\mathcal{X}(n)\times\mathcal{Y}(n)} \min\{x_n, y_n\} &= \min\{g(n), f(n)\} - \epsilon. \end{aligned} \tag{2.70}$$

Therefore,

$$\min\{g(n), f(n)\} - \epsilon < \min\{x_n, y_n\} < \min\{g(n), f(n)\} + \epsilon, \ \forall (x_n, y_n) \in \mathcal{X}(n) \times \mathcal{Y}(n).$$

Accordingly, it is possible to bound:

$$\begin{aligned} &\Pr\{x_n \in \mathcal{X}(n) \,\cap\, y_n \in \mathcal{Y}(n)\} \leq \\ &\qquad \Pr\{\min\{g(n), f(n)\} - \epsilon < \min\{x_n, y_n\} < \min\{g(n), f(n)\} + \epsilon\} \\ &\qquad\qquad = \Pr\{|\min\{x_n, y_n\} - \min\{g(n), f(n)\}| < \epsilon\}. \end{aligned} \tag{2.71}$$

First inequality follows from the fact that it might exist $(x_n, y_n) \notin \mathcal{X}(n) \times \mathcal{Y}(n)$ such that:

$$\min\{g(n), f(n)\} - \epsilon < \min\{x_n, y_n\} < \min\{g(n), f(n)\} + \epsilon.$$

Finally, considering (2.71) and (2.69), we can claim that:

$$\Pr\{|x_n - g(n)| < \epsilon \ \cap \ |y_n - f(n)| < \epsilon\} \leq \Pr\{|\min\{x_n, y_n\} - \min\{g(n), f(n)\}| < \epsilon\}. \tag{2.72}$$

Therefore, given (2.68), we obtain:

$$\lim_{n\to\infty} \Pr\{|\min\{x_n, y_n\} - \min\{g(n), f(n)\}| < \epsilon\} = 1, \ \forall\epsilon > 0. \tag{2.73}$$

which concludes the proof. ■

Lemma 2.3 *Let x_n be a sequence of random variables and $g(n)$, $f(n)$ be two non-random sequences. Furthermore, let*

$$x_n - g(n) \xrightarrow{P} 0 \ \text{and} \ \lim_{n\to\infty} g(n) - f(n) = 0. \tag{2.74}$$

Then, $x_n - f(n) \xrightarrow{P} 0$.

Proof: Let us first state the convergence $x_n - g(n) \xrightarrow{P} 0$ as:

$$\lim_{n\to\infty} \Pr\{|x_n - g(n)| < \epsilon\} = 1 \ \forall\epsilon > 0. \tag{2.75}$$

Likewise, the limit (2.74) can be rewritten as:

$$\forall\epsilon' > 0, \ \exists N_o \ \text{such that} \ |g(n) - f(n)| < \epsilon', \ \forall n \geq N_o. \tag{2.76}$$

Consider now a fixed value ϵ', for which $N_o = N_o(\epsilon')$. It is clear that, for $n \geq N_o$

$$\mathcal{X}_n = (g(n) - \epsilon, g(n) + \epsilon) \subseteq \mathcal{X}'_n = (f(n) - \epsilon - \epsilon', f(n) + \epsilon + \epsilon'). \tag{2.77}$$

Accordingly, for $n \geq N_o$:

$$\begin{aligned} \Pr\{|x_n - g(n)| < \epsilon\} &= \Pr\{x_n \in \mathcal{X}_n\} \\ &\leq \Pr\{x_n \in \mathcal{X}'_n\} \\ &= \Pr\{|x_n - f(n)| < \epsilon + \epsilon'\} \end{aligned} \tag{2.78}$$

Now, considering (2.78) and (2.75), we can show that

$$\lim_{n\to\infty} \Pr\{|x_n - f(n)| \leq \epsilon + \epsilon'\} = 1 \ \forall\epsilon > 0. \tag{2.79}$$

Finally, as ϵ' can be arbitrarily chosen within the interval $(0, \infty)$, it concludes the proof. ■

Lemma 2.4 *[55] Let the random sequences x_n and y_n converge in probability to x and y, respectively. Then,*

- $x_n \pm y_n \xrightarrow{P} x \pm y$,
- $x_n \cdot y_n \xrightarrow{P} x \cdot y$,
- *if* $y_n > 0, \forall n$, *then* $\frac{x_n}{y_n} \xrightarrow{P} \frac{x}{y}$.

2.3.1 The Law of Large Numbers

The law of large numbers is the most relevant representative of convergence in probability. It will be of definite help when deriving scaling laws for relay networks (as *e.g.*, asymptotic performance with the number of relays and/or users), and reads as follows:

Law of Large Numbers: Let $\{a_n\}$ be a sequence of independent random variables, with mean $\mu_n = \mathrm{E}\{a_n\} < \infty$ and bounded variance. Then, for $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} a_n - \frac{1}{n}\sum_{i=1}^{n} \mu_n \xrightarrow{P} 0. \tag{2.80}$$

Intuitively, it states that the sample mean converges to the average mean. The law can take the equivalent form, much more useful for the rest of the thesis:

$$\frac{1+\sum_{i=1}^{n} a_n}{n} - \frac{1+\sum_{i=1}^{n} \mu_n}{n} \xrightarrow{P} 0. \tag{2.81}$$

From it, and utilizing Lemma 2.1, the following corollary can be derived.

Corollary 2.3 *Let $\chi_n \geq 0, \forall n$ be a sequence of independent random variables with positive, bounded mean $\bar{\chi}_n = \mathrm{E}\{\chi_n\} \in (0,\infty)$ and bounded variance. Then:*

$$\log_2\left(1+\sum_{i=1}^{n}\chi_i\right) - \log_2\left(1+\sum_{i=1}^{n}\bar{\chi}_i\right) \xrightarrow{P} 0. \tag{2.82}$$

Proof: First, applying the law of large numbers (2.81), the following convergence holds:

$$\frac{1+\sum_{i=1}^{n}\chi_i}{n} - \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n} \xrightarrow{P} 0. \tag{2.83}$$

That is,

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{n} - \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}\right| < \epsilon\right\} = 1, \ \ \forall\epsilon. \tag{2.84}$$

Next, notice that given $\chi_i \geq 0$:

$$\frac{1+\sum_{i=1}^{n}\chi_i}{n} > 0, \ \ \forall n, \ \ \text{and} \ \ \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n} > 0, \ \ \forall n$$

Therefore, the probability above can be transformed as:

$$\begin{aligned}\Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{n} - \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}\right| < \epsilon\right\} &= \Pr\left\{\left|\frac{\frac{1+\sum_{i=1}^{n}\chi_i}{n}}{\frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}} - 1\right| \cdot \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n} < \epsilon\right\} \\ &= \Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i} - 1\right| < \frac{\epsilon}{\frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}}\right\}\end{aligned}$$

Let us now define $\bar{\chi}_{\min} = \min_i \bar{\chi}_i > 0$, since all random variables have non-zero mean. Hence, it is possible to bound:

$$\frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n} \geq \frac{1}{n} + \bar{\chi}_{\min} > \bar{\chi}_{\min}, \ \ \forall n. \tag{2.85}$$

Therefore,

$$\frac{\epsilon}{\frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}} < \frac{\epsilon}{\bar{\chi}_{\min}}, \ \ \forall n, \tag{2.86}$$

which implies that, for all n:

$$\Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i} - 1\right| < \frac{\epsilon}{\frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}}\right\} \leq \Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i} - 1\right| < \frac{\epsilon}{\bar{\chi}_{\min}}\right\}. \tag{2.87}$$

Plugging this onto the probability above allows to bound:

$$\Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{n} - \frac{1+\sum_{i=1}^{n}\bar{\chi}_i}{n}\right| < \epsilon\right\} \ \leq \ \Pr\left\{\left|\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i} - 1\right| < \frac{\epsilon}{\bar{\chi}_{\min}}\right\}$$

Now, defining $\epsilon' = \frac{\epsilon}{\bar{\chi}_{\min}} \in (0, \infty)$ and considering (2.84), it is clear that

$$\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i} - 1 \xrightarrow{P} 0. \tag{2.88}$$

Therefore, we may apply Lemma 2.1, to derive that

$$\log_2\left(\frac{1+\sum_{i=1}^{n}\chi_i}{1+\sum_{i=1}^{n}\bar{\chi}_i}\right) - \log_2(1) \xrightarrow{P} 0. \tag{2.89}$$

That is,

$$\log_2\left(1+\sum_{i=1}^{n}\chi_i\right) - \log_2\left(1+\sum_{i=1}^{n}\bar{\chi}_i\right) \xrightarrow{P} 0. \tag{2.90}$$

which concludes the proof. ∎

Chapter 3

Multiple-Parallel Relay Channel

3.1 Introduction

Relay channels, cooperative communications and distributed multiple-antenna processing are the *philosopher's stone* of wireless communications, called upon to transform wireless into wired. Like in gold rush, researchers worldwide wrestle to discover their true potential, and... patent it. Unfortunately, coming up with panaceas (and making them profitable for one's own enterprise) can take a lifetime of research and no success is guaranteed. Aiming, thus, at limiting the scope of my efforts to attainable ventures, this chapter merely focuses on a single query:

- How much can the point-to-point channel capacity be enlarged by placing N intermediate relay nodes?

All the contributions in this chapter try to address this question.

3.1.1 Overview

This chapter studies the multiple-parallel relay channel under AWGN (see Fig. 3.1). We consider single-antenna source, destination and relays, in a *time-invariant* and *memory-less* channel. Transmit and receive CSI are available at the source and destination, respectively, and all network nodes operate in the same frequency band. Finally, relaying is full-duplex: relays transmit and receive, simultaneously, using the same frequency.

Our focus is the channel capacity. We first provide an upper bound and then lower bounds by means of the achievable rates with different relaying techniques. We organize our contributions as follows:

- First, the *max-flow-min-cut* Theorem is used to provide an upper bound the capacity of the channel. The obtained bound turns out to scale as $\log_2 N$ in Rayleigh-fading (N is the number of relays). Likewise, it is shown to be reciprocal: source-relay channels and relay-destination channels can be interchanged without modifying the upper bound. That is, the bound does not distinguish whether source is transmitting to destination or *viceversa*.

- *Decode-and-forward* (D&F) is the first relaying technique studied. We derive its achievable rate as the generalization, to arbitrary number of relays N, of the single-relay result [10, Theorem 5]. Moreover, for Rayleigh-distributed fading, we obtain that as $N \to \infty$:
$$R_{\mathrm{D\&F}} - \mathcal{C}\left(2 \cdot W_0\left(\frac{\sqrt{N}}{2}\right) \cdot \frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{3.1}$$
This scaling law is shown to be always lower than $\log_2 \log N$. As argued in the sequel, this fact comes from the source-relays broadcast limitation.

- *Two-level partial decode-and-forward* (PD&F) is presented next as a generalized version of D&F. We derive its achievable rate and show that, for large number of relays, it does not outperform D&F. Hence, new techniques need to be considered.

- *Compress-and-forward* (C&F) is studied considering distributed Wyner-Ziv compression at the relays [39]. The achievable rate turns out to be the solution of a non-convex optimization problem with $1 + \sum_{i=1}^{N} \binom{N}{i} = 2^N$ constraints, which becomes untractable for large number of relays. As a benchmark to this technique, we thus provide a computable upper bound that can be calculated using known optimization tools. Finally, we demonstrate that the C&F rate scales as $\log_2 \log(N)$. In this case, it is due to the relays-destination MAC limitation.

- *Linear Relaying* (LR) is studied last. It consists of relay nodes transmitting, on every channel use, a linear combination of previously received signals [24]. We first derive the source temporal covariance that maximizes the achievable rate. Once the optimum source signal is obtained, the search for the optimum linear functions at the relays is an untractable optimization problem. To overcome this limitation, we propose a suboptimum approach. Regarding asymptotic performance, Dana *et al.* have previously shown that the achievable rate with amplify-based relaying scales as $\log_2(N)$ [56]. We present an alternative proof. Finally, we show that the LR rate is also reciprocal, in the same manner as the *max-flow-min-cut*, and unlike all other relaying techniques.

- To conclude the chapter, all techniques are adapted to half-duplex operation at the relays.

The remainder of the chapter is organized as follows: Section 3.2 presents the channel model, main assumptions and the *max-flow-min-cut* bound. Sections 3.3 and 3.4 study D&F and PD&F, respectively, while Sections 3.5 and 3.6 are devoted to C&F and LR, respectively. Section 3.7 presents numerical results on all techniques. Finally, Section 3.8 extends the analysis to half-duplex relaying, and Section 3.9 states conclusions.

3.2 Channel Model

The multiple-parallel relay channel (MPRC) is a channel in which there is a single source s communicating to a single destination d with the aid of a set $\mathcal{N} = \{1, \cdots, N\}$ of parallel relay nodes (see Fig. 3.1). Source, relays and destination transmit/receive scalar signals, and wireless channels are *time-invariant*, *memoryless*, modeled using a complex coefficient. The complex channel between source and destination is denoted by a, and the channels from source to relay $i \in \mathcal{N}$, and from relay $i \in \mathcal{N}$ to destination by b_i and c_i, respectively. Unlike the *multi-level* model [25], no wireless connectivity between relays exists in our model: we assume that relays use transmit directional antennas to the destination.

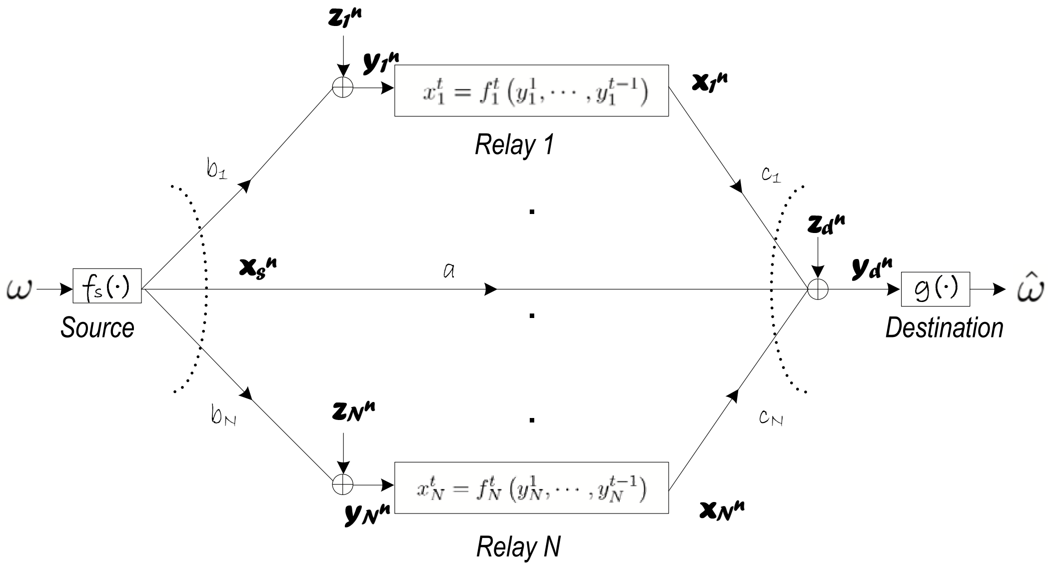

Figure 3.1: Multiple-Parallel Relay Channel.

We refer to the signal transmitted by the source as $\boldsymbol{x}_s^n = \{x_s^t\}_{t=1}^n \in \mathbb{C}^n$, where x_s^t is the transmitted symbol during channel use t and n the codeword length. This is received at the relays embedded into observations $\boldsymbol{y}_i^n = \{y_i^t\}_{t=1}^n \in \mathbb{C}^n$, $i \in 1, \cdots, N$:

$$\boldsymbol{y}_i^n = b_i \cdot \boldsymbol{x}_s^n + \boldsymbol{z}_i^n, \tag{3.2}$$

where $z_i \sim \mathcal{CN}(0, N_o)$ is additive white Gaussian noise (AWGN) at relay i. The signal transmitted by relay $i = 1, \cdots, N$ is denoted by $\boldsymbol{x}_i^n = \{x_i^t\}_{t=1}^n \in \mathbb{C}^n$, and is defined by means of a causal relaying function: $x_i^t = f^{r_i}\left(y_i^1, \cdots, y_i^{t-1}\right)$. Finally, the destination node $\boldsymbol{y}_d^n = \{y_d^t\}_{t=1}^n \in \mathbb{C}^n$ receives the superposition of signals transmitted by source and relays

$$\boldsymbol{y}_d^n = a \cdot \boldsymbol{x}_s^n + \sum_{i=1}^{N} c_i \cdot \boldsymbol{x}_i^n + \boldsymbol{z}_d^n, \tag{3.3}$$

with $z_d \sim \mathcal{CN}(0, N_o)$. For the channel, we assume the following:

(*A1*). *Full-duplex operation*: relays transmit and receive simultaneously in the same frequency band. It can be implemented in practice using different, uncoupled antennas for transmission and reception.

(*A2*). Transmit *channel state information* (CSI) and receive CSI at the source and destination, respectively. Channel awareness not only includes the source-destination channel, but also the source-relay and relay-destination channels. It may be difficult to satisfy this assumption in practice. However, distributed knowledge can be obtained in slow varying channels via channel reciprocity, and feedback, during a set up phase.

(*A3*). The total transmitted power in the network is constrained to P:

$$\frac{1}{n}\sum_{t=1}^{n}\left(\mathrm{E}\left\{|x_s^t|^2\right\}+\sum_{i=1}^{N}\mathrm{E}\left\{|x_i^t|^2\right\}\right)\leq P. \tag{3.4}$$

The aim is to provide a fair comparison with the no-relay channel with source power P.

3.2.1 Preliminaries

Definition 3.1 *A $\left(n, 2^{nR}\right)$ code for the MPRC is defined by:*

- *a set of messages $\mathcal{W}=\left\{1,\cdots,2^{nR}\right\}$, a signal space $\mathcal{X}_s$ and a source encoding function*

$$f_s : \mathcal{W}\rightarrow\mathcal{X}_s^n,$$

- *N signal spaces $\mathcal{X}_i,\ i=1,\cdots,N$ and N causal relay functions*

$$f_{r_i} : \mathcal{Y}_i^n\rightarrow\mathcal{X}_i^n,\ i=1,\cdots,N,$$

- *a decoding function $g:\mathcal{Y}_d^n\rightarrow\mathcal{W}$.*

Definition 3.2 *A rate R is achievable if there exists a sequence of codes $\left(n, 2^{nR}\right)$ for which $\lim_{n\rightarrow\infty}P_e^n=0$, where*

$$P_e^n=\frac{1}{2^{nR}}\sum_{\omega}\Pr\left\{g\left(\boldsymbol{y}_d^n\right)\neq\omega|\omega\text{ was sent}\right\}. \tag{3.5}$$

Definition 3.3 *The capacity of the MPRC is the supremum of all transmission rates that are achievable. Recently, it has been derived as the mutual information between the source input and the destination output, maximized over the input distribution and over the relay functions:*

$$\mathcal{C}=\lim_{n\rightarrow\infty}\max_{p_{\boldsymbol{x}_s^n},\left\{f_{r_i}\right\}_{i=1}^{N}}\frac{1}{n}\cdot I\left(\boldsymbol{x}_s^n;\boldsymbol{y}_d^n\right), \tag{3.6}$$

and subject to causal relaying in Definition 3.1 and power constraint (3.4) [57, Theorem 2.1].

However, deriving a single letter expression for the capacity (3.6) is still an open problem. In this section, we present an upper bound on it.

3.2.2 Max-flow-min-cut Upper Bound

The *max-flow-min-cut* theorem defines a necessary (but not sufficient) condition that any transmission rate within a connected network must satisfy in order to be achievable. It is evaluated here in order to provide an upper bound on (3.6). The condition is expressed using a $\max\min$ operation, and for our problem is formulated as follows [21, Theorem 14.10.1]:

$$\mathcal{C} \leq \max_{p(x_s,\boldsymbol{x}_{1:N})} \left\{ \min_{\mathcal{T}\subset\{s,1,\cdots,N,d\}:s\in\mathcal{T},d\in\mathcal{T}^c} I\left(\boldsymbol{x}_{\mathcal{T}};\boldsymbol{y}_{\mathcal{T}^c}|\boldsymbol{x}_{\mathcal{T}^c}\right)\right\}, \tag{3.7}$$

where maximization is subject to (3.4). The bound can be interpreted resorting to data flows; clearly, (3.7) states that the maximum transmission rate between the source s and the destination d is always lower than or equal to the maximum data flow between the set of inputs $\boldsymbol{x}_{\mathcal{T}}$ and the set of outputs $\boldsymbol{y}_{\mathcal{T}^c}$, whenever $s\in\mathcal{T}$ and $d\in\mathcal{T}^c$. This is, somehow, obvious. Notice that the minimization in (3.7) is taken over $1+\sum_{i=1}^{N}\binom{N}{i}$ elements, as much as possible subsets $\mathcal{T}$ (*also known as* network cuts). However, here, we only consider the MAC and BC cuts, which yields the following bound.

Upper Bound: The capacity of the AWGN multiple-parallel relay channel satisfies

$$\begin{aligned}\mathcal{C} &\leq \max_{0\leq\rho\leq1}\min\left\{\mathcal{C}\left(\left(|a|^2+\rho\sum_{i=1}^N|c_i|^2\right)\frac{P}{N_o}\right),\mathcal{C}\left(\left(|a|^2+\sum_{i=1}^N|b_i|^2\right)(1-\rho)\frac{P}{N_o}\right)\right\} \\ &= \mathcal{C}\left(\left(|a|^2+\frac{\sum_{i=1}^N|b_i|^2}{|a|^2+\sum_{i=1}^N|b_i|^2+\sum_{i=1}^N|c_i|^2}\sum_{i=1}^N|c_i|^2\right)\frac{P}{N_o}\right)\end{aligned} \tag{3.8}$$

Remark 3.1 *Interestingly, the obtained upper bound is reciprocal. That is, interchanging channels $b_i\leftrightarrow c_i$ maintains (3.8) unaltered. This has a nice, and democratic, interpretation: the upper bound does not distinguish whether s is transmitting to d or viceversa. Although yet undemonstrated, the MPRC capacity (3.6) is expected not to distinguish either and to perform equally from s to d than from d to s. This is an open question that we are not able to answer in this dissertation.*

Proof: Let us consider only two network cuts at the minimization of (3.7): $\mathcal{T}=\{s,1,\cdots,N\}$ and $\mathcal{T}=\{s\}$. With them, we can bound

$$\mathcal{C}\leq\max_{p(x_s,\boldsymbol{x}_{1:N})}\min\left\{I\left(x_s,\boldsymbol{x}_{1:N};y_d\right),I\left(x_s;y_d,\boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right)\right\}, \tag{3.9}$$

where maximization is subject to power constraint (3.4). We build upon [11, Section VII.B] to evaluate such an upper bound to our network. Firstly, we notice that both parts of minimization are maximized, under AWGN, with $p\left(x_s, \boldsymbol{x}_{1:N}\right)$ jointly Gaussian. Hence, Gaussian distribution is the optimum input distribution. Furthermore, it is easy to show that the two elements of the minimization are individually maximized (and so it is the minimum of them) for $x_1, \cdots, x_N$ fully correlated:

$$\left|\frac{\mathrm{E}\left\{x_i x_j^*\right\}}{\sqrt{\mathrm{E}\left\{|x_i|^2\right\}}\sqrt{\mathrm{E}\left\{|x_j|^2\right\}}}\right| = 1,\ \forall i, j \in \mathcal{N}. \tag{3.10}$$

This is due to two facts: *i*) the multi-access cut $I\left(x_s, \boldsymbol{x}_{1:N}; y_d\right)$ takes maximum value when random variables add coherently, *i.e.* when they are fully correlated. *ii*) Conditioning reduces entropy [21, Theorem 2.6.5]. That is, $H\left(A|B\right) \geq H\left(A|B,C\right)$, with equality whenever B and C are fully correlated or A is conditionally independent of C, given B. Therefore, since

$$\begin{aligned} I\left(x_s; y_d, \boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right) &= H\left(y_d, \boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right) - H\left(y_d, \boldsymbol{y}_{1:N}|x_s, \boldsymbol{x}_{1:N}\right) \\ &= H\left(y_d, \boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right) - H\left(z_d, \boldsymbol{z}_{1:N}\right), \end{aligned} \tag{3.11}$$

it is clear that the broadcast cut $I\left(x_s; y_d, \boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right)$ takes maximum value when all $x_1, \cdots, x_N$ are fully correlated.

With such an input distribution, and defining $\rho_{s,r} = |\mathrm{E}\left\{x_s x_j^*\right\}/\sqrt{\mathrm{E}\left\{|x_s|^2\right\}}\sqrt{\mathrm{E}\left\{|x_j|^2\right\}}|, \forall j \in \mathcal{N}$ as the source-relays correlation, we can evaluate:

$$\begin{aligned} I\left(x_s; y_d, \boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}\right) &= \mathcal{C}\left(\left(|a|^2 + \sum_{i=1}^{N}|b_i|^2\right)\frac{\mathrm{E}\left\{|x_s|^2|\boldsymbol{x}_{1:N}\right\}}{N_o}\right) \\ &= \mathcal{C}\left(\left(|a|^2 + \sum_{i=1}^{N}|b_i|^2\right)\frac{\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}}{N_o}\right) \\ &= \mathcal{C}\left(\left(|a|^2 + \sum_{i=1}^{N}|b_i|^2\right)\left(1-\rho\right)\frac{P}{N_o}\right), \end{aligned} \tag{3.12}$$

where we have defined

$$\left(1-\rho\right)P = \left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}. \tag{3.13}$$

Furthermore, it is also possible to bound:

$$\begin{aligned} I\left(x_s, \boldsymbol{x}_{1:N}; y_d\right) &= I\left(x_s; y_d|\boldsymbol{x}_{1:N}\right) + I\left(\boldsymbol{x}_{1:N}; y_d\right) \\ &\leq \mathcal{C}\left(\frac{|a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}}{N_o}\right) \\ &\quad + \mathcal{C}\left(\frac{\left(|a|^2 + \sum_{i=1}^{N}|c_i|^2\right)\left(\rho_{s,r}^2\mathrm{E}\left\{|x_s|^2\right\} + \sum_{i=1}^{N}\mathrm{E}\left\{|x_i|^2\right\}\right)}{N_o + |a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}}\right) \end{aligned} \tag{3.14}$$

$$= \mathcal{C}\left(\frac{|a|^2(1-\rho)P}{N_o}\right)+\mathcal{C}\left(\frac{\left(|a|^2+\sum_{i=1}^N|c_i|^2\right)\rho P}{N_o+|a|^2(1-\rho)P}\right)$$

$$= \mathcal{C}\left(\left(|a|^2+\rho\sum_{i=1}^N|c_i|^2\right)\frac{P}{N_o}\right),$$

where second equality is derived by noting that $\rho_{s,r}^2 \mathrm{E}\left\{|x_s|^2\right\} = \mathrm{E}\left\{|x_s|^2\right\} - (1-\rho)P$ from definition (3.13), and $\mathrm{E}\left\{|x_s|^2\right\} + \sum_{i=1}^N \mathrm{E}\left\{|x_i|^2\right\} = P$ from power constraint (3.4). Furthermore, the inequality has been derived by noting that

$$\begin{aligned} I(\boldsymbol{x}_{1:N}; y_d) &= H(y_d) - H(y_d|\boldsymbol{x}_{1:N}) \qquad (3.15)\\ &= \log_2\left(|a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\} + \left(|a|\rho_{s,r}\sqrt{\mathrm{E}\left\{|x_s|^2\right\}}+\sum_{i=1}^N|c_i|\sqrt{\mathrm{E}\left\{|x_i|^2\right\}}\right)^2 + N_o\right)\\ &\quad -\log_2\left(|a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}+N_o\right)\\ &= \mathcal{C}\left(\frac{\left(|a|\rho_{s,r}\sqrt{\mathrm{E}\left\{|x_s|^2\right\}}+\sum_{i=1}^N|c_i|\sqrt{\mathrm{E}\left\{|x_i|^2\right\}}\right)^2}{N_o+|a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}}\right)\\ &\leq \mathcal{C}\left(\frac{\left(|a|^2+\sum_{i=1}^N|c_i|^2\right)\left(\rho_{s,r}^2\mathrm{E}\left\{|x_s|^2\right\}+\sum_{i=1}^N\mathrm{E}\left\{|x_i|^2\right\}\right)}{N_o+|a|^2\left(1-\rho_{s,r}^2\right)\mathrm{E}\left\{|x_s|^2\right\}}\right). \end{aligned}$$

The second equality comes from the fact that $x_1, \cdots, x_N$ are fully correlated, having all correlation $\rho_{s,r}$ with the source. Finally, the inequality holds since maximal ratio combining upper bounds the square value of the sum.

At this point, it remains to maximize over ρ the minimum of the two cuts. For that purpose, notice that (3.14) is an increasing function of ρ, while (3.12) is decreasing. Therefore, the maximum over ρ of the minimum of both is given at the value where both are equal:

$$\rho^* = \frac{\sum_{i=1}^N|b_i|^2}{|a|^2+\sum_{i=1}^N|b_i|^2+\sum_{i=1}^N|c_i|^2}. \qquad (3.16)$$

This concludes the proof. As a remark, it can be shown that, for $N \to \infty$ and unitary-mean, Rayleigh fading, the bound converges in probability to $\mathcal{C}\left(\frac{N}{2}\frac{P}{N_o}\right)$. ■

The above result is also an upper bound on the capacity of the *multi-level* relay channel, as we did not take any assumption on the inter-relay connectivity. Likewise, notice that Gastpar also considers the MAC and BC cuts in [32]; however, in that work, both are optimized separately over the input distributions, resulting in a less tight bound than (3.8).

3.3 Decode-and-forward

Decode-and-forward (D&F) is the first relaying technique considered in this chapter. With it, the relay nodes fully decode the source message. Then, they re-encode it and retransmit it coherently to destination. The relays, thus, help the source by mimicking a transmit antenna array towards the receiver. This technique is presented for a single relay in [10, Theorem 5] and extended here to multiple relays. The extension is based upon a key fact: the higher the cardinality of relays that decode the source message, the higher the multiple-antenna gain towards the destination. However, the more relays, the lower the source's rate that all can decode. Clearly, then, there is an optimum subset of relays to be active: the one who better trade among the two effects.

Theorem 3.1 *The AWGN multiple-parallel relay channel achieves the rate*

$$\begin{aligned} R_{\mathrm{D\&F}} &= \max_{1\leq m\leq N}\max_{0<\eta\leq 1}\min\left\{\mathcal{C}\left(\left(|a|^2+(1-\eta)\sum_{i=1}^{m}|c_i|^2\right)\frac{P}{N_o}\right),\mathcal{C}\left(|b_m|^2\eta\frac{P}{N_o}\right)\right\} \quad (3.17)\\ &= \max_{1\leq m\leq N}\mathcal{C}\left(|b_m|^2\min\left\{1,\frac{|a|^2+\sum_{i=1}^{m}|c_i|^2}{|b_m|^2+\sum_{i=1}^{m}|c_i|^2}\right\}\frac{P}{N_o}\right) \end{aligned}$$

with decode-and-forward. The source-relay channels have been ordered as:

$$|b_1|\geq\cdots\geq|b_m|\geq\cdots\geq|b_N|. \quad (3.18)$$

Proof: Let the N relay nodes be ordered as in (3.18), and assume that only the subset $\mathcal{R}_m=\{1,\cdots,m\}\subseteq\mathcal{N}$ is active. The source selects message $\omega\in\left\{1,\cdots,2^{nR}\right\}$ for transmission, and splits it into B sub-messages of κR bits, with $\kappa=\frac{n}{B}$, *i.e.,* $\omega=\left[\omega^1,\cdots,\omega^B\right]$. The submessages are then pipelined into $B+1$ channel blocks, with κ channels uses per block. We consider $n,\kappa,B\gg1$, so that $\frac{B}{B+1}\approx1$.

The source encodes the message using a block-Markov approach [21, Sec. 14.7]: on every block b, it transmits the sub-message ω^b to the relays and destination. Simultaneously, it cooperates with the relays in $\mathcal{R}_m$ to retransmit its previously transmitted sub-message ω^{b-1}. To that end, source and relays transmit

$$\begin{aligned} \boldsymbol{x}_s^{\kappa}[b] &= \boldsymbol{s}^{\kappa}\left(\omega^b\right)+\frac{a^*}{\sqrt{|a|^2+\sum_{i\in\mathcal{R}_m}|c_i|^2}}\boldsymbol{v}^{\kappa}\left(\omega^{b-1}\right),\\ \boldsymbol{x}_i^{\kappa}[b] &= \frac{c_i^*}{\sqrt{|a|^2+\sum_{i\in\mathcal{R}_m}|c_i|^2}}\cdot\boldsymbol{v}^{\kappa}\left(\omega^{b-1}\right),\ \ \forall\, i\in\mathcal{R}_m. \end{aligned}$$

Here, we have selected $\boldsymbol{s}^{\kappa}(\cdot)$ to be a random Gaussian codebook, generated *i.i.d.* from $s \sim \mathcal{CN}(0, \eta P)$. In turn, $\boldsymbol{v}^{\kappa}(\cdot)$ is a random Gaussian codebook, generated from $v \sim \mathcal{CN}(0, (1-\eta) P)$. Being $0 < \eta \leq 1$, we notice that the power constraint (3.4) is satisfied with equality. The received signal at the relays and destination thus reads:

$$\begin{aligned}
\boldsymbol{y}_i^{\kappa}[b] &= b_i \cdot \boldsymbol{s}^{\kappa}\left(\omega^b\right) + \frac{b_i \cdot a^*}{\sqrt{|a|^2 + \sum_{i \in \mathcal{R}_m} |c_i|^2}} \boldsymbol{v}^{\kappa}\left(\omega^{b-1}\right) + N_o, \ \ \forall\, i \in \mathcal{R}_m \\
\boldsymbol{y}_d^{\kappa}[b] &= a \cdot \boldsymbol{s}^{\kappa}\left(\omega^b\right) + \sqrt{|a|^2 + \sum_{i \in \mathcal{R}_m} |c_i|^2} \cdot \boldsymbol{v}^{\kappa}\left(\omega^{b-1}\right) + N_o.
\end{aligned}$$

On every block b, all nodes in $\mathcal{R}_m$ are able to decode ω^b and therefore retransmit it coherently during block $b+1$ *iff* (assume they have all estimated ω^{b-1} correctly):

$$\begin{aligned}
R &\leq \min_{i \in \mathcal{R}_m} I\left(s; y_i | v\right) \qquad (3.19) \\
&= \mathcal{C}\left(|b_m|^2 \eta \frac{P}{N_o}\right),
\end{aligned}$$

The equality is due to ordering (3.18). Next, consider the decoding at the destination in the same block b. Assume that sub-message ω^{b-2} has been successfully decoded. Then, using *parallel channel* decoding as in [58, Sec III.B], the destination uses signals $\boldsymbol{y}_d^{\kappa}[b-1]$ and $\boldsymbol{y}_d^{\kappa}[b]$ to estimate message ω^{b-1}. It can do so reliably *iff*

$$\begin{aligned}
R &\leq I\left(s; y_d | v\right) + I\left(v; y_d\right) \qquad (3.20) \\
&= I\left(s, v; y_d\right) \\
&= \mathcal{C}\left(\left(|a|^2 + (1-\eta) \sum_{i=1}^{m} |c_i|^2\right) \frac{P}{N_o}\right).
\end{aligned}$$

where $I\left(s; y_d | v\right)$ is the mutual information that destination extracts from $\boldsymbol{y}_d^{\kappa}[b-1]$ to decode ω^{b-1}, given *a priori* knowledge of ω^{b-2}. In turn, $I\left(v; y_d\right)$ is the mutual information extracted from $\boldsymbol{y}_d^{\kappa}[b]$ to decode ω^{b-1}. Now, considering both (3.19) and (3.20) we obtain the minimization in (3.17), which shows the maximum achievable rate for a given set $\mathcal{R}_m$ and η. However, note that (3.19) is a strictly decreasing function with η, while (3.20) is an increasing function. Hence, the maximum over η of the minimum of the two is given at the η^* that makes both equal, *i.e.,*

$$\eta^* = \frac{|a|^2 + \sum_{i=1}^{m} |c_i|^2}{|b_m|^2 + \sum_{i=1}^{m} |c_i|^2}. \qquad (3.21)$$

However, η^* has to be lower than or equal to one. Hence, second equality in (3.17) holds. Finally, we may arbitrarily choose the decoding set $\mathcal{R}_m$ from $\{\mathcal{R}_1, \cdots, \mathcal{R}_N\}$. ■

As mentioned in the proof, parameter m in Theorem 3.1 denotes the cardinality of the decoding set, that is, the number of relays that the source node selects to decode and retransmit. Notice that $\mathcal{C}\left(\left(|a|^2+(1-\eta)\sum_{i=1}^{m}|c_i|^2\right)\frac{P}{N_o}\right)$, increases with the cardinality m, while $\mathcal{C}\left(|b_m|^2\eta\frac{P}{N_o}\right)$ decreases. Therefore, for the optimum number of relays m^* both quantities will be similar. Or, in other words, m^* is the cardinality of active relays that better trades among the two of them.

3.3.1 Asymptotic Performance

Consider unitary-mean, Rayleigh fading[1] within the network: $a, b_i, c_i \sim \mathcal{CN}(0,1), i=1,\cdots,N$. The following convergence in probability can be proven for the rate in Theorem 3.1.

Theorem 3.2 *Let $W_0(x)$ be the branch zero of the Lambert W function*[2]*. Then, for $N\to\infty$:*

$$R_{\text{D\&F}}-\mathcal{C}\left(2\cdot W_0\left(\frac{\sqrt{N}}{2}\right)\cdot\frac{P}{N_o}\right)\xrightarrow{P}0. \tag{3.22}$$

Proof: We focus on the first equality in (3.17). For $N\gg 1$, it can be re-stated as:

$$R_{\text{D\&F}}=\max_{k\in(0,1]}\max_{0<\eta\leq 1}\min\left\{\mathcal{C}\left(\left(|a|^2+(1-\eta)\sum_{i=1}^{kN}|c_i|^2\right)\frac{P}{N_o}\right),\mathcal{C}\left(|b_{kN}|^2\eta\frac{P}{N_o}\right)\right\}, \tag{3.23}$$

where, as mentioned, $|b_{kN}|^2$ is the kN^{th} ordered channel as in (3.18). Firstly, considering $\eta\neq 1$, we can apply Corollary 2.3 to show that:

$$\mathcal{C}\left(\left(|a|^2+(1-\eta)\sum_{i=1}^{kN}|c_i|^2\right)\frac{P}{N_o}\right)-\mathcal{C}\left((1+(1-\eta)kN)\frac{P}{N_o}\right)\xrightarrow{P}0. \tag{3.24}$$

Furthermore, we can use Lemma 2.3 to derive:

$$\mathcal{C}\left(\left(|a|^2+(1-\eta)\sum_{i=1}^{kN}|c_i|^2\right)\frac{P}{N_o}\right)-\mathcal{C}\left((1-\eta)kN\frac{P}{N_o}\right)\xrightarrow{P}0. \tag{3.25}$$

Now, recall that the non-ordered source-relay channels $|b|^2$ are *i.i.d.*, unitary-mean, exponentially distributed. Therefore, they all share cdf $F(x)=Pr\{|b|^2<x\}=1-e^{-x}$, whose inverse function is $Q(y)=F^{-1}(x)=-\log(1-y)$. Hence, we may apply ordered statistics to show the following convergence in probability, for $N\to\infty$ [60, Section 10.2]:

$$|b_{kN}|^2\xrightarrow{P}Q(1-k)=-\log(k). \tag{3.26}$$

[1] Channels are *time-invariant* and *memory-less*.

[2] The function is defined as $W_0(x)=\{y\in\mathbb{R}: ye^y=x\}$. It satisfies $W_0(x)<\log(x)\ \forall x>e$ [59].

Due to Lemma 2.1, we thus show:

$$\mathcal{C}\left(|b_{kN}|^2\eta\frac{P}{N_o}\right)\xrightarrow{P}\mathcal{C}\left(-\log\left(k\right)\eta\frac{P}{N_o}\right) \tag{3.27}$$

Now, notice that $\max$ and $\min$ functions keep convergence unaltered (see Lemma 2.2). Therefore, we can plug (3.25) and (3.27) into definition (3.23) to derive

$$R_{\mathrm{D\&F}}-\max_{k\in(0,1]}\max_{0<\eta\leq 1}\min\left\{\mathcal{C}\left((1-\eta)\,kN\frac{P}{N_o}\right),\mathcal{C}\left(-\log\left(k\right)\eta\frac{P}{N_o}\right)\right\}\xrightarrow{P}0. \tag{3.28}$$

It is easy to notice that the optimum value of η in (3.28) yields $(1-\eta^*)\,kN=-\log\left(k\right)\eta^*$, *i.e,* $\eta^*=\frac{kN}{kN-\log(k)}$. Therefore, we can rewrite:

$$R_{\mathrm{D\&F}}-\max_{k\in(0,1]}\mathcal{C}\left(-\frac{\log\left(k\right)kN}{kN-\log\left(k\right)}\frac{P}{N_o}\right)\xrightarrow{P}0. \tag{3.29}$$

Finally, using standard arguments for non-convex optimization, the maximization over k is shown to be attained at $k^*=\left\{k:-\log\left(k\right)=\sqrt{kN}\right\}=4\frac{W_0\left(\sqrt{N}/2\right)^2}{N}$. As a result,

$$R_{\mathrm{D\&F}}-\mathcal{C}\left(-\frac{\log\left(k^*\right)k^*N}{k^*N-\log\left(k^*\right)}\frac{P}{N_o}\right)\xrightarrow{P}0. \tag{3.30}$$

Now, the following can be derived:

$$\begin{aligned}\mathcal{C}\left(-\frac{\log\left(k^*\right)k^*N}{k^*N-\log\left(k^*\right)}\frac{P}{N_o}\right) &= \mathcal{C}\left(\frac{\sqrt{k^*N}k^*N}{k^*N+\sqrt{k^*N}}\frac{P}{N_o}\right) \\ &= \mathcal{C}\left(\frac{\sqrt{k^*N}}{1+(k^*N)^{-\frac{1}{2}}}\frac{P}{N_o}\right) \\ &= \mathcal{C}\left(\frac{2W_0\left(\sqrt{N}/2\right)}{1+\left(2W_0\left(\sqrt{N}/2\right)\right)^{-1}}\frac{P}{N_o}\right),\end{aligned} \tag{3.31}$$

where first equality comes from definition of k^* and third from its closed form expression. Therefore, noting that $\lim_{N\to\infty}\left(2W_0\left(\sqrt{N}/2\right)\right)^{-1}=0$, we can claim that

$$\lim_{N\to\infty}\mathcal{C}\left(-\frac{\log\left(k^*\right)k^*N}{k^*N-\log\left(k^*\right)}\frac{P}{N_o}\right)-\mathcal{C}\left(2W_0\left(\frac{\sqrt{N}}{2}\right)\frac{P}{N_o}\right)=0 \tag{3.32}$$

Hence, considering (3.32) and (3.30) on Lemma 2.3, it concludes the proof. ■

This result shows that $R_{\mathrm{D\&F}}$ performs worse than $\mathcal{C}\left(\log\left(\frac{N}{\sqrt{2}}\right)\frac{P}{N_o}\right)$. Hence, for large N, it is far from the *max-flow-min-cut* bound, which scales as $\mathcal{C}\left(\frac{N}{2}\frac{P}{N_o}\right)$. This fact is explained by the source-relays broadcast limitation: the maximum transmission rate with D&F is constrained by the maximum rate at which the source can communicate with relays. This rate is lower than, or equal to, the source to the best-relay channel capacity. It is well-known that the highest capacity of a set of N *i.i.d.* Rayleigh-faded capacities scales as $\mathcal{C}\left(\log(N)\frac{P}{N_o}\right) < \mathcal{C}\left(N\frac{P}{N_o}\right)$ [61, Lemma 2]. This result suggests that other techniques may be more effective for large N.

3.4 Partial Decode-and-Forward

Partial decode-and-forward (PD&F) was introduced in [10] for the single-relay channel and is a generalization of *decode-and-forward*. The relays are only required to partially decode the source data, which is useful given that the amount of decoded data can be adapted to the source-relay channel quality.

In this dissertation, we consider *two-level* partial decoding, which is defined as follows: the source splits its transmitted message ω into two independent messages, and simultaneously transmits them using superposition coding. The first message is directly sent to the destination. The second message is transmitted through the relays using block-Markov encoding. The destination, in turn, estimates both messages using successive decoding. The performance of PD&F is optimized by optimally allocating power on the two messages. Previous results regarding partial decoding for a single relay can be found in [11, 22, 57] and in the original work [10]. We extend these results here to the multi-relay case.

Theorem 3.3 *The AWGN multiple-parallel relay channel achieves the rate*

$$R_{\mathrm{PD\&F}} = \max_{1\leq m\leq N} \max_{0\leq(\eta,\beta)\leq 1} \min\left\{\mathcal{C}\left(\left(|a|^2+\beta(1-\eta)\sum_{i=1}^{m}|c_i|^2\right)\frac{P}{N_o}\right),\right.$$
$$\left.\mathcal{C}\left(\frac{|b_m|^2\beta\eta P}{N_o+|b_m|^2(1-\beta)P}\right)+\mathcal{C}\left(|a|^2(1-\beta)\frac{P}{N_o}\right)\right\} \quad (3.33)$$

with two-level partial decode-and-forward. The source-relay channels have been ordered as in (3.18).

Remark 3.2 *As previously, m denotes the cardinality of the decoding set. Additionally, parameter $1-\beta$ is the fraction of power dedicated to the non-relayed message, and $1-\eta$ the* *correlation between the source and relays. The maximization is not convex on η, β and shall be solved using exhaustive two-dimensional algorithms.*

Proof: The proof follows equivalent arguments to the proof of Theorem 3.1. First, let the N relay nodes be ordered as in (3.18) and assume that only the subset $\mathcal{R}_m = \{1, \cdots, m\} \subseteq \mathcal{N}$ is active. The source divides its transmitted message $\omega \in \left\{1, \cdots, 2^{nR}\right\}$ into two independent messages $\omega_d \in \left\{1, \cdots, 2^{nR_d}\right\}$ and $\omega_r \in \left\{1, \cdots, 2^{nR_r}\right\}$, and transmits them using superposition coding. The total transmission rate is $R = R_d + R_r$. Each of the messages is then split (as for D&F) into B sub-messages, and pipelined onto $B + 1$ channel blocks. The first message is directly sent to destination, while the other is transmitted through the relays using block-Markov encoding. On a given block b, source and relays transmit:

$$\boldsymbol{x}_s^\kappa [b] = \boldsymbol{u}^\kappa \left(\omega_d^b\right) + \boldsymbol{s}^\kappa \left(\omega_r^b\right) + \frac{a^*}{\sqrt{|a|^2 + \sum_{i\in\mathcal{R}_m} |c_i|^2}} \boldsymbol{v}^\kappa \left(\omega_r^{b-1}\right),$$

$$\boldsymbol{x}_i^\kappa [b] = \frac{c_i^*}{\sqrt{|a|^2 + \sum_{i\in\mathcal{R}_m} |c_i|^2}} \cdot \boldsymbol{v}^\kappa \left(\omega_r^{b-1}\right), \quad \forall\, i \in \mathcal{R}_m$$

where we have selected $\boldsymbol{u}^\kappa(\cdot)$ to be a random Gaussian codebook, generated *i.i.d.* from $u \sim \mathcal{CN}(0, (1-\beta) P)$ and $\boldsymbol{s}^\kappa(\cdot)$ to be a random Gaussian codebook, generated *i.i.d.* from $s \sim \mathcal{CN}(0, \beta\eta P)$. In turn, $\boldsymbol{v}^\kappa(\cdot)$ is also random Gaussian, generated from $v \sim \mathcal{CN}(0, \beta(1-\eta) P)$. Since $0 \le \beta \le 1$,$0 \le \eta \le 1$, notice that the power constraint (3.4) is satisfied with equality.

On block b, all relay nodes in $\mathcal{R}_m$ attempt to decode ω_r^b. They can (all) do so *iff* (assume ω_r^{b-1} has been detected correctly):

$$\begin{aligned} R_r &\le \min_{i\in\mathcal{R}_m} I(s; y_i|v) \\ &= \mathcal{C}\left(\frac{|b_m|^2\beta\eta P}{N_o + |b_m|^2(1-\beta) P}\right), \end{aligned} \tag{3.34}$$

where the inequality is due to the ordering in (3.18). At the same block b, the destination is able to decode ω_r^{b-1} from $\boldsymbol{y}_d^\kappa[b-1]$ and $\boldsymbol{y}_d^\kappa[b]$ *iff* (assuming ω_r^{b-2} detected correctly) [58, Section III.B]:

$$\begin{aligned} R_r &\le I(s; y_d|v) + I(v; y_d) \\ &= I(s, v; y_d) \\ &= \mathcal{C}\left(\frac{\beta\eta|a|^2 + \beta(1-\eta)\left(|a|^2 + \sum_{i=1}^m |c_i|^2\right)}{N_o + |a|^2(1-\beta) P} P\right) \end{aligned} \tag{3.35}$$

Moreover, once decoded ω_r^{b-1}, the destination removes its contribution onto $\boldsymbol{y}_d^{\kappa}[b-1]$ and decodes message ω_d^{b-1} (assumed ω_r^{b-2} well estimated). It can do so *iff*:

$$\begin{aligned} R_d &\leq I\left(u; y_d | s, v\right) \\ &= \mathcal{C}\left(|a|^2 \left(1-\beta\right) \tfrac{P}{N_o}\right). \end{aligned} \tag{3.36}$$

Finally, the transmission rate equals $R = R_r + R_d$. Thus, adding (3.35) and (3.36) we obtain the left part of minimization in (3.33), while by adding (3.34) and (3.36) we obtain the right part. This shows the achievable rate for a given $\mathcal{R}_m$. However, notice that we may arbitrarily choose the decoding set $\mathcal{R}_m$ from $\{\mathcal{R}_1, \cdots, \mathcal{R}_N\}$. So we can do with the power allocated on codes. ■

3.4.1 Asymptotic Performance

The main result of this subsection is that, assuming Rayleigh-fading, Theorem 3.2 also holds for $R_{\text{PD\&F}}$:

$$R_{\text{PD\&F}} - \mathcal{C}\left(2 \cdot W_0\left(\frac{\sqrt{N}}{2}\right) \cdot \frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{3.37}$$

That is, the rates with D&F and with two-level partial decoding are asymptotically equal.

In order to explain this result, let us briefly discuss the impact of $|a|^2$ on the achievable rate. On the one hand, it is clear that, when the source-destination channel $|a|$ is of similar magnitude than the source-relay channels $|b_m|$, the *two-level* approach provides a gain of almost $\mathcal{C}\left(|a|^2 \left(1-\beta\right) \frac{P}{N_o}\right)$ with respect to D&F. On the other hand, when the source-destination channel is much weaker than the source-relay channels (*i.e.* $|a|^2 \ll |b_m|^2$), it is possible to show that:

$$\mathcal{C}\left(|b_m|^2 \left(1-\beta\right) \frac{P}{N_o}\right) - \mathcal{C}\left(|a|^2 \left(1-\beta\right) \frac{P}{N_o}\right) \approx \mathcal{C}\left(|b_m|^2 \left(1-\beta\right) \frac{P}{N_o}\right). \tag{3.38}$$

Therefore, we can approximate:

$$\begin{aligned}
&\mathcal{C}\left(\frac{|b_m|^2\beta\eta P}{N_o+|b_m|^2(1-\beta)P}\right)+\mathcal{C}\left(|a|^2(1-\beta)\frac{P}{N_o}\right)= \\
&\quad\mathcal{C}\left(|b_m|^2(1-\beta+\beta\eta)\frac{P}{N_o}\right)-\mathcal{C}\left(|b_m|^2(1-\beta)\frac{P}{N_o}\right)+\mathcal{C}\left(|a|^2(1-\beta)\frac{P}{N_o}\right)\approx \\
&\quad\mathcal{C}\left(|b_m|^2(1-\beta+\beta\eta)\frac{P}{N_o}\right)-\mathcal{C}\left(|b_m|^2(1-\beta)\frac{P}{N_o}\right)= \\
&\quad\mathcal{C}\left(\frac{|b_m|^2\beta\eta P}{N_o+|b_m|^2(1-\beta)P}\right),
\end{aligned} \tag{3.39}$$

where we assumed $\eta \neq 0$. That is, the contribution of the direct link $|a|^2(1-\beta)P$ is negligible compared to the interference caused: $|b_m|^2(1-\beta)P$. Hence, considering (3.39), the rate optimization (3.33) remains:

$$\max_{0\leq(\eta,\beta)\leq 1}\min\left\{\mathcal{C}\left(\left(|a|^2+\beta(1-\eta)\sum_{i=1}^{m}|c_i|^2\right)\frac{P}{N_o}\right),\mathcal{C}\left(\frac{|b_m|^2\beta\eta P}{N_o+|b_m|^2(1-\beta)P}\right)\right\}$$

which is clearly attained at:

$$\begin{aligned}
\beta^* &= 0 \\
\eta^* &= \min\left\{1,\frac{|a|^2+\sum_{i=1}^{m}|c_i|^2}{|b_m|^2+\sum_{i=1}^{m}|c_i|^2}\right\},
\end{aligned}$$

yielding an identical rate as that of D&F. Accordingly, PD&F does not provide gain. Intuitively, this is easy to explain: the weaker the source-destination channel is with respect to source-relays channels, the less worthy is to transmit data directly to destination and the better is to transmit it via coherent beamforming among relays. Hence, relays do not partially but totally decode the source message.

