Alma mater studiorum a.d. 1088 - Università di Bologna

Dottorato di Ricerca in
Ingegneria Elettronica, delle Telecomunicazioni e
Tecnologie dell'Informazione (ET-IT)

Ciclo XXXI

Settore Concursuale: 09/F2 - Telecomunicazioni
Settore Scientifico Disciplinare: ING-INF/03

Caching in Heterogeneous Networks

Autore
Estefanía
Recayte

Supervisori
Alessandro Vanelli Coralli
Giuseppe Cocco

Coordinatore Dottorato
Alessandra Costanzo

Esame finale anno 2019

A mami Carmencita y a papi Edu...

Acknowledgements

This work is the outcome of three years of PhD made of research, investigations and simulations. I would like to thank the people which actively contributed in this journey. I am very grateful for the supervision and all the support received from Giuseppe Cocco, Gianluigi Liva and Francisco Lázaro. This thesis was possible thanks to their useful explanations, comments and suggestions. Additionally, I would like to acknowledge Alessandro Vanelli and Sandro Scalise for their help and for doing this PhD possible. My thanks go also to the nice colleagues from DLR: Balaz, Benni, Fede, Fran, Gianluigi, Hannes, Manuel, Marcel, Mustafa, Thomas, Stefan and Svilen. It was always a pleasure to share working days with them.

I want to thank all my friends from München, Italia and Mar del Plata, simply for always being present.

Finally, the most special thanks are for unconditional support and love received from my family: mami Carmencita, papi Edu, Seba and Marina, Ale, Leo and Vale. Thanks for teaching me every day the real values and the simplicity of life.

Abstract

A promising solution in order to cope with the massive request of wireless data traffic consists of having replicas of the potential requested content memorized across the network. In cache-enabled heterogeneous networks, content is pre-fetched close to the users during network off-peak periods in order to directly serve the users when the network is congested. In fact, the main idea behind caching is the replacement of backhaul capacity with storage capabilities, for example, at the edge of the network. Caching content at the edge of heterogeneous networks not only leads to significantly reduce the traffic congestion in the backhaul link but also leads to achieve higher levels of energy efficiency. However, the good performance of a system foresees a deep analysis of the possible caching techniques. Due to the physical limitation of the caches' size and the excessive amount of content, the design of caching policies which define how the content has to be cached and select the likely data to store is crucial.

Within this thesis, caching techniques for storing and delivering the content in heterogeneous networks are investigated from two different aspects. The first part of the thesis is focused on the reduction of the power consumption when the cached content is delivered over an Gaussian interference channel and per-file rate constraints are imposed. Cooperative approaches between the transmitters in order to mitigate the interference experienced by the users are analyzed. Based on such approaches, the caching optimization problem for obtaining the best cache allocation solution (in the sense of minimizing the average power consumption) is proposed. The second part of the thesis is focused on caching techniques at packet level with the aim of reducing the transmissions from the core of an heterogeneous network. The design of caching schemes based on rate-less codes for storing and delivering the cached content are proposed. For each design, the placement optimization problem which minimizes the transmission over the backhaul link is formulated.

Table of contents

List of figures

List of tables

Acronyms

5G fifth generation

AWGN additive white Gaussian noise

BC broadcast channel

BEC binary erasure channel

CSI channel state information

DPC dirty paper coding

GEO geostationary Earth orbit

GIC Gaussian interference channel

GIC-C Gaussian interference channel with common information

GIC-WC Gaussian interference channel without common information

GIN Gaussian interference as noise

HAPS high altitude platform stations

HK Han-Kobayashi

i.i.d. independent and identically distributed

ISD ideal soliton distribution

KKT Karush-Kuhn-Tucker

LEO low Earth orbit

LRFC linear random fountain codes

LT Luby Transform

MAC multiple access channel

MBS macro base station

MC multicast channel

MDS maximum distance separable

MIMO-BC multiple input multiple output Gaussian broadcast channel

MISO multiple input single output

MISO-OC multiple input single output orthogonal channel

ML maximum likelihood

mmWaves millimeter-wave

OC orthogonal channel

q-EC q-ary erasure channel

QoE quality of experience

QoS quality of service

RSD robust soliton distribution

SBS small base station

SVD single value decomposition

UAV unmanned aerial vehicle

Chapter 1

Introduction

1.1 Introduction to Caching in Heterogeneous Networks

The increasing demand for massive multimedia services give rise to new challenges to face in wireless communication networks. It is foreseen that traffic demand in the fifth generation (5G) mobile communication networks will increase by 1000-fold by the year 2020 [1]. This huge increase in demand will correspond in part to high-definition video contents and streaming applications, which not only require a significant amount of bandwidth but also have stringent quality of service (QoS) requirements.

In order to counteract such exponential growth in content traffic, the most appropriate countermeasure consists of *bringing the content close to the users* [2], by deploying a tier of transmitters near to the end users. The two-tier heterogeneous networks have attracted significant attention in literature [3–6] and they are a promising technology solution for meeting 5G requirements [7]. For instance, in terrestrial networks an architecture where small base stations (SBSs) with short-range and low-power are combined with a macro base station (MBS) achieves higher levels of energy efficiency [8–12] and increases the system capacity significantly [13, 14].

The benefits of heterogeneous networks can be further exploited when transmitters are equipped with storage capacity. Caching contents at the edge of heterogeneous networks (e.g. SBSs [15, 16], unmanned aerial vehicles (UAVs) [17, 18], low Earth orbit (LEO) satellites [19], hubs, etc.) can help to use more efficiently the available resources avoiding the network congestion. Duplicated transmissions of the same content are avoided and in consequence resources are more efficiently used. Specially,

the network load can be reduced at time of high data traffic by directly serving the user from the edge. Furthermore, the communication delay can be reduced significantly when the processing and storage capabilities are in proximity of the end user. Applications such as video content delivery can extremely benefit from such edge caching and improve users' quality of experience (QoE).

The caching process consists on two phases, namely the *placement phase* and the *delivery phase* [20]. During the placement phase the content is pre-fetched close to the users. This phase occurs during off-peak traffic periods. Due to the limited size of the physical memory, a placement policy for selecting which content should be stored has to be decided. During this phase, the network has not a deterministic knowledge of the users future demands. Instead, the delivery phase usually occurs during peak traffic periods i.e, when the network is congested. At this stage, the network has to fulfill the requests arriving from the users. The main target is to serve the users as much as possible with the cached content in order to reduce the transmission from the core of the network. Reducing such transmissions leads to reduce both the network congestion and power consumption.

In this work, we investigate the two phases of the caching process when contents are memorized at the edge of the heterogeneous networks. In particular, we focus on designing caching solutions from two different perspectives. This dissertation is based on our work published in [21–24].

In the first part of the thesis, the study is focused on caching techniques to minimize the average power consumption during the delivery phase when content are delivered over an interference channel. As a matter of fact, users in heterogeneous networks might experience interference from the surrounding transmitters. Different interference channel models from the literature are initially analyzed. We propose, for each of the transmission models, a solution for optimizing the power consumption when the content have transmission rate constraints (in terms of bits/s/Hz). We further formulate the optimization problem to select the cache allocation which minimize the average power consumption in the delivery phase. While interference in cache-aided networks have been studyed in different works [25–29], our main contribution is based on considering that files are delivered over a Gaussian interference channel and different rate requirements per file have to be respected. A two-user network is considered where transmitters might cooperate for serving the users and three levels of cooperation are assumed: (i) non cooperation, (ii) limited cooperation and (iii) full cooperation.

The second part of the thesis is focused on the study of caching schemes base on encoded content. In literature, caching content which has been encoded using optimal

codes were proposed in [30–34]. However, the practical use of optimal codes is limited due to the high encoding and decoding complexity. Thus, in this work we propose to cache and to deliver content which has been encoded using rate-less codes. We present the design of the caching scheme based in rate-less codes and we evaluate its performance in terms of backhaul transmission rate. Our results are analyzed and compered with respect to the optimal caching scheme. A satellite heterogeneous network is considered and two different caching schemes based on rate-less codes are considered, i.e. linear random fountain codes based caching scheme and Luby Transform based caching scheme. For each scheme, the optimization of the placement phase is proposed with the aim of reducing transmissions from the core of the network (from the satellite).

1.2 Thesis Outline

This thesis is organized as follows

- Chapter 2: Introduces the basics from information theory, providing definitions and concepts for understanding transmissions over the interference channel.
- Chapter 3: Describes the different transmission channels considered for delivering cached content. In particular, for each transmission the channel model and the achievable transmission rate is given.
- Chapter 4: Presents a deep analysis of caching in Gaussian interference channels in heterogeneous networks. The optimization problem to select the best caching allocation which minimizes the average transmission power is formulated. The complexity and algorithm for solving the optimization problem are provided. Finally, simulation results for each approach considered are analyzed.
- Chapter 5: Introduces the basics from coding theory, providing definitions and concepts to understand rate-less codes.
- Chapter 6: Proposes the use of rate-less caching scheme in heterogeneous networks. Performance is evaluated and compared with respect to the caching scheme based on optimal codes. To this end, an optimization problem which aims at reducing transmissions from the core of the network is formulated for each caching scheme considered.
- Chapter 7: Addresses the final conclusions of the thesis.

Chapter 2

Information Theory Background

Within this chapter, we introduce background concepts and basics from information theory which are used through this thesis. We start by recalling the concept of entropy and mutual information and then we move to the concept of achievable rate and channel capacity [35].

2.1 Definitions

Definition 1. (Entropy) Let X be a discrete random variable which takes values from a finite alphabet $\mathcal{X}$ and has probability mass function $p_X(x) = \Pr[X = x], x \in \mathcal{X}$. The *entropy* of X, denoted by $H(X)$ is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x). \tag{2.1}$$

and it is measured in *bits*.

Definition 2. (Joint entropy) Let X and Y be two discrete random variables with a joint distribution $p_{XY}(x, y)$. The *joint entropy* $H(X, Y)$ of (X, Y) is defined as

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log_2 p_{XY}(x, y). \tag{2.2}$$

Definition 3. (Conditional entropy) The *conditional entropy* $H(Y|X)$ of a pair of discrete random variables (X, Y) with a joint distribution $p_{XY}(x, y)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} p_X(x) H(Y|X=x) \\ &= -\sum_{x\in\mathcal{X}} p_X(x) \sum_{y\in\mathcal{Y}} p_{Y|X}(y|x) \log_2 p_{Y|X}(y|x) \\ &= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x,y) \log_2 p_{Y|X}(y|x). \end{aligned} \tag{2.3}$$

Definition 4. (Differential entropy) The *differential entropy* $h(X)$ of a continuous random variable X with probability density function $f(x)$ and support S is defined as

$$h(X) = -\int_S f(x) \log_2 f(x) dx. \tag{2.4}$$

Definition 5. (Mutual information) The *mutual information* $I(X;Y)$ between two continuous random variables X and Y with marginal distributions $f_X(x)$ and $f_Y(y)$, respectively, and with joint density function $f_{XY}(x, y)$ is defined as

$$I(X;Y) = \int\int f_{XY}(x,y) \log_2 \frac{f_{XY}(x,y)}{f_X(x) f_Y(y)} dxdy \tag{2.5}$$

and is measured in bits per channel use.

2.2 The Gaussian Channel

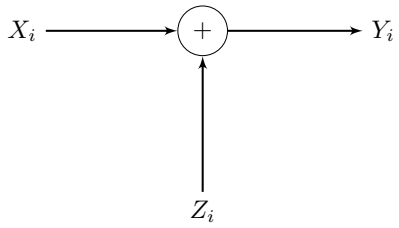


Fig. 2.1 Gaussian channel.

A representation of the time discrete Gaussian channel is depicted in Figure 2.1. At time i, the output Y_i of the channel is the sum of the input X_i and the noise Z_i. The symbols X_i, $i = 1, ..., n$, are independent and identically distributed (i.i.d.) random variables. The noise terms Z_i, $i = 1, ..., n$, i.i.d. with Z_i distributed according to a

zero-mean Gaussian distribution with variance N. Hence

$$Y_i = X_i + Z_i, \qquad Z_i \sim \mathcal{N}(0, N). \tag{2.6}$$

A message over the Gaussian channel is transmitted in the form of a codeword $(x_1, x_2, ..., x_n)$. If we assume an average power constraint P on the input signal then any codeword transmitted has to satisfy

$$\mathbb{E}[x^2] \leq P. \tag{2.7}$$

2.2.1 Capacity of the Gaussian channel

Definition 6. (Information capacity) The *information capacity*, or simply *capacity*, of the Gaussian channel is defined as the maximum of the mutual information between the input and the output taken over all distributions of the input message that satisfy the power constraint. That is

$$C = \max_{\mathbb{E}[x^2] \leq P} I(X, Y), \tag{2.8}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. The capacity of the Gaussian channel with average power constraint constraint P and noise variance N can be derived as follows

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X + Z|X) \\ &= h(Y) - h(Z) \end{aligned} \tag{2.9}$$

where the last equality is because Z is independent from X.

Since $Z \sim \mathcal{N}(0, N)$, the differential entropy $h(Z)$ can be calculated as follows

$$\begin{aligned} h(Z) &= -\int f_Z(z) \log_2 f_Z(z) dz \\ &= -\int \frac{1}{\sqrt{2\pi N}} \exp\left\{-\frac{z^2}{2N}\right\} \left(\log_2 \frac{1}{\sqrt{2\pi N}} - \frac{z^2}{2N} \log_2 e \right) dz \\ &= \frac{1}{2} \log_2(2\pi N) + \frac{\log_2 e}{2N} \mathbb{E}[Z] \\ &= \frac{1}{2} \log_2(2\pi e N). \end{aligned} \tag{2.10}$$

The variance random variable $Y = X + Z$ can be calculated as follows

$$\begin{aligned}\mathbb{E}[Y^2] &= \mathbb{E}[(X+Z)^2] \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Z] + \mathbb{E}[Z^2] \\ &= P + N.\end{aligned} \tag{2.11}$$

where the second equality follows from the independence of the variables while the last equality is due from the fact that the mean of Z is zero.

The maximum differential entropy $h(Y)$ is obtained when Y follows a Gaussian distribution, thus

$$h(Y) = \frac{1}{2}\log_2 2\pi e(P+N). \tag{2.12}$$

Hence, if we insert (2.12) and (2.10) into (2.9) we obtain that the mutual information of the Gaussian channel is bounded by

$$I(X;Y) \leq \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \tag{2.13}$$

and the capacity of the Gaussian channel is

$$C = \max_{\mathbb{E}\{x^2\}\leq P} I(X;Y) = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right), \tag{2.14}$$

which is obtained when $X \sim \mathcal{N}(0, P)$.

2.2.2 Achievable Rate and Capacity Region

A rate R is said to be *achievable* in a point to point communication for a Gaussian channel with power constraint P, if there exists a sequence of $(2^{nR}, n)$ codes such that the maximal probability of error $P_e^{(n)}$ can be made arbitrary small when n is sufficiently large, i.e. ,

$$P_e^{(n)} \to 0 \text{ as } n \to \infty.$$

Hence, the capacity C is the maximum rate that can be reliably achieved. Achieving capacity requires coding over arbitrarily large blocks of data. In other words, there exist channel codes that guarantees a reliable point to point transmission when R < C.

The results obtained for a point to point communication can be extended to the multiuser case. In a multiple access channel (MAC) channel with N_t transmitters, the rate vector $(R_1, R_2, ..., R_{N_t})$ is said to be *achievable* if there exists a sequence of $(2^{nR_1}, n), (2^{nR_2}, n), ..., (2^{nR_{N_t}}, n)$ codes, indexed by n, such that

$$P_e^{(n)} \to 0 \text{ as } n \to \infty.$$

An $N-$dimensional region $\mathcal{R}$ is called an *achievable rate region* for a memoryless discrete channel if every point in that region is achievable.

2.2.3 Capacity Region

The *capacity region* $\mathcal{C}$ of a discrete memoryless channel is a closure of the set of achievable rate vectors.
The *sum-capacity* C_{sum} of a discrete memoryless channel is defined as

$$C_{sum} = \max \Big\{ \sum_{k=1}^{N_t} R_k | (R_1, R_2, ..., R_{N_t}) \in \mathcal{C} \Big\}. \tag{2.15}$$

2.2.4 Interference Channel

The two-user *interference channel* consists of two senders and two receivers. In this type of channels, each transmitter desires to communicate with its correspondent receiver. However, since the channel is shared among the users, each receiver perceives interference from the other transmission.

The discrete memoryless interference channel is defined by a quintuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{P}, \mathcal{Y}_1, \mathcal{Y}_2)$, where $\mathcal{X}_1, \mathcal{X}_2$ are the channel input alphabets sets, $\mathcal{Y}_1, \mathcal{Y}_2$ are the channel output alphabets sets and $\mathcal{P}$ denotes the collection of channel conditional probabilities $p_{Y|X}(y_1, y_2|x_1, x_2)$ on $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$.
The channel assumed to be memoryless so that for n channel use, we have

$$p_{Y|X}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^{n} p_{Y|X}(y_{1i}, y_{2i}|x_{1i}, x_{2i}) \tag{2.16}$$

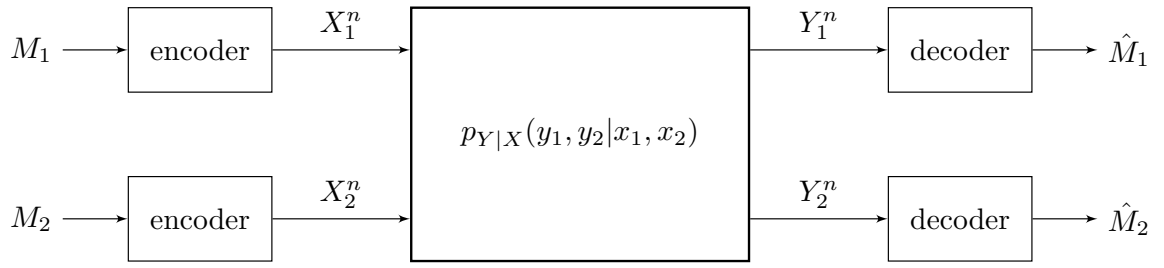


Fig. 2.2 Interference channel.

where $\mathbf{x}_1 = (x_{11}, ..., x_{1n}) \in \mathcal{X}_1^n$, $\mathbf{x}_2 = (x_{21}, ..., x_{2n}) \in \mathcal{X}_2^n$, $\mathbf{y}_1 = (y_{11}, ..., y_{1n}) \in \mathcal{Y}_1^n$ and $\mathbf{y}_2 = (y_{21}, ..., y_{2n}) \in \mathcal{Y}_2^n$. The marginal distributions p_1, p_2 of $\mathcal{P}$ are given by

$$\begin{aligned} p_1(y_1|x_1, x_2) &= \sum_{y_2 \in \mathcal{Y}_2} p_{Y|X}(y_1, y_2|x_1, x_2) \\ p_2(y_2|x_1, x_2) &= \sum_{y_1 \in \mathcal{Y}_2} p_{Y|X}(y_1, y_2|x_1, x_2). \end{aligned} \tag{2.17}$$

The representation of the discrete interference channel is illustrated in Figure 2.2.

One of the main problem of this kind of channels lies on how to deal with mutual interference. Based on the level of interference experienced, a different strategy can be applied. The receiver can address the interference in the following three different ways

(i) treat interference as noise,

(ii) partially decode interference,

(iii) fully decode interference.

Rate splitting and *superposition encoding* are two encoding techniques which were proposed in order to partially or fully decode interference and they are explained next.

2.2.4.1 Rate Splitting

The *rate splitting* coding technique was introduced by Carlial in [36] with the aim of allowing each receiver to decode part of the message intended to the other receiver, hence minimizing the effect of the interference.

In this approach each encoder splits its own message into a *common* message and in a *private* message. The available transmission power is divided between the two split messages and a transmission rate is assigned for each of the messages. The goal of this technique is to allow each receiver to decode the undesired common message,

therefore reducing the interference on its desired message. In the splitting process, each transmitter lowers the rate of the common part of the message in order to guarantee decoding at the opposite receiver.

Han and Kobayashi in [37] use the rate splitting technique in order to derive the best-known achievable rate region for the interference channel. Further details are explained in Section 3.2.

2.2.4.2 Superposition Coding

The superposition coding technique was studied by Cover in [38]. Cover proposed this technique in a scenario where a single transmitter wants to simultaneously send independent messages to multiple receivers. Considering the two-user case, the receiver who has higher channel coefficient is referred as *strong user* while the receiver who has lower channel coefficient is considered the *weaker user*. The idea behind this technique is that the transmitter superimposes the message destined to the weaker receiver on top of the message destined to the stronger receiver. As in the rate splitting case, the transmitter splits the available transmission power between the two messages and selects the transmission rate for each of the message. At the receiver side, each user uses different approaches for decoding the own message. The weaker user decodes its message treating the superimposed signal as noise. Instead, the strongest user decodes first the undesired superposed message and subtracts it from the original signal by applying successive interference cancellation. At end, the stronger user decodes the own message which is interference-free.

Chapter 3

Channel Models

In this chapter we briefly introduce the channel models considered in the thesis. In particular, we present eight different transmission models for the content delivery over the Gaussian channel: the Gaussian interference as noise (GIN), the Gaussian interference channel without common information (GIC-WC), the Gaussian interference channel with common information (GIC-C), the broadcast channel (BC), the multiple input multiple output Gaussian broadcast channel (MIMO-BC), the multicast channel (MC), the orthogonal channel (OC) and the multiple input single output orthogonal channel (MISO-OC).

3.1 Gaussian Interference Channel

The two-users Gaussian interference channel model is represented in Fig. 3.1 and consists of two transmitters and two receivers. The study of this kind of channel was iniziated by Shannon in [39] and has been further studied by Sato in [40] and Carleial in [36].
The memoryless Gaussian interference channel is a quintuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{P}, \mathcal{Y}_1, \mathcal{Y}_2)$ with $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \mathbb{R}$ where $\mathbb{R}$ is the real numbers field, and the channel probability in $\mathcal{P}$ are specified by

$$y_1 = \sqrt{a_{11}} \cdot x_1 + \sqrt{a_{12}} \cdot x_2 + z_1 \tag{3.1}$$

$$y_2 = \sqrt{a_{22}} \cdot x_1 + \sqrt{a_{21}} \cdot x_2 + z_2 \tag{3.2}$$

for $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y_1 \in \mathcal{Y}_1$ and $y_2 \in \mathcal{Y}_2$, where z_1, z_2 are independent Gaussian random variables with mean zero and power variance N_1 and N_2, respectively. In this

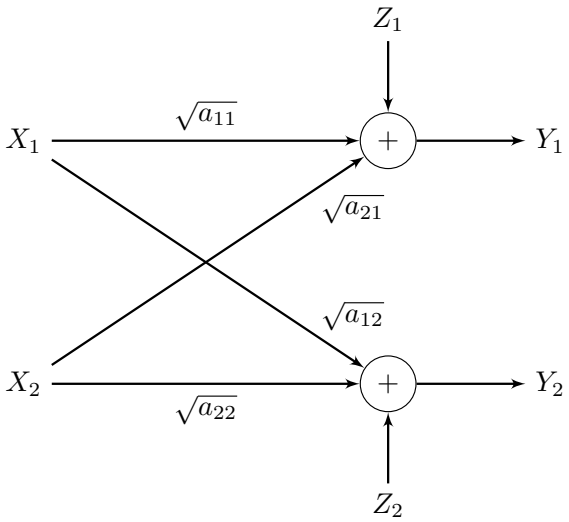


Fig. 3.1 Gaussian interference channel (GIN).

work, unitary variance is considered ($N_1 = N_2 = 1$). The quantities $\sqrt{a_{ij}}$ represent the real value channel gain from transmitter i to receiver j while a_{ij} represent the power transmission coefficients.

3.1.1 Scaling Transformation of the Gaussian Interference Channel

The physical model of the Gaussian interference channel in (3.1)-(3.2) can be transformed into an equivalent model which is capable of exactly the same communication performance.

We call *scaling transformation* of a channel on which changes in output signals due to a transformation on the parameters and input signals can be removed by constant gain factors, i.e. attenuation or amplification, at the output terminals.

3.1.2 Standard Gaussian Interference Channel

We say that the channel has the *standard form* when all direct coefficients are equal to unity and all noise powers are also equal to unity, i.e. $N_1 = N_2 = 1$ and $a_{11} = a_{22} = 1$. Through a scaling transformation it is possible to convert the channel into the standard form.

For example, from the point of view of achievable rates, the channel model (3.1)-(3.2)

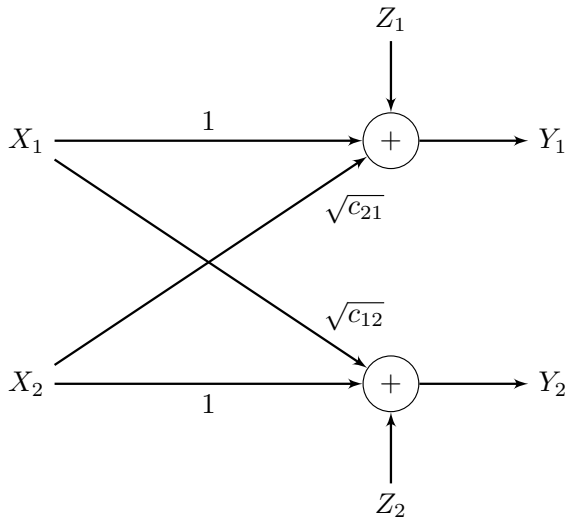


Fig. 3.2 Standard Gaussian interference channel (GIN).

is equivalent to the standard model

$$y_1 = x_1 + \sqrt{c_{12}} \cdot x_2 + z_1 \tag{3.3}$$

$$y_2 = \sqrt{c_{21}} \cdot x_1 + x_2 + z_2 \tag{3.4}$$

where the scaled coefficients c_{12} and c_{21} are calculated as follows

$$c_{12} = \frac{a_{12}}{a_{22}} \qquad c_{21} = \frac{a_{21}}{a_{11}}. \tag{3.5}$$

If we denote with $\tilde{P}_i$ the power of the codeword x_i in the standard model and with P_i the corresponding physical power then follows

$$P_i = \frac{\tilde{P}_i}{a_{ii}}.$$

The standard form of the two-user Gaussian interference channel is represented in Fig. 3.2.

3.1.3 Achievable Rates of the Gaussian as Noise Interference Channel

We define the GIN channel the transmission channel scheme where each receiver considers the whole interfere signal originated from the other transmitter as noise. From the equations of the model (3.1)-(3.2) we can derive the following rate conditions

for transmitting over a standard GIN channel

$$\begin{aligned} R_1 &\leq \frac{1}{2}\log_2\left(1+\frac{a_{11}\cdot P_1}{1+a_{12}\cdot P_2}\right) \\ R_2 &\leq \frac{1}{2}\log_2\left(1+\frac{a_{22}\cdot P_2}{1+a_{21}\cdot P_2}\right). \end{aligned} \tag{3.6}$$

3.2 Gaussian Interference Channel without Common Information

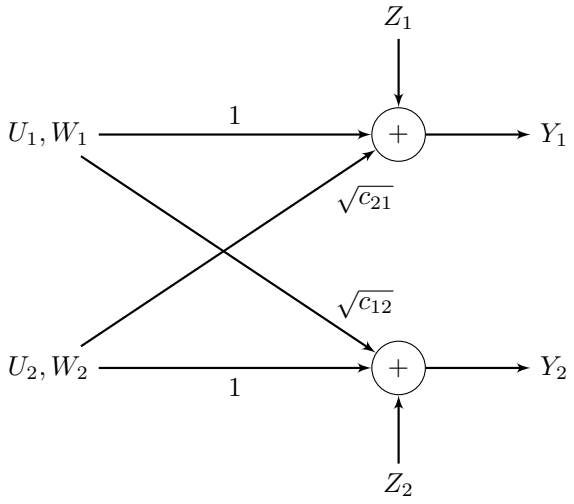


Fig. 3.3 Standard Gaussian interference channel without common information (GIC-WC).

We denote as the GIC-WC the channel model which applies the Han-Kobayashi (HK) transmission approach proposed in [37]. In such work, authors derive the best-known achievable rate region for the two-user Gaussian interference channel by optimizing over two power splitting variables. The scheme is based on rate splitting and joint decoding at the receivers. The representation of the GIC-WC is depicted in Fig. 3.5.

In the HK coding strategy, each message X_1 and X_2 is divided into two parts, a *private* message denoted with U_i and in a *public* message denoted with W_i for $i = 1, 2$. The two variables λ_1 and λ_2 are the power splitting coefficients with

$$0 \leq \lambda_i \leq 1$$

for $i = 1, 2$. If P_i is the power constraint of message X_i then the power of the private P_{U_i} and of the public message P_{W_i} are defined as follows

$$P_{U_i} = \lambda_i P_i$$
$$P_{W_i} = \bar{\lambda}_i P_i$$

where

$$\bar{\lambda}_i = 1 - \lambda_i.$$

Each encoder uses superposition to transmit its message. The splitting technique proposed allows each of the receivers to decode part of the information from its non intended sender. At the receiver side, public messages are jointly decoded along with the intended private message. Decoding part of the interfering signal allows to achieve higher rates with respect to the GIN transmission case.

The behavior of the HK scheme changes depending on the level of mutual interference suffered by the system. Three different interference regimes can be distinguished, namely *weak interference*, *moderate interference* and *strong interference*. For instance, when the interference is weak, the receiver gets low level of power of the undesired message, which makes it difficult to distinguish it from the noise. In this case, the GIC-WC technique treats the interfering signal as noise. When the interference coefficients are sufficiently high, the interfering message becomes the dominant component of the received signal. In this case, the best choice is to fully decode the interfering message and afterwards decode the desired one. When the interference is moderate, the idea is to partially decode the interfering signal and partially treat it as noise. All these three cases are considered with the HK scheme which splits the messages into two parts according to the channel conditions.

The achievable rates (R_1, R_2) for the standard model form of the HK scheme are given by

$$R_1 \leq \rho_1 \tag{3.7}$$
$$R_2 \leq \rho_2 \tag{3.8}$$
$$R_1 + R_2 \leq \rho_{12} \tag{3.9}$$
$$2 \cdot R_1 + R_2 \leq \rho_{10} \tag{3.10}$$
$$R_1 + 2 \cdot R_2 \leq \rho_{20} \tag{3.11}$$

where

$$
\begin{aligned}
\rho_1(c_{12},c_{21}) &= \sigma_1 + I(Y_1;U_1|W_1W_2),\\
\rho_2(c_{12},c_{21}) &= \sigma_2 + I(Y_2;U_2|W_1W_2),\\
\rho_{12}(c_{12},c_{21}) &= I(Y_1;U_1|W_1W_2) + I(Y_2;U_2|W_1W_2) + \min\Big\{I(Y_1;W_1W_2),\\
&\qquad I(Y_2;W_1W_2), I(Y_1;W_1|W_2) + I(Y_2;W_2|W_1), I(Y_2;W_1|W_2) + I(Y_1;W_2|W_1)\Big\},\\
\rho_{10}(c_{12},c_{21}) &= 2\sigma_1 + 2I(Y_1;U_1|W_1W_2) + I(Y_2;U_2|W_1W_2) - [\sigma_1 - I(Y_2;W_1|W_2)]^+ +\\
&\qquad \min\Big\{I(Y_2;W_2|W_1), I(Y_2;W_2) + [I(Y_2;W_1|W_2) - \sigma_1]^+, I(Y_1;W_2|W_1),\\
&\qquad I(Y_1,W_1W_2) - \sigma_1\Big\},\\
\rho_{20}(c_{12},c_{21}) &= 2\sigma_2 + 2I(Y_2;U_2|W_1W_2) + I(Y_1;U_1|W_1W_2) - \Big[\sigma_2 - I(Y_1;W_2|W_1)\Big]^+ +\\
&\qquad \min\Big\{I(Y_1;W_1|W_2), I(Y_1;W_1) + \Big[I(Y_1;W_2|W_1) - \sigma_2\Big]^+, I(Y_2;W_1|W_1),\\
&\qquad I(Y_2,W_1W_2) - \sigma_2\Big\},\\
[x]^+ &= \max\{0,x\}.
\end{aligned}
$$

while

$$
\begin{aligned}
\sigma_1 &= \min\Big\{I(Y_1;W_1|W_2), I(Y_2;W_1|U_2W_2)\Big\},\\
\sigma_2 &= \min\Big\{I(Y_2;W_1|W_1), I(Y_1;W_1|U_1W_1)\Big\},\\
\sigma_{12} &= \min\Big\{I(Y_1;W_1W_2), I(Y_2;W_1W_1), I(Y_1;W_1|W_2) + I(Y_2;W_2|W_1),\\
&\qquad I(Y_2;W_1|W_2) + I(Y_1;W_2|W_1)\Big\}.
\end{aligned}
$$

Denoting by $C(x) = \log_2(1+x)$ the Gaussian capacity function, then the mutual information in the expression above for the channel in standard form are given by

$$
\begin{aligned}
I(Y_1;U_1|W_1W_2) &= C\left(\lambda_1\tilde{P}_1/(1+c_{12}\lambda_2\tilde{P}_2)\right),\\
I(Y_2;U_2|W_1W_2) &= C\left(\lambda_2\tilde{P}_2/(1+c_{21}\lambda_1\tilde{P}_1)\right),\\
I(Y_1;W_1|W_2) &= C\left(\bar{\lambda}_1\tilde{P}_1/(1+\lambda_1\tilde{P}_1+c_{12}\lambda_2\tilde{P}_2)\right),\\
I(Y_1;W_2|W_1) &= C\left(c_{12}\bar{\lambda}_2\tilde{P}_2/(1+\lambda_1\tilde{P}_1+c_{12}\lambda_2\tilde{P}_2)\right),\\
I(Y_1;W_1W_2) &= C\left((\bar{\lambda}_1\tilde{P}_1+c_{12}\bar{\lambda}_2\tilde{P}_2)/(1+\lambda_1\tilde{P}_1+c_{12}\lambda_2\tilde{P}_2)\right),\\
I(Y_2;W_2|W_1) &= C\left(\bar{\lambda}_2\tilde{P}_2/(1+\lambda_2\tilde{P}_2+c_{21}\lambda_1\tilde{P}_1)\right),\\
I(Y_2;W_1|W_2) &= C\left(c_{21}\bar{\lambda}_1\tilde{P}_1/(1+\lambda_2\tilde{P}_2+c_{21}\lambda_1\tilde{P}_1)\right),\\
I(Y_2;W_1W_2) &= C\left((\bar{\lambda}_2\tilde{P}_2+c_{21}\bar{\lambda}_1\tilde{P}_1)/(1+\lambda_2\tilde{P}_2+c_{21}\lambda_1\tilde{P}_1)\right),
\end{aligned}
$$

$$
\begin{aligned}
I(Y_1; W_1) &= C\left(\bar{\lambda}_1 \tilde{P}_1/(1 + \lambda_1 \tilde{P}_1 + c_{12}\lambda_2 \tilde{P}_2)\right), \\
I(Y_2; W_2) &= C\left(\bar{\lambda}_2 \tilde{P}_2/(1 + \lambda_2 \tilde{P}_2 + c_{21}\lambda_1 \tilde{P}_1)\right), \\
I(Y_1; W_2|U_1 W_1) &= C\left(c_{12}\bar{\lambda}_2 \tilde{P}_2/(1 + c_{12}\lambda_2 \tilde{P}_2)\right), \\
I(Y_2; W_1|U_2 W_2) &= C\left(c_{21}\bar{\lambda}_1 \tilde{P}_1/(1 + c_{21}\lambda_1 \tilde{P}_1)\right).
\end{aligned}
$$

3.3 Gaussian Interference Channel with Common Information

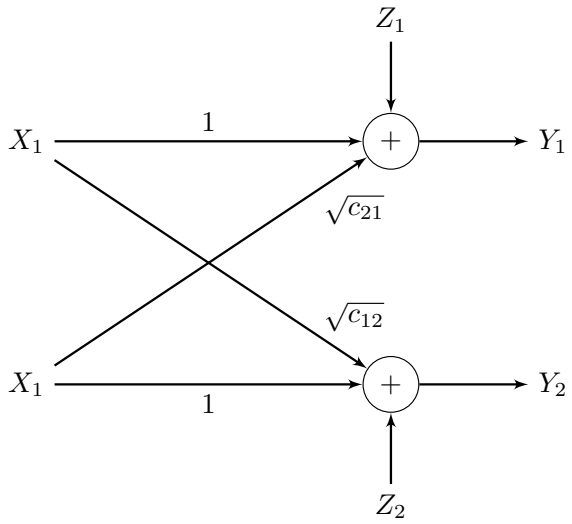


Fig. 3.4 Standard Gaussian interference channel with common information (GIC-C).

In the GIC-C each sender delivers the same message to the own receiver, i.e. $X_1 = X_2$ is transmitter by both senders. The representation of the standard form of the GIC-C is illustrated in Fig.3.4.

The achievable capacity region of a model where two senders need to deliver private messages as well as a common message to their corresponding receivers has been investigated in [41] by Jiang *et al.*. In such work, the achievable rate region is obtained by applying successive encoding and simultaneous decoding. Starting from the results derived in [41], we consider the case in which only a common message for both receivers is present. Thus, the achievable rates can be written as

$$
\begin{aligned}
R_1 &\leq \frac{1}{2}\log_2\left(1 + \left(\sqrt{\tilde{P}_1} + \sqrt{c_{21} \cdot \tilde{P}_2}\right)^2\right) \\
R_2 &\leq \frac{1}{2}\log_2\left(1 + \left(\sqrt{\tilde{P}_2} + \sqrt{c_{12} \cdot \tilde{P}_1}\right)^2\right).
\end{aligned}
\tag{3.12}
$$

3.4 Gaussian Broadcast Channel

The BC with power constraint P was studied by Cover in [35]. The two-user Gaussian broadcast channel is illustrated in Fig. 3.5 and consists of two receivers and a single transmitter.
For deriving the achievable rates of the BC channel, let us define the following quantities

$$a_+ = \max\left\{\sqrt{a_{12}}, \sqrt{a_{21}}\right\} \tag{3.13}$$

$$a_- = \min\left\{\sqrt{a_{12}}, \sqrt{a_{21}}\right\}. \tag{3.14}$$

The transmitter, to encode the messages, generates the output signal as the sum of two codewords. One codeword is generated with power P_+ at rate R_+ and the other codeword with power P_- at rate R_-. The message at rate $R_+(R_-)$ is destined to the receiver with channel coefficient $a_+(a_-)$. When the message arrives at the receivers, the worst user, i.e that with channel coefficient a_-, considers the codeword at power P_+ as noise. The good receiver, i.e. that with channel coefficient a_+, first decodes the codeword destined to the worst user and subtracts such codeword from the message. Successively, the best user decodes the desired codeword.

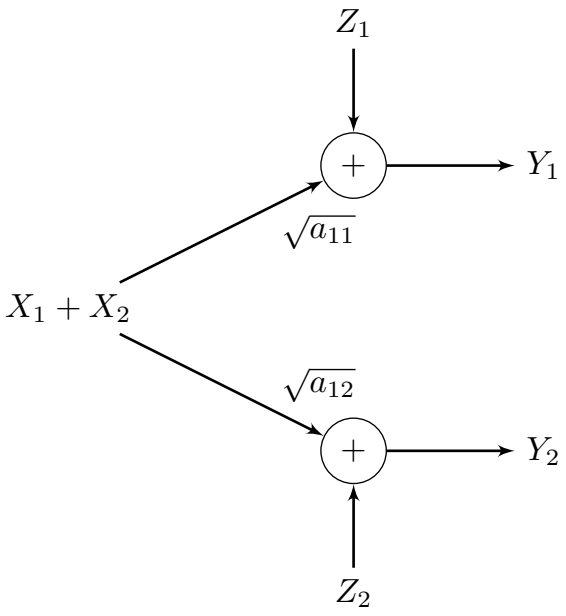


Fig. 3.5 Gaussian broadcast channel (BC).

The achievable rates for the Gaussian BC are

$$R_+ \leq \frac{1}{2}\log_2(1 + a_+ \cdot P_+) \tag{3.15}$$

$$R_- \leq \frac{1}{2}\log_2\left(\frac{a_- \cdot P_-}{1 + a_+ \cdot P_+}\right). \tag{3.16}$$

3.5 Gaussian Multiple Input Multiple Output Broadcast Channel

In this work, we consider as a two-user MIMO-BC the transmission channel model composed by two single-antenna transmitters which are coordinated for sending the messages to two single antenna users, as illustrated in Figure 3.6. The achievable region for this channel is derived by assuming perfect knowledge of the channel state information (CSI) at both ends and by applying dirty paper coding (DPC) which is briefly explained in the next subsection.

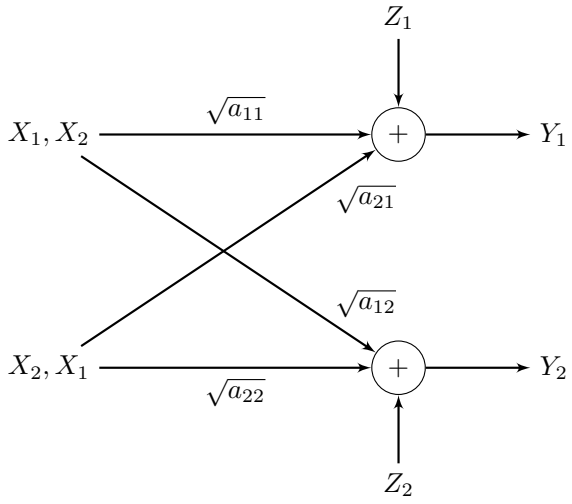


Fig. 3.6 Gaussian multiple input multiple output broadcast channel (MIMO-BC).

3.5.1 Dirty Paper Coding

The DPC is a coding scheme introduced by Costa in [42]. DPC consists in a layered precoding of the data so that the effect of the self-induced interference can be canceled out. In this technique, user messages are sequentially encoded. The decoding is such that a message experiences interference only from messages encoded after him. In

this case, the message for u_1 is interfered by the message for u_2 and u_2 experiences no interference. This technique allows to partially cancel out interference, which is known at the transmitter.

The DPC technique can also be applied in a system where each transmitter has to send a message to the respective users and the transmitters are in posses of all the messages. For example, in a two-user network with cache capabilities when both transmitters have both messages requested by the users. In this case, transmitters can coordinate in order to apply the DPC technique.

3.5.2 MIMO-BC Achievable Rate

The achievable rates for the MIMO-BC was studied by Shamai *et. al* in [43]. The achievable rates in this channel depend on the order in which the messages are encoded. The possible permutation encoding orders for the DPC technique are indicated with π_m with $m = \{1, 2\}$ such that

$$\pi_1 = \{2, 1\} \qquad \pi_2 = \{1, 2\}.$$

The permutation π_1 indicates that the message for the user two is encoded before the message for the user one while the permutation π_2 indicates that the message for the user one is encoded before the message for the user two.
We indicate with $\pi_m(l)$ the l-th element of the encoding order m such that

$$\begin{aligned} \pi_1(1) &= \{2\} & \pi_2(1) &= \{1\} \\ \pi_1(2) &= \{1\} & \pi_2(2) &= \{2\}. \end{aligned}$$

We denote with $\mathbf{H}_1$ the channel matrix between transmitters and receiver one and with $\mathbf{H}_2$ the channel matrix between transmitters and receiver two. The capacity region of

the two-user MIMO-BC channel is

$$\mathcal{C}(P, \mathbf{I}, \mathbf{H}_1, \mathbf{H}_2) = \bigcup_{\substack{S\succeq 0 \\ P_1+P_2\leq P}} conv\Bigg\{ \bigcup_{\pi\in\{\pi_1,\pi_2\}} \Bigg\{ (R_1, R_2) \Bigg| R_1^{\pi_1} = \frac{1}{2}\log_2 \frac{|\mathbf{H}_1\mathbf{B}_1\mathbf{H}_1^T + \mathbf{I}|}{|\mathbf{I}|},$$
$$R_2^{\pi_1} = \frac{1}{2}\log_2 \frac{|\mathbf{H}_2(\mathbf{B}_1+\mathbf{B}_2)\mathbf{H}_2^T + \mathbf{I}|}{|\mathbf{H}_2\mathbf{B}_1\mathbf{H}_2^T + \mathbf{I}|},$$
$$R_1^{\pi_2} = \frac{1}{2}\log_2 \frac{|\mathbf{H}_1(\mathbf{B}_1+\mathbf{B}_2)\mathbf{H}_1^T + \mathbf{I}|}{|\mathbf{H}_1\mathbf{B}_2\mathbf{H}_1^T + \mathbf{I}|},$$
$$R_2^{\pi_2} = \frac{1}{2}\log_2 \frac{|\mathbf{H}_2\mathbf{B}_2\mathbf{H}_2^T + \mathbf{I}|}{|\mathbf{I}|},$$
$$\text{for some } \mathbf{B}_1, \mathbf{B}_2 \text{ s.t. } \mathbf{S} = \mathbf{B}_1 + \mathbf{B}_2 \succeq 0, \mathbf{B}_1 \succeq 0, \mathbf{B}_2 \succeq 0 \Bigg\} \Bigg\}. \quad (3.17)$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are precoding matrices, $\mathbf{S}_i$ is the input covariance matrix for user i, $\mathbf{I}$ determines the channel noise variance and $|\cdot|$ denotes the determinant. The total capacity is given by the convex closure of this region.

3.6 Gaussian Multicast Channel

The MC consists in one transmitter and two receivers, as illustrated in Fig. 3.7. In this channel the receivers are interested in the same transmitted message which has to be transmitted at rate R_1.
The capacity region of the multicast channel is defined by

$$R_1 \leq \frac{1}{2}\log_2(1 + a_{11}\cdot P_1) \quad (3.18)$$
$$R_1 \leq \frac{1}{2}\log_2(1 + a_{12}\cdot P_2). \quad (3.19)$$

3.7 Orthogonal Channels

The two user OC is depicted in Fig. 3.8. In this channel model each transmitter operates in a different frequency band such that no interference is received from the other transmitter.

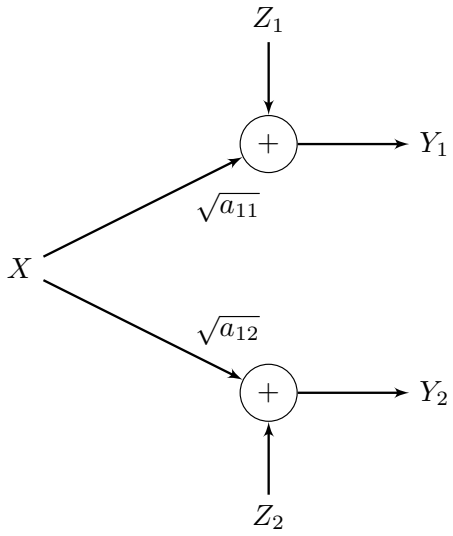


Fig. 3.7 Gaussian multicast channel (MC).

The capacities of the orthogonal channels are

$$R_1 \leq \frac{1}{2}\log_2(1 + a_{11} \cdot P_1) \tag{3.20}$$

$$R_2 \leq \frac{1}{2}\log_2(1 + a_{22} \cdot P_2). \tag{3.21}$$

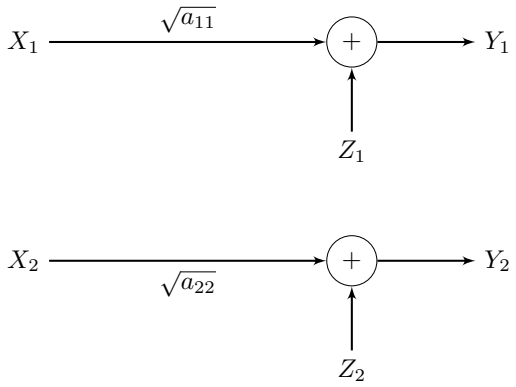


Fig. 3.8 Gaussian orthogonal channel (OC).

3.8 Multiple Input Single Output Orthogonal Channel

The two-user MISO-OC is represented in Fig. 3.9. In this channel model, it is assumed that there are three transmitters, two of them are coordinated to act as an antenna terminal and transmit the message to one receiver while the other transmitter operates in a different frequency band in order to serve the second user.
The achievable rates in the MISO-OC orthogonal channel are given by [44]

$$R_1 \leq \frac{1}{2}\log_2\Big(1 + (a_{11} + a_{21}) \cdot P_1\Big) \tag{3.22}$$

$$R_2 \leq \frac{1}{2}\log_2\Big(1 + a_{22} \cdot P_2\Big). \tag{3.23}$$

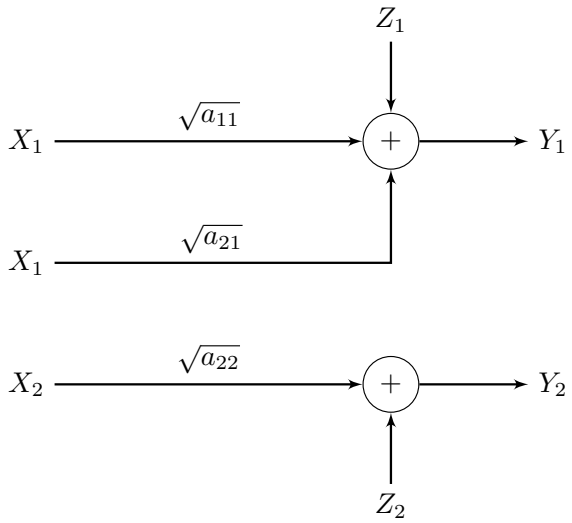


Fig. 3.9 Gaussian multiple input multiple output orthogonal channel (MIMO-OC).

Chapter 4

Caching in the Gaussian Channel

In this chapter, we analyze the placement and the delivery phase of caching when content is delivered over a Gaussian channel in presence of interference. Unlike previously considered in literature, we consider the case in which each message has to be transferred at a specific transmission rate in a one-shot delivery phase. The problem of delivering each content at a specific transmission rate and potentially with interference calls for selecting the correct level of power. Interference poses a challenge when the placement caching allocation problem has to be formulated. We analyze the power transmission consumption and we derive the optimization problem which selects the best caching policy to apply when cache-aided transmitters have to deliver contents over the Gaussian interference channel. To this end, the first part of the chapter is concentrated on the study of the power consumption in the delivery phase. For each transmission channel, a solution for minimizing the power consumption is derived. Then, the placement phase is studied, the caching optimization problem which finds the best cache allocation that minimize the average power expenditure is proposed.

4.1 Caching in Heterogeneous Networks

In cellular terrestrial networks, a two-tier heterogeneous network consists of deploying a large number of low power SBSs controlled by a central node denoted as MBS or master node. Such architecture is predicted for the next generation of mobile networks, 5G, where the MBS operates at lower frequencies to provide continuous coverage and mobility to users while SBSs use millimeter-wave (mmWaves) for fast data transfer. The SBSs can be connected to the MBS via wired or wireless backhaul. SBSs may not only help to serve the users but also can reduce the power consumption and increase

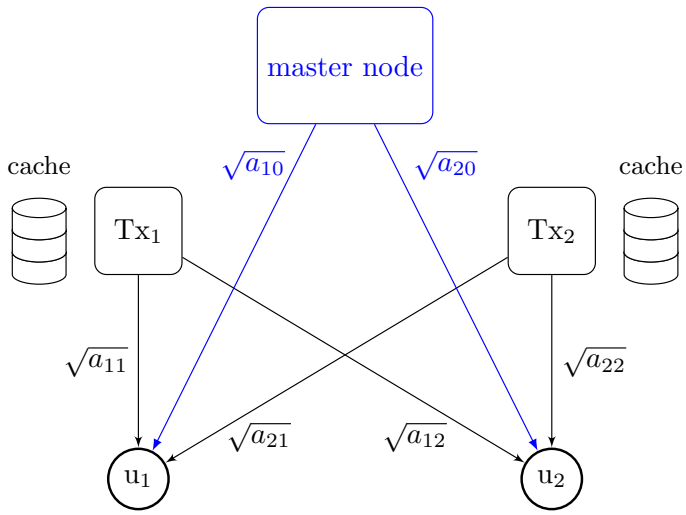


Fig. 4.1 System model of the considered heterogeneous network

the efficiency with which network resources are used. In this kind of heterogeneous network, significant gains can be obtained if contents are cached at the edge of the network. In fact, when contents are closer to end users, higher throughput and higher energy efficiency can be achieved. However, the dense deployment of SBSs introduces interference between small cells. One of the main problem, in this architectures is the interference management of the signals coming from the surrounding SBSs. We have introduced in the previous chapter techniques which either treat the interference as noise or partially/totally decode the undesired signal. Such techniques can be implemented by transmitters to deliver the cached content to users. Transmitters must be coordinated and the correct level of power should be selected for each message considering the presence of interference, when necessary. In a heterogeneous network, coordination can be carried out by the master node when it has perfect knowledge of the channel state information and knowledge of the cached content at each transmitter.

In the rest of this chapter, we derive the optimization problem to find the best caching policy for delivery content over a Gaussian interference channels when power consumption has to be minimized and a rate per-file constraint has to be respected. The two-user case is analyzed in this thesis, which represents a necessary preliminary step for extending the study to a generic number of users.

We want to highlight that the idea of memorizing content at the edge of an heterogeneous network not only concerns cellular terrestrial networks. Indeed, this concept can be extended to others types of heterogeneous networks. For example, in satellite networks, a cached heterogeneous network can consist of a geostationary Earth orbit (GEO) satellite acting as master node and LEO satellites connected to

respective cache-enabled ground stations. Another possibility is to have the GEO satellite connected directly to high altitude platform stations (HAPS) with cache capabilities serving the users. One last possibility is to employ UAV as mobile base stations. In such setting, a high altitude UAV, a satellite or an MBS could act as a master node while drones flying at lower altitude take the task to transmit cached content to the users.

The network depicted in Fig. 4.1 shows a general case of caching in a two-user heterogeneous network. We assume to have a master node, which has access to all the content that can be requested by the users, two transmitters with limited storage capability and two users. The master node has knowledge of all channel coefficients and requests of the users such that can coordinate transmitters in order to serve the users.

In the following, we study the placement and delivery phase optimization for the above mentioned scenarios. For ease of exposition, a terrestrial network scenario is assumed. Three different delivery setups in the two-user heterogeneous network are distinguished, namely

(i) *non-cooperative setup*

(ii) *limited cooperative setup*

(iii) *full cooperative setup*

In the non-cooperative setup, transmitters do not cooperate for serving the users, this means that there is no interference mitigation and all the signal coming for the other transmitter is considered as noise.

In the limited cooperative setup, transmitters cooperate for serving own users and mitigate interference. In this approach, each transmitter is allowed to serve only the own user.

In the full cooperative setup, transmitters do cooperate for counteracting the mutual interference and also the transmitters are allowed to serve the neighbor user. Serving the other user allows to spare resources of the master node.

4.1.1 System Model

The caching optimization problem is studied in a terrestrial heterogeneous network composed by an MBS, two SBS_n $n = \{1, 2\}$ and two users u_n where all the nodes in the network are equipped with a single antenna, as depicted in Fig. 4.2. The transmitters

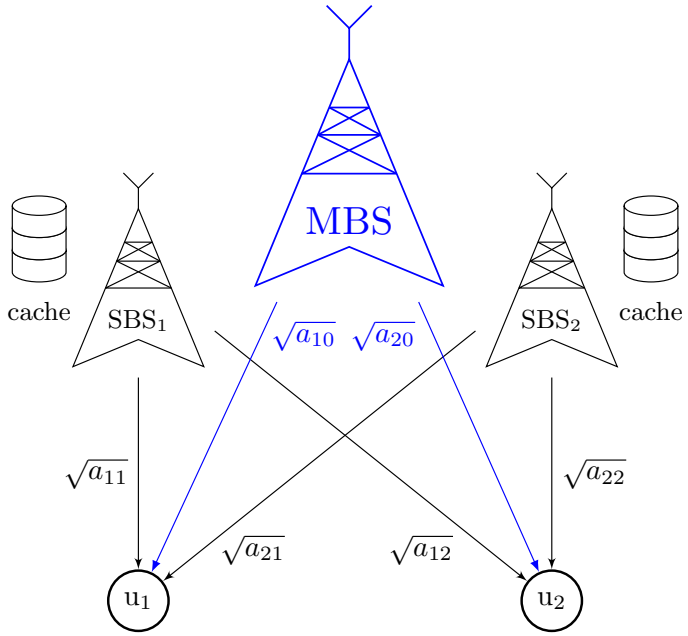


Fig. 4.2 System model of the considered heterogeneous network

nodes (SBSs) are equipped with a local cache with memory size M, the content of which is denoted by $\mathcal{M}_n$.

The MBS has access to a collection of N files (e.g. video files) $\mathcal{F} = \{f_1, ..., f_N\}$. Each file f_i has a minimum required transmission rate R_i, i.e., the minimum rate measured in bits/s/Hz at which the file has to be transmitted to the user in order to satisfy the QoS constraints. Although in the following we consider a constraint in terms of transmission rate over the channel, this can be easily translated into a constraint in terms of delay and quality, which is particularly suited to model video and image transmissions.

4.1.1.1 Notation

The channel coefficients are denoted by

- $\sqrt{a_{n0}}$: channel coefficient between u_n and the MBS,
- $\sqrt{a_{nm}}$: interference channel coefficient between u_n and SBS_m,
- $\sqrt{a_{nn}}$: direct channel coefficient between u_n and SBS_n.