Taking this into account, we can derive the asymptotic performance of $R_{\text{PD\&F}}$. Consider, as previously, unitary-mean, Rayleigh fading within the network: $a, b_i, c_i \sim \mathcal{CN}(0,1)$. For such a channel distribution, we showed in subsection 3.3.1, equation (3.26), that the ordered channel b_m can be approximated (for sufficiently large N) as

$$|b_m|^2 \approx -\log\left(\frac{m}{N}\right) \tag{3.40}$$

Hence, for small $m \ll N$, the source-relay channel satisfies $|b_m|^2 \gg |a|^2$ and partial decoding does not provide gain. Using this fact, it is demonstrated that Theorem 3.2 also holds for $R_{\text{PD\&F}}$. We omit the complete proof as it does not provide new arguments. Therefore, D&F and PD&F are asymptotically equivalent.

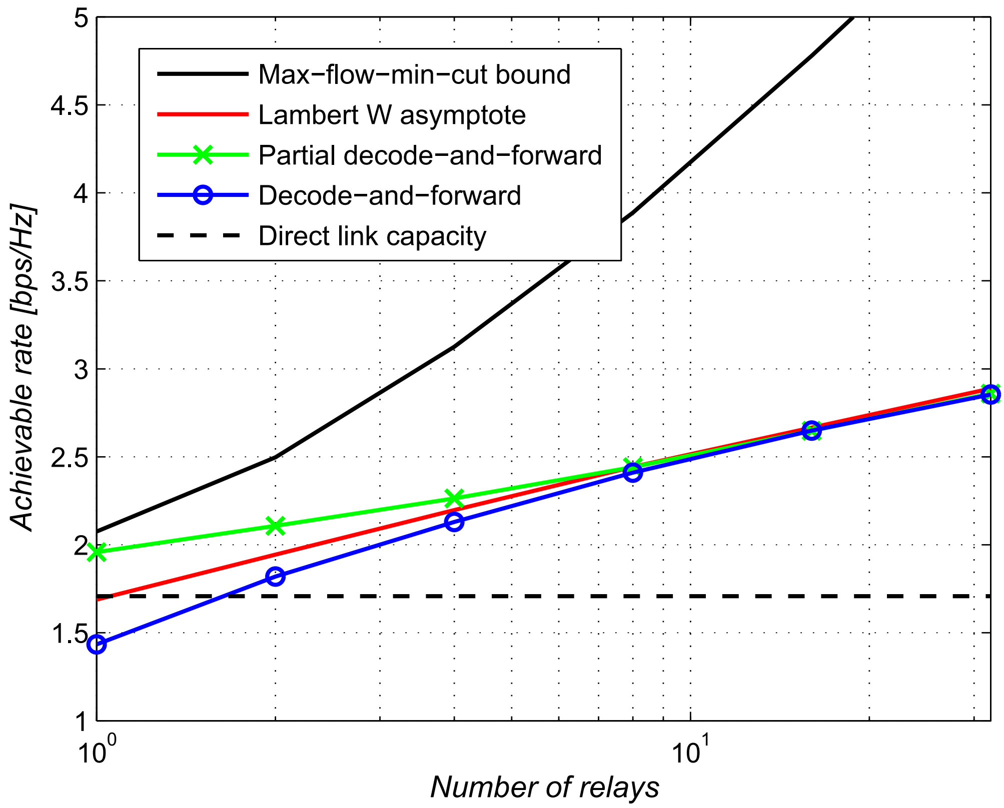

Figure 3.2: Comparison of the *max-flow-min-cut* bound and the achievable rates with D&F and PD&F. Transmit SNR = 5dB. Wireless channels are *i.i.d.*, unitary-power, Rayleigh distributed. The expected value of the achievable rate, averaged over the joint channel distribution, is shown.

Nonetheless, asymptotical equivalence does not mean that two-level PD&F and D&F perform always equal. As an example, both achievable rates are plotted in Fig. 3.2 versus the total number of relays, and compared with the *max-flow-min-cut* bound. We clearly notice that for low number of relays, PD&F indeed provides significant gains, being a very competitive approach for $N = 1$. However, for increasing number of relays both grow following the same Lambert W scaling law (as derived in Theorem 3.2). Hence, both diverge from the upper bound that scales as $\log_2 N$. This suggests that non-regenerative techniques shall be studied in order to improve performance at the large N regime.

3.5 Compress-and-Forward

Unlike previous techniques, with *compress-and-forward* (C&F) the relays mimic a receive antenna array [10]: they send their received signals $\boldsymbol{y}_i^n$, $i = 1, \cdots, N$ to the destination, which makes use of them (within a *coherent detector*) to estimate the source's message. Note that relays do not decode the source signal; hence, the broadcast limitation discussed earlier is eliminated. However, the communication is indeed constrained by the relay-destination channels: since they are of limited capacity, relays need to compress their signals before transmitting them to destination.

Operation is as follows. Let $\boldsymbol{y}_i^n$, $i = 1, \cdots, N$ be the observations at relays of the signal transmitted by the source, following (3.2). The relays perform two consecutive steps: first, they compress their observations, in a distributed manner and without cooperation among them. This is referred to as multi-source compression [62]. Later, they send the compressed signals towards the destination; to do so, they map them onto a multiple-access channel (MAC) code. In turn, the destination performs three consecutive tasks: first, it decodes the MAC code and estimates the compressed signals transmitted by the relays. Next, it decompresses the signals using its own received signal $\boldsymbol{y}_d^n$ as side information; and, finally, the destination coherently combines the de-compressed vectors, along with its own received signal, to estimate the source's message.

Notice that the proposed strategy imposes source-channel separation at the relays, which is not shown to be optimal [63]. However, it includes the coding scheme that, to the best of author knowledge, has largest known performance: Distributed Wyner-Ziv (D-WZ) coding [39]. Such an architecture is the direct extension of Berger-Tung coding to the decoder side information case [64]. In turn, Berger-Tung can be thought as the lossy counterpart of Slepian-Wolf lossless coding [65]. D-WZ is thus the coding scheme proposed to be used.

We start by defining the multiple-source compression code. Then, we present recent results on the minimum compression rates with D-WZ coding [39]. Finally, we derive the achievable rate of the MPRC with C&F at the relays.

Definition 3.4 (Multiple-source compression code) *A* $\left(n, 2^{n\phi_1}, \cdots, 2^{n\phi_N}\right)$ *compression code with side information* y_d *at the decoder is defined by* $N + 1$ *mappings,* $f_n^i(\cdot)$, $i = 1, \cdots, N$, *and* $g_n(\cdot)$, *and* $2N+1$ *spaces* $\mathcal{Y}_i, \hat{\mathcal{Y}}_i$, $i=1,\cdots,N$ *and* $\mathcal{Y}_d$, *where*

$$f_n^i : \mathcal{Y}_i^n \rightarrow \left\{1,\cdots,2^{n\phi_i}\right\}, \;\; i=1,\cdots,N$$
$$g_n : \left\{1,\cdots,2^{n\phi_1}\right\} \times \cdots \times \left\{1,\cdots,2^{n\phi_N}\right\} \times \mathcal{Y}_d^n \rightarrow \hat{\mathcal{Y}}_1^n \times \cdots \times \hat{\mathcal{Y}}_N^n.$$

Proposition 3.1 (Distributed Wyner-Ziv Coding [39]) *Consider* y_i *defined in (3.2), and let the random variables* $\hat{y}_i$, $i=1,\cdots,N$, *have conditional probability* $p\left(\hat{y}_i|y_i\right)$ *and satisfy the Markov chain* $\left(y_d, y_i^c, \hat{y}_i^c\right) \rightarrow y_i \rightarrow \hat{y}_i$*. Then, considering a sequence of compression codes* $\left(n, 2^{n\phi_1},\cdots,2^{n\phi_N}\right)$ *with side information* y_d *at the decoder, the following holds:*

$$\lim_{n\rightarrow\infty} \frac{1}{n} I\left(\boldsymbol{x}_s^n; \boldsymbol{y}_d^n, g_n\left(\boldsymbol{y}_d^n, f_n^1\left(\boldsymbol{y}_1^n\right),\cdots,f_n^N\left(\boldsymbol{y}_N^n\right)\right)\right) = I\left(x_s; y_d, \hat{y}_1,\cdots,\hat{y}_N\right) \tag{3.41}$$

if:

- *each compression codebook* $\mathfrak{C}_i$, $i=1,\cdots,N$ *consists of* $2^{n\phi_i}$ *random sequences* $\hat{\boldsymbol{y}}_i^n$ *drawn i.i.d. from* $\prod_{t=1}^n p\left(\hat{y}_i\right)$, *where* $p\left(\hat{y}_i\right) = \sum_{y_i} p\left(y_i\right) p\left(\hat{y}_i|y_i\right)$,
- *for every* $i=1,\cdots,N$, *the encoding* $f_n^i\left(\cdot\right)$ *outputs the bin-index of codewords* $\hat{\boldsymbol{y}}_i^n$ *that are jointly typical with the source sequence* $\boldsymbol{y}_i^n$*. In turn,* $g_n\left(\cdot\right)$ *outputs the codewords* $\hat{\boldsymbol{y}}_i^n$, $i=1,\cdots,N$ *that, belonging to the bins selected by the encoders, are all jointly typical with* $\boldsymbol{y}_d^n$,
- *the compression rates* $\phi_1,\cdots,\phi_N$ *satisfy*

$$I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \leq \sum_{i\in\mathcal{G}} \phi_i \qquad \forall \mathcal{G} \subseteq \left\{1,\cdots,N\right\}. \tag{3.42}$$

Proof: The proposition is proven for discrete sources and discrete side information in [39, Theorem 2]. Also, the extension to the Gaussian case is conjectured therein. The conjecture can be proven by noting that D-WZ coding is equivalent to Berger-Tung coding with side information at the decoder [64]. In turn, Berger-Tung coding can be implemented through time-sharing of successive Wyner-Ziv compressions [66], for which introducing side information y_d at the decoder reduces the compression rate as in (3.42). ∎

We are now able to derive the achievable rate with such a compression scheme at the relays.

Theorem 3.4 *The AWGN multiple-parallel relay channel achieves the rate*

$$R_{\mathrm{C\&F}} = \max_{\gamma_s, \boldsymbol{\gamma}_{1:N}, \boldsymbol{\rho}_{1:N}} \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i}\right)\gamma_s\right) \tag{3.43}$$

$$\text{s.t. } \mathcal{C}\left(\frac{\gamma_s \sum_{u \in \mathcal{G}} \frac{|b_u|^2}{N_o + \rho_u}}{1 + \gamma_s\left(\frac{|a|^2}{N_o} + \sum_{j \notin \mathcal{G}} \frac{|b_j|^2}{N_o + \rho_j}\right)}\right) + \sum_{u \in \mathcal{G}} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) \leq \mathcal{C}\left(\sum_{u \in \mathcal{G}} \frac{|c_u|^2 \gamma_u}{N_o + |a|^2 \gamma_s}\right), \forall \mathcal{G} \subseteq \mathcal{N}.$$

$$\gamma_s + \sum_{i=1}^{N} \gamma_i \leq P$$

with D-WZ compress-and-forward relaying.

Remark 3.3 *In the maximization, ρ_i and γ_i, $i = 1, \cdots, N$ stand for the compression noise and power allocated to relay $i = 1, \cdots, N$, respectively. In turn, γ_s denotes the power allocated to the source. Clearly, the optimization is not convex, and needs to be solved using exhaustive search. The search is, unfortunately, unfeasible for large N.*

Proof: Let the source select message $\omega \in \left\{1, \cdots, 2^{nR}\right\}$ for transmission, and divide it into B sub-messages of κR bits each, with $\kappa = \frac{n}{B}$: $\omega = \left[\omega^1, \cdots, \omega^B\right]$. The sub-messages are then pipelined onto $B + 1$ channel blocks, of κ channel uses each, and transmitted using block-Markov encoding as in [21, Sec. 14.7]. First, on every block b, the source transmits sub-message ω^b to relays and destination

$$\boldsymbol{x}_s^\kappa [b] = \boldsymbol{s}^\kappa \left(\omega^b\right), \tag{3.44}$$

where we have selected $\boldsymbol{s}^\kappa(\cdot)$ to be a random Gaussian codebook, generated *i.i.d.* from $s \sim \mathcal{CN}(0, \gamma_s)$. The signal is received at the relays following (3.2):

$$\boldsymbol{y}_i^\kappa[b] = b_i \boldsymbol{s}^\kappa\left(\omega^b\right) + \boldsymbol{z}_i^\kappa, \;\; i = 1, \cdots, N. \tag{3.45}$$

The signals are then distributedly compressed at the relays using a compression code as that in Proposition 3.1. That is, each relay maps $\boldsymbol{y}_i^\kappa[b]$ using functions

$$f_\kappa^i : \mathcal{Y}_i^\kappa \rightarrow \left\{1, \cdots, 2^{\kappa \phi_i}\right\}, \tag{3.46}$$

where ϕ_i is the compression rate of relay i. On the next block $b+1$, the relays send to the destination (via the MAC channel) the indexes $s_i[b] = f_\kappa^i\left(\boldsymbol{y}_i^\kappa[b]\right), i = 1, \cdots, N$. To do so, they map them onto multi-access channel codebooks $\boldsymbol{v}_i^\kappa(\cdot), i = 1, \cdots, N$:

$$\boldsymbol{x}_i^\kappa[b+1] = \boldsymbol{v}_i^\kappa\left(s_i[b]\right), \; i = 1, \cdots, N. \tag{3.47}$$

Gaussian codebooks are well known to be optimal for the AWGN MAC. In this case, we generate them *i.i.d.* from $v_i \sim \mathcal{CN}(0, \gamma_i), i = 1, \cdots, N$. The received signal at the destination on this block thus follows (3.3):

$$\boldsymbol{y}_d^\kappa[b+1] = a\boldsymbol{s}^\kappa\left(\omega^{b+1}\right) + \sum_{i=1}^N c_i \boldsymbol{v}_i^\kappa\left(s_i[b]\right) + \boldsymbol{z}_d^\kappa. \tag{3.48}$$

Let us now define the decoding at the destination in block $b+1$. First, it recovers indexes $s_{1:N}[b]$ from its received signal $\boldsymbol{y}_d^\kappa[b+1]$.The destination can do so *iff* transmission rates ϕ_i lie within the capacity region of its MAC channel (being $\boldsymbol{s}^\kappa\left(\omega^{b+1}\right)$ interference):

$$\sum_{u\in\mathcal{G}} \phi_u \leq \mathcal{C}\left(\frac{\sum_{u\in\mathcal{G}}|c_u|^2\gamma_u}{N_o + |a|^2\gamma_s}\right), \forall \mathcal{G} \subseteq \mathcal{N}. \tag{3.49}$$

Once the indexes $s_{1:N}[b]$ has been estimated, destination removes their contribution on $\boldsymbol{y}_d^\kappa[b+1]$:

$$\begin{aligned} \boldsymbol{y}_d^{'\kappa}[b+1] &= \boldsymbol{y}_d^\kappa[b+1] - \sum_{i=1}^N c_i \cdot \boldsymbol{x}_i^\kappa[b+1] \\ &= a\boldsymbol{s}^\kappa\left(\omega^{b+1}\right) + \boldsymbol{z}_d^\kappa. \end{aligned} \tag{3.50}$$

After that, the destination decompresses indexes $s_{1:N}[b]$ using its received signal $\boldsymbol{y}_d^{'\kappa}[b]$ as side information. Following Proposition 3.1, the indexes are then de-mapped by means of function

$$g_\kappa : \left\{1, \cdots, 2^{\kappa\phi_1}\right\} \times \cdots \times \left\{1, \cdots, 2^{\kappa\phi_N}\right\} \times \mathcal{Y}_d^{'\kappa} \rightarrow \hat{\mathcal{Y}}_{1:N}^\kappa. \tag{3.51}$$

Finally, the de-mapped vectors $\hat{\boldsymbol{y}}_{1:N}^\kappa[b] = g_\kappa\left(s_1[b], \cdots, s_N[b], \boldsymbol{y}_d^{'\kappa}[b]\right)$ are used, along with $\boldsymbol{y}_d^{'\kappa}[b]$, to decode message ω^b, which is correctly estimated *iff* [38, 67]:

$$\begin{aligned} R &\leq \lim_{\kappa\to\infty} \frac{1}{\kappa} I\left(\boldsymbol{x}_s^\kappa; \boldsymbol{y}_d^{'\kappa}, g_\kappa\left(f_\kappa^1\left(\boldsymbol{y}_1^\kappa\right), \cdots, f_\kappa^N\left(\boldsymbol{y}_N^\kappa\right), \boldsymbol{y}_d^{'\kappa}\right)\right) \\ &= I\left(x_s; y_d^{'}, \hat{y}_1, \cdots, \hat{y}_N\right). \end{aligned} \tag{3.52}$$

The second equality follows from (3.41) in Proposition 3.1. However, equality only holds for compression rates satisfying the set of constraints (3.42):

$$I\left(\boldsymbol{y}_\mathcal{G}; \hat{\boldsymbol{y}}_\mathcal{G} | y_d^{'}, \hat{\boldsymbol{y}}_\mathcal{G}^c\right) \leq \sum_{u\in\mathcal{G}} \phi_u \qquad \forall \mathcal{G} \subseteq \mathcal{N}. \tag{3.53}$$

Therefore, taking into account the constraints (3.53) and (3.49) and the rate (3.52), the maximum achievable rate R with C&F is:

$$\begin{aligned} R_{\text{C\&F}} &= \max_{\prod_{i=1}^{N} p(\hat{y}_i|y_i), \gamma_s, \gamma_{1:N},} I\left(x_s; y_d^{'}, \hat{\boldsymbol{y}}_{1:N}\right) \\ &\quad \text{s.t. } I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \leq \mathcal{C}\left(\frac{\sum_{u\in\mathcal{G}} |c_u|^2\gamma_u}{N_o + |a|^2\gamma_s}\right), \forall \mathcal{G} \subseteq \mathcal{U}, \\ &\quad \gamma_s + \sum_{i=1}^{N} \gamma_i \leq P, \end{aligned} \tag{3.54}$$

where the last constraint accounts for the power constraint (3.4). As mentioned, the source signal is Gaussian $x_s \sim \mathcal{CN}(0, \gamma_s)$. Despite that, the solution of maximization in (3.54) is still an open problem, which has remained unsolvable for us. We propose, thus, a suboptimum approach, which is $p(\hat{y}_i|y_i)$ $i = 1, \cdots, N$ to be Gaussian too: $p(\hat{y}_i|y_i) = \frac{1}{\sqrt{\pi\rho_i}}\exp\left(-\frac{|\hat{y}_i - y_i|^2}{\rho_i}\right)$, where ρ_i is hereafter referred to as compression noise. With such a codebook distribution, we can derive:

$$I\left(x_s; y_d^{'}, \hat{\boldsymbol{y}}_{1:N}\right) = \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i}\right)\gamma_s\right). \tag{3.55}$$

Furthermore, it is possible to evaluate:

$$\begin{aligned} I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) &= H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) - H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|\boldsymbol{y}_{\mathcal{G}}, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \\ &= I\left(x_s; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) + H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|x_s, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) - H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|\boldsymbol{y}_{\mathcal{G}}, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \\ &= I\left(x_s; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) + H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|x_s, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) - H\left(\hat{\boldsymbol{y}}_{\mathcal{G}}|x_s, \boldsymbol{y}_{\mathcal{G}}, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \\ &= I\left(x_s; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) + I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|x_s, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \\ &= I\left(x_s; y_d^{'}, \hat{\boldsymbol{y}}_{1:N}\right) - I\left(x_s; y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) + I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|x_s, y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \\ &= \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{u=1}^{N} \frac{|b_u|^2}{N_o + \rho_u}\right)\gamma_s\right) - \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{u\in\mathcal{G}^c} \frac{|b_u|^2}{N_o + \rho_u}\right)\gamma_s\right) \\ &\quad + \sum_{u\in\mathcal{G}} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) \\ &= \mathcal{C}\left(\frac{\gamma_s \sum_{u\in\mathcal{G}} \frac{|b_u|^2}{N_o + \rho_u}}{1 + \gamma_s\left(\frac{|a|^2}{N_o} + \sum_{j\notin\mathcal{G}} \frac{|b_j|^2}{N_o + \rho_j}\right)}\right) + \sum_{u\in\mathcal{G}} \mathcal{C}\left(\frac{N_o}{\rho_u}\right). \end{aligned} \tag{3.56}$$

where the second equality follows from the definition of mutual information and the third from the Markov chain in Proposition 3.1. Finally, the fifth equality comes from the chain rule for mutual information. Plugging (3.55) and (3.56) into (3.54), and optimizing over $\boldsymbol{\rho}_{1:N}$ concludes the proof. ∎

As mentioned, the maximum achievable rate with C&F (3.43) is the solution of a non-convex optimization problem with $1 + \sum_{i=1}^{N} \binom{N}{i}$ constraints. Its computation, thus, requires non-convex algorithms that (for large N) become unfeasible. The next subsection proposes an upper bound on the C&F achievable rate that can be easily computed using known approaches in the literature. This will be used as a benchmark for C&F.

3.5.1 Upper Bound

In Theorem 3.4, the relay compression rates have been constrained to lie within the MAC capacity region of the destination, thus forced to satisfy $\sum_{i=1}^{N} \binom{N}{i}$ constraints. In order to relax the problem and provide a computable upper bound, we consider the compression rates satisfying the sum-rate constraint only. This model applies to scenarios where relays are not connected to the destination via a wireless medium but via *e.g.,* common wired Ethernet [51]. Let us then eliminate $\sum_{i=1}^{N} \binom{N}{i} - 1$ constraints of (3.43) to define:

$$\bar{R}_{\text{C\&F}} = \max_{\gamma_s, \boldsymbol{\gamma}_{1:N}, \boldsymbol{\rho}_{1:N}} \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i}\right)\gamma_s\right). \tag{3.57}$$

$$\text{s.t. } \mathcal{C}\left(\frac{\gamma_s \sum_{u=1}^{N} \frac{|b_u|^2}{N_o+\rho_u}}{1 + \gamma_s \frac{|a|^2}{N_o}}\right) + \sum_{u=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) \leq \qquad \gamma_s + \sum_{i=1}^{N} \gamma_i \leq P$$

Clearly, $\bar{R}_{\text{C\&F}}$ is an upper bound on $R_{\text{C\&F}}$. Moreover, notice that optimization variables $\gamma_1, \cdots, \gamma_N$ only appear at the right hand side of the first constraint, for which is easy to show:

$$\max_{\gamma_1, \cdots, \gamma_N} \sum_{u=1}^{N} |c_u|^2 \gamma_u = (1-\beta)\, P \cdot \max\ |c_u|^2 \tag{3.58}$$

$$\text{s.t} \sum_{u=1}^{N} \gamma_u \leq (1-\beta)\, P$$

Let us then define $j = \arg\max |c_u|^2$. Considering (3.58) into the maximization in (3.57), it is clear that the optimum point of the latter $\gamma_s *, \boldsymbol{\gamma}_{1:N}^*, \boldsymbol{\rho}_{1:N}^*$ satisfies $\gamma_i^* = 0, \forall i \neq j$. Accordingly, the following equality holds:

$$\bar{R}_{\text{C\&F}} = \max_{\beta, \boldsymbol{\rho}_{1:N}} \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i}\right)\beta P\right) \tag{3.59}$$

$$\text{s.t. } \mathcal{C}\left(\frac{\beta P \sum_{u=1}^{N} \frac{|b_u|^2}{N_o+\rho_u}}{1 + \beta P \frac{|a|^2}{N_o}}\right) + \sum_{u=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) \leq \mathcal{C}\left(\frac{|c_j|^2 (1-\beta)\, P}{N_o + |a|^2 \beta P}\right), \quad 0 < \beta \leq 1$$

We have then reduced the problem in (3.57) to a simpler computable one. Now, the optimization over $\boldsymbol{\rho}_{1:N}$, for a fixed value of β, can be solved using a *dual decomposition approach*. Afterwards, the maximization over β can be carried out using a simple one-dimensional exhaustive search. The algorithm is described in the next subsection.

Iterative Algorithm to Compute the Upper Bound

Let us rewrite $\bar{R}_{\text{C\&F}} = \max_{\beta\in(0,1]} r(\beta)$, where:

$$\begin{aligned} r(\beta) &= \max_{\boldsymbol{\rho}_{1:N}} \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o+\rho_i}\right)\beta P\right) \\ & \quad \text{s.t. } \mathcal{C}\left(\frac{\beta P \sum_{u=1}^{N} \frac{|b_u|^2}{N_o+\rho_u}}{1+\beta P\frac{|a|^2}{N_o}}\right) + \sum_{u=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) \leq \Phi_T, \end{aligned} \tag{3.60}$$

with $\Phi_T = \mathcal{C}\left(\frac{|c_j|^2(1-\beta)P}{N_o+|a|^2\beta P}\right)$. We first notice that maximization (3.60) is not concave in standard form: although the constraint defines a convex, regular set, the objective function is not concave but convex. Therefore, KKT conditions become necessary but not sufficient for optimality.

In order to solve the optimization, we need to resort to other strategies. In particular, we proceed as follows: first, we show that the duality gap for the problem is zero. Afterwards, we propose an iterative algorithm that solves the dual problem, thus solving the primal problem too. As mentioned in Chapter 2, the interesting property of the dual problem is that the coupling constraint in (3.60) is decoupled.

First of all, let us rewrite the objective function of (3.60) as:

$$\mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o+\rho_i}\right)\beta P\right) = \mathcal{C}\left(\frac{|a|^2}{N_o}\beta P\right) + \mathcal{C}\left(\frac{\beta P \sum_{u=1}^{N} \frac{|b_u|^2}{N_o+\rho_u}}{1+\beta P\frac{|a|^2}{N_o}}\right). \tag{3.61}$$

Therefore, the Lagrangian of (3.60) can be thus defined on $\boldsymbol{\rho}_{1:N} \geq 0$ and $\lambda \geq 0$ as:

$$\mathcal{L}(\rho_1, \cdots, \rho_N, \lambda) = (1-\lambda)\,\mathcal{C}\left(\frac{\beta P \sum_{u=1}^{N} \frac{|b_u|^2}{N_o+\rho_u}}{1+\beta P\frac{|a|^2}{N_o}}\right) - \lambda\left(\sum_{u=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_u}\right) - \Phi_T\right). \tag{3.62}$$

The dual function $g(\lambda)$ is, thus, constructed for $\lambda \geq 0$ following subsection 2.2.2:

$$g(\lambda) = \max_{\rho_1,\cdots,\rho_N \geq 0} \mathcal{L}(\rho_1, \cdots, \rho_N, \lambda). \tag{3.63}$$

Finally, the solution of the dual problem is then obtained from

$$\mathcal{C}' = \min_{\lambda \geq 0} g(\lambda). \tag{3.64}$$

Lemma 3.1 *The duality gap for optimization (3.60) is zero, i.e., the primal problem (3.60) and the dual problem (3.64) have the same solution.*

Proof: As derived in Proposition 2.10, the duality gap for problems of the form of (3.60), and satisfying the time-sharing property, is zero [53, Theorem 1]. Time-sharing property is satisfied if the solution of (3.60) is concave with respect to the compression sum-rate Φ_T. It is well known that time-sharing of compressions can neither decrease the resulting distortion [21, Lemma 13.4.1] nor improve the mutual information obtained from the reconstructed vectors. Hence, the property holds for (3.60), and the duality gap is zero. ■

We then solve the dual problem in order to obtain the solution of the primal. First, consider maximization (3.63). As expected, the maximization can not be solved in closed form. However, as the feasible set (*i.e.*, $\rho_1, \cdots, \rho_N \geq 0$) is the cartesian product of convex sets, then a *block coordinate ascent* algorithm can be used to search for the maximum (see subsection 2.2.1). The algorithm iteratively optimizes the function with respect to one ρ_n while keeping the others fixed, and it is defined for our problem as:

$$\rho_n^{t+1} = \arg\max_{\rho_n \geq 0} \mathcal{L}\left(\rho_1^{t+1}, \cdots, \rho_{n-1}^{t+1}, \rho_n, \rho_{n+1}^{t}, \cdots, \rho_N^{t}, \lambda\right), \tag{3.65}$$

where t is the iteration index. As shown in Proposition 3.2, the maximization (3.65) is uniquely attained.

Proposition 3.2 *Let the optimization* $\rho_n^* = \arg\max_{\rho_n \geq 0} \mathcal{L}\left(\rho_1, \cdots, \rho_N, \lambda\right)$ *and define*

$$s = \frac{|b_n|^2 \beta P}{1 + \left(\frac{|a|^2}{N_o} + \sum_{i \neq n} \frac{|b_i|^2}{N_o + \rho_i}\right) \beta P} + N_o. \tag{3.66}$$

The optimization is uniquely attained at

$$\rho_n^* = \left(\left[\frac{1}{\lambda}\left(\frac{1}{N_o} - \frac{1}{s}\right) - \frac{1}{N_o}\right]^+\right)^{-1}. \tag{3.67}$$

Remark 3.4 *Within the proposition, we consider the definition* $\frac{1}{0} \triangleq \infty$.

Proof: First, let us expand the function under maximization as:

$$
\begin{aligned}
f(\rho_n) &= \mathcal{L}(\rho_1, \cdots, \rho_N, \lambda) \\
&= (1-\lambda)\,\mathcal{C}\left(\frac{\beta P \sum_{i\neq n} \frac{|b_i|^2}{N_o+\rho_i}}{1+\beta P \frac{|a|^2}{N_o}}\right) - \lambda\left(\sum_{i\neq n} \mathcal{C}\left(\frac{N_o}{\rho_i}\right) - \Phi_T\right) \\
&\quad + (1-\lambda)\,\mathcal{C}\left(\frac{s-N_o}{N_o+\rho_n}\right) - \lambda\mathcal{C}\left(\frac{N_o}{\rho_n}\right).
\end{aligned}
\tag{3.68}
$$

Clearly, only the last two terms depend on ρ_n. Therefore, it is enough to maximize the function:

$$
\tilde{f}(\rho_n) = (1-\lambda)\,\mathcal{C}\left(\frac{s-N_o}{N_o+\rho_n}\right) - \lambda\mathcal{C}\left(\frac{N_o}{\rho_n}\right).
$$

which is continuous, but neither concave nor convex. However, for $\lambda \geq 1$, it is clear that $\tilde{f}(\rho_n) \leq 0, \forall \rho_n$, and therefore $\rho_n^* = \infty$.

On the other hand, for $\lambda < 1$, we can change variables $\rho_n = a_n^{-1} \geq 0$ and transform:

$$
\begin{aligned}
\tilde{f}(a_n) &= (1-\lambda)\,\mathcal{C}\left(a_n \frac{s-N_o}{a_n N_o + 1}\right) - \lambda\mathcal{C}(a_n N_o) \\
&= (1-\lambda)\,\mathcal{C}(a_n s) - \mathcal{C}(a_n N_o).
\end{aligned}
\tag{3.69}
$$

For which is can be shown that there is no more than one stationary point:

$$
\frac{d\tilde{f}}{da_n} = 0 \rightarrow a_n^* = \frac{1}{\lambda}\left(\frac{1}{N_o} - \frac{1}{s}\right) - \frac{1}{N_o}. \tag{3.70}
$$

For the stationary point, we can prove that its second derivative exists and is lower than zero; accordingly, it is a local maximum of the function, unique because there is no other. Moreover, it is easy to obtain that: *i*) $\tilde{f}(0) = 0$, and *ii*) since $\lambda < 1$, then $\lim_{a_n\rightarrow\infty} \tilde{f}(a_n) = -\infty$. That is, $a_n = \infty$ is the global minimum of the problem. Making use of *i*) and *ii*), we can claim that the local maximum a_n^* is the global maximum. However, we restricted the optimization to the values $a_n \geq 0$. Hence, function $\tilde{f}(a_n)$ takes maximum at:

$$
a_n^* = \left[\frac{1}{\lambda^*}\left(\frac{1}{N_o} - \frac{1}{s}\right) - \frac{1}{N_o}\right]^+. \tag{3.71}
$$

Finally, we change variables again to recover ρ_n^* as (3.67) which concludes the proof. Notice that for $\lambda \geq 1$, (3.67) evaluates as $\eta_n^* = \infty$ which was the optimum value derived above. Therefore the solution is valid for all λ. ■

The function $\mathcal{L}(\rho_1, \cdots, \rho_N, \lambda)$ is continuously differentiable, and the maximization (3.65) is uniquely attained. Hence, the limit point of the sequence $\{\rho_1^t, \cdots, \rho_N^t\}$ is proven to converge to a stationary point of $\mathcal{L}(\cdot)$ (see Proposition 2.6). However, in order to demonstrate convergence to the global maximum (that is, to $g(\lambda)$), it is necessary to show that the mapping $T(\rho_1, \cdots, \rho_N) = \left[\rho_1 + \gamma \frac{\partial \mathcal{L}}{\partial \rho_1}, \cdots, \rho_N + \gamma \frac{\partial \mathcal{L}}{\partial \rho_N}\right]$ is a block-contraction for some γ (Proposition 2.7). Unfortunately, we were not able to demonstrate the contraction property on the Lagrangian, although simulation results suggest global convergence of our algorithm always.

Once $g(\lambda)$ is obtained through the Gauss-Seidel Algorithm[3], it must be minimized on $\lambda \geq 0$. First, recall that $g(\lambda)$ is a convex function, defined as the pointwise maximum of a family of affine functions [52]. Hence, to minimize it, we may use a subgradient approach [68]. As mentioned in Section 2.2.1, the subgradient search consists on following search direction $-h(\lambda)$ such that

$$\frac{g(\lambda') - g(\lambda)}{\lambda' - \lambda} \geq h(\lambda) \qquad \forall \lambda'. \tag{3.72}$$

Such a search is proven to converge to the global minimum for diminishing step-size rules [69, Section II-B]. Moreover, given the analysis in Section 2.2.2, the following $h(\lambda)$ satisfies (3.72):

$$h(\lambda) = \Phi_T - \mathcal{C}\left(\frac{\beta P \sum_{u=1}^{N} \frac{|b_u|^2}{N_o + \rho_u(\lambda)}}{1 + \beta P \frac{|a|^2}{N_o}}\right) - \sum_{u=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_u(\lambda)}\right). \tag{3.73}$$

where $\boldsymbol{\rho}_{1:N}(\lambda)$ is the limiting point of algorithm (3.65) for λ. This is used to search for the optimum λ as:

$$\text{increase } \lambda \text{ if } \ h(\lambda) \leq 0 \ \text{ or } \ \text{decrease } \lambda \text{ if } \ h(\lambda) \geq 0. \tag{3.74}$$

Consider now $\lambda^0 = 1$ as the initial value of the Lagrange multiplier. For such a multiplier, it is easy to show that the optimum solution of (3.63) is $\{\rho_1^*. \cdots, \rho_N^*\} = \infty$. Hence, the subgradient (3.73) is $h(\lambda) = \Phi_T$. Following (3.74), the optimum value of λ is then strictly lower than one. Algorithm 1 takes all this into account in order to solve the dual problem, hence solving the primal too. Recall that we can only claim convergence of the algorithm to a stationary point.

3.5.2 Asymptotic Performance

The C&F achievable rate (3.43) is, as pointed out previously, a computationally untractable optimization problem for $N \to \infty$. Deriving, thus, its asymptotic performance in closed form

[3] Assume hereafter that the algorithm has converged to the global maximum of $\mathcal{L}(\rho_1, \cdots, \rho_N, \lambda)$.

Algorithm 1 Computation of the Upper Bound

for $\beta = 0 : 1$ **do**
  Define $\Phi_T = \mathcal{C}\left(\frac{|c_j|^2(1-\beta)P}{N_o+|a|^2\beta P}\right)$
  Initialize $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$
  repeat
    $\lambda = \frac{\lambda_{\max}-\lambda_{\min}}{2}$
    Obtain $\{\rho_1(\lambda), \cdots, \rho_N(\lambda)\} = \arg\max_{\rho_{1:N}} \mathcal{L}(\rho_1, \cdots, \rho_N, \lambda)$ from Algorithm 2
    Evaluate $h(\lambda)$ as in (3.73).
    If $h(\lambda) \le 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
  until $\lambda_{\max} - \lambda_{\min} \le \epsilon$
  $r(\beta) = \mathcal{L}(\rho_1(\lambda), \cdots, \rho_N(\lambda), \lambda)$
end for
$\bar{R}_{\text{C\&F}} = \max\ r(\beta)$

is mathematically unfeasible. However, it is indeed possible to use the upper bound in the preceding subsection to bound the asymptotic performance of the achievable rate.

In particular, assuming unitary-mean, Rayleigh fading within the network: $a, b_i, c_i \sim \mathcal{CN}(0,1)$, the following convergence in probability can be proven.

Theorem 3.5 *Let $R_{\text{C\&F}}$ be defined in (3.43), and consider $\hat{R}_{\text{C\&F}} = \mathcal{C}\left(\max_{i=1\cdots,N}\{|a|^2, |c_i|^2\}\frac{P}{N_o}\right)$. Then:*

$$R_{\text{C\&F}} \le \hat{R}_{\text{C\&F}}, \tag{3.75}$$

and for $N \to \infty$:

$$\hat{R}_{\text{C\&F}} - \mathcal{C}\left(\log(N+1)\frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{3.76}$$

Remark 3.5 *This result demonstrates that the achievable rate with C&F is always upper bounded by the sum-capacity of the destination's MAC, given the sum-power constraint among transmitters.*

Proof: The first part of the proof relies directly on the optimization in (3.59). Let $\beta^*, \rho^*_{1:N}$ be the optimum values for this optimization, and recall that $j = \arg\max |c_u|^2$. Then, we can

Algorithm 2 *Block Coordinate Algorithm* to Obtain $g(\lambda)$

Initialize $\rho_n^0 = 0, n = 1, \cdots, N$ and $t = 0$

repeat

  for $n = 1$ to N **do**

    Compute

$$s = \frac{|b_n|^2 \beta P}{1 + \left(\frac{|a|^2}{N_o} + \sum_{i<n} \frac{|b_i|^2}{N_o + \rho_i^{t+1}} + \sum_{i>n} \frac{|b_i|^2}{N_o + \rho_i^t}\right) \beta P} + N_o.$$

    Update $\rho_n^{t+1} = 1 / \left[\frac{1}{\lambda}\left(\frac{1}{N_o} - \frac{1}{s}\right) - \frac{1}{N_o}\right]^+$.

  end for

  $t = t + 1$

until The sequence converges $\{\rho_1^t, \cdots, \rho_N^t\} \to \{\rho_1^*, \cdots, \rho_N^*\}$

Return $\{\rho_1^*, \cdots, \rho_N^*\}$

show that:

$$\begin{aligned} R_{\text{C\&F}} &\leq \bar{R}_{\text{C\&F}} \\ &= \mathcal{C}\left(\left(\frac{|a|^2}{N_o} + \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i^*}\right) \beta^* P\right) \\ &= \mathcal{C}\left(\frac{|a|^2 \beta^* P}{N_o}\right) + \mathcal{C}\left(\frac{\beta^* P \sum_{i=1}^{N} \frac{|b_i|^2}{N_o + \rho_i^*}}{1 + \frac{|a|^2}{N_o} \beta^* P}\right) \\ &\leq \mathcal{C}\left(\frac{|a|^2 \beta^* P}{N_o}\right) + \mathcal{C}\left(\frac{|c_j|^2 (1 - \beta^*) P}{N_o + |a|^2 \beta^* P}\right) \\ &= \mathcal{C}\left(\frac{|a|^2 \beta^* P}{N_o} + \frac{|c_j|^2 (1 - \beta^*) P}{N_o}\right) \\ &\leq \mathcal{C}\left(\frac{\max_{u=1,\cdots,N} \{|a|^2, |c_u|^2\} P}{N_o}\right), \end{aligned} \tag{3.77}$$

where the second inequality follows from the fact that, at the optimum β^*, $\rho_{1:N}^*$, the constraint of (3.59) is satisfied. The third inequality is explained by noting that $\beta \in (0, 1]$. Furthermore, we repeatedly use the equality $\mathcal{C}(a + b) = \mathcal{C}(a) + \mathcal{C}\left(\frac{b}{1+a}\right)$. This proves (3.75). Now, we need to prove (3.76). This can be done by first proving the following convergence in probability:

$$\lim_{N \to \infty} \Pr\left\{\left|\frac{\max_{u=1,\cdots,N} \{|a|^2, |c_u|^2\}}{\log(N+1)} - 1\right| < \epsilon\right\} = 1, \ \ \forall \epsilon > 0. \tag{3.78}$$

To prove the limit, we first expand:

$$\Pr\left\{\left|\frac{\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{\log\left(N+1\right)}-1\right|<\epsilon\right\}= \qquad (3.79)$$
$$\Pr\left\{\left[\left(1-\epsilon\right)\log\left(N+1\right)\right]^+<\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}<\left(1+\epsilon\right)\log\left(N+1\right)\right\}$$

Notice that $a, c_1, \cdots, c_N$ are independent, unitary mean, Rayleigh distributed. Therefore, all share the same cdf, $F(x) = 1 - e^{-x},\ x \geq 0$, which allows to compute the *c.d.f.* of the maximum as $\Pr\left\{\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\} \leq x\right\} = \left(1-e^{-x}\right)^{N+1}$. Accordingly,

$$\Pr\left\{\left|\frac{\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{\log\left(N+1\right)}-1\right|<\epsilon\right\}= \qquad (3.80)$$
$$\left(1-e^{-(1+\epsilon)\log(N+1)}\right)^{N+1}-\left(1-e^{-\left[(1-\epsilon)\log(N+1)\right]^+}\right)^{N+1}.$$

Now, it is possible to show that

$$\left(1-e^{-(1+\epsilon)\log(N+1)}\right)^{N+1} = \left(1-\left(\frac{1}{N+1}\right)^{1+\epsilon}\right)^{N+1}\stackrel{\epsilon>0,N\to\infty}{\rightarrow} 1 \qquad (3.81)$$

$$\left(1-e^{-\left[(1-\epsilon)\log(N+1)\right]^+}\right)^{N+1} = \begin{cases}\left(1-\left(\frac{1}{N+1}\right)^{1-\epsilon}\right)^{N+1}\stackrel{N\to\infty}{\rightarrow}0 & \epsilon\in(0,1)\\ 0 & \epsilon\geq 1\end{cases} \qquad (3.82)$$

Therefore, plugging (3.81) and (3.82) into (3.80) demonstrates that (3.78) holds. That is,

$$\frac{\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{\log\left(N+1\right)}-1\stackrel{P}{\rightarrow}0. \qquad (3.83)$$

Now, let us upper bound:

$$\begin{aligned}
&\Pr\left\{\left|\frac{\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{\log\left(N+1\right)}-1\right|<\epsilon\right\} \qquad (3.84)\\
&=\Pr\left\{\left|\frac{\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{\frac{P}{N_o}\log\left(N+1\right)}-1\right|<\epsilon\right\}\\
&=\Pr\left\{\left|\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}-\frac{P}{N_o}\log\left(N+1\right)\right|<\epsilon\cdot\frac{P}{N_o}\log\left(N+1\right)\right\}\\
&=\Pr\left\{\left|1+\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}-\left(1+\frac{P}{N_o}\log\left(N+1\right)\right)\right|<\epsilon\cdot\frac{P}{N_o}\log\left(N+1\right)\right\}\\
&\leq\Pr\left\{\left|1+\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}-\left(1+\frac{P}{N_o}\log\left(N+1\right)\right)\right|<\epsilon\cdot\left(1+\frac{P}{N_o}\log\left(N+1\right)\right)\right\}\\
&=\Pr\left\{\left|\frac{1+\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{1+\frac{P}{N_o}\log\left(N+1\right)}-1\right|<\epsilon\right\},
\end{aligned}$$

where the inequality comes from the fact that, augmenting the interval, increases the probability. Therefore, given (3.78) and inequality (3.84), the following convergence is proven

$$\frac{1+\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{1+\frac{P}{N_o}\log\left(N+1\right)}-1\overset{P}{\to}0. \tag{3.85}$$

Finally, making use of Lemma 2.1, we thus claim that

$$\log_2\frac{1+\frac{P}{N_o}\max_{u=1,\cdots,N}\left\{|a|^2,|c_u|^2\right\}}{1+\frac{P}{N_o}\log\left(N+1\right)}-\log_2 1\overset{P}{\to}0. \tag{3.86}$$

Or, in other words,

$$\mathcal{C}\left(\max_{u=1\cdots,N}\left\{|a|^2,|c_u|^2\right\}\frac{P}{N_o}\right)-\mathcal{C}\left(\log\left(N+1\right)\frac{P}{N_o}\right)\overset{P}{\to}0, \tag{3.87}$$

which concludes the proof. ■

As mentioned, Theorem 3.5 demonstrates that *compress-and-forward* is limited by the relays-destination MAC. This has a straightforward cause: after the compression step, the signals transmitted by the relays are independent and uncorrelated. Therefore, among them, no coherent transmission is possible, and they all have to compete for the power resources P and for access to the destination. Such an access architecture is known to have a sum-capacity scaling law equal to $\log_2\log\left(N\right)$ [70].

Aiming, thus, at making relaying more spectrally efficient, we turn now to amplify-based relaying. In particular, we consider *linear relaying*.

3.6 Linear Relaying

Linear relaying (LR) consists of relay nodes transmitting, on every channel use t, a linear combination of previously received signals, *i.e.* $x_i^t=\sum_{j=1}^{t-1}\phi_i\left(t,j\right)y_i^j$, where $\phi_i\left(t,j\right)$ is the amplifying factor at the relay i of received signal y_i^j.

This technique was proposed by Zahedi and El Gamal in [24] as the natural extension of *amplify-and-forward* to full-duplex operation. From a practical standpoint, is the simplest and most easily deployed scheme: it only requires buffering and amplifying capabilities. In this section, we will show that it is also effective, mainly because it is constrained neither by the source-relays broadcast channel nor by the relays-destination multiple access channel.

It operates as follows: let the source select message $\omega \in \left\{1, \cdots, 2^{nR}\right\}$ for transmission and map it onto codeword $\boldsymbol{x}_s^n$. The codeword is then transmitted into B channel blocks of $\kappa = n/B$ channel uses each. On every block b, the source transmits the sequence of symbols $\boldsymbol{x}_s^\kappa[b]$, which is received at the relays and destination following (3.2) and (3.3), respectively. Simultaneously, the relays linearly combine the received signal during this block, and transmit the sequence of symbols $\boldsymbol{x}_i^\kappa[b]$, $i = 1, \cdots, N$, which can be compactly written as:

$$\begin{aligned}\boldsymbol{x}_i^\kappa[b] &= \boldsymbol{\Phi}_i \cdot \boldsymbol{y}_i^\kappa[b] \\ &= \boldsymbol{\Phi}_i \cdot \left(b_i \cdot \boldsymbol{x}_s^\kappa[b] + \boldsymbol{z}_i^\kappa\right), \end{aligned} \tag{3.88}$$

where $\boldsymbol{\Phi}_i \in \mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}$ is referred to as the linear relaying matrix. It defines the linear combination of inputs, and is strictly lower triangular to preserve causality. The received signal at the destination is given by[4]:

$$\begin{aligned}\boldsymbol{y}_d^\kappa &= a \cdot \boldsymbol{x}_s^\kappa + \sum_{i=1}^N c_i \cdot \boldsymbol{x}_i^\kappa + \boldsymbol{z}_d^\kappa \\ &= \left(a \cdot \boldsymbol{I} + \sum_{i=1}^N b_i c_i \boldsymbol{\Phi}_i\right) \cdot \boldsymbol{x}_s^\kappa + \left(\boldsymbol{z}_d^\kappa + \sum_{i=1}^N c_i \boldsymbol{\Phi}_i \boldsymbol{z}_i^\kappa\right). \end{aligned} \tag{3.89}$$

As for previous relaying techniques, the communication is constrained to satisfy (3.4). This can be stated as follows: let $\boldsymbol{Q}^\kappa = \mathrm{E}\left\{\boldsymbol{x}_s^\kappa \left(\boldsymbol{x}_u^\kappa\right)^\dagger\right\} \succeq 0$ be the source temporal covariance matrix, and $\boldsymbol{Q}_i^\kappa = \mathrm{E}\left\{\boldsymbol{x}_i^\kappa \left(\boldsymbol{x}_i^\kappa\right)^\dagger\right\} = |b_i|^2 \boldsymbol{\Phi}_i \boldsymbol{Q}^\kappa \boldsymbol{\Phi}_i^\dagger + N_o \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\dagger$, the relays temporal covariance matrix. Then, the total transmitted power in the network is:

$$\begin{aligned}\mathsf{P}\left(\boldsymbol{Q}^{\boldsymbol{\kappa}}, \boldsymbol{\Phi}_{1:N}\right) &= \frac{1}{\kappa}\left(\mathrm{tr}\left\{\boldsymbol{Q}^\kappa\right\} + \sum_{i=1}^N \mathrm{tr}\left\{\boldsymbol{Q}_i^\kappa\right\}\right) \\ &= \frac{1}{\kappa}\mathrm{tr}\left\{\boldsymbol{Q}^\kappa\left(\boldsymbol{I} + \sum_{i=1}^N |b_i|^2 \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right) + N_o \sum_{i=1}^N \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right\}, \end{aligned} \tag{3.90}$$

which is enforced to satisfy (3.4), *i.e.,* $\mathsf{P}\left(\boldsymbol{Q}^{\boldsymbol{\kappa}}, \boldsymbol{\Phi}_{1:N}\right) \leq P$. Clearly, the signal model (3.89) is equivalent to that of the Gaussian channel with block memory and colored noise [71]. Using this analogy, its maximum achievable rate can be derived as the maximum mutual information between the sequences transmitted and received by source and destination, respectively, when

[4]For clarity of exposition, we remove index b.

the number of channel uses per block $\kappa \to \infty$ (see *e.g.*, [24, 71]). That is:

$$\begin{aligned} R_{\text{LR}} &= \lim_{\kappa\to\infty} \max_{\mathbf{\Phi}_{1:N}\in\mathbb{C}^{\kappa\times\kappa}_{\text{SLT}}} \max_{p_{\boldsymbol{x}_s^\kappa}:\mathsf{P}(\boldsymbol{Q}^\kappa,\mathbf{\Phi}_{1:N})\leq P} \frac{1}{\kappa}\cdot I\left(\boldsymbol{x}_s^\kappa;\boldsymbol{y}_d^\kappa\right) \\ &= \lim_{\kappa\to\infty} \max_{\mathbf{\Phi}_{1:N}\in\mathbb{C}^{\kappa\times\kappa}_{\text{SLT}}} \max_{\boldsymbol{Q}^\kappa\succeq 0:\mathsf{P}(\boldsymbol{Q}^\kappa,\mathbf{\Phi}_{1:N})\leq P} \frac{1}{\kappa}\cdot \log_2\det\left(\boldsymbol{I}+\boldsymbol{H}_e\boldsymbol{Q}^\kappa\boldsymbol{H}_e^\dagger\right) \end{aligned} \tag{3.91}$$

where

$$\boldsymbol{H}_e = \frac{1}{\sqrt{N_o}}\left(\boldsymbol{I}+\sum_{i=1}^N |c_i|^2\mathbf{\Phi}_i\mathbf{\Phi}_i^\dagger\right)^{-\frac{1}{2}}\left(a\cdot\boldsymbol{I}+\sum_{i=1}^N b_ic_i\mathbf{\Phi}_i\right). \tag{3.92}$$

The second equality is derived using standard arguments for Gaussian channels in [21, Section 10.5] [71]. Unfortunately, unlike the Gaussian channel with memory, no closed-form expression can be given for optimization (3.91). In fact, the optimization on $\boldsymbol{Q}^\kappa$ is convex and can be solved. However, the other one is not convex on $\mathbf{\Phi}_{1:N}$ and no solution can be given. Therefore, in order to tackle the problem, we proceed in two steps: first, we analyze the optimum source signalling, considering fixed κ and a fixed (not necessarily optimum) set $\mathbf{\Phi}_{1:N}$. We refer to this result as conditional capacity with LR. Afterwards, we propose suboptimum relaying matrices.

3.6.1 Design of the Source Temporal Covariance

The conditional capacity of LR, given κ and the set $\mathbf{\Phi}_{1:N}\in\mathbb{C}^{\kappa\times\kappa}_{\text{SLT}}$, is the supremum of all rates that are achievable when the relay nodes use the fixed set of relaying matrices $\mathbf{\Phi}_{1:N}$. It is defined from (3.91) as :

$$\begin{aligned} R_{\text{LR}}^\kappa\left(\mathbf{\Phi}_{1:N}\right) &= \max_{\boldsymbol{Q}^\kappa\succeq 0} \frac{1}{\kappa}\cdot\log_2\det\left(\boldsymbol{I}+\boldsymbol{H}_e\boldsymbol{Q}^\kappa\boldsymbol{H}_e^\dagger\right) \\ &\quad\ \text{s.t. } \mathsf{P}\left(\boldsymbol{Q}^\kappa,\mathbf{\Phi}_{1:N}\right)\leq P \end{aligned} \tag{3.93}$$

Notice that we have defined it as a capacity since it is the maximum transmission rate that is achievable when relays can only transmit by means of amplifying matrices $\mathbf{\Phi}_{1:N}$. The optimization is clearly convex in standard form: the objective function is concave and differentiable, and the constraint defines a convex, regular set. It can be thus solved by means of KKT conditions, obtaining the optimum source signalling.