The popularity ranking vector and the transmission rate vector are denoted by

- $\mathbf{q}_i = \{q_1, ..., q_i, ..., q_N\}$ where q_i represents the probability that f_i is requested.

- $\mathbf{R}_i = \{R_1, ..., R_i, ..., R_N\}$ where R_i represents the minimum rate requirement of file f_i in bits/s/Hz.

The presence or not of a content in the memory cache of a determined transmitter is indicated with the binary variable $x_{i,n}$. In particular, the following holds

$$x_{i,n} = \begin{cases} 1 \text{ if } f_i \in \mathcal{M}_n \\ 0 \text{ if } f_i \notin \mathcal{M}_n. \end{cases} \tag{4.1}$$

The complementary value of $x_{i,n}$ is defined as

$$\bar{x}_{i,n} \triangleq 1 - x_{i,n}.$$

In general, we indicate with $\bar{x}$ the complementary value of a binary variable x.
The allocation vectors containing the values for $x_{i,1}$ and $x_{i,2}$ are denoted with $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. As an example, if SBS_1 has a cache with size $M = 2$, the library of files has cardinality $N = 5$, i.e. $\mathcal{F} = \{f_1, f_2, f_3, f_4, f_5\}$, and the content of the cache of SBS_1 is $\mathcal{M}_1 = \{f_1, f_4\}$ then the allocation vector is expressed as

$$\mathbf{x}_1 = [10010].$$

4.1.2 Comparison of the Approaches

An important characteristic of cache-aided heterogeneous network is the presence of a central node which can coordinate the transmitters. The central node may coordinate the transmitters either for cooperative or non-cooperative transmissions. In fact, coordination is needed in both cases because in presence of interference each transmitter needs to know which is the level of transmit power to ensure a reliable communication at both receivers.

The caching policy which decides which content has to be cached at each SBSs strongly depends on whether cooperation between transmitters is in place. Three cases are investigated in order to understand which caching strategy has to be chosen to minimize the power consumption.

There is no cooperation when each transmitter delivers the cached content over the interference channel and each user considers the content addressed to the other user as noise. This study case corresponds to the description of the GIN channel model.

Instead, there is cooperation between transmitters when each of them adjusts the level of power in order to mitigate the interference caused to the other user. We distinguish two types of cooperation: limited cooperation and full cooperation. In the former case, it is assumed that each transmitter serves only its own user and cooperative techniques can be applied in order to reduce interference. In the latter case, it is assumed that a user can be served by any of the transmitters. In the limited cooperative case the following cooperative transmission models can be applied: GIC-C, GIC-WC. In the full cooperative case the following cooperative transmission models can be applied: GIC-C, GIC-WC, MISO-OC and MIMO-BC. Note that in the last setups, since it is assumed that each transmitter can serve both users, they can implement BC and MC transmissions.

In all approaches, when requests arrive from the users, the MBS decides who serve the users and how the users are served. The decision is made based in the following two considerations: whether the requested content is present in the SBSs' caches and whether the level of mutual interference allows a reliable communication. Let us explain in the following when each delivery technique is selected.

4.1.2.1 Non-Cooperative Approach

In the non-cooperative delivery, users are coordinated for sending the contents to the user. Since each receiver considers the signal coming from the other transmitter as noise, the level of power has to be coordinated between them such that QoS constraints are fulfilled contemporary for both users.

In this approach, because of the strong interference the SBSs might not be able to send the files at the required transmission rate and in such case the MBS is the responsible for the transmission. We use the indicator function $\mathbb{1}_{\alpha}$ which determines if content can be delivered by the SBSs over the GIN channel, where α is a parameter which depends on the interference level and it is analytically derived in sub-section 4.2.1.1. In case of high levels of interference, it might be that a solution does not exist for sending the content over the GIN channel at the required per-file rate. In such cases, the MBS serves the users either by a broadcast or a multicast transmission.

In the decision Tables 4.1 and 4.2 it is summarized how the MBS determines the sender and the transmission model to chose based on the cached content of the transmitters. Table 4.1 is the delivery strategy when $i \neq j$ (users ask for different content) while Table 4.2 considers the case when $i = j$ (users ask for the same content). The first column in Table 4.1 indicates the presence or not of the file requested in SBS_1' caches while the second column indicates if the files are present in SBS_2's cache, as

defined in eq. (4.1), where $\mathcal{M}_i$ represents the cache memory of SBS_i. The third column is the indication function which specifies if content can be delivered by applying the GIN transmission technique. The last two columns of the table indicate who is serving the users and which transmission technique has to be applied.

Table 4.1 Delivery strategy in the cooperative approach when different content is requested

$x_{i,1}$	$x_{j,2}$	$\mathbb{1}_\alpha$	Transmitter	Model
0	0	-	MBS	BC
0	1	-	SBS_2 - MBS	OC
1	0	-	SBS_1 - MBS	OC
1	1	1	SBS_1 - SBS_2	GIN
1	1	0	SBS	BC

Table 4.2 Delivery strategy in the non-cooperative approach when the same content is requested

$x_{i,1}$	$x_{i,2}$	$\mathbb{1}_\alpha$	Transmitter	Model
0	0	-	MBS	MC
0	1	-	SBS_2 - MBS	OC
1	0	-	SBS_1 - MBS	OC
1	1	1	SBS_1 - SBS_2	GIN
1	1	0	MBS	MC

4.1.2.2 Limited Cooperative Approach

In the limited cooperative approach, the SBSs are coordinated for sending the contents to the users at the minimum rate required. This approach is based on mitigating the mutual interference by applying the coding techniques studied. In this setup, transmitted may be coordinated for applying the GIC-WC and GIC-C transmissions models.

However, in this setup, it is assumed that a transmitter can only serve the own user. When a pair of requests arrives to the network then the MBS decides who and how the contents have to be delivered to the users. The decision is made based only on the caches' content. In Tables 4.3 and 4.4 are summarized how the MBS determines the sender and the transmission model for delivering the content. We recall that it is supposed that u_1 requires f_i and u_2 requires f_j. Table 4.5 represents the delivery

strategy when $i \neq j$ (users asks for different content) while Table 4.6 represents the delivery strategy when $i = j$ (users ask for the same content).

Table 4.3 Delivery strategy in the cooperative approach when different content is requested

$x_{i,1}$	$x_{j,2}$	Transmitter	Model
0	0	MBS	BC
1	0	SBS_1 - MBS	OC
0	1	SBS_2 - MBS	OC
1	1	SBS_1 - SBS_2	GIC-WC

Table 4.4 Delivery strategy in the cooperative approach when the same content is requested

$x_{i,1}$	$x_{i,2}$	Transmitter	Model
0	0	MBS	MC
1	0	SBS_1 - MBS	OC
0	1	SBS_2 - MBS	OC
1	1	SBS_1 -SBS_2	GIC-C

4.1.2.3 Full Cooperative Approach

In the full cooperative approach the SBSs are coordinated for sending the contents to the users at the minimum rate required. The full cooperation approach consists on mitigating the mutual interference by applying coding techniques and the possibility from the transmitters to serve any user.

When a pair of requests arrives to the network then the MBS decides who and how the contents have to be delivered to the users. In this approach, the decision is made based only on the caches' content. This setup represents a scenario where the main interest of the MBS is to spare own resources such that the MBS always decides for serving the users with the cached content at transmitters. Hence, each SBS serves the respective user through the direct link and might also serve the other user through the interference link. When content is not available at the SBSs, it is transmitted by the MBS.

In Tables 4.5 and 4.6 is summarized how the MBS determines the sender and the transmission model for delivering the content. Table 4.5 represents the delivery strategy when $i \neq j$ (users asks for different content) while Table 4.6 represents the delivery strategy when $i = j$ (users ask for the same content). Note that in this setup, since each transmitter can serve both users, it is necessary to identify if the content requested by both users is present or not in the cache of each transmitter. Hence, we add two columns in the delivery strategy table.

In the full cooperative approach, if none of the files requested are present in the transmitters' caches then the MBS serves both users by sending the messages through a broadcast/multicast transmission. Instead, for example, if f_i requested by u_1 is not present in the caches and SBS_1 has the content f_j requested from u_2, then SBS_1 serves u_2 while u_1 is served by the MBS. In this case the MBS and SBS_1 implement an orthogonal transmission channel.

4.2 Delivery Phase: Power Optimization

While a large part of works in literature focus on maximizing the achievable rates given a channel model, little research has been dedicated to the inverse problem: how to minimize the transmit power when the data rate for each receiver given a channel model is specified. The minimization of the power when transmission rates are fixed is related to the QoS requirement in a downlink transmission.

In this subsection, we concentrate on the study of the delivery phase of a cache-aided network from the power consumption point of view. It is assumed that the placement

Table 4.5 Delivery strategy in the cooperative approach when different content is requested

$x_{i,1}$	$x_{j,1}$	$x_{i,2}$	$x_{j,2}$	Transmitter	Model
0	0	0	0	MBS	BC
0	0	0	1	SBS_2 - MBS	OC
0	0	1	0	SBS_2 - MBS	OC
0	0	1	1	SBS_2	BC
0	1	0	0	SBS_1 - MBS	OC
0	1	0	1	SBS_1 - SBS_2-MBS	MISO-OC
0	1	1	0	SBS_1 - SBS_2	GIC-WC
0	1	1	1	SBS_1 - SBS_2	GIC-WC
1	0	0	0	SBS_1 - MBS	OC
1	0	0	1	SBS_1 - SBS_2	GIC-WC
1	0	1	0	SBS_1 - SBS_2-MBS	MISO-OC
1	0	1	1	SBS_1 - SBS_2	GIC-WC
1	1	0	0	SBS_1	BC
1	1	0	1	SBS_1 - SBS_2	GIC-WC
1	1	1	0	SBS_1 - SBS_2	GIC-WC
1	1	1	1	SBS_1 - SBS_2-MBS	MIMO-BC

Table 4.6 Delivery strategy in the cooperative approach when the same content is requested

$x_{i,1}$	$x_{i,2}$	Transmitter	Model
0	0	MBS	MC
0	1	SBS_2	MC
1	0	SBS_1	MC
1	1	SBS_1-SBS_2	GIC-C
1	1	MBS	BC

phase has been already carried out and each transmitter has the content requested by the own user.

4.2.1 Power Optimization

We assume without loss of generality that R_i is the rate requirement of the cached content f_i destined to receiver u_1 while R_j is the rate requirement of the cached content f_j destined to receiver u_2.

We introduce the cost function $c_T(i,j)$ as the minimum power consumed during the delivery phase by the system for transmitting the content f_i to receiver u_1 and the content f_j to receiver u_2 at rate (R_1, R_2), when it is implemented the transmission.

Finally, we want to highlight that our purpose is to minimize the physical power P_i, P_j. However, for some transmission scheme models it is easier to work in their standard form. In such cases, we recall the correspondence of the physical power in the standard form as

$$P_i = \frac{\tilde{P}_i}{a_{11}} \quad \text{and} \quad P_j = \frac{\tilde{P}_j}{a_{22}}.$$

4.2.1.1 GIN Power Minimization

The power optimization problem for transmissions in the GIN channel can be formulated as follows

$$\begin{aligned} \underset{P_i,P_j}{\text{minimize}}\ & P_i + P_j \\ \text{subject to}\ & \frac{1}{2}\log_2\left(1+\frac{a_{11}\cdot P_i}{1+a_{12}\cdot P_j}\right) \geq R_i \\ & \frac{1}{2}\log_2\left(1+\frac{a_{22}\cdot P_j}{1+a_{21}\cdot P_i}\right) \geq R_j \\ & P_i, P_j \geq 0. \end{aligned} \tag{4.2}$$

The problem can be solved analytically. We derive, as an example, all the passages the power in the first inequality. Such procedure is omitted in the rest of the thesis.

Given the first inequality, we have that

$$\begin{aligned} \frac{1}{2}\log_2\left(1+\frac{a_{11}\cdot P_i}{1+a_{12}\cdot P_j}\right) &\geq R_i \\ \log_2\left(1+\frac{a_{11}\cdot P_i}{1+a_{12}\cdot P_j}\right) &\geq 2R_i \\ \frac{a_{11}\cdot P_i}{1+a_{12}\cdot P_j} &\geq 2^{2R_i}-1 \\ a_{11}\cdot P_i &\geq (2^{2R_i}-1)\cdot(1+a_{12})\cdot P_j. \end{aligned}$$

where the power P_i is minimized when

$$P_i = \frac{1}{a_{11}}\cdot(2^{2R_i}-1)\cdot(a_{12}\cdot P_j+1) \tag{4.3}$$

and if we insert eq. (4.3) in the second inequality of eq. (4.2) we have

$$P_j \geq \frac{(2^{2R_j}-1)\cdot\left[\frac{a_{21}}{a_{11}}\cdot(2^{2R_i}-1)+1\right]}{a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2R_j}-1)\cdot(1-2^{2R_i})}. \tag{4.4}$$

Let call α the denominator of (4.4), such that

$$\alpha = a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2Rj}-1)\cdot(1-2^{2R_i}). \tag{4.5}$$

The eq. (4.4) admits a feasible solution when P_j is positive so that $\alpha > 0$. When $\alpha \leq 0$ do not exist P_i and P_j such that the (R_1, R_2) can be achieved so the problem has not solution. Let us define the indicator function $\mathbb{1}_\alpha$ as

$$\mathbb{1}_\alpha = \begin{cases} 1 & \alpha > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

The cost of the GIN channel when $\mathbb{1}_\alpha = 1$ is

$$\begin{aligned} c_{GIN}(i,j) =& \frac{(2^{2R_i}-1)}{a_{11}}\left(\frac{a_{12}(2^{2R_j}-1)\cdot\left[\frac{a_{21}}{a_{11}}\cdot(2^{2R_i}-1)+1\right]}{a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2R_j}-1)\cdot(1-2^{2R_i})}+1\right)+ \\ & \frac{(2^{2R_j}-1)\cdot\left[\frac{a_{21}}{a_{11}}\cdot(2^{2R_i}-1)+1\right]}{a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2R_j}-1)\cdot(1-2^{2R_i})}. \end{aligned} \tag{4.7}$$

We assume that if $\mathbb{1}_\alpha = 0$ the MBS applies broadcast or multicast transmission technique.

4.2.1.2 GIC-WC Power Minimization

The power minimization problem for transmissions in the GIC-WC channel can be formulated as follows

$$\underset{\tilde{P}_i,\tilde{P}_j}{\text{minimize}}\ \frac{\tilde{P}_i}{a_{11}}+\frac{\tilde{P}_j}{a_{22}} \tag{4.8}$$

$$\text{subject to } \rho_1 \geq R_1 \tag{4.9}$$

$$\rho_2 \geq R_2 \tag{4.10}$$

$$\rho_{12} \geq R_1 + R_2 \tag{4.11}$$

$$\rho_{10} \geq 2R_1 + R_2 \tag{4.12}$$

$$\rho_{20} \geq R_1 + 2R_2 \tag{4.13}$$

$$P_i, P_j \geq 0. \tag{4.14}$$

where $\rho_1, \rho_2, \rho_{12}, \rho_{10}, \rho_{20}$ are defined in Section 3.2. Solving the problem presented in eqs. (4.8) to (4.14) is a challenging task due to the intricacy of the HK conditions. In order to calculate the minimum cost $c_{\text{GIC-WC}}(i,j) = P_i + P_j$, we developed the algorithm *calculate_min_cost_HK*, shown in Algorithm 1 in pseudocode.
Let us define the following values[1]

- ΔP^{tot} the granularity at which the sum power is evaluated,
- ΔP the granularity at which the sum power is divided between the transmitters P_i and P_j,
- $\Delta\lambda$ the granularity in which power is split between the private and the common message.

The algorithm takes as input the minimum and maximum granularities of ΔP^{tot}, ΔP and $\Delta\lambda$ as well as the required file rates R_i and R_j and returns the minimum value of $c_{\text{GIC-WC}}$ that satisfies the rate constraints for the two cached files.
At initialization the algorithm performs the search over a grid of points[2] spaced apart by the largest granularity specified in the input in order to speed-up convergence. The granularity is progressively decreased to the minimum value specified in the input to enhance the accuracy of the solution.
The algorithm is much more efficient than exhaustive search over a grid with same minimum granularity. Specifically, if

$$\Delta P^{\text{tot}}_{\text{min}} \ll \Delta P^{\text{tot}}_{\text{max}},$$

the gain in terms of number of operations is on the order of

$$\left(\Delta P^{\text{tot}}_{\text{max}} / \Delta P^{\text{tot}}_{\text{min}} - 1\right) \times 1/\Delta\lambda_{\text{min}} \times 1/\Delta P_{\text{min}}.$$

The algorithm is based on the fact that, if a given sum power P^{tot} does not allow to satisfy the rate constraints for any rate- and power-split, then no $\tilde{P}^{\text{tot}} < P^{\text{tot}}$ can

[1] The three parameters are bounded as follows: $\Delta P^{\text{tot}} > 0$, $0 < \Delta P < 1$ and $0 < \Delta\lambda < 1$.

[2] As an example, the first point of the grid for a given P^{tot} is $(P_1,\ P_2,\ \lambda_1,\ \lambda_2) = (\Delta P \cdot P^{\text{tot}},\ (1-\Delta P)P^{\text{tot}},\ \Delta\lambda,\ \Delta\lambda)$, the second point is $(P_1,\ P_2,\ \lambda_1,\ \lambda_2) = (2\Delta P \cdot P^{\text{tot}},\ (1-2\Delta P)P^{\text{tot}},\ \Delta\lambda,\ \Delta\lambda)$ and so on.

satisfy such constraints.

The algorithm converges to the absolute minimum asymptotically as

$$\Delta P_{\min}^{\text{tot}} \to 0, \Delta\lambda_{\min} \to 0, \Delta P_{\min} \to 0.$$

Algorithm 1 $c_{\text{GIC-WC}}$ =calculate_min_cost_HK($\Delta P_{\min}^{\text{tot}}$, $\Delta P_{\max}^{\text{tot}}$, $\Delta P_{\min}$, $\Delta P_{\max}$, $\Delta\lambda_{\min}$, $\Delta\lambda_{\max}$, R_i, R_j)

$c_{GIwc} = -1$;
$P_{\text{tot}} = \Delta P_{\max}^{\text{tot}}$;
$\Delta P_{\text{tot}} = \Delta P_{\max}^{\text{tot}}$;
while $\Delta P^{\text{tot}} > \Delta P_{\min}^{\text{tot}}$ **do** ▷ Go on until the minimum granularity for ΔP^{tot} is reached
    flag = 0
    $\Delta P = \Delta P_{\max}$ ▷ Reset ΔP
    $\Delta\lambda = \Delta\lambda_{\max}$ ▷ Reset $\Delta\lambda$
    while flag == 0 **do**
        create $G(\Delta P, \Delta\lambda)$ ▷ Create search grid with current granularity excluding points previously checked
        for $g \in G(\Delta P, \Delta\lambda)$ **do** ▷ Each point g is defined by coordinates (P_i, P_j, λ_1, λ_2)
            if g allows to achieve the pair (R_i, R_j) **then**
                flag = 1
                $c_{GICWC} = P^{\text{tot}}$
                $\Delta P^{\text{tot}} = \Delta P^{\text{tot}}/2$
                $P^{\text{tot}} = P^{\text{tot}} - \Delta P^{\text{tot}}$ ▷ Check if (R_i, R_j) are achievable with a smaller P^{tot}
                break for
            end if
        end for
        if (flag == 0) **then**
            if $(\Delta P > \Delta P_{\min})$ **then**
                $\Delta P = \Delta P/2$
            else if $(\Delta\lambda > \Delta\lambda_{\min})$ **then**
                $\Delta\lambda = \Delta\lambda/2$
            else if $c_{GICWC} \neq -1$ **then** ▷ Check if so far (R_i, R_j) has been achieved
                $\Delta P^{\text{tot}} = \Delta P^{\text{tot}}/2$
                $P^{\text{tot}} = P^{\text{tot}} + \Delta P^{\text{tot}}$
            else
                $P^{\text{tot}} = P^{\text{tot}} + \Delta P^{\text{tot}}$ ▷ If (R_i, R_j) is not achievable with P^{tot}, the same holds for a smaller P^{tot}
                break inner while
            end if
        end if
    end while
end while
return $c_{\text{GIC-WC}}$

4.2.1.3 GIC-C Power Minimization

The power minimization for transmissions in the standard form of the GIC-C can be formulated as follows

$$\underset{\tilde{P}_i,\tilde{P}_j}{\text{minimize}} \ \frac{\tilde{P}_i}{a_{11}} + \frac{\tilde{P}_j}{a_{22}} \tag{4.15}$$

$$\text{subject to } \frac{1}{2}\log_2\left(1+\left(\sqrt{\tilde{P}_i}+\sqrt{c_{12}\cdot\tilde{P}_j}\right)^2\right) \geq R_i \tag{4.16}$$

$$\frac{1}{2}\log_2\left(1+\left(\sqrt{\tilde{P}_j}+\sqrt{c_{21}\cdot\tilde{P}_i}\right)^2\right) \geq R_i \tag{4.17}$$

$$P_i, P_j \geq 0. \tag{4.18}$$

We solve this problem by applying the Karush-Kuhn-Tucker (KKT) conditions [45]. The optimization problem can be rewritten as follows

$$\underset{P_i,P_j}{\text{minimize}} \ \frac{\tilde{P}_i}{a_{11}} + \frac{\tilde{P}_j}{a_{22}} \tag{4.19}$$

$$\text{subject to } -\sqrt{2^{2R_i}-1}+\sqrt{\tilde{P}_i}+\sqrt{c_{12}\cdot\tilde{P}_j} \geq 0 \tag{4.20}$$

$$-\sqrt{2^{2R_i}-1}+\sqrt{\tilde{P}_j}+\sqrt{c_{21}\cdot\tilde{P}_i} \geq 0 \tag{4.21}$$

$$\tilde{P}_i, \tilde{P}_j \geq 0. \tag{4.22}$$

The KKT conditions to be satisfied are:

$$\begin{aligned}
\nabla_{\mathbf{P}} L(\mathbf{P}^*, \lambda^*) &= 0, \\
g_m(\mathbf{P}^*) &\geq 0, \quad m = 1, ..., n \\
\lambda_m^* &\geq 0, \quad m = 1, ..., n \\
\lambda_m^* \cdot g_m(\mathbf{P}^*) &= 0, \quad m = 1, ..., n.
\end{aligned}$$

The derivation of KKT conditions for this problem are given in Appendix A.

The cost of this channel is

$$c_{GIC-C}(i,j) = \tilde{P}_i + \tilde{P}_j,$$

where $\tilde{P}_i + \tilde{P}_j$ is derived in Appendix A.

4.2.1.4 BC Power Minimization

The power optimization problem in the BC can be written as follows

$$\underset{P_-,P_+}{\text{mininize}}\ P_- + P_+ \tag{4.23}$$

$$\text{subject to } \frac{1}{2}\log_2\left(1 + a_+ \cdot P_+\right) \geq R_+ \tag{4.24}$$

$$\frac{1}{2}\log_2\left(1 + \frac{a_- \cdot P_-}{1 + a_- \cdot P_+}\right) \geq R_- \tag{4.25}$$

$$P_-, P_+ \geq 0. \tag{4.26}$$

The overall problem can be solved analytically as follows

$$P_+ = \frac{2^{2R_+} - 1}{a_{+0}} \tag{4.27}$$

$$P_- = \frac{(2^{2R_-} - 1)(1 + a_{-0} \cdot P_+)}{a_{-0}}. \tag{4.28}$$

Plugging P_+ in the expression of P_-, we obtain

$$P_- = \frac{1}{a_-}(2^{2R_-} - 1)\left(1 + \frac{a_-}{a_+}(2^{2R_+} - 1)\right). \tag{4.29}$$

The overall cost of this channel $c_{BC}(i,j)$ is

$$c_{BC}(i,j) = \frac{2^{2R_+} - 1}{a_+} + \frac{(2^{2R_-} - 1)(1 + a_- P_+)}{a_-}. \tag{4.30}$$

4.2.1.5 MIMO-BC Power Minimization

The minimization of the power in the MIMO-BC transmission channel was studied by Cioffi *et. al* in [46]. In the two-user MIMO-BC channel each user requires a specific transmission rate, R_1 for user one and R_2 for user two. It is assumed that transmitters know the channel information perfectly. The transmitter minimizes the power by determining the covariance matrix for the users such that the transmission rates are respected.

In this transmission channel we assume that the two single-antenna transmitters can collaborate to transmit both messages acting as a single multi-antenna terminal. The

channel matrices for our model are given by

$$\mathbf{H}_1 = \begin{bmatrix} \sqrt{a_{11}} \\ \sqrt{a_{21}} \end{bmatrix} \qquad \mathbf{H}_2 = \begin{bmatrix} \sqrt{a_{12}} \\ \sqrt{a_{22}} \end{bmatrix}. \tag{4.31}$$

In order to solve the minimization power we need to introduce a modification of the channel matrices and transform them into square matrices. The modification is suggested in [46] and consists in introducing a virtual receive antenna at each user with virtual channel gain equal to ϵ, with ϵ much smaller than any of the channel coefficients. Thus the modified channel matrices $\widetilde{\mathbf{H}}_1$ and $\widetilde{\mathbf{H}}_2$ are given by

$$\widetilde{\mathbf{H}}_1 = \begin{bmatrix} \sqrt{a_{11}} & \epsilon \\ \sqrt{a_{21}} & \epsilon \end{bmatrix} \qquad \widetilde{\mathbf{H}}_2 = \begin{bmatrix} \sqrt{a_{12}} & \epsilon \\ \sqrt{a_{22}} & \epsilon \end{bmatrix}. \tag{4.32}$$

Calling $\mathbf{S}_1^B$ and $\mathbf{S}_2^B$ the input covariance matrix for receiver one and receiver two, respectively, then the minimization problem for the two-user MIMO-BC can be written as

$$\begin{aligned} \underset{\mathbf{S}_1,\mathbf{S}_2,\pi_m}{\text{minimize}}\ & \text{tr}\left(\sum_{j=1}^{2} \mathbf{S}_j^B\right) \\ \text{subject to}\ & \frac{1}{2}\log_2 |\mathbf{I} + \widetilde{\mathbf{H}}_{\pi_m(2)}^T \mathbf{S}_{\pi_m(2)}^B \widetilde{\mathbf{H}}_{\pi_m(2)}| = R_{\pi_m(2)}^B \\ & \frac{1}{2}\log_2 \frac{|\mathbf{I} + \sum_{j=1}^{2} \widetilde{\mathbf{H}}_{\pi_m(1)} \mathbf{S}_j^B \widetilde{H}_{\pi_m(1)}|}{|\mathbf{I} + \widetilde{\mathbf{H}}_{\pi_m(1)}^T \mathbf{S}_{\pi_m(2)}^B \widetilde{\mathbf{H}}_{\pi_m(1)}|} = R_{\pi_m(1)}^B \\ & m = 1, 2, \end{aligned} \tag{4.33}$$

where $\text{tr}(\cdot)$ indicates the trace of the matrix, $\mathbf{A}^T$ indicates the transpose of the matrix $\mathbf{A}$ while the subscript π_m in the channel and input covariance matrices indicates the encoding order has to be also optimized since dirty paper coding is used.

The problem in eq. (4.33) is not easy to solve. However, it can be converted into a simple convex form by usign the BC-to-MAC transformation proposed in [47]. In such work, Vishwanath formulates the mathematical transformation BC-to-MAC (MAC-to-BC). By duality, maximizing the sum rate of a BC is equivalent to maximizing the sum rate in the dual MAC whose sum power is equal to the BC power constraints.

The equivalent convex optimization problem can be written as

$$\begin{aligned} \underset{\mathbf{S}_1^M, \mathbf{S}_2^M, \pi_m}{\text{minimize}} \quad & \text{tr}\left(\sum_{j=1}^{2} \mathbf{S}_j^M\right) \\ \text{subject to} \quad & \frac{1}{2}\log_2 |\mathbf{I} + \sum_{j=1}^{2} \widetilde{\mathbf{H}}_j \mathbf{S}_j^M \widetilde{\mathbf{H}}_j^T| = \bar{R}_{\pi_m(2)}^M \\ & \frac{1}{2}\log_2 |\mathbf{I} + \widetilde{\mathbf{H}}_{\pi_m(1)} \mathbf{S}_{\pi_m(1)}^M \widetilde{\mathbf{H}}_{\pi_m(1)}^T| = \bar{R}_{\pi_m(1)}^M \\ & m = 1, 2, \end{aligned} \tag{4.34}$$

where $\mathbf{S}_j^M$ indicates the dual MAC covariance matrix. The following holds

$$\text{tr}\left(\sum_{j=1}^{2} \mathbf{S}_j^B\right) = \text{tr}\left(\sum_{j=1}^{2} \mathbf{S}_j^M\right) \tag{4.35}$$

and

$$\bar{R}_i^M = \sum_{j=i}^{2} R_i^B. \tag{4.36}$$

It is not straightforward to find the solution to problem eq. (4.34). A suboptimal solution $\left(\mathbf{S}_1^M, \mathbf{S}_2^M\right)$ can be obtained using the iterative Algorithm 2 which has been suggested in [46][48] and we adopted to our model. We illustrate in the following how the problem is iterative solved and further details of the derivation of the Algorithm 2 are given in Appendix B.

The algorithm was built based on the *water-filling* problem and the following consideration. The problem in eq. (4.34) can be rewritten as equivalent problem which satisfies the KKT conditions. If we define

$$\tilde{\lambda}_i^{KKT} = \frac{1}{2}\sum_{j=1}^{i} \lambda_j^{KKT}$$

where λ_i^{KKT} are the Lagrange multipliers with $i = \{1, 2\}$ then solving eq. (4.34) for each permutation π_m is equivalent to solve

$$\text{find } \mathbf{S}_i^M \tag{4.37}$$

$$\text{subject to } \mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^T + \mathbf{N}_i = \tilde{\lambda}_i^{KKT}\mathbf{P}_i \tag{4.38}$$

$$\frac{1}{2}\log_2 \frac{|\mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^T + \mathbf{N}_i|}{|\mathbf{N}_i|} = R_i \tag{4.39}$$

where

$$\mathbf{N}_i = \mathbf{I} + \sum_{l=i+1}^{2} \mathbf{H}_l\mathbf{S}_l^M\mathbf{H}_l^T.$$

Algorithm 2 solves the problem described in eqs. (4.37) to (4.39) by applying single value decomposition (SVD) which is briefly explained next.

Algorithm 2 $(\mathbf{S}_1^M, \mathbf{S}_2^M)$ =calculate_min_power$(\mathbf{N}_1, \mathbf{N}_2, \mathbf{H}_1, \mathbf{H}_2)$

$\mathbf{S}_i^M$ = null matrix, $i = 1, 2$
while not further increase on tr$(\mathbf{S}_i)$ **do**
  for $i = 2$ to 1 **do**
    $\mathbf{Q}_i\mathbf{\Delta}_i\mathbf{Q}_i^T = \text{SVD}(\mathbf{N}_i)$
    $\hat{\mathbf{H}}_i = \Delta_i^{-\frac{1}{2}}\mathbf{Q}_i^T\mathbf{H}_i$
    $\mathbf{A}_i = \mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^T + \mathbf{I}$
    $\mathbf{P}_i = \{(\mathbf{H}_i\mathbf{H}_i^T)^{-1} + \mathbf{A}_i^{-1}\sum_{j=1}^{i-1}(\mathbf{H}_j\mathbf{S}_j^M\mathbf{H}_j^T)(\mathbf{H}_j\mathbf{H}_j^T)^{-1}\}^{-1}$
    $\mathbf{F}_i\hat{\mathbf{\Delta}}_i\mathbf{F}_i^T = \text{SVD}(\hat{\mathbf{P}}_i)$
    $\hat{\mathbf{S}}_i^M = \text{Water\_filling}(R, \hat{\mathbf{\Delta}}_i)$
    $\mathbf{S}_i^M = \hat{\mathbf{H}}_i^{-1}(\mathbf{F}_i\hat{\mathbf{S}}_i^M\mathbf{F}_i^{-1} - \mathbf{I})\hat{\mathbf{H}}_i^{-T}$
    return $\mathbf{S}_1^M, \mathbf{S}_2^M$
  end for
end while

Single Value Decomposition

Given the matrix $\mathbf{M}$ of dimension $m \times n$ the entries of which are real or complex numbers, then there exists a factorization of $\mathbf{M}$ which can be written as follows

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$$

where

- $\mathbf{U}$ is an $m \times m$ unitary matrix in the considered field
- $\mathbf{\Sigma}$ is a diagonal $m \times n$ matrix with non negative real numbers in the diagonal

- $\mathbf{V}$ is an $n \times n$ unitary matrix over the field considered, and $\mathbf{V}^T$ is the conjugate transpose of $\mathbf{V}$.

The minimum power for transmitting is this channel is obtained by evaluating each permutation π_m. For a given permutation, Algorithm 2 gives as result $P_i + P_j = \mathrm{tr}\left(\mathbf{S}_1^B + \mathbf{S}_2^B\right) = \mathrm{tr}\left(\mathbf{S}_1^M + \mathbf{S}_2^M\right)$. The minimum is taken over the results of all possible permutations. The overall cost of the channel is

$$c_{\text{MIMO-BC}}(i,j) = P_i + P_j.$$

4.2.1.6 MC Power Minimization

In the MC the same message has to be sent to both receivers at rate R_i. The optimization problem for minimizing the power can be formulated as follows

$$\begin{aligned} \underset{P_i}{\text{minimize}}\ & P_i \\ \text{subject to:}\ & \frac{1}{2}\log_2\left(1 + a_{11}\cdot P_i\right) \geq R_i \\ & \frac{1}{2}\log_2\left(1 + a_{12}\cdot P_i\right) \geq R_i \\ & P_i \geq 0. \end{aligned} \tag{4.40}$$

It can be easily shown that the minimum power required for u_1 is

$$P_i = \frac{2^{2R_i} - 1}{a_{11}}$$

while for u_2 is

$$P_i = \frac{2^{2R_i} - 1}{a_{12}}.$$

The cost of the channel $c_{\text{MC}}(i,i)$ is:

$$\begin{aligned} c_{\text{MC}}(i,j) =& \max\left\{\frac{2^{2R_i} - 1}{a_{11}}, \frac{2^{2R_i} - 1}{a_{12}}\right\} \\ =& \frac{2^{2R_-} - 1}{a_-}. \end{aligned} \tag{4.41}$$

4.2.1.7 OC Power Minimization

The optimization problem to be solved in the OC channel is

$$\begin{aligned} \underset{P_i,P_j}{\text{minimize}}\ & P_i + P_j \\ \text{subject to:}\ & \frac{1}{2}\log_2\left(1 + a_{11}\cdot P_i\right) \geq R_i \\ & \frac{1}{2}\log_2\left(1 + a_{22}\cdot P_j\right) \geq R_j \\ & P_i, P_j \geq 0, \end{aligned} \tag{4.42}$$

where the solution is

$$P_i = \frac{2^{2R_i} - 1}{a_{11}}$$

and

$$P_j = \frac{2^{2R_j} - 1}{a_{22}}.$$

The cost $c_{\text{OC}}(i,j)$ of a transmission over the OC is

$$c_{\text{OC}}(i,j) = \frac{2^{2R_i} - 1}{a_{11}} + \frac{2^{2R_j} - 1}{a_{22}}. \tag{4.43}$$

4.2.1.8 MISO-OC Power Minimization

The optimization problem in the MISO-OC channel to be solved is

$$\begin{aligned} \underset{P_i,P_j}{\text{minimize}}\quad & P_i + P_j \\ \text{subject to}\quad & \frac{1}{2}\log_2\left(1 + (a_{11} + a_{12})\cdot P_i\right) \geq R_i \\ & \frac{1}{2}\log_2\left(1 + a_{20}\cdot P_j\right) \geq R_j \\ & P_i, P_j \geq 0, \end{aligned} \tag{4.44}$$

where the solution is

$$P_i = \frac{2^{2R_i} - 1}{a_{11} + a_{12}}$$

and

$$P_j = \frac{2^{2R_j} - 1}{a_{20}}.$$

The cost $c_{\text{MISO-OC}}(i,j)$ for the MISO-OC is

$$c_{\text{MISO-OC}}(i,j) = \frac{2^{2R_i}-1}{a_{11}+a_{12}} + \frac{2^{2R_j}-1}{a_{20}}.$$

4.2.2 Channel Model Summary Table

In Table 4.7 the results of the optimization obtained for each channel model are summarized. The first column specifies the transmission channel, the second column specifies the cost of transmitting in that channel while the last column indicates the method we propose for calculating the channel cost.

Table 4.7 Transmission channel model costs

Channel model	$c_T(i,j)$	Method
GIN	$c_{\text{GIN}}(i,j) = \left(\frac{a_{12}(2^{2R_j}-1)\cdot\left[\frac{a_{21}}{a_{11}}\cdot(2^{2R_i}-1)+1\right]}{a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2R_j}-1)\cdot(1-2^{2R_i})} + 1 \right) \frac{(2^{2R_i}-1)}{a_{11}} + \frac{(2^{2R_j}-1)\cdot\left[\frac{a_{21}}{a_{11}}\cdot(2^{2R_i}-1)+1\right]}{a_{22}+\frac{a_{21}\cdot a_{12}}{a_{11}}\cdot(2^{2R_j}-1)\cdot(1-2^{2R_i})}$	analytically
GIC-WC	$c_{\text{GIC-WC}}(i,j) = P_i + P_j$	algorithm 1
GIC-C	$c_{\text{GIC-C}}(i,j) = P_i + P_j$	KKT conditions
BC	$c_{\text{BC}}(i,j) = \frac{2^{2R_+}-1}{a_+} + \frac{(2^{2R_-}-1)(1+a_-P_+)}{a_-}$	analytically
MIMO-BC	$c_{\text{MIMO-BC}}(i,j) = P_i + P_j$	algorithm 2
MC	$c_{\text{MC}}(i,j) = \frac{2^{2R_-}-1}{a_-}$	analytically
OC	$c_{\text{OC}}(i,j) = \frac{2^{2R_i}-1}{a_{11}} + \frac{2^{2R_j}-1}{a_{22}}$	analytically
MISO-OC	$c_{\text{MISO-OC}}(i,j) = \frac{2^{2R_i}-1}{a_{11}+a_{12}} + \frac{2^{2R_j}-1}{a_{20}}$	analytically

4.3 Placement Phase: Cache Allocation Optimization

In this section, we propose the optimization of the content placement, i.e. we determinate which is the content to be stored in the transmitters' caches such that during the delivery phase the average power is minimized. The novel aspect in this study is to

consider that each content not only has a popularity ranking, i.e each content has a different probability to be requested, but also has a transmission rate constraint and a one-shot delivery phase is considered. This supplementary assumption leads to a new case study. In fact, if only popularity ranking is considered then the best approach for filling up the caches is based on the expectation of the requests. However, it is not straightforward to conclude that the same consideration holds when data transmission rates have to be respected.

In the following, we define the system model and the notation used in the next chapter. From the results obtained in the previous section, we formulate the optimization problem for the three the different approaches

(i) non-cooperative approach,

(ii) limited cooperative approach,

(iii) full cooperative approach.

The three approaches are described more in detail in the following.

4.3.1 Non-Cooperative Approach

In Fig. 4.3 the active links for each transmission technique in the non-cooperative approach are represented.

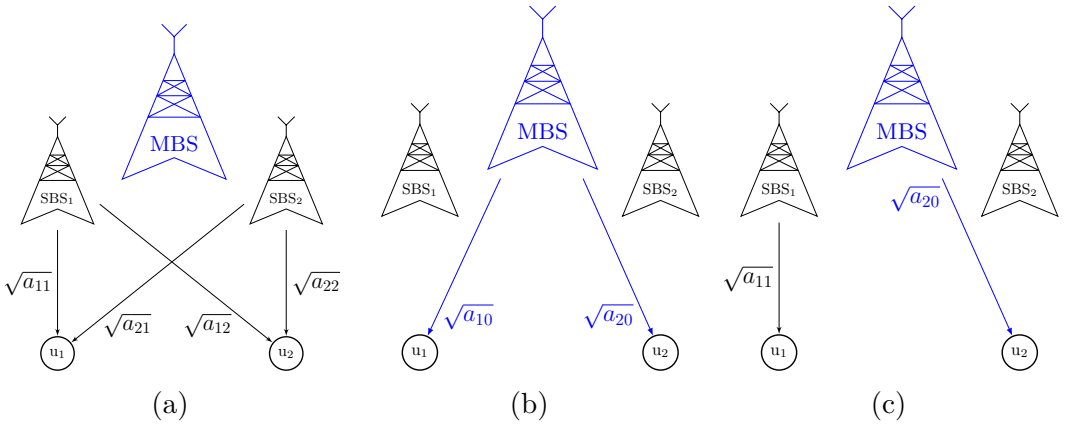


Fig. 4.3 System model for the non-cooperative approach: (a) interference as noise, (b) broadcast or multicast channel and (c) orthogonal channel

Fig. 4.3-(a) represents GIN transmissions from the SBSs, Fig. 4.3-(b) represents broadcast of multicast transmissions from the MBS while Fig. 4.3-(c) represents orthogonal transmissions.

4.3.1.1 Average Delivery Cost in Non-Cooperative Approach

Given the allocations vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, the expression of the average delivery cost in the non-cooperative approach can be written as follows

$$\begin{aligned} Q_{\text{nc}}(\mathbf{x}_1,\mathbf{x}_2) = \sum_{f_i\in\mathcal{F}} \sum_{\substack{f_j\in\mathcal{F} \\ f_i\neq f_j}} q_i q_j \Big\{ & c_{\text{GIN}}(i,j)\Big[\mathbb{1}_\alpha x_{i,1}x_{j,2}\Big] + c_{\text{OC}}(i,j)\Big[x_{i,1}\bar{x}_{j,2} + \bar{x}_{i,1}x_{j,2}\Big] \\ & + c_{\text{BC}}^{\text{MBS}}(i,j)\Big[\bar{\mathbb{1}}_\alpha x_{i,1}x_{j,2} + \bar{x}_{i,1}\bar{x}_{j,2}\Big]\Big\} + \sum_{\substack{f_j\in\mathcal{F} \\ f_i=f_j}} q_i^2 \Big\{ c_{\text{GIN}}(i,j)\Big[\mathbb{1}_\alpha x_{i,1}x_{i,2}\Big] \\ & + c_{\text{MC}}^{\text{MBS}}(i,j)[\bar{x}_{i,1}\bar{x}_{i,2} + \bar{\mathbb{1}}_\alpha x_{i,1}x_{i,2}\Big] + c_{\text{OC}}(i,j)\Big[\bar{x}_{i,2}x_{i,1} + \bar{x}_{i,1}x_{i,2}\Big]\Big\}. \end{aligned} \tag{4.45}$$

In eq. (4.45), the average is calculated over every possible pair of file requested in the library and weighted by the probability that the pair of files is requested, i.e. $q_i \cdot q_j$. The first addend of the expression is the average cost when different files are delivered while the second addend calculates the average cost when the same file is delivered by the transmitters.

We clarify that for a fixed cache allocation ($\mathbf{x}_1$, $\mathbf{x}_2$), the cost for transmitting a pair of files (f_i, f_j) is the cost of the transmission of the channel selected according to Table 4.1 and Table 4.2. For each request only one transmission cost contributes in the average. For instance, if the file f_i requested by u_1 is not present in the cache of SBS_1, i.e. $x_{i,1} = 0$, $\bar{x}_{i,1} = 1$, and f_j requested by u_2 is not present in the cache of SBS_2, i.e. $x_{j,2} = 0$, $\bar{x}_{j,2} = 1$, then the cost of the transmission is $c_{\text{BC}}^{MBS}(i,j)$ given by the broadcast channel and all the other costs are zero.

The expression on eq. (4.45) represents the objection function of the optimization problem in which we are interested. In particular, we want to find the vectors allocation $\mathbf{x}_1$ and $\mathbf{x}_2$ for which eq. (4.45) assumes the minimum value.

4.3.1.2 Non-Cooperative Cache Placement Optimization

The optimization of the cache content placement in the non-cooperative approach aims to determinate the allocation vectors which minimize the average power consumption in the delivery phase. The objective function to minimize is given by the function (4.45) which has to be optimized over the allocation vectors $\mathbf{x}_1, \mathbf{x}_2$ and the power dimension. Denoting the capacity of a point to point Gaussian channel with signal-to-noise ratio x as

$$C(x) = \frac{1}{2}\log_2(1+x),$$

the proposed optimization problem for the non-cooperative approach can be written as

$$\underset{P,P,\mathbf{x}_1,\mathbf{x}_2}{\text{minimize}}\ Q_{nc}(\mathbf{x}_1,\mathbf{x}_2), \tag{4.46}$$

$$\text{subject to}: \sum x_{in} \leq M_n, \qquad n=\{1,2\}, \tag{4.47}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_- \leq C\left(\frac{a_{-0}\cdot P_-}{1+a_{-0}P_+}\right), \qquad i\neq j, \tag{4.48}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_+ \leq C\left(a_{+0}\cdot P_+\right), \qquad i\neq j, \tag{4.49}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_i \leq C\left(a_{11}\cdot P_i\right), \qquad i=j, \tag{4.50}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_i \leq C\left(a_{22}\cdot P_j\right), \qquad i=j, \tag{4.51}$$

$$\bar{x}_{j,2}\cdot x_{i,1}\cdot R_i \leq C(a_{11}\cdot P_i), \tag{4.52}$$

$$\bar{x}_{j,2}\cdot x_{i,1}\cdot R_j \leq C(a_{22}\cdot P_j), \tag{4.53}$$

$$\bar{x}_{i,1}\cdot x_{j,2}\cdot R_i \leq C(a_{11}\cdot P_i), \tag{4.54}$$

$$\bar{x}_{i,1}\cdot x_{j,2}\cdot R_j \leq C(a_{22}\cdot P_j), \tag{4.55}$$

$$\mathbb{1}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_i \leq C\left(\frac{P_i}{1+a_{12}P_j}\right), \tag{4.56}$$

$$\mathbb{1}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_j \leq C\left(\frac{P_j}{1+a_{21}\cdot P_i}\right), \tag{4.57}$$

$$\bar{\mathbb{1}}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_- \leq C\left(\frac{a_{-0}\cdot P_-}{1+a_{-0}\cdot P_+}\right), \qquad i\neq j, \tag{4.58}$$

$$\bar{\mathbb{1}}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_+ \leq C\left(a_{+0}\cdot P_+\right), \qquad i\neq j, \tag{4.59}$$

$$\bar{\mathbb{1}}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_i \leq C\left(a_{11}\cdot P_i\right), \qquad i=j, \tag{4.60}$$

$$\bar{\mathbb{1}}_\alpha\cdot x_{i,1}\cdot x_{j,2}\cdot R_i \leq C\left(a_{22}\cdot P_j\right), \qquad i=j, \tag{4.61}$$

$$P_i>0, \tag{4.62}$$

$$P_j>0, \tag{4.63}$$

$$x_{i,n}\in\{0,1\}, \qquad n=\{1,2\}, \tag{4.64}$$

$$f_i, f_j \in \mathcal{F}. \tag{4.65}$$

Inequality in (4.47) ensures that the total amount of data stored in a cache will not exceed its size, (4.48)-(4.49) denote the capacity region of the Gaussian broadcast channel, (4.50)-(4.51) denote the capacity region of the multicast channel, (4.52)-(4.55) represent the capacity region of the orthogonal channels. Instead, conditions (4.58)-(4.59) denote the capacity region of the Gaussian broadcast channel and (4.60)-(4.61) the capacity region of the multicast channel when does not exist a solution for

transmitting in the GIN. Hence, these channels are considered only when the SBSs are not capable to deliver the files at the requested rates. Conditions (4.56)-(4.57) represent the constraints for the capacity in the GIN channel.

4.3.2 Limited Cooperative Approach

In Fig. 4.4 the active links for each transmission technique in the limited cooperative approach are represented.

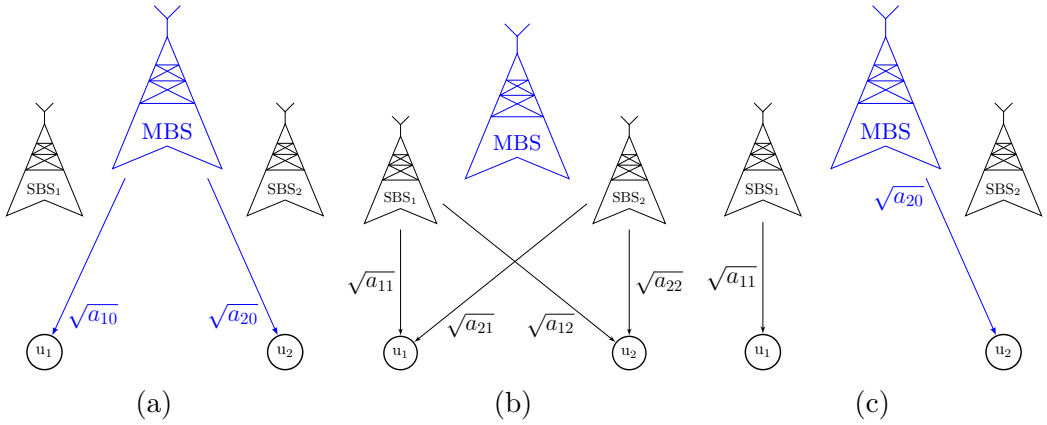


Fig. 4.4 System models for the limited cooperative approach: (a) broadcast or multicast channel from MBSs, (b) interference channel with common information or interference channel without common information (c) orthogonal channel

4.3.2.1 Average Delivery Cost in limited Cooperative Approach

The expected power cost in the cooperative case $Q_{\text{l-c}}(\mathbf{x}_1, \mathbf{x}_2)$ can be written as

$$
\begin{aligned}
Q_{\text{l-c}}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{f_i \in \mathcal{F}} \sum_{\substack{f_j \in \mathcal{F} \\ f_i \neq f_j}} q_i \cdot q_j \cdot \Big[& \bar{x}_{i,1} \cdot \bar{x}_{j,2} \cdot c_{\text{BC}} + \\
x_{i,1} \cdot x_{j,2} \cdot c_{\text{GIC-WC}} + \bar{x}_{j,2} \cdot x_{i,1} \cdot c_{\text{OC}} + \bar{x}_{i,1} \cdot x_{j,2} \cdot c_{\text{OC}} \Big] + & \\
\sum_{f_i \in \mathcal{F}} q_{ii}^2 \cdot \Big[& \bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot c_{\text{MC}} + \cdot x_{i,1} \cdot x_{i,2} \cdot c_{\text{GIC-C}} + \\
& \bar{x}_{i,2} \cdot x_{i,1} \cdot c_{\text{BC}} + \bar{x}_{i,1}) \cdot x_{i,2} \cdot c_{\text{OC}} \Big],
\end{aligned}
\tag{4.66}
$$

where c_{BC}, $c_{\text{GIC-WC}}$, $c_{\text{GIC-C}}$, c_{OC}, c_{MC}, denote the minimum power consumption $(P_i + P_j)$ required in the broadcast, Gaussian interference channel without common information, Gaussian interference channel with common information, orthogonal and multicast channels, respectively.

The expression on eq. (4.66) represents the objective function of the optimization problem we are interested in. In particular, we want to find the vectors allocation $\mathbf{x}_1$ and $\mathbf{x}_2$ for which eq. (4.66) assumes the minimum value.

4.3.2.2 Limited Cooperative Cache Placement Optimization

The minimization problem for the limited cooperative case can be formulated as follows

$$\underset{P_i,P_j,\mathbf{x}_1,\mathbf{x}_2}{\text{minimize}}\ Q_{\text{l-c}}(\mathbf{x}_1,\mathbf{x}_2), \tag{4.67}$$

$$\text{s.t.} \sum x_{in} \leq M_n, \qquad n=\{1,2\}, \tag{4.68}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_- \leq C\left(\frac{a_{-0}\cdot P_-}{1+a_{-0}\cdot P_+}\right), \qquad i\neq j, \tag{4.69}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_+ \leq C\left(a_{+0}\cdot P_+\right), \qquad i\neq j, \tag{4.70}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_i \leq C\left(a_{10}\cdot P_i\right), \qquad i=j, \tag{4.71}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{j,2}\cdot R_i \leq C\left(a_{20}\cdot P_j\right), \qquad i=j, \tag{4.72}$$

$$\bar{x}_{j,2}\cdot x_{i,1}\cdot R_i \leq C(a_{11}\cdot P_i), \tag{4.73}$$

$$\bar{x}_{j,2}\cdot x_{i,1}\cdot R_j \leq C(a_{20}\cdot P_j), \tag{4.74}$$

$$\bar{x}_{i,1}\cdot x_{j,2}\cdot R_i \leq C(a_{10}\cdot P_i), \tag{4.75}$$

$$\bar{x}_{i,1}\cdot x_{j,2}\cdot R_j \leq C(a_{22}\cdot P_j), \tag{4.76}$$

$$x_{i,1}\cdot x_{j,2}\cdot R_i \leq \rho_1, \qquad i\neq j, \tag{4.77}$$

$$x_{i,1}\cdot x_{j,2}\cdot R_j \leq \rho_2, \qquad i\neq j, \tag{4.78}$$

$$x_{i,1}\cdot x_{j,2}\cdot (R_i+R_j) \leq \rho_{12}, \qquad i\neq j, \tag{4.79}$$

$$x_{i,1}\cdot x_{j,2}\cdot (2R_i+R_j) \leq \rho_{10}, \qquad i\neq j, \tag{4.80}$$

$$x_{i,1}\cdot x_{j,2}\cdot (R_i+2R_j) \leq \rho_{20}, \qquad i\neq j, \tag{4.81}$$

$$x_{i,1}\cdot x_{j,2}\cdot R_i \leq C\left(\left(\sqrt{\tilde{P}_i}+\sqrt{a_{12}\tilde{P}_j}\right)^2\right), \qquad i=j, \tag{4.82}$$

$$x_{i,1}\cdot x_{j,2}\cdot R_i \leq C\left(\left(\sqrt{\tilde{P}_j}+\sqrt{a_{21}\tilde{P}_i}\right)^2\right), \qquad i=j, \tag{4.83}$$

$$P_i > 0, \tag{4.84}$$

$$P_j > 0, \tag{4.85}$$

$$x_{i,n}\in\{0,1\}, \qquad n=\{1,2\}. \tag{4.86}$$

Inequality in (4.68) ensures that the total amount of data stored in a cache will not exceed its size, (4.69)-(4.70) denote the capacity region of the Gaussian broadcast

channel, (4.71)-(4.72) denote the capacity region of the multicast channel, (4.73)-(4.76) represent the conditions for the case of orthogonal channels, (4.77)-(4.81) are the Han-Kobayashi constraints and define an achievable rate of the GIC-WC , (4.82)-(4.83) define an achievable rate region of the GIC-C when both transmitters have to transmit the same file and (4.84)-(4.85) are imposed to guarantee the non-negativity of the transmit power while (4.86) represent the caching choice and account for the discrete nature of the optimization variable. Note that the power variables in constraints (4.77)-(4.83) are related to the equivalent system ($\tilde{P}_i$, $\tilde{P}_j$). Since we are interested in evaluating the physical power consumption, then the power values in the calculation of the cost has to be scaled as $P_i = \tilde{P}_i/a_{11}$ and $P_j = \tilde{P}_j/a_{22}$.

4.3.3 Full Cooperative Approach

In Fig. 4.3 are illustrated the active links for each transmission delivery channel in the full cooperative approach.