Theorem 3.6 *Let us define* $\boldsymbol{A}=\left(\boldsymbol{I}+\sum_{i=1}^N|b_i|^2\mathbf{\Phi}_i^\dagger\mathbf{\Phi}_i\right)$*, and compute the SVD-Decomposition* $\boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}}=\boldsymbol{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^\dagger$ *with* $\mathbf{\Lambda}=\text{diag}(\lambda_1,\cdots,\lambda_\kappa)$*. Then:*

$$R_{\text{LR}}^\kappa\left(\mathbf{\Phi}_{1:N}\right)=\frac{1}{\kappa}\cdot\sum_{j=1}^\kappa\mathcal{C}\left(\lambda_j\psi_j\right) \tag{3.94}$$

where

$$\psi_j = \left[\frac{1}{\mu} - \frac{1}{\lambda_j}\right]^+, \text{ so that } \operatorname{tr}\left\{N_o \sum_{i=1}^N \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right\} + \sum_{j=1}^\kappa \psi_j = \kappa P. \tag{3.95}$$

The conditional capacity is attained with source covariance $\boldsymbol{Q}_{\text{opt}}^\kappa = \boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{V}\mathbf{\Psi}\boldsymbol{V}^\dagger\boldsymbol{A}^{-\frac{1}{2}}$, *being* $\mathbf{\Psi} = \operatorname{diag}(\psi_1, \cdots, \psi_\kappa)$.

Remark 3.6 *Surprisingly, the optimum covariance does not diagonalize the equivalent* $\boldsymbol{H}_e$, *but its modified form* $\boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}}$. *This is computed from*

$$\boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}} = \frac{1}{\sqrt{N_o}}\left(\boldsymbol{I} + \sum_{i=1}^N |c_i|^2 \mathbf{\Phi}_i \mathbf{\Phi}_i^\dagger\right)^{-\frac{1}{2}} \left(a \cdot \boldsymbol{I} + \sum_{i=1}^N b_i c_i \mathbf{\Phi}_i\right) \left(\boldsymbol{I} + \sum_{i=1}^N |b_i|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right)^{-\frac{1}{2}}.$$

From it, we clearly notice that the source-relay channels b_i *impact on the conditional capacity in the same manner than the relay-destination channels* c_i *do. Thereby, both can be interchanged (i.e.,* $b_i \leftrightarrow c_i$*) without modifying the achievable rate. This has an important interpretation in practice: with LR, the achievable rate from* s *to* d *is equal to that from* d *to* s. *Linear relaying is thus reciprocal, as the max-flow-min-cut bound and unlike D&F, PD&F and C&F.*

Proof: We first write the Lagrangian for optimization (3.93):

$$\mathcal{L}\left(\boldsymbol{Q}^\kappa, \mathbf{\Omega}, \mu\right) = \log\left(\det\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{Q}^\kappa\boldsymbol{H}_e^\dagger\right)\right) + \operatorname{tr}\left\{\mathbf{\Omega}\boldsymbol{Q}^\kappa\right\} - \mu\left(\kappa \mathsf{P}\left(\boldsymbol{Q}^{\boldsymbol{\kappa}}, \mathbf{\Phi}_{1:N}\right) - \kappa P\right)$$

where $\mu \geq 0$ and matrix $\mathbf{\Omega} \succeq 0$ are the Lagrange multipliers for the power and semi-definite positive constraints, respectively. The KKT conditions for the problem, which are sufficient and necessary for optimality (due to convexity, and regularity of the feasible set) are [52]:

$$i)\ \mu\left(\boldsymbol{I} + \sum_{i=1}^N |b_i|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right) - \mathbf{\Omega} = \boldsymbol{H}_e^\dagger\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{Q}^\kappa\boldsymbol{H}_e^\dagger\right)^{-1}\boldsymbol{H}_e \tag{3.96}$$

$$ii)\ \mu\left(\operatorname{tr}\left\{\boldsymbol{Q}^\kappa\left(\boldsymbol{I} + \sum_{i=1}^N |b_i|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right) + N_o\sum_{i=1}^N \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right\} - \kappa P\right) = 0, \tag{3.97}$$

$$iii)\ \operatorname{tr}\left\{\mathbf{\Omega}\boldsymbol{Q}^\kappa\right\} = 0$$

Let us now define $\boldsymbol{A} = \left(\boldsymbol{I} + \sum_{i=1}^N |b_i|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right)$, which is clearly singular and Hermitian, semi-definite positive. Also, consider the change of variables $\tilde{\boldsymbol{Q}} = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{Q}^\kappa\boldsymbol{A}^{\frac{1}{2}}$ and $\tilde{\mathbf{\Omega}} = \boldsymbol{A}^{-\frac{1}{2}}\mathbf{\Omega}\boldsymbol{A}^{-\frac{1}{2}}$, both Hermitian, semidefinite positive. With them, we can turn the KKT conditions into:

$$i)\ \mu - \tilde{\mathbf{\Omega}} = \boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{H}_e^\dagger\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{H}_e^\dagger\right)^{-1}\boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}} \tag{3.98}$$

$$ii)\ \mu\left(\operatorname{tr}\left\{\tilde{\boldsymbol{Q}} + N_o\sum_{i=1}^N \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right\} - \kappa P\right) = 0,$$

$$iii)\ \operatorname{tr}\left\{\tilde{\mathbf{\Omega}}\tilde{\boldsymbol{Q}}\right\} = 0$$

Considering the SVD-decomposition $\boldsymbol{H}_e\boldsymbol{A}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^{\dagger}$, it can be derived (as for the capacity of MIMO channels [3]) that $\tilde{\boldsymbol{Q}}^* = \boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{V}^{\dagger}$, with $\psi_j = \left[\frac{1}{\mu^*} - \frac{1}{\lambda_j}\right]^+, \ j = 1, \cdots, k$ satisfies KKTs. μ^* is such that $\sum_{j=1}^{k}\psi_j + N_o\mathrm{tr}\left\{\sum_{i=1}^{N}\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right\} - \kappa P = 0$, and $\tilde{\boldsymbol{\Omega}}^*$ computed from KKT *i*) as: $\tilde{\boldsymbol{\Omega}}^* = \boldsymbol{V}\left(\mu - \boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{I} + \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Psi}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\boldsymbol{V}^{\dagger}$, which is semidefinite positive. Finally, we recover the optimum source temporal covariance as

$$\boldsymbol{Q}_{\mathrm{opt}}^{\kappa} = \boldsymbol{A}^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}^*\boldsymbol{A}^{-\frac{1}{2}}, \tag{3.99}$$

for which we evaluate $\log_2\det\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{Q}_{\mathrm{opt}}^{\kappa}\boldsymbol{H}_e^{\dagger}\right)$. ■

3.6.2 Linear Relaying Matrix Design

In the previous subsection, we presented the optimum source temporal covariance, given the fixed set $\boldsymbol{\Phi}_{1:N}$. Now, in order to search for the optimum set of relaying matrices, we need to solve (3.91):

$$\boldsymbol{\Phi}_{1:N}^* = \lim_{\kappa\to\infty}\arg\max_{\boldsymbol{\Phi}_{1:N}\in\mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}} R_{\mathrm{LR}}^{\kappa}\left(\boldsymbol{\Phi}_{1:N}\right). \tag{3.100}$$

As mentioned above, such a maximization is not convex. To solve it, we need to resort to non-convex optimization algorithms that, for $N \gg 1$ and $\kappa \to \infty$, become unfeasible.

A suboptimum approach is thus considered. In particular, we propose the relays to use the suboptimum set of relaying matrices, $\bar{\boldsymbol{\Phi}}_{1:N}$, based upon *amplify-and-forward* extended to $\kappa > 2$: on every channel use t, relays only amplify and retransmit the signal received on previous channel use $t-1$. With them, the required memory at the relays is reduced to one sample. They are defined as:

$$\begin{aligned}&\bar{\boldsymbol{\Phi}}_i = \eta_i\boldsymbol{\Phi}_0, \ i = 1, \cdots, N. \ \text{with} \\ &\left[\boldsymbol{\Phi}_0\right]_{p,q} \triangleq \begin{cases}\sqrt{\beta} & p = q+1; \ 1 \le q \le \kappa - 1 \\ 0 & \text{elsewhere.}\end{cases},\end{aligned} \tag{3.101}$$

where $\eta_i \in \mathbb{C}$ are the beamforming weights and satisfy $\sum_{i=1}^{N}|\eta_i|^2 = 1$. Notice that β in (3.101) needs to satisfy $\kappa P > N_o\beta(\kappa - 1)$ so that $\sum_{j=1}^{\kappa}\psi_j > 0$ in (3.95). That is, $\beta < \frac{P}{N_o}\frac{\kappa}{\kappa-1}$.

We need now to select the beamforming weights η_i that maximize the achievable rate. Unfortunately, as relays not only amplify the source signal but also thermal noise, the beamforming optimization turns out be non-convex, and unsolvable for us. To simplify the framework, we propose a suboptimum solution, which we call *Maximal Ratio Transmission* (MRT). It consists of relays beamforming towards destination as if they were a transmit antenna array, regardless of noise amplification:

$$\eta_i \triangleq \frac{b_i^* \cdot c_i^*}{\sqrt{\sum_{i=1}^{N} |b_i \cdot c_i|^2}}. \tag{3.102}$$

This is indeed the optimum beamforming with high SNR at the relays. Let us now apply Theorem 3.6 to the $\bar{\mathbf{\Phi}}_{1:N}$ proposed in (3.101). Firstly, we define $A = 1 + \sum_{i=1}^{N} |b_i|^2 |\eta_i|^2 \beta$ and $B = 1 + \sum_{i=1}^{N} |c_i|^2 \eta_i|^2 \beta$. Then, it is easy to show that for $k \to \infty$:

$$\begin{aligned} \boldsymbol{I} + \sum_{i=1}^{N} |b_i|^2 \bar{\mathbf{\Phi}}_i^\dagger \bar{\mathbf{\Phi}}_i &= \mathrm{diag}\left([A\mathbf{1}_{k-1}, 1]\right) \approx A\boldsymbol{I}_k \\ \boldsymbol{I} + \sum_{i=1}^{N} |c_i|^2 \bar{\mathbf{\Phi}}_i \bar{\mathbf{\Phi}}_i^\dagger &= \mathrm{diag}\left([1, B\mathbf{1}_{k-1}]\right) \approx B\boldsymbol{I}_k \end{aligned} \tag{3.103}$$

where $\boldsymbol{T} \approx \boldsymbol{R}$ stands for $\lim_{\kappa\to\infty} ||\boldsymbol{T} - \boldsymbol{R}||/||\boldsymbol{T}|| = 0$. Therefore, $\boldsymbol{H}_e \boldsymbol{A}^{-\frac{1}{2}}$ in Theorem 3.6 satisfies:

$$[\boldsymbol{H}_e \boldsymbol{A}^{-\frac{1}{2}}]_{p,q} \approx \begin{cases} (N_o AB)^{-\frac{1}{2}} \cdot a & p = q,\ p = 1, \cdots, \kappa \\ (N_o AB)^{-\frac{1}{2}} \cdot \sqrt{\beta \sum_{i=1}^{N} |b_i c_i|^2} & p = q+1;\ 1 \leq q \leq \kappa - 1 \\ 0 & \text{elsewhere} \end{cases} \tag{3.104}$$

With it, $\boldsymbol{H}_e \boldsymbol{A}^{-1} \boldsymbol{H}_e^\dagger$ can be approximated by a Toeplitz matrix of the form:

$$[\boldsymbol{H}_e \boldsymbol{A}^{-1} \boldsymbol{H}_e^\dagger]_{p,q} \approx \frac{t_{p-q}}{N_o A \cdot B}, \tag{3.105}$$

with $t_{-1} = a^* \sqrt{\beta \sum_{i=1}^{N} |b_i c_i|^2}$, $t_0 = |a|^2 + \beta \sum_{i=1}^{N} |b_i c_i|^2$, $t_1 = a\sqrt{\beta \sum_{i=1}^{N} |b_i c_i|^2}$ and $t_p = 0$ elsewhere. Let us now denote by $\mathbf{\Lambda}$ the eigenvalues of $\boldsymbol{H}_e \boldsymbol{A}^{-1} \boldsymbol{H}_e^\dagger$, and compute the Fourier Transform of $\frac{t_p}{N_o A \cdot B}$:

$$\begin{aligned} f(\omega) &= \tfrac{1}{N_o A \cdot B} \cdot \textstyle\sum_{p=-1}^{p=1} t_p e^{-j\omega p} \\ &= \frac{|a|^2 + \beta \sum_{i=1}^{N} |b_i c_i|^2 + 2\mathrm{Re}\left\{ a\sqrt{\beta \sum_{i=1}^{N} |b_i c_i|^2} e^{-j\omega} \right\}}{N_o \left(1 + \sum_{i=1}^{N} |b_i|^2 |\eta_i|^2 \beta\right) \left(1 + \sum_{i=1}^{N} |c_i|^2 |\eta_i|^2 \beta\right)}. \end{aligned} \tag{3.106}$$

Thereby, we can make use of [72, Theorem 4.1] to show that, for large κ, the eigenvalue distribution of (3.105) coincides with the Fourier transform of $\frac{t_p}{N_o A \cdot B}$. As a result:

$$\begin{aligned}\lambda_n &\approx f\left(2\pi\frac{n}{\kappa}\right) \\ &= \frac{|a|^2 + \beta\sum_{i=1}^N |b_ic_i|^2 + 2\mathrm{Re}\left\{a\sqrt{\beta\sum_{i=1}^N |b_ic_i|^2}e^{-j2\pi\frac{n}{\kappa}}\right\}}{N_o\left(1+\sum_{i=1}^N |b_i|^2|\eta_i|^2\beta\right)\left(1+\sum_{i=1}^N |c_i|^2|\eta_i|^2\beta\right)}, \ n = 1, \cdots, \kappa.\end{aligned} \tag{3.107}$$

We can now introduce (3.107) in Theorem 3.6 to derive the achievable rate for a given value of β. Finally, optimizing over $\beta \in [0, \frac{P}{N_o}\frac{\kappa}{\kappa-1})$, we obtain the maximum transmission rate with the proposed linear relaying functions. Last optimization can be carried out using one-dimensional exhaustive search.

3.6.3 Asymptotic Performance

Let us analyze now the asymptotic performance of LR. As for previous sections, we consider unitary-mean, Rayleigh-fading within the network: $a, b_i, c_i \sim \mathcal{CN}(0, 1)$. For such a channel distribution, D&F, PD&F and C&F were shown to be limited by the source-relays broadcast channel and by the relays-destination multiple-access channel, respectively. Both limitations constrain their achievable rates to scale even slower than $\log_2 \log N$. In this section, we will show that LR is not constrained by none of them; therefore, its achievable rate is able to scale as $\log_2 N$, seizing all the beamforming gain of the network:

$$\lim_{N\to\infty} \frac{R_{\mathrm{LR}}}{\log_2 N} \stackrel{P}{=} K, \tag{3.108}$$

with K a constant. This result has been previously published by Dana *et. al.* in [56]. Upper and lower bounds are used therein to demonstrate the scaling law. In the following, we provide a more intuitive proof so that the reader can internalize the result; however, we refer to [56] for the first mathematically rigorous proof.

Consider the linear relaying functions $\bar{\mathbf{\Phi}}_{1:N}$ presented in (3.101)-(3.102). With them, the achievable rate follows (3.94) in Theorem 3.6. The rate is computed therein in terms of the eigenvalues of the matrix $\boldsymbol{H}_e\boldsymbol{A}^{-1}\boldsymbol{H}_e^\dagger$, which (for $\kappa \gg 1$) can be approximated using (3.107):

$$\lambda_n = \frac{|a|^2 + \beta\sum_{i=1}^N |b_ic_i|^2 + 2\mathrm{Re}\left\{a\sqrt{\beta\sum_{i=1}^N |b_ic_i|^2}e^{-j2\pi\frac{n}{\kappa}}\right\}}{N_o\left(1+\sum_{i=1}^N |b_i|^2|\eta_i|^2\beta\right)\left(1+\sum_{i=1}^N |c_i|^2|\eta_i|^2\beta\right)}, \ n = 1, \cdots, \kappa, \tag{3.109}$$

Notice that, from the specific values of η_i in (3.102) we compute:

$$\sum_{i=1}^{N} |b_i|^2|\eta_i|^2 = \frac{\sum_{i=1}^{N} |b_i|^4|c_i|^2}{\sum_{i=1}^{N} |b_ic_i|^2} \text{ and } \sum_{i=1}^{N} |c_i|^2|\eta_i|^2 = \frac{\sum_{i=1}^{N} |c_i|^4|b_i|^2}{\sum_{i=1}^{N} |b_ic_i|^2}. \tag{3.110}$$

Let us now show some convergence results related to (3.109) and (3.110). First, we notice that random variables $|b_ic_i|^2 = |b_i|^2|c_i|^2$, $i = 1, \cdots, N$, are unitary mean. Therefore, applying the law of large numbers we claim that, for $N \to \infty$:

$$\frac{\sum_{i=1}^{N} |b_ic_i|^2}{N} \xrightarrow{P} 1. \tag{3.111}$$

Furthermore, notice that $\mathrm{E}\left\{|b_i|^4|c_i|^2\right\} = \mathrm{E}\left\{|b_i|^4\right\}$ since b_i, c_i are independent and Rayleigh distributed. Likewise:

$$\begin{aligned} \mathrm{E}\left\{|b_i|^4\right\} &= \int_0^\infty x^2e^{-x}\mathrm{d}x \qquad (3.112)\\ &= -x^2e^{-x}\Big]_0^\infty + 2\int_0^\infty xe^{-x}\mathrm{d}x \\ &= 2. \qquad (3.113)\end{aligned}$$

Therefore, $\mathrm{E}\left\{|b_i|^4|c_i|^2\right\} = \mathrm{E}\left\{|c_i|^4|b_i|^2\right\} = 2$, $i = 1, \cdots, N$, and making use of the law of large numbers allows we obtain:

$$\frac{\sum_{i=1}^{N} |b_i|^4|c_i|^2}{N} \xrightarrow{P} 2 \quad \text{and} \quad \frac{\sum_{i=1}^{N} |c_i|^4|b_i|^2}{N} \xrightarrow{P} 2. \tag{3.114}$$

Accordingly, we can plug (3.111) and (3.114) into (3.110), and apply Lemma 2.4 to derive:

$$\sum_{i=1}^{N} |b_i|^2|\eta_i|^2 \xrightarrow{P} 2 \text{ and } \sum_{i=1}^{N} |c_i|^2|\eta_i|^2 \xrightarrow{P} 2. \tag{3.115}$$

Thereby, introducing (3.115) in (3.109), along with (3.111), we can obtain:

$$\frac{\lambda_j}{\left(\frac{\beta N}{N_o\cdot(1+2\beta)^2}\right)} \xrightarrow{P} 1, \ \ j = 1, \cdots, \kappa. \tag{3.116}$$

Finally, plugging such an asymptotic performance of the eigenvalues into Theorem 3.6, the $\log_2 N$ scaling law of *Linear Relaying* is quickly grasped.

3.7 Numerical Results

The achievable rates of the four relaying schemes presented in this chapter are evaluated in a block-fading environment. We assume the channel fading to take a zero-mean, complex,

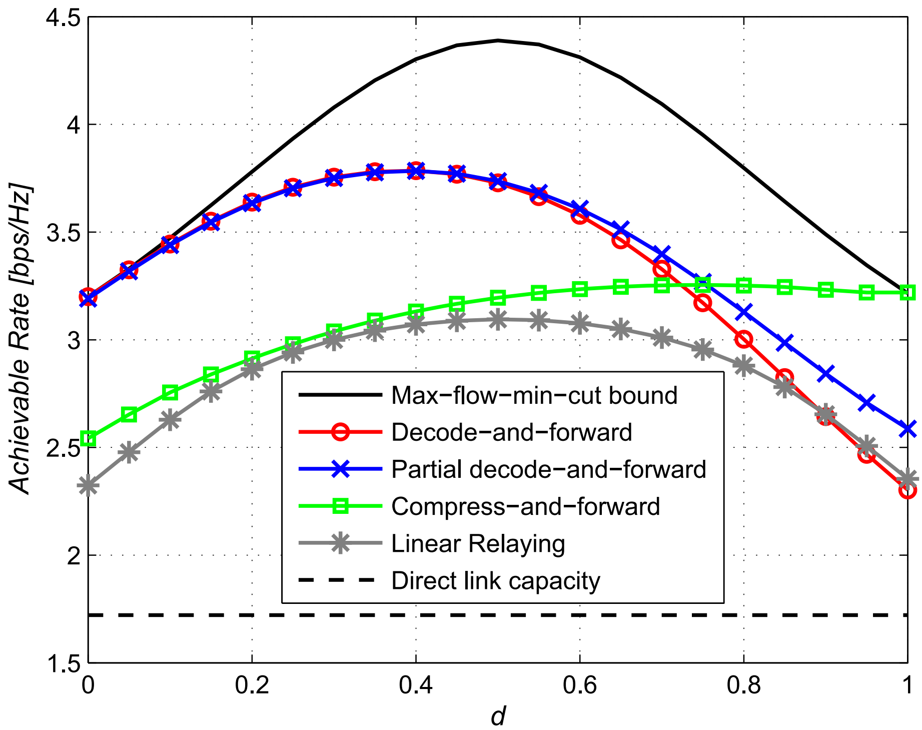

Figure 3.3: Achievable rates versus source-relay distance d. We consider transmit SNR = 5dB and $N = 2$ relays. Wireless channels are Rayleigh faded.

Gaussian distribution and to be *time-invariant*. In particular, defining $d_{s,d} = 1$ the source-destination distance and d the source-relays distance, the channel gains are distributed as $a \sim \mathcal{CN}(0,1)$, $b_i \sim \mathcal{CN}(0, d^{-\alpha})$, and $c_i \sim \mathcal{CN}\left(0, (1-d)^{-\alpha}\right)$, $i = 1, \cdots, N$. We have set $\alpha = 3$, the path-loss exponent. All plots in this section show the expected value of the achievable rate, averaged over the joint channel distribution via Monte-Carlo.

Relaying strategies are compared in Figs. 3.3-3.5 with respect to the source-relays distance d, for number of relays $N = 2, 8$ and 32, respectively. We have considered $\frac{P}{N_o} = 5$ dB. In the case of C&F, only the upper bound derived in subsection 3.5.1 is plotted. Moreover, the *max-flow-min-cut* bound and the source-destination direct link capacity are depicted as reference. Additionally, the achievable rate versus the number of relays are depicted in Figs. 3.6-3.7, for distances $d = 0.05$ and $d = 0.95$, respectively. Interesting conclusions can be drawn from the plots:

- *Decode-and-forward.* Surprisingly, and despite its asymptotic performance, it is one of the most competitive approaches. To start with, for low number of relays (*e.g.* $N = 2$ in

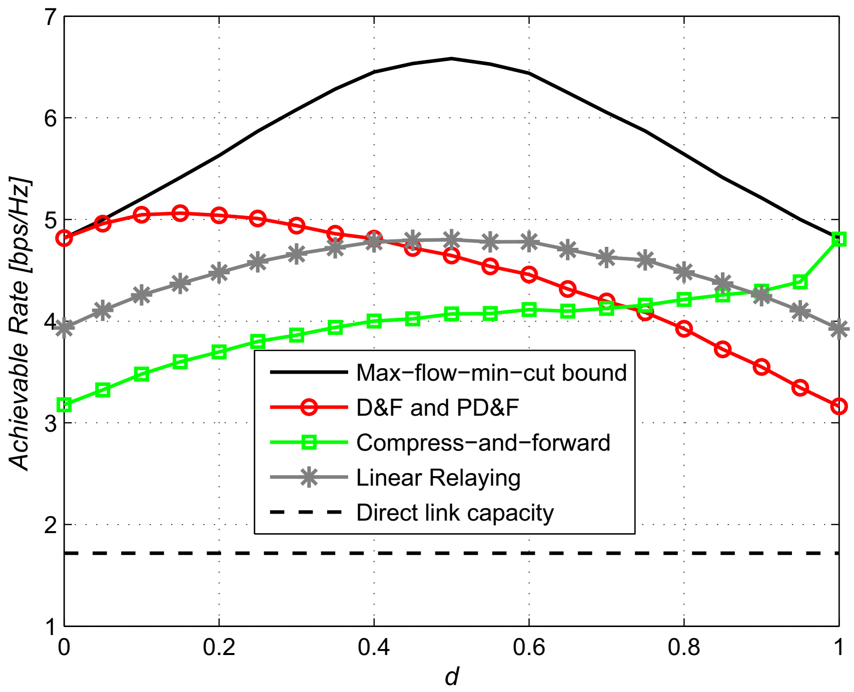

Figure 3.4: Achievable rates versus source-relay distance d. We consider transmit SNR = 5dB and $N = 8$ relays. Wireless channels are Rayleigh faded.

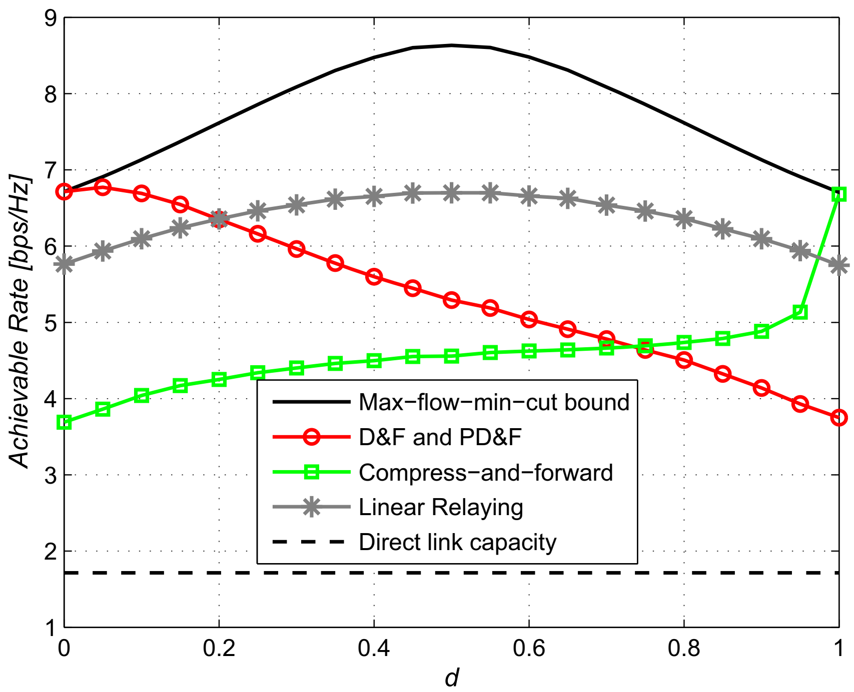

Figure 3.5: Achievable rates versus source-relay distance d. We consider transmit SNR = 5dB and $N = 32$ relays. Wireless channels are Rayleigh faded.

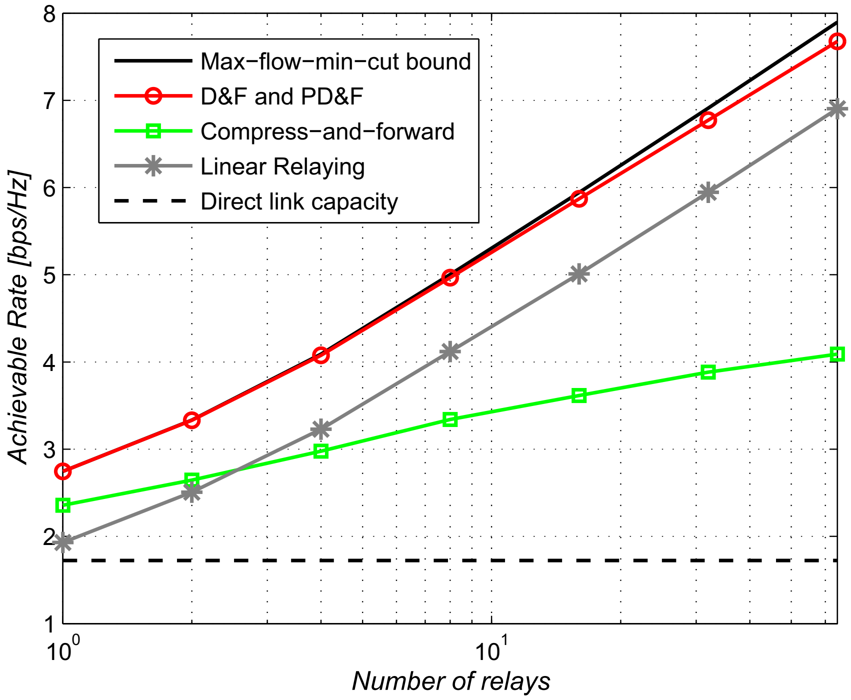

Figure 3.6: Achievable rates versus the number of relays N. We consider transmit SNR = 5dB and $d = 0.05$. Wireless channels are Rayleigh faded.

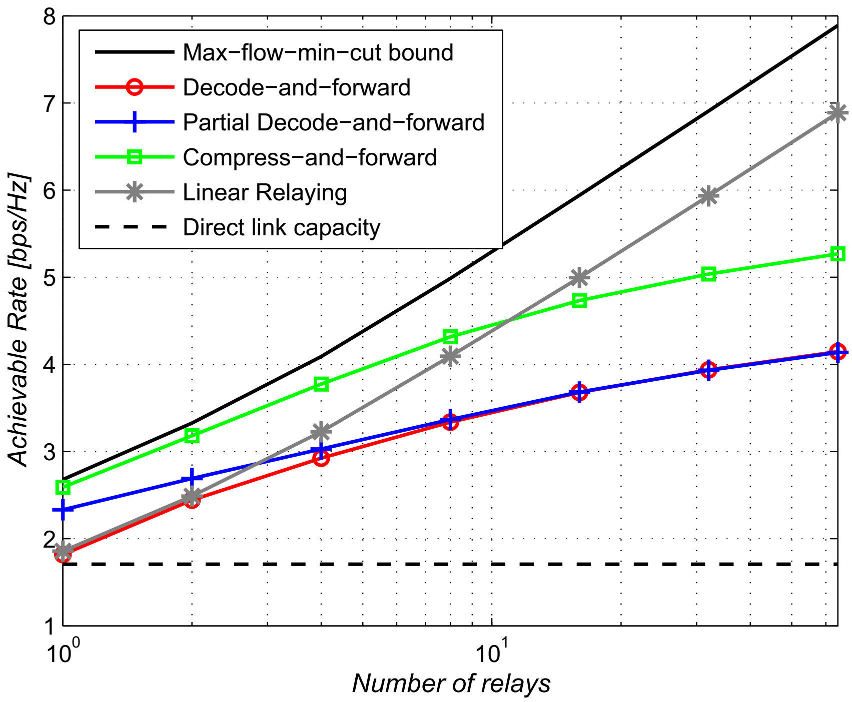

Figure 3.7: Achievable rates versus the number of relays N. We consider transmit SNR = 5dB and $d = 0.95$. Wireless channels are Rayleigh faded.

Fig. 3.3) it is the dominant technique[5] on the 60% of the distance range: $0 \leq d \leq 0.60$. Likewise, for short-mid distances ($0 \leq d \leq 0.2$ in Fig. 3.3), it perfectly mimics a transmit antenna array, having an achievable rate identical to the *max-flow-min-cut* upper bound. Thus, it is capacity-achieving for a wide range of distances.

Unfortunately, for increasing number of relays (*e.g.* $N = 8$ and $N = 32$ in Fig. 3.4 and 3.5, respectively) such a promising features are gotten worse. In particular, it remains the best technique only for the 20% of the range closer to the source. This is consistent with the analysis carried out in Section 3.3.1, where significant losses are shown for $N \gg 1$. In fact, when analyzing D&F for increasing number of relays, we notice that: *i*) even for very low source-relay distances ($d = 0.05$ in Fig. 3.6) the broadcast-limitation moves the D&F rate further away from the capacity upper bound (on its defense, we must say that more than 15 relays are needed to glimpse the Lambert W asymptotic performance) *ii*) the worst scenario for D&F is jointly having high source-relay distances and high number of relays (see $d = 0.95$ in Fig. 3.7). For it, D&F performs substantially far apart from all other techniques.

- *Partial decoding*. It is the most disappointing technique among all considered in this dissertation. Numerical results show that, in general, it produces negligible gains with respect to D&F. As derived in the chapter, for large N (*e.g.* for $N = 8$ and $N = 32$ relays in Fig. 3.4 and 3.5 respectively) both regenerative techniques perform exactly equal. In fact, via simulation, we realize that partial decoding is only worthwhile within a unique scenario: low number of relays plus large source-relays distance (see *e.g.,* the range $0.7 \leq d \leq 1$ in Fig. 3.3 and $N \leq 8$ in Fig. 3.7). Within such a setup, it produces moderate gains since the source-destination channel is of similar magnitude to that from source to relays ($|a| \approx |b_m|$). For all other cases, it is reveled useless (see *e.g.* Fig. 3.6 with $d = 0.05$).

- *Linear relaying*. Interestingly, it is at the same time the head and the tails of the coin. On the one hand, it has an extremely poor performance with low number of relays; *e.g.*, Fig. 3.3, where it is outperformed by all other relaying techniques, and Fig. 3.6-3.7, where it presents the worst performance for $N = 1$ and $N = 2$ relays. This can be easily understood: the most important feature of LR is its capability to seize all the beamforming

[5] Along with PD&F.

gain of the system. In contrast, its main drawback is the noise amplification. With low number of relays, the degree of noise amplification is not compensated by the beamforming gain obtained from relays, which is small for $N \leq 2$. Hence, it performs poorly. On the other hand, with high number of relays the beamforming gain is large and powerful enough to overcome noise amplification. Hence, LR becomes the most profitable relaying techniques, outperforming all others, as shown in *e.g.* Fig. 3.5 and Fig. 3.7. Finally, it is clearly shown in Fig. 3.3-3.5 that linear relaying is reciprocal with respect to the source-relay distance, as shown in Theorem 3.6.

- *Compress-and-forward*. As expected, it is the better relaying technique for large d (see Fig. 3.3-3.5). However, the range of distances for which it is optimal is much smaller than that of D&F, and only collapses with the upper bound for exactly $d = 1$. In this sense, we can claim that D&F mimics more perfectly a transmit antenna array than C&F mimics a receive one. Additionally, the $\log_2 \log N$ scaling law of C&F is clearly grasped from Fig. 3.7. It is shown therein that, even for large d, C&F is optimal only for small number of relays.

3.8 The Half-duplex Model

In previous sections, it was assumed that the relays transmit towards the destination and receive from the source simultaneously, and in the same frequency band. This mode of operation is known as *full-duplex*. We justified it assuming that relays use different antennas to transmit and receive, both perfectly isolated. However, such an assumption is not always realistic in practice, mainly, due to two reasons:

1. Transmit and receive antennas are, normally, not perfectly isolated. Therefore, the self-interference can be much stronger than the useful signal received from the source. The reception at the relay would be thus limited by its own interference, and not by AWGN as we assumed up to now. As an example, consider a relay transmitting with power 23 dBm. Assume an isolation among antennas of 80 dB. Hence, the self-interference level is

$$P_{int} = 23 - 80 = -57 \text{ (dBm)}. \tag{3.117}$$

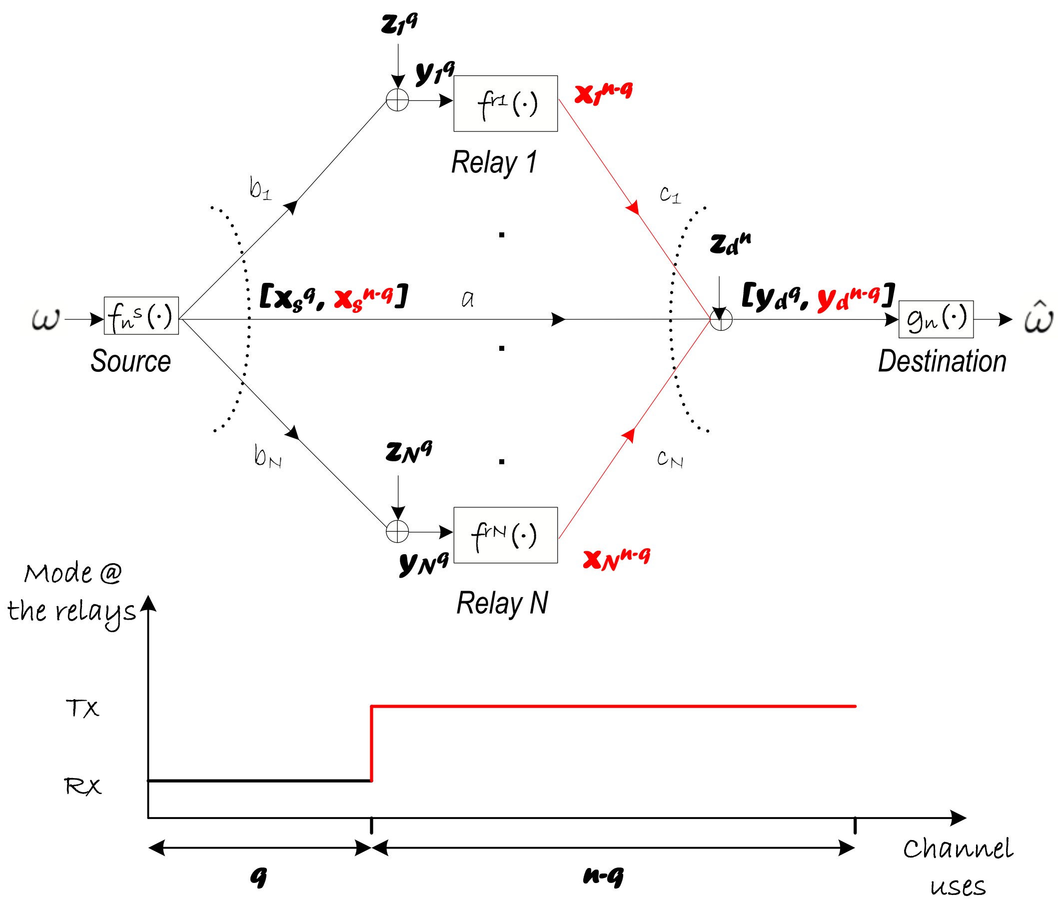

Figure 3.8: Multiple-Parallel Relay Channel

On the other hand, consider a typical noise power at the relay of -110 dBm, and a signal-to-noise-ratio for the source signal of $SNR = 20$ dB. Therefore, the power received at the relay from the source is

$$P_s = -110 + 20 = -90 \text{ (dBm)}. \tag{3.118}$$

Hence, the signal-to-interference ratio is $SIR = -90 + 57 = -33$ dB, much weaker than the $SNR = 20$ dB that we assumed for analysis.

2. It is not always possible, or worthwhile, to use two antennas at the relays. When a single antenna is used, the transmit and receive chains are usually isolated by a circulator. Typical state-of-the-art circulators have an isolation of up to 70 dB. Hence, considering the previous example, it is clear that self-interference limits the performance.

When the conditions for full-duplex operation are not met, relays must transmit and receive using a multiplexing mode, either time-division (TD) or frequency-division (FD). This is referred to as *half-duplex* operation. In this section, we consider a TD scheme at the relays, although results are perfectly applicable to FD. We arrange the communication in two consecutive time slots, of q and $n-q$ channel uses, respectively (see Fig. 3.8). During the first slot, the source node transmits the sequence of symbols $\boldsymbol{x}_s^q[1]$ to the set of relays and destination:

$$\boldsymbol{y}_i^q[1] = b_i \boldsymbol{x}_s^q[1] + \boldsymbol{z}_i^q, \; i = 1, \cdots, N. \tag{3.119}$$

$$\boldsymbol{y}_d^q[1] = a \boldsymbol{x}_s^q[1] + \boldsymbol{z}_d^q, \tag{3.120}$$

where $z_i, z_d \sim \mathcal{CN}(0,1)$. During the second slot, relays switch to transmission mode and send to destination the signals $\boldsymbol{x}_i^{n-q}[2]$, $i = 1, \cdots, N$. Simultaneously, the source transmits a new sequence of symbols $\boldsymbol{x}_s^{n-q}[2]$. The superposition of signals is then received at the destination as

$$\boldsymbol{y}_d^{n-q}[2] = a \boldsymbol{x}_s^{n-q}[2] + \sum_{i=1}^{N} c_i \boldsymbol{x}_i^{n-q}[2] + \boldsymbol{z}_d^q, \tag{3.121}$$

Finally, the destination attempts to decode making use of the signal received during the two phases.

Our analysis is constructed upon the same assumptions *(A2)-(A3)* presented in Section 3.2. However, in this case, the power constraint is rewritten as:

$$\frac{1}{n}\left(\sum_{t=1}^{q} \mathrm{E}\left\{|x_s^t[1]|^2\right\} + \sum_{t=1}^{n-q}\left(\mathrm{E}\left\{|x_s^t[2]|^2\right\} + \sum_{i=1}^{N} \mathrm{E}\left\{|x_i^t[2]|^2\right\}\right)\right) \leq P. \tag{3.122}$$

In the sequel, we derive the achievable rates of D&F, PD&F and C&F with such a TD scheme. The main difference with the full-duplex mode is that, now, block-Markov coding is not necessary as relays transmit and receive in different time-slots. This simplifies the practical implementation. We will not derive the achievable of LR, as it can be directly obtained by applying the results in Section 3.6, setting $\kappa = 2$. This is referred to as *amplify-and-forward* (A&F).

Finally, let us remark that we have assumed a *fixed slot* structure. That is, the number of channel uses per time-slot, q and $n-q$, are fixed and designed before transmission. Recently, Kramer has shown that when the relays receive and transmit randomly, the point-to-point achievable rate is increased [73]. We omit this possibility.

3.8.1 Decode-and-Forward

Theorem 3.7 *The AWGN multiple-parallel relay channel achieves the rate*

$$R_{\mathrm{D\&F}} = \max_{1\leq m\leq N} \max_{\alpha,P_1,P_2} \min\left\{\alpha\cdot\mathcal{C}\left(|a|^2\frac{P_1}{N_o}\right) + (1-\alpha)\cdot\mathcal{C}\left(\left(|a|^2+\sum_{i=1}^{m}|c_i|^2\right)\frac{P_2}{N_o}\right),\right.$$
$$\left.\alpha\cdot\mathcal{C}\left(|b_m|^2\frac{P_1}{N_o}\right)\right\}$$
$$\text{s.t. } \alpha P_1 + (1-\alpha)P_2 \leq P,$$

with half-duplex decode-and-forward. The source-relay channels have been ordered as in (3.18).

Remark 3.7 *The optimization above is not jointly convex on P_1, P_2 and α. However, it can be easily transformed into a convex, differentiable one by first changing variables $E_1 = \alpha P_1$ and $E_2 = (1-\alpha)P_2$ and then substituting the minimum of functions by its equivalent epigraph form [52]. To solve such an optimization we propose dual decomposition, constructed through Gradient Projection to obtain the dual function, and subgradient search to minimize it.*

Proof: Let the N relay nodes be ordered as in (3.18), and assume that only the subset $\mathcal{R}_m = \{1,\cdots,m\} \subseteq \mathcal{N}$ is active. The message $\omega \in \left\{1,\cdots,2^{nR}\right\}$ is encoded and transmitted to relay, entirely, during time slot 1. We denote the transmitted signal by:

$$\boldsymbol{x}_s^q[1] = \boldsymbol{s}^q(\omega), \tag{3.123}$$

where we have selected $\boldsymbol{s}^q(\cdot)$ to be a random, Gaussian codebook generated from $s \sim \mathcal{CN}(0, P_1)$. At the end of the slot, all relay nodes in $\mathcal{R}_m$ are able to decode ω *iff*:

$$\begin{aligned} R &\leq \frac{q}{n}\cdot\min_{i\in\mathcal{R}_m} I(s;y_i) \\ &= \frac{q}{n}\cdot\mathcal{C}\left(|b_m|^2\frac{P_1}{N_o}\right), \end{aligned} \tag{3.124}$$

where the equality is due to the ordering in (3.18). Once the message has been decoded, both relays and source re-encode it and send it coherently to the destination. To that end, they transmit as an antenna array with optimum beamforming:

$$\begin{aligned} \boldsymbol{x}_s^{n-q}[2] &= \frac{a^*}{\sqrt{|a|^2+\sum_{i\in\mathcal{R}_m}|c_i|^2}}\boldsymbol{v}^{n-q}(\omega), \\ \boldsymbol{x}_i^{n-q}[2] &= \frac{c_i^*}{\sqrt{|a|^2+\sum_{i\in\mathcal{R}_m}|c_i|^2}}\cdot\boldsymbol{v}^{n-q}(\omega), \ \ \forall\, i\in\mathcal{R}_m, \end{aligned}$$

where $\boldsymbol{v}^{n-q}(\cdot)$ is a random, Gaussian codebook, generated from $v \sim \mathcal{CN}(0, P_2)$. Notice that whenever

$$\frac{q}{n}P_1 + \frac{n-q}{n}P_2 \leq P, \tag{3.125}$$

the power constraint (3.122) is satisfied. At the end of slot 2, the destination attempt to estimate the transmitted message ω from $\boldsymbol{y}_d^q[1]$ and $\boldsymbol{y}_d^{n-q}[2]$. Consider now that codebooks $\boldsymbol{s}^q(\cdot)$ and $\boldsymbol{v}^{n-q}(\cdot)$ form an incremental redundancy code, built up using *e.g.,* parity forwarding [74]. With such a codebook design, the destination error-free decodes the message *iff*:

$$\begin{aligned} R &\leq \frac{q}{n} \cdot I\left(s; y_d[1]\right) + \frac{n-q}{n} \cdot I\left(v; y_d[2]\right) \\ &= \frac{q}{n} \cdot \mathcal{C}\left(|a|^2 \frac{P_1}{N_o}\right) + \frac{n-q}{n} \cdot \mathcal{C}\left(\left(|a|^2 + \sum_{i=1}^{m} |c_i|^2\right) \frac{P_2}{N_o}\right). \end{aligned} \tag{3.126}$$

Defining $\alpha = \frac{q}{n}$, we obtain minimization in Theorem 3.7 from (3.124) and (3.126). Finally, P_1, P_2 as well as the slot duration can be arbitrarily designed. So it can the number of active relays, which concludes the proof. ∎

3.8.2 Partial Decode-and-Forward

Theorem 3.8 *The AWGN multiple-parallel relay channel achieves the rate*

$$\begin{aligned} R_{\text{PD\&F}} = \max_{1 \leq m \leq N} \max_{\alpha, \beta, P_1, P_2} \min \Bigg\{ & \alpha \cdot \mathcal{C}\left(|a|^2 \frac{P_1}{N_o}\right) + (1-\alpha) \cdot \mathcal{C}\left(\left(|a|^2 + \beta \sum_{i=1}^{m} |c_i|^2\right) \frac{P_2}{N_o}\right), \\ & \alpha \cdot \mathcal{C}\left(|b_m|^2 \frac{P_1}{N_o}\right) + (1-\alpha) \cdot \mathcal{C}\left(|a|^2 \frac{(1-\beta) P_2}{N_o}\right) \Bigg\} \\ \text{s.t. } & \alpha P_1 + (1-\alpha) P_2 \leq P \end{aligned}$$

with half-duplex partial decode-and-forward. The source-relay channels have been ordered as (3.18).

Proof: Let the N relay nodes be ordered as in (3.18), and assume that only the subset $\mathcal{R}_m = \{1, \cdots, m\} \subseteq \mathcal{N}$ is active. The source selects message $\omega \in \{1, \cdots, 2^{nR}\}$ for transmission, and splits it into two independent messages $\omega_r \in \{1, \cdots, 2^{nR_r}\}$ and $\omega_d \in \{1, \cdots, 2^{nR_d}\}$, with $R = R_r + R_d$. The first message is sent to relays and destination during slot 1, encoded using a Gaussian codebook, $\boldsymbol{s}^q(\cdot)$, generated from $s \sim \mathcal{CN}(0, P_1)$. The transmitted signal is denoted by:

$$\boldsymbol{x}_s^q[1] = \boldsymbol{s}^q(\omega_r). \tag{3.127}$$

All the relays in $\mathcal{R}_m$ can decode the message at the end of the slot *iff*:

$$R_r \leq \frac{q}{n} \cdot \mathcal{C}\left(|b_m|^2 \frac{P_1}{N_o}\right). \tag{3.128}$$

During slot 2, source and relays re-encode ω_r and send it coherently to destination. In parallel, the source transmits message ω_d using superposition coding. The transmitted signals are thus:

$$\begin{aligned}
\boldsymbol{x}_s^{n-q}[2] &= \boldsymbol{u}^{n-q}(\omega_d) + \frac{a^*}{\sqrt{|a|^2 + \sum_{i\in\mathcal{R}_m}|c_i|^2}}\boldsymbol{v}^{n-q}(\omega_r), \\
\boldsymbol{x}_i^{n-q}[2] &= \frac{c_i^*}{\sqrt{|a|^2 + \sum_{i\in\mathcal{R}_m}|c_i|^2}} \cdot \boldsymbol{v}^{n-q}(\omega_r), \ \ \forall\, i \in \mathcal{R}_m,
\end{aligned}$$

where we have selected $\boldsymbol{u}^{n-q}(\cdot)$ and $\boldsymbol{v}^{n-q}(\cdot)$ to be random, Gaussian codebooks, generated from $u \sim \mathcal{CN}(0, (1-\beta)P_2)$ and $v \sim \mathcal{CN}(0, \beta P_2)$, respectively. At the end of the second slot, the destination first makes use of $\boldsymbol{y}_d^q[1]$ and $\boldsymbol{y}_d^{n-q}[2]$ to estimate ω_r. As previously, we consider that $\boldsymbol{s}^q(\cdot)$ and $\boldsymbol{v}^{n-q}(\cdot)$ form an incremental redundancy code. Hence, decoding is performed with arbitrary small error probability if:

$$\begin{aligned}
R_r &\leq \frac{q}{n} \cdot I(s; y_d[1]) + \frac{n-q}{n} \cdot I(v; y_d[2]) \\
&= \frac{q}{n} \cdot \mathcal{C}\left(|a|^2 \frac{P_1}{N_o}\right) + \frac{n-q}{n} \cdot \mathcal{C}\left(\frac{\left(|a|^2 + \sum_{i=1}^m |c_i|^2\right)\beta P_2}{N_o + |a|^2(1-\beta)P_2}\right).
\end{aligned} \tag{3.129}$$

ω_r can be removed from $\boldsymbol{y}_d^{n-q}[2]$ after it has been decoded, and then ω_d can be estimated *iff*:

$$\begin{aligned}
R_d &\leq \frac{n-q}{n} \cdot I(u; y_d[2]|v) \\
&= \frac{n-q}{n} \cdot \mathcal{C}\left(|a|^2 \frac{(1-\beta)P_2}{N_o}\right).
\end{aligned} \tag{3.130}$$

The transmission rate is $R = R_r + R_d$. Thus, adding (3.129) and (3.130) we obtain the left part of minimization in Theorem 3.8, while by adding (3.128) and (3.130) we obtain the right part. This shows the achievable rate for a given $\mathcal{R}_m$. Finally, notice that we may arbitrarily choose the decoding set $\mathcal{R}_m$ from $\{\mathcal{R}_1, \cdots, \mathcal{R}_N\}$. Moreover, the power and slot duration can be arbitrarily allocated on codes. ∎

As in D&F, the resource optimization is not convex. Moreover, the achievable rate in Theorem 3.8 cannot be transformed into a convex problem. Therefore, we resort to exhaustive search to solve it.

3.8.3 Compress-and-Forward

Theorem 3.9 *The AWGN multiple-parallel relay channel achieves the rate*

$$R_{\text{C\&F}} = \max_{\alpha,\gamma_s^{(1)},\gamma_s^{(2)},\gamma_{1:N},\rho_{1:N}} \alpha\cdot\mathcal{C}\left(\left(\frac{|a|^2}{N_o}+\sum_{i=1}^N\frac{|b_i|^2}{N_o+\rho_i}\right)\gamma_s^{(1)}\right)+(1-\alpha)\cdot\mathcal{C}\left(\frac{|a|^2\gamma_s^{(2)}}{N_o}\right)$$

$$\text{s.t. } \mathcal{C}\left(\frac{\gamma_s^{(1)}\sum_{u\in\mathcal{G}}\frac{|b_u|^2}{N_o+\rho_u}}{1+\gamma_s^{(1)}\left(\frac{|a|^2}{N_o}+\sum_{j\notin\mathcal{G}}\frac{|b_j|^2}{N_o+\rho_j}\right)}\right)+\sum_{u\in\mathcal{G}}\mathcal{C}\left(\frac{N_o}{\rho_u}\right)\leq$$

$$\frac{1-\alpha}{\alpha}\cdot\mathcal{C}\left(\sum_{u\in\mathcal{G}}\frac{|c_u|^2\gamma_u}{N_o+|a|^2\gamma_s^{(2)}}\right), \forall\mathcal{G}\in\mathcal{N}.$$

$$\alpha\gamma_s^{(1)}+(1-\alpha)\left(\gamma_s^{(2)}+\sum_{i=1}^N\gamma_i\right)\leq P,$$

with half-duplex D-WZ compress-and-forward relaying.

Proof: Let the source select message $\omega\in\left\{1,\cdots,2^{nR}\right\}$, and split it into two independent messages $\omega_r\in\left\{1,\cdots,2^{nR_r}\right\}$ and $\omega_d\in\left\{1,\cdots,2^{nR_d}\right\}$, with $R=R_r+R_d$. During slot 1, the source transmits ω_r to relays and destination; during slot 2, it transmits message ω_d. We denote the transmitted signals by:

$$\begin{aligned}\boldsymbol{x}_s^q[1] &= \boldsymbol{s}^q(\omega_r) \\ \boldsymbol{x}_s^{n-q}[2] &= \boldsymbol{u}^{n-q}(\omega_d),\end{aligned} \tag{3.131}$$

where we have selected $\boldsymbol{s}^q(\cdot)$ and $\boldsymbol{u}^{n-q}(\cdot)$ to be random, Gaussian codebooks, generated *i.i.d.* from $s\sim\mathcal{CN}\left(0,\gamma_s^{(1)}\right)$ and $u\sim\mathcal{CN}\left(0,\gamma_s^{(2)}\right)$, respectively.