4.3.3.1 Average Delivery Cost in the full Cooperative Approach

Given the allocations vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, the expression of the average delivery power cost $Q_{\mathrm{c}}(\mathbf{x}_1, \mathbf{x}_2)$ in the cooperative approach can be written as follows

$$
\begin{aligned}
Q_{\mathrm{c}}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{f_i \in \mathcal{F}} \sum_{\substack{f_j \in \mathcal{F} \\ f_i \neq f_j}} q_i q_j \Big\{ & c_{\text{GIC-C}}(i,j) \Big[\sum_{t=i,j} \sum_{n=1,2} x_{t,n} \cdot x_{\bar{t},n} \cdot x_{t,\bar{n}} \cdot \bar{x}_{\bar{t},\bar{n}} \Big] + c_{\mathrm{BC}}^{\mathrm{SBS}}(i,j) \\
& \Big[\sum_{n=1,2} x_{i,n} \cdot x_{j,n} \cdot \bar{x}_{i,\bar{n}} \cdot \bar{x}_{j,\bar{n}} \Big] + c_{\mathrm{BC}}^{\mathrm{MBS}}(i,j) \Big[\bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot \bar{x}_{j,1} \cdot \bar{x}_{j,2} \Big] + c_{\mathrm{OC}}(i,j) \Big[\sum_{t=i,j} \sum_{n=1,2} x_{tn} \\
& \bar{x}_{\bar{t},n} \cdot \bar{x}_{t,\bar{n}} \cdot \bar{x}_{\bar{t},\bar{n}} \Big] + c_{\text{GIC-WC}}(i,j) \Big[\sum_{n=1,2} x_{i,n} \cdot x_{j,\bar{n}} \cdot \bar{x}_{i,\bar{n}} \cdot \bar{x}_{j,n} \Big] + c_{\text{ MISO-OC}}(i,j) \Big[\sum_{t=i,j} x_{t,1} \\
& x_{t,2} \cdot \bar{x}_{\bar{t},1} \cdot \bar{x}_{\bar{t},2} \Big] + c_{\text{ MIMO-BC}}(i,j) \Big[x_{i,1} \cdot x_{i,2} \cdot x_{j,1} \cdot x_{j,2} \Big] \Big\} + \sum_{\substack{f_j \in \mathcal{F} \\ f_i = f_j}} q_i^2 \Big\{ c_{\mathrm{MC}}^{\mathrm{MBS}}(i,j) [\bar{x}_{i,1} \cdot \bar{x}_{i,2}] + \\
& c_{\mathrm{MC}}^{\mathrm{SBS}} [x_{i,1} \cdot \bar{x}_{i,2} + x_{i,2} \cdot \bar{x}_{i,1}] + c_{\text{GIC-C}}(i,j) [x_{i,1} \cdot x_{i,2}] \Big\},
\end{aligned}
\tag{4.87}
$$

where we recall that $c_T(i,j)$ is the cost for delivering a pair of files over the channel T, which was calculated for every channel model in subsection 4.2.1, and $x_{i,n}$ is the binary indicator which denotes the presence or not of f_i in the cache of SBS_n.
The expression in eq. (4.87) represents the average power consumed for serving a random request for a fixed cache allocation ($\mathbf{x}_1$, $\mathbf{x}_2$). The average is calculated over

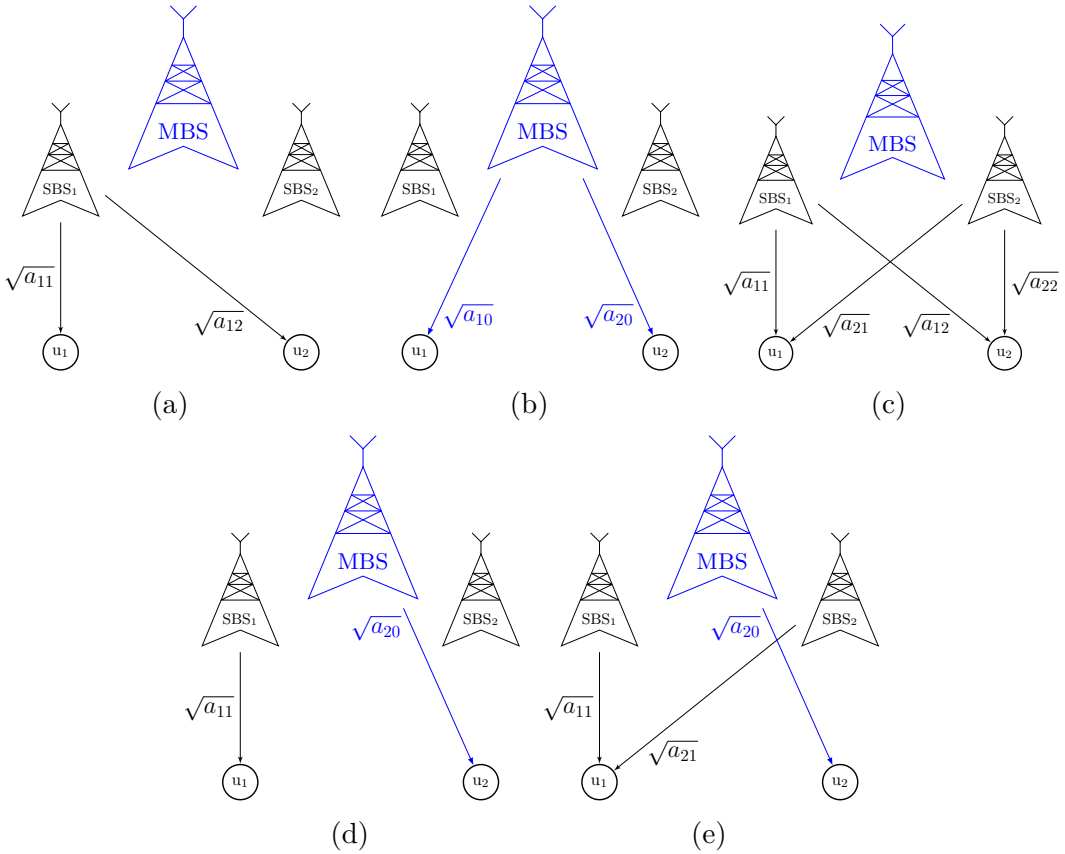


Fig. 4.5 System models for the full cooperative approach: (a) broadcast or multicast channel from SBSs, (b) broadcast or multicast channel from MBSs, (c) interference channel with common information or interference channel without common information or multiple input multiple output broadcast chanel (d) orthogonal channel and (e) multiple input single ouput-ortoghonal channel

every possible pair of files requested from the library and weighted by the probability that the pair of files is requested, i.e. $q_i \cdot q_j$. The first addend of the expression is the average cost when different files are delivered while the second addend is the average cost when the same file is delivered by the transmitters.

Given a pair of requested files and a fixed cache allocation, only one transmission has an impact on the average. For example, if the file f_i requested by u_1 is not present in the caches of the SBSs, i.e. $x_{i,1} = 0$ $x_{i,2} = 0$ or equally $\bar{x}_{i,1} = 1$ $\bar{x}_{i,2} = 1$, and f_j requested by u_2 is not present in the caches of SBSs, i.e. $x_{j,1} = 0$ $x_{j,2} = 0$ or equally $\bar{x}_{i,2} = 1$ $\bar{x}_{j,2} = 1$, then the transmission technique will be the broadcast channel from the MBS at cost $c_{BC}^{MBS}(i,j)$. In other words, the cost of the channel T can be *active* or

inactive. For example, the cost of the GIC-C is inactive when

$$c_{\text{GIC-C}}(i,j)\Big[\sum_{t=i,j}\sum_{n=1,2} x_{t,n}\cdot x_{\bar{t},n}\cdot x_{t,\bar{n}}\cdot \bar{x}_{\bar{t},\bar{n}}\Big]=0$$

and that is when

$$\sum_{t=i,j}\sum_{n=1,2} x_{t,n}\cdot x_{\bar{t},n}\cdot x_{t,\bar{n}}\cdot \bar{x}_{\bar{t},\bar{n}}=0,$$

instead, when

$$\sum_{t=i,j}\sum_{n=1,2} x_{t,n}\cdot x_{\bar{t},n}\cdot x_{t,\bar{n}}\cdot \bar{x}_{\bar{t},\bar{n}}=1$$

the cost of the GIC-C is active. Thus, the activation/deactivation of a cost depends on which content is cached.

The expression on eq. (4.87) represents the objective function of the optimization problem in which we are interested. In particular, we want to find the vectors allocation $\mathbf{x}_1$ and $\mathbf{x}_2$ for which eq. (4.87) assumes the minimum value.

4.3.3.2 Full Cooperative Cache Placement Optimization

The cache content placement optimization aims at determining the vector allocation which minimizes the average power consumption in the delivery phase. The objective function to minimize is given by the function (4.87) which has to be optimized over the allocation vectors $\mathbf{x}_1, \mathbf{x}_2$ and the power dimension.

The proposed optimization problem for the full cooperative approach can be written as

$$\underset{P_i,P_j,\mathbf{x}_1,\mathbf{x}_2}{\text{minimize}}\ Q_{\text{c}}(\mathbf{x}_1,\mathbf{x}_2), \tag{4.88}$$

$$\text{subject to: } \sum_{f_i\in\mathcal{F}} x_{in}\le M_n, \qquad n=\{1,2\}, \tag{4.89}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{i,2}\cdot\bar{x}_{j,1}\cdot\bar{x}_{j,2}\cdot R_{-}\le C(a_{-0}P_{-}) \tag{4.90}$$

$$\bar{x}_{i,1}\cdot\bar{x}_{i,2}\cdot\bar{x}_{j,1}\cdot\bar{x}_{j,2}\cdot R_{+}\le C\left(a_{+0}\cdot P_{+}\right), \tag{4.91}$$

$$x_{l,w}\cdot\bar{x}_{l,\bar{w}}\cdot\bar{x}_{\bar{l},w}\cdot\bar{x}_{\bar{l},\bar{w}}\cdot R_{\bar{l}}\le C\left(a_{\bar{w}0}P_{\bar{l}}\right), \qquad w=\{1,2\}, l=\{i,j\}, \tag{4.92}$$

$$x_{l,w}\cdot\bar{x}_{l,\bar{w}}\cdot\bar{x}_{\bar{l},w}\cdot\bar{x}_{\bar{l},\bar{w}}\cdot R_{l}\le C\left(a_{wk}P_{l}\right), \qquad \text{if } l=i \text{ then } k=1 \text{ else } k=2, \tag{4.93}$$

$$x_{i,r} \cdot x_{j,\bar{r}} \cdot \bar{x}_{i,\bar{r}} \cdot \bar{x}_{j,r} \cdot R_i \leq \rho_1(c_{r\bar{r}}, c_{\bar{r}r}, \tilde{P}_i, \tilde{P}_j), \qquad r = \{1,2\}, \tag{4.94}$$

$$x_{i,r} \cdot x_{j,\bar{r}} \cdot \bar{x}_{i,\bar{r}} \cdot \bar{x}_{j,r} \cdot R_j \leq \rho_2(c_{r\bar{r}}, c_{\bar{r}r}, \tilde{P}_i, \tilde{P}_j) \tag{4.95}$$

$$x_{i,r} \cdot x_{j,\bar{r}} \cdot \bar{x}_{i,\bar{r}} \cdot \bar{x}_{j,r} \cdot (R_i + R_j) \leq \rho_{12}(c_{r\bar{r}}, c_{\bar{r}r}, \tilde{P}_i, \tilde{P}_j) \tag{4.96}$$

$$x_{i,r} \cdot x_{j,\bar{r}} \cdot \bar{x}_{i,\bar{r}} \cdot \bar{x}_{j,r} \cdot (2R_i + R_j) \leq \rho_{10}(c_{r\bar{r}}, c_{\bar{r}r}, \tilde{P}_i, \tilde{P}_j) \tag{4.97}$$

$$x_{i,r} \cdot x_{j,\bar{r}} \cdot \bar{x}_{i,\bar{r}} \cdot \bar{x}_{j,r} \cdot (R_i + 2R_j) \leq \rho_{20}(c_{r\bar{r}}, c_{\bar{r}r}, \tilde{P}_i, \tilde{P}_j) \tag{4.98}$$

$$x_{i,1} \cdot x_{i,2} \cdot x_{j,1} \cdot x_{j,2} \cdot R_i \leq C(\pi_m, \mathbf{H}_1, \mathbf{H}_2, \mathbf{S}_1, \mathbf{S}_2), \tag{4.99}$$

$$x_{i,1} \cdot x_{i,2} \cdot x_{j,1} \cdot x_{j,2} \cdot R_j \leq C(\pi_m, \mathbf{H}_1, \mathbf{H}_2, \mathbf{S}_1, \mathbf{S}_2) \tag{4.100}$$

$$x_{i,z} \cdot x_{j,z} \cdot \bar{x}_{i,\bar{z}} \cdot \bar{x}_{j,\bar{z}} \cdot R_- \leq C\left(a_{-z}P_-/(1 + a_{-z}P_+)\right), \qquad z = \{1,2\}, \tag{4.101}$$

$$x_{i,z} \cdot x_{j,z} \cdot \bar{x}_{i,\bar{z}} \cdot \bar{x}_{j,\bar{z}} \cdot R_+ \leq C\left(a_{+z}P_+\right), \tag{4.102}$$

$$x_{i,1} \cdot x_{i,2} \cdot \bar{x}_{j,1} \cdot \bar{x}_{j,2} \cdot R_i \leq C\left((a_{11} + a_{12}) \cdot P_i\right), \tag{4.103}$$

$$x_{i,1} \cdot x_{i,2} \cdot \bar{x}_{j,1} \cdot \bar{x}_{j,2} \cdot R_j \leq C(a_{20} \cdot P_j), \tag{4.104}$$

$$x_{j,1} \cdot x_{j,2} \cdot \bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot R_i \leq C(a_{10} \cdot P_i), \tag{4.105}$$

$$x_{j,1} \cdot x_{j,2} \cdot \bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot R_j \leq C\left((a_{22} + a_{21}) \cdot P_j\right), \tag{4.106}$$

$$x_{ip} \cdot \bar{x}_{i\bar{p}} \cdot R_i \leq C(a_{pp} \cdot P_i), p = \{1,2\}, \tag{4.107}$$

$$x_{i,p} \cdot \bar{x}_{i,\bar{p}} \cdot R_i \leq C(a_{\bar{p}p} \cdot P_j, \tag{4.108}$$

$$\bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot \bar{x}_{j,1} \cdot \bar{x}_{j,2} \cdot R_i \leq C\left(a_{10} \cdot P_i\right) \qquad i = j, \tag{4.109}$$

$$\bar{x}_{i,1} \cdot \bar{x}_{i,2} \cdot \bar{x}_{j,1} \cdot \bar{x}_{j,2} \cdot R_j \leq C\left(a_{20} \cdot P_j\right), \qquad i = j, \tag{4.110}$$

$$x_{i,1} \cdot x_{i,2} \cdot R_i \leq C\left(\left(\sqrt{\tilde{P}_i} + \sqrt{c_{21} \cdot \tilde{P}_j}\right)^2\right), \qquad i = j, \tag{4.111}$$

$$x_{i,1} \cdot x_{i,2} \cdot R_j \leq C\left(\left(\sqrt{\tilde{P}_j} + \sqrt{c_{12} \cdot \tilde{P}_i}\right)^2\right), \qquad i = j, \tag{4.112}$$

$$P_i \geq 0, P_j \geq 0, \tag{4.113}$$

$$P_- \geq 0, P_+ \geq 0, \tag{4.114}$$

$$x_{i,n} \in \{0,1\}, \tag{4.115}$$

$$x_{j,n} \in \{0,1\}, \tag{4.116}$$

$$f_i, f_j \in \mathcal{F}. \tag{4.117}$$

Inequality (4.89) ensures that the total amount of data stored in a cache does not exceed its size, (4.90)-(4.91) denote the capacity region of the Gaussian broadcast channel of the MBS, while (4.92)-(4.93) represent the conditions for the case of orthogonal channels, (4.94)-(4.98) are the Han-Kobayashi conditions and define the best known achievable rate of the GIC-WC, (4.99)-(4.100) define the capacity region for the MIMO broadcast channel, (4.101)-(4.102) denote the capacity region for the broadcast channel of SBS_n, (4.103)-(4.104) and (4.105)-(4.106) define the capacity region of the orthogonal channel when the SBSs implement a MISO-OC channel to serve u_1 (u_2) while u_2 (u_1) is served by the MBS, (4.107)-(4.108) denote the capacity region of the multicast channel of SBS_n, (4.111)-(4.112) define an achievable rate region of the GIC-C when both SBSs have to transmit the same file, (4.109)-(4.110) define the capacity region of the multicast channel of the MBS, (4.113) is imposed to guarantee the non negativity of the transmit power, while (4.115) represents the caching choice and accounts for the discrete nature of the optimization variable. Note that the power variables in constraints (4.94)-(4.98) and (4.111)-(4.112) refer to the normalized channel model ($\tilde{P}_i$, $\tilde{P}_j$). Since we are interested in evaluating the physical power consumption, the power values in the calculation of the cost have to be scaled as $P_i = \tilde{P}_i/a_{11}$ and $P_j = \tilde{P}_j/a_{22}$.

4.3.4 Algorithm Implementation and Complexity

The optimization problems defined for the non-cooperative, limited cooperative and full cooperative approaches are non-convex. The optimal cache allocation for each approach can be derived by minimizing the cost function in (4.46),(4.67) and (4.87), independently along the power dimension and the cache allocation dimension.

First, the minimum average power for a given cache allocation is calculated. For each of the N^2 possible requests the power has to be minimized. Each request, according to whether the file is present or not in a cache, requires to solve one of the power minimization sub-problems presented in Section 4.2.1. The optimal cache allocation is obtained by exhaustive search (greedy algorithm) within the set of possible allocations.

Let us clarify what just stated with an example.

Example: Let us consider the cooperative case and let us assume to have $N = 3$ files f_1, f_2, f_3 and memory size $M = 2$. First we fix a memory allocation in both SBSs, e.g. $\mathbf{x}_1 = [1 \quad 1 \quad 0]$ and $\mathbf{x}_2 = [1 \quad 1 \quad 0]$. For this memory allocation, we calculate the cost function $Q(\mathbf{x}_1, \mathbf{x}_2)$. That is, for each of the nine possible requests, the corresponding minimization sub-problem (see Section III) is solved. Note that for each request only the corresponding cost is activated in each equation (4.46),(4.67) and (4.87). For

instance, for the full cooperative approach if f_1 is requested by u_1 and f_2 is requested by u_2, then the corresponding channel is a broadcast MIMO and all the other terms in the sum are zero.

Complexity

The complexity for minimizing the average power for a given cache allocation is $O(N^2)$ while the cost for finding the optimal cache allocation is $O(N!)$.

Although such algorithm is useful to evaluate the performance limit of the considered setup, its complexity makes it unsuited for a practical application when the number of files is large. Therefore, we propose in the following a sub-optimal algorithm with reduced complexity.

4.3.5 Sub-Optimal Algorithm

In the following, we propose low-complexity algorithm for finding a sub-optimal solution of the cache allocation problem. Note that the sub-optimal solution given by the algorithm may coincide with the optimal solution in some cases.

Let us analyze the expression of the average power consumption in the full cooperative approach given by eq.(4.87). The cost of each channel depends in a non-linear way on the channel coefficients a_{nm} and on the rate of both requested files R_i, R_j. It also depends linearly on the probabilities of the files requests q_i, q_j. Let us now consider genie-aided receivers having access to the message destined to the other user. In this case, the interference can be entirely removed and each user sees an interference-free channel. In such channel, the cost of transmitting f_i is equal to its popularity ranking q_i times a term proportional to 2^{R_i}. Since we consider a setup in which the SBS-user channel is on average better than that between MBS-user one, having files with higher cost transmitted by SBSs rather than the MBS minimizes the overall power expenditure.
We define the utility function

$$w_k(R_k, q_j) = 2^{2R_k} \cdot q_k$$

associated to f_k.

The proposed algorithm, described in Algorithm 3, caches the files with highest $w(R, q)$. The complexity of the algorithm grows linearly with N. In the result subsec-

Algorithm 3 $[\mathbf{x}_1 \quad \mathbf{x}_2]$ =low_complexity$(\mathbf{R}, \mathbf{q})$

$\mathbf{x}_1 = \mathbf{0}$ and $\mathbf{x}_2 = \mathbf{0}$ ▷ $\mathbf{0}$ is an N-dimensional all-zero vector
$W = \{w_1, \ldots, w_k, \ldots, w_N\}$ ▷ $w_k = 2^{2R_k} \cdot q_k$ is the k^{th} element of the set
for $l = 1, \ldots, M$ **do**
    $w_m = \arg\max_l \{w_l\}$
    $x_{m,1} = 1$
    $W = W \setminus \{w_m\}$
    $w_m = \arg\max_l \{w_l\}$
    $x_{m,2} = 1$
    $W = W \setminus \{w_m\}$
end for
return $[\mathbf{x}_1 \quad \mathbf{x}_2]$

tion, we compare the performance of the proposed algorithm with the optimal one and with two benchmarks.

4.4 Caching Power Optimization Analysis

In this section, we analyze the non-cooperative, limited cooperative and full cooperative approaches in the following three scenarios

(i) *static scenario*, we also refer this scenario as the additive white Gaussian noise (AWGN) case, where all channel coefficients are fixed so no fading is present,

(ii) *mobile transmitters scenario* where the transmitter-user links experience fading while the master node-user links are AWGN channels. This models a heterogeneous satellite network in which two LEO satellites act as transmitters while a GEO satellite acts as master node. Due to the LEOs orbital motion and the presence of obstacles such as tall buildings fading is present in the LEO-user channels[3] while the link to the GEO, assumed to be line of sight, does not change significantly over time, which is the case in clear-sky conditions. This setup also models a hybrid network with a GEO satellite acting as master node while hovering UAVs play the role of transmitters. Also in this case, the UAV-user links are affected by fading due to UAVs motion [49],

(iii) *mobile user scenario* where each link is affected by an independent block fading process.

[3]In practice the equivalent of an SBS (MBS) would be a LEO (GEO) satellite together with the respective ground station, where the cache would be physically located. This is because on-board storage in satellites is not common practice.

4.4.1 Preliminary Results for AWGN

In this subsection, we study the optimization for a problem of small dimension in order to have an insight on the problems. Then, in the next subsection, we extend the problem by introducing more variables in the design and give a deeper analysis.

Let us start comparing the non-cooperative and limited cooperative approaches. We assume a library of carinality $N = 3$ and each SBS has a memory of size $M_n = 2$. We consider that each file has a probability to be requested and a rate constraint given by the values on Table 4.8.

Table 4.8 Rate and popularity rankings per-file.

$\mathcal{F}$	R_i [bits/sec/Hz]	q_i
f_1	1.20	0.15
f_2	0.40	0.35
f_3	0.60	0.50

We consider that the channel coefficients are fixed. Powers are exponentially distributed with mean values $\mathbb{E}[a_{11}] = 1$, $\mathbb{E}[a_{22}] = 1$, $\mathbb{E}[a_{01}] = \mathbb{E}[a_{20}] = 0.01$ while $\mathbb{E}[c_{12}] = \mathbb{E}[c_{21}]$ take values in $[0, 1]$.

In Fig. 4.6 the minimum average power cost $Q(\mathbf{x}_1, \mathbf{x}_2)$ is plotted versus the interference coefficients (assumed to be the same for both users) in an AWGN environment for the non-cooperative and limited cooperative approaches. The limited cooperative approach is indicated with the blue curve while the non-cooperative approach with the red curve. Note that the optimization is done for each interference level. This means that the problem described in Section 4.3.1.2 and Section 4.3.1.2 were solved for each level of mutual interference.

The letter "A" indicates that the optimal file caching for that specific interference level in the two SBSs are $\mathcal{M}_1 = \{f_1, f_3\}$ and $\mathcal{M}_2 = \{f_1, f_3\}$ while the letter "B" indicates the storage solution $\mathcal{M}_1 = \{f_2, f_3\}$ $\mathcal{M}_2 = \{f_2, f_3\}$.

Two conclusions can be drawn out from this results. First, we observe that cooperation leads not only to counteract the mutual interference but also to decrease the average power consumption for delivering cached content. The second conclusion is that memorizing the most popular files is not always the most convenient strategy when transmitters have to respect rate requirements. As can be seen, the cache allocation denoted with "A" represents the best solution and contains f_3 which has lower probability to be requested with respect to f_2 which is not cached.

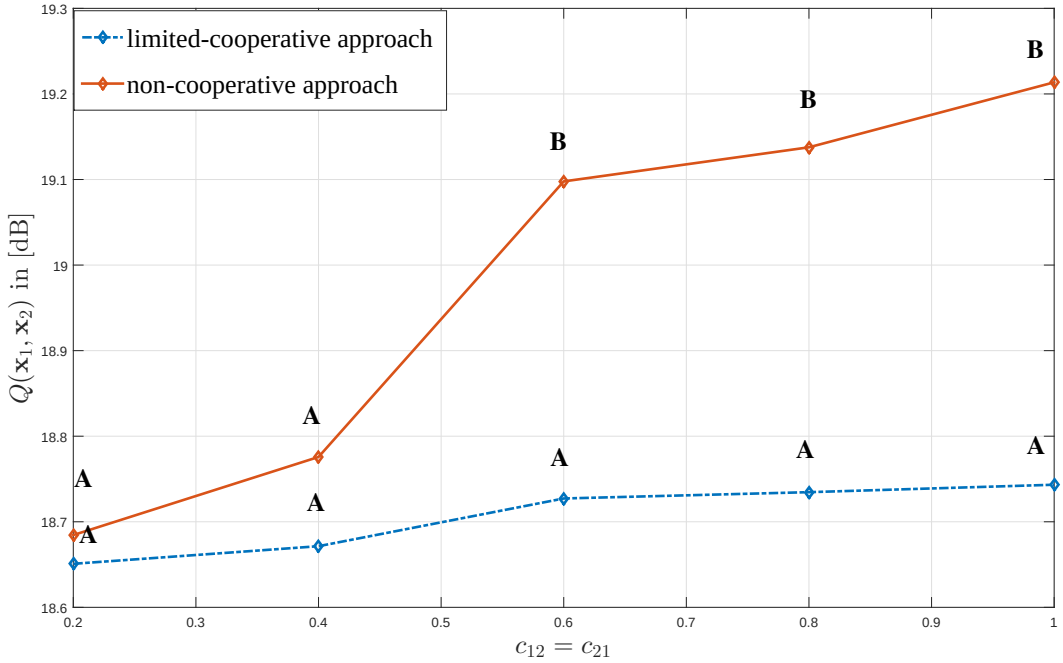


Fig. 4.6 Minimum average power cost in dB in the AWGN case plotted against the interference coefficient. The optimization is done for each average interference coefficient expressed in the standard form. In the simulation we set $a_{10} = a_{20} = 0.01$. "A" points represent the memory allocation: $\mathcal{M}_1 = \{f_1, f_3\}, \mathcal{M}_2 = \{f_1, f_3\}$ while "B" points represent $\mathcal{M}_1 = \{f_2, f_3\}, \mathcal{M}_2 = \{f_2, f_3\}$.

4.4.2 Simulation Parameters in Fading Scenario

We model the channel coefficients in the links affected by fading as Lognormal random variables, i.e., given a channel coefficient a, we have $\log(a) \sim \mathcal{N}(\mu, \sigma^2)$, $\log(.)$ being the Napierian logarithm. The parameters μ and σ have been chosen so that in the three scenarios we have $\mathbb{E}[a_{11}] = \mathbb{E}[a_{22}] = 1$ and $\mathbb{E}[a_{10}] = \mathbb{E}[a_{20}] = 0.01$. We plot our results against the mean value of the interfering link coefficients as presented in the standard model (c_{12}, c_{21}), where direct channel coefficients have unit mean. The parameters for such coefficients are chosen so that $\mathbb{E}[c_{12}] = \mathbb{E}[c_{21}]$ take values in $[0, 1]$. This allows to compare different scenarios against a single parameter. In order to do a fair comparison, this is done for both the non-cooperative as well as the full cooperative scheme.

Table 4.9 Rate and popularity rankings for each file in the direct probability and inverse probability setups.

$\mathcal{F}$	R_i [bits/sec/Hz]	q_i *direct*	q_i *inverse*
f_1	1.20	0.45	0.05
f_2	1.00	0.20	0.15
f_3	0.50	0.15	0.15
f_4	0.40	0.15	0.20
f_5	0.20	0.05	0.45

We consider the system depicted in Fig. 4.5 for the non-cooperative approach and Fig. 4.3 for the full cooperative approach where each SBS has a memory of size $M = 2$ and the library has cardinality $N = 5$.

Two setups are considered. In the first one, files with higher rate have a higher popularity ranking while in the second one files with higher rate have a lower popularity ranking. We refer to these setups as *direct probability* and *inverse probability*, respectively. The corresponding values for R_i and q_i are shown in Table 4.9. We plot for each scenario the minimum power cost $Q(\mathbf{x}_1, \mathbf{x}_2)$ of the most significant cache allocations in the full cooperative approach (solid curves) and in the non-cooperative approach (dashed curves) versus the mean value of the interference coefficients, assumed to be the same for both users[4]. Note that, by the symmetry of the problem, the results do not change if the content of the cache of SBS_1 and that of SBS_2 are switched, i.e., the performance of the solution $\mathcal{M}_1 = \{f_i, f_j\}$, $\mathcal{M}_2 = \{f_k, f_l\}$ and $\mathcal{M}_1 = \{f_k, f_l\}$, $\mathcal{M}_2 = \{f_i, f_j\}$ coincide.

4.4.3 Numerical Results Analysis

Static Scenario

In Fig. 4.7 the most significant cache allocations for the static scenario with direct probabilities for the non-cooperative and full cooperative approach are plotted. When no cooperation is in place, the optimal caching solution consists in each SBS storing the files with highest probabilities/transmission rate ($\mathcal{M}_1 = \mathcal{M}_2 = \{f_1, f_2\}$) which is indicated with the orange dashed curve. In this case, when mutual interference is low the MBS only transmits files with low probability and low transmission rate. On the contrary, when the level of interference increases the SBSs are not able to serve their own users even if they have the requested files and transmissions are delegated to the

[4]Note that, while in the static case this is equivalent to having a symmetric channel matrix, this is not necessarily the case in the mobile transmitter case due to the independence of fading.