At the end of the first slot, the received signals at the relays $\boldsymbol{y}_i^q[1], i=1,\cdots,N$ are compressed using a D-WZ code as that in Definition 3.4 and Proposition 3.1. As a results, $\boldsymbol{y}_i^q[1]$ are mapped using compression functions

$$f_q^i:\mathcal{Y}_i^q\rightarrow\left\{1,\cdots,2^{n\phi_i}\right\}, \tag{3.132}$$

where ϕ_i is the compression rate of relay i. On the following time slot, the relays send to the destination (via its MAC channel) indexes $s_i=f_\kappa^i\left(\boldsymbol{y}_i^q[1]\right)$, $i=1,\cdots,N$. Indices are mapped onto multi-access channel codebooks $\boldsymbol{v}_i^{n-q}(\cdot)$ as follows:

$$\boldsymbol{x}_i^{n-q}[2]=\boldsymbol{v}_i^{n-q}(s_i),\ i=1,\cdots,N. \tag{3.133}$$

We select the MAC codebooks to be Gaussian, generated from $v_i \sim \mathcal{CN}(0, \gamma_i)$, $i = 1, \cdots, N$. Notice that the power constraint is satisfied if

$$\frac{q}{n}\gamma_s^{(1)} + \frac{n-q}{n}\left(\gamma_s^{(2)} + \sum_{i=1}^{N}\gamma_i\right) \leq P, \tag{3.134}$$

The received signal at the destination during the second time slot is (3.121):

$$\boldsymbol{y}_d^{n-q}[2] = a\boldsymbol{u}^{n-q}(\omega_d) + \sum_{i=1}^{N} b_i \boldsymbol{v}_i^{n-q}(s_i) + \boldsymbol{z}_d^{\kappa}. \tag{3.135}$$

Let us define the decoding at the destination at the end of slot 2. First, it estimates the transmitted indexes $s_{1:N}$. This can be done with arbitrary small error probability *iff* the transmission rates ϕ_i lie within the capacity region, given by:

$$\sum_{u\in\mathcal{G}}\phi_u \leq \frac{n-q}{n}\cdot\mathcal{C}\left(\frac{\sum_{u\in\mathcal{G}}|c_u|^2\gamma_u}{N_o + |a|^2\gamma_s^{(2)}}\right), \forall\mathcal{G}\subseteq\mathcal{U}. \tag{3.136}$$

Once the indexes $s_{1:N}$ have been estimated, their contribution on $\boldsymbol{y}_d^{n-q}[2]$ can be removed:

$$\begin{aligned}\boldsymbol{y}_d^{\prime n-q}[2] &= \boldsymbol{y}_d^{n-q}[2] - \sum_{i=1}^{N} c_i \cdot \boldsymbol{x}_i^{n-q}[2] \\ &= a\boldsymbol{u}^{n-q}(\omega_d) + \boldsymbol{z}_d^{n-q}.\end{aligned} \tag{3.137}$$

Hence, ω_d can be correctly decoded *iff*:

$$R_d \leq \frac{n-q}{n}\cdot\mathcal{C}\left(\frac{|a|^2\gamma_s^{(2)}}{N_o}\right). \tag{3.138}$$

Afterwards, the destination decompresses indexes $s_{1:N}$ using $\boldsymbol{y}_d^q[1]$ as side information. the decompression is carried out by means of the demapping function (Proposition 3.1)

$$g_q : \left\{1, \cdots, 2^{n\phi_1}\right\} \times \cdots \times \left\{1, \cdots, 2^{n\phi_N}\right\} \times \mathcal{Y}_d^q \to \hat{\mathcal{Y}}_{1:N}^q. \tag{3.139}$$

The de-mapped vectors $\hat{\boldsymbol{y}}_{1:N}^q = g(s_1, \cdots, s_N, \boldsymbol{y}_d^q[1])$ are then used, along with $\boldsymbol{y}_d^q[1]$, to decode message ω_r, which is correctly decoded *iff* [67]:

$$R_r \leq \frac{q}{n}\cdot I(s; y_d[1], \hat{y}_1, \cdots, \hat{y}_N), \tag{3.140}$$

where the inequality follows from (3.41) in Proposition 3.1. However, (3.140) only holds for compression rates satisfying the set of constraints (3.42):

$$\frac{q}{n}\cdot I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{\prime}, \hat{\boldsymbol{y}}_{\mathcal{G}}^c\right) \leq \sum_{u\in\mathcal{G}}\phi_u \qquad \forall\mathcal{G}\subseteq\mathcal{N}. \tag{3.141}$$

Therefore, taking into account the constraints (3.141) and (3.136) and the rate (3.140), the maximum achievable rate $R = R_r + R_d$ with C&F is:

$$R_{\text{C\&F}} = \max_{\prod_{i=1}^{N} p(\hat{y}_i|y_i), \gamma_s, \gamma_{1:N}, n} \frac{q}{n} \cdot I\left(s; y_d^{'}, \hat{\boldsymbol{y}}_{1:N}\right) + \frac{n-q}{n} \cdot \mathcal{C}\left(\frac{|a|^2\gamma_s^{(2)}}{N_o}\right) \tag{3.142}$$

$$\text{s.t. } \frac{q}{n} \cdot I\left(\boldsymbol{y}_{\mathcal{G}}; \hat{\boldsymbol{y}}_{\mathcal{G}}|y_d^{'}, \hat{\boldsymbol{y}}_{\mathcal{G}}^{c}\right) \leq \frac{n-q}{n} \cdot \mathcal{C}\left(\frac{\sum_{u\in\mathcal{G}} |c_u|^2\gamma_u}{N_o + |a|^2\gamma_s}\right), \ \forall \mathcal{G} \subseteq \mathcal{U},$$

$$\frac{q}{n}\gamma_s^{(1)} + \frac{n-q}{n}\left(\gamma_s^{(2)} + \sum_{i=1}^{N}\gamma_i\right) \leq P.$$

We may now define $\alpha = \frac{q}{n}$, and apply the same arguments in (3.55)-(3.56) to conclude the proof. ■

3.8.4 Numerical Results and Comparison

In this subsection, we briefly compare the performance of techniques with and without half-duplex (HD) constraint. Unfortunately, we are not able to compare the achievable rates with C&F due to the non-convexity of the optimization in Theorem 3.9. For simulation purposes, the setup and channel distribution are the same as those in Section 3.7.

Fig. 3.9 compares full-duplex (FD) relaying versus HD versus de source-relay distance d. We assume $N = 2$ relays and transmit $\text{SNR} = P/N_o$ of 5 dB. We notice the following:

- All relaying techniques substantially get worse performance, penalized by the transmit-receive multiplexing at the relays.
- The one with greatest penalization is D&F. In particular, it losses the most of its gains when relays are close to the source. As an example, for $d = 0.4$, the HD constraint induces a 0.8 bps/Hz loss.
- The same analysis holds for PD&F. However, with the HD constraint, PD&F outperforms D&F within a wider range ($0.45 \leq d \leq 1$) than with FD: $0.6 \leq d \leq 1$. This is one of the main conclusions of this section: *PD&F is a more interesting approach for HD relaying than for FD relaying*. This is explained by noting that the message directly transmitted to destination induces interferences onto relays when working on full-duplex, while it does not when they work on half-duplex. Likewise, the achievable rate with HD PD&F

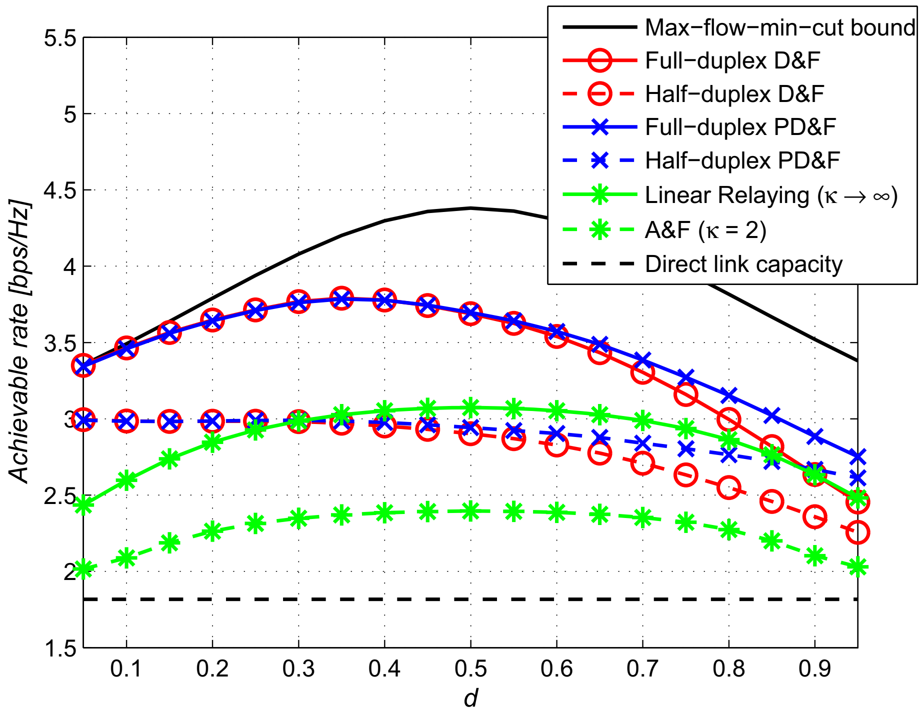

Figure 3.9: Comparison of the half-duplex and full-duplex achievable rates versus the source-relay distance. We consider transmit SNR = 5dB and $N = 2$ relays. Wireless channels are Rayleigh faded.

is almost flat with respect to the source-relay distance d. This poses PD&F as a very good alternative when relays have random, uncontrolled distance to the source.

- A&F (that is, LR with $\kappa = 2$) reduces significantly its performance compared to LR. In particular, with $N = 2$, A&F never outperforms D&F, not even for relays close to the destination.

Fig. 3.10 depicts the achievable rates versus the number of relays, for $d = 0.5$. From the plot, we infer the following: *i*) with D&F, the difference between the FD achievable rate and HD achievable rate augments when increasing the number of relays; that is, FD and HD D&F scale differently. *ii*) the A&F achievable rate is shown to be almost half of that achieved with LR (see blue curve). Therefore, unlike other techniques where losses are lower, the HD constraint penalizes amplify-based relaying with half of its rate gain. In consequence, HD makes amplifying less worthy. As an example, LR outperforms FD D&F with $N > 7$ relays, while A&F needs $N > 17$ to outperform HD D&F.

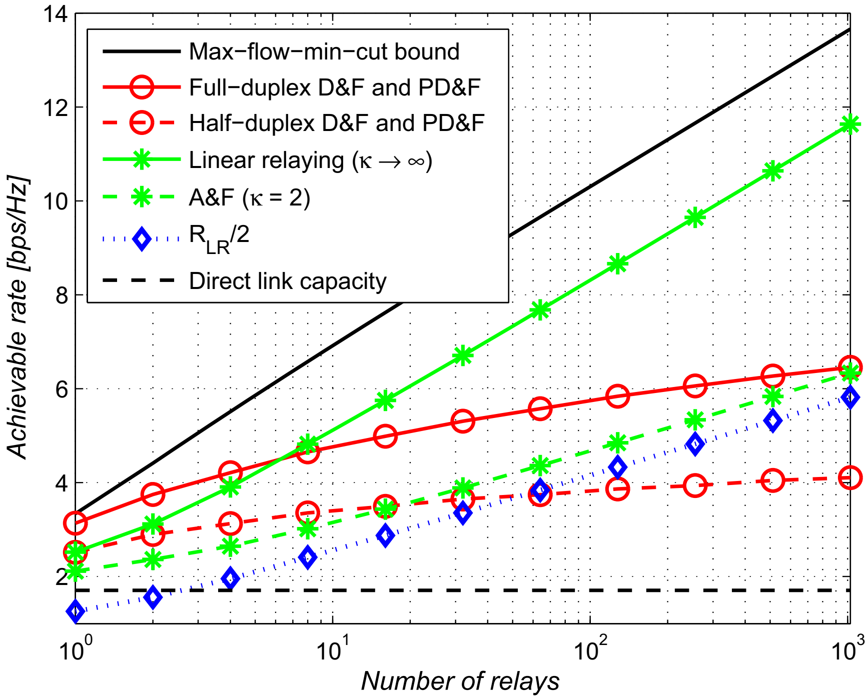

Figure 3.10: Comparison of the half-duplex and full-duplex achievable rates versus the number of relays N. We consider transmit SNR = 5dB and $d = 0.5$. Wireless channels are Rayleigh faded.

3.9 Conclusions

This chapter has studied the multiple-parallel relay channel (MPRC), where a single source communicates to a single destination with the aid of N parallel relay nodes. The capacity of the channel has been upper bounded using the *max-flow-min-cut* Theorem. The obtained upper bound turned out to be reciprocal, and was shown to scale as $\mathcal{C}\left(\frac{N}{2}\frac{P}{N_o}\right)$ under *time-invariant* (unitary-mean) Rayleigh fading.

Besides, the capacity of the channel has been lower bounded by means of the achievable rates with: D&F, two-level PD&F, C&F and LR. The first two consisted of relay nodes totally or partially decoding the source's signal. Both achievable rates were shown to perform equally for low source-relay distances, being capacity achieving. Likewise, for large number of relays, both follow the same scaling law: $\mathcal{C}\left(2\cdot W_0\left(\frac{\sqrt{N}}{2}\right)\cdot\frac{P}{N_o}\right)$. Such a law is shown to diverge from the upper bound, making both techniques clearly suboptimal. We explained this fact resorting to the source-relay broadcast limitation.

With C&F, the relay nodes compress their receiving signals using a Distributed Wyner-Ziv (D-WZ) code, and then send them to the destination. The destination, in turn, utilizes the relay signals within a coherent detector to estimate the source's message. The achievable rate of this technique was presented in terms of a non-convex optimization, which turned out unsolvable for us. Hence, we proposed a computable upper bound as a benchmark for this technique. Moreover, we showed that the achievable rate with C&F scales as $\mathcal{C}\left(\frac{P}{N_o}\log_2 N\right)$, in this case due to the relay-destination MAC limitation.

Finally, we studied LR, which consisted of relay nodes transmitting, on every channel use, a linear combination of previously received signals. The optimum source temporal covariance for this technique was derived, and suboptimum linear relaying matrices proposed. Furthermore, we showed that the achievable rate scales as $\mathcal{C}\left(\frac{N\cdot P}{N_o}\right)$, in the same manner as the upper bound and unlike all previous techniques. Hence, it is the only known technique that seize all the beamforming gain of the system.

Numerical results have shown that:

1. D&F is a very competitive approach (and in some cases capacity-achieving) for low number of relays and short/mid source-relay distances. On the other hand, it dramatically

losses performance with high number of relays and/or relays close to the destination.

2. PD&F does not produce, in general, rate gains with respect to D&F. In particular, for moderate number of relays $N > 5$ and/or relays closer to the source than to the destination, it is useless. Only for low number of relays and/or relays extremely close to destination, it is an interesting approach.

3. Unlike D&F, C&F is not able to mimic a receive antenna array unless relays are extremely close to destination. Indeed, even for short relay-destination distance, it presents significant losses with respect to the *max-flow-min-cut* bound. Furthermore, its behavior when increasing the number of relays is even worst than that of $D\&F$: it follows the scaling law more rapidly.

4. LR is, simultaneously, the head and tails of the coin. It performs extremely well when moderate/high number of relays are available. In such a setup, it rapidly shows the $\log N$ scale, running parallel to the capacity upper bound. On the other hand, with low number of relays, it performs extremely poor due to the noise amplification, its main drawback. Finally, it is reciprocal with respect to source and destination.

In this chapter, the extension of results to half-duplex (HD) operation were also performed. The numerical analysis showed that:

1. All relaying techniques substantially get worse performance with half-duplex constraint. Among all, D&F is the one with higher penalization. It losses great part of its gains for short/mid source-relay distances.

2. PD&F produces greater gains with respect to D&F when operating in HD. This is because the message directly sent form source to destination does not induces interference on the relays.

3. A&F (*i.e.*, half-duplex linear relaying) looses half of the gains obtained with full-duplex LR, independently of the number of relays and/or source-relay distance.

Chapter 4

Multi-Access Channel with Multiple-Parallel Relays

4.1 Introduction

Relaying has been placed at the forefront of the struggle against multi-path fading, mainly due to its ability to provide single-antenna communications with *extra* spatial diversity [14]. However, it has plenty more features besides; among the most significants, its capability to reduce the harmful impact of static channel impairments *i.e.* path-loss and shadowing [13, 57]. In this sense, *relaying* is able enlarge the coverage area of access networks or, equivalently, to increase their throughput per square meter [75]. However, is this worthwhile in multi-user access?. Or, more specifically: *i*) how much can the capacity region of the MAC be enlarged using multiple relays?. *ii*) Does multiuser diversity undermine relaying diversity?

Let us extend the results in Chapter 3 to multiple sources, trying to address those questions.

4.1.1 Overview

In this chapter, the Gaussian MAC with multiple-parallel relays (see Fig. 4.1) is studied. As previously, we consider single-antenna sources, relays and destination, in a *time-invariant* and *memory-less* channel. Also, we assume that transmit and receive CSI are available at the sources and destination, respectively, and all network nodes operate in the same frequency band. Finally, relaying is full-duplex.

We define two main goals: 1) to study the capacity region of the channel. 2) To analyze its sum-capacity when the number of sources U grows to infinity. In other words, to estimate the impact of multiuser diversity into access networks with relays.

Aiming at keeping exposition simple, we restrain the analysis of the capacity region to the two-user MAC. Our contributions are organized as follows:

- First, we provide the infinite-letter characterization of the capacity region, which is non-computable in practice. In order to bound it, we derive a computable outer region using the *max-flow-min-cut* Theorem [21, Theorem 14.10.1]. Then, we provide an upper bound on the asymptotic[1] sum-capacity of the channel under Rayleigh-distributed fading. We show that a greater sum-rate can be obtained even at the asymptote, which encourages to think that multiuser diversity does not undermine the impact of relaying.

- Next, we derive the achievable rate region of the channel with D&F, as an extension of the single-relay result in [76]. Additionally, we show that, under Rayleigh-fading, the achievable sum-rate satisfies:

$$\mathcal{SR}_{\text{D\&F}}(U) - \mathcal{C}\left(U\frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{4.1}$$

  Therefore, asymptotically, D&F is not capable to increase the sum-capacity of the MAC without relays. As explained in the sequel, such a result is due to the distributed channel hardening effect at the relays' received signal.

- Linear relaying is studied next. The rate region is derived using theory for vector MACs and characterized using the sum-rate (SR) and weighted sum-rate (WSR) optimization. Both optimizations are shown to be convex on the transmitter-side signalling, given a

[1] Asymptotic performance in the number of users, keeping the set of relays fixed.

fixed set of linear functions at the relays. We solve the optimizations using iterative algorithms, namely *block-coordinate* approach for the SR, and *gradient projection* for the WSR. After designing the sources' signals, the optimization of the linear functions at the relays is not convex. Therefore, we propose a suboptimum approach.

- C&F with distributed Wyner-Ziv coding at the relays is studied last. As with the previous technique, the rate region is characterized by means of the WSR optimization. The optimization, though, is not convex and has $2 \cdot N + 2$ optimization variables; numerical resolution is thus unfeasible. Therefore, we propose a computable upper bound on it that will serve as benchmark. Finally, an upper bound on the asymptotic performance of C&F (for the number of users growing to infinity) is found. We show that, unlike D&F, rate gains can be obtained even at the asymptote.

The analysis of PD&F is omitted as it was shown in the previous chapter to perform exactly as D&F for moderate/high number of relays. The rest of the chapter is organized as follows: Section 4.2 presents the channel model and the capacity outer region. D&F is analyzed in Section 4.3, LR in Section 4.4, and C&F in Section 4.5. Finally, Section 4.6 shows numerical results and Section 4.7 enumerates conclusions.

4.2 Channel Model

The multi-access channel with multiple-parallel relays (MPR-MAC) is a channel in which two sources, s_1 and s_2, communicate simultaneously to a single destination d with the aid of a set $\mathcal{N} = \{1, \cdots, N\}$ of parallel relay nodes (see Fig. 4.1). All network nodes transmit/receive scalar signals, and wireless channels are *time-invariant*, *memoryless*, modeled using a complex scalar. We denote by a_u the complex channel between source u and destination, while $b_{u,i}$ and c_i stand for source u to relay i, and relay i to destination channels, respectively. As for the MPRC in Chapter 3, it is assumed that relays transmit directionally to destination and are not able to communicate among them.

We denote by $\boldsymbol{x}_u^n = \{x_u^t\}_{t=1}^n \in \mathbb{C}^n$ the signal transmitted by the source $u = 1, 2$, where x_u^t is the transmitted symbol during channel use t and n is the codeword length. The received signals

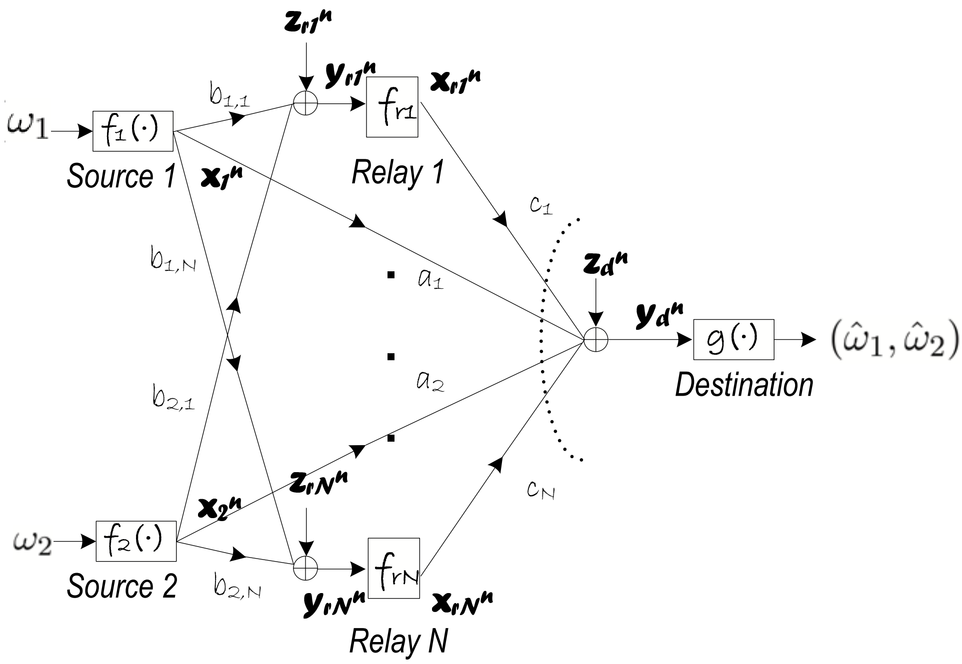

Figure 4.1: Gaussian MAC with multiple-parallel relays.

at the relay nodes are

$$\boldsymbol{y}_{r_i}^n = \sum_{u=1}^{2} b_{u,i} \cdot \boldsymbol{x}_u^n + \boldsymbol{z}_{r_i}^n, \quad i = 1, \cdots, N \tag{4.2}$$

where $z_{r_i} \sim \mathcal{CN}(0, N_o)$ is additive white Gaussian noise (AWGN). The signals transmitted by the relays are $\boldsymbol{x}_{r_i}^n = \left\{x_{r_i}^t\right\}_{t=1}^n \in \mathbb{C}^n$, $i = 1, \cdots, N$, and are defined by means of causal relaying functions: $x_{r_i}^t = f_{r_i}\left(y_{r_i}^1, \cdots, y_{r_i}^{t-1}\right)$. Accordingly, the destination node receives

$$\boldsymbol{y}_d^n = \sum_{u=1}^{2} a_u \cdot \boldsymbol{x}_u^n + \sum_{i=1}^{N} c_i \cdot \boldsymbol{x}_{r_i}^n + \boldsymbol{z}_d^n, \tag{4.3}$$

with $z_d \sim \mathcal{CN}(0, N_o)$. As for the previous chapter, we assume the following:

(*A1*). *Full-duplex operation*: relays transmit and receive simultaneously in the same frequency band. This can be implemented using different, uncoupled antennas for transmission and reception.

(*A2*). Transmit *channel state information* (CSI) and receive CSI at the sources and destination, respectively. Channel awareness includes source-destination, source-relay and relay-destination channels, and can be obtained (and fedback) during a setup phase.

(*A3*). The total power used in the communication of user u is constrained to P_u. This can be arbitrarily distributed among the user and relays. Such a constraint aims at a fair comparison with the no-relay MAC with per-user constraint P_u.

4.2.1 Preliminaries

In order to present rate results, we define the following:

Definition 4.1 *A* $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ *code for the MPR-MAC is defined by:*

- *two sets of messages* $\mathcal{W}_u = \left\{1, \cdots, 2^{nR_u}\right\}$*,* $u = 1, 2$*, two signal spaces* $\mathcal{X}_u$*,* $u = 1, 2$ *and two source encoding functions*

$$f_u : \mathcal{W}_u \rightarrow \mathcal{X}_u^n, \; u = 1, 2, \tag{4.4}$$

- N *signal spaces* $\mathcal{X}_i, \; i = 1, \cdots, N$ *and* N *causal relay functions*

$$f_{r_i} : \mathcal{Y}_i^n \rightarrow \mathcal{X}_i^n, \; i = 1, \cdots, N, \tag{4.5}$$

- *a decoding function* $g : \mathcal{Y}_d^n \rightarrow \mathcal{W}_1 \times \mathcal{W}_2$.

Definition 4.2 *A rate duple* (R_1, R_2) *is achievable if there exists a sequence of codes* $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ *for which* $\lim_{n\rightarrow\infty} P_e^n = 0$*, where*

$$P_e^n = \frac{1}{2^{nR_1} 2^{nR_2}} \sum_{\omega_1, \omega_2} \Pr\left\{g\left(\boldsymbol{y}_d^n\right) \neq (\omega_1, \omega_2) | (\omega_1, \omega_2) \text{ was sent}\right\}. \tag{4.6}$$

Definition 4.3 *The capacity region of the MPR-MAC is the closure of all rates duples that are achievable. It is characterized by means of the infinite-letter definition:*

$$\mathcal{C} = \lim_{n\rightarrow\infty} \operatorname{coh}\left(\bigcup_{p_{\boldsymbol{x}_1^n} p_{\boldsymbol{x}_2^n}, \left\{f_{r_i}\right\}_{i=1,\cdots,N}} \left\{ R_{1,2} : \begin{array}{c} R_1 \leq \frac{1}{n} \cdot I\left(\boldsymbol{x}_1^n; \boldsymbol{y}_d^n | \boldsymbol{x}_2^n\right) \\ R_2 \leq \frac{1}{n} \cdot I\left(\boldsymbol{x}_2^n; \boldsymbol{y}_d^n | \boldsymbol{x}_1^n\right) \\ R_1 + R_2 \leq \frac{1}{n} \cdot I\left(\boldsymbol{x}_1^n, \boldsymbol{x}_2^n; \boldsymbol{y}_d^n\right) \end{array} \right\}\right), \tag{4.7}$$

where the union is taken over the sources distributions and relaying functions that satisfy the power constraint (A3) and the causal relaying in Definition 4.1, respectively.

The infinite-letter characterization (4.7) has been obtained using arguments in [57, Theorem 2.1]: Fano's inequality for the weak converse, and random coding/joint typical decoding for achievability. Unfortunately, deriving a computable, single-letter expression is still an open problem. In this section, we present an outer region.

4.2.2 Outer Region

As for the previous chapter, we present a necessary condition that any rate duple (R_1, R_2) must satisfy in order to be achievable. The set of rates that satisfy the condition defines an outer region on the capacity region of the channel.

Outer Region: All rate duples belonging to the capacity region (4.7) satisfy:

$$R_1 \leq \mathcal{C}\left(\left(|a_1|^2 + \frac{\sum_{i=1}^N |b_{1,i}|^2 \sum_{i=1}^N |c_i|^2}{|a_1|^2 + \sum_{i=1}^N |b_{1,i}|^2 + \sum_{i=1}^N |c_i|^2}\right)\frac{P_1}{N_o}\right) \quad (4.8)$$

$$R_2 \leq \mathcal{C}\left(\left(|a_2|^2 + \frac{\sum_{i=1}^N |b_{2,i}|^2 \sum_{i=1}^N |c_i|^2}{|a_2|^2 + \sum_{i=1}^N |b_{2,i}|^2 + \sum_{i=1}^N |c_i|^2}\right)\frac{P_2}{N_o}\right) \quad (4.9)$$

$$R_1 + R_2 \leq \max_{0 \leq \rho_1, \rho_2 < 1} \min \left\{ \mathcal{C}\left(\sum_{u=1}^2 \left(|a_u|^2 + \rho_u \sum_{i=1}^N |c_i|^2\right)\frac{P_u}{N_o}\right), \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^2 (1-\rho_u)\frac{P_u}{N_o}\boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right)\right\} \quad (4.10)$$

where $\boldsymbol{h}_u = [a_u, b_{u,1}, \cdots, b_{u,N}]^T$, $u = 1, 2$.

Proof: Achievable rates for user 1 and user 2 must satisfy, separately, the single-user upper bound presented in Section 3.2.2. Hence, equations (4.8) and (4.9) hold. In order to demonstrate (4.10), we again resort to the *max-flow-min-cut* Theorem [21, Theorem 14.10.1]. In particular, considering the multi-access and broadcast cuts, it is possible to bound:

$$R_1 + R_2 \leq \max_{p\left(\boldsymbol{x}_{r_{1:N}}|x_1,x_2\right)p(x_1)p(x_2)} \min\left\{I\left(x_1, x_2, \boldsymbol{x}_{r_{1:N}}; y_d\right), I\left(x_1, x_2; y_d, \boldsymbol{y}_{r_{1:N}}|\boldsymbol{x}_{r_{1:N}}\right)\right\}, \quad (4.11)$$

where the factorization of the distribution comes from non-cooperative nature of the sources' signalling; that is, x_1 and x_2 are independent. As for the single-source case, we can easily show that: *i*) a Gaussian distribution is optimal, *ii*) full correlation among the relays' signals is also optimal, as it maximizes both sides of the minimization simultaneously, *i.e.,*

$$\left|\frac{\mathrm{E}\left\{x_{r_i}x_{r_j}^*\right\}}{\sqrt{\mathrm{E}\left\{|x_{r_i}|^2\right\}}\sqrt{\mathrm{E}\left\{|x_{r_j}|^2\right\}}}\right| = 1,\ i,j\in\mathcal{N}. \tag{4.12}$$

Hence, defining

$$\rho_{u,r} = \left|\frac{\mathrm{E}\left\{x_u x_{r_j}^*\right\}}{\sqrt{\mathrm{E}\left\{|x_u|^2\right\}}\sqrt{\mathrm{E}\left\{|x_{r_j}|^2\right\}}}\right|,\ u=1,2,\ \ \forall j\in\mathcal{N} \tag{4.13}$$

as the source-relays correlation, and taking *i*) and *ii*) into account, we evaluate the broadcast cut as:

$$\begin{aligned} I\left(x_1,x_2;y_d,\boldsymbol{y}_{r_{1:N}}|\boldsymbol{x}_{r_{1:N}}\right) &= \log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\frac{\mathrm{E}\left\{|x_u|^2|\boldsymbol{x}_{1:N}\right\}}{N_o}\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right) \\ &= \log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\frac{\left(1-\rho_{u,r}^2\right)\mathrm{E}\left\{|x_u|^2\right\}}{N_o}\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right) \\ &= \log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\left(1-\rho_u\right)\frac{P_u}{N_o}\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right), \end{aligned} \tag{4.14}$$

where $\boldsymbol{h}_u$ is defined above, and we have set

$$\left(1-\rho_u\right)P_u = \left(1-\rho_{u,r}^2\right)\mathrm{E}\left\{|x_u|^2\right\},\ u=1,2. \tag{4.15}$$

Furthermore, for the multiple access cut it is possible to compute:

$$\begin{aligned} I\left(x_1,x_2,\boldsymbol{x}_{r_{1:N}};y_d\right) &= I\left(x_1,x_2;y_d|\boldsymbol{x}_{r_{1:N}}\right)+I\left(\boldsymbol{x}_{r_{1:N}};y_d\right) \\ &= \mathcal{C}\left(\sum_{u=1}^{2}\frac{|a_u|^2\left(1-\rho_{u,r}^2\right)\mathrm{E}\left\{|x_u|^2\right\}}{N_o}\right)+I\left(\boldsymbol{x}_{r_{1:N}};y_d\right) \\ &= \mathcal{C}\left(\sum_{u=1}^{2}\frac{|a_u|^2\left(1-\rho_u\right)P_u}{N_o}\right)+I\left(\boldsymbol{x}_{r_{1:N}};y_d\right). \end{aligned} \tag{4.16}$$

Following equivalent arguments to those in (3.15), and considering the power constraint (*A3*), we can bound

$$I\left(\boldsymbol{x}_{r_{1:N}};y_d\right) \le \mathcal{C}\left(\frac{\sum_{u=1}^{2}\left(|a_u|^2+\sum_{i=1}^{N}|c_i|^2\right)\rho_u P_u}{N_o+\sum_{u=1}^{2}|a_u|^2\left(1-\rho_u\right)P_u}\right), \tag{4.17}$$

which plugged into (4.16), allows us to obtain:

$$I\left(x_1,x_2;y_d,\boldsymbol{y}_{r_{1:N}}|\boldsymbol{x}_{r_{1:N}}\right) \le \mathcal{C}\left(\sum_{u=1}^{2}\left(|a_u|^2+\rho_u\sum_{i=1}^{N}|c_i|^2\right)\frac{P_u}{N_o}\right). \tag{4.18}$$

Finally, to obtain (4.10) the minimum of (4.14) and (4.18) must be maximized over ρ_1,ρ_2, which concludes the proof. ■

Notice that, as for the single-user case, the outer region also bounds the capacity region of MACs with inter-relay connectivity. Indeed, we did not take any assumption on the connectivity among relays.

4.2.3 Multiuser Diversity and Relaying

Throughout the chapter, we consider the MPR-MAC composed uniquely of two users. However, results can be extended to $U > 2$ using standard arguments in [21, Section 14.3.5]. Specifically, the sum-capacity bound (4.10) generalizes for more than two sources as:

$$\sum_{u=1}^{U} R_u \leq \mathcal{SR}_{\text{ub}}(U) \triangleq \max_{0\leq\rho_{1:U}<1} \min\left\{\mathcal{C}\left(\sum_{u=1}^{U}\left(|a_u|^2+\rho_u\sum_{i=1}^{N}|c_i|^2\right)\frac{P_u}{N_o}\right),\right. \left.\log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{U}(1-\rho_u)\frac{P_u}{N_o}\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right)\right\} \quad (4.19)$$

In this subsection, we derive the scaling law of such a sum-capacity bound for number of users growing to ∞. In other words, we analyze whether multiuser diversity reduces the potential gains of relaying or not. In the analysis it is assumed that, when increasing the number of users, the set of relays remains unaltered, as well as the relay-destination channels.

Consider unitary-mean, Rayleigh fading so that $a_u, b_{u,i} \sim \mathcal{CN}(0,1)$. Also, for notational simplicity, assume $P_u = P$, $u = 1, \cdots, U$. Then, the following convergence in probability can be proven.

Theorem 4.1 *Consider the MPR-MAC with a fixed set of relays $\mathcal{N}$. Then, for $U \to \infty$:*

$$\mathcal{SR}_{\text{ub}}(U) - \max_{\xi\in[0,U)}\min\left\{\mathcal{C}\left(\frac{P}{N_o}\left(U+\sum_{i=1}^{N}|c_i|^2\xi\right)\right),(N+1)\,\mathcal{C}\left(\frac{P}{N_o}(U-\xi)\right)\right\}\xrightarrow{P}0.$$

Proof: Let us first focus on the definition of $\mathcal{SR}_{\text{ub}}(U)$ in (4.19), and apply Corollary 2.3 to derive that, for $U \to \infty$:

$$\mathcal{C}\left(\sum_{u=1}^{U}\left(|a_u|^2+\rho_u\sum_{i=1}^{N}|c_i|^2\right)\frac{P}{N_o}\right)-\mathcal{C}\left(\frac{P}{N_o}\left(U+\sum_{i=1}^{N}|c_i|^2\sum_{u=1}^{U}\rho_u\right)\right)\xrightarrow{P}0. \quad (4.20)$$

For the other part of the minimization, it is possible to compute:

$$\begin{aligned}\log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{U} (1-\rho_u)\frac{P}{N_o}\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right) &= \log_2 \det\left(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \\ &= \sum_{i=1}^{N+1} \mathcal{C}\left(\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)\right),\end{aligned} \tag{4.21}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_U]$, $\boldsymbol{Q} = \text{diag}\left(\frac{P}{N_o}(1-\rho_1), \cdots, \frac{P}{N_o}(1-\rho_U)\right)$ and $\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)$ is the i^{th} eigenvalue of $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H$. Notice that all the entries of $\boldsymbol{H}$ are *i.i.d.*, zero-mean, Gaussian with unit variance. Therefore, we can apply [77, Corollary 1] in order to show that, for $U \to \infty$:

$$\lambda_i\left(\frac{\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H}{U}\right) - \frac{\text{tr}\{\boldsymbol{Q}\}}{U} \xrightarrow{P} 0. \quad i = 1, \cdots, N+1. \tag{4.22}$$

Such a convergence holds since the elements of diagonal matrix $\boldsymbol{Q}$ are bounded. Notice now that $\lambda_i\left(\frac{\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H}{U}\right) = \frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)}{U}$. Therefore,

$$\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)}{U} - \frac{\text{tr}\{\boldsymbol{Q}\}}{U} \xrightarrow{P} 0. \quad i = 1, \cdots, N+1. \tag{4.23}$$

This can be transformed without modifying convergence

$$\frac{1+\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)}{U} - \frac{1+\text{tr}\{\boldsymbol{Q}\}}{U} \xrightarrow{P} 0. \; i = 1, \cdots, N+1. \tag{4.24}$$

Consider the computation of the trace:

$$\text{tr}\{\boldsymbol{Q}\} = \frac{P}{N_o}\sum_{u=1}^{U}(1-\rho_u) = \frac{P}{N_o}\left(U - \sum_{u=1}^{U}\rho_u\right). \tag{4.25}$$

It is clear that, being $\rho_u < 1, \forall u$, it satisfies

$$\text{tr}\{\boldsymbol{Q}\} \geq U\frac{P}{N_o}\cdot\left(\min_u\{1-\rho_u\}\right) > 0, \;\; \forall U \tag{4.26}$$

Accordingly, it is possible to reproduce steps (2.83)-(2.89) in Corollary 2.3, to derive that for $U \to \infty$:

$$\mathcal{C}\left(\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)\right) - \mathcal{C}\left(\text{tr}\{\boldsymbol{Q}\}\right) \xrightarrow{P} 0. \; i = 1, \cdots, N+1. \tag{4.27}$$

On such a convergence, we can introduce the value of the trace (4.25) to claim

$$\mathcal{C}\left(\lambda_i\left(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right)\right) - \mathcal{C}\left(\frac{P}{N_o}\left(U - \sum_{u=1}^{U}\rho_u\right)\right) \xrightarrow{P} 0. \; i = 1, \cdots, N+1. \tag{4.28}$$

Finally, it is possible to plug (4.28) into (4.21) to derive:

$$\log_2 \det \left(\boldsymbol{I} + \sum_{u=1}^{U} (1-\rho_u) \frac{P}{N_o} \boldsymbol{h}_u \boldsymbol{h}_u^{\dagger} \right) - (N+1)\, \mathcal{C} \left(\frac{P}{N_o} \left(U - \sum_{u=1}^{U} \rho_u \right) \right) \stackrel{P}{\rightarrow} 0. \quad (4.29)$$

Let us then denote $\xi = \sum_{u=1}^{U} \rho_u$, $\xi \in [0, U)$. Therefore, *i*) using convergences (4.20) and (4.29) into definition (4.19); and, *ii*) noting that max and min keep convergence unaltered (see Lemma 2.2), it concludes the proof. ∎

An interesting conclusion can be drawn from Theorem 4.1: relays do not provide multiplexing gain on the sum-rate. This is concluded by noticing that $\mathcal{SR}_{\text{ub}}(U)$ converges to the minimum of two functions, one of which has pre-log term equal to 1. Since the multiplexing gain of the system is equal to the pre-log term, this is clearly 1 in our channel [9] [78]. As mentioned, this result holds even for MACs with inter-relay connectivity. Nonetheless, even though multiplexing capabilities are not provided by the relays, they increase the sum-capacity bound by means of an in-log term. This encourages to think that, even with infinite users, relays improve the sum-capacity of the MAC; or, in other words, multiuser diversity and relaying diversity are compatible.

4.3 Decode-and-forward

D&F is the first technique considered. As studied in Chapter 3, with D&F the relay nodes fully decode the users' messages. Later, relays reencode them and retransmit them coherently towards destination. The technique was proposed by Cover in [10, Theorem 5], and later applied by Kramer to the MAC with one relay in [76]. In this section, we extend those results to the MPR-MAC.

Before deriving the rate region, let us briefly discuss the relevance of relay selection in the setup. As mentioned in Theorem 3.1, the extension of D&F to multiple relays is based upon a key fact: the higher the cardinality of relays that decode the users' messages, the higher the multiple-antenna gain towards the destination. However, the more relays, the smaller the rate region that all can decode. Therefore, there is always an optimum subset of relays that should be be active: the one who better trade among the two effects. As shown below, for the

MAC, such an optimum subset is (or may be) different for any boundary point of the region. Therefore, all possible subsets $\mathcal{S} \subseteq \mathcal{N}$ must be considered and (possibly) *time-shared.*

Theorem 4.2 *The MPR-MAC achieves the rate region:*

$$\mathcal{R}_{\text{D\&F}}(P_1, P_2) = \text{coh}\left(\bigcup_{\mathcal{S}\subseteq\mathcal{N}} \bigcup_{0\leq\eta_1,\eta_2\leq 1} \{R_{1,2} : \right. \tag{4.30}$$

$$\left.\left. \sum_{u\in\mathcal{G}} R_u \leq \min\left\{\Phi\left(\mathcal{S},\eta_1,\eta_2,\mathcal{G}\right), \Psi\left(\mathcal{S},\eta_1,\eta_2,\mathcal{G}\right)\right\} \mid \forall \mathcal{G} \subseteq \{1,2\} \right\}\right),$$

where

$$\begin{aligned}\Phi\left(\mathcal{S},\eta_1,\eta_2,\mathcal{G}\right) &= \min_{i\in\mathcal{S}} \mathcal{C}\left(\sum_{u\in\mathcal{G}} |b_{u,i}|^2 \eta_u \frac{P_u}{N_o}\right)\\ \Psi\left(\mathcal{S},\eta_1,\eta_2,\mathcal{G}\right) &= \mathcal{C}\left(\sum_{u\in\mathcal{G}} \left(|a_u|^2 + (1-\eta_u)\sum_{i\in\mathcal{S}} |c_i|^2\right) \frac{P_u}{N_o}\right)\end{aligned}$$

with D&F relaying.

Remark 4.1 *$\mathcal{S}$ is a subset of $\mathcal{N}$, and η_1, η_2 are the correlations between users' signals and relays' signals.*

Proof: : Let only the subset of relays $\mathcal{S} \subseteq \mathcal{N}$ be active, and let the users select messages $\omega_u \in \left\{1, \cdots, 2^{nR_u}\right\}$, $u = 1, 2$ for transmission. Each message is divided into B blocks of κR_u bits, with $\kappa = \frac{n}{B}$, *i.e.,* $\omega_u = \left[\omega_u^1, \cdots, \omega_u^B\right]$ $u = 1, 2$. The submessages are then pipelined into $B + 1$ channel blocks, with κ channels uses per block. We consider $n, \kappa, B \gg 1$, so that $\frac{B}{B+1} \approx 1$.

Messages are transmitted using a block-Markov approach [21, Sec. 14.7]: on every block b, users transmit the new sub-messages ω_u^b to the relays in $\mathcal{S}$, and to the destination. Simultaneously, they cooperate with the relays in $\mathcal{S}$ to retransmit their previously transmitted sub-messages ω_u^{b-1}. To that end, users and relays transmit (consider $\omega_u^0 = \omega_u^{B+1} = 0$):

$$\boldsymbol{x}_u^\kappa[b] = \boldsymbol{s}_u^\kappa\left(\omega_u^b, \omega_u^{b-1}\right) + \frac{a_u^*}{\sqrt{|a_u|^2 + \sum_{i\in\mathcal{S}} |c_i|^2}} \boldsymbol{v}_u^\kappa\left(\omega_u^{b-1}\right), \quad u = 1, 2 \tag{4.31}$$

$$\boldsymbol{x}_{r_i}^\kappa[b] = \sum_{u=1}^{2} \frac{c_i^*}{\sqrt{|a_u|^2 + \sum_{i\in\mathcal{S}} |c_i|^2}} \cdot \boldsymbol{v}_u^\kappa\left(\omega_u^{b-1}\right), \quad \forall\, i \in \mathcal{S} \tag{4.32}$$

where we have selected $\boldsymbol{v}_u^\kappa(\cdot)$, $u = 1,2$ to be Gaussian codebooks, generated *i.i.d.* from $v_u \sim \mathcal{CN}(0,(1-\eta_u)P_u)$. In turn, we have chosen $\boldsymbol{s}_u^\kappa(\cdot,\cdot)$, $u=1,2$ to be multiplexed codes, generated *i.i.d.* from $s_u \sim \mathcal{CN}(0,\eta_u P_u)$. Multiplexed codes consist of $\left(\kappa, 2^{\kappa R}\cdot 2^{\kappa R}\right)$ codes indexed by two entries ω_t, ω_d. A receiver can reliably decode both ω_t and ω_d if the channel capacity satisfies $C > 2R$. However, whenever it knows ω_t it can decode ω_d if $C > R$, and, similarly, if it knows ω_d it can decode ω_t (see [79, Section III.A] for full details on the code construction and definition). Notice that, for $0 < \eta_u \leq 1$, $u = 1,2$, the power constraint (*A3*) is satisfied, since the power transmitted by user u and by the relays to cooperate with him add up to P_u.

The received signals at the relays and destination thus read:

$$\begin{aligned}
\boldsymbol{y}_{r_i}^\kappa[b] &= \sum_{u=1}^{2} b_{u,i}\cdot \boldsymbol{s}_u^\kappa\left(\omega_u^b,\omega_u^{b-1}\right) + \sum_{u=1}^{2}\frac{b_{u,i}\cdot a_u^*}{\sqrt{|a_u|^2+\sum_{i\in\mathcal{S}}|c_i|^2}}\boldsymbol{v}_u^\kappa\left(\omega_u^{b-1}\right) + N_o, \ \ \forall\, i\in\mathcal{S}\\
\boldsymbol{y}_d^\kappa[b] &= \sum_{u=1}^{2} a_u\cdot \boldsymbol{s}_u^\kappa\left(\omega_u^b,\omega_u^{b-1}\right) + \sum_{u=1}^{2}\sqrt{|a_u|^2+\sum_{i\in\mathcal{S}}|c_i|^2}\cdot \boldsymbol{v}_u^\kappa\left(\omega_u^{b-1}\right) + N_o.
\end{aligned}$$

On every given block b, the relay node $i\in\mathcal{S}$ is able to decode both ω_u^b, $u=1,2$ if and only if (assuming ω_u^{b-1}, $u=1,2$ well estimated) [21, Section 14.3]:

$$\begin{aligned}
R_1 &\leq I\left(s_1; y_{r_i}|s_2,v_1,v_2\right) = \mathcal{C}\left(|b_{1,i}|^2\eta_1\frac{P_1}{N_o}\right) \\
R_2 &\leq I\left(s_2; y_{r_i}|s_1,v_1,v_2\right) = \mathcal{C}\left(|b_{2,i}|^2\eta_2\frac{P_2}{N_o}\right) \\
R_1+R_2 &\leq I\left(s_2,s_1; y_{r_i}|v_1,v_2\right) = \mathcal{C}\left(\textstyle\sum_{u=1}^2 |b_{u,i}|^2\eta_u\frac{P_u}{N_o}\right).
\end{aligned} \tag{4.33}$$

Therefore, all relay nodes in $\mathcal{S}$ correctly decode the messages, and then retransmit them in block $b+1$, if and only if:

$$\begin{aligned}
R_1 &\leq \min_{i\in\mathcal{S}} \mathcal{C}\left(|b_{1,i}|^2\eta_1\frac{P_2}{N_o}\right) \\
R_2 &\leq \min_{i\in\mathcal{S}} \mathcal{C}\left(|b_{2,i}|^2\eta_2\frac{P_2}{N_o}\right) \\
R_1+R_2 &\leq \min_{i\in\mathcal{S}} \mathcal{C}\left(\textstyle\sum_{u=1}^2 |b_{u,i}|^2\eta_u\frac{P_u}{N_o}\right).
\end{aligned} \tag{4.34}$$

Next, consider the decoding at the destination node, implemented through *backward* decoding [58, Sec III.B]. That is, the destination starts decoding from the last block and proceeds backward. Assume that, on a given block b, the destination has successfully decoded

$\omega_u^b, \cdots, \omega_u^B$, $u = 1,2$. Then, it uses $\boldsymbol{y}_d^\kappa[b]$ to estimate ω_u^{b-1}:

$$\boldsymbol{y}_d^\kappa[b] = \sum_{u=1}^{2} a_u \boldsymbol{s}_u^\kappa\left(\omega_u^b, \omega_u^{b-1}\right) + \sum_{u=1}^{2}\sqrt{|a_u|^2 + \sum_{i\in\mathcal{S}}|c_i|} \cdot \boldsymbol{v}_u^\kappa\left(\omega_u^{b-1}\right) + \boldsymbol{z}_d^\kappa. \tag{4.35}$$

Being $\boldsymbol{s}^\kappa(\cdot,\cdot)$ a multiplexed code, and assuming ω_b well estimated, the destination reliably decodes ω_u^{b-1} $u = 1,2$ *iff*:

$$\begin{aligned} R_1 &\leq I\left(s_1, v_1; y_d | s_2, v_2\right) = \mathcal{C}\left(\left(|a_1|^2 + (1-\eta_1)\sum_{i\in\mathcal{S}}|c_i|^2\right)\frac{P_1}{N_o}\right) \\ R_2 &\leq I\left(s_2, v_2; y_d | s_1, v_1\right) = \mathcal{C}\left(\left(|a_2|^2 + (1-\eta_2)\sum_{i\in\mathcal{S}}|c_i|^2\right)\frac{P_2}{N_o}\right) \\ R_1 + R_2 &\leq I\left(s_1, s_2, v_1, v_2; y_d\right) = \mathcal{C}\left(\sum_{u=1}^{2}\left(|a_u|^2 + (1-\eta_u)\sum_{i\in\mathcal{S}}|c_i|^2\right)\frac{P_u}{N_o}\right) \end{aligned} \tag{4.36}$$

Now, considering the two limitations in (4.34) and (4.36), we obtain the minimization in (4.30). However, notice that correlations η_1, η_2 can be arbitrarily chosen, as well as the subset of active relays $\mathcal{S} \subseteq \mathcal{N}$. Hence, the rate region becomes the union over subsets and $0 < \eta_1, \eta_2 \leq 1$. Finally, due to time-sharing the convex hull operation is demonstrated. ■

Notice that the boundary points of the rate region can be attained using superposition coding (SC), successive interference cancelation (SIC) at the decoders, and (optionally) *time-sharing* (TS). In fact, for fixed subset of relays $\mathcal{S}$ and fixed correlations η_1, η_2, the rate region is the intersection of the capacity regions at the relays in $\mathcal{S}$ and at the destination (as shown in the proof). Since SC, SIC and TS are thus optimal for the individual regions, they are optimal for the intersection of them as well.