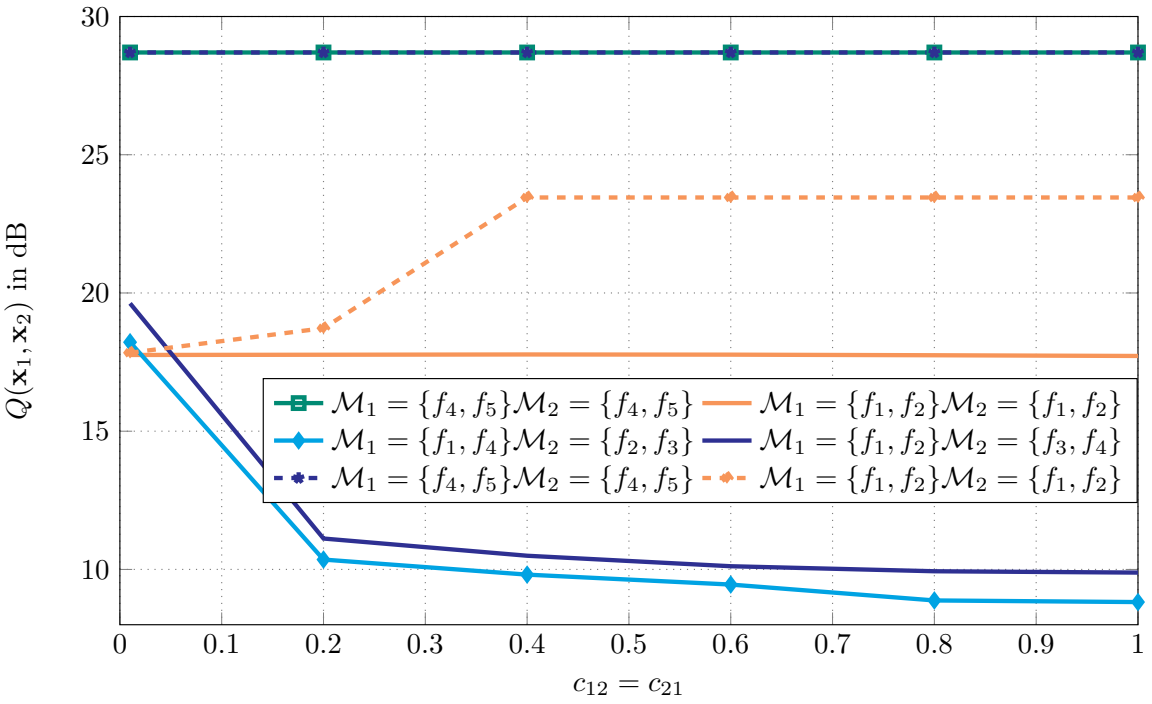


Fig. 4.7 Minimum average power cost in dB in the static case plotted against the interference coefficient with direct file probabilities. The result for the cooperative approach is represented with solid curves while dashed curves are used for the non-cooperative approach. The optimization is done for each average interference coefficient. In the simulations we set $a_{10} = a_{20} = 0.01$ and $a_{11} = a_{22} = 1$.

MBS. Thus, the usage of the MBS becomes more frequent and more power is consumed due to the high rate of the files and lower channel gain. This explains the step of roughly 4 dB from $c_{12} = c_{21} = 0.2$ to $c_{12} = c_{21} = 0.4$. For $c_{12} = c_{21}$ larger than or equal to 0.4 the curve flattens out because, from that point on, the mutual interference is so strong that the required rates cannot be achieved by the SBSs and the MBS takes over all transmissions. In the full cooperative approach the best performance is obtained when one of the SBS stores the files with highest probability or transmission rate and the two SBSs store different sets of files. The result is shown in the light blue curve with diamonds. A slight increase in power (from 0.5 dB up to 1 dB) is observed when the caches store different files but the two most probable files are in the same cache ($\mathcal{M}_1 = \{f_1, f_2\}$, $\mathcal{M}_2 = \{f_3, f_4\}$). In both full cooperative and non-cooperative approaches, the most power consuming solution coincides and corresponds to each SBS storing the two files with lower probabilities/rates ($\mathcal{M}_1 = \mathcal{M}_2 = \{f_4, f_5\}$). The flatness of the curves in both the cooperative approach and the non-cooperative approach are

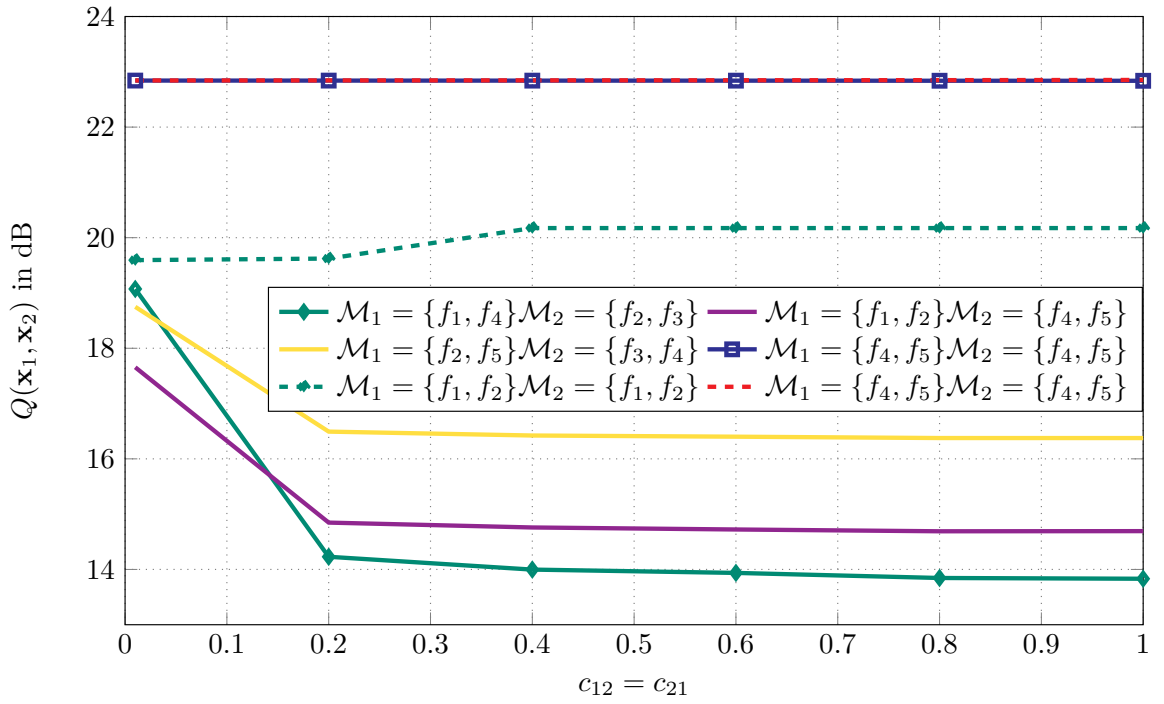


Fig. 4.8 Minimum average power cost in dB in the static scenario plotted against the interference coefficient with inverse probabilities. The optimization is done for each average interference coefficient. In the simulations we set $a_{10} = a_{20} = 0.01$ and $a_{11} = a_{22} = 1$.

due to the fact that the power consumed by the MBS has a strong weight in the overall mean, since transmissions at high data rate from MBS are the most probable.

Let us now consider the low interference regime. In the full cooperative approach, all the caching schemes at the lowest interference level consume at least 5 dB more with respect to the other levels. This is because, when a single SBS serves both users, it applies either the broadcast or the multicast channel and such channels are penalized when one of the two channel coefficients are relatively small. Later on in this section we show that, despite the higher power consumed in some cases, full cooperation has an advantage in terms of MBS resources sparing.

In Fig. 4.8 the power cost versus interference level in the static scenario and for inverse probabilities (i.e. files with higher R_i have lower q_i) is plotted. The plot shows that for both approaches the most power-efficient cache allocation (solid and dashed green curves) is not the one including the file with the highest popularity but rather the one with the highest rate. The reason for this is that the weight of MBS transmissions on the overall power budget is higher with respect to the SBSs' due to the lower channel gains. For the considered rates and popularity ranking it is better to have the

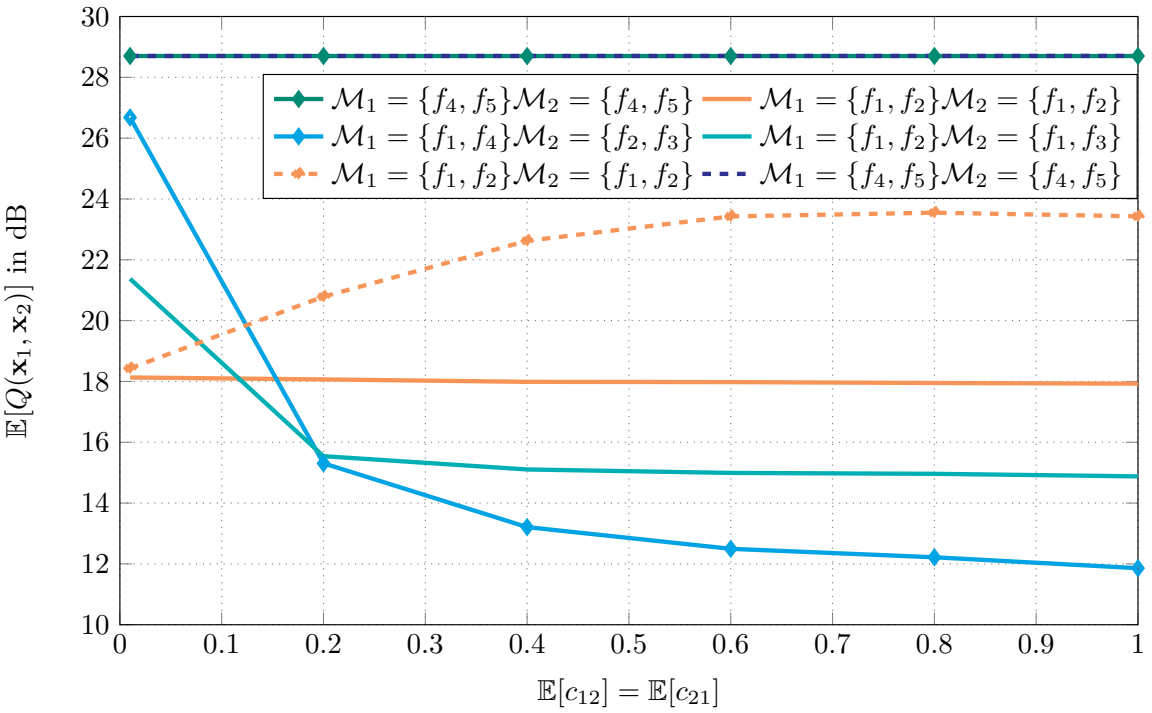


Fig. 4.9 Average power cost in dB versus interference link coefficients for different memory allocations in the mobile transmitters setup in a fading environment with direct probabilities. In the simulations we set $a_{10} = a_{20} = 0.01$ and $\mathbb{E}[a_{11}] = \mathbb{E}[a_{22}] = 1$.

MBS transmitting low rate messages more often rather than transmitting higher rate messages with lower popularity. In the full cooperative case, if the file with highest rate is not cached there is a loss larger than or equal to 2.5 dB with respect to the best curve. This holds also for the allocation including the most popular file (yellow curve), which confirms that it is better to store the file with highest rate rather than the one with highest popularity. As a matter of facts, the worst performance in the both the full cooperative and the non-cooperative case is obtained when each SBS stores the two files with largest probabilities/lowest data rate.

We examine now the mobile transmitters scenario (fading on the SBS-user link, no fading on the MBS-user link). Note that, due to the difference between two Lognormal random variables at the denominator of equation (4.4), the average transmission power in the GIN channel is not finite [50]. Comparing the full cooperative approach and the non-cooperative approach in such conditions always leads to an infinite power gain of the former with respect to the latter as long as the probability that the two SBSs transmit together is non-zero. In order to get some interesting insight in the presence of fading, in the simulations of the mobile transmitters scenario we introduced

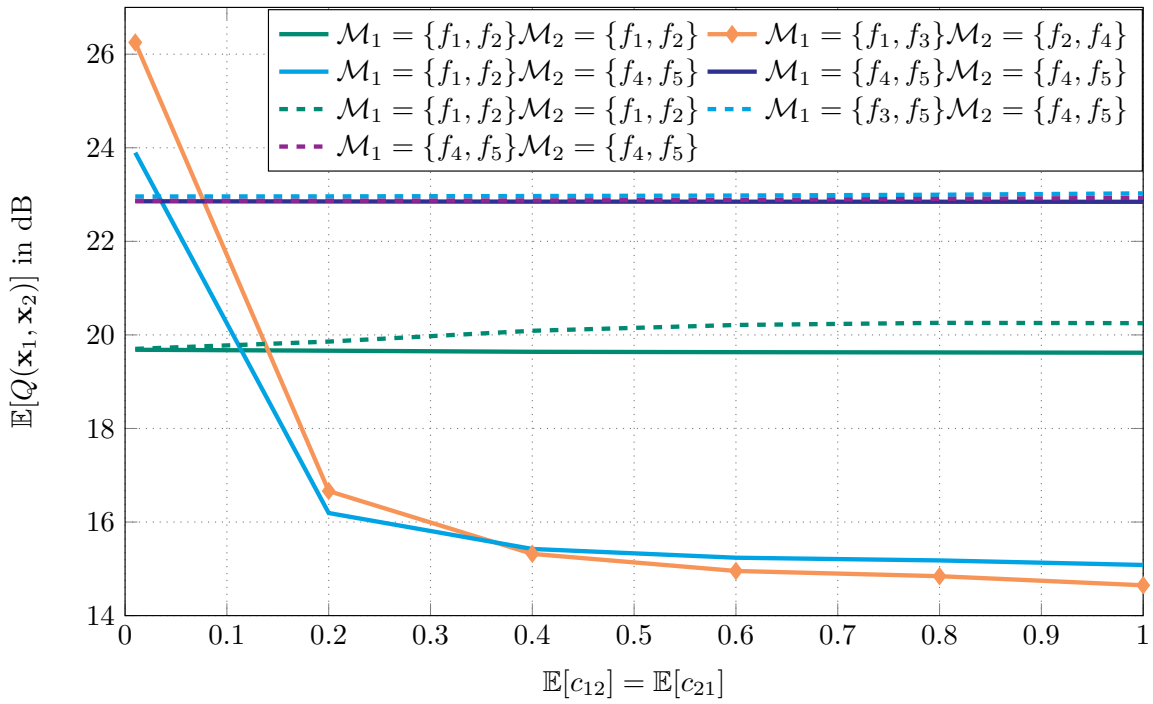


Fig. 4.10 Average power cost in dB versus interference link coefficients for different memory allocations for the mobile transmitters setup in a fading environment with inverse probabilities. In the simulations we set $a_{10} = a_{20} = 0.01$ and $\mathbb{E}[a_{11}] = \mathbb{E}[a_{22}] = 1$.

a constraint on the maximum transmit power. Such constraint is chosen so that the probability that the SBSs cannot transmit due to it is around 10^{-5} in the worst case scenario (i.e., GIN with maximum rates and $\mathbb{E}[c_{12}] = \mathbb{E}[c_{21}] = 0.4$[5]).

Mobile Transmitter Scenario

In Fig. 4.9 the minimum cost $\mathbb{E}[Q(\mathbf{x}_1, \mathbf{x}_2)]$ for different average interference levels in the mobile transmitters case for direct probabilities is plotted. Let us first consider the full cooperative approach. Similarly to the static scenario, when interference is not close to zero the best performance is obtained when each SBS stores different files and the most probable ones are in $\mathcal{M}_1 \cup \mathcal{M}_2$. The power consumption is constant (18 dB) for all levels of interference when the caches of the SBSs are equal and each SBS stores the files with highest rate/popularity. Note that such caching solution is the best for very low levels of interference. In fact, when both SBSs cache the same files, broadcast and multicast transmissions, which are the most expensive ones, are not implemented.

[5]Note that $\mathbb{E}[c_{12}] = \mathbb{E}[c_{21}] = 1$ would lead to a higher outage probability for the SBSs in the non-cooperative approach, but the outage in such case is due almost exclusively to the lack of a feasible solution to equation (4.4) rather than to the constraint on the maximum power.

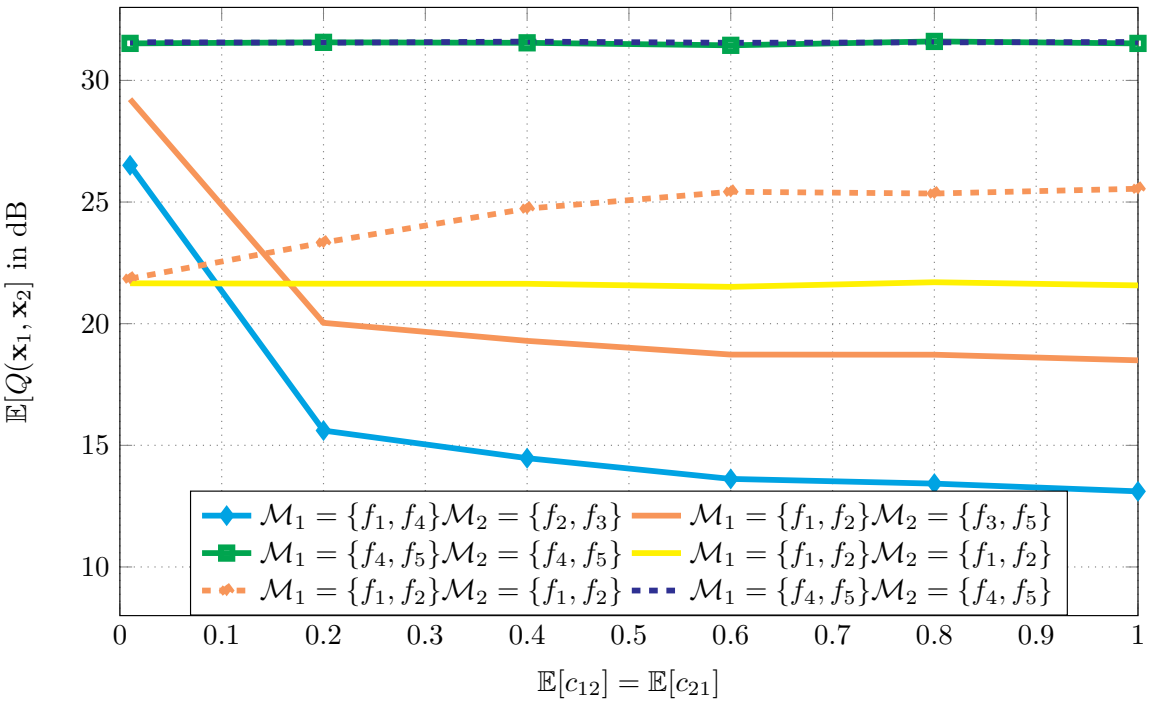


Fig. 4.11 Minimum average power cost in dB in the mobile users plotted against the interference coefficient with direct file probabilities. The optimization is done for each average interference coefficient. In the simulation we set $\mathbb{E}[a_{10}] = \mathbb{E}[a_{20}] = 0.01$ and $\mathbb{E}[a_{11}] = \mathbb{E}[a_{22}] = 1$.

In the non-cooperative approach the caching scheme which consumes less power is the one in which each SBS stores the most probable files.

Mobile Users Scenario

Let us consider the mobile user case. In Fig. 4.11 and Fig. 4.12, $\mathbb{E}[Q(\mathbf{x}_1, \mathbf{x}_2)]$ is plotted against $\mathbb{E}[c_{12}] = \mathbb{E}[c_{21}]$ for direct and inverse probability, respectively. The best and worst caching solutions in terms of power consumption are similar to the AWGN and mobile transmitters scenarios.

From the plots presented so far we can make the following observations. In the full cooperative case and for interference levels significantly larger than zero the best caching solution is to have different files stored by different transmitters. In particular, if files have inverse probability then it is convenient to store those with highest rate requirement. In the full cooperative approach the power consumption varies little with the interference level if each SBS transmitter has the same memory allocation. If no interference is present, the best solution in the full cooperative case for fixed and mobile scenarios is to have the same memory allocation in both SBSs, storing the two

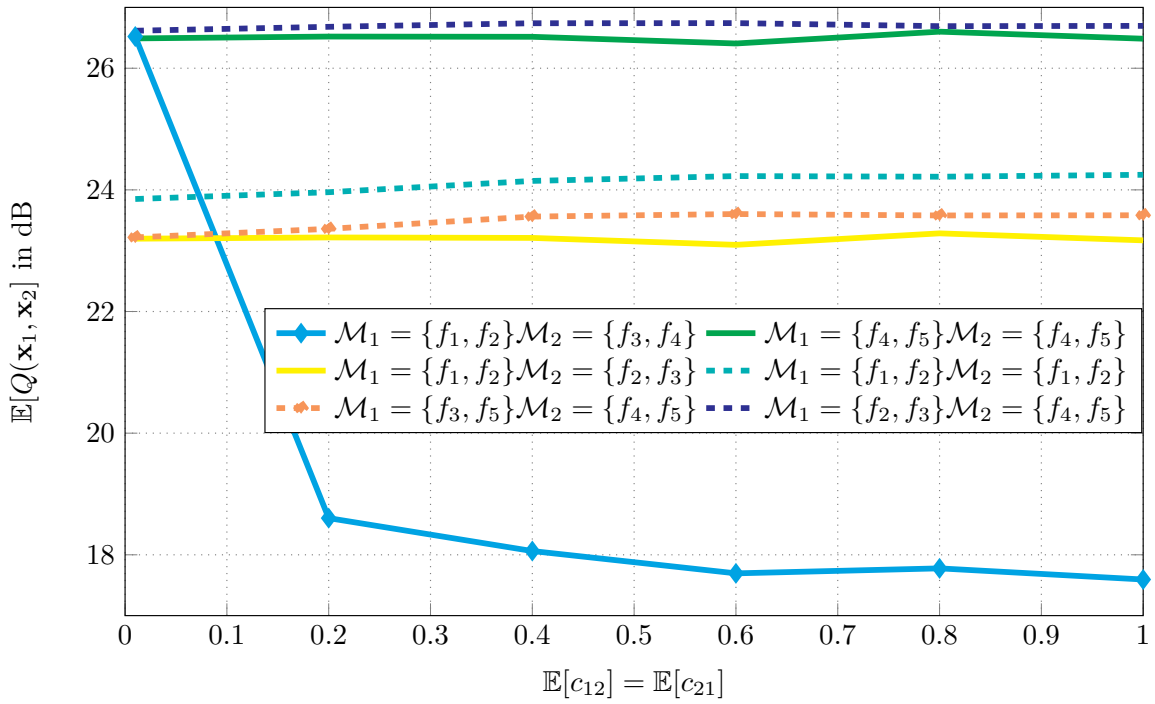


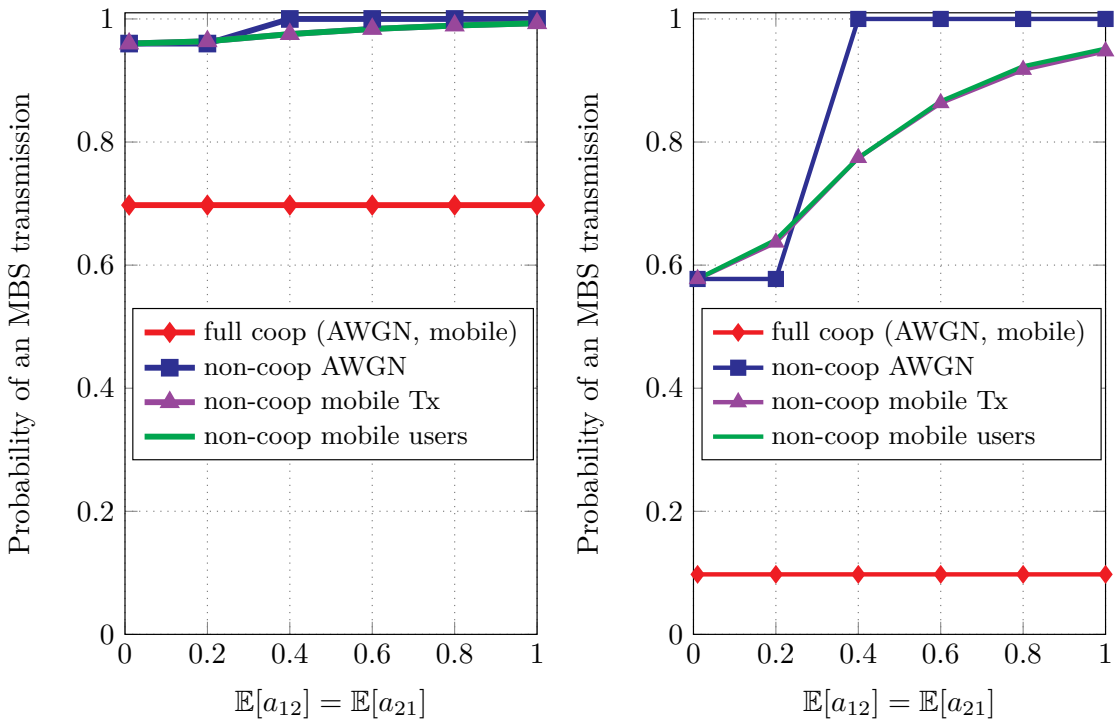
Fig. 4.12 Minimum average power cost in dB in the mobile case plotted against the interference coefficient with inverse file probabilities. The optimization is done for each average interference coefficient. In the simulation we set $\mathbb{E}[a_{10}] = \mathbb{E}[a_{20}] = 0.01$ and $\mathbb{E}[a_{11}] = \mathbb{E}[a_{22}] = 1$.

files with highest rates. Such memory allocation presents also the best solution for the non-cooperative approach in all scenarios.

Moving from the static to the mobile transmitters case and from this to the mobile users the minimum required power increases. This is not surprising since the number of links affected by fading increases. However, we note that in the direct file probability case, the optimal cache allocation in all three cases remains the same. A similar observation is also true in the inverse probability case. This may suggest that imperfect knowledge of channel statistics may be tolerated at least up to a certain extent. This aspect can have important practical implications.

MBS Utilization

So far we saw that in some scenarios and for low levels of interference the power gain deriving from full cooperation is limited. This is because we are trying to spare the usage of the MBS as much as possible. Since from the plots showed so far the MBS utilization cannot be appreciated, in Fig. 4.13a and Fig. 4.13b we plot, for a given caching allocation and for different levels of interference, the probability that a transmission



(a) Inverse (b) Direct

Fig. 4.13 Probability of having an MBS transmission at different levels of interference for a given allocation memory. $\mathcal{M}_1 = \{f_1, f_3\}, \mathcal{M}_2 = \{f_2, f_4\}$ for the full cooperative approach and $\mathcal{M}_1 = \{f_1, f_2\}, \mathcal{M}_2 = \{f_1, f_2\}$ for the non-cooperative approach.

is served by the MBS for inverse and direct probabilities, respectively. For the full cooperative approach we consider the cache allocation: $\mathcal{M}_1 = \{f_1, f_3\}, \mathcal{M}_2 = \{f_2, f_4\}$ while for the non-cooperative approach: $\mathcal{M}_1 = \mathcal{M}_2 = \{f_1, f_2\}$.

In the full cooperative approach, represented by the red curve, the probability of having a transmission from the MBS is independent from the level of interference and in the static and mobile transmitters setups the probability coincides. This is because in the full cooperative case the usage of the MBS only depends on the cache allocation of the SBSs while in the non-cooperative approach it depends also on the mutual interference level. In fact, in the non-cooperative case the probability of MBS transmissions grows with the level of interference. This is due to the fact that, in the GIN channel, transmissions are forwarded to the MBS when a feasible solution does not exists and the probability of this happening increases with the interference level. In the inverse file probabilities case, using full cooperation reduces the MBS usage of at least 25% with respect to the non-cooperative case. In Fig. 4.13b we see how the usage of the MBS is greatly reduced with full cooperation if files have direct probabilities. In

this case, since the files stored in the SBSs' caches are those with higher popularity ranking (and rates), transmissions from the MBS occur with very low probability. In fact, in the full cooperative approach MBS transmissions occur only in 10% of the cases when the best caching allocation is considered, with a gain of up to 90% with respect to the best allocation of the non-cooperative case. However, as mentioned previously in the present section, such gain implies, in some case, a higher cost in terms of power.

Low-Complexity Algorithm

In the following, we compare the performance of the optimal algorithm with that of the low-complexity algorithm proposed in Section 4.3.5. Two other benchmarks are also considered, namely, one in which the files with highest popularity ranking are cached and one in which the files with the highest rate requirements are cached. The probabilities and rate of each file are given in Table 4.10.

Table 4.10 Rate and popularity ranking.

$\mathcal{F}$	R_i	q_i [bits/sec/Hz]
f_1	1.20	0.5911
f_2	0.20	0.1697
f_3	1.40	0.0818
f_4	0.80	0.0487
f_5	1.80	0.0326
f_6	1.00	0.0235
f_7	1.60	0.0178
f_8	0.60	0.0140
f_9	2.00	0.0113
f_{10}	0.40	0.0094

We consider $N = 10$ files and a memory size for each SBS of $M = 2$. The results are shown in Fig. 4.14 for the AWGN scenario. The black curve represents the optimal caching strategy. The result of the proposed low-complexity algorithm is plotted in red. As can be seen from the plot the suboptimal solution outperforms the benchmark strategies that store files with either the highest rates or the highest popularity rankings. We recall that this is achieved with a complexity that grows linearly with N. As a last remark, note that the low-complexity algorithm and the two benchmark allocate the cache independently of the average interference, while the optimal algorithm takes it into account. Although the low complexity algorithm consumes more power than the

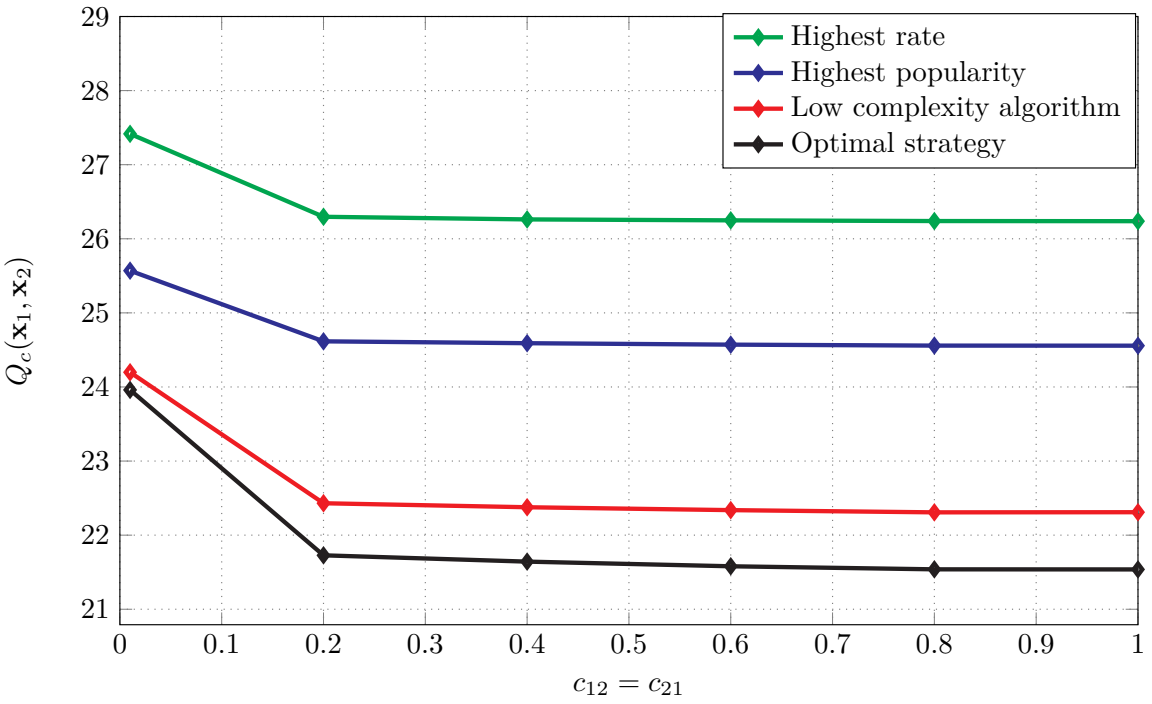


Fig. 4.14 Minimum average power cost in dB in the static scenario plotted against the interference coefficient. For $N = 10$ and $M = 2$. R_k and q_k for each file are described on Table 4.10. The optimization is done for each average interference coefficient. In the simulations we set $a_{10} = a_{20} = 0.01$ and $a_{11} = a_{22} = 1$.

optimal one, it has the advantage of not requiring previous knowledge of the average interference level.

4.5 Conclusions

In this first part of the thesis, we addressed the problem of caching optimization and content delivery over Gaussian delivery channels in the presence of interference and different per-file rate constraints and a one-shot delivery phase. Up to our knowledge this problem has not been previously studied in literature.

We formulated and analyzed the caching problem in case of no cooperation between SBSs as well as in the case of limited and full cooperation. One among a set of different cooperative strategies is chosen by the transmitters depending on the cache allocation and file requests. We derived for each of the considered transmission strategies the minimum power needed for satisfying the rate constraints. An optimal cache allocation algorithm as well as a low-complexity sub-optimal algorithm were proposed.

The results show that memorizing the most popular files is not always the most convenient strategy when rate requirements have to be respected. Furthermore, we showed that in the cooperative approaches the probability of a transmission by the MBS only depends on the cache allocation and file popularity ranking while in the non-cooperative approach such probability also depends on the mutual interference. Thanks to the full cooperation, up to 90% of the MBS resources can be saved with respect to the non-cooperative case. This implies a trade-off in terms of SBSs power.

Chapter 5

Coding Background

The presence of noise implies that the digital information message received at the user side might be different from that one delivered from the source side. In other words, the noisy channel introduces errors over the desired data of the receiver. The information data transmitted can be protected by opportunely using channel coding techniques.

5.1 Definitions and Basics

We denote with $\mathbb{F}_q$ a Galois field (or finite field) of order q.

Definition 7. (Vector space) A *vector space* $\mathbb{F}_q^n$ over the field $\mathbb{F}_q$ is the set of n-tuples of elements from $\mathbb{F}_q$, along with an addition operation and a scalar multiplication. The vector addition $\mathbf{x} + \mathbf{y}$, being $\mathbf{x} \in \mathbb{F}_q^n$ and $\mathbf{y} \in \mathbb{F}_q^n$, defined as the *component-wise addition* in $\mathbb{F}_q$ of $\mathbf{x} + \mathbf{y}$, i.e.

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n).$$

The scalar multiplication $\beta \cdot \mathbf{x}$ in $\mathbb{F}_q$, being $\beta \in \mathbb{F}_q$, is

$$\beta \cdot \mathbf{x} = (\beta \cdot x_1, \beta \cdot x_2, ..., \beta \cdot x_n).$$

Definition 8. (Matrix rank) The *rank of a matrix* $\mathbf{G}$ of dimension $m \times n$, with $m \geq n$ denoted by rank($\mathbf{G}$), is defined as the number of its linearly independent rows or columns. The matrix G is said to be full-rank, if rank($\mathbf{G}$) $= n$.

Definition 9. (Linear block code) A *linear block code* of cardinality q^k over $\mathbb{F}_q$ with codeword length n and information length k, denoted by $\mathcal{C}(n,k)$, is a k-dimensional linear subspace of the n-dimensional vector space $\mathbb{F}_q^n$. The rate of a linear block code is $r = k/n$.

Theorem 1. *Let us consider a random system of m linear equations over $\mathbb{F}_q$ in k unknowns, with $m = k + \delta$ given by*

$$\mathbf{y} = \mathbf{x}\mathbf{A}, \tag{5.1}$$

where the vector $\mathbf{y}$ is known, $\mathbf{x}$ is the unknown vector with components in $\mathbb{F}_q$ and $\mathbf{A}$ is a $k \times m$ random matrix whose entries are uniformly at random distributed in $\mathbb{F}_q$. The probability $P_f(k,\delta,q)$ that the system in (5.1) does not admit a unique solution

$$P_f(k,\delta,q) = \Pr\{rank(A) < k\}$$

is given by [51] [52]

$$P_f(k,\delta,q) = 1 - \prod_{i=1}^{\delta} \Big(1 - \frac{q^{i-1}}{q^{k+\delta}}\Big). \tag{5.2}$$

5.2 Binary Erasure Channel

The binary erasure channel (BEC) is the discrete memoryless channel illustrated in Figure 5.1. The BEC is characterized by a binary input alphabet $\mathcal{X} = \{0,1\}$ and ternary output alphabet $\mathcal{Y} = \{0,1,E\}$, where the symbol E denotes an erasure event.

The channel transition probabilities are given by

$$\begin{aligned} &P_{Y|X}(0|0) = 1-\epsilon \qquad P_{Y|X}(E|0) = \epsilon \qquad P_{Y|X}(1|0) = 0 \\ &P_{Y|X}(1|1) = 1-\epsilon \qquad P_{Y|X}(E|1) = \epsilon \qquad P_{Y|X}(0|1) = 0 \end{aligned} \tag{5.3}$$

where ϵ denotes the erasure probability.

5.2.1 Erasures Correcting Codes

Typical codes used for transmission over the BEC can be classified as *fixed-rate* codes or as *rate-less* codes. Fixed-rate codes are those codes whose code rate is fixed prior to transmission. An example of fixed-rate code are MDS codes. Instead, the rate of a rate-less code does not need to be a priori fixed, since it can be lowered on demand, i.e. an unlimited amount of redundancy can be generated.

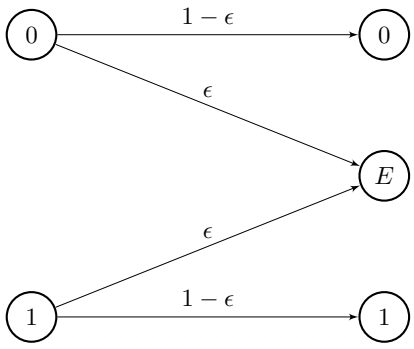


Fig. 5.1 The binary erasure channel (BEC).

5.2.2 BEC Performance Bounds

The block error probability P_B for a binary linear block code with parameters (n, k), where n is the codeword lenght and k the informartion lenght, over a BEC with erasure probability ϵ is lower bounded by the Singleton bound $P_B^{(S)}$ [53], defined as

$$P_B^{(S)} = \sum_{i=n-k+1}^{n} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i}.$$

The Singleton bound is achievable only by MDS codes.

5.3 q-ary Erasure Channel

The q-ary erasure channel (q-EC) is the discrete memoryless channel illustrated in Figure 5.2. The q-EC is characterized by a q-ary input alphabet $\mathcal{X} = \mathbb{F}_q$ and an output alphabet $\mathcal{Y} = \mathbb{F}_q + \{E\}$ of cardinality $q + 1$. The channel transition probabilities are given by

$$\begin{aligned} P_{Y|X}(y|x) &= 1 - \epsilon \qquad \text{for } y = x \\ P_{Y|X}(E|x) &= \epsilon \\ P_{Y|X}(y|x) &= 0 \qquad\quad \text{for } y \neq x. \end{aligned} \tag{5.4}$$

5.3.1 q-EC Performance Bounds

The lower bound on the block error probability P_B of a (n, k) linear block code over $\mathbb{F}_q$ on a memoryless q-EC with erasure probability ϵ is given by the Singleton bound in equation (5.4).

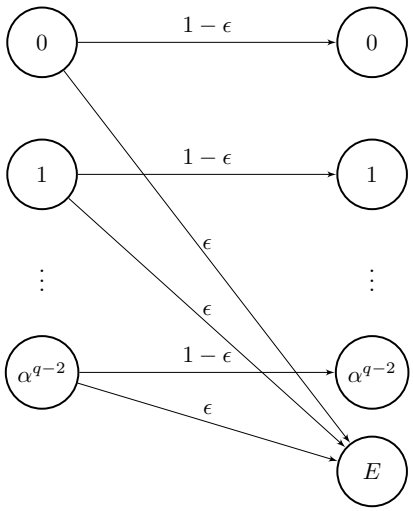


Fig. 5.2 The q-ary erasure channel (q-EC).

5.4 Rate-less Codes

Rate-less codes are ideally suited for scenarios where a large number of users have to be served through multicast or broadcast transmissions. With rate-less codes, the transmitter can generate coded symbols *on the fly* whenever necessary. With rate-less codes, a user needs to collect a certain number of encoded symbols to retrieve the original content. A feedback is used only when the decoder success in order to acknowledge the transmitter that not more symbols are needed.

Rate-less codes are applicable in scenarios where the channel can be modelled as a BEC and the probability of erasure is not known a-priori.

5.4.1 Linear Random Fountain Codes

LRFC are a simple class of rate-less codes. The encoding and decoding processes are explained in the following.

Encoding

Suppose that the transmitter has to send the message

$$\mathbf{u} = (u_1, ..., u_k)$$

to a user, where each component belongs to $\mathbb{F}_q$ and it is referred as *input symbol* or *source symbol*. A *coded symbol* or *output symbol* is generated by a random linear

combination of the input symbols. For each message the transmitter can generate a potentially unlimited number of output symbols. In particular, the i-th output symbol is generated as

$$c_i = \sum_{j=1}^{k} g_{j,i} u_j,$$

where the coefficients $g_{j,i}$ are picked with uniform probability in $\mathbb{F}_q$.
Let $\mathbf{c} = (c_1, c_2, ..., c_l)$ be the sequence of output symbols generated by the transmitter side then, for a fixed l, the encoding can be written as a system of linear equations given by the vector matrix multiplication

$$\mathbf{c} = \mathbf{uG},$$

where $\mathbf{G}$ is a $k \times l$ matrix with elements $g_{j,i}$ are picked uniformly at random in $\mathbb{F}_q$. When the $q = 2$, i.e. a binary LRFC, then an output symbol is obtained by bit-wise a subset of input symbols.

Decoding

Let us assume that the user collects m output symbols, denoted by the vector

$$\mathbf{y} = (y_1, y_2, ..., y_m)$$

with m not erased output symbols.
We denote with $\mathcal{I} = \{i_1, i_2, ..., i_m\}$ the set of indices corresponding to the m coded symbols received, i.e.

$$y_j = c_{i_j}$$

The following system of equation expresses the dependency of the collected output symbols y on the vector of input symbols

$$\mathbf{y} = \mathbf{u}\tilde{\mathbf{G}}, \tag{5.5}$$

where $\tilde{\mathbf{G}}$ is a $k \times m$ matrix given by m columns of $\mathbf{G}$ with indices in $\mathcal{I}$.

The user is able to successfully decode the desired message only if the system of linear equations in (5.5) admits a unique solution, i.e. if and only if $\tilde{\mathbf{G}}$ is full rank.
Let us define the *overhead* δ as the number of symbols in excess to k that the user correctly receives, i.e.

$$\delta = m - k,$$

the probability that the user fails to decode the message is given by Theorem 1. Hence, the probability of failure $P_{LRFC}(k, \delta, q)$ of an LRFC built over $\mathbb{F}_q$ is

$$P_{\text{LRFC}}(k, \delta, q) = 1 - \prod_{i=1}^{\delta} \Big(1 - \frac{q^{i-1}}{q^{k+\delta}}\Big). \tag{5.6}$$

The following tight bounds for P_{LRFC} were derived in [54]

$$l(\delta, q) \leq P_{\text{LRFC}}(k, \delta, q) < u(\delta, q) \tag{5.7}$$

where

$$l(\delta, q) := q^{-\delta-1} \tag{5.8}$$

and

$$u(\delta, q) := \frac{q^{-\delta}}{q-1}. \tag{5.9}$$

Both bounds are independent from the number of input symbols k.

In Fig. 5.3 bounds of the probability of failure of the LRFC decoder $P_{\text{LRFC}}(k, \delta, q)$ as a function of the overhead δ for $q = 2, 4, 8, 16, 32, 64$ is plotted. As can be observed in the plot, the bounds become tighter for increasing q.

The decoding process consists in the inversion of the matrix $\tilde{\mathbf{G}}$ which can be done using, for example, Gaussian elimination.

5.4.2 LT Codes

Luby transform (LT) are the first pratical rate-less codes for the BEC and were introduced in [55]. The encoding and decoding process is explained in the following. We shall consider only binary LT codes.

Encoding

LT codes are characterized by the output symbol degree distribution

$$\Omega = (\Omega_1, \Omega_2, ..., \Omega_{d_{\max}}),$$

where $d_{\max}$ is the maximum output degree. Being $\mathbf{u} = (u_1, u_2, ..., u_k)$ the vector of the input symbols, the i-th output symbol c_i is generated by linearly combining D distinct

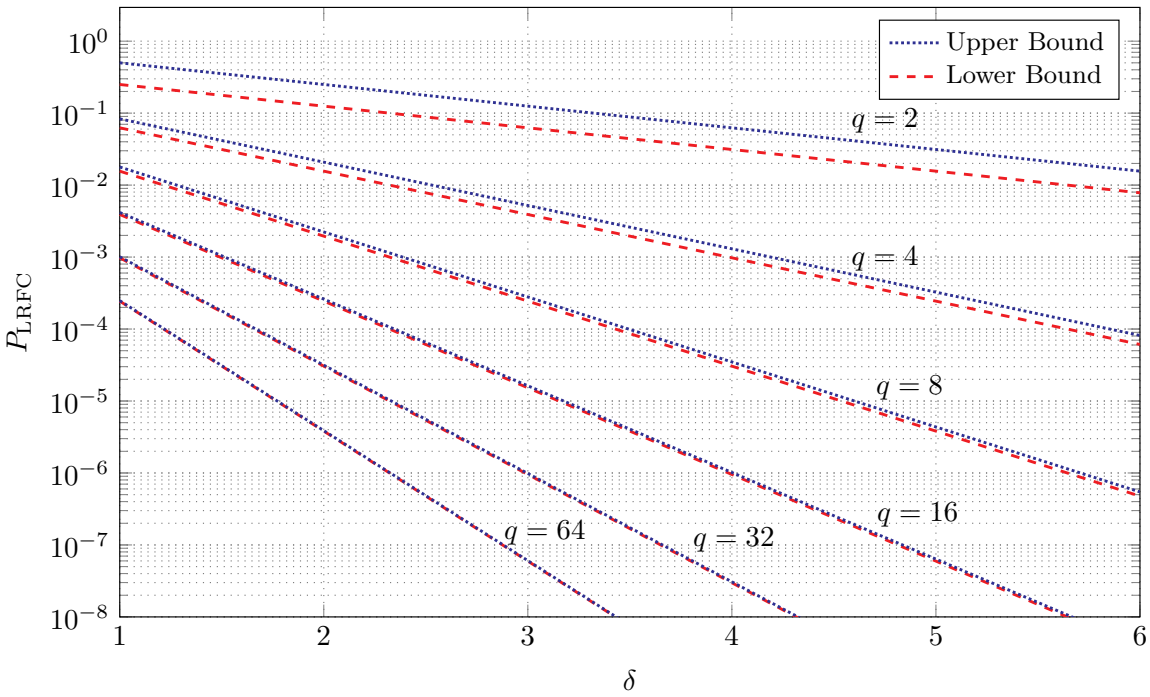


Fig. 5.3 Lower and upper bound of the decoding error probability for LRFC over $\mathbb{F}_q$, for $q = 2, 4, 8, 16, 32, 64$.

input symbols selected at random among the k source symbols. The output symbol degree D is a random variable with probability

$$\Omega_d = \Pr[D = d].$$

An unlimited number of output symbols can be generated by the transmitter encoder. Let $\mathbf{c} = (c_1, c_2, ..., c_l)$ be the sequence of output symbols generated at the transmitter side. For a fixed l, the encoding can be written as a system of linear equations given by the vector matrix multiplication

$$\mathbf{c} = \mathbf{uG},$$

where $\mathbf{G}$ is the generator matrix of the LT code with dimension $k \times l$, each column of $\mathbf{G}$ is associated to an output symbol.

LT codes can be represented by a bipartite graph, as shown in Figure 5.4. Matrix $\mathbf{G}$ denotes the incidence matrix of the bipartite graph. The two classes of nodes in the bipartite graph correspond to input and output symbols. The matrix $\mathbf{G}$ has one row for each element of $\mathbf{u}$ and one column for each element of $\mathbf{c}$. The entry in row i and

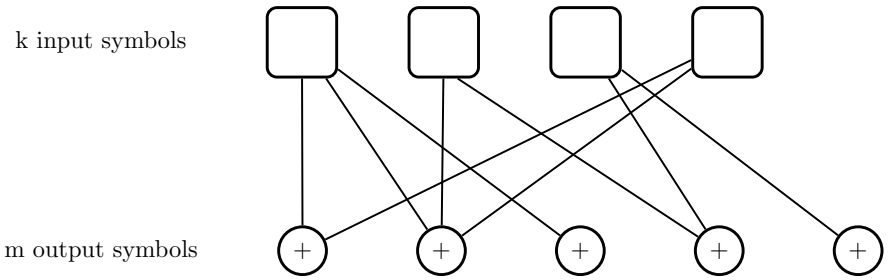


Fig. 5.4 LT bipartite graph.

column j is 1 if u_i is an addend on the linear combination for generating c_j and 0 if it is not.

For example, the LT bipartite graph of Figure 5.4 consists on $k = 4$ input symbols and $m = 5$ output symbol, we can write the corresponding system of linear equations as follows

$$\begin{cases} c_1 &= u_1 + u_4 \\ c_2 &= u_1 + u_2 + u_4 \\ c_3 &= u_1 \\ c_4 &= u_2 + u_3 \\ c_5 &= u_3 \end{cases} \tag{5.10}$$

where the incidence matrix $\mathbf{G}$ is

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{5.11}$$

LT codes can be decoded with a suboptimal low complexity algorithm, *iterative decoding*, which was proposed in [55] and it is described next.

Iterative Decoding

Let us now assume that the user receives m output symbols $\mathbf{y} = (y_1, y_2, ..., y_m)$ with m not erased output symbols and let us denote by $\mathcal{I} = \{i_1, i_2, ..., i_m\}$ the set of indices which correspond to the $m \geq k$ coded symbols received, so that

$$y_j = c_{i_j}.$$

The collected output symbols can be related to the input symbols by means of the system of equations

$$\mathbf{y} = \mathbf{u}\tilde{\mathbf{G}}, \tag{5.12}$$

where $\tilde{\mathbf{G}}$ is a $k \times m$ matrix given by m columns of $\mathbf{G}$ with indices in $\mathcal{I}$.

Iterative peeling decoding consists of initially marking all input symbols as unresolved and then executing k decoding steps. At every decoding step, the decoder searches for degree one output symbols. If a degree one output symbol y is found, the input symbol u connected by an edge with y is declared as resolved and all the edges attached to u are erased from the bipartite graph. Erase an edge which connects u to y corresponds on doing the xor operation between u and y.

This operation reduces the degree of all the output symbols attached to u. If no degree one output symbol is found a decoding failure is declared. The decoding succeeds when k input symbols are recovered after k decoding stages.

The probability of failure at the user's decoder can be derived as follows [56, 57]. Let us define

- *ripple* $\mathscr{R}$: as the set of output symbols of degree $d \geq 1$,
- *cloud* $\mathscr{C}$: as the set of output symbols of degree $d \geq 2$,
- $\mathtt{C}$: as the random variable associated to the cardinality of the cloud and $\mathtt{c}$ its realization,
- $\mathtt{R}$: as the random variable associated to the cardinality of the ripple and $\mathtt{r}$ its realization.

The probability of decoding failure $P_{\mathrm{LT}}(\delta)$ can be derived by modelling the LT decoder as a finite state machine, as in [56, 57].

Let us introduce subscript u to indicate the number of unresolved input symbols, so that $\mathtt{C}_u$, for example, represents the cardinality of the cloud when u input symbols are unresolved. Given the overhead $\delta = m - k$, the decoder is represented as a finite state machine with state

$$\mathsf{S}_u := (\mathtt{C}_u, \mathtt{R}_u).$$

Based on this model, a recursion can be obtained to derive $\Pr\{\mathsf{S}_{u-1} = (\mathtt{c}_{u-1}, \mathtt{r}_{u-1})\}$ as a function of $\Pr\{\mathsf{S}_u = (\mathtt{c}_u, \mathtt{r}_u)\}$.

Such recursion allows to obtain the exact probability of the decoder being at state $\mathsf{S}_u := (\mathtt{c}_u, \mathtt{r}_u)$ for $u = k, k-1, \ldots, 1$. By observing that decoding fails whenever the

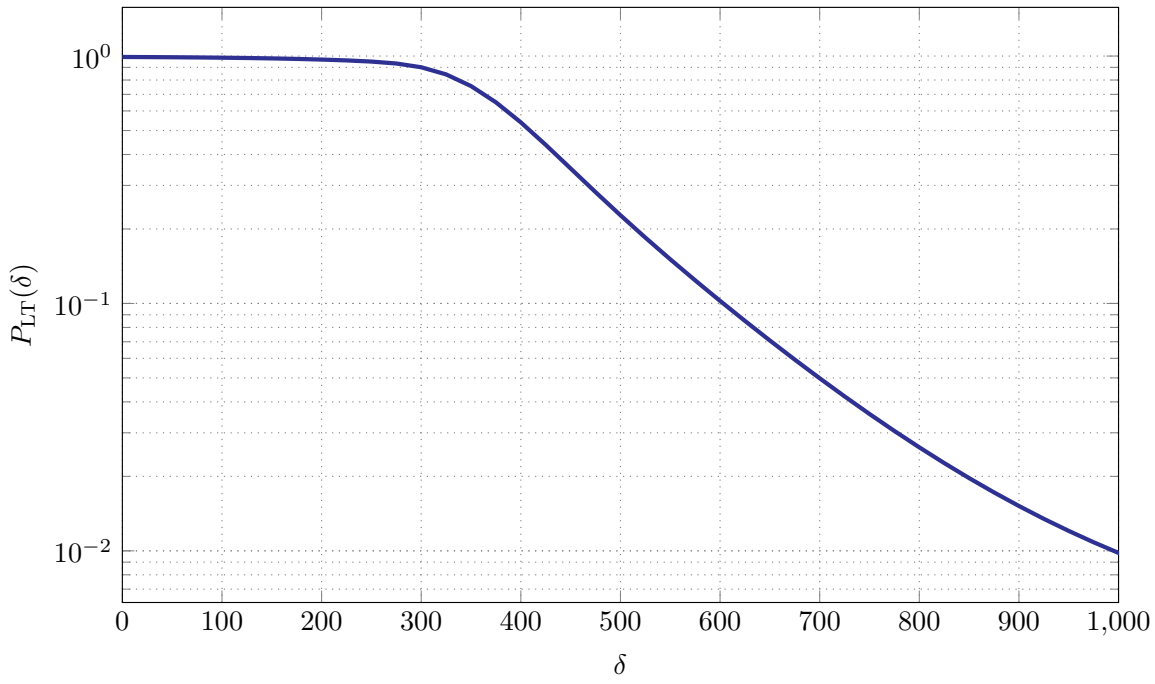


Fig. 5.5 Failure probability in function of overhead δ for a robust Soliton distribution with parameters $c = 0.05$ and $d = 3$ for k = 10000 under iterative decoding.

ripple is empty ($\mathtt{r}_u = 0$), the probability of decoding failure can be obtained as

$$P_{\mathrm{LT}}(\delta) = \sum_{u=1}^{k} \sum_{\mathtt{c}_u} \Pr\{\mathsf{S}_u = (\mathtt{c}_u, 0)\}, \tag{5.13}$$

where $P_{\mathrm{LT}}(\delta) = 1$ for $\delta < 0$.

Aiming at reducing the probability of failure of the LT decoder, two degree distributions were designed and proposed in [55]: the ideal soliton distribution (ISD) and, its improved version, the robust soliton distribution (RSD). The idea behind those designs is to maintain with high probability the ripple not empty during the whole decoding process. Both distributions are detailed in Appendix C.

As an example, in Fig. 5.5 the probability of decoding failure as a function of the overhead δ for RSD under iterative peeling decoding is plotted for k = 10000.

Chapter 6

Caching with Rate-Less Codes in Heterogeneous Networks

Within this chapter, rate-less code based caching schemes for heterogeneous networks are proposed. The performance of the considered scheme is compared with a benchmark scheme that employs MDS codes. To this end, we propose to optimize the cached content during the placement phase.

6.1 Rate-Less Codes based Caching Scheme in Heterogeneous Networks

The caching process in heterogeneous networks based on encoded content can be explained as follows. During the placement phase, each content is first fragmented and encoded by the master node and then placed in the network caches. During the delivery phase, the encoded fragments of the requested content are delivered to the user and he reconstructs the original content.