Let us compute now the maximum sum-rate of the rate region (4.30). For $\mathcal{G} = \{1\}$ and $\mathcal{G} = \{2\}$, it is clear that:

$$\begin{aligned} R_1 \leq \bar{R}_1 &\triangleq \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\leq\eta_1,\eta_2\leq 1} \min\left\{\Phi\left(\mathcal{S},\eta_1,\eta_2,\{1\}\right), \Psi\left(\mathcal{S},\eta_1,\eta_2,\{1\}\right)\right\} \\ R_2 \leq \bar{R}_2 &\triangleq \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\leq\eta_1,\eta_2\leq 1} \min\left\{\Phi\left(\mathcal{S},\eta_1,\eta_2,\{2\}\right), \Psi\left(\mathcal{S},\eta_1,\eta_2,\{2\}\right)\right\} \end{aligned} \tag{4.37}$$

which are the maximum single-user rates of both sources when they are assisted by relays, which can be computed from Theorem 3.1. Additionally, from the sum-rate constraint (*i.e.*, $\mathcal{G} = \{1,2\}$):

$$\sum_{u=1}^{2} R_u \leq \bar{R}_{1,2} \triangleq \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\leq\eta_1,\eta_2\leq 1} \min\left\{\Phi\left(\mathcal{S},\eta_1,\eta_2,\{1,2\}\right), \Psi\left(\mathcal{S},\eta_1,\eta_2,\{1,2\}\right)\right\}. \tag{4.38}$$

Unfortunately, no closed-form expression can be obtained for $\bar{R}_{1,2}$. Notice now that, unlike the sum-rate of the MAC without relays, $\bar{R}_{1,2}$ can be either greater than, lower than or equal to $\bar{R}_1 + \bar{R}_2$. Therefore, we need to consider both (4.37) and (4.38) in order to compute the maximum sum-rate of the channel with D&F, which is

$$SR_{\text{D\&F}} = \min\left\{\bar{R}_1 + \bar{R}_2, \bar{R}_{1,2}\right\}. \tag{4.39}$$

4.3.1 Multiuser Diversity and D&F

We study now the impact of multiuser diversity on the achievable rate region with D&F. In particular, the scaling law (on the number of users) of its achievable sum-rate is obtained. First, we trivially extend (4.38) to $U > 2$:

$$\sum_{u=1}^{U} R_u \leq \mathcal{SR}_{\text{D\&F}}(U) \triangleq \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\prec\eta_{\mathcal{U}}\preceq 1} \min\left\{\Phi\left(\mathcal{S},\eta_{\mathcal{U}},\mathcal{U}\right), \Psi\left(\mathcal{S},\eta_{\mathcal{U}},\mathcal{U}\right)\right\}, \tag{4.40}$$

where $\mathcal{U} = \{1,\cdots,U\}$, and $\eta_{\mathcal{U}} = \{\eta_1,\cdots,\eta_U\}$. The interpretation of (4.40) is as follows: the maximum sum-rate with D&F is always lower than or equal to the maximum, over the subset of active relays and over the source-relay correlation, of the minimum of: *i*) the sum-capacity towards the active relays, $\Phi\left(\mathcal{S},\eta_{\mathcal{U}},\mathcal{U}\right)$ and *ii*) the sum-capacity towards the destination with the aid of the active relays, $\Psi\left(\mathcal{S},\eta_{\mathcal{U}},\mathcal{U}\right)$. However, as pointed out in (4.39), such a bound does not necessarily hold with equality. The explanation comes from the fact that relays and destination may be able to decode the same sum-capacity, but at different points of the rate region (See Fig. 4.2). Accordingly, it may occur that such a sum-capacity cannot be simultaneously maintained with the relays and with the destination, and therefore it is not achievable.

Assume unitary-mean, Rayleigh-distributed fading coefficients $a_u \sim \mathcal{CN}(0,1)$, and $b_{u,i} \sim \mathcal{CN}(0,1)$, $\forall u, i$. Also, for simplicity of exposition, assume that all power constraints are identical: $P_u = P$, $u = 1,\cdots,U$. Then, the following convergence in probability can be proven.

Theorem 4.3 *Consider the MPR-MAC with a finite set of relays $\mathcal{N}$. Then, for U $\rightarrow \infty$,*

$$\mathcal{SR}_{\text{D\&F}}(U) - \mathcal{C}\left(U\cdot\frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{4.41}$$

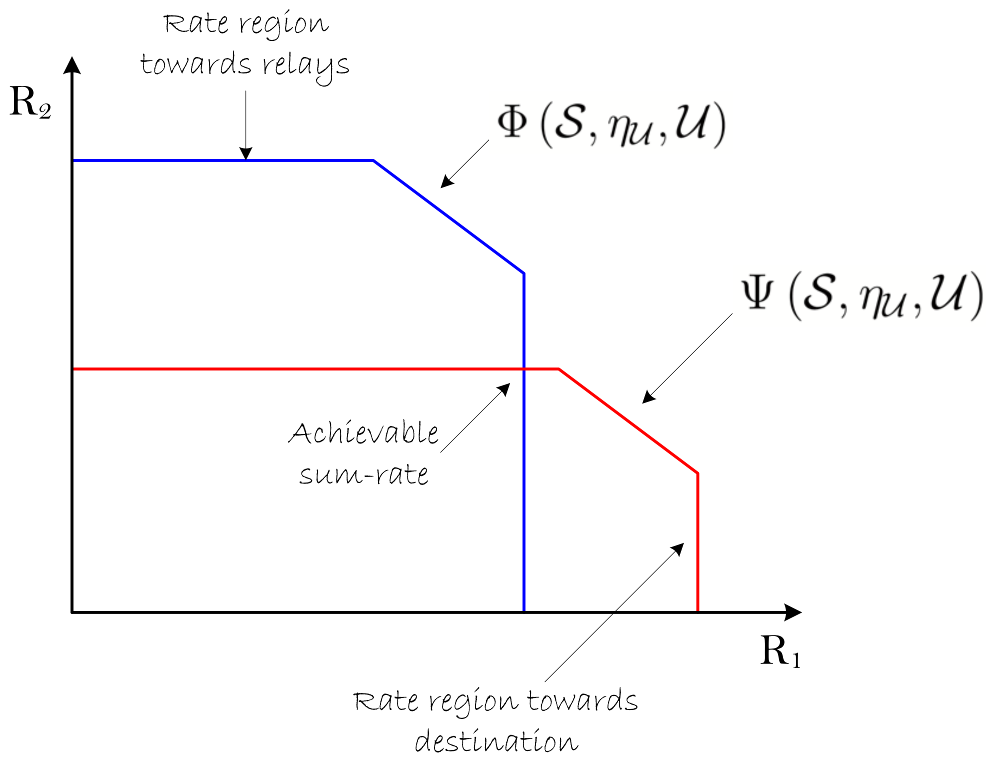

Figure 4.2: Achievable sum-rate with D&F.

Proof: In (4.40), we defined

$$\begin{aligned} \Phi\left(\mathcal{S},\boldsymbol{\eta}_{\mathcal{U}},\mathcal{U}\right) &= \min_{i\in\mathcal{S}} \mathcal{C}\left(\frac{P}{N_o}\sum_{u\in\mathcal{U}}|b_{u,i}|^2\eta_u\right) \\ \Psi\left(\mathcal{S},\boldsymbol{\eta}_{\mathcal{U}},\mathcal{U}\right) &= \mathcal{C}\left(\frac{P}{N_o}\sum_{u\in\mathcal{U}}\left(|a_u|^2+(1-\eta_u)\sum_{i\in\mathcal{S}}|c_i|^2\right)\right). \end{aligned} \tag{4.42}$$

First of all, given unitary-mean, Rayleigh fading, we may apply Corollary 2.3 to derive:

$$\mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}|b_{u,i}|^2\eta_u\right)-\mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}\eta_u\right)\xrightarrow{P}0.$$

Furthermore, given the subset of relays $\mathcal{S}$ of finite cardinality and noting from Lemma 2.2 that the $\min$ operator keeps convergence unaltered, the following relationship also holds:

$$\Phi\left(\mathcal{S},\boldsymbol{\eta}_{\mathcal{U}},\mathcal{U}\right)-\mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}\eta_u\right)\xrightarrow{P}0. \tag{4.43}$$

Consider now the convergence of $\Psi\left(\mathcal{S},\boldsymbol{\eta}_{\mathcal{U}},\mathcal{U}\right)$. Using Corollary 2.3 we can derive that:

$$\Psi\left(\mathcal{S},\boldsymbol{\eta}_{\mathcal{U}},\mathcal{U}\right)-\log_2\left(1+\frac{P}{N_o}\left(U+\sum_{i\in\mathcal{S}}|c_i|^2\sum_{u=1}^{U}(1-\rho_u)\right)\right)\xrightarrow{P}0, \tag{4.44}$$

Therefore, using convergence (4.43) and (4.44) in definition (4.40), and recalling that the $\max$ and $\min$ operators do not altere convergence (see Lemma 2.2), we derive that:

$$\mathcal{SR}_{\mathrm{DF}}\left(U\right) - \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\prec\eta_{\mathcal{U}}\preceq 1} \min\left\{\mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}\eta_u\right), \right.$$
$$\left. \mathcal{C}\left(\frac{P}{N_o}\left(U+\sum_{i\in\mathcal{S}}|c_i|^2\sum_{u=1}^{U}(1-\rho_u)\right)\right)\right\} \xrightarrow{P} 0, \tag{4.45}$$

However, notice that $\sum_{u=1}^{U}\eta_u \leq U$ always. Therefore,

$$\mathcal{SR}_{\mathrm{DF}}\left(U\right) - \max_{\mathcal{S}\subseteq\mathcal{N}} \max_{0\prec\eta_{\mathcal{U}}\preceq 1} \mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}\eta_u\right) \xrightarrow{P} 0, \tag{4.46}$$

which is evaluated as

$$\mathcal{SR}_{\mathrm{DF}}\left(U\right) - \mathcal{C}\left(U\cdot\frac{P}{N_o}\right) \xrightarrow{P} 0. \tag{4.47}$$

This concludes the proof. ■

Theorem 4.3 states that the sum-rate of the MAC with D&F relays and the MAC without relays are asymptotically equal. Therefore, D&F relaying is not useful, in terms of capacity, when there is enough multiuser diversity in the network. We can interpret this resorting to the *Distributed Channel Hardening* effect [80]. To explain this effect, let us first recall that for D&F with a single source (see Section 3.3.1), the ordered SNRs at the relays were approximated by the inverse function of the *c.d.f.* of the channel distribution: see *e.g.*, (3.26) and (3.27). That is,

$$\mathrm{SNR}_{r_m} \approx -\log\left(\frac{m}{N}\right)\frac{P}{N_o}, \ \ m = 1,\cdots,N. \tag{4.48}$$

The result was obtained using ordered statistics in [60]; according to it, the relays with low $m \ll N$ (*i.e.*, the relays with highest source-relay channels following (3.18)) enjoyed links with the source of much higher capacity than the destination did. Hence, the virtual antenna was reliably mimicked and D&F became useful.

However, with multiple sources, the received SNR at the relays is the result of the power contribution from all users, which is (as for the single-user case) random due to fading:

$$\mathrm{SNR}_{r_m} = \frac{P}{N_o}\sum_{u=1}^{U}\eta_u|b_{u,m}|^2, \ \ m = 1,\cdots,N. \tag{4.49}$$

Nevertheless, with infinite users this sum satisfies the law of large numbers: the received SNR at the relays converges in probability to a fixed constant, which is indeed equal for all relays:

$$\mathrm{SNR}_{r_m} - \frac{P}{N_o}\sum_{u=1}^{U}\eta_u \xrightarrow{P} 0, \;\; \forall m. \tag{4.50}$$

Thereby, in this case, the ordered statistics are not able to provide any SNR gain among the relays, and all have the same received power. Moreover, since $\eta_u \leq 1 \; \forall u$, the asymptotic sum-rate of the relays is clearly lower than or equal to that of the destination without relays:

$$\frac{P}{N_o}\sum_{u=1}^{U}\eta_u \leq \frac{P}{N_o}U. \tag{4.51}$$

Therefore, relays become the bottleneck of the network and no rate gain is obtained. This is what we refer to distributed channel hardening effect. The effect was discovered by Hochwald *et al.* in the context of opportunistic MISO-MAC [80], and complemented by Larsson in [81].

Finally, notice that the *max-flow-min-cut* bound (Theorem 4.1) indeed suggests rate gains even for high number of users. Therefore, D&F is clearly suboptimal for densely-populated MACs.

4.4 Linear Relaying

With *LR*, the relays do not decode the users' signals, but amplify them instead. Therefore, it is expected that this technique is not affected by the channel hardening effect. At the MPR-MAC, *LR* operates as follows [24]: the two sources select messages $\omega_u \in \left\{1, \cdots, 2^{R_u}\right\}$, $u = 1, 2$ for transmission and map them onto two independent codewords $\boldsymbol{x}_u^n$, $u = 1, 2$ of sufficiently large length n. The codewords are then transmitted into B channel blocks of $\kappa = \frac{n}{B}$ channel uses each. On every block b, the sequence of symbols $\boldsymbol{x}_u^\kappa[b]$, $u = 1, 2$ are transmitted and received at the relays and destination following (4.2) and (4.3), respectively. Simultaneously, the relays linearly combine the received signal during this block, and send the sequence of symbols $\boldsymbol{x}_{r_i}^\kappa[b]$, given by:

$$\begin{aligned} \boldsymbol{x}_{r_i}^\kappa[b] &= \boldsymbol{\Phi}_i \cdot \boldsymbol{y}_{r_i}^\kappa[b] \\ &= \boldsymbol{\Phi}_i \cdot \left(\sum_{u=1}^{2} b_{u,i} \cdot \boldsymbol{x}_u^\kappa[b] + \boldsymbol{z}_{r_i}^\kappa\right), \;\; i = 1, \cdots, N. \end{aligned} \tag{4.52}$$

As for previous chapter, $\boldsymbol{\Phi}_i \in \mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}$ is referred to as linear relaying matrix of relay i. It defines the linear combination of past inputs, and is strictly lower triangular to preserve causality. The received signal at the destination is[2]:

$$\begin{aligned}\boldsymbol{y}_d^\kappa &= \sum_{u=1}^{2} a_u \cdot \boldsymbol{x}_u^\kappa + \sum_{i=1}^{N} c_i \cdot \boldsymbol{x}_{r_i}^\kappa + \boldsymbol{z}_d^\kappa, \\ &= \sum_{u=1}^{2}\left(a_u \cdot \boldsymbol{I} + \sum_{i=1}^{N} b_{u,i} c_i \boldsymbol{\Phi}_i\right)\cdot \boldsymbol{x}_u^\kappa + \left(\boldsymbol{z}_d^\kappa + \sum_{i=1}^{N} c_i \boldsymbol{\Phi}_i \boldsymbol{z}_{r_i}^\kappa\right). \end{aligned} \tag{4.53}$$

As previously, the communication is constrained to satisfy the per-user constraint *(A3)*, which can be stated as follows: let $\boldsymbol{Q}_u^\kappa = \mathrm{E}\left\{\boldsymbol{x}_u^\kappa \left(\boldsymbol{x}_u^\kappa\right)^\dagger\right\} \succeq 0$ be the source $u = 1,2$ temporal covariance matrix, and $\boldsymbol{Q}_{r_i}^\kappa = \mathrm{E}\left\{\boldsymbol{x}_{r_i}^\kappa \left(\boldsymbol{x}_{r_i}^\kappa\right)^\dagger\right\} = \sum_{u=1}^{2} |b_{u,i}|^2 \boldsymbol{\Phi}_i \boldsymbol{Q}_u^\kappa \boldsymbol{\Phi}_i^\dagger + N_o \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\dagger$, the relays temporal covariance. Then, the total transmitted power in the network is computed as:

$$\begin{aligned} P &= \frac{1}{\kappa}\left(\sum_{u=1}^{2} \mathrm{tr}\left\{\boldsymbol{Q}_u^\kappa\right\} + \sum_{i=1}^{N} \mathrm{tr}\left\{\boldsymbol{Q}_{r_i}^\kappa\right\}\right) \\ &= \frac{1}{\kappa}\sum_{u=1}^{2} \mathrm{tr}\left\{\boldsymbol{Q}_u^\kappa\left(\boldsymbol{I} + \sum_{i=1}^{N} |b_{u,i}|^2 \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right) + \frac{N_o}{2}\sum_{i=1}^{N} \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right\} \end{aligned} \tag{4.54}$$

Therefore, we can easily infer that the power used in the network to transmit the message of user $u = 1,2$ is:

$$\mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \boldsymbol{\Phi}_{1:N}\right) = \frac{1}{\kappa}\mathrm{tr}\left\{\boldsymbol{Q}_u^\kappa\left(\boldsymbol{I} + \sum_{i=1}^{N} |b_{u,i}|^2 \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right) + \frac{N_o}{2}\sum_{i=1}^{N} \boldsymbol{\Phi}_i^\dagger \boldsymbol{\Phi}_i\right\}. \tag{4.55}$$

Accordingly, assumption *(A3)* is formulated as $\mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \boldsymbol{\Phi}_{1:N}\right) \leq P_u$, $u = 1,2$. The rate region with such a relaying architecture is given in the following theorem.

Theorem 4.4 *With LR, the achievable rate region of the MPR-MAC is*

$$\mathcal{R}_{\mathrm{LR}}\left(P_1, P_2\right) = \lim_{\kappa\to\infty} \mathrm{coh}\left(\bigcup_{\boldsymbol{\Phi}_{1:N} \in \mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}} \mathcal{R}_\kappa\left(\boldsymbol{\Phi}_{1:N}\right)\right) \tag{4.56}$$

where $\mathcal{R}_\kappa\left(\boldsymbol{\Phi}_{1:N}\right) \triangleq$

$$\bigcup_{\substack{\boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa \succeq 0 \\ \mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \boldsymbol{\Phi}_{1:N}\right) \leq P_u,\ u=1,2}} \left\{ R_{1,2} : \begin{array}{c} R_1 \leq \frac{1}{\kappa}\log_2\det\left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_1 \boldsymbol{Q}_1^\kappa \boldsymbol{H}_1^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \\ R_2 \leq \frac{1}{\kappa}\log_2\det\left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_2 \boldsymbol{Q}_2^\kappa \boldsymbol{H}_2^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \\ R_1 + R_2 \leq \frac{1}{\kappa}\log_2\det\left(\boldsymbol{I} + \sum_{u=1}^{2} \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \end{array}\right\} \tag{4.57}$$

and $\boldsymbol{R}_d = N_o\left(\boldsymbol{I} + \sum_{i=1}^{N} |c_i|^2 \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\dagger\right)$, $\boldsymbol{H}_u = \left(a_u \cdot \boldsymbol{I} + \sum_{i=1}^{N} b_{u,i} c_i \boldsymbol{\Phi}_i\right)$, $u = 1,2$.

[2]For notational simplicity, we remove index b.

Remark 4.2 *Hereafter, we refer to $\mathcal{R}_\kappa\left(\mathbf{\Phi}_{1:N}\right)$ as the conditional rate region of the MPR-MAC. It is defined as the closure of all rate duples $\left(R_1, R_2\right)$ that are achievable when the relay nodes use the fixed set of relaying matrices $\mathbf{\Phi}_{1:N} \in \mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}$.*

Proof: Let the sources transmit on blocks of κ channel uses, and the relays use the given set of matrices $\mathbf{\Phi}_{1:N}$. The received signal is given in (3.89), for which the capacity region is derived using theory of vector MAC [21]:

$$\mathcal{R}_\kappa\left(\mathbf{\Phi}_{1:N}\right) = \mathrm{coh}\left(\bigcup_{p_{\boldsymbol{x}_1^\kappa}p_{\boldsymbol{x}_2^\kappa}:\mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa,\mathbf{\Phi}_{1:N}\right)\leq P_u,\ u=1,2}\left\{R_{1,2}: \begin{array}{c} R_1 \leq \frac{1}{\kappa}\cdot I\left(\boldsymbol{x}_1^\kappa;\boldsymbol{y}_d^\kappa|\boldsymbol{x}_2^\kappa\right) \\ R_2 \leq \frac{1}{\kappa}\cdot I\left(\boldsymbol{x}_2^\kappa;\boldsymbol{y}_d^\kappa|\boldsymbol{x}_1^\kappa\right) \\ R_1+R_2 \leq \frac{1}{\kappa}\cdot I\left(\boldsymbol{x}_1^\kappa,\boldsymbol{x}_2^\kappa;\boldsymbol{y}_d^\kappa\right)\end{array}\right\}\right),$$

where achievability is proven using random coding arguments with codewords of length $n = B\kappa \to \infty$, generated *i.i.d* from $p_{\boldsymbol{x}_u^\kappa}$ $u = 1, 2$, and joint typical decoding. For the converse, Fano's inequality applies. Considering Gaussian noise, the rate region above is directly (4.57) [21, Section 10.5]. This demonstrates the greater achievable rate region with a fixed set of relaying functions $\mathbf{\Phi}_{1:N}$. However, the relaying matrices can be arbitrarily chosen at the relay nodes; thus, any duple $\left(R_1, R_2\right)$ belonging to $\bigcup_{\mathbf{\Phi}_{1:N}\in\mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}}\mathcal{R}_\kappa\left(\mathbf{\Phi}_{1:N}\right)$ is also achievable. Furthermore, so is any point at the convex hull of the union through *time-sharing* of sets of linear relaying matrices:

$$\mathcal{R}_\kappa = \mathrm{coh}\left(\bigcup_{\mathbf{\Phi}_{1:N}\in\mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}}\mathcal{R}_\kappa\left(\mathbf{\Phi}_{1:N}\right)\right). \quad (4.58)$$

Finally, the number of channel uses per block, κ, is also designed arbitrarily. Therefore:

$$\mathcal{R}_{\mathrm{LR}} = \bigcup_\kappa \mathcal{R}_\kappa = \lim_{\kappa\to\infty}\mathcal{R}_\kappa, \quad (4.59)$$

where the second equality holds from the non-decreasing nature of $\mathcal{R}_\kappa$ with κ [24, Sec. III]. Hence, introducing (4.58) in (4.59), concludes the proof. ■

In order to evaluate the convex region $\mathcal{R}_{\mathrm{LR}}\left(P_1, P_2\right)$, we can resort to the weighted sum-rate (WSR) optimization, which consists of describing the region by means of its bounding hyperplanes [82, Sec. III-C]:

$$\mathcal{R}_{\mathrm{LR}}\left(P_1,P_2\right) = \left\{R_{1,2}: \alpha R_1 + \left(1-\alpha\right)R_2 \leq R\left(\alpha\right), \forall\alpha\in[0,1]\right\}, \quad (4.60)$$

where $R(\alpha)$ is the maximum WSR of channel, assuming weights α and $1-\alpha$ for users s_1 and s_2, respectively. Notice that given relationship (4.56), $R(\alpha)$ can be expressed in terms of the WSR of the conditional rate region. Specifically, defining

$$\mathcal{R}_\kappa(\mathbf{\Phi}_{1:N}) = \{R_{1,2} : \alpha R_1 + (1-\alpha) R_2 \leq R_\kappa(\alpha, \mathbf{\Phi}_{1:N}), \forall \alpha \in [0,1]\}, \tag{4.61}$$

it is clear that:

$$R(\alpha) = \lim_{\kappa\to\infty} \max_{\mathbf{\Phi}_{1:N}\in\mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}} R_\kappa(\alpha, \mathbf{\Phi}_{1:N}). \tag{4.62}$$

Hereafter, we refer to $R_\kappa(\alpha, \mathbf{\Phi}_{1:N})$ as the conditional WSR, achieved with equality at the boundary of (4.57). Clearly, it can be attained using superposition coding at the users, successive interference cancelation (SIC) at the destination, and *time-sharing*.

In order to tackle optimization (4.62), we proceed in two steps:

- First, we devise an iterative algorithm to compute $R_\kappa(\alpha, \mathbf{\Phi}_{1:N})$, which is the maximum value of $\alpha R_1 + (1-\alpha) R_2$ when the relays use the fixed set of relaying functions $\mathbf{\Phi}_{1:N}$.
- Afterwards, we plug it on (4.62), and study the maximization. This turns out to be non-convex and not solvable in practice. We thus propose a suboptimum approach.

Let us consider then $R_\kappa(\alpha, \mathbf{\Phi}_{1:N})$. As mentioned, the conditional WSR can be attained using SIC at the destination. This consists in first decoding the user with lowest weight, then subtracting its contribution on the received signal, and finally decoding the other without interference. Therefore, assuming *w.l.o.g.* $\alpha \geq 0.5$, the user s_1 has higher priority at the SIC and thus is decoded last. With such a priority, the transmission rates at the users read:

$$\begin{aligned} R_1 &= \frac{1}{\kappa}\log_2\det\left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_1\boldsymbol{Q}_1^\kappa\boldsymbol{H}_1^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\right) \\ R_2 &= \frac{1}{\kappa}\log_2\det\left(\boldsymbol{I} + \left(\boldsymbol{R}_d + \boldsymbol{H}_1\boldsymbol{Q}_1^\kappa\boldsymbol{H}_1^\dagger\right)^{-1}\boldsymbol{H}_2\boldsymbol{Q}_2^\kappa\boldsymbol{H}_2^\dagger\right) \end{aligned} \tag{4.63}$$

Accordingly, the maximum value of $\alpha R_1 + (1-\alpha) R_2$ is computed from

$$\begin{aligned} R_\kappa(\alpha, \mathbf{\Phi}_{1:N}) &= \max_{\boldsymbol{Q}_1^\kappa,\boldsymbol{Q}_2^\kappa\succeq 0} \sum_{p=1}^{2}\theta_p\cdot\log_2\det\left(\boldsymbol{I} + \sum_{u=1}^{p}\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_u\boldsymbol{Q}_u^\kappa\boldsymbol{H}_u^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\right), \\ &\quad \text{s.t. } \mathsf{P}_u(\boldsymbol{Q}_u^\kappa, \mathbf{\Phi}_{1:N}) \leq P_u,\ u = 1,2 \end{aligned} \tag{4.64}$$

where $\theta_1 = \frac{2\alpha-1}{\kappa}$ and $\theta_2 = \frac{1-\alpha}{\kappa}$. Notice that the objective function is the sum of convex functions and the constraints define a convex set. Therefore, the optimization is convex in standard form.

4.4.1 Sum-Rate Maximization

Let us first consider maximization (4.64) with $\alpha = 0.5$. For it, $\theta_1 = 0$ and $\theta_2 = \frac{1}{2\kappa}$:

$$R_\kappa \left(0.5, \mathbf{\Phi}_{1:N}\right) = \max_{\boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa \succeq 0} \frac{1}{2\kappa} \log_2 \det \left(\boldsymbol{I} + \sum_{u=1}^{2} \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger \boldsymbol{R}_d^{-\frac{1}{2}} \right) \quad \text{s.t. } \mathsf{P}_u \left(\boldsymbol{Q}_u^\kappa, \mathbf{\Phi}_{1:N}\right) \leq P_u, \; u = 1, 2 \tag{4.65}$$

To solve the optimization, we propose the use of a *block-coordinate ascent algorithm* (see Section 2.2.1). It consists in iteratively optimizing the objective function with respect to one $\boldsymbol{Q}_u^\kappa$ while keeping the other fixed, and is defined for our problem as:

$$\boldsymbol{Q}_u^\kappa (t+1) = \arg \max_{\boldsymbol{Q}_u^\kappa \succeq 0} \log_2 \det \left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}} \left(\boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger + \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa (t+2-\bar{u}) \boldsymbol{H}_{\bar{u}}^\dagger \right) \boldsymbol{R}_d^{-\frac{1}{2}} \right), \quad \text{s.t. } \mathsf{P}_u \left(\boldsymbol{Q}_u^\kappa, \mathbf{\Phi}_{1:N}\right) \leq P_u \tag{4.66}$$

where t is the iteration index, $u = 1, 2$, and $\bar{u} = \{1, 2\}/u$. It can be shown that not only the original maximization (4.65) is convex, but also its feasible set is the cartesian product of convex sets and (4.66) is strictly convex, *i.e.* it is uniquely attained. Hence, the limit point of the sequence $\{\boldsymbol{Q}_1^\kappa(t), \boldsymbol{Q}_2^\kappa(t)\}$ is guaranteed to converge to the sum-rate (see Corollary 2.1). The solution of maximization (4.66) is presented in the following Proposition.

Proposition 4.1 *Consider the optimization in (4.66), define* $\boldsymbol{A}_u = \left(\boldsymbol{I} + \sum_{i=1}^{N} |b_{u,i}|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i \right)$ *and compute the SVD-decomposition*

$$\left(\boldsymbol{R}_d + \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa (t+2-\bar{u}) \boldsymbol{H}_{\bar{u}}^\dagger \right)^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{A}_u^{-\frac{1}{2}} = \boldsymbol{U}_u \mathbf{\Lambda}^{\frac{1}{2}} \boldsymbol{V}_u^\dagger. \tag{4.67}$$

Then, $\boldsymbol{Q}_u^\kappa(t+1) = \boldsymbol{A}_u^{-\frac{1}{2}} \boldsymbol{V}_u \mathbf{\Psi} \boldsymbol{V}_u^\dagger \boldsymbol{A}_u^{-\frac{1}{2}}$, *where*

$$\psi_j = \left[\frac{1}{\mu} - \frac{1}{\lambda_j} \right]^+, \quad \textit{with} \quad \sum_{j=1}^{\kappa} \psi_j = \kappa P_u - \operatorname{tr}\left\{ \frac{N_o}{2} \sum_{i=1}^{N} \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i \right\}. \tag{4.68}$$

Remark 4.3 *With such a solution, algorithm (4.66) is equivalent to the iterative waterfilling algorithm for the MIMO-MAC sum-rate [50]. However, in our case, the SVD decomposition takes also into account the matrix* $\boldsymbol{A}_u^{-\frac{1}{2}}$.

Proof: First, let us rewrite the objective function of (4.66) as:

$$\log_2 \det \left(\boldsymbol{I} + \sum_{u=1}^{2} \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger \boldsymbol{R}_d^{-\frac{1}{2}} \right) = \log_2 \det \left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa \boldsymbol{H}_{\bar{u}}^\dagger \boldsymbol{R}_d^{-\frac{1}{2}} \right) + \log_2 \det \left(\boldsymbol{I} + \boldsymbol{H}_e \boldsymbol{Q}_u^\kappa \boldsymbol{H}_e^\dagger \right), \tag{4.69}$$

where $\boldsymbol{H}_e = \left(\boldsymbol{R}_d + \boldsymbol{H}_{\bar{u}}\boldsymbol{Q}_{\bar{u}}^{\kappa}\boldsymbol{H}_{\bar{u}}^{\dagger}\right)^{-\frac{1}{2}}\boldsymbol{H}_u$. Notice that the first term does not depend on $\boldsymbol{Q}_u^{\kappa}$. Hence, the Lagrangian for the optimization is given by [48]:

$$\mathcal{L}\left(\boldsymbol{Q}_u^{\kappa}, \boldsymbol{\Omega}, \mu\right) = \log\left(\det\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{Q}_u^{\kappa}\boldsymbol{H}_e^{\dagger}\right)\right) + \operatorname{tr}\left\{\boldsymbol{\Omega}\boldsymbol{Q}_u^{\kappa}\right\} - \mu\left(\kappa\mathsf{P}_u\left(\boldsymbol{Q}_u^{\kappa}, \boldsymbol{\Phi}_{1:N}\right) - \kappa P_u\right)$$

where $\mu \geq 0$ and matrix $\boldsymbol{\Omega} \succeq 0$ are the Lagrange multipliers for the power and semi-definite positive constraints, respectively. The KKT conditions for the problem (sufficient and necessary for optimality due to convexity, and regularity of the feasible set) are [48]:

$$\begin{aligned} &i)\ \mu\left(\boldsymbol{I} + \textstyle\sum_{i=1}^{N}|b_{u,i}|^2\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right) - \boldsymbol{\Omega} = \boldsymbol{H}_e^{\dagger}\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{Q}_u^{\kappa}\boldsymbol{H}_e^{\dagger}\right)^{-1}\boldsymbol{H}_e \\ &ii)\ \mu\left(\operatorname{tr}\left\{\boldsymbol{Q}_u^{\kappa}\left(\boldsymbol{I} + \textstyle\sum_{i=1}^{N}|b_{u,i}|^2\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right) + \frac{N_o}{2}\sum_{i=1}^{N}\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right\} - \kappa P_u\right) = 0 \\ &iii)\ \operatorname{tr}\left\{\boldsymbol{\Omega}\boldsymbol{Q}_u^{\kappa}\right\} = 0 \end{aligned} \tag{4.70}$$

Let us now define $\boldsymbol{A}_u = \boldsymbol{I} + \sum_{i=1}^{N}|b_{u,i}|^2\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i$, which is singular and semidefinite positive. Also, consider the change of variables $\tilde{\boldsymbol{Q}}_u = \boldsymbol{A}_u^{\frac{1}{2}}\boldsymbol{Q}_u^{\kappa}\boldsymbol{A}_u^{\frac{1}{2}}$ and $\tilde{\boldsymbol{\Omega}} = \boldsymbol{A}_u^{-\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{A}_u^{-\frac{1}{2}}$, both semidefinite positive, as $\boldsymbol{Q}_u^{\kappa}$ and $\boldsymbol{\Omega}$. With them, we can turn the KKT conditions into:

$$\begin{aligned} &i)\ \mu - \tilde{\boldsymbol{\Omega}} = \boldsymbol{A}_u^{-\frac{1}{2}}\boldsymbol{H}_e^{\dagger}\left(\boldsymbol{I} + \boldsymbol{H}_e\boldsymbol{A}_u^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}_u\boldsymbol{A}_u^{-\frac{1}{2}}\boldsymbol{H}_e^{\dagger}\right)^{-1}\boldsymbol{H}_e\boldsymbol{A}_u^{-\frac{1}{2}} \\ &ii)\ \mu\left(\operatorname{tr}\left\{\tilde{\boldsymbol{Q}}_u + \frac{N_o}{2}\textstyle\sum_{i=1}^{N}\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right\} - \kappa P_u\right) = 0 \\ &iii)\ \operatorname{tr}\left\{\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{Q}}_u\right\} = 0 \end{aligned} \tag{4.71}$$

From the SVD-decomposition $\boldsymbol{H}_e\boldsymbol{A}_u^{-\frac{1}{2}} = \boldsymbol{U}_u\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}_u^{\dagger}$, we can obtain (as for the capacity of MIMO channels [3]) that $\tilde{\boldsymbol{Q}}_u^{*} = \boldsymbol{V}_u\boldsymbol{\Psi}\boldsymbol{V}_u^{\dagger}$, with $\psi_j = \left[\frac{1}{\mu^*} - \frac{1}{\lambda_j}\right]^{+}$, $j = 1, \cdots, k$ satisfies KKT conditions. μ^* is such that $\sum_{j=1}^{k}\psi_j + \frac{N_o}{2}\operatorname{tr}\left\{\sum_{i=1}^{N}\boldsymbol{\Phi}_i^{\dagger}\boldsymbol{\Phi}_i\right\} - \kappa P_u = 0$, and $\tilde{\boldsymbol{\Omega}}^*$ computed from KKT condition *i*) as: $\tilde{\boldsymbol{\Omega}}^* = \boldsymbol{V}_u\left(\mu - \boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{I} + \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Psi}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\boldsymbol{V}_u^{\dagger}$, which is semidefinite positive. Finally, we recover the optimum source temporal covariance as $\boldsymbol{Q}_u^{\kappa} = \boldsymbol{A}_u^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}_u^{*}\boldsymbol{A}_u^{-\frac{1}{2}}$.

■

4.4.2 Weighted Sum-Rate Maximization

We study now the maximization (4.64) with $\alpha > 0.5$. In the previous subsection, we showed that such an optimization can be tackled using a *block-coordinate* approach. However, in this case, it is not clear that the maximization over one $\boldsymbol{Q}_u^{\kappa}$ keeping the other fixed is uniquely attained. Hence, the algorithm is not guaranteed to converge [48, Proposition 2.7.1]. In order to solve the problem with guaranteed convergence, we propose the use of *Gradient Projection* (GP), which is detailed in Section 2.2.1. Firstly, we introduce the change of variables $\tilde{\boldsymbol{Q}}_u^\kappa = \boldsymbol{A}_u^{\frac{1}{2}}\boldsymbol{Q}_u^\kappa\boldsymbol{A}_u^{\frac{1}{2}}$, $u=1,2$, where $\boldsymbol{A}_u \succeq 0$ is defined in Proposition 4.1. With it, maximization (4.64) turns into:

$$\begin{aligned} R_\kappa\left(\alpha, \boldsymbol{\Phi}_{1:N}\right) &= \max_{\tilde{\boldsymbol{Q}}_1^\kappa, \tilde{\boldsymbol{Q}}_2^\kappa \succeq 0} \sum_{p=1}^{2} \theta_p \cdot \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{p} \boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_u\boldsymbol{A}_u^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}_u^\kappa\boldsymbol{A}_u^{-\frac{1}{2}}\boldsymbol{H}_u^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\right), \\ &\text{s.t. } \operatorname{tr}\left\{\tilde{\boldsymbol{Q}}_u\right\} + \tfrac{N_o}{2}\operatorname{tr}\left\{\textstyle\sum_{i=1}^N \boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^\dagger\right\} \le \kappa P_u, \ u=1,2 \end{aligned} \tag{4.72}$$

Now, the feasible set is defined in terms of a trace constraint, for which the GP method can be more easily applied. Likewise, notice that the feasible set is the cartesian product of convex sets.

The algorithm iterates as follows: denote by $f\left(\tilde{\boldsymbol{Q}}_1^\kappa, \tilde{\boldsymbol{Q}}_2^\kappa\right)$ the objective of (4.72), and consider the initial point $\left\{\tilde{\boldsymbol{Q}}_1^\kappa(0), \tilde{\boldsymbol{Q}}_2^\kappa(0)\right\}$, then

$$\tilde{\boldsymbol{Q}}_u^\kappa(t+1) = \tilde{\boldsymbol{Q}}_u^\kappa(t) + \gamma_t\left(\boldsymbol{K}_u - \tilde{\boldsymbol{Q}}_u^\kappa(t)\right), \quad u=1,2. \tag{4.73}$$

where t is the iteration index, $\gamma_t \in (0,1]$ the step-size, and

$$\boldsymbol{K}_u = \left[\tilde{\boldsymbol{Q}}_u^\kappa(t) + s_t \cdot \nabla_{\tilde{\boldsymbol{Q}}_u^\kappa} f\left(\tilde{\boldsymbol{Q}}_1^\kappa(t), \tilde{\boldsymbol{Q}}_2^\kappa(t)\right)\right]^\perp, \quad u=1,2. \tag{4.74}$$

$s_t \ge 0$ is a scalar, $\nabla_{\tilde{\boldsymbol{Q}}_u^\kappa} f(\cdot)$ is the gradient of $f(\cdot)$ with respect to $\tilde{\boldsymbol{Q}}_u^\kappa$ and $[\cdot]^\perp$ is the projection (with respect to the Frobenius norm) onto the feasible set. As shown in Corollary 2.2, whenever γ_t and s_t are chosen appropriately (using *e.g.*, the limited maximization rule or Armijo's rule [48, Section 2.3.1]), the sequence $\left\{\tilde{\boldsymbol{Q}}_1^\kappa(t), \tilde{\boldsymbol{Q}}_2^\kappa(t)\right\}$ is proven to converge to the maximum of (4.72). In order to make the algorithm work for our problem, we need to: *i*) compute the gradient of $f(\cdot)$ and *ii*) calculate the projection of Hermitian, semi-definite positive, matrices onto the feasible set. First, the gradient is equal to twice the conjugate of the partial derivative of the function with respect $\tilde{\boldsymbol{Q}}_u^\kappa$ [83]:

$$\nabla_{\tilde{\boldsymbol{Q}}_u^\kappa} f\left(\tilde{\boldsymbol{Q}}_1^\kappa, \tilde{\boldsymbol{Q}}_2^\kappa\right) = \left(2\left[\frac{\partial f\left(\tilde{\boldsymbol{Q}}_1^\kappa, \tilde{\boldsymbol{Q}}_2^\kappa\right)}{\partial \tilde{\boldsymbol{Q}}_u^\kappa}\right]^T\right)^\dagger \tag{4.75}$$

$$= \left(2\sum_{p=u}^{2}\theta_p \cdot \boldsymbol{A}_u^{-\frac{1}{2}}\boldsymbol{H}_u^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\left(\boldsymbol{I} + \sum_{j=1}^{p}\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_j\boldsymbol{A}_j^{-\frac{1}{2}}\tilde{\boldsymbol{Q}}_j^\kappa\boldsymbol{A}_j^{-\frac{1}{2}}\boldsymbol{H}_j^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\right)^{-1}\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_u\boldsymbol{A}_u^{-\frac{1}{2}}\right)^\dagger \log_2 e$$

The gradient is Hermitian and semi-definite positive. In order to project it onto the feasible set, we recall that this is defined in (4.72) by means of two trace constraints. Hence, we can make use of [84, Section IV.C.2] to project Hermitian matrices $\tilde{\boldsymbol{Q}}_u \succeq 0$, $u = 1, 2$ (with eigen-decomposition $\tilde{\boldsymbol{Q}}_u = \boldsymbol{U}_u \boldsymbol{\eta}_u \boldsymbol{U}_u^\dagger$) as:

$$\left[\tilde{\boldsymbol{Q}}_u\right]^\perp = \begin{cases} \tilde{\boldsymbol{Q}}_u & \text{if} \quad \operatorname{tr}\left\{\tilde{\boldsymbol{Q}}_u\right\} + \frac{N_o}{2}\operatorname{tr}\left\{\sum_{i=1}^N \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\dagger\right\} \leq \kappa P_u \\ \boldsymbol{U}_u \left[\boldsymbol{\eta} - \delta_u \boldsymbol{I}\right]^+ \boldsymbol{U}_u^\dagger & \text{otherwise} \end{cases} \tag{4.76}$$

where $\delta_u \geq 0$ is such that $\operatorname{tr}\left\{\left[\tilde{\boldsymbol{Q}}_u\right]^\perp\right\} + \frac{N_o}{2}\operatorname{tr}\left\{\sum_{i=1}^N \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\dagger\right\} = \kappa P_u$.

4.4.3 Linear Relaying Matrix Design

In the previous subsections, we devised iterative algorithms to compute the conditional SR and WSR of the system. Now, in order to search for the non-conditioned ones, we need to solve (4.62). From such a maximization, the optimum linear functions at the relays are obtained. Unfortunately, the maximization is not convex and requires unfeasible non-convex optimization procedures to be solved.

We thus propose a suboptimum approach. In particular, and motivated by the convex hull operation and the union at (4.56), we propose that the relays *time-share* among three suboptimum sets of relaying matrices: $\boldsymbol{\Phi}_{1:N}^a$, $\boldsymbol{\Phi}_{1:N}^b$ and $\boldsymbol{\Phi}_{1:N}^c$. All three based upon the extension of *amplify-and-forward* to $\kappa > 2$: on every channel use t, relays only amplify and retransmit the signal received on previous channel use $t-1$. As mentioned, with them the required memory at the relays is reduced to one sample. The resulting relaying matrices are given by:

$$\begin{aligned} &\boldsymbol{\Phi}_i^\varepsilon = \eta_i^\varepsilon \boldsymbol{\Phi}_0, \; i = 1, \cdots, N, \; \varepsilon = a, b, c \text{ with} \\ &\left[\boldsymbol{\Phi}_0\right]_{p,q} \triangleq \begin{cases} \sqrt{\beta} & p = q+1; \;\; 1 \leq q \leq \kappa - 1 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \tag{4.77}$$

where $\eta_i^\varepsilon \in \mathbb{C}$ are the beamforming weights that allow for coherent transmission and satisfy $\sum_{i=1}^N |\eta_i^\varepsilon|^2 = 1$. Notice that β in (4.77) must satisfy $\frac{1}{\kappa}\operatorname{tr}\left\{\frac{N_o}{2}\sum_{i=1}^N \boldsymbol{\Phi}_i^\varepsilon \boldsymbol{\Phi}_i^{\varepsilon\dagger}\right\} < \min\left\{P_1, P_2\right\}$ so that power constraint (4.55) can hold. That is, $\beta < \frac{\min\{P_1,P_2\}}{N_o/2}\frac{\kappa}{\kappa-1}$. Finally, we select the weights η_i^ε to be:

$$\eta_i^a \triangleq \frac{b_{1,i}^* \cdot c_i^*}{g^a}, \quad \eta_i^b \triangleq \frac{b_{2,i}^* \cdot c_i^*}{g^b}, \quad \eta_i^c \triangleq \frac{\sum_{u=1}^2 b_{u,i}^* \cdot c_i^*}{g^c}, \quad i = 1, \cdots, N. \tag{4.78}$$

where g^ε is such that $\sum_{i=1}^N |\eta_i^\varepsilon|^2 = 1$. The idea is that the weights maximize (via Maximal Ratio Transmission) the individual rates of user 1, user 2 and the sum-rate, respectively. Finally, in order to select the optimum value of β, we propose the use of a simple one-dimensional exhaustive search, considering R_1, R_2 and $R_1 + R_2$ as the objective metrics for $\boldsymbol{\eta}^a_{1:N}$, $\boldsymbol{\eta}^b_{1:N}$ and $\boldsymbol{\eta}^c_{1:N}$, respectively.

4.5 Compress-and-forward

In the previous sections, we studied the two relaying techniques that performed better for the single-user scenario. We now turn to the third relaying technique: C&F. We extend it to multiple users and show that, unlike the $N \to \infty$ case, when $U \to \infty$ C&F and D&F have different scaling laws. Interestingly, C&F is not affected by the distributed channel hardening effect.

As for Section 3.5, we assume that relays distributedly compress their received signals by means of a distributed Wyner-Ziv code [39]. Later, they send the compressed versions to the destination, which makes use of them within a coherent detector. See the previous chapter for a complete operational overview of this technique.

Theorem 4.5 *With C&F, the MPR-MAC achieves the rate region*

$$\mathcal{R}_{\text{C\&F}}\left(P_1, P_2\right) = \text{coh}\left(\bigcup_{\substack{\gamma_1,\gamma_2,\gamma_{r_1},\cdots,\gamma_{r_N}: \\ \gamma_u + \frac{\sum_{i=1}^N \gamma_{r_i}}{2} \leq P_u, \ u=1,2}} \ \bigcup_{\substack{\boldsymbol{\rho}_{1:N} \geq 0: \\ \boldsymbol{\rho}_{1:N} \in \mathcal{J}\left(\gamma_{1:2}, \gamma_{r_{1:N}}\right)}} \left\{ R_{1,2} : \right. \right. \tag{4.79}$$

$$\left.\left. \sum_{u \in \mathcal{G}} R_u \leq \log_2 \det \left(\boldsymbol{I} + \sum_{u \in \mathcal{G}} \gamma_u \boldsymbol{h}_u \boldsymbol{h}_u^\dagger \right) \mid \mathcal{G} \subseteq \{1, 2\} \right\} \right),$$

where

$$\boldsymbol{h}_u = \left[\frac{a_u}{\sqrt{N_o}}, \frac{b_{u,1}}{\sqrt{N_o + \rho_1}}, \cdots, \frac{b_{u,N}}{\sqrt{N_o + \rho_N}} \right]^T, \tag{4.80}$$

and $\mathcal{J}\left(\boldsymbol{\gamma}_{1:2}, \boldsymbol{\gamma}_{r_{1:N}}\right) \triangleq$

$$\left\{ \boldsymbol{\rho}_{1:N} : \log_2 \frac{\det\left(\boldsymbol{I} + \sum_{u=1}^2 \gamma_u \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right)}{\det\left(\boldsymbol{I} + \sum_{u=1}^2 \gamma_u \boldsymbol{h}_u\left(\mathcal{S}^c\right) \boldsymbol{h}_u\left(\mathcal{S}^c\right)^\dagger\right)} + \sum_{i \in \mathcal{S}} \mathcal{C}\left(\frac{N_o}{\rho_i}\right) \leq \right.$$

$$\left. \mathcal{C}\left(\sum_{i \in \mathcal{S}} \frac{|c_i|^2 \gamma_{r_i}}{N_o + \sum_{u=1}^2 |a_u|^2 \gamma_u} \right), \forall \mathcal{S} \subseteq \mathcal{N}. \right\} \tag{4.81}$$

Remark 4.4 *In the theorem, we used notation*

$$\boldsymbol{h}_u(\mathcal{S}) : [\boldsymbol{h}_u(\mathcal{S}\}]_{p+1} = \begin{cases} [\boldsymbol{h}_u]_{p+1} & \textit{if } p = 0 \textit{ or } p \in \mathcal{S} \\ 0 & \textit{elsewhere} \end{cases}$$

Also, γ_u and γ_{r_i} denote the power allocated to user $u = 1, 2$ and relay $i = 1, \cdots, N$, respectively. Finally, ρ_i stands for the compression noise of relay i.

Proof: Let the two sources select messages $\omega_u \in \left\{1, \cdots, 2^{nR_u}\right\}$, $u = 1, 2$ for transmission. They divide them it into B sub-messages of κR bits each, with $\kappa = \frac{n}{B}$: $\omega_u = \left[\omega_u^1, \cdots, \omega_u^B\right]$. The messages are then transmitted using block-Markov encoding within $B+1$ channel blocks. On every block b, the sources transmit the new sub-messages ω_u^b, $u = 1, 2$ using signals

$$\boldsymbol{x}_u^\kappa[b] = \boldsymbol{s}_u^\kappa\left(\omega_u^b\right), \ u = 1, 2, \tag{4.82}$$

where we have selected $\boldsymbol{s}_u^\kappa(\cdot)$ to be random Gaussian codebooks, generated *i.i.d.* from $s_u \sim \mathcal{CN}(0, \gamma_u)$. Gaussian signalling has not been shown to be optimal. The signals received at the relays are given by:

$$\boldsymbol{y}_{r_i}^\kappa[b] = \sum_{u=1}^{2} b_{u,i}\boldsymbol{s}_u^\kappa\left(\omega_u^b\right) + \boldsymbol{z}_{r_i}^\kappa, \quad i = 1, \cdots, N. \tag{4.83}$$

Those are distributedly compressed using a multi-source compression code as that in Proposition 3.1, by means of which relays map the signals $\boldsymbol{y}_{r_i}^\kappa[b]$ using functions

$$f_\kappa^{r_i} : \mathcal{Y}_{r_i}^\kappa \rightarrow \left\{1, \cdots, 2^{\kappa\phi_i}\right\}, \tag{4.84}$$

where ϕ_i is the compression rate of relay i. On block $b+1$, the relays send the indexes $s_{r_i}[b] = f_\kappa^{r_i}\left(\boldsymbol{y}_{r_i}^\kappa[b]\right)$, $i = 1, \cdots, N$ to the destination via the MAC channel. To that end, they map them onto multi-access channel codebooks $\boldsymbol{v}_{r_i}^\kappa(\cdot)$, $i = 1, \cdots, N$:

$$\boldsymbol{x}_{r_i}^\kappa[b+1] = \boldsymbol{v}_{r_i}^\kappa\left(s_{r_i}[b]\right), \ i = 1, \cdots, N. \tag{4.85}$$

We select the MAC codebooks to be Gaussian, generated *i.i.d.* from $v_{r_i} \sim \mathcal{CN}(0, \gamma_{r_i})$, $i = 1, \cdots, N$. Notice that the power utilized by the relays to cooperate with both users is $\sum_{i=1}^N \gamma_{r_i}$. Therefore, the power constraint (*A3*) is satisfied whenever

$$\gamma_u + \frac{\sum_{i=1}^N \gamma_{r_i}}{2} \leq P_u, \ u = 1, 2. \tag{4.86}$$

The received signal at the destination on block $b+1$ thus reads (4.3):

$$\boldsymbol{y}_d^\kappa[b+1] = \sum_{u=1}^{2} a_u \cdot \boldsymbol{s}_u^\kappa\left(\omega_u^{b+1}\right) + \sum_{i=1}^{N} c_i \boldsymbol{v}_{r_i}^\kappa\left(s_{r_i}[b]\right) + \boldsymbol{z}_d^\kappa. \quad (4.87)$$

Let us now define the decoding at the destination in block $b+1$. Indexes $s_{r_{1:N}}[b]$ are first recovered from $\boldsymbol{y}_d^\kappa[b+1]$. It is possible to do so *iff* the transmission rates ϕ_i lie within its capacity region (considering $\boldsymbol{s}_u^\kappa\left(\omega^{b+1}\right)$, $u=1,2$ as interference):

$$\sum_{i\in\mathcal{S}} \phi_i \leq \mathcal{C}\left(\frac{\sum_{i\in\mathcal{S}}|c_i|^2\gamma_{r_i}}{N_o + \sum_{u=1}^{2}|a_u|^2\gamma_u}\right), \forall\mathcal{S}\subseteq\mathcal{N}. \quad (4.88)$$

Once the indexes $s_{r_{1:N}}[b]$ have been estimated, the destination removes their contribution on $\boldsymbol{y}_d^\kappa[b+1]$:

$$\begin{aligned} \boldsymbol{y}_d^{'\kappa}[b+1] &= \boldsymbol{y}_d^\kappa[b+1] - \sum_{i=1}^{N} c_i \cdot \boldsymbol{x}_{r_i}^\kappa[b+1] \\ &= \sum_{u=1}^{2} a_u \boldsymbol{s}_u^\kappa\left(\omega_u^{b+1}\right) + \boldsymbol{z}_d^\kappa. \end{aligned} \quad (4.89)$$

Afterwards, the indexes $s_{r_{1:N}}[b]$ are decompressed using the received signal $\boldsymbol{y}_d^{'\kappa}[b]$ as side information. That is, following Proposition 3.1,the destination de-maps by means of function

$$g_\kappa : \left\{1,\cdots,2^{\kappa\phi_1}\right\} \times \cdots \times \left\{1,\cdots,2^{\kappa\phi_N}\right\} \times \mathcal{Y}_d^{'\kappa} \rightarrow \hat{\mathcal{Y}}_{r_{1:N}}^\kappa. \quad (4.90)$$

The de-mapped vectors $\hat{\boldsymbol{y}}_{r_{1:N}}^\kappa[b] = g_\kappa\left(s_{r_1}[b],\cdots,s_{r_N}[b],\boldsymbol{y}_d^{'\kappa}[b]\right)$ are finally used, along with $\boldsymbol{y}_d^{'\kappa}[b]$, to decode the users' messages ω_u^b, $u=1,2$. Those are correctly estimated *iff* [21,38]:

$$\sum_{u\in\mathcal{G}} R_u \leq I\left(x_\mathcal{G}; y_d^{'}, \hat{y}_{r_1},\cdots,\hat{y}_{r_N}\right) \quad \forall\mathcal{G}\subseteq\{1,2\}, \quad (4.91)$$

where the inequality follows from (3.41) in Proposition 3.1. Assume that Gaussian compression codebooks are used at the relays, although they have not been shown to be optimal. That is, $p\left(\hat{y}_i|y_i\right) = \frac{1}{\sqrt{\pi\rho_i}}\exp\left(-\frac{|\hat{y}_i-y_i|^2}{\rho_i}\right)$, where ρ_i is the variance of $\hat{y}_{r_i}$ conditioned on y_{r_i}:

$$\rho_i = \mathrm{E}\left\{|\hat{y}_{r_i} - \mathrm{E}\left\{\hat{y}_{r_i}|y_{r_i}\right\}|^2|y_{r_i}\right\}. \quad (4.92)$$

Hereafter, is referred to as compression noise. With such a Gaussian codebooks, the rate region (4.91) equals that of the AWGN SIMO-MAC:

$$\sum_{u\in\mathcal{G}} R_u \leq \log_2\det\left(\boldsymbol{I} + \sum_{u\in\mathcal{G}}\gamma_u \boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right) \quad \forall\mathcal{G}\subseteq\{1,2\}, \quad (4.93)$$

where $\boldsymbol{h}_u$ is defined in (4.80). However, this only holds for compression rates satisfying the set of constraints imposed by the D-WZ coding (3.42), *i.e.,*

$$I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\leq\sum_{i\in\mathcal{S}}\phi_i\qquad\forall\mathcal{S}\subseteq\mathcal{N}.\tag{4.94}$$

Nonetheless, notice that compression rates are also constrained by (4.88), which plugged into (4.94) makes (4.93) to hold if:

$$I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\leq\mathcal{C}\left(\frac{\sum_{i\in\mathcal{S}}|c_i|^2\gamma_{r_i}}{N_o+\sum_{u=1}^{2}|a_u|^2\gamma_u}\right)\qquad\forall\mathcal{S}\subseteq\mathcal{N}.\tag{4.95}$$

Such a constraint can be evaluated for the Gaussian compression codebook as:

$$\begin{aligned}
I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right) &= H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)-H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|\boldsymbol{y}_{\mathcal{S}},y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right) \qquad (4.96)\\
&= I\left(x_1,x_2;\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)+H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|x_1,x_2,y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)-H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|\boldsymbol{y}_{\mathcal{S}},y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\\
&= I\left(x_1,x_2;\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)+H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|x_1,x_2,y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)-H\left(\hat{\boldsymbol{y}}_{\mathcal{S}}|x_1,x_2,\boldsymbol{y}_{\mathcal{S}},y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\\
&= I\left(x_1,x_2;\hat{\boldsymbol{y}}_{\mathcal{S}}|y_d^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)+I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|x_1,x_2,y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\\
&= I\left(x_1,x_2;y_d^{'},\hat{\boldsymbol{y}}_{1:N}\right)-I\left(x_1,x_2;y_d^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)+I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|x_1,x_2,y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^{c}\right)\\
&= \log_2\frac{\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\gamma_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\gamma_u\boldsymbol{h}_u\left(\mathcal{S}^c\right)\boldsymbol{h}_u\left(\mathcal{S}^c\right)^{\dagger}\right)}+\sum_{i\in\mathcal{S}}\mathcal{C}\left(\frac{N_o}{\rho_i}\right).
\end{aligned}$$

where second equality follows from the definition of mutual information and third from the Markov chain in Proposition 3.1. Finally, fifth equality comes from the chain rule for mutual information.

Let us now define $\mathcal{J}\left(\gamma_{1:2},\gamma_{r_{1:N}}\right)$ as the set of compression noises that satisfies (4.95). The relays may arbitrary choose any codebook design within this set, and also *time-share* them. Therefore the achievable rate region remains:

$$\mathrm{coh}\left(\bigcup_{\boldsymbol{\rho}_{1:N}\in\mathcal{J}\left(\boldsymbol{\gamma}_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right)}\left\{R_{1,2}:\sum_{u\in\mathcal{G}}R_u\leq\log_2\det\left(\boldsymbol{I}+\sum_{u\in\mathcal{G}}\gamma_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)\ \forall\mathcal{G}\subseteq\{1,2\}\right\}\right).$$

Finally, the power can be arbitrarily allocated at the sources and relays, which concludes the proof. ■

Looking at the rate region (4.79), we clearly notice the parallelism with the SIMO-MAC. From it, we can show that the boundary points of (4.79) can be attained using SIC at the decoder, which will be used to compute its weighted sum-rate.

4.5.1 Weighted Sum-Rate Maximization

As pointed out for $\mathcal{R}_{\text{LR}}$, any convex region can be described using its bounding hyperplanes [82, Sec. III-C]. Therefore, it is possible to resort to the maximum weighted sum-rate (WSR) in order to characterize the region:

$$\mathcal{R}_{\text{C\&F}}\left(P_1,P_2\right)=\left\{R_{1,2}:\alpha R_1+\left(1-\alpha\right)R_2\leq R\left(\alpha\right),\forall\alpha\in\left[0,1\right]\right\}, \tag{4.97}$$

where $R\left(\alpha\right)$ is indeed the maximum WSR with C&F, given weights α and $1-\alpha$ for users s_1 and s_2, respectively.

Such a WSR is achieved at the boundary of the rate region, and attained (as mentioned) using SIC at the destination. Considering, *w.l.o.g.*, priority $\alpha \geq 0.5$ the user's achievable rates thus read:

$$\begin{aligned} R_1 &= \log_2\det\left(\boldsymbol{I}+\gamma_1\boldsymbol{h}_1\boldsymbol{h}_1^\dagger\right) \\ R_2 &= \log_2\det\left(\boldsymbol{I}+\left(\boldsymbol{I}+\gamma_1\boldsymbol{h}_1\boldsymbol{h}_1^\dagger\right)^{-1}\gamma_2\boldsymbol{h}_2\boldsymbol{h}_2^\dagger\right). \end{aligned} \tag{4.98}$$

Therefore, the maximum value of $\alpha R_1+\left(1-\alpha\right)R_2$ is

$$\begin{aligned} R\left(\alpha\right)=\max_{\gamma_{1:2},\gamma_{r_{1:N}}}\ \max_{\boldsymbol{\rho}_{1:N}\in\mathcal{J}\left(\gamma_{1:2},\gamma_{r_{1:N}}\right)} &\sum_{j=1}^{2}\theta_j\cdot\log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{j}\gamma_u\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right). \\ \text{s.t. } \gamma_u+\frac{\sum_{i=1}^{N}\gamma_{r_i}}{2}\leq P_u,\ u=1,2 \end{aligned} \tag{4.99}$$

where $\theta_1=2\alpha-1$ and $\theta_2=1-\alpha$. Clearly, the maximization is not concave: the objective function, which has to be maximized is convex on $\boldsymbol{\rho}_{1:N}$. Furthermore, the feasible set $\mathcal{J}\left(\gamma_{1:2},\gamma_{r_{1:N}}\right)$ is not convex. Therefore, non-convex methods (such as exhaustive search) need to be used. Unfortunately, the maximization involves $2\cdot N+2$ optimization variables and $2+\sum_{i=1}^{N}\binom{N}{i}$ constraints, which makes a numerical resolution unfeasible.

For the single-user case (Section 3.5), the impossibility of solving the actual rate was handled by proposing an upper bound. The upper bound was used to benchmark this technique and compare it with other relaying protocols. For the MAC, we shall proceed similarly and compute an upper bound on the weighted sum-rate, which will be used to obtain an outer region on the achievable rate region. The bound is derived by eliminating all the constrains in (4.81) but only one: the sum-rate constraint. Such a relaxation models a scenario where relays are not connected to the destination via a MAC channel, but via a common wired backhaul as *e.g.*, ethernet. Let us then define:

$$\bar{R}(\alpha) = \max_{\boldsymbol{\gamma}_{1:2},\boldsymbol{\gamma}_{r_{1:N}},\boldsymbol{\rho}_{1:N}} \sum_{j=1}^{2}\theta_j \cdot \log_2\det\left(\boldsymbol{I} + \sum_{u=1}^{j}\gamma_u \boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right). \tag{4.100}$$

$$\text{s.t. } \log_2\frac{\det\left(\boldsymbol{I}+\sum_{u=1}^{2}\gamma_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\left(1+\sum_{u=1}^{2}\gamma_u|a_u|^2/N_o\right)} + \sum_{i=1}^{N}\mathcal{C}\left(\frac{N_o}{\rho_i}\right) \leq \mathcal{C}\left(\sum_{i=1}^{N}\frac{|c_i|^2\gamma_{r_i}}{N_o+\sum_{u=1}^{2}|a_u|^2\gamma_u}\right)\Bigg\}$$

$$\gamma_u + \frac{\sum_{i=1}^{N}\gamma_{r_i}}{2} \leq P_u,\ u = 1,2$$

This is clearly an upper bound on (4.99), as we have eliminated $\sum_{i=1}^{N}\binom{N}{i} - 1$ constraints. Now, following equivalent arguments to those in (3.57)-(3.60), it is possible to define $p = \arg\max|c_i|^2$, $\gamma_{r_p} = 2\beta\cdot\min\{P_1,P_2\}$, $\gamma_{r_i}^{*} = 0,\ i\neq p$ and $\gamma_u = P_u^{'} \triangleq P_u - \frac{\gamma_{r_p}}{2}$, in order to show that:

$$\bar{R}(\alpha) = \max_{0\leq\beta\leq1} r(\beta), \tag{4.101}$$

with

$$r(\beta) = \max_{\boldsymbol{\rho}_{1:N}} \sum_{j=1}^{2}\theta_j\cdot\log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{j}P_u^{'}\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right) \tag{4.102}$$

$$\text{s.t. } \log_2\frac{\det\left(\boldsymbol{I}+\sum_{u=1}^{2}P_u^{'}\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\left(1+\sum_{u=1}^{2}P_u^{'}|a_u|^2/N_o\right)} + \sum_{i=1}^{N}\mathcal{C}\left(\frac{N_o}{\rho_i}\right) \leq \Phi_T.$$

We have defined $\Phi_T = \mathcal{C}\left(\frac{|c_p|^2 2\beta\cdot\min\{P_1,P_2\}}{N_o+\sum_{u=1}^{2}|a_u|^2P_u^{'}}\right)$. Still, the optimization is not convex. However, (4.102) can be solved using *dual decomposition*, as shown below, and (4.101) can be solved by means of a single-dimension exhaustive search.

Iterative Algorithm to Compute the Upper Bound.

This subsection proposes an algorithm to compute $\bar{R}(\alpha)$. We proceed as follows: First, we show that the duality gap for the problem (4.102) is zero. Later, we propose an iterative algorithm that solves the dual problem, thus solving the primal too. As mentioned in Section 2.2.2, the key property of the dual problem is that the coupling constraint is decoupled.

Let the Lagrangian of (4.102) be defined on $\rho_{1:N} \geq 0$ and $\lambda \geq 0$ as:

$$\begin{aligned}\mathcal{L}\left(\rho_1, \cdots, \rho_N, \lambda\right) &= \sum_{j=1}^{2} \theta_j \cdot \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{j} P_u' \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right) \\ &\quad -\lambda\left(\log_2 \frac{\det\left(\boldsymbol{I} + \sum_{u=1}^{2} P_u' \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right)}{\left(1 + \sum_{u=1}^{2} P_u' |a_u|^2 / N_o\right)} + \sum_{i=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_i}\right) - \Phi_T\right).\end{aligned} \tag{4.103}$$

The dual function $g\left(\lambda\right)$ for $\lambda \geq 0$ follows:

$$g\left(\lambda\right) = \max_{\rho_1, \cdots, \rho_N \geq 0} \mathcal{L}\left(\rho_1, \cdots, \rho_N, \lambda\right). \tag{4.104}$$

The solution of the dual problem is then obtained from

$$R' = \min_{\lambda \geq 0} g\left(\lambda\right). \tag{4.105}$$

Lemma 4.1 *The duality gap for (4.102) is zero. That is,* $r\left(\beta\right) = R'$.

Proof: We may apply the time-sharing property in Proposition 2.10 to demonstrate the zero-duality gap is demonstrated. The procedure is equivalent to that presented for the single-user case. ■

The first step in order to solve the dual problem is to compute the dual function $g\left(\lambda\right)$. In Section 3.5.1, we showed that such an optimization can be performed using a *block-coordinate* algorithm. Unfortunately, now, the maximization with respect to a single ρ_n is not clear to be uniquely attained. Hence, in order to solve (4.104), we propose to use *gradient projection* (GP) as in Section 2.2.1. Consider the initial point $\{\rho_1^0, \cdots, \rho_n^0\} > 0$, then, the iterations are defined as:

$$\rho_n^{t+1} = \rho_n^t + \gamma_t\left(\bar{\rho}_n^t - \rho_n^t\right), \quad n = 1, \cdots, N \tag{4.106}$$

where t is the iteration index and $0 < \gamma_t \leq 1$ is the step size. Also,

$$\bar{\rho}_n^t = \left[\rho_n^t + s_t \cdot \nabla_{\rho_n} \mathcal{L}\left(\lambda, \rho_1^t, \cdots, \rho_N^t\right)\right]^\perp, \quad n = 1, \cdots, N \tag{4.107}$$

with $s_t \geq 0$ a scalar and $\nabla_{\rho_n} \mathcal{L}\left(\lambda, \rho_1^t, \cdots, \rho_N^t\right)$ the gradient of $\mathcal{L}\left(\cdot\right)$ with respect to ρ_n, evaluated at $\rho_1^t, \cdots, \rho_N^t$. Finally, $[\cdot]^\perp$ denotes the projection (with respect to the Frobenius norm) onto the positive axis.

As derived in Proposition 2.8, whenever γ_t and s_t are chosen appropriately, the sequence $\{\rho_1^t, \cdots, \rho_n^t\}$ is proven to converge to a stationary of the Lagrangian. However, for global convergence to hold (*i.e.*, convergence to $g(\lambda)$), the contraction property must be satisfied [49, Proposition 3.10]. Unfortunately, we were not able to prove this property for our optimization.

In order to define the algorithm, we need to: *i*) compute the projection of a given ρ onto the positive axis, which is:

$$[\rho]^{\perp} = \max\{\rho, 0\}, \tag{4.108}$$

and *ii*) obtain the gradient of the Lagrangian with respect to a single ρ_n:

$$\nabla_{\rho_n}\mathcal{L}(\boldsymbol{\rho}_{1:N}, \lambda) = \frac{\partial \mathcal{L}(\boldsymbol{\rho}_{1:N}, \lambda)}{\partial \rho_n} \tag{4.109}$$

The Lagrangian is defined in (4.103), with $\boldsymbol{h}_u$ following (4.80):

$$\begin{aligned} \boldsymbol{h}_u &= \left[\frac{a_u}{\sqrt{N_o}}, \frac{b_{u,1}}{\sqrt{N_o+\rho_1}}, \cdots, \frac{b_{u,N}}{\sqrt{N_o+\rho_N}}\right]^T \\ &= (N_o\boldsymbol{I}+\boldsymbol{\rho})^{-\frac{1}{2}}\cdot \boldsymbol{g}_u, \end{aligned} \tag{4.110}$$

where we have set $\boldsymbol{\rho} = \operatorname{diag}(1, \rho_1, \cdots, \rho_N)$ and $\boldsymbol{g}_u = [a_u,\ b_{u,1}, \cdots, b_{u,N}]^T$. Therefore, in order to compute its derivative, we can first notice that

$$\begin{aligned} \det\left(\boldsymbol{I} + \sum_{u=1}^{2} P'_u \boldsymbol{h}_u \boldsymbol{h}_u^{\dagger}\right) &= \det\left(\boldsymbol{I} + (N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}[\boldsymbol{g}_1,\ \boldsymbol{g}_2]\operatorname{diag}(P'_1,\ P'_2)[\boldsymbol{g}_1,\ \boldsymbol{g}_2]^{\dagger}\right) \\ &= \det\left(\boldsymbol{I} + (N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\boldsymbol{G}\boldsymbol{P}\boldsymbol{G}^{\dagger}\right) \end{aligned} \tag{4.111}$$

with $\boldsymbol{G} = [\boldsymbol{g}_1,\ \boldsymbol{g}_2]$ and $\boldsymbol{P} = \operatorname{diag}(P'_1,\ P'_2)$. Accordingly, it is clear that [83]:

$$\begin{aligned} &\left[\frac{\partial \log\det\left(\boldsymbol{I}+\sum_{u=1}^{2} P'_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\partial \boldsymbol{\rho}}\right]^T = \\ &\qquad -(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\boldsymbol{G}\boldsymbol{P}\boldsymbol{G}^{\dagger}\left(\boldsymbol{I}+(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\boldsymbol{G}\boldsymbol{P}\boldsymbol{G}^{\dagger}\right)^{-1}(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}. \end{aligned} \tag{4.112}$$

From it, with $\boldsymbol{\rho}$ defined above, we can easily obtain:

$$\frac{\partial \log\det\left(\boldsymbol{I}+\sum_{u=1}^{2} P'_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\partial \rho_n} = \left[\frac{\partial \log_2\det\left(\boldsymbol{I}+\sum_{u=1}^{2} P'_u\boldsymbol{h}_u\boldsymbol{h}_u^{\dagger}\right)}{\partial \boldsymbol{\rho}}\right]_{n+1,n+1}. \tag{4.113}$$

In parallel, the same reasoning can be carried out to derive that

$$\begin{aligned} &\frac{\partial \log\det\left(\boldsymbol{I}+P'_1\boldsymbol{h}_1\boldsymbol{h}_1^{\dagger}\right)}{\partial \rho_n} = \\ &\left[-(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\boldsymbol{g}_1P'_1\boldsymbol{g}_1^{\dagger}\left(\boldsymbol{I}+(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\boldsymbol{g}_1P'_1\boldsymbol{g}_1^{\dagger}\right)^{-1}(N_o\boldsymbol{I}+\boldsymbol{\rho})^{-1}\right]_{n+1,n+1}. \end{aligned} \tag{4.114}$$

Taking this into account, it is easy to obtain the derivative of the Lagrangian in (4.103) as

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\rho}_{1:N},\lambda\right)}{\partial \rho_n} = \tag{4.115}$$
$$\left(\theta_1 \cdot \left[-\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\boldsymbol{g}_1 P_1' \boldsymbol{g}_1^\dagger\left(\boldsymbol{I}+\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\boldsymbol{g}_1 P_1'\boldsymbol{g}_1^\dagger\right)^{-1}\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\right]_{n+1,n+1}\right.$$
$$+(\theta_2-\lambda)\cdot\left[-\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\boldsymbol{G}\boldsymbol{P}\boldsymbol{G}^\dagger\left(\boldsymbol{I}+\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\boldsymbol{G}\boldsymbol{P}\boldsymbol{G}^\dagger\right)^{-1}\left(N_o\boldsymbol{I}+\boldsymbol{\rho}\right)^{-1}\right]_{n+1,n+1}$$
$$\left.+\lambda\left(\frac{N_o}{\rho_n}\frac{1}{N_o+\rho_n}\right)\right)\log 2\left(e\right).$$

(4.115) is used in the GP algorithm to obtain $g\left(\lambda\right)$. Notice that for $\alpha \leq \frac{1}{2}$, the roles of users s_1 and s_2 are interchanged, and user 1 is decoded first. This roles would also need to be interchanged in the computation of the gradient.

Once $g\left(\lambda\right)$ has been obatined, we still need to minimize it in order to solve the dual problem (4.105). This can be done following the same sub-gradient approach as that in Section 3.5.1. In this case, the subgradient is computed following (2.44) as:

$$h\left(\lambda\right) = \Phi_T - \log_2 \frac{\det\left(\boldsymbol{I}+\sum_{u=1}^2 P_u'\boldsymbol{h}_u\boldsymbol{h}_u^\dagger\right)}{\left(1+\sum_{u=1}^2 P_u'|a_u|^2/N_o\right)} - \sum_{i=1}^N \mathcal{C}\left(\frac{N_o}{\rho_i\left(\lambda\right)}\right), \tag{4.116}$$

where $\boldsymbol{\rho}_{1:N}\left(\lambda\right)$ are the limiting points of algorithm (4.106), given λ. The subgradient search is guaranteed to converge to the global minimum, and thus to $r\left(\beta\right)$ [69, Section II-B]. Finally, we propose an exhaustive search to maximize the bound over β. Algorithm 3 describes all the necessary steps.

4.5.2 Multiuser Diversity and C&F

Let us study now C&F when the number of users grows without bound. Previously in this dissertation, we have shown that: *i*) multiuser diversity makes D&F relaying asymptotically useless, and *ii*) D&F and C&F behave equally when $N \to \infty$. From both facts, a doubt assails us here: do C&F and D&F perform equally for $U \to \infty$? or, in order words, is C&F affected by the distributed channel hardening effect? In this section, we show that the answer is no.

First, we extend the result in Theorem 4.5 to $U > 2$. In particular, considering the sum-rate

Algorithm 3 Computation of the WSR Upper Bound

for $\beta = 0 : 1$ **do**
  Let $p = \arg\max |c_i|^2$ and define $\Phi_T = \mathcal{C}\left(\frac{|c_p|^2 2\beta \cdot \min\{P_1, P_2\}}{N_o + \sum_{u=1}^{2} |a_u|^2 P_u'}\right)$
  Initialize $\lambda_{\min} = 0$ and $\lambda_{\max}$
  repeat
    $\lambda = \frac{\lambda_{\max} - \lambda_{\min}}{2}$
    Obtain $\{\rho_1(\lambda), \cdots, \rho_N(\lambda)\} = \arg\max_{\boldsymbol{\rho}_{1:N}} \mathcal{L}(\rho_1, \cdots, \rho_N, \lambda)$ from Algorithm 4
    Evaluate $h(\lambda)$ as in (4.116).
    If $h(\lambda) \leq 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
  until $\lambda_{\max} - \lambda_{\min} \leq \epsilon$
  $r(\beta) = \mathcal{L}(\rho_1(\lambda), \cdots, \rho_N(\lambda), \lambda)$
end for
$\bar{R}(\alpha) = \max\ r(\beta)$

constraint only, we can generalize:

$$\sum_{u=1}^{U} R_u \leq \mathcal{SR}_{\text{C\&F}}(U) \triangleq \max_{\gamma_{1:U}, \boldsymbol{\gamma}_{r_{1:N}}, \boldsymbol{\rho}_{1:N}} \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{U} \gamma_u \boldsymbol{h}_u \boldsymbol{h}_u^{\dagger}\right). \qquad (4.117)$$
$$\text{s.t. } \boldsymbol{\rho}_{1:N} \in \mathcal{J}(\boldsymbol{\gamma}_{1:2}, \boldsymbol{\gamma}_{r_{1:N}})$$
$$\gamma_u + \frac{\sum_{i=1}^{N} \gamma_{r_i}}{U} \leq P_u,\ u = 1, \cdots, U$$

where in $\mathcal{J}(\boldsymbol{\gamma}_{1:2}, \boldsymbol{\gamma}_{r_{1:N}})$ is defined in (4.81) , and:

$$\boldsymbol{h}_u = \left[\frac{a_u}{\sqrt{N_o}}, \frac{b_{u,1}}{\sqrt{N_o + \rho_1}}, \cdots, \frac{b_{u,N}}{\sqrt{N_o + \rho_N}}\right]^T,\ u = 1, \cdots, U. \qquad (4.118)$$

As previously, assume unitary-mean, Rayleigh fading. Also, assume identical power constraints: $P_u = P$, $u = 1, \cdots, U$. The following convergence in probability can be proven[3].

Theorem 4.6 *Consider $\mathcal{SR}_{\text{C\&F}}(U)$ in (4.117), and define*

$$T(U) = \mathcal{C}\left(\frac{P}{N_o} \sum_{u=1}^{U} \max\left\{|a_u|^2, |c_{1:N}|^2\right\}\right). \qquad (4.119)$$

Then, $\mathcal{SR}_{\text{C\&F}}(U) \leq T(U)$, and for $U \to \infty$:

$$T(U) - \mathcal{C}\left(U \cdot \left(\max\left\{|c_{1:N}|^2\right\} + e^{-\max\left\{|c_{1:N}|^2\right\}}\right) \cdot \frac{P}{N_o}\right) \xrightarrow{P} 0. \qquad (4.120)$$

[3]As mentioned, when obtaining convergence for $U \to \infty$, we assume that the relay-destination channels c_i are static.

Algorithm 4 GP to obtain $g(\lambda)$

Initialize $\rho_n^0 = 10^6$, $n = 1, \cdots, N$ and $t = 0$
repeat
  Compute the gradient $g_n^t = \nabla_{\rho_n} \mathcal{L}(\lambda, \rho_1^t, \cdots, \rho_N^t)$, $n = 1, \cdots, N$ from (4.115).
  Choose appropriate s_t
  Set $\hat{\rho}_n^t = \rho_n^t + s_t \cdot g_n^t$. Then, $\bar{\rho}_n^t = \max\{\hat{\rho}_n^t, 0\}$, $n = 1, \cdots, N$.
  Choose appropriate γ_t
  Update $\rho_n^{t+1} = \rho_n^t + \gamma_t(\bar{\rho}_n^t - \rho_n^t)$, $n = 1, \cdots, N$
  $t = t + 1$
until The sequence converges $\{\rho_1^t, \cdots, \rho_N^t\} \rightarrow \{\rho_1^*, \cdots, \rho_N^*\}$
Return $\{\rho_1^*, \cdots, \rho_N^*\}$

Remark 4.5 *Notice that the function $y + e^{-y} > 1$ for all $y > 0$. Therefore, $T(U)$ converges to a value greater than the sum-capacity without relays: $\mathcal{C}\left(U \cdot \frac{P}{N_o}\right)$.*

Proof: In order to prove the theorem, we first focus on the definition of $\mathcal{SR}_{\text{C\&F}}(U)$ in (4.117). Specifically, we consider its first constraint $\boldsymbol{\rho}_{1:N} \in \mathcal{J}(\boldsymbol{\gamma}_{1:2}, \boldsymbol{\gamma}_{r_{1:N}})$, and take its particularization to $\mathcal{S} = \mathcal{N}$:

$$\log_2 \frac{\det\left(\boldsymbol{I} + \sum_{u=1}^{U} \gamma_u \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right)}{\left(1 + \sum_{u=1}^{U} \gamma_u \frac{|a_u|^2}{N_o}\right)} + \sum_{i=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_i}\right) \leq \mathcal{C}\left(\sum_{i=1}^{N} \frac{|c_i|^2 \gamma_{r_i}}{N_o + \sum_{u=1}^{2} |a_u|^2 \gamma_u}\right) \tag{4.121}$$

It is clear that, at the optimum point of maximization (4.117), *i.e.*, $\gamma_{1:U}^*, \boldsymbol{\gamma}_{r_{1:N}}^*, \boldsymbol{\rho}_{1:N}^*$, the constraint is satisfied. Therefore:

$$\begin{aligned}
\log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{U} \gamma_u^* \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right) &\leq \mathcal{C}\left(\sum_{u=1}^{U} \gamma_u^* \frac{|a_u|^2}{N_o}\right) + \left(\mathcal{C}\left(\sum_{i=1}^{N} \frac{|c_i|^2 \gamma_{r_i}^*}{N_o + \sum_{u=1}^{U} |a_u|^2 \gamma_u^*}\right)\right. \\
&\quad \left. - \sum_{i=1}^{N} \mathcal{C}\left(\frac{N_o}{\rho_i^*}\right)\right) \\
&\leq \mathcal{C}\left(\sum_{u=1}^{U} \gamma_u^* \frac{|a_u|^2}{N_o}\right) + \mathcal{C}\left(\sum_{i=1}^{N} \frac{|c_i|^2 \gamma_{r_i}^*}{N_o + \sum_{u=1}^{U} |a_u|^2 \gamma_u^*}\right) \\
&= \mathcal{C}\left(\frac{\sum_{u=1}^{U} |a_u|^2 \gamma_u^* + \sum_{i=1}^{N} |c_i|^2 \gamma_{r_i}^*}{N_o}\right) \\
&= \mathcal{C}\left(\frac{\sum_{u=1}^{U}\left(|a_u|^2 \gamma_u^* + \sum_{i=1}^{N} |c_i|^2 \frac{\gamma_{r_i}^*}{U}\right)}{N_o}\right)
\end{aligned}$$

where the first inequality follows from basic manipulation of (4.121). Consider now the power constraint of (4.117):

$$\gamma_u^* + \frac{\sum_{i=1}^{N} \gamma_{r_i}^*}{U} \leq P, \;\; u = 1, \cdots, U. \tag{4.122}$$

It is clear that, making an analogous derivation of that in (3.58), such a constraint forces

$$|a_u|^2 \gamma_u^* + \sum_{u=1}^{N} |c_u|^2 \frac{\gamma_{r_u}^*}{U} \leq P \cdot \max\left\{|a_u|^2, |c_{1:N}|^2\right\}. \tag{4.123}$$

Therefore, using (4.123) into (4.122), we can obtain:

$$\log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{U} \gamma_u^* \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right) \leq \mathcal{C}\left(\frac{\sum_{u=1}^{U} P \cdot \max\{|a_u|^2, |c_{1:N}|^2\}}{N_o}\right) \tag{4.124}$$

Finally, notice that the sum-rate is defined in (4.117) as

$$\mathcal{SR}_{\mathrm{C\&F}}(U) \triangleq \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{U} \gamma_u^* \boldsymbol{h}_u \boldsymbol{h}_u^\dagger\right). \tag{4.125}$$

Therefore, making use of (4.124) and (4.125), we demonstrate that $\mathcal{SR}_{\mathrm{C\&F}}(U) \leq T(U)$.

Now, it remains to prove the asymptotic convergence of $T(U)$ in the number of users. Recall that, when computing the asymptotic performance on U, we assume the relay-destination channels $c_{1:N}$ fixed. Let us then define the random variables

$$\chi_u = \max\left\{|a_u|^2, |c_{1:N}|^2\right\}, \; u = 1, \cdots, U, \tag{4.126}$$

so that $T(U) = \mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U} \chi_u\right)$. Notice that, for fixed $c_{1:N}$, the only randomness on χ_u is contributed by $|a_u|^2$. Hence, the *p.d.f.* of those random variables can be easily computed (conditioned on the given realization of $c_{1:N}$) as:

$$f_{\chi_u|c_{1:N}}(x) = \begin{cases} e^{-x} & x > \max\{|c_{1:N}|^2\} \\ \left(1 - e^{-\max\left\{|c_{1:N}|^2\right\}}\right) \cdot \delta\left(x - \max\{|c_{1:N}|^2\}\right) & x = \max\{|c_{1:N}|^2\} \\ 0 & x < \max\{|c_{1:N}|^2\} \end{cases} \tag{4.127}$$

which is obtained by noting that $|a_u|^2$ is unitary mean, exponentially distributed. The mean of

χ_u, $u = 1, \cdots, U$ can be thus computed as

$$
\begin{aligned}
\mathrm{E}\left\{\chi_u | c_{1:N}\right\} &= \int_0^{\infty} x \cdot f_{\chi_u|c_{1:N}}(x)\, dx \qquad (4.128)\\
&= \max\left\{|c_{1:N}|^2\right\}\left(1 - e^{-\max\left\{|c_{1:N}|^2\right\}}\right) + \int_{\max\{|c_{1:N}|^2\}}^{\infty} x \cdot e^{-x} dx\\
&= \max\left\{|c_{1:N}|^2\right\}\left(1 - e^{-\max\left\{|c_{1:N}|^2\right\}}\right)\\
&\quad + \left(\int_{\max\{|c_{1:N}|^2\}}^{\infty} e^{-x} dx - xe^{-x}\right]_{\max\{|c_{1:N}|^2\}}^{\infty}\Bigg)\\
&= \max\left\{|c_{1:N}|^2\right\}\left(1 - e^{-\max\left\{|c_{1:N}|^2\right\}}\right)\\
&\quad + \left(e^{-\max\left\{|c_{1:N}|^2\right\}} + \max\left\{|c_{1:N}|^2\right\} e^{-\max\left\{|c_{1:N}|^2\right\}}\right)\\
&= \max\left\{|c_{1:N}|^2\right\} + e^{-\max\left\{|c_{1:N}|^2\right\}},
\end{aligned}
$$

where third equality comes from integration by parts. We realize now that all $\boldsymbol{\chi}_{1:N}$ are conditionally independent, given $c_{1:N}$, and all have the same non-zero mean. Therefore, we can apply Corollary 2.3 to derive:

$$
\mathcal{C}\left(\frac{P}{N_o}\sum_{u=1}^{U}\chi_u\right) - \mathcal{C}\left(\frac{P}{N_o}U\mathrm{E}\left\{\chi_u|c_{1:N}\right\}\right) \xrightarrow{P} 0. \qquad (4.129)
$$

That is,

$$
T(U) - \mathcal{C}\left(U \cdot \left(\max\left\{|c_{1:N}|^2\right\} + e^{-\max\left\{|c_{1:N}|^2\right\}}\right) \cdot \frac{P}{N_o}\right) \xrightarrow{P} 0, \qquad (4.130)
$$

which concludes the proof. ∎

The obtained result suggests that, unlike D&F, multiuser diversity does not undermine C&F. In other words, distributed channel hardening is not harmful to this relaying strategy. Unfortunately, the asymptotic performance is only demonstrated for the upper bound $T(U)$, which reduces the value of our result. Indeed, we were not able to derive any asymptotic convergence of the true achievable sum-rate: $\mathcal{SR}_{\mathrm{C\&F}}(U)$. However, we conjecture that it will have the same scaling law as $T(U)$.

4.6 Numerical Results

The rate region of the MPR-MAC with the three relaying techniques analyzed in this chapter are evaluated numerically for fading channels. We model channel fading as zero-mean,

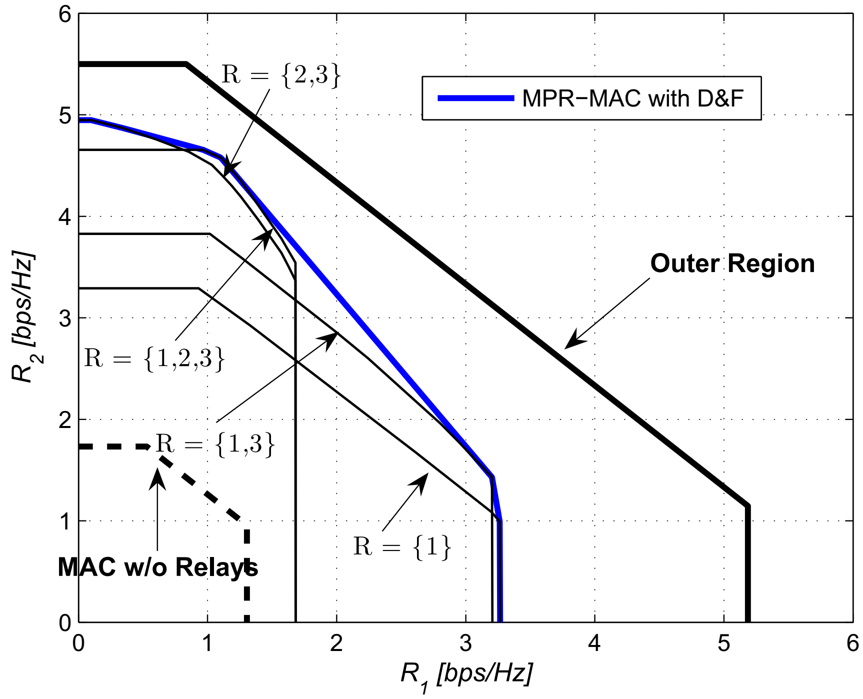

Figure 4.3: Achievable rate region of the two-user MPR-MAC with DF, $N = 3$ relays, and per-user constraint $\frac{P_u}{N_o} = 5$ dB.

complex, Gaussian-distributed and *time-invariant*. In particular, we consider $a_u \sim \mathcal{CN}(0,1)$, $b_{u,i} \sim \mathcal{CN}(0, d^{-\alpha})$, and $c_i \sim \mathcal{CN}\left(0, (1-d)^{-\alpha}\right)$, $u = 1, 2$, $i = 1, \cdots, N$, where $d_{s,d} = 1$ is the sources-destination distance and d the users-relays distance. Finally, we set the path-loss exponent to $\alpha = 3$.

The rate regions with D&F, LR and C&F are depicted in Fig. 4.3-4.5, respectively. We have considered $N = 3$ relays, and per-user constraint $P_u/N_o = 5$ dB, $u = 1, 2$. Moreover, relays are placed at a distance $d = 0.5$, and the region is plotted for the following channel realization: $[a_1,\ a_2] = [0.68e^{-j2.47},\ 0.85e^{-j0.89}]$, $[c_1,\ c_2,\ c_3] = [1.63e^{j0.08},\ 3.35e^{-j1.17},\ 1.54e^{-j3.01}]$ and

$$[b_{u,i}] = \begin{bmatrix} 5.17e^{j2.35} & 0.85e^{j2.63} & 2.17e^{j1.25} \\ 3.72e^{j1.23} & 5.19e^{j1.53} & 5.73e^{j1.80} \end{bmatrix}$$

As reference, we also plot the capacity region of the MAC without relays and the outer region presented in Sec. 4.2.2. First, Fig. 4.3 plots results for D&F, whose rate region is computed from Theorem 4.2 as the union of the achievable rate regions with all possible subsets of active relays $\mathcal{S} \subseteq \mathcal{N}$. We notice that the maximum achievable rate for user 1 is given when only the relay $\mathcal{S} = \{1\}$ is active, while user 2 maximizes its achievable rate with relays $\mathcal{S} = \{2, 3\}$

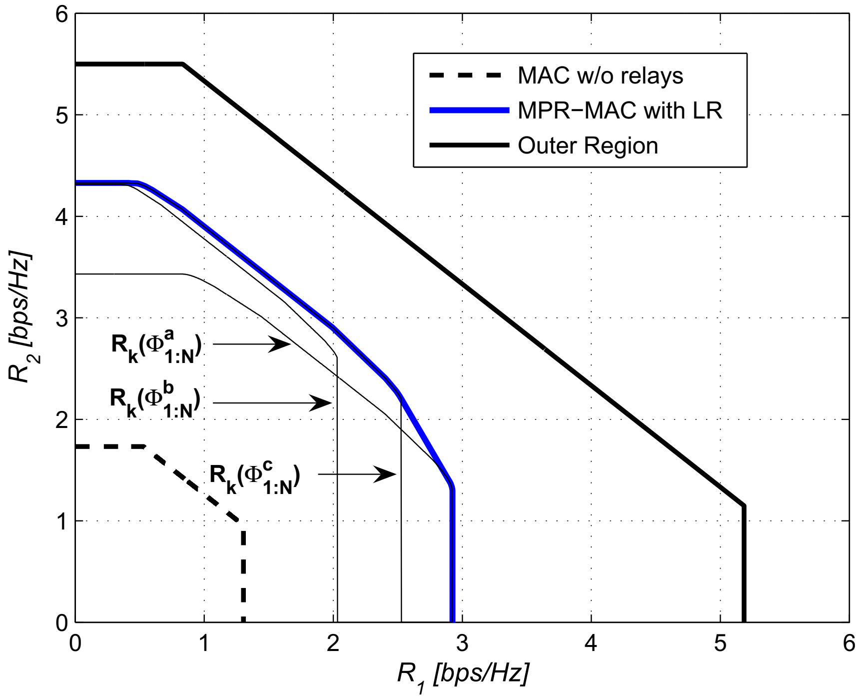

Figure 4.4: Achievable rate region of the two-user MPR-MAC with LR, $N = 3$ relays, and per-user constraint $\frac{P_u}{N_o} = 5$ dB. $\kappa = 60$.

active. Finally, it is shown that the rate region with D&F is closer to the outer bound than to the capacity region of the MAC without relays, providing more than two-fold gains with respect to the latter.

Next, Fig. 4.4 depicts the achievable rate region with LR. We compute it as the union of the conditional rate regions with linear functions $\mathbf{\Phi}^a_{1:N}$, $\mathbf{\Phi}^b_{1:N}$, $\mathbf{\Phi}^c_{1:N}$, all three presented in Section 4.4.3. We first notice that *LR* performs farther from the outer region than D&F does. This can be seen as a penalty for its simplicity, and a consequence of the noise amplification. However, LR provides significant gains with respect to the MAC with no relays. Finally, Fig. 4.5 concludes the analysis depicting the performance of C&F. In particular, the outer region presented in Section 4.5.1 is plotted, which was derived using the WSR upper bound (4.100). Those upper bounds on the bounding hyperplanes are also plotted as reference. Again, two-fold gains are shown with respect to the MAC without relays, which demonstrates that relaying is a valuable approach to overcome channel impairments in multiuser MAC.

Fig. 4.6 plots the achievable sum-rate of the two-user MPR-MAC with D&F, LR and C&F,

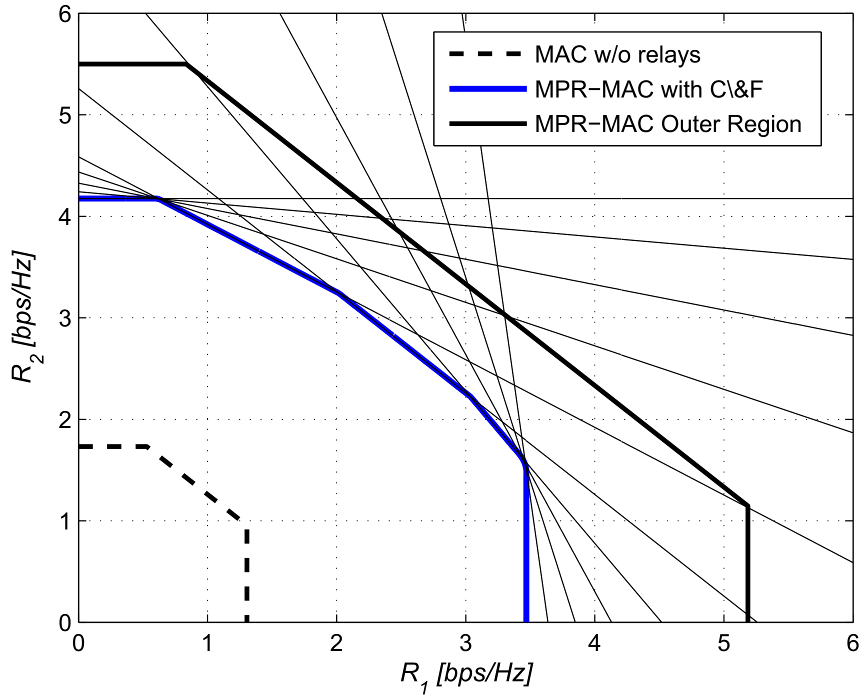

Figure 4.5: Outer bound on the achievable rate region of the two-user MPR-MAC with C&F, $N = 3$ relays, and per-user constraint $\frac{P_u}{N_o} = 5$ dB.

respectively, versus the source-relay distance d. In particular, it depicts the expected sum-rate, averaged over the joint channel distribution, assuming $N = 3$ and $P_u/N_o = 5$ dB, $u = 1, 2$. For the case of C&F, we present the sum-rate upper bound computed in (4.100). Likewise, the sum-capacity bound in (4.10) and the sum-capacity without relays are plotted as reference. Results and conclusions are slightly different to those in the previous chapter:

- Now, the *max-flow-min-cut* bound (4.10) is not reciprocal with respect to the source-relay distance d. Hence, it does not present its maximum at $d = 0.5$, but at larger d. It is clearly shown, thus, that the maximum sum-rate of the MPR-MAC is attained when relays are closer to the destination than to the users.

- As for the MPRC, D&F perfectly merges the sum-capacity upper bound for low d, thus being capacity-achieving. However, for increasing source-relay distance d, it significantly diverges from the upper bound. Hence, it becomes clearly suboptimal, which can be explained by the harmful effect of channel hardening on D&F operation.

- LR performs extremely poor. This is due, as for the MPRC, to the low number of relays

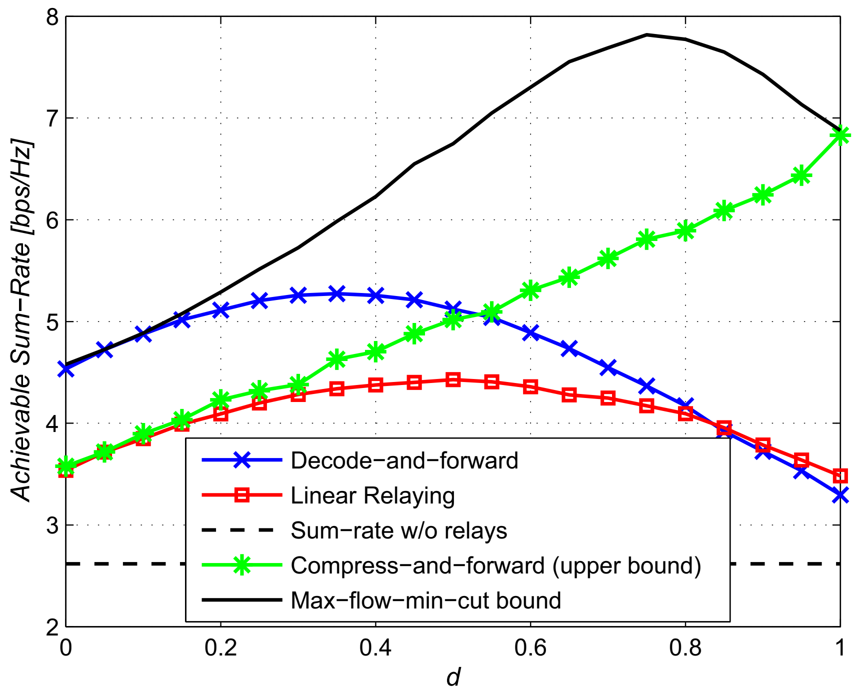

Figure 4.6: Achievable sum-rates of the two-user MPR-MAC versus source-relay distance d, considering $N = 3$ relays and per-user constraint $\frac{P_u}{N_o} = 5$ dB.

and the harmful impact of noise amplification. Also, we notice that the sum-rate with LR is reciprocal with respect to d, taking the maximum at $d = 0.5$. This invites to think that LR does not distinguish whether the users are communicating with destination or *viceversa*. In other words, that the MAC and BC with LR are dual. We demonstrate in the next chapter that this claim holds.

- C&F presents an extraordinary performance, being the most competitive approach within half of the distance range: $0.5 \leq d \leq 1$. In particular, it provides almost three-fold gains with respect to the no-relay MAC when the relays are close to the destination ($d \approx 1$). Unfortunately, the green curve is only an upper bound on the achievable sum-rate with C&F. Therefore, definite conclusions cannot be taken.

Finally, the asymptotic analysis in the number of users U is depicted in Fig. 4.7. We have considered unitary-mean, Rayleigh fading (*i.e.*, path-loss exponent $\alpha = 0$), and plotted the asymptotic sum-rates in Section 4.2.3 (outer region), Section 4.3.1 (D&F) and Section 4.5.2 (C&F). The performance of LR is obtained through simulation. From the asymptotes, we

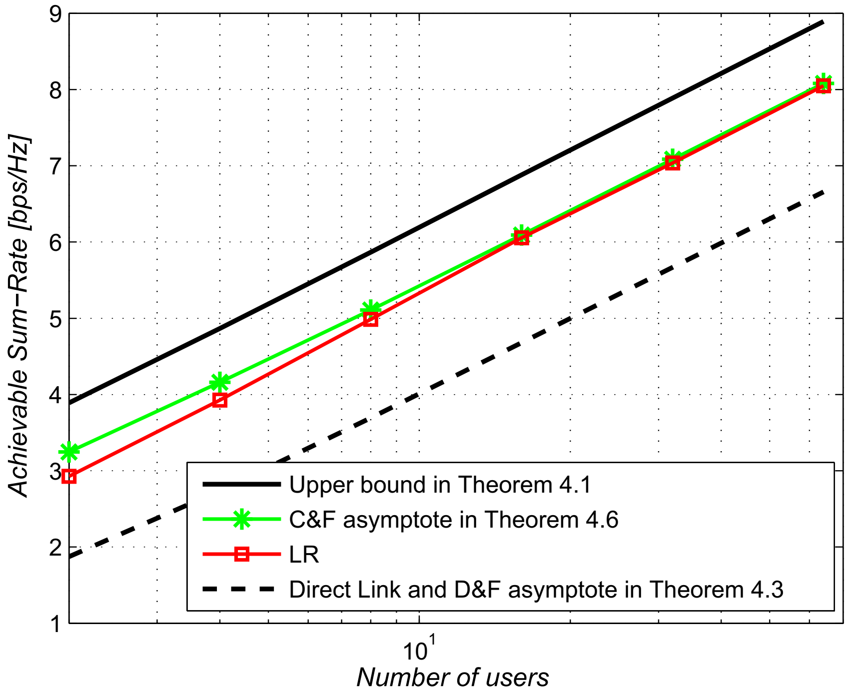

Figure 4.7: Achievable sum-rate of the MPR-MAC versus the number of users. We consider *i.i.d.*, zero-mean, Rayliegh fading and per-user constraint $\frac{P_u}{N_o} = 5$ dB.

conclude that: *i*) The sum-capacity upper bound in (4.20) presents an almost constant gain with respect to the sum-capacity without relays, independently of the number of users. This gain has to be shared among all users; hence, the higher the number of users, the lower the individual rate gain. *ii*) C&F slightly decreases its impact for higher number of users, and *iii*) LR presents a constant gain with respect to the no-relay MAC. That is, it is not affected at all by the hardening effect. However, as for the *max-flow-min-cut* bound, the per-user gain decreases for increasing number of users.

4.7 Conclusions

This chapter has presented an outer region on the capacity region of the MPR-MAC and derived inner regions by means of the achievable rates with D&F, LR and C&F.

The first achievable rate region consisted of the extension, to $N > 1$ relays, of the result presented by Kramer in [76]. We have shown that, in order to attain any point of the D&F

region, all possible subsets of active relays has to be considered, and possibly *time-shared*.

The achievable rate region with LR was next derived using theory of vector channels, and presented by means of its sum-rate and weighted sum-rate optimization. Both optimizations were shown to be convex on the sources' signalling, given a fixed set of linear functions at the relays. We obtain the optimmu osurce temporal covariances using iterative algorithms, namely: *block-coordinate* approach for the SR and *gradient projection* for the WSR. After designing the sources' signals, the optimization of the linear functions at the relays was not convex. To overcome this limitation, we proposed a suboptimum approach.

Last, C&F was studied considering distributed Wyner-Ziv coding at the relays. Again, the rate region was characterized by means of the WSR optimization. The optimization, though, was not convex and had $2 \cdot N + 2$ optimization variables; numerical resolution was thus unfeasible. Thereby, we proposed a computable outer region on it that allowed for benchmarking.

Finally, the achievable sum-rates of the channel were studied when the number of users grew to infinity. We assumed unitary-mean, Rayleigh-fading, and showed that in the asymptote, D&F does not provide gain with respect to the no-relay MAC. This was due to the distributed channel hardening effect. Additionally, the asymptotic performance of C&F was derived and demonstrated to outperform the no-relay MAC always.

Numerical results for (*time-invariant*) Rayleigh fading, have shown that:

1. The maximum sum-capacity of the MPR-MAC is attained when relays are closer to the destination than to the users.

2. D&F is sum-capacity achieving for low source-relay distances. However, it performs poorly when relays are placed close to the destination and/or the number of users increases.

3. C&F has become a very competitive approach, providing threefold sum-rate gains for relays located close to the destination.

4. LR has the same characteristics as in the MPRC: *i*) the achievable sum-rate is reciprocal with respect to the user/relay distance, *ii*) it performs poorly with low number of relays, and *iii*) it seizes all the beamforming gain of the system with high number of them.

Chapter 5

Broadcast Channel with Multiple-Parallel Relays

5.1 Introduction

Broadcasting is one of the most widespread transmission modes in wireless networks. Well-known examples are the 3G cellular downlink channel and the radio/television broadcasting. In current times, the required (and commercialized) data rates for broadcasting services are several orders of magnitude higher than those of the uplink, as *e.g.* in UMTS and wired XDSL. Accordingly: *i*) more capable and reliable links are needed, and *ii*) more harmful the impact of fading and shadowing is. In this framework, multi-antenna processing takes paramount importance to strengthen the performance of the system. However, typical broadcast receivers are low-size handsets for which having more than one antenna is not always possible. For these cases, relays can be used in order to combat channel impairments, thus increasing the coverage, capacity and robustness of the network.

5.1.1 Overview

The Gaussian BC assisted by multiple-parallel relays (See Fig. 5.1) is studied in this chapter. Single-antenna source, relays and destinations are considered, and the channel is assumed *time-invariant* and *memory-less*. Also, as previously, transmit and receive CSI are available at the source and destinations, respectively, and all relay nodes operate in full-duplex, using the same frequency band. Notice that, unlike [43], we assume direct link between source and destinations.

We aim at analyzing how much the capacity region of the broadcast is enlarged in the presence of relays. For simplicity of exposition, the analysis is restrained to the two-user case. Unfortunately, we were not able to derive the capacity region of the channel; instead, we provide inner regions using the achievable rates with D&F, LR and C&F.

In the absence of relays, dirty paper coding is well-known to be the capacity-achieving source encoding for SISO and MIMO broadcast channels [45, 47, 85]. It consists of encoding the first user as if it were alone in the network; next, the source encodes the second user, utilizing the (non-causal) knowledge of the interference caused on it by the user 1 signal. It was shown by Costa in [47] that such a coding allows the second user to decode as if it had not interference from the first user. In contrast, the first user decodes with the signal intended to the second user as interference.

In the presence of relays such an encoding at the source has not been shown to be optimal; indeed, the MPR-BC capacity region remains unknown and so does the optimal encoding. Despite that, we select dirty paper coding as the channel encoding for the source, and adapt it here to the different relaying techniques. The contributions of this chapter can be summarized as follows:

- An achievable rate region with D&F is provided first. In order to implement D&F, we propose a source encoding composed of block-Markov plus dirty paper coding. We were not able to demonstrate the optimality of this approach.

- The achievable rate region with LR is derived next. In particular, we demonstrate that it is equal to that of the MPR-MAC under a sum-power constraint. That is, MAC-BC duality holds for LR. This result extends to BCs with direct link between source and users the

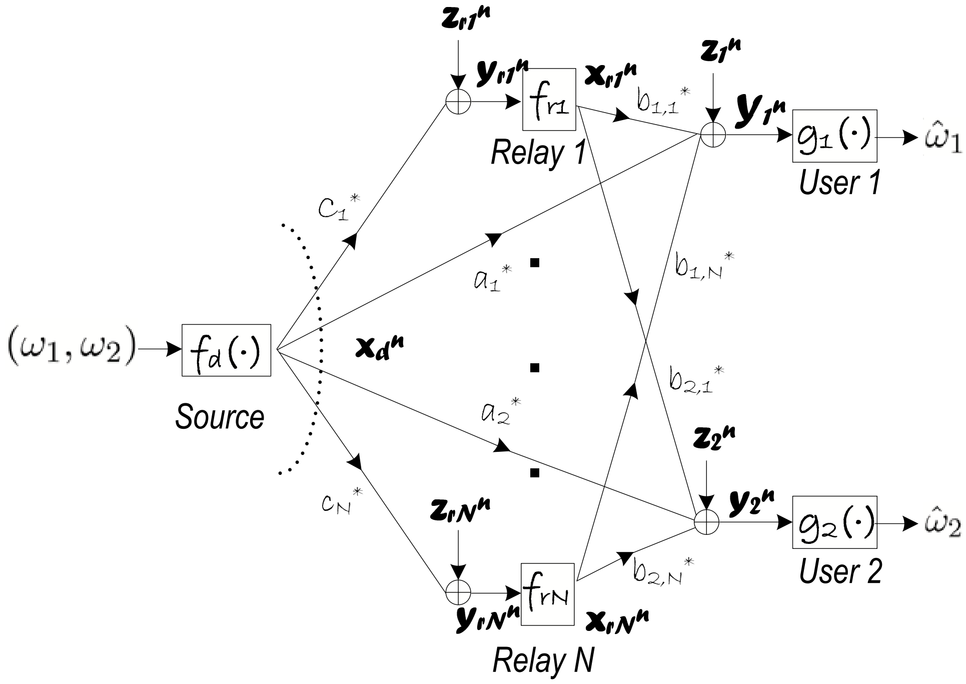

Figure 5.1: Gaussian BC with multiple-parallel relays.

result presented by Jafar *et. al.* in [43] for two-hop BC. We build upon the MIMO MAC-BC duality in [45] to prove our result. Intuitively, the obtained duality is explained by the reciprocity pointed out in Theorem 3.6, where it is shown that LR does not distinguish whether the source is transmitting to the users or *viceversa*. Finally, to maximize the weighted sum-rate and, therefore, characterize the region, we devise iterative algorithms, based upon *dual decomposition*.

- C&F is studied, and its rate region derived considering dirty paper encoding at the source, and distributed Wyner-Ziv compression at the relays.

The rest of the chapter is organized as follows: Section 5.2 presents the channel model. D&F is analyzed in Section 5.3, LR in Section 5.4, and C&F in Section 5.5. Finally, Section 5.6 shows numerical results and Section 5.7 enumerates conclusions.

5.2 Channel Model

The broadcast channel with multiple-parallel relays (MPR-BC) is a channel in which a single source d transmits simultaneously to two destinations, s_1 and s_2, with the aid of a set $\mathcal{N} = \{1, \cdots, N\}$ of parallel relays (see Fig. 5.1). All network nodes transmit/receive scalar signals, and wireless channels are *time-invariant*, *memoryless*, modeled using a complex scalar: a_u^* represents the source to user u channel, while c_i^* and $b_{u,i}^*$ represent the source to relay i and relay i to user u channels, respectively. Notice that channels are dual (*i.e.*, complex conjugate) to those of the MAC, and there is no wireless connectivity between relays.

We denote by $\boldsymbol{x}_d^n = \{x_d^t\}_{t=1}^n \in \mathbb{C}^n$ the signal transmitted by the source, where x_d^t is the transmitted symbol during channel use t and n the codeword length. The received signals at the relays thus read

$$\boldsymbol{y}_{r_i}^n = c_i^* \cdot \boldsymbol{x}_d^n + \boldsymbol{z}_{r_i}^n, \; i \in 1, \cdots, N \tag{5.1}$$

where $z_{r_i} \sim \mathcal{CN}(0, N_o)$ is additive white Gaussian noise (AWGN). The relays transmit signals $\boldsymbol{x}_{r_i}^n = \left\{x_{r_i}^t\right\}_{t=1}^n \in \mathbb{C}^n$, $i = 1, \cdots, N$, which are defined by means of causal relaying functions: $x_{r_i}^t = f_{r_i}\left(y_{r_i}^1, \cdots, y_{r_i}^{t-1}\right)$. Accordingly, the received signal at the two destination nodes is given by

$$\boldsymbol{y}_u^n \;\; = \;\; a_u^* \boldsymbol{x}_d^n + \sum_{i=1}^N b_{u,i}^* \boldsymbol{x}_{r_i}^n + \boldsymbol{z}_u^n, \;\; u = 1, 2, \tag{5.2}$$

where $z_u \sim \mathcal{CN}(0, N_o)$, $u = 1, 2$. Furthermore, the same assumptions made for the MPRC and MPR-MAC apply in this chapter. Specifically, we assume:

(*A1*). *Full-duplex operation*: relays transmit and receive simultaneously in the same frequency band. This can be implemented using different antennas for transmission and reception, the latter directly pointed to the source of the broadcast.

(*A2*). Transmit *channel state information* (CSI) and receive CSI at the source and destinations, respectively. Channel awareness includes source-destination, source-relay and relay-destination channels, and can be obtained (and fedback) during a setup phase.

(*A3*). A sum-power constraint is enforced:

$$\frac{1}{n} \sum_{t=1}^{n} \left(\mathrm{E}\left\{|x_d^t|\right\} + \sum_{i=1}^{N} \mathrm{E}\left\{|x_{r_i}^t|\right\} \right) \leq P. \tag{5.3}$$

With such a constraint, we can compare the MPR-BC with a BC without relays and source power constraint P.

5.2.1 Preliminaries

In order to present achievable rates for the channel, we define the following.

Definition 5.1 *A $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ code for the MPR-BC is defined by:*

- *two sets of messages $\mathcal{W}_u = \left\{1, \cdots, 2^{nR_u}\right\}$, $u = 1, 2$, one signal space $\mathcal{X}_d$, and a source encoding function*

$$f_d : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{X}_d^n, \tag{5.4}$$

- *N signal spaces $\mathcal{X}_i$, $i = 1, \cdots, N$ and N causal relay functions*

$$f_{r_i} : \mathcal{Y}_i^n \to \mathcal{X}_i^n,\ i = 1, \cdots, N, \tag{5.5}$$

- *two decoding functions $g_u : \mathcal{Y}_u^n \to \mathcal{W}_u$, $u = 1, 2$.*

Definition 5.2 *A rate duple (R_1, R_2) is achievable if there exists a sequence of codes $\left(n, 2^{nR_1}, 2^{nR_2}\right)$ for which $\lim_{n\to\infty} P_e^n = 0$, where*

$$P_e^n = \frac{1}{2^{nR_1}2^{nR_2}} \sum_{\omega_1,\omega_2} \Pr\left\{ \bigcup_{u=1,2} g_u\left(\boldsymbol{y}_u^n\right) \neq \omega_u | (\omega_1, \omega_2) \text{ was sent} \right\}. \tag{5.6}$$

Definition 5.3 *The capacity region of the MPR-BC is the closure of all rates duples that are achievable, given the power constraint (5.3) and the causal relaying functions in (5.5). Its characterization, even using an infinite-letter expression, is still an open problem.*

Hereafter, we provide inner regions on the capacity region of the MPR-BC by means of the achievable rates with D&F, LR and C&F.

5.3 Decode-and-forward

As mentioned throughout the dissertation, D&F consists of relay nodes fully decoding the source messages, prior to re-encoding and retransmitting them to the users. In this sense, the relays cooperate with the source forming a transmit antenna array towards the users. Therefore, D&F makes the MPR-BC to somehow mimic a MISO BC, which is a widely studied communication scenario [86, 87].

It is well-known that the capacity region of the MISO BC is achievable using dirty paper coding (DPC) at the source [47, 85]. However, when mimicking it by means of D&F relays, such a claim has not been yet demonstrated, and the optimum channel encoding remains unknown. In this dissertation, we were not able to derive the best source encoding for D&F relaying. Instead of this, we propose the use of the mixture of block-Markov encoding and DPC. The first is used to implement full-duplex relaying, while the second is used for the wireless broadcasting.

Theorem 5.1 *With D&F, the MPR-BC achieves the rate region:*

$$\mathcal{R}_{\mathrm{D\&F}}(P) = \mathrm{coh}\left(\bigcup_{m=1,\cdots,N} \bigcup_{\pi} \bigcup_{\substack{\gamma_1,\gamma_2\geq 0, \boldsymbol{g}_1,\boldsymbol{g}_2 \in \mathbb{C}^{m+1\times 1}: \\ \sum_{u=1}^{2}\gamma_u + \boldsymbol{g}_u^{\dagger}\boldsymbol{g}_u \leq P}} \left\{ R_{1,2} : \begin{array}{l} R_{\pi(1)} \leq \max_{\alpha,\alpha'} \left\{ \min\left\{ \mathcal{C}\left(|c_m|^2 \frac{\gamma_{\pi(1)}}{N_o}\right), T(\alpha,\alpha') \right\} - \right. \\ \left. \qquad \max\left\{ \frac{1}{2}\log_2\left(\frac{\gamma_{\pi(1)}+\alpha^2\gamma_{\pi(1)}}{\gamma_{\pi(1)}}\right), \log_2\left(1+\alpha'^2\right) \right\} \right\} \\ R_{\pi(2)} \leq \min\left\{ \mathcal{C}\left(\frac{|c_m|^2\gamma_{\pi(2)}}{N_o+|c_m|^2\gamma_{\pi(1)}}\right), \mathcal{C}\left(\frac{|a_{\pi(2)}|^2\gamma_{\pi(2)}+|\boldsymbol{h}_{\pi(2)}(m)\boldsymbol{g}_{\pi(2)}|^2}{N_o+|a_{\pi(2)}|^2\gamma_{\pi(1)}+|\boldsymbol{h}_{\pi(2)}(m)\boldsymbol{g}_{\pi(1)}|^2}\right) \right\} \end{array} \right\} \right) \tag{5.7}$$

where

$$T(\alpha,\alpha') = \log_2\left(\frac{\left(\sum_{u=1}^{2}\left(|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)\right)^2 (\gamma_1+\alpha^2\gamma_2)\left(1+\alpha'^2\right)}{\left(\sum_{u=1}^{2}|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)(\gamma_1+\alpha^2\gamma_2) - |a_1\gamma_1 + a_1\alpha\gamma_2|^2} \cdot \frac{1}{\left(\sum_{u=1}^{2}|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)(1+\alpha^{2\prime}1) - |\boldsymbol{h}_1(m)\,\boldsymbol{g}_1 + \boldsymbol{h}_1(m)\,\boldsymbol{g}_2\alpha'|^2} \right), \tag{5.8}$$

π *is any permutation of* $\{1,2\}$ *and the source-relay channels have been ordered as:*

$$|c_1| \geq \cdots \geq |c_m| \geq \cdots \geq |c_N|. \tag{5.9}$$

Moreover, we have defined:

$$\boldsymbol{h}_u(m) = \left[a_u^*,\ b_{u,1}^*,\cdots,b_{u,m}^*\right],\ \ u=1,2,\ m=1,\cdots,N. \tag{5.10}$$

Remark 5.1 *In the Theorem, m denotes the cardinality of the decoding set. That is, the number of relays that are active. γ_u, $u = 1, 2$ stand for the power allocated by the source to transmit new "user u" data to the relays. In turn, $\boldsymbol{g}_u$ is the distributed beamforming performed by the source and active relays to relay the data to user u. Finally, π is a given order of the dirty paper encoding.*

Proof: First of all, let the N relay nodes be ordered as in (5.9), and assume that only the subset $\mathcal{R}_m = \{1, \cdots, m\} \subseteq \mathcal{N}$ is active. Furthermore, consider the source selecting messages $\omega_u \in \left\{1, \cdots, 2^{nR_u}\right\}$, $u = 1, 2$ for transmission to user 1 and user 2, respectively. Each message is divided into B blocks of κR_u bits, with $\kappa = \frac{n}{B}$, *i.e.*, $\omega_u = \left[\omega_u^1, \cdots, \omega_u^B\right]$ $u = 1, 2$. The sub-blocks ω_u^b are then pipelined into $B + 1$ channel blocks of κ channel uses, as explained below. We consider $n, \kappa, B \gg 1$ so that $\frac{B}{B+1} \approx 1$.

The source encodes both messages using the combination of block-Markov encoding [21, Sec. 14.7] and dirty paper encoding [47], as follows: on every given block b, the source transmits the new sub-messages ω_u^b, $u = 1, 2$ to the relays in $\mathcal{R}_m$, and to the users. Simultaneously, it cooperates with the relays in $\mathcal{R}_m$ to retransmit its previously transmitted sub-messages ω_u^{b-1}, $u = 1, 2$. Assume that the source first encodes the messages intended to user 2, and then encodes the messages for user 1. Notice that such an ordering is arbitrary and can be interchanged, as shown at the end of the proof. Let us make the source and relays transmit during block b:

$$\boldsymbol{x}_d^{\kappa}[b] = \boldsymbol{s}_2^{\kappa}\left(\omega_2^b, \omega_2^{b-1}\right) + g_{d,2} \cdot \boldsymbol{v}_2^{\kappa}\left(\omega_2^{b-1}\right) + \boldsymbol{x}_1^{\kappa}[b], \tag{5.11}$$

$$\boldsymbol{x}_{r_i}^{\kappa}[b] = g_{i,2} \cdot \boldsymbol{v}_2^{\kappa}\left(\omega_2^{b-1}\right) + \boldsymbol{x}_{i,1}^{\kappa}[b], \ \ \forall\, i \in \mathcal{R}_m. \tag{5.12}$$

We have generically labeled $\boldsymbol{x}_1^{\kappa}[b]$ and $\boldsymbol{x}_{i,1}^{\kappa}[b]$ the signals intended to user one, which will be discussed latter. Focussing on the signals intended to user 2 (who is encoded first) we choose $\boldsymbol{v}_2^{\kappa}(\cdot)$ to be a Gaussian codebook, generated *i.i.d.* from $v_2 \sim \mathcal{CN}(0, 1)$. Likewise, we define $g_{d,2}$ and $g_{i,2}$, $i \in \mathcal{R}_m$ as the weights used to beamform the codeword $\boldsymbol{v}_2^{\kappa}\left(\omega_2^{b-1}\right)$ towards user 2. On the other hand, $\boldsymbol{s}_2^{\kappa}(\cdot, \cdot)$ is a multiplexed code, generated *i.i.d.* from $s_2 \sim \mathcal{CN}(0, \gamma_2)$. As mentioned, multiplexed codes consist of $\left(\kappa, 2^{\kappa R_2} \cdot 2^{\kappa R_2}\right)$ codes indexed by two entries ω_t, ω_d [79, Section III.A]. A receiver can reliably decode both ω_t and ω_d if the channel capacity is $C > 2R_2$. However, whenever it knows ω_t, it can decode ω_d if $C > R_2$, and *viceversa*. Until now, the encoding for user 2 is exactly the same as the encoding for the MPR-MAC.

Let us define now the encoding of the messages of user 1: that is, the way to create sequences $\boldsymbol{x}_1^{\kappa}[b]$ and $\boldsymbol{x}_{i,1}^{\kappa}[b]$. We consider a dirty paper approach, and follow step-by-step the reasoning in [47, Section II]. First, we define two random variables $s_1 \sim \mathcal{CN}(0, \gamma_1)$ and $v_1 \sim \mathcal{CN}(0, 1)$, independent of s_2 and v_2, and create two auxiliary variables:

$$\begin{aligned} u &= s_1 + \alpha \cdot s_2 \sim \mathcal{CN}\left(0, \gamma_1 + \alpha^2 \gamma_2\right) \\ u' &= v_1 + \alpha' \cdot v_2 \sim \mathcal{CN}\left(0, 1 + \alpha'^2\right) \end{aligned} \tag{5.13}$$

where α, α' are two given parameters. We construct now two random codebooks $\boldsymbol{u}^{\kappa}(\cdot,\cdot)$, $\boldsymbol{u}'^{\kappa}(\cdot)$, generated *i.i.d.* from u and u', respectively, and known at the source, relays and destinations. The first one is a multiplexed code of size $\left(\kappa, 2^{\kappa\Phi} \cdot 2^{\kappa\Phi}\right)$, with Φ a given constant, and has all its codewords grouped into $2^{\kappa R_1} \cdot 2^{\kappa R_1}$ bins. We denote each bin as $i_u(m, t)$, with $m, t = 1, \cdots, 2^{\kappa R_1}$. The second code is a $\left(\kappa, 2^{\kappa\Phi}\right)$ random code, and has all the sequences placed into $2^{\kappa R_1}$ bins; each bin is denoted by $i_{u'}(j)$, $j = 1, \cdots, 2^{\kappa R_1}$.

Using this codes, the source encodes the message intended to user 1 during block b as follows: first, it obtains the two bin indexes $i_u\left(\omega_1^b, \omega_1^{b-1}\right)$ and $i_{u'}\left(\omega_1^{b-1}\right)$. Next, within those bins (the first belonging to code $\boldsymbol{u}^{\kappa}(\cdot,\cdot)$ and the second to code $\boldsymbol{u}'^{\kappa}(\cdot)$), the source selects two codewords (one per codebook) which are jointly typical with $\boldsymbol{s}_2^{\kappa}\left(\omega_2^b, \omega_2^{b-1}\right)$ and with $\boldsymbol{v}_2^{\kappa}\left(\omega_2^{b-1}\right)$, respectively. As noted by Costa [47], there will be such a joint typical codeword whenever the number of codewords-per-bin on both codebooks ($N_{c.p.b}^{U}$ and $N_{c.p.b}^{U'}$) satisfy:

$$\begin{aligned} N_{c.p.b}^{U} &\geq 2^{\kappa I(u;s_2)} \\ N_{c.p.b}^{U'} &\geq 2^{\kappa I(u';v_2)}. \end{aligned} \tag{5.14}$$

Or, in other words, if the number of bins-per-codebook satisfy

$$\begin{aligned} \#\text{bins}_{\boldsymbol{u}} &\leq \frac{2^{\kappa\Phi} \cdot 2^{\kappa\Phi}}{2^{\kappa I(u;s_2)}} = 2^{\kappa(2\Phi - I(U;S_2))} \\ \#\text{bins}_{\boldsymbol{u}'} &\leq \frac{2^{\kappa\Phi}}{2^{\kappa I(u';v_2)}} = 2^{\kappa(\Phi - I(U;V_2))} \end{aligned} \tag{5.15}$$

Remember that $\boldsymbol{u}^{\kappa}(\cdot,\cdot)$ is a $\left(\kappa, 2^{\kappa\Phi} \cdot 2^{\kappa\Phi}\right)$ code, while $\boldsymbol{u}'^{\kappa}(\cdot)$ is a $\left(\kappa, 2^{\kappa\Phi}\right)$ code. However, we have previously set $\#\text{bins}_{\boldsymbol{u}} = 2^{\kappa R_1} \cdot 2^{\kappa R_1}$, and $\#\text{bins}_{\boldsymbol{u}'} = 2^{\kappa R_1}$. Therefore, we may restate (5.15) as:

$$\begin{aligned} 2R_1 &\leq 2\Phi - I(u; s_2) = 2\Phi - \log_2\left(\frac{\gamma_1 + \alpha^2\gamma_2}{\gamma_1}\right) \\ R_1 &\leq \Phi - I(u'; v_2) = \Phi - \log_2\left(1 + \alpha'^2\right) \end{aligned} \tag{5.16}$$

and, as a result

$$R_1 \quad \leq \quad \Phi - \max\left\{\frac{1}{2}\log_2\left(\frac{\gamma_1+\alpha^2\gamma_2}{\gamma_1}\right), \log_2\left(1+\alpha^{'2}\right)\right\}. \tag{5.17}$$

Now, let us denote by $\boldsymbol{u}_o$ the codeword that, belonging to bin $i_u\left(\omega_1^b, \omega_1^{b-1}\right)$, is jointly typical with $\boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right)$; and by $\boldsymbol{u}'_o$ the codeword that, belonging to bin $i_{u'}\left(\omega_1^{b-1}\right)$, is jointly typical with $\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right)$. Assume that no errors occurred during the joint typical process (*i.e.*, constraint (5.17) is satisfied). The source then constructs [47]:

$$\boldsymbol{s}_1^\kappa[b] \quad = \quad \boldsymbol{u}_o - \alpha \cdot \boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right). \tag{5.18}$$

$$\boldsymbol{v}_1^\kappa[b] \quad = \quad \boldsymbol{u}'_o - \alpha' \cdot \boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right). \tag{5.19}$$

Three important remarks must be made here: *i*) the signal $\boldsymbol{v}_1^\kappa[b]$ can also constructed by the relays. Indeed, it only depends on previous $\omega_1^{b-1}, \omega_2^{b-1}$ and not on current ω_1^b, ω_2^b, *ii*) $\boldsymbol{s}_1^\kappa[b]$ and $\boldsymbol{v}_1^\kappa[b]$ are statistically independent of $\boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right)$ and $\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right)$, due to the code construction and the joint typicality expressed before [47], and *iii*) the power constraint at the codes is satisfied with high probability: $\frac{1}{n}\sum_{t=1}^{\kappa}|s_1^t| \leq \gamma_1$ and $\frac{1}{n}\sum_{t=1}^{\kappa}|v_1^t| \leq 1$ [47].

Now, taking into account these sequences of symbols, we force the source and relays to transmit:

$$\boldsymbol{x}_d^\kappa[b] \quad = \quad \boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right) + \boldsymbol{s}_1^\kappa[b] + g_{d,2}\cdot\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right) + g_{d,1}\cdot\boldsymbol{v}_1^\kappa[b], \tag{5.20}$$

$$\boldsymbol{x}_{r_i}^\kappa[b] \quad = \quad g_{i,2}\cdot\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right) + g_{i,1}\cdot\boldsymbol{v}_1^\kappa[b], \quad \forall\, i \in \mathcal{R}_m, \tag{5.21}$$

where factors $g_{\cdot,u}$ $u = 1,2$ are, as mentioned, beamforming weights. Recall that $v_2, v_1 \sim \mathcal{CN}(0,1)$ and $s_u \sim \mathcal{CN}(0,\gamma_u)$, $u = 1,2$. Also, let us define $\boldsymbol{g}_u = [g_{d,u},\ g_{1,u},\ \cdots, g_{m,u}]^T$, $u = 1,2$ the spatial beamforming. Hence, it is easy to show that the power constraint (*A3*) is satisfied if

$$\gamma_1 + \gamma_2 + \sum_{u=1}^{2}\boldsymbol{g}_u^\dagger\boldsymbol{g}_u \leq P. \tag{5.22}$$

The received signals by relays and users is then given by

$$\boldsymbol{y}_{r_i}^\kappa[b] \quad = \quad c_i^*\cdot\left(\boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right) + \boldsymbol{s}_1^\kappa[b] + g_{d,2}\cdot\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right) + g_{d,1}\cdot\boldsymbol{v}_1^\kappa[b]\right) + z_{r_i}^\kappa,\ \forall\, i \in \mathcal{R}_m.$$

$$\boldsymbol{y}_u^\kappa[b] \quad = \quad a_u^*\cdot\left(\boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right) + \boldsymbol{s}_1^\kappa[b]\right) + \boldsymbol{h}_u(m)\cdot\left(\boldsymbol{g}_2\cdot\boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right) + \boldsymbol{g}_1\cdot\boldsymbol{v}_1^\kappa[b]\right) + \boldsymbol{z}_u^\kappa,$$

where $\boldsymbol{h}_u(m)$ was defined in (5.10). Consider now the decoding at relay $i \in \mathcal{R}_m$ during block b. Assume that the relay has reliably decoded ω_u^{b-1}, $u = 1, 2$ during the previous block. Hence, it first decodes ω_2^b from $\boldsymbol{s}_2^\kappa$. It can do so, reliably, if:

$$\begin{aligned} R_2 &\leq I(s_2; y_{r_i}|v_1, v_2) && (5.23)\\ &= \mathcal{C}\left(\frac{|c_i|^2\gamma_2}{N_o + |c_i|^2\gamma_1}\right). && (5.24)\end{aligned}$$

Therefore, all relay nodes in $\mathcal{R}_m$ can decode the message and retransmit it during the next block if:

$$\begin{aligned} R_2 &\leq \min_{i\in\mathcal{R}_m} \mathcal{C}\left(\frac{|c_i|^2\gamma_2}{N_o + |c_i|^2\gamma_1}\right). && (5.25)\\ &= \mathcal{C}\left(\frac{|c_m|^2\gamma_2}{N_o + |c_m|^2\gamma_1}\right)\end{aligned}$$

where second equality is due to the ordering in (5.9). Next, relays attempt to decode ω_1^b from $\boldsymbol{s}_1^\kappa[b]$. They look for a sequence $\boldsymbol{u}_o$ within $\boldsymbol{u}^\kappa\left(\cdot, \omega_1^{b-1}\right)$ which is joint typical with y_{r_i} given the knowledge of $\boldsymbol{s}_2^\kappa\left(\omega_2^b, \omega_2^{b-1}\right), \boldsymbol{v}_2^\kappa\left(\omega_2^{b-1}\right)$ and $\boldsymbol{v}_1^\kappa\left(\omega_1^{b-1}\right)$. Such a codeword can be found (given multiplexed codes and upon knowing ω_1^{b-1}) if:

$$\begin{aligned} \Phi &\leq \min_{i\in\mathcal{R}_m} I(u; y_{r_i}|s_2, v_1, v_2) && (5.26)\\ &= \mathcal{C}\left(|c_m|^2\frac{\gamma_1}{N_o}\right)\end{aligned}$$

where equality is due to ordering, and given the relationship in (5.13). Consider now the decoding at the users during block b. As for the MPR-MAC with D&F, users utilize backward decoding [58, Sec III.B]. That is, the users start decoding from the last block and proceed backward. Assume that, on the given block b, each user $u = 1, 2$ has successfully decoded $\omega_u^b, \cdots, \omega_u^B$. First, user 2 attempts to decode ω_2^{b-1} from $\boldsymbol{y}_2^\kappa[b]$. It can do so reliably, if and only if:

$$\begin{aligned} R_2 &\leq I(s_2, v_2; y_2) && (5.27)\\ &= \mathcal{C}\left(\frac{|a_2|^2\gamma_2 + |\boldsymbol{h}_2(m)\,\boldsymbol{g}_2|^2}{N_o + |a_2|^2\gamma_1 + |\boldsymbol{h}_2(m)\,\boldsymbol{g}_1|^2}\right)\end{aligned}$$

In turn, user 1 attempts to decode ω_1^{b-1} from $\boldsymbol{y}_1^\kappa[b]$. Recall that ω_1^{b-1} was embedded onto $\boldsymbol{u}^\kappa\left(\omega_1^b, \cdot\right)$ and $\boldsymbol{u}'^\kappa(\cdot)$. Hence, in order to decode the message, user 1 looks for the codewords $\boldsymbol{u}_o$ and $\boldsymbol{u}'_o$ that, belonging to those codebooks, are jointly typical with $\boldsymbol{y}_1^\kappa[b]$. It will find only one $\boldsymbol{u}_o$ and one $\boldsymbol{u}'_o$, whenever the codebooks length satisfy: $2^{n\Phi} \leq 2^{nI(u,u';y_1)}$. That is:

$$\begin{aligned} \Phi &\leq I(u,u';y_1) \\ &= H(y_1) - H(y_1|u,u') \\ &= H(y_1) + H(u|u') - H(y_1,u|u') \\ &= H(y_1) + H(u) - H(y_1,u|u') \\ &= H(y_1) + H(u) - H(y_1|u') - H(u|y_1,u') \\ &= H(y_1) + H(u) - H(y_1|u') - H(u|y_1) \\ &= H(y_1) + H(u) - (H(y_1,u') - H(u')) - (H(y_1,u) - H(y)) \\ &= 2 \cdot H(y_1) + H(u) + H(u') - H(y_1,u') - H(y_1,u) \end{aligned} \tag{5.28}$$

The first equality follows from the definition of mutual information, while the second, forth and sixth equalities hold due to the chain rule for entropy. The third and firth equalities are satisfied given that u and u' are independent.

Recall now that $u = s_1 + \alpha s_2$, $u' = v_1 + \alpha' v_2$, and y_1 is defined above. Therefore:

$$\begin{aligned} H(y_1) &= \log_2\left((2\pi e)\sum_{u=1}^{2}\left(|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)\right) \\ H(u) &= \log_2\left(2\pi e\left(\gamma_1 + \alpha^2\gamma_2\right)\right) \\ H(u') &= \log_2\left(2\pi e\left(1 + \alpha^{'2}\right)\right) \\ H(y_1,u) &= \log_2\left((2\pi e)^2\left(\left(\sum_{u=1}^{2}|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)\left(\gamma_1 + \alpha^2\gamma_2\right) - \right.\right. \\ &\qquad \left.\left. |a_1\gamma_1 + a_1\alpha\gamma_2|^2\ \right)\right) \\ H(y_1,u') &= \log_2\left((2\pi e)^2\left(\left(\sum_{u=1}^{2}|a_1|^2\gamma_u + |\boldsymbol{h}_1(m)\,\boldsymbol{g}_u|^2\right)\left(1 + \alpha^{2'}1\right) - \right.\right. \\ &\qquad \left.\left. |\boldsymbol{h}_1(m)\,\boldsymbol{g}_1 + \boldsymbol{h}_1(m)\,\boldsymbol{g}_2\alpha'|^2\ \right)\right) \end{aligned} \tag{5.29}$$

where forth and fifth equalities are obtained following [47, Eq. (3)].

Therefore, taking into account both constraints (5.25) and (5.27), the maximum transmission rate of user 2 is demonstrated. Furthermore, from (5.26) and (5.28), it is shown that:

$$\Phi \leq \min\left\{\mathcal{C}\left(|c_m|^2\frac{\gamma_1}{N_o}\right), I(u,u'|y_1)\right\}, \tag{5.30}$$

which plugged in (5.17) allows us to obtain

$$R_1 \leq \max_{\alpha,\alpha'} \left\{ \min \left\{ \mathcal{C}\left(|c_m|^2 \frac{\gamma_1}{N_o} \right), I\left(u, u'|y_1\right) \right\} - \max \left\{ \frac{1}{2}\log_2\left(\frac{\gamma_1 + \alpha^2 \gamma_2}{\gamma_1} \right), \log_2\left(1 + \alpha'^2\right) \right\} \right\} \quad (5.31)$$

This concludes the proof for a given DPC ordering, as well as for a given power allocation γ_u, $u = 1, 2$ and distributed beamforming $\boldsymbol{g}_u$, $u = 1, 2$. However, the source may arbitrarily select the encoding order, and may allocate power arbitrarily. Finally, the set of active relays $\mathcal{R}_m$ can be chosen from $\{\mathcal{R}_1, \cdots, \mathcal{R}_N\}$, and *time-sharing* between different sets may be used. This concludes the proof. ■

The achievable rate region with D&F is convex. However, we have not been able to characterize it. The reason is that the weighted sum-rate optimization is not convex, and involves the optimization of two vectors $\boldsymbol{g}_1, \boldsymbol{g}_2$, two scalars γ_1, γ_2, and two dirty paper orderings. Optimization turned to be unsolvable for us. Hence, in this chapter, we could not provide numerical results for D&F.

5.4 Linear Relaying

Throughout the dissertation, we have shown that previous results for vector channels can used in order to derive achievable rates with LR. In this section, we proceed similarly. In particular, utilizing the MIMO MAC-BC duality presented by Vishwanath *et al.* in [45], we prove that duality also holds for the MPR-BC and MPR-MAC with LR. Afterwards, making use of such a duality, we characterize the achievable rate region of the BC by means of its maximum weighted sum-rate.

Let the source of the BC select messages $\omega_u \in \left\{1, \cdots, 2^{nR_u}\right\}$ for transmission to user $u = 1, 2$, respectively, and map them onto two independent codewords $\boldsymbol{x}_u^n$, $u = 1, 2$. Both codewords are then transmitted simultaneously, plugged into $B = \frac{n}{\kappa}$ channel blocks of κ channel uses per block. On each block b, the source transmits the sequence $\boldsymbol{x}_d^\kappa[b] = \sum_{u=1}^2 \boldsymbol{x}_u^\kappa[b]$, which is received at the users and relays according to (5.2) and (5.1). The signals are linearly combined at the relays using the set of linear functions $\mathbf{\Phi}_{1:N} \in \mathbb{C}_{\text{SLT}}^{\kappa\times\kappa}$:

$$\boldsymbol{x}_{r_i}^{\kappa}[b] = \mathbf{\Phi}_i \left(c_i^* \cdot \sum_{u=1}^{2} \boldsymbol{x}_u^{\kappa}[b] + \boldsymbol{z}_{r_i}^{\kappa} \right), \; i = 1, \cdots, N. \tag{5.32}$$

All transmitted signals are then superimposed at the users, and received under AWGN:

$$\begin{aligned} \boldsymbol{y}_u^{\kappa}[b] &= a_u^* \boldsymbol{x}_d^{\kappa}[b] + \sum_{i=1}^{N} b_{u,i}^* \boldsymbol{x}_{r_i}^{\kappa}[b] + \boldsymbol{z}_u^{\kappa} && (5.33) \\ &= \left(a_u^* \boldsymbol{I} + \sum_{i=1}^{N} b_{u,i}^* c_i^* \mathbf{\Phi}_i \right) \cdot \sum_{j=1}^{2} \boldsymbol{x}_j^{\kappa}[b] + \left(\boldsymbol{z}_u^{\kappa} + \sum_{i=1}^{N} b_{u,i}^* \mathbf{\Phi}_i \boldsymbol{z}_{r_i}^{\kappa} \right), \quad u = 1, 2. \end{aligned}$$

Following the same arguments in (4.54), the sum-power constraint (5.3) is restated as:

$$\sum_{u=1}^{2} \mathsf{P}'_u \left(\boldsymbol{W}_u^{\kappa}, \mathbf{\Phi}_{1:N} \right) \leq P, \tag{5.34}$$

where $\boldsymbol{W}_u^{\kappa} = \mathrm{E}\left\{ \boldsymbol{x}_u^{\kappa} \left(\boldsymbol{x}_u^{\kappa} \right)^{\dagger} \right\} \succeq 0$ is the temporal covariance matrix of codeword $u = 1, 2$, and:

$$\mathsf{P}'_u \left(\boldsymbol{W}_u^{\kappa}, \mathbf{\Phi}_{1:N} \right) = \frac{1}{\kappa} \mathrm{tr} \left\{ \boldsymbol{W}_u^{\kappa} \left(\boldsymbol{I} + \sum_{i=1}^{N} |c_i|^2 \mathbf{\Phi}_i^{\dagger} \mathbf{\Phi}_i \right) + \frac{N_o}{2} \sum_{i=1}^{N} \mathbf{\Phi}_i^{\dagger} \mathbf{\Phi}_i \right\}, \; u = 1, 2. \tag{5.35}$$

The rate region of signal model (5.33) with power constraint (5.34) is presented in the next theorem.

Theorem 5.2 *With Linear relaying, the achievable rate region of the MPR-BC is*

$$\mathcal{R}_{\text{LR}}(P) = \lim_{\kappa \to \infty} \text{coh} \left(\bigcup_{\mathbf{\Phi}_{1:N} \in \mathbb{C}_{\text{SLT}}^{\kappa\times\kappa}} \mathcal{R}'_{\kappa} \left(\mathbf{\Phi}_{1:N} \right) \right) \tag{5.36}$$

where

$$\mathcal{R}'_{\kappa} \left(\mathbf{\Phi}_{1:N} \right) = \bigcup_{P_{1,2}: P_1 + P_2 = P} \mathcal{R}_{\kappa} \left(\mathbf{\Phi}_{1:N}^{\sharp} \right) \tag{5.37}$$

Remark 5.2 *Operator $[\cdot]^{\sharp}$ stands for the conjugate transpose with respect to the opposite diagonal (see the Notation). Notice that if $\boldsymbol{A}$ is strictly lower triangular, so is $\boldsymbol{A}^{\sharp}$. Also, recall that $\mathcal{R}_{\kappa}\left(\mathbf{\Phi}_{1:N}\right)$ is presented in Theorem 4.4, and is the conditional rate region of the MPR-MAC with LR.*

Remark 5.3 *We refer to $\mathcal{R}'_{\kappa}\left(\mathbf{\Phi}_{1:N}\right)$ as conditional rate region of the MPR-BC. It is the closure of rate duples that are achievable when the relays use the fixed set of linear functions $\mathbf{\Phi}_{1:N}$.*

Clearly, we notice that the MPR-BC is dual to the MPR-MAC with a sum-power constraint. Hence, Theorem 5.2 generalizes the duality presented by Jafar in [43] to BCs with direct link between source and destinations. Duality can be interpreted resorting to the well-known reciprocity of LR. Indeed, as mentioned for the MPRC (see Section 3.6.1), LR does not distinguish whether the source is transmitting to the users or viceversa.

Proof: We first prove (5.37) following two steps: *i*) show that the MPR-MAC performance is the same with functions $\mathbf{\Phi}^{\sharp}_{1:N}$ and with functions $\mathbf{\Phi}^{\dagger}_{1:N}$ (notice that the latter functions are non-causal, *i.e.* they are strictly upper-triangular matrices, and therefore non-usable in practice). *ii*) demonstrate that the MPR-MAC with linear functions $\mathbf{\Phi}^{\dagger}_{1:N}$ and sum-power constraint has the same performance than the MPR-BC with functions $\mathbf{\Phi}_{1:N}$. To prove it, we will make use of results presented by Vishwanath *et al.* in [45].

Let us consider step *i*. Assume that the two sources transmit symbol sequences $\boldsymbol{x}_u^\kappa = [x_u^1, \cdots, x_u^\kappa]^T$ $u = 1, 2$, when relays use functions $\mathbf{\Phi}^{\sharp}_{1:N}$. Also, consider that the two sources transmit a time-reversed version of previous sequences, *i.e.,* $\tilde{\boldsymbol{x}}_u^\kappa = [x_u^\kappa, \cdots, x_u^1]^T$, when relays use the non-causal functions $\mathbf{\Phi}^{\dagger}_{1:N}$. It is clear that the received signals in both cases are equal, but time-reversed. Hence, their rate performance is identical. Consider now the proof of step *ii*. First, at the MPR-MAC with functions $\mathbf{\Phi}^{\dagger}_{1:N}$, the received signal at the base station reads (4.53):

$$\boldsymbol{y}_d^\kappa = \sum_{u=1}^{2} \boldsymbol{H}_u \boldsymbol{x}_u^\kappa + \tilde{\boldsymbol{z}}_d^\kappa, \tag{5.38}$$

where the equivalent channels and noise are $\boldsymbol{H}_u = \left(a_u \boldsymbol{I} + \sum_{i=1}^N b_{u,i} c_i \mathbf{\Phi}_i^\dagger\right), u = 1, 2, \tilde{\boldsymbol{z}}_d^\kappa = \boldsymbol{z}_d^\kappa + \sum_{i=1}^N c_i \mathbf{\Phi}_i^\dagger \boldsymbol{z}_{r_i}^\kappa$. The noise temporal covariance at the destination is thus evaluated as

$$\boldsymbol{R}_d = \mathrm{E}\left\{\tilde{\boldsymbol{z}}_d^\kappa \left(\tilde{\boldsymbol{z}}_d^\kappa\right)^\dagger\right\} = N_o \left(\boldsymbol{I} + \sum_{i=1}^N |c_i|^2 \mathbf{\Phi}_i^\dagger \mathbf{\Phi}_i\right). \tag{5.39}$$

As established in Theorem 4.4, the rate region for signal model (5.38) is (4.57). Such a region $\mathcal{R}_\kappa\left(\mathbf{\Phi}^{\dagger}_{1:N}\right)$ is shown to be achievable through superposition coding at the users and SIC at the destination. Let us now consider the BC with linear functions $\mathbf{\Phi}_{1:N}$. The received signal at the users follows (5.33):

$$\boldsymbol{y}_u^\kappa = \sum_{j=1}^{2} \boldsymbol{H}_u^\dagger \boldsymbol{x}_j^\kappa + \tilde{\boldsymbol{z}}_u^\kappa, \; u = 1, 2 \tag{5.40}$$

where $\boldsymbol{H}_u$ is defined above and the equivalent noise is $\tilde{\boldsymbol{z}}_u^\kappa = \boldsymbol{z}_u^\kappa + \sum_{i=1}^N b_{u,i}^* \mathbf{\Phi}_i \boldsymbol{z}_{r_i}^\kappa$. Therefore,

$$\boldsymbol{R}_u = \mathrm{E}\left\{\tilde{\boldsymbol{z}}_u^\kappa \left(\tilde{\boldsymbol{z}}_u^\kappa\right)^\dagger\right\} = N_o \left(\boldsymbol{I} + \sum_{i=1}^N |b_{u,i}|^2 \mathbf{\Phi}_i \mathbf{\Phi}_i^\dagger\right). \tag{5.41}$$

Clearly, signal model (5.40) is equivalent to that of the MIMO BC, for which dirty paper coding (DPC) is capacity achieving [85]. Hence, the same holds for the MPR-BC: DPC achieves the conditional rate region $\mathcal{R}'_{\kappa}\left(\mathbf{\Phi}_{1:N}\right)$, defined as the set of all rates that are achievable when relays use the fixed set $\mathbf{\Phi}_{1:N}$.

Hence, in order to prove (5.37), we first need to show that $\mathcal{R}'_{\kappa}\left(\mathbf{\Phi}_{1:N}\right) \supseteq \bigcup_{\mathrm{P}_{1,2}:\sum_{u=1}^{2}\mathrm{P}_u\leq P}\mathcal{R}_{\kappa}\left(\mathbf{\Phi}_{1:N}^{\dagger}\right)$. In words: show that for every set of covariances $\boldsymbol{Q}_{1,2}^{\kappa}$ at the MAC, there exists a set of covariance at the BC $\boldsymbol{W}_{1,2}^{\kappa}$, that achieves the same rate duple and satisfies the same sum-power constraint. Let us focus on the MAC, where rate duples are achieved using SIC. Assume *w.l.o.g* that user 1 is decoded first:

$$\begin{aligned} R_u^{\mathrm{MAC}} &= \tfrac{1}{\kappa}\log_2\det\left(\boldsymbol{I}+\left(\boldsymbol{R}_d+\textstyle\sum_{j=u+1}^{2}\boldsymbol{H}_j\boldsymbol{Q}_j^{\kappa}\boldsymbol{H}_j^{\dagger}\right)^{-1}\boldsymbol{H}_u\boldsymbol{Q}_u^{\kappa}\boldsymbol{H}_u^{\dagger}\right) \\ &= \tfrac{1}{\kappa}\log_2\det\left(\boldsymbol{I}+\left(\boldsymbol{I}+\textstyle\sum_{j=u+1}^{2}\hat{\boldsymbol{H}}_j\hat{\boldsymbol{Q}}_j^{\kappa}\hat{\boldsymbol{H}}_j^{\dagger}\right)^{-1}\hat{\boldsymbol{H}}_u\hat{\boldsymbol{Q}}_u^{\kappa}\hat{\boldsymbol{H}}_u^{\dagger}\right) \end{aligned} \tag{5.42}$$

with $\hat{\boldsymbol{H}}_u=\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_u\boldsymbol{R}_u^{-\frac{1}{2}}$ and $\hat{\boldsymbol{Q}}_u^{\kappa}=\boldsymbol{R}_u^{\frac{1}{2}}\boldsymbol{Q}_u^{\kappa}\boldsymbol{R}_u^{\frac{1}{2}}$, $u=1,2$. Consider now the BC with DPC, and select user 1 to be encoded last and user 2 first. The rates achieved by the users are [45, Section III-B]:

$$\begin{aligned} R_u^{\mathrm{BC}} &= \tfrac{1}{\kappa}\log_2\left(\det\left(\boldsymbol{I}+\left(\boldsymbol{R}_u+\textstyle\sum_{j=1}^{u-1}\boldsymbol{H}_u^{\dagger}\boldsymbol{W}_j^{\kappa}\boldsymbol{H}_u\right)^{-1}\boldsymbol{H}_u^{\dagger}\boldsymbol{W}_u^{\kappa}\boldsymbol{H}_u\right)\right) \\ &= \tfrac{1}{\kappa}\log_2\left(\det\left(\boldsymbol{I}+\left(\boldsymbol{I}+\textstyle\sum_{j=1}^{u-1}\hat{\boldsymbol{H}}_u^{\dagger}\hat{\boldsymbol{W}}_j^{\kappa}\hat{\boldsymbol{H}}_u\right)^{-1}\hat{\boldsymbol{H}}_u^{\dagger}\hat{\boldsymbol{W}}_u^{\kappa}\hat{\boldsymbol{H}}_u\right)\right) \end{aligned} \tag{5.43}$$

where $\hat{\boldsymbol{H}}_u$ is defined above and $\hat{\boldsymbol{W}}_u^{\kappa}=\boldsymbol{R}_d^{\frac{1}{2}}\boldsymbol{W}_u^{\kappa}\boldsymbol{R}_d^{\frac{1}{2}}$. As mentioned, the BC rate duple must be achieved satisfying the same sum-power constraint of the MAC. Recall that the power consumed by the the MPR-MAC is computed in (4.55) and the power consumed by the MPR-BC in (5.34). Hence, the following must hold:

$$\begin{aligned} \textstyle\sum_{u=1}^{2}\mathsf{P}'_u\left(\boldsymbol{W}_u^{\kappa},\mathbf{\Phi}_{1:N}\right) &\leq \textstyle\sum_{u=1}^{2}\mathsf{P}_u\left(\boldsymbol{Q}_u^{\kappa},\mathbf{\Phi}_{1:N}^{\dagger}\right)\Leftrightarrow \\ \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\boldsymbol{W}_u^{\kappa}\left(\boldsymbol{I}+\sum_{i=1}^{N}|c_i|^2\mathbf{\Phi}_i^{\dagger}\mathbf{\Phi}_i\right)\right\} &\leq \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\boldsymbol{Q}_u^{\kappa}\left(\boldsymbol{I}+\sum_{i=1}^{N}|b_{u,i}|^2\mathbf{\Phi}_i\mathbf{\Phi}_i^{\dagger}\right)\right\}\Leftrightarrow \\ \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\boldsymbol{W}_u^{\kappa}\boldsymbol{R}_d\right\} &\leq \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\boldsymbol{Q}_u^{\kappa}\boldsymbol{R}_u\right\}\Leftrightarrow \\ \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\hat{\boldsymbol{W}}_u^{\kappa}\right\} &\leq \textstyle\sum_{u=1}^{2}\mathrm{tr}\left\{\hat{\boldsymbol{Q}}_u^{\kappa}\right\}. \end{aligned} \tag{5.44}$$

Now, we can directly apply the duality derivation for MIMO channels in [45, Eq. (8)-(11)]. It is shown therein that given $\hat{\boldsymbol{Q}}_{1,2}^{\kappa}$, it is possible to define $\boldsymbol{B}_u=\left(\boldsymbol{I}+\sum_{j=u+1}^{2}\hat{\boldsymbol{H}}_j\hat{\boldsymbol{Q}}_j^{\kappa}\hat{\boldsymbol{H}}_j^{\dagger}\right)$

and $\boldsymbol{A}_u = \left(\boldsymbol{I} + \sum_{j=1}^{u-1} \hat{\boldsymbol{H}}_u^\dagger \hat{\boldsymbol{W}}_j^\kappa \hat{\boldsymbol{H}}_u\right)$, and to take the SVD-decomposition $\boldsymbol{B}_u^{-\frac{1}{2}} \hat{\boldsymbol{H}}_u \boldsymbol{A}_u^{-\frac{1}{2}} = \boldsymbol{F}_u \boldsymbol{\Lambda}_u \boldsymbol{G}_u^\dagger$, in order to demonstrate that matrices[1]

$$\hat{\boldsymbol{W}}_u^\kappa = \boldsymbol{B}_u^{-\frac{1}{2}} \boldsymbol{F}_u \boldsymbol{G}_u^\dagger \boldsymbol{A}_u^{\frac{1}{2}} \hat{\boldsymbol{Q}}_u^\kappa \boldsymbol{A}_u^{\frac{1}{2}} \boldsymbol{G}_u \boldsymbol{F}_u^\dagger \boldsymbol{B}_u^{-\frac{1}{2}} \rightarrow \boldsymbol{W}_u^\kappa = \boldsymbol{R}_d^{-\frac{1}{2}} \hat{\boldsymbol{W}}_u^\kappa \boldsymbol{R}_d^{-\frac{1}{2}}, \; u = 1, 2. \tag{5.45}$$

satisfies that $R_u^{\mathrm{BC}} = R_u^{\mathrm{MAC}}$, $u = 1, 2$, and for them the power constraint (5.44) holds. This demonstrates that any rate duple at the MAC can be achieved at the BC. Now the converse should be proven. That is, $\mathcal{R}'_\kappa(\boldsymbol{\Phi}_{1:N}) \subseteq \bigcup_{\mathrm{P}_{1,2}:\sum_{u=1}^2 \mathrm{P}_u \leq P_T} \mathcal{R}_\kappa\left(\boldsymbol{\Phi}_{1:N}^\dagger\right)$. This is shown using equivalent arguments as those for the MAC-to-BC. We skip the proof as it does not provide new ideas. This concludes the proof of equality (5.37).

Finally, to demonstrate (5.36), we apply arguments in Theorem 4.4: linear relaying functions can be arbitrarily chosen and time-shared, and the region is non decreasing with κ. ■

In order to describe the region in Theorem 5.2, we can again resort to the WSR optimization. That is, characterize it by means of its bounding hyperplanes [82]:

$$\mathcal{R}'(P_T) = \left\{R_{1,2} : \alpha R_1 + (1-\alpha) R_2 \leq R'(\alpha), \forall \alpha \in [0,1]\right\}. \tag{5.46}$$

We refer to $R'(\alpha)$ as the maximum WSR of the broadcast channel with LR. It is clear that, given the relationship (5.36), $R'(\alpha)$ can be computed in terms of the maximum WSR of the conditional rate region. In particular, defining:

$$\mathcal{R}'_\kappa(\boldsymbol{\Phi}_{1:N}) = \left\{R_{1,2} : \alpha R_1 + (1-\alpha) R_2 \leq R'_\kappa(\alpha, \boldsymbol{\Phi}_{1:N}), \forall \alpha \in [0,1]\right\}, \tag{5.47}$$

we can state that:

$$R'(\alpha) = \lim_{\kappa\to\infty} \max_{\boldsymbol{\Phi}_{1:N} \in \mathbb{C}_{\mathrm{SLT}}^{\kappa\times\kappa}} R'_\kappa(\alpha, \boldsymbol{\Phi}_{1:N}). \tag{5.48}$$

At this point, we proceed into two consecutive steps:

- First, we provide an iterative algorithm to compute the conditional WSR $R'_\kappa(\alpha, \boldsymbol{\Phi}_{1:N})$, which is the maximum value of $\alpha R_1 + (1-\alpha) R_2$ when the relays use the fixed set $\boldsymbol{\Phi}_{1:N}$. Due to duality in Theorem 5.2, $R'_\kappa(\alpha, \boldsymbol{\Phi}_{1:N})$ is equal to the conditional WSR of the MPR-MAC with a sum-power constraint and linear functions $\boldsymbol{\Phi}_{1:N}^\sharp$.

[1] See [45, Sect. IV-B] for full details.

- Next, we study the optimization in (5.48). As it turns out to be non-convex and non-solvable, we propose a set of suboptimum linear functions.

Let us then study $R'_\kappa\left(\alpha, \mathbf{\Phi}_{1:N}\right)$, considering *w.l.o.g.* $\alpha \geq 0.5$. As mentioned, due to duality, the weighted sum-rate of the BC is equal to that of the MAC (4.64), but in this case with a sum-power constraint:

$$
\begin{aligned}
R'_\kappa\left(\alpha, \mathbf{\Phi}_{1:N}\right) &= \max_{\boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa \succeq 0} \sum_{p=1}^{2} \theta_p \cdot \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{p} \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \\
&\quad \text{s.t. } \sum_{u=1}^{2} \mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \mathbf{\Phi}_{1:N}^\sharp\right) \leq P
\end{aligned}
\tag{5.49}
$$

where we have defined $\theta_1 = \frac{2\alpha-1}{\kappa}$, $\theta_2 = \frac{1-\alpha}{\kappa}$, and $\boldsymbol{R}_d = N_o\left(\boldsymbol{I} + \sum_{i=1}^N |c_i|^2 \mathbf{\Phi}_i^\sharp \left(\mathbf{\Phi}_i^\sharp\right)^\dagger\right)$, $\boldsymbol{H}_u = \left(a_u \cdot \boldsymbol{I} + \sum_{i=1}^N b_{u,i} c_i \mathbf{\Phi}_i^\sharp\right)$, $u = 1, 2$. The optimization is continuously differentiable and convex, and the feasible set regular. Hence, it can be solved using standard convex methods. However, the constraint couples both optimization variables, preventing the direct application of the algorithms presented in Section 4.4.

Hence, if we wish to iteratively solve the optimization, we first need to decouple the constraint. We do so by using *dual decomposition*, consisting of solving the dual problem. As the optimization is convex, it has zero duality-gap and the dual problem solution solves the primal too (see Proposition 2.9). Let then the Lagrangian of (5.49) be defined on $\mu \geq 0$ and $\boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa \succeq 0$ as:

$$
\begin{aligned}
\mathcal{L}\left(\mu, \boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa\right) &= \sum_{p=1}^{2} \theta_p \cdot \log_2 \det\left(\boldsymbol{I} + \sum_{u=1}^{p} \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{Q}_u^\kappa \boldsymbol{H}_u^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \\
&\quad -\mu\left(\sum_{u=1}^{2} \mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \mathbf{\Phi}_{1:N}^\sharp\right) - P\right)
\end{aligned}
\tag{5.50}
$$

The dual function is then computed as:

$$
g\left(\mu\right) = \max_{\boldsymbol{Q}_2^\kappa, \boldsymbol{Q}_2^\kappa \succeq 0} \mathcal{L}\left(\mu, \boldsymbol{Q}_1^\kappa, \boldsymbol{Q}_2^\kappa\right). \tag{5.51}
$$

For which the constrains are now decoupled. Finally, the solution of the dual problem is

$$
R'_\kappa\left(\alpha, \mathbf{\Phi}_{1:N}\right) = \min_{\mu \geq 0} g\left(\mu\right). \tag{5.52}
$$

5.4.1 Sum-Rate Maximization

Consider first both users having the same priority $\alpha = 0.5$. For this case, $\theta_1 = 0$, and $\theta_2 = \frac{1}{2\kappa}$ in (5.50). The first task in order to obtain (5.52) is to solve maximization (5.51). For that purpose, we propose the use of a *block-coordinate ascent* algorithm, which iterates:

$$\boldsymbol{Q}_u^\kappa (t+1) = \arg \max_{\boldsymbol{Q}_u^\kappa \succeq 0} \mathcal{L}\left(\mu, \boldsymbol{Q}_u^\kappa, \boldsymbol{Q}_{\bar{u}}^\kappa (t+2-\bar{u})\right), \quad u = 1, 2, \quad \bar{u} = \{1,2\}/u. \tag{5.53}$$

The maximization above is uniquely attained, as shown in the next proposition; also, the Lagrangian is convex and the feasible set a cartesian product of semidefinite cones. Hence, the limit point of the sequence $\{\boldsymbol{Q}_1^\kappa (t), \boldsymbol{Q}_2^\kappa (t)\}$ is proven to converge to $g(\mu)$, as demonstrated in Corollary 2.1.

Proposition 5.1 *Consider the problem (5.53), define* $\boldsymbol{A}_u = \left(\boldsymbol{I} + \sum_{i=1}^N |b_{u,i}|^2 \left(\boldsymbol{\Phi}_i^\sharp\right)^\dagger \boldsymbol{\Phi}_i^\sharp\right)$ *and compute the SVD-decomposition*

$$\left(\boldsymbol{R}_d + \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa (t+2-\bar{u}) \boldsymbol{H}_{\bar{u}}^\dagger\right)^{-\frac{1}{2}} \boldsymbol{H}_u \boldsymbol{A}_u^{-\frac{1}{2}} = \boldsymbol{U}_u \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{V}_u. \tag{5.54}$$

Then, the maximization is uniquely attained at $\boldsymbol{Q}_u^\kappa (t+1) = \boldsymbol{A}_u^{-\frac{1}{2}} \boldsymbol{V}_u \boldsymbol{\Psi} \boldsymbol{V}_u^\dagger \boldsymbol{A}_u^{-\frac{1}{2}}$, *where*

$$\psi_j = \left[\frac{\theta_2 \cdot \log_2 e}{\mu} - \frac{1}{\lambda_j}\right]^+, \quad j = 1, \cdots, \kappa. \tag{5.55}$$

Proof: Consider the convex problem $\arg\max_{\boldsymbol{Q}_u^\kappa \succeq 0} \mathcal{L}\left(\mu, \boldsymbol{Q}_u^\kappa, \boldsymbol{Q}_{\bar{u}}^\kappa\right)$. The function $\mathcal{L}(\cdot)$ is defined in (5.50), where we have set $\theta_1 = 0$ and $\theta_2 = \frac{1}{2\kappa}$. First, as for (4.69), we can rewrite the objective function as:

$$\begin{aligned}\mathcal{L}\left(\mu, \boldsymbol{Q}_u^\kappa, \boldsymbol{Q}_{\bar{u}}^\kappa\right) &= \theta_2 \log_2 \det\left(\boldsymbol{I} + \boldsymbol{R}_d^{-\frac{1}{2}} \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa \boldsymbol{H}_{\bar{u}}^\dagger \boldsymbol{R}_d^{-\frac{1}{2}}\right) \\ &\quad -\mu\left(\sum_{j=1}^2 \mathsf{P}_j\left(\boldsymbol{Q}_j^\kappa, \boldsymbol{\Phi}_{1:N}^\sharp\right) - P\right) + \theta_2 \log_2 \det\left(\boldsymbol{I} + \boldsymbol{H}_e \boldsymbol{Q}_u^\kappa \boldsymbol{H}_e^\dagger\right).\end{aligned} \tag{5.56}$$

where $\boldsymbol{H}_e = \left(\boldsymbol{R}_d + \boldsymbol{H}_{\bar{u}} \boldsymbol{Q}_{\bar{u}}^\kappa \boldsymbol{H}_{\bar{u}}^\dagger\right)^{-\frac{1}{2}} \boldsymbol{H}_u$. Notice that the first term in (5.56) does not depend on $\boldsymbol{Q}_u^\kappa$. Hence, we may state the Lagrangian for the optimization as:

$$\mathcal{J}\left(\boldsymbol{\Omega}, \boldsymbol{Q}_u^\kappa\right) = \theta_2 \log_2 \det\left(\boldsymbol{I} + \boldsymbol{H}_e \boldsymbol{Q}_u^\kappa \boldsymbol{H}_e^\dagger\right) - \mu\left(\kappa \mathsf{P}_u\left(\boldsymbol{Q}_u^\kappa, \boldsymbol{\Phi}_{1:N}^\sharp\right) - \kappa P_T\right) + \operatorname{tr}\left\{\boldsymbol{\Omega}, \boldsymbol{Q}_u^\kappa\right\},$$

where $\boldsymbol{\Omega} \succeq 0$ is the Lagrange multiplier for the semidefinite constraint. The KKT conditions for the problem are thus:

$$\begin{aligned} &i) \ \frac{\mu}{\theta_2 \cdot \log_2 e}\left(\boldsymbol{I} + \sum_{i=1}^N |b_{u,i}|^2 \left(\boldsymbol{\Phi}_i^\sharp\right)^\dagger \boldsymbol{\Phi}_i^\sharp\right) - \frac{\boldsymbol{\Omega}}{\theta_2 \cdot \log_2 e} = \boldsymbol{H}_e^\dagger \left(\boldsymbol{I} + \boldsymbol{H}_e \boldsymbol{Q}_u^\kappa \boldsymbol{H}_e^\dagger\right)^{-1} \boldsymbol{H}_e \\ &ii) \ \operatorname{tr}\left\{\boldsymbol{\Omega} \boldsymbol{Q}_u^\kappa\right\} = 0 \end{aligned} \tag{5.57}$$

which are exactly identical to those in (4.70), except for the second one. Therefore, we can easily reproduce the steps (4.71) and thereafter, to prove the theorem. ■

Once the dual function has been obtained through the *block-coordinate* approach, it must be minimized over μ. As the function is convex by definition, it can be optimized using a subgradient search [68]. The search consists of following direction $-h(\mu)$, where:

$$\frac{g(\mu') - g(\mu)}{\mu' - \mu} \geq h(\mu), \quad \forall \mu' \neq \mu. \tag{5.58}$$

Such a search is proven to converge for diminishing step-size rules [48]. Considering the definition of $g(\mu)$, and revisiting Section 2.2.2, the following $h(\mu)$ satisfies the subgradient condition (5.58):

$$h(\mu) = P - \sum_{u=1}^{2} \mathsf{P}_u \left(\boldsymbol{Q}_u^{\kappa}(\mu), \boldsymbol{\Phi}_{1:N}^{\sharp} \right), \tag{5.59}$$

where $\boldsymbol{Q}_1^{\kappa}(\mu), \boldsymbol{Q}_2^{\kappa}(\mu)$ are the limiting points of (5.53). Therefore, we use it to search for the optimum μ:

$$\text{Increase } \mu \text{ if } h(\mu) \leq 0 \text{ or decrease } \mu \text{ if } h(\mu) > 0. \tag{5.60}$$

5.4.2 Weighted Sum-Rate Maximization

Consider now (5.52) for $\alpha > 0.5$. As previously, we first focus on solving (5.51). In this case, a *block-coordinate ascent* algorithm is not guaranteed to converge, as the individual optimization of the function with respect to a single covariance matrix is not proven to be uniquely attained. Thereby, in order to solve it with proven convergence, we propose the use of *Gradient Projection*. The algorithm is similar to that presented for the WSR of the MAC, and iterates as follows: let the initial point $\left\{ \tilde{\boldsymbol{Q}}_1^{\kappa}(0), \tilde{\boldsymbol{Q}}_2^{\kappa}(0) \right\}$, then

$$\boldsymbol{Q}_u^{\kappa}(t+1) = \boldsymbol{Q}_u^{\kappa}(t) + \gamma_t \left(\boldsymbol{K}_u - \boldsymbol{Q}_u^{\kappa}(t) \right), \quad u = 1, 2. \tag{5.61}$$

where t is the iteration index, $\gamma_t \in (0, 1]$ the step-size, and:

$$\boldsymbol{K}_u = \left[\boldsymbol{Q}_u^{\kappa}(t) + s_t \cdot \nabla_{\boldsymbol{Q}_u^{\kappa}} \mathcal{L}\left(\mu, \boldsymbol{Q}_1^{\kappa}(t), \boldsymbol{Q}_2^{\kappa}(t)\right) \right]^{\perp}, \quad u = 1, 2. \tag{5.62}$$

As previously, $[\cdot]^{\perp}$ denotes the projection onto the feasible set. In this case, onto the cone of semidefinite positive matrices. As mentioned, whenever γ_t and s_t are chosen appropriately, the GP converges to the maximum[2], and hence obtains $g(\mu)$ (see Corollary 2.2). In order to

[2] Recall that the optimization is convex.

implement the algorithm, we need to compute the gradient of the Lagrangian with respect to $\boldsymbol{Q}_u^\kappa$ [83]:

$$\begin{aligned}\nabla_{\boldsymbol{Q}_u^\kappa}\mathcal{L}\left(\mu,\boldsymbol{Q}_1^\kappa,\boldsymbol{Q}_2^\kappa\right) &= \left(2\left[\frac{\partial\mathcal{L}\left(\mu,\boldsymbol{Q}_1^\kappa,,\boldsymbol{Q}_2^\kappa\right)}{\partial\tilde{\boldsymbol{Q}}_u^\kappa}\right]^T\right)^\dagger \\ &= \left(2\sum_{p=u}^{2}\theta_p\cdot\boldsymbol{H}_u^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\left(\boldsymbol{I}+\sum_{j=1}^{p}\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_j\boldsymbol{Q}_j^\kappa\boldsymbol{H}_j^\dagger\boldsymbol{R}_d^{-\frac{1}{2}}\right)^{-1}\boldsymbol{R}_d^{-\frac{1}{2}}\boldsymbol{H}_u\right)^\dagger\cdot\log_2 e \\ &\quad -\mu\cdot\frac{2}{\kappa}\left(\boldsymbol{I}+\sum_{i=1}^{N}|b_{u,i}|^2\left(\boldsymbol{\Phi}_i^\sharp\right)^\dagger\boldsymbol{\Phi}_i^\sharp\right)^\dagger.\end{aligned}\tag{5.63}$$

The gradient is Hermitian. Finally, we need to project a Hermitian matrix $\boldsymbol{S}$ (with eigen-decomposition $\boldsymbol{S}=\boldsymbol{U}\boldsymbol{\eta}\boldsymbol{U}^\dagger$) onto the semidefinite positive cone. This can be done from [88, Theorem 2.1] as:

$$[\boldsymbol{S}]^\perp=\boldsymbol{U}[\boldsymbol{\eta}]^+\boldsymbol{U}^\dagger.\tag{5.64}$$

Once obtained $g\left(\mu\right)$ through GP, we minimize it using the same subgradient approach described for the sum-rate. The conditional WSR is thus obtained with guaranteed convergence.

5.4.3 Linear Relaying Matrix Design

As previously, in order to obtain the non-conditioned WSR, $R'\left(\alpha\right)$, we need to solve (5.48). However, the optimization is not convex, and no solution can be given. We thus consider a sub-optimum approach. In particular, we propose the relays to *time-share* among the suboptimum linear functions defined for the MPR-MAC in (4.77) and (4.78):

$$\begin{aligned}&\boldsymbol{\Phi}_i^\varepsilon=\eta_i^\varepsilon\boldsymbol{\Phi}_0,\ i=1,\cdots,N,\ \varepsilon=a,b,c\text{ with}\\ &[\boldsymbol{\Phi}_0]_{p,q}\triangleq\begin{cases}\sqrt{\beta} & p=q+1;\ \ 1\leq q\leq\kappa-1\\ 0 & \text{elsewhere.}\end{cases}\end{aligned}\tag{5.65}$$

As in the previous chapter, $\eta_i^\varepsilon\in\mathbb{C}$ are the beamforming weights among relays, which allow for coherent transmission and satisfy $\sum_{i=1}^N|\eta_i^\varepsilon|^2=1$. Notice that β in (5.65) need to satisfy $\frac{1}{\kappa}\mathrm{tr}\left\{N_o\sum_{i=1}^N\boldsymbol{\Phi}_i^\varepsilon\boldsymbol{\Phi}_i^{\varepsilon\dagger}\right\}<P$ so that power constraint (5.34) can hold. That is, $\beta<\frac{P}{N_o}\frac{\kappa}{\kappa-1}$. We select the weights η_i^ε to be:

$$\eta_i^a\triangleq\frac{b_{1,i}^*\cdot c_i^*}{g^a},\quad\eta_i^b\triangleq\frac{b_{2,i}^*\cdot c_i^*}{g^b},\quad\eta_i^c\triangleq\frac{\sum_{u=1}^2 b_{u,i}^*\cdot c_i^*}{g^c},\quad i=1,\cdots,N.\tag{5.66}$$

where g^{ε} is such that $\sum_{i=1}^{N} |\eta_i^{\varepsilon}|^2 = 1$. The beamformings aim at maximizing (via Maximal Ratio Transmission) the individual rates of user 1, user 2 and the sum-rate, respectively. Now, the optimum value β must be selected. For that purpose, we propose the use of a simple one-dimensional exhaustive search considering R_1, R_2 and $R_1 + R_2$ as the objective metrics for η_i^a, η_i^b and η_i^c, respectively.

5.5 Compress-and-forward

Unlike previous techniques, with C&F the relay nodes mimic a receive antenna array with the users, helping them to decode the base station signal. Hence, the MPR-BC performs as a SIMO BC, for which dirty paper coding is the optimal source encoding strategy.

There is a fundamental difference between the implementation of C&F at the MPR-BC and the implementation at the MPRC or MPR-MAC: in the broadcast case more than one destination have to decompress the relay signals. As mentioned in Section 3.5 and Section 4.5, C&F is constructed using distributed Wyner-Ziv (D-WZ) compression codes at the relay nodes [39]. Those codes are decompressed at the destination using its own received signal as side information (See Proposition 3.1). Therefore, when having two different destinations (that is, two different decoders of the compressed signals) the one with worst side information will limit the operational point. In other words, the relays must select compression codes that can be simultaneously decompressed by both users.

Theorem 5.3 *With D-WZ C&F, the MPR-BC achieves the rate region:*

$$\mathcal{R}_{\mathrm{C\&F}}(P) = \mathrm{coh}\left(\bigcup_{\substack{\gamma_1,\gamma_2,\gamma_{r_1},\cdots,\gamma_{r_N}: \\ \sum_{u=1}^{2}\gamma_u + \sum_{i=1}^{N}\gamma_{r_i} \le P}} \bigcup_{\pi} \bigcup_{\substack{\boldsymbol{\rho}_{1:N} \ge 0: \\ \boldsymbol{\rho}_{1:N} \in \mathcal{J}_u\left(\gamma_{1:2},\gamma_{r_{1:N}}\right),\ u=1,2}} \left\{ R_{1,2} : \begin{array}{l} R_{\pi(1)} \le \log_2 \det\left(\boldsymbol{I} + \gamma_{\pi(1)} \boldsymbol{q}_{\pi(1)} \boldsymbol{q}_{\pi(1)}^{\dagger}\right) \\ R_{\pi(2)} \le \log_2 \det\left(\boldsymbol{I} + \left(\boldsymbol{I} + \gamma_{\pi(1)} \boldsymbol{q}_{\pi(2)} \boldsymbol{q}_{\pi(2)}^{\dagger}\right)^{-1} \gamma_{\pi(2)} \boldsymbol{q}_{\pi(2)} \boldsymbol{q}_{\pi(2)}^{\dagger}\right) \end{array} \right\} \right), \quad (5.67)$$

where

$$\boldsymbol{q}_u = \left[\frac{a_u^*}{\sqrt{N_o}}, \frac{c_1^*}{\sqrt{N_o + \rho_1}}, \cdots, \frac{c_N^*}{\sqrt{N_o + \rho_N}} \right]^T, \quad (5.68)$$

π is any permutation of $\{1,2\}$ and $\mathcal{J}_u\left(\boldsymbol{\gamma}_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right)\triangleq$

$$\left\{\boldsymbol{\rho}_{1:N}:\mathcal{C}\left(\frac{(\gamma_1+\gamma_2)\left(\sum_{i\in\mathcal{S}}\frac{|c_i|^2}{N_o+\rho_i}\right)}{(\gamma_1+\gamma_2)\left(\frac{|a_u|^2}{N_o}+\sum_{i\notin\mathcal{S}}\frac{|c_i|^2}{N_o+\rho_i}\right)}\right)+\sum_{i\in\mathcal{S}}\mathcal{C}\left(\frac{N_o}{\rho_i}\right)\leq \mathcal{C}\left(\sum_{i\in\mathcal{S}}\frac{|b_{u,i}|^2\gamma_{r_i}}{N_o+|a_u|^2(\gamma_1+\gamma_2)}\right),\forall\mathcal{S}\subseteq\mathcal{N}.\right\} \quad (5.69)$$

Proof: Let the messages $\omega_u\in\left\{1,\cdots,2^{nR_u}\right\}$, $u=1,2$ be selected for transmission to user 1 and user 2, respectively. It divides them it into B sub-messages of κR bits each, with $\kappa=\frac{n}{B}$: $\omega_u=\left[\omega_u^1,\cdots,\omega_u^B\right]$. The messages are then pipelined using block-Markov encoding within $B+1$ channel blocks.

On each block b, the source transmits the sub-messages ω_u^b, $u=1,2$ to the users and relays. To do it, the source maps them onto a dirty paper code (similar to that explained for D&F). First, it encodes ω_2^b using a random codeword $\boldsymbol{s}_2^\kappa(\cdot)$ generated *i.i.d.* from $s_2\sim\mathcal{CN}(0,\gamma_u)$. Later, it encodes ω_1^b utilizing non-causal knowledge of the interference that signal $\boldsymbol{s}_2^\kappa\left(\omega_2^b\right)$ produces in user 1 [47, Section II]. The procedure follows equivalent arguments to those in (5.13)-(5.18), and outputs a signal $\boldsymbol{s}_1^\kappa$. The source thus transmits

$$\boldsymbol{x}_d^\kappa[b]=\boldsymbol{s}_2^\kappa\left(\omega_2^b\right)+\boldsymbol{s}_1^\kappa[b]. \quad (5.70)$$

The elements of $\boldsymbol{s}_1^\kappa[b]$ are distributed following $s_1\sim\mathcal{CN}(0,\gamma_1)$, as shown in the proof of D&F [47]. Notice that the encoding order at the DPC is arbitrary and can be reversed. Furthermore, the Gaussian codebook, although used, has not been shown to be optimal. The two transmitted codewords are then received at the relays under AWGN:

$$\boldsymbol{y}_{r_i}^\kappa[b]=c_i^*\sum_{u=1}^{2}\boldsymbol{s}_u^\kappa\left(\omega_u^b\right)+\boldsymbol{z}_{r_i}^\kappa,\ \ i=1,\cdots,N. \quad (5.71)$$

Upon receiving the signals, relays distributedly compress them using a multi-source compression code as that in Proposition 3.1. Therefore, signals $\boldsymbol{y}_{r_i}^\kappa[b]$ are mapped at the relays using functions

$$f_\kappa^{r_i}:\mathcal{Y}_{r_i}^\kappa\to\left\{1,\cdots,2^{\kappa\phi_i}\right\}, \quad (5.72)$$

where ϕ_i is the compression rate of relay i. The mapping consists of finding the bin-index of compression codewords that are jointly typical with $\boldsymbol{y}_{r_i}^\kappa[b]$ (see Proposition 3.1 for full details).

On the next block $b+1$, the relays send to the users (via their MAC channels) the indexes $t_{r_i}[b] = f_\kappa^{r_i}\left(\boldsymbol{y}_{r_i}^\kappa[b]\right)$, $i = 1, \cdots, N$. To that end, the indexes are mapped onto multi-access channel codebooks $\boldsymbol{v}_{r_i}^\kappa(\cdot)$, $i = 1, \cdots, N$:

$$\boldsymbol{x}_{r_i}^\kappa[b+1] = \boldsymbol{v}_{r_i}^\kappa\left(t_{r_i}[b]\right), \; i = 1, \cdots, N. \tag{5.73}$$

We select the MAC codebooks to be Gaussian, generated *i.i.d.* from $V_{r_i} \sim \mathcal{CN}(0, \gamma_{r_i})$, $i = 1, \cdots, N$. Notice that the power utilized by the relays in order to cooperate with the source is $\sum_{i=1}^N \gamma_{r_i}$. Therefore, the power constraint (*A3*) is satisfied whenever:

$$\gamma_1 + \gamma_2 + \sum_{i=1}^N \gamma_{r_i} \leq P. \tag{5.74}$$

The received signal at the two users during the block $b+1$ is given by:

$$\boldsymbol{y}_u^\kappa[b+1] = a_u^* \sum_{p=1}^2 \cdot \boldsymbol{s}_p^\kappa\left(\omega_p^{b+1}\right) + \sum_{i=1}^N b_{u,i}^* \boldsymbol{v}_{r_i}^\kappa\left(t_{r_i}[b]\right) + \boldsymbol{z}_d^\kappa, \;\; u = 1, 2. \tag{5.75}$$

Let us now explain the decoding at the users in block $b+1$. First, they recover indexes $t_{r_{1:N}}[b]$ from its received signal $\boldsymbol{y}_u^\kappa[b+1]$, $u = 1, 2$. Both users can do so *iff* the transmission rates ϕ_i lie within their capacity region (being $\boldsymbol{s}_p^\kappa\left(\omega^{b+1}\right)$, $p = 1, 2$ interference):

$$\sum_{i\in\mathcal{S}} \phi_i \leq \mathcal{C}\left(\frac{\sum_{i\in\mathcal{S}} |b_{u,i}|^2 \gamma_{r_i}}{N_o + |a_u|^2\left(\gamma_1 + \gamma_2\right)}\right), \forall \mathcal{S} \subseteq \mathcal{N}, \;\; u = 1, 2. \tag{5.76}$$

Once the indexes $t_{r_{1:N}}[b]$ have been estimated, the two users remove their contribution on $\boldsymbol{y}_u^\kappa[b+1]$:

$$\begin{aligned} \boldsymbol{y}_u^{\prime\kappa}[b+1] &= \boldsymbol{y}_u^\kappa[b+1] - \sum_{i=1}^N b_{u,i} \cdot \boldsymbol{x}_{r_i}^\kappa[b+1] \\ &= a_u \sum_{p=1}^2 \boldsymbol{s}_p^\kappa\left(\omega_p^{b+1}\right) + \boldsymbol{z}_d^\kappa. \end{aligned} \tag{5.77}$$

Afterwards, each user decompresses the indexes $t_{r_{1:N}}[b]$ using its own received signal $\boldsymbol{y}_u^{\prime\kappa}[b]$ as side information. As shown in Proposition 3.1, the decompression step consists of finding compression codewords $\hat{\boldsymbol{y}}_{r_{1:N}}$ that, belonging to the bins selected by the compression encoders, are jointly typical with $\boldsymbol{y}_u^{\prime\kappa}[b]$. Each user can find at least one typical codeword if and only if (see Proposition 3.1)

$$I\left(\boldsymbol{y}_\mathcal{S}; \hat{\boldsymbol{y}}_\mathcal{S} | y_u^\prime, \hat{\boldsymbol{y}}_\mathcal{S}^c\right) \leq \sum_{i\in\mathcal{S}} \phi_i \qquad \forall \mathcal{S} \subseteq \mathcal{N}, \;\; u = 1, 2. \tag{5.78}$$

Therefore, taking into account constraints (5.76) and (5.78), both users can decompress the relays signals if and only if:

$$I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|y_u',\hat{\boldsymbol{y}}_{\mathcal{S}}^c\right)\leq\mathcal{C}\left(\frac{\sum_{i\in\mathcal{S}}|b_{u,i}|^2\gamma_{r_i}}{N_o+|a_u|^2\left(\gamma_1+\gamma_2\right)}\right)\ \ \forall\mathcal{S}\subseteq\mathcal{N},\ \ u=1,2. \tag{5.79}$$

Once decompressed, vectors $\hat{\boldsymbol{y}}_{r_{1:N}}^{\kappa}[b]$ are used by the receivers, along with their own signals $\boldsymbol{y}_u^{'\kappa}[b]$, to decode their intended messages. Let us select the compression codebooks at the relays to be Guassian (even though they have not been shown to be optimal). That is, $p\left(\hat{y}_i|y_i\right)=\frac{1}{\sqrt{\pi\rho_i}}\exp\left(-\frac{|\hat{y}_i-y_i|^2}{\rho_i}\right)$, where we refer to ρ_i as the compression noise of relay i. Therefore, given the dirty paper encoding, the users can correctly estimate their messages *iff* [45, Section III]:

$$R_1 \leq \log_2\det\left(\boldsymbol{I}+\gamma_1\boldsymbol{q}_1\boldsymbol{q}_1^{\dagger}\right) \tag{5.80}$$

$$R_2 \leq \log_2\det\left(\boldsymbol{I}+\left(\boldsymbol{I}+\gamma_1\boldsymbol{q}_2\boldsymbol{q}_2^{\dagger}\right)^{-1}\gamma_2\boldsymbol{q}_2\boldsymbol{q}_2^{\dagger}\right) \tag{5.81}$$

where $\boldsymbol{q}_u$ is defined in (5.68). This is equivalent to the achievable rates of the AWGN SIMO BC. Let us now particularize the constraint (5.79) for the Gaussian compression codewords. In particular, using equivalent arguments to those in (3.56), it is possible to derive that:

$$\begin{aligned}I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|y_u',\hat{\boldsymbol{y}}_{\mathcal{S}}^c\right) &= I\left(x_d;y_u^{'},\hat{\boldsymbol{y}}_{r_{1:N}}\right)-I\left(x_d;y_u^{'},\hat{\boldsymbol{y}}_{\mathcal{S}}^c\right)+I\left(\boldsymbol{y}_{\mathcal{S}};\hat{\boldsymbol{y}}_{\mathcal{S}}|x_d,y_d',\hat{\boldsymbol{y}}_{\mathcal{S}}^c\right)\\ &= \mathcal{C}\left(\frac{\left(\gamma_1+\gamma_2\right)\left(\sum_{i\in\mathcal{S}}\frac{|c_i|^2}{N_o+\rho_i}\right)}{\left(\gamma_1+\gamma_2\right)\left(\frac{|a_u|^2}{N_o}+\sum_{i\notin\mathcal{S}}\frac{|c_i|^2}{N_o+\rho_i}\right)}\right)+\sum_{i\in\mathcal{S}}\mathcal{C}\left(\frac{N_o}{\rho_i}\right)\end{aligned} \tag{5.82}$$

Therefore, plugging (5.82) into (5.79), the rates (5.80) are achievable if: $\boldsymbol{\rho}_{1:N}\in\mathcal{J}_1\left(\gamma_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right)$ and $\boldsymbol{\rho}_{1:N}\in\mathcal{J}_2\left(\gamma_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right)$. However, the relays may arbitrary choose any codebook design belonging to both $\mathcal{J}_u\left(\gamma_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right)$, $u=1,2$ and also *time-share* them. Therefore the achievable rate region remains:

$$\operatorname{coh}\left(\bigcup_{\boldsymbol{\rho}_{1:N}\in\mathcal{J}_u\left(\gamma_{1:2},\boldsymbol{\gamma}_{r_{1:N}}\right),u=1,2}\left\{R_{1,2}:\begin{array}{c}R_1\leq\log_2\det\left(\boldsymbol{I}+\gamma_1\boldsymbol{q}_1\boldsymbol{q}_1^{\dagger}\right)\\ R_2\leq\log_2\det\left(\boldsymbol{I}+\left(\boldsymbol{I}+\gamma_1\boldsymbol{q}_2\boldsymbol{q}_2^{\dagger}\right)^{-1}\gamma_2\boldsymbol{q}_2\boldsymbol{q}_2^{\dagger}\right)\end{array}\right\}\right).$$

Finally, the power can be arbitrarily allocated at the sources and relays. Moreover, the dirty paper encoding order can be arbitrarily set. ■

As for D&F, it was impossible for us to derive the weighted sum-rate of region (5.67), not even to upper bound it. Indeed, the WSR optimization is not convex. Therefore, we are not able to characterize and draw the region.

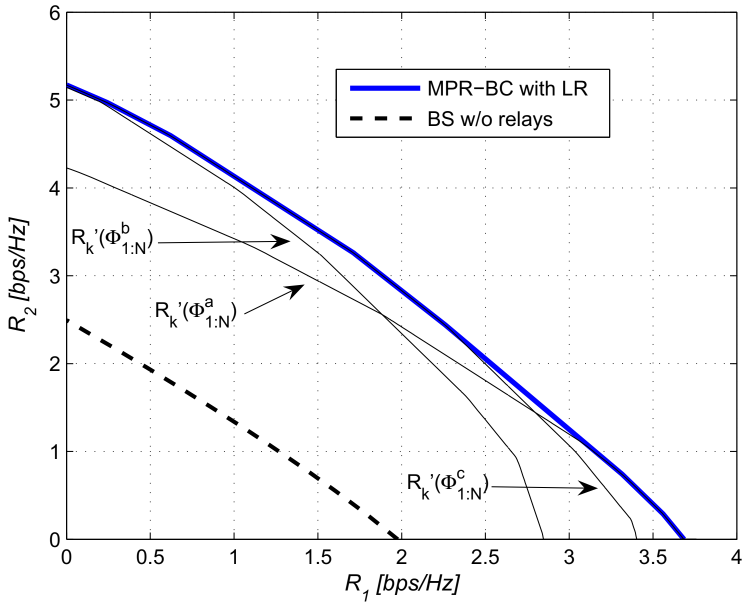

Figure 5.2: Rate region of the MPR-BC with LR. We consider $N = 3$ relays, and overall power constraint $\frac{P}{N_o} = 8$ dB. $\kappa = 60$.

5.6 Numerical Results

As mentioned previously, we have not been capable to evaluate the rate regions with D&F and C&F. We thus evaluate in the section the one with LR. Equivalently to Section 4.6, we model the channel fading as zero-mean, complex, Gaussian-distributed and *time-invariant*. Defining $d_{s,u} = 1$ the source-users distance and d the users-relays distance, we assume $a_u \sim \mathcal{CN}(0, 1)$, $b_{u,i} \sim \mathcal{CN}(0, d^{-\alpha})$, and $c_i \sim \mathcal{CN}\left(0, (1-d)^{-\alpha}\right)$, $u = 1, 2$, $i = 1, \cdots, N$. Finally, the path-loss exponent is set to $\alpha = 3$.

The achievable rate region of the MPR-BC with LR is depicted in Fig. 5.2. We consider $N = 3$ relays, and overall transmit power constraint $P/N_o = 8$ dB. Relays are placed at a distance $d = 0.5$, and the region is plotted for the given realization of the channel: $[a_1,\ a_2] = [0.68e^{-j2.47},\ 0.85e^{-j0.89}]$, $[c_1,\ c_2,\ c_3] = [1.63e^{j0.08},\ 3.35e^{-j1.17},\ 1.54e^{-j3.01}]$ and

$$[b_{u,i}] = \begin{bmatrix} 5.17e^{j2.35} & 0.85e^{j2.63} & 2.17e^{j1.25} \\ 3.72e^{j1.23} & 5.19e^{j1.53} & 5.73e^{j1.80} \end{bmatrix}$$

As explained in Section 5.4.3, the *LR* region is approximated by *time-sharing* of the conditional rate regions with $\mathbf{\Phi}^a_{1:N}, \mathbf{\Phi}^b_{1:N}, \mathbf{\Phi}^c_{1:N}$. Results show that both users increase significantly their transmission rates, and, particularly user 2 doubles it. Besides, the sum-rate of the system is achieved when no rate is allocated to user 1, as in BC without relays. However, this claim may not hold for all the channel realizations.

Notice that for $N = 3$, the LR achievable rates in the MPRC and MPR-MAC were clearly outperformed by those of D&F and C&F. It is expected then that so they do in the MPR-BC. Accordingly, given the two-fold gains with LR (which are expected to be improved with the other two techniques) we conclude that relaying (even with low number of relays) is very powerful approach to increase coverage of broadcast networks.

5.7 Conclusions

This chapter studied the BC assisted by multiple-parallel relays and presented its achievable rate regions with D&F, LR and C&F, respectively. The region with D&F was derived considering block-Markov plus dirty paper encoding at the source. Mixing both encodings, the full-duplex relaying and the wireless broadcasting were simultaneously implemented. Unfortunately, we were not able to draw the region since the weighted sum-rate optimization remained unsolvable for us.

The achievable rate region with LR was next presented, and shown to be equal to that of the MPR-MAC with a sum-power constraint. That is, the MPR-MAC and MPR-BC are dual channels with LR. This results extended to BCs with direct link between source and users the duality presented by Jafar *et. al.* in [43] for two-hop MAC/BC. To prove duality, we made use of duality results for vector channels [45]. Furthermore, in order to characterize the rate region, we resorted to the sum-rate and weighted sum-rate optimizations. To compute them, we devised iterative algorithms, based upon *dual decomposition*.

Finally, the BC with C&F was studied. Its rate region was derived considering dirty paper encoding at the source, and distributed Wyner-Ziv compression at the relays. As for D&F, the region could not be drawn since the weighted sum-rate optimization were not convex, and remained unknown.

Chapter 6

Conclusions

This dissertation has studied the capacity of Gaussian channels with multiple, parallel[1], relays. In particular, our analysis has focused on three well-known channels: the *point-to-point* channel, the *multi-access* channel and the *broadcast* channel. For them, we have presented achievable rates and capacity outer regions. All over the dissertation, we have assumed: *i*) full-duplex operation at the relays, *ii*) transmit and receive channel state information available at all network nodes and *iii*) *time-invariant*, *memory-less* fading. The research results has been presented as follows:

Chapter 3 has studied the multiple-parallel relay channel (MPRC), where a single source communicates to a single destination with the aid of N parallel relay nodes. The capacity of the channel has been upper bounded using the *max-flow-min-cut* Theorem. The obtained bound is shown to scale as $\mathcal{C}\left(\frac{N}{2}\frac{P}{N_o}\right)$ under *time-invariant* (unitary-mean) Rayleigh fading. Besides, the capacity has been lower bounded by means of the achievable rates with: D&F, two-level

[1]As previously defined, two relays are said to be *parallel* if there is no direct link between them, while both have direct link from the source and towards the destination.

PD&F, C&F and LR. The first two consisted of relay nodes totally or partially decoding the source's signal, respectively. Partial decoding was shown to outperform D&F for low number of relays; on the contrary, for large number of them, both were shown to follow the same scaling law: $\mathcal{C}\left(2 \cdot W_0\left(\frac{\sqrt{N}}{2}\right) \cdot \frac{P}{N_o}\right)$. Such a law diverged from the upper bound, which made both techniques clearly suboptimal; we explained this fact resorting to the source-relay broadcast limitation. Finally, D&F and PD&F also performed equally for low source-relay distances, being both capacity-achieving.

C&F consisted of relay nodes transmitting towards destination a compressed version of their received signals. In turn, the destination utilizes the compressed signals in order to estimate the source's message via coherent detection. Distributed Wyner-Ziv (D-WZ) compression was assumed at the relays. The achievable rate of this technique was presented in terms of a non-convex optimization, which turned out unsolvable for us. Hence, we proposed a computable upper bound as a benchmark for it. Moreover, we showed that the achievable rate with C&F scales as $\mathcal{C}\left(\frac{P}{N_o}\log_2 N\right)$, which is due to the relay-destination MAC limitation.

Finally, we studied LR, which consisted of relay nodes transmitting, on every channel use, a linear combination of previously received signals. The optimum source temporal covariance for this technique was derived, and suboptimum linear relaying matrices proposed. Furthermore, we showed that the achievable rate scales as $\mathcal{C}\left(\frac{N \cdot P}{N_o}\right)$, in the same manner as the upper bound and unlike all previous techniques. Hence, LR was shown to be the only known technique that seizes all the beamforming gain of the system.

Numerical results for Rayleigh-faded networks showed that: *i*) D&F is capacity achieving for short/mid source-relay distances and low number of relays, *ii*) PD&F only outperforms D&F for low number of relays and large source-relay distances, *iii*) C&F provides small rate gains, unless the relays are extremely close to the source, and *iv*) LR is clearly the best technique for large number of relays, given that its beamforming capability allows to overcome noise amplification. On the contrary, for low number of relays, the degree of noise amplification is not compensated by the beamforing gain, and thus, it performs poorly.

Chapter 4 extended previous results to the multi-access channels with multiple-parallel relays (MPR-MAC). On it, multiple sources communicate simultaneously to a single destination, in the presence of N parallel relay nodes. For the channel, we first provided a capacity outer region based upon the *max-flow-min-cut* Theorem. Likewise, we presented achievable rate regions with D&F, C&F and LR. The first was derived as the extension to N relays of the single-relay result [76]. The second was obtained assuming D-WZ compression at the relays. Finally, the latest was derived using theory of vector MAC channels, and further characterized by means of the weighted sum-rate maximization. Such a maximization allowed us to derive the optimum sources' temporal covariance for every boundary point of the rate region. In order to design the linear functions at the relays, we proposed a suboptimum approach.

In parallel, we also carried out the asymptotic analysis of the channel, considering the number of users U growing to infinity. The analysis aimed at studying the impact of multi-user diversity onto the sum-capacity of the MPR-MAC. First, we showed that D&F does not provide sum-rate gains when $U \to \infty$. This was explained resorting to the distributed channel hardening effect at the input of the relays. In contrast, we demonstrated that the *max-flow-min-cut* upper bound indeed suggests gains even at the asymptote. Therefore, D&F is clearly suboptimal with large number of users. Finally, we presented the scaling behavior of C&F. This was shown not to be affected by the channel hardening effect and to provide rate gains.

Numerical results for Rayleigh fading showed that: *i*) the sum-capacity upper bound is maximized for relays close to the destination. This result invites us to think that placing relays close to the base station is more interesting in terms of achievable sum-rates. *ii*) With multiple users, C&F presents an improved performance compared to that of the single-user case, and *iii*) D&F is sum-capacity achieving for low/mid source-relay distances.

Finally, **Chapter 5** ended the analysis by considering the broadcast channel with multiple-parallel relays. For the channel, we presented achievable rates with D&F, C&F and LR, all three based upon dirty paper encoding at the source.

The most significant result of this chapter is that MAC-BC duality holds for linear relaying. Such a claim was demonstrated using duality arguments for MIMO channels, and with it, we extended the duality presented by Jafar *et al.* in [43] to MPR-BCs with direct link between source and destinations.

6.1 Future Work

The interference channel with multiple-parallel relays (MPR-IC) is the great absent of this dissertation [89, 90]. With it, our work would have covered the four main channels found in wireless networks. It is, thus, at the top of the *future-work* stack.

The MPR-IC is a channel in which M independent sources communicate, simultaneously and independently, to M different destinations with the aid of a set $\mathcal{N} = \{1, \cdots, N\}$ of parallel relay nodes. The capacity region of the interference channel is a long standing open problem. So is, then, the capacity when assisted by multiple relays. Accordingly, a huge field of research appears at a first glance: from capacity outer regions and sum-capacity upper bounds, to achievable rate regions (with *e.g. decode-and-forward*, *compress-and-forward*, *linear relaying*, etc.) and asymptotic achievable sum-rates. This task has not been undertaken in this dissertation due to two main reasons:

- *Partial knowledge of the Interference Channel*. The IC is one of the most unknown channels within the field of Information Theory. Even in the absence of relays, its capacity region (as well as the optimum encoding at the sources) is still an open problem. Thus, unlike the point-to-point, *multi-access* and *broadcast* channels, we did not find a clear framework to benchmark the gains obtained via relaying with.

  Besides, the tightest inner bound on the capacity region of the IC without relays, the Han-Kobayashi achievable rate region [90], requires a computationally prohibitive source encoding. This makes it unfeasible in practice, and hard to be extended to the IC with multiple relays.

- *Time*. As mentioned, the extension of the MPRC results to the MPR-IC is not straightforward and requires a separate analysis. On it, we need not only to address different relaying protocols, but also different sources' encoding functions. In addition, the *state-of-the-art* on the topic is emaciated. Hence, the analysis of the MPR-IC would take, at least, another year of research. We decided, then, to leave it for post-doc research.

The research line that next leads the *yet-to-do* list is: *Extension to MIMO channels of the single-antenna results presented in this dissertation*. MIMO relay channels are introduced, among other references, in [91, 92] and consider all network nodes equipped with multiple antennas. Throughout the thesis, we have not assumed multi-antenna nodes mainly due to:

- *Effectiveness*. Clearly, the impact of relays is most significant in networks with single-antenna sources and single-antenna destinations, where neither spatial diversity nor beamforming capabilities are available in the absence of relays. We thus concentrated our efforts on the setup for which relaying is more profitable.

- *Simplicity*. Placing multiple antennas at the network nodes introduces new degrees of freedom: the spatial covariances at the source and relays [91]. Those has to be optimized in order to obtain highest achievable rates. Regarding channels with relays, such an optimization is not convex in most of the cases. Moreover, for the cases where it is, the computation is generally exhaustive and no closed-form expression can be given. Aiming, thus, at eliminating the collateral damage of not been able to provide closed-form expressions for the optimum covariances, neither for the achievable rates nor for the asymptotic performance, we decided to consider single-antenna nodes only.

Finally, regarding the results presented in this dissertation, the most important loose ends have been:

1. *Resource allocation within the compress-and-forward setup*. The achievable rate of the MPRC with C&F is presented in Theorem 3.4 in terms of a non-convex optimization. We have not been able to provide its solution *i.e.,* the optimum power allocation among relays and the optimum compression noises at them. Instead, we devised an iterative resource allocation algorithm based upon a rate upper bound, yielding a suboptimum solution. The same happened for the weighted sum-rate of the MPR-MAC with C&F (see Section 4.5.1). New efforts shall be, thus, invested in order to solve the optimizations and to devise the correct resource allocation algorithm.

2. *Design of spectral efficient linear relaying matrices*. This dissertation was not able to derive, for linear relaying, the optimum linear functions at the relays, neither for MPRC (Section 3.6), nor for the MPR-MAC or MPR-BC (Section 4.4 and Section 5.4, respectively). This was due to the non-convexity of the matrices optimization. Instead, we proposed suboptimum approaches, based upon: *i*) *amplify-and-forward* extended to more

than two channel uses, and *ii*) maximal ratio transmission among the relays. New approaches shall be thus studied in order to improve performance, specifically at the low number of relays regime.

3. *Characterization of the MPR-BC achievable rate regions.* We could not characterize of the achievable rate regions of the channel with D&F and C&F. For both cases, the weighted sum-rate was a non-convex optimization with unmanageable number of optimization variables. Thus, it turned unsolvable for us. Research need to be carried out to characterize the achievable weighted sum-rate with both techniques.

Bibliography

[1] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 623–656, Jul., Oct. 1948.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, 1^{st} Edition, Cambridge University Press, 2005.

[3] I. Telatar, "Capacity of multi-antenna Gaussian channel," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[4] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[5] V. Tarokh, N. Seshadi, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[6] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-time block codes for orthogonal designs," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[7] E.G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, 1^{st} Edition, Cambridge University Press, 2003.

[8] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall Signal Processing Series, 1993.

[9] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[10] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Information Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[11] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[12] E. Zimermann, P. Herhold, and G. Fettweis, "On the performance of cooperative relaying protocols in wireless networks," *European Trans. on Telecommunications*, vol. 16, no. 1, pp. 5–16, Jan. 2005.

[13] M. Dohler, *Virtual Antenna Arrays*, PhD Dissertation, King's College London, UK, 2003.

[14] J.N. Laneman, *Cooperative Diversity in Wireless Networks: Algorithms and Architectures*, PhD Dissertation, Massachusetts Institute of Technology, 2002.

[15] D. Chen and J.N. Laneman, "The diversity-multiplexing tradeoff for the multiaccess relay channel," in *Proc. Conference in Information Sciences and Systems*, Princeton, NJ, USA, Mar. 2006.

[16] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity – part I: system description," *IEEE Trans. on Communications*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[17] A. Bletsas, A. Khisti, D.P. Reed, and A.Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[18] B.M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. on Communications*, vol. 48, no. 12, pp. 2042–2052, Dec. 2000.

[19] E.C. Van der Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

[20] E.C. Van der Meulen, *Transmission of information in a* T*-terminal discrete memoryless channel*, PhD Dissertation, University of California, Berkeley, 1968.

[21] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.

[22] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[23] R. Nabar, H. Bölcskei, and F.W. Kneuböhler, "Fading relay channels: performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[24] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[25] L.L. Xie and P.R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[26] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. International Symposium on Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[27] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multi-hopping for relay networks," in *Proc. International Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, Feb. 2004, pp. 192–195.

[28] L.L. Xie and P.R. Kumar, "A network infromation theory for wireless communication: scaling laws and operation," *IEEE Trans. on Information Theory*, vol. 50, pp. 748–767, May 2004.

[29] I. Maric and R. Yates, "Forwarding strategies for Gaussian parallel-relay networks," in *Proc. International Symposium On Information Theory*, Chicago, IL, USA, Jun. 2004.

[30] I. Maric and R. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," in *Proc. 38th Annual Asilomar Conference on Signals, Systems, and Computers*, Asilomar, CA, USA, Nov. 2004, pp. 1912–1916.

[31] A. del Coso and C. Ibars, "Distributed antenna channels with regenerative relaying: relay selection and asymptotic capacity," *EURASIP Journal on Wireless Communications and Networking*, vol. Nov. 2007.

[32] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. on Information Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

[33] P. Gupta and P.R. Kumar, "Towards an Information Theory of large networks: an achievable rate region," *IEEE Trans. on Information Theory*, vol. 49, no. 8, pp. 1877–1893, Aug. 2003.

[34] A.F. Dana, M. Sharif, R. Gowaikar, B. Hassibi, and M. Effros, "Is broadcast channel plus multi-access optimal for Gaussian wireless networks?," in *Proc. 37th Asilomar Conf. on Signals, Systems, and Computers*, Asilomar, CA, USA, Nov. 2003.

[35] B. Schein, *Distributed coordination in network information theory*, PhD Dissertation, Massachussets Institute of Technology, 2001.

[36] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg, and G. Kramer, "Communication via decentralized processing," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Jun. 2005.

[37] A. Sanderovich, O. Somekh, and S. Shamai (Shitz), "Uplink macro diversity with limited backhaul capacity," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007.

[38] A. Sanderovich, S. Shamai (Shitz), and Y. Steinberg, "Distributed MIMO receiver - achievable rates and upper bounds," *submitted to IEEE Trans. on Information Theory*, [arXiv:0710.0116v1].

[39] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. on Information Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.

[40] L. Sankaranarayanan, G. Kramer, and N.B. Mandayam, "Capacity theorems for the multiple-access relay channel,," in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, USA, Sep. 2004.

[41] D. Chen, K. Azarian, and J.N. Laneman, "A case for amplify-forward relaying in the block-fading multiaccess channel," *submitted IEEE Trans. on Information Theory*, 2007.

[42] Y. Liang and V. Veeravalli, "Cooperative relay broadcast channels," in *Proc. International Conference on Wireless Networks, Communications and Mobile Computing*, Jun. 2005, pp. 1449–1454.

[43] S.A. Jafar, K.S. Gomadam, and C. Huang, "Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays," *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3350–3370, 2007.

[44] L.L. Xie and P.R. Kumar, "Multisource, multidestination, multirelay wireless networks," *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3586–3595, Oct. 2007.

[45] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[46] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. on Information Theory*, vol. 35, no. 3, pp. 605–619, May 1989.

[47] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, May. 1983.

[48] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[49] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Belmont, MA, 1997.

[50] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi, "Iterative water-filling for Gaussian multiple-access channels," *IEEE Trans. on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[51] A. del Coso and S. Simoens, "Distributed compression for the uplink of a backhaul-constrained coordinated cellular network," *submitted to IEEE Trans. on Wireless Communications.*, 2008. arXiv:0802.0776.

[52] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1^{st} Edition, Cambridge University Press, 2004.

[53] W. Yu and R. Lui, “Dual methods for nonconvex spectrum optimization of multicarrier systems,” *IEEE Trans. on Communications*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[54] W. Gardner, *Introduction to Random Processes, with Applications to Signals and Systems*, 2nd Ed., McGraw-Hill, 1990.

[55] Y. Eldar and A.M. Chan, “On the asymptotic performance of the decorrelator,” *IEEE Trans. on Information Theory*, vol. 49, no. 9, pp. 2309–2313, Dic. 2003.

[56] A.F.. Dana and B. Hassibi, “On the power efficiency of sensory and ad-hoc wireless networks,” *IEEE Trans. on Information Theory*, vol. 52, no. 7, pp. 2890–2914, Jul. 2006.

[57] S. Zahedi, *On reliable communication over relay channels*, PhD Dissertation, Stanford University, 2005.

[58] A. Host-Madsen, “Capacity bounds for cooperative diversity,” *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.

[59] R.M. Corless, G.H. Gonnet, D.E.G. Hare, and D.J. Jeffrey, “On the Lambert W function,” *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[60] H. David, *Order Statistics*, 3^{rd} Edition, John Wiley & Sons, 1981.

[61] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. on Information Theory*, vol. 48, no. 6, Jun. 2002.

[62] Y. Oohama, “Gaussian multiterminal source coding,” *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1912–1923, 1997.

[63] M. Gastpar, M. Vetterli, and P.L. Dragotti, “Sensing reality and communicating bits: a dangerous liaison,” *IEEE Signal Processing Magazine*, pp. 70–83, Jul. 2006.

[64] S.Y. Tung, *Multiterminal source coding*, PhD Dissertation, Cornell University, 1978.

[65] D. Slepian and J.K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–481, Jul. 1973.

[66] J. Chen and T. Berger, “Successive Wyner-Ziv coding scheme and its implications to the quadratic Gaussian CEO problem,” *submitted IEEE Trans. on Information Theory*, 2006.

[67] R.G. Gallager, *Information theory and reliable communication*, John Wiley and Sons, 1968.

[68] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple-access channel sum capacity problem," in *Proc. Conference on Information Sciences and Systems*, The Johns Hopkins University, Mar. 2003.

[69] D.P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[70] R. Knopp and P.A. Humblet, "Information capacity and power control in single-cell multi-user communications," in *Proc. IEEE International Conference on Communications*, Jun. 1995.

[71] W. Hirt, *Capacity and information rates of discrete-time channels with memory*, PhD Dissertation, Swiss Federal Institute of Technology, Zurich, 1988.

[72] R.M. Gray, "On the aymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. on Information Theory*, vol. 18, no. 6, pp. 725–730, Nov. 1972.

[73] G. Kramer, "Models and theory for relay channel with receive constraints," in *Proc. 42nd Allerton Conference on communications, Control and Computing*, Allerton, IL, Sept. 2004.

[74] P. Razaghi and Wei Yu, "Parity forwarding for multiple-relay networks," *submitted to IEEE Trans. on Information Theory*, Nov. 2007.

[75] C. Bordenave, "Spatial capacity of multiple-access wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 4977–4988, Nov. 2006.

[76] G. Kramer and A.J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. International Symposium on Information Theory*, Sorrento, Italy, Jun. 2000, p. 40.

[77] J. Evans and D. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. on Information Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.

[78] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff in cooperative wireless systems," in *Proc. Conference in Information Sciences and Systems*, Princeton, NJ, USA, Mar. 2006.

[79] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.

[80] B.M. Hochwald, T.L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[81] E.G. Larsson, "On the combination of spatial diversity and multiuser diversity," *IEEE Communication Letters*, vol. 8, no. 8, pp. 517–519, Aug. 2004.

[82] R.G. Cheng and S. Verdú, "Gaussian multiple-access channels with ISI: Capacity region and multi-user water-filling," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[83] K.B. Petersen and M.S. Pedersen, *The Matrix Cookbook*, 2007.

[84] S. Ye and R.S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, pp. 2839–2847, Nov. 2003.

[85] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sep 2006.

[86] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[87] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Dec. 2006.

[88] J. Malick and H.S. Sendov, "Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices," *Springer Set-Valued Analysis*, vol. 14, no. 3, pp. 273–293, Sep. 2006.

[89] A.B. Carleial, "Interference channels," *IEEE Transactions on Information Theory*, vol. 24, pp. 60–70, Jan. 1978.

[90] T.S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, pp. 49–60, Jan. 1981.

[91] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. on Information Theory*, vol. 51, no. 1, pp. 2020–2040, Jan. 2005.

[92] S. Simoens, O. Munoz, J. Vidal, and A. del Coso, "Capacity bounds for gaussian MIMO relay channel with channel state information," in *Proc. IEEE* 9^{th} *Workshop on Signal Processing Advances in Wireless Communications*, Recife, Brazil, Jul. 2008.