Despite the fact that, the best performance of caching schemes in heterogeneous network is obtained when optimal codes are used (i.e. the MDS based caching schemes), in this work, we propose to analyze the behavior of caching based on rate-less codes. The main reason is based on the possibility to apply our scheme is those heterogeneous network where its topology is unknown or varies in time. Since rate-less codes allow generate an undefined amount of encoded fragments, the master node can provide extra fragments to the users or fill up the network caches whenever is needed. This

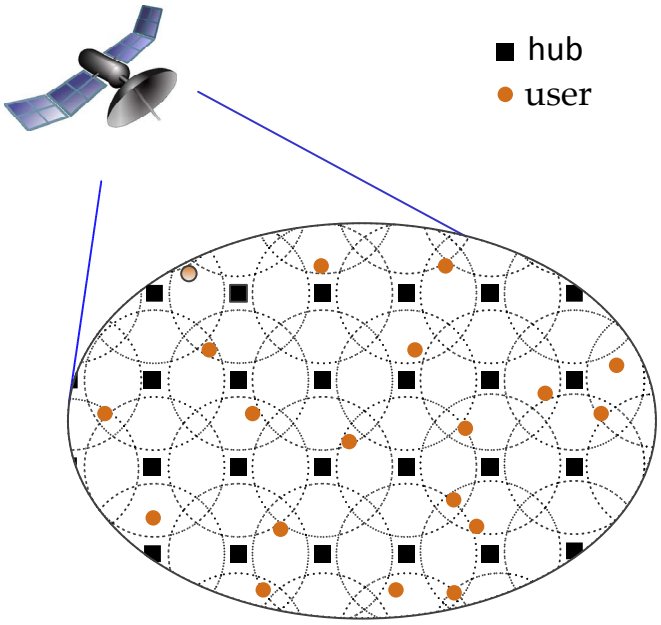


Fig. 6.1 Satellite Heterogeneous Network.

is not possible with the MDS caching scheme in which the code rate has to be fixed a-priory as a function of the number of caches present in the network [31].

The fixed-rate MDS codes can achieve the best possible performance, however, they also have a high encoding and decoding complexity, and one has to choose their rate code in advance. In contrast, in scenarios where the probability of erasures is relatively high or/and when is not known a-priory how much redundancy is needed for recovering the original message then caching schemes based on rate-less codes present the most suitable choice to apply.

In order to understand how rate-less caching schemes perform compared to an MDS caching scheme, in the following we start presenting the design and the analysis of a simple caching scheme based in LRFC in a heterogeneous network. Motivated by the results obtained we extend our studies to practical rate-less caching schemes based on LT codes.

6.2 System Model

We consider a heterogeneous satellite network as shown in Fig. 6.1. The heterogeneous network is composed by a single GEO, a number of *hubs* (e.g. terrestrial repeaters or high altitude platform stations (HAPS)) which are represented with black squares and

a number of users represented with orange circles. Each user according its location is connected to one or multiple hubs but no direct connection exists between a user and the satellite. Instead, each hub is connected to the satellite through a backhaul link.

The satellite has access to a library of content (e.g. video content) that users may ask for. The content is first fragmented and then encoded by the satellite using rate-less codes. Each hub is equipped with a cache of limited size which is used to store coded content. When a user requests content, he is first served by the hubs that he is connected to and whenever the content delivered by the hubs is not enough to fulfill the request, the satellite might provide the rest of the content through the backhaul link.

The scheme consists of two phases, placement and delivery, which are explained in the following.

Placement Phase

During the placement phase, the GEO fragments and encodes each file using rate-less codes. All files are fragmented into the same number of packets/input symbols. Then, the satellite sends output symbols of each file to each hub. For each file, the number of output symbols stored is the same across all hubs, however, the sets of output symbols cached at different hubs are different. The placement phase is assumed to be carried out offline.

Delivery Phase

During the delivery phase, user requests files at random. In a first stage, the user downloads output symbols of the requested file cached in the hubs he is connected to. If the number of symbols is not enough for decoding the file successfully, then the GEO satellite generates additional output symbols until an acknowledgment of successful in decoding is received. The use of the satellite during the delivery phase might happen because the user fails on decoding the content requested and extra symbols are needed.

The GEO sends the additional output symbols to the user via one of the hubs he is connected to.

Notation

The following notation will be is used

- $\mathcal{F} = \{\mathsf{f}_1 \ldots, \mathsf{f}_n\}$: as the library of n equally long files,

- M: the memory size of a hub in files,
- k: the number of packets that a file is fragmented into,
- w_j: the number of encoded packet cached at each hub for file f_j,
- θ_j: probability of file f_j being requested,
- γ_h: the probability of a user to be connected to h hubs,
- z: total number of coded packet that a user receives from the hubs he is connected to.

6.3 Average Backhaul Transmission Rate

In order to evaluate the performance of the proposed caching schemes, we shall use as metric the *average backhaul transmission rate*, $\mathbb{E}[T]$, which is defined as the average number of output coded symbols that the GEO has to send during the delivery phase in order to fulfill a user request. In the following we derive the analytical expression of the average backhaul transmission rate for a caching scheme based on rate-less codes. Let us define the following variables for the deriving $\mathbb{E}[T]$

- Z: the random variable associated to the number of output symbols stored in the hubs that the user is connected to and let us indicate with z its realization,
- H: the random variable associated to the number of hubs a user is connected to and let us indicate with h its realization,
- J: the random variable associated to the index of the file requested by a user and let us indicate with j its realization,
- T: the random variable denoting the number of output symbols which have to be sent over the backhaul link to serve the request of a user and let us indicate with t its realization.

Recalling that w_j is the number of coded symbols from file j stored in every hub, we have that

$$P_{Z|J,H}(z|j,h) = \begin{cases} 1 & \text{if } z = w_j\, h \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

The probability mass function of Z is

$$\begin{aligned} P_Z(z) &= \sum_j \sum_h P_{Z|J,H}(z|j,h) P_J(j) P_H(h) \\ &= \sum_j \sum_h \theta_j \, \gamma_h \, P_{Z|J,H}(z|j,h). \end{aligned} \tag{6.2}$$

We are interested in deriving the distribution of the backhaul transmission rate T. If we condition T to Z, the probability of $T = t$ corresponds to the probability that decoding succeeds when the user has received exactly t output symbols from the backhaul link in excess to the z output symbols it received from the hubs through local links, that is, when $m = z + t$.
The expectation of T is obtained as

$$\begin{aligned} \mathbb{E}[T] &= \sum_t t P(t) \\ &= \sum_t t \sum_{z=0}^{\infty} P_{T|Z}(t|z) P_Z(z) \\ &= \sum_t t \left(\sum_{z=0}^{k} P_{T|Z}(t|z) P_Z(z) + \sum_{z=k+1}^{\infty} P_{T|Z}(t|z) P_Z(z) \right), \end{aligned} \tag{6.3}$$

where in the last equality two different cases: (i) $z \leq k$ and (ii) $z > k$, are distinguished. In the former case, the user has collected less than k output symbols from the neighboring hubs while in the latter, at least k output symbols has been collected. Let us define $\bar{T}_1$ and $\bar{T}_2$ as

$$\bar{T}_1 := \sum_t t \sum_{z=0}^{k} P_{T|Z}(t|z) P_Z(z) \tag{6.4}$$

$$\bar{T}_2 := \sum_t t \sum_{z=k+1}^{\infty} P_{T|Z}(t|z) P_Z(z) \tag{6.5}$$

so that

$$\mathbb{E}[T] = \bar{T}_1 + \bar{T}_2. \tag{6.6}$$

Let us define $P_{\mathrm{F}}(\delta)$ the probability of failure of the decoder for a rate-less code when δ symbols in excess to k has been received, such that for a fixed order q and a fixed

number of fragments k we have

$$P_{\mathrm{F}}(\delta) = \begin{cases} P_{\mathrm{LRFC}}(k, \delta, q) & \text{if LRFC is used} \\ P_{\mathrm{LT}}(\delta) & \text{if LT is used.} \end{cases} \tag{6.7}$$

Now, we derive $P_{T|Z}(t|z)$ for $z \leq k$ and for $z > k$. If we introduce the variable change $\delta = z - k + t$ in the expression of $\bar{T}_1$, we obtain

$$\begin{aligned} \bar{T}_1 &= \sum_{z=0}^{k} P_Z(z) \sum_{\delta=z-k}^{\infty} (\delta - z + k)\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &\overset{\text{(a)}}{=} \sum_{z=0}^{k} P_Z(z) \sum_{\delta=0}^{\infty} (\delta - z + k)\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &= \sum_{z=0}^{k} P_Z(z) \left(\sum_{\delta=0}^{\infty} \delta\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] + (k-z) \sum_{\delta=0}^{\infty} \Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \right) \\ &\overset{\text{(b)}}{=} \sum_{z=0}^{k} P_Z(z) \Big(\mathbb{E}[\Delta] + k - z \Big) \\ &= (\mathbb{E}[\Delta] + k) \Pr\{Z \leq k\} - \sum_{z=0}^{k} z P_Z(z) \end{aligned} \tag{6.8}$$

where equality (a) is due to $[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)] = 0$ for $\delta < 0$ and equality (b) is due to

$$\sum_{\delta=0}^{\infty} [P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)] = 1.$$

Introducing the same variable change in the expression of $\bar{T}_2$ we have

$$\begin{aligned} \bar{T}_2 &= \sum_{z=k+1}^{\infty} P_Z(z) \sum_{\delta=z-k}^{\infty} (\delta - z + k)\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &= \sum_{z=k+1}^{\infty} P_Z(z) \left(\sum_{\delta=z-k}^{\infty} \delta\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] + \sum_{\delta=z-k}^{\infty} (k-z)\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \right). \end{aligned} \tag{6.9}$$

Let us rewrite (6.9) as follow

$$\bar{T}_2 = \sum_{z=k+1}^{\infty} P_Z(z) \Big(\bar{T}_{21}(z) + \bar{T}_{22}(z) \Big), \tag{6.10}$$

where

$$\begin{aligned}\bar{T}_{21}(z) &= \sum_{\delta=z-k}^{\infty} \delta\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &= \sum_{\delta=0}^{\infty} \delta\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] - \sum_{\delta=0}^{z-k-1} \delta\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &= \sum_{\delta=0}^{\infty} P_{\mathrm{F}}(\delta) - \left[\sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta) - (z-k)P_{\mathrm{F}}(z-k-1)\right] \\ &= \mathbb{E}[\Delta] - \sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta) + (z-k)P_{\mathrm{F}}(z-k-1) \end{aligned} \tag{6.11}$$

and

$$\begin{aligned}\bar{T}_{22}(z) &= \sum_{\delta=z-k}^{\infty} (k-z)\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] \\ &= (k-z)\left\{\sum_{\delta=0}^{\infty}\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big] - \sum_{\delta=0}^{z-k-1}\Big[P_{\mathrm{F}}(\delta-1) - P_{\mathrm{F}}(\delta)\Big]\right\} \\ &= (k-z)\Big\{1 - \Big[1 - P_{\mathrm{F}}(z-k-1)\Big]\Big\} \\ &= (k-z)P_{\mathrm{F}}(z-k-1). \end{aligned} \tag{6.12}$$

By inserting (6.11) and (6.12) in (6.10) and sum we obtain

$$\bar{T}_2 = \sum_{z=k+1}^{\infty} P_Z(z)\Big[\mathbb{E}[\Delta] - \sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta)\Big]. \tag{6.13}$$

Average Backhaul Transmission Rate

Finally, if we replace (6.8) and (6.13) in (6.6), the expression of the average backhaul transmission rate becomes

$$\begin{aligned}\mathbb{E}[T] &= \Big(\mathbb{E}[\Delta] + k\Big)\Pr\{Z \le k\} - \sum_{z=0}^{k} zP_Z(z) + \sum_{z=k+1}^{\infty} P_Z(z)\Big[\mathbb{E}[\Delta] - \sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta)\Big] \\ &= \mathbb{E}[\Delta] + \sum_{z=0}^{k}(k-z)P_Z(z) - \sum_{z=k+1}^{\infty} P_Z(z)\left[\sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta)\right]. \end{aligned} \tag{6.14}$$

Upper Bound of the Average Backhaul Transmission Rate

In expression (6.35), since the following term is always non negative

$$\sum_{z=k+1}^{\infty} P_Z(z) \left[\sum_{\delta=0}^{z-k-1} P_{\mathrm{F}}(\delta) \right]$$

we can define

$$T_{\mathrm{UP}} = \mathbb{E}[\Delta] + \sum_{z=0}^{k} (k-z) P_Z(z) \tag{6.15}$$

such that (6.35) is upper bounded by

$$\mathbb{E}[T] \leq T_{\mathrm{UP}}. \tag{6.16}$$

Note that the upper bound (6.15) is independent of the probability of failure P_{F}.

6.4 LRFC Caching Scheme

In this section, we analyze the caching scheme based on linear random fountain codes.

We specify the expression of the average backhaul rate for the LRFC caching scheme as well as a tight bound to it. The placement optimization problem to minimize the backhaul rate for the LRFC caching scheme is as well formulated.

We consider the system model described in Section 6.2 where the GEO fragments each $\mathsf{f}_j \in \mathcal{F}$ into k input symbols and encode each file by using independently a LRFC, as it explained in Section 5.4.1.

6.4.1 LRFC Overhead Decoding Probability

In order to calculate the average backhaul transmission rate, we first need to derive the overhead decoding probability σ_δ. That is, the probability that a user needs exactly $k + \delta$ coded symbols to successfully decode the requested file.

Recovering the requested file by means of maximum likelihood (ML) decoding reduces to solving the following system of linear equations when $m = k + \delta$ output symbols has been received

$$\mathbf{y} = \mathbf{u}\tilde{\mathbf{G}}.$$

We denote denote with S_δ the event that the matrix $\tilde{\mathbf{G}}$ is full rank when $m = k + \delta$ output symbols have been collected, where

$$\Pr\{S_\delta\} = 1 - P_{\mathrm{LRFC}}(k, \delta, q).$$

Let us denote the complementary event, i.e. the rank of $\tilde{\mathbf{G}}$ is less than k, with $\bar{S}_\delta$. We are interested in deriving

$$\sigma_\delta := \Pr\{S_\delta \mid \bar{S}_{\delta-1}\}.$$

Starting from

$$\begin{aligned}\Pr\{S_\delta\} &= \Pr\{S_\delta \mid S_{\delta-1}\}\Pr\{S_{\delta-1}\} + \Pr\{S_\delta \mid \bar{S}_{\delta-1}\}\Pr\{\bar{S}_{\delta-1}\} \\ &= \Pr\{S_{\delta-1}\} + \Pr\{S_\delta \mid \bar{S}_{\delta-1}\}\Pr\{\bar{S}_{\delta-1}\}\end{aligned}$$

we have that

$$\Pr\{S_\delta \mid \bar{S}_{\delta-1}\} = \frac{\Pr\{S_\delta\} - \Pr\{S_{\delta-1}\}}{\Pr\{\bar{S}_{\delta-1}\}}$$

$$= \frac{1 - \Pr\{\bar{S}_\delta\} - (1 - \Pr\{\bar{S}_{\delta-1}\})}{\Pr\{\bar{S}_{\delta-1}\}}$$
$$= 1 - \frac{\Pr\{\bar{S}_\delta\}}{\Pr\{\bar{S}_{\delta-1}\}} \tag{6.17}$$

which can be rewritten as

$$\sigma_\delta = 1 - \frac{P_{\text{LRFC}}(k, \delta, q)}{P_{\text{LRFC}}(k, \delta - 1, q)}. \tag{6.18}$$

The expression in (6.18) holds for $\delta \geq 0$, whereas for $\delta < 0$ we have $\sigma_\delta = 0$.
Bounds on equation (6.18) can be obtained from (5.7). In particular, for $\delta \geq 0$ we can write

$$1 - \frac{u(\delta, q)}{l(\delta - 1, q)} < \sigma_\delta < 1 - \frac{l(\delta, q)}{u(\delta - 1, q)} \tag{6.19}$$

yielding

$$1 - \frac{1}{q - 1} < \sigma_\delta < 1 - \frac{q - 1}{q^2}. \tag{6.20}$$

We define the bounds of σ_δ with

$$\Psi_l := 1 - \frac{1}{q - 1} \tag{6.21}$$

and

$$\Psi_u := 1 - \frac{q - 1}{q^2}. \tag{6.22}$$

Note that bounds are independent of the overhead δ (for non negative δ) and become tight as q grows. Note also that for $q = 2$ the lower bound becomes 0 and, hence, it loses significance.

6.4.2 LRFC Overhead Average

We denote with Δ_{LRFC} the random variable associated to the average number of symbols in excess to k that a user needs to recover the requested content and let us also denote as δ its realization.
The decoding process can be modelled using a Markov chain of infinite states where the initial state represents the event that the user has collected $m = k$ output symbols. From each state there are two transition probabilities.

At step i, i.e. when $\delta = i$, the transition probability to go to the final state (successfully decode) is given by σ_i. Thus, the average number of step for decoding

can be simple written as

$$\mathbb{E}[\Delta_{\text{LRFC}}] = \sum_{\delta=0}^{\infty} \delta \left[\prod_{i=0}^{\delta-1} (1-\sigma_i) \right] \sigma_\delta. \tag{6.23}$$

By using (6.21) and (6.22) in (6.23), $\mathbb{E}[\Delta_{\text{LRFC}}]$ can be upper bounded as

$$\begin{aligned} \mathbb{E}[\Delta_{\text{LRFC}}] &\leq \sum_{\delta=0}^{\infty} \delta \left(1-\Psi_l\right)^{\delta} \Psi_u \\ &= \Psi_u \sum_{\delta=0}^{\infty} \delta \left(1-\Psi_l\right)^{\delta} \\ &= \Psi_u \frac{1-\Psi_l}{\Psi_l^2} \\ &= \frac{q-1}{(q-2)^2}\Big(1-\frac{q-1}{q^2}\Big) \end{aligned} \tag{6.24}$$

We define

$$\delta_u := \frac{q-1}{(q-2)^2}\Big(1-\frac{q-1}{q^2}\Big). \tag{6.25}$$

6.4.3 LRFC Average Backhaul Transmission Rate

From (6.14), we can write the average backhaul transmission rate for the linear random fountain codes based caching scheme as

$$\mathbb{E}[T_{\text{LRFC}}] = \mathbb{E}[\Delta_{\text{LRFC}}] + \sum_{z=0}^{k}(k-z)P_Z(z) - \sum_{z=k+1}^{\infty} P_Z(z) \left[\sum_{\delta=0}^{z-k-1} P_{\text{LRFC}}(\delta) \right]. \tag{6.26}$$

Based on the bound derived in (6.16), the average backhaul transmission rate in (6.26) is upper bounded by

$$\mathbb{E}[T_{\text{LRFC}}] \leq \mathbb{E}[\Delta_{\text{LRFC}}] + \sum_{z=0}^{k}(k-z)P_Z(z). \tag{6.27}$$

6.4.4 LRFC Cache Placement Optimization

The LRFC placement problem calls for defining the number of output symbols per file that each hub has to cache during the placement phase such that the average transmission backhaul rate during the delivery phase is minimized.

We present in this section the placement optimization problem adapted to a LRFC cached scheme based on the optimization problem proposed for MDS codes in [31].

The LRFC Placement optimization problem can be written as

$$\min_{w_1,\ldots,w_n} \quad \mathbb{E}[T_{\text{LRFC}}] \tag{6.28}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} w_j = Mk \tag{6.29}$$

$$w_j \in \mathbb{N}_0. \tag{6.30}$$

The first constraint determines that the total number of stored coded symbols should be equal to the size cache. The second constraint specifies the discrete nature of the optimization variable.

Solving exactly the optimization problem requires evaluating the equation (6.26), which is computationally expensive. Hence, as an alternative to minimizing the average backhaul rate, we propose minimizing its upper bound in (6.16), which leads to the following optimization problem

$$\min_{w_1,\ldots,w_n} \quad \left[\mathbb{E}[\Delta_{\text{LRFC}}] + \sum_{z=0}^{k}(k-z)P_Z(z)\right] \tag{6.31}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} w_j = Mk \tag{6.32}$$

$$w_j \in \mathbb{N}_0. \tag{6.33}$$

Since the upper bound on $\mathbb{E}[T_{\text{LRFC}}]$ in (6.26) relies on the upper bound in (6.16), which is a tight bound, we expect the result of the optimization problem in (6.31)-(6.33) to be close to the result of the optimization problem in (6.28)-(6.30).

A suboptimal solution of the problem in (6.31)-(6.33) can be obtained by using linear programming methods when the constraint (6.33) is relaxed.

6.5 LT Caching Scheme

In this section, we analyze the caching scheme based on LT. We specify the expression of the average backhaul rate for the LT caching scheme as well as a tight bound to it. The placement optimization problem to minimize the backhaul rate for the LT caching scheme is as well formulated.

6.5.1 LT Overhead Average

We denote with Δ_{LT} the random variable associated to the number of LT output symbols in excess to k that a user needs to successfully decode the requested content. We denote with δ the realization of Δ_{LT}. We can calculate the expected value of Δ_{LT}, i.e. the average overhead, as follows

$$\begin{aligned}\mathbb{E}[\Delta_{\text{LT}}] &= \sum_{\delta=1}^{\infty} \delta \cdot \Big[P_{\text{LT}}(\delta-1) - P_{\text{LT}}(\delta)\Big] \\ &= \sum_{\delta=0}^{\infty} (\delta+1) \cdot P_{\text{LT}}(\delta) - \sum_{\delta=0}^{\infty} \delta \cdot P_{\text{LT}}(\delta) \\ &= \sum_{\delta=0}^{\infty} P_{\text{LT}}(\delta), \end{aligned} \tag{6.34}$$

where $P_{\text{LT}}(\delta)$ was derived in Section 5.4.2.

6.5.2 LT Average Backhaul Transmission Rate

From (6.14), we can write the average backhaul transmission rate for the LT based caching scheme as

$$\mathbb{E}[T_{\text{LT}}] = \mathbb{E}[\Delta_{\text{LT}}] + \sum_{z=0}^{k} (k-z) P_Z(z) - \sum_{z=k+1}^{\infty} P_Z(z) \left[\sum_{\delta=0}^{z-k-1} P_{\text{LT}}(\delta) \right]. \tag{6.35}$$

Based on the bound derived in (6.16)-(6.35) is upper bounded by)

$$\mathbb{E}[T_{\text{LT}}] \leq \mathbb{E}[\Delta_{\text{LT}}] + \sum_{z=0}^{k} (k-z) P_Z(z). \tag{6.36}$$

6.5.3 LT Cache Placement Optimization

The LT placement problem calls for defining the number of output symbols per files that each hub has to cache during the placement phase such that the average backhaul rate during the delivery phase is minimized.

The LT placement optimization problem can be formulated as follows

$$\min_{w_1,\dots,w_n} \quad \mathbb{E}[T_{\mathrm{LT}}] \tag{6.37}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} w_j = Mk \tag{6.38}$$

$$w_j \in \mathbb{N} \quad j = 1,\dots,n. \tag{6.39}$$

The first constraint determines that the total number of stored coded symbols should be equal to the size cache. The second constraint specifies from the discrete nature of the optimization variable.

Solving exactly the optimization problem requires evaluating the complex equation (6.35), which is computationally expensive. Hence, as an alternative to minimize the average backhaul transmission rate, we propose minimizing its upper bound in (6.16), which leads to the following optimization problem

$$\min_{w_1,\dots,w_n} \quad \left[\mathbb{E}[\Delta_{\mathrm{LT}}] + \sum_{z=0}^{k}(k-z)P_Z(z)\right] \tag{6.40}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} w_j = Mk \tag{6.41}$$

$$w_j \in \mathbb{N} \quad j = 1,\dots,n. \tag{6.42}$$

A suboptimal solution of the problem in (6.40)-(6.42) can be obtained by using linear programming methods when the constraint (6.42) is relaxed. Note that w_j represents the number of encoded packet which is a real number.

6.6 LRFC and LT Caching Scheme Performance Results

In this section, we evaluate the performance of both caching schemes proposed and we compare them with the performance of an MDS caching scheme.

We consider the system model described in Section 6.2 where users are uniformly distributed within the coverage area of the satellite and border effects are neglected. We consider that all transmissions are error-free.

In the following, we derive the connectivity distribution γ_h based on the geometry of the network considered and we describe the Zipf distribution which is considered for modelling the probability of being requested of each file θ_j.

Connectivity Distribution

We assume that hubs are arranged according to a uniform two dimensional grid with spacing d and each hub has a radius of coverage r. The connectivity distribution γ_h, i.e. the probability that a user is connected to h hubs can be derived geometrically or with the Monte-Carlo method. In Appendix D, we describe the algorithm to derive the distribution via Monte-Carlo.

For our model, we consider the distance between two hubs to be equal to $d = 60$ km and the radius of coverage of each hub equal to $r = 45$ km and we obtain that the connectivity distribution is given by

$$\gamma_1 = 0.2907,\ \gamma_2 = 0.6591,\ \gamma_3 = 0.0430,\ \gamma_4 = 0.0072. \tag{6.43}$$

Zipf File Probability Distribution

We assume that the probability of file f_j being requested θ_j follows a Zipf distribution with parameter α leading to

$$\theta_j = \frac{1/j^\alpha}{\sum_{i=1}^{n} 1/i^\alpha} \quad \forall j = 1, ..., n,$$

where α denotes the skewness of the distribution. The file f_j is characterized by the ranking number j, such that $j = 1$ represents the most popular content. For $\alpha = 0$ the distribution becomes uniform.

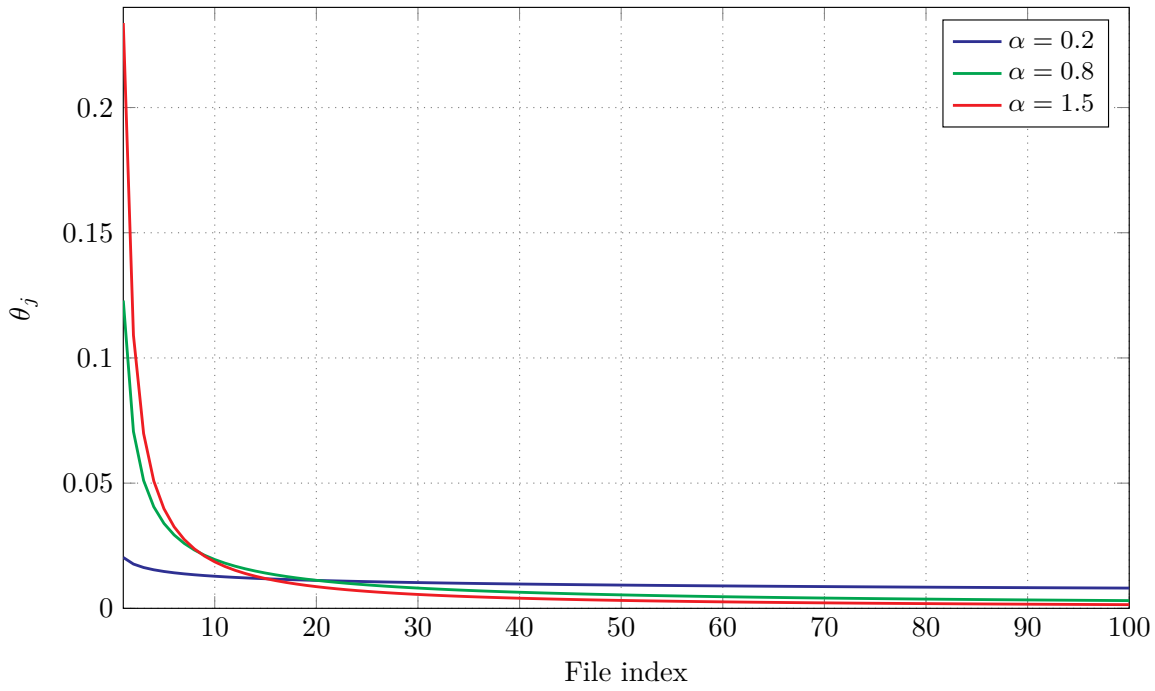


Fig. 6.2 Zipf distribution for $n = 100$ and parameter distribution $\alpha = 0.8$.

The Zipf is a distribution which models the popularity of Internet pages [58] as well as the demand of video content in the Internet [59]. As an example, in Fig. 6.2 is plotted for each file index j the probability of being requested θ_j for a library of $n = 100$ files and for three different values of α. Note that by increasing the value of α the requests tend to concentrate more and more on the most popular files.

6.6.1 LRFC Numerical Results

In this section, the normalized backhaul transmission rate defined as $\mathbb{E}[T_{\text{LRFC}}]/k$ for the LRFC caching scheme is evaluated.

Given the connectivity γ_h and the requested θ_j probability distribution for each scenario, we optimize the number of fragments w_j to be cached during the placement phase by solving the optimization problem (6.33)-(6.42). Finally, given w_j we calculate $\mathbb{E}[T_{\text{LRFC}}]/k$ from (6.26).

Bound Evaluation

Table 6.1 shows $\mathbb{E}[\Delta_{\text{LRFC}}]$ and its upper bound δ_u for different values of q. The values in the second column were numerically derived from equation (6.23) while values in the third column were derived from the bound in equation (6.25). We can observe how the bound becomes tighter for increasing q.

Table 6.1 Average overhead required for successful decoding for a LRFC with $k = 10$.

q	$\mathbb{E}[\Delta_{\text{LRFC}}]$	δ_u
2	1.6047	-
4	0.4211	0.6094
8	0.1610	0.1792
16	0.0708	0.0720
32	0.0333	0.0334
64	0.0159	0.0161
128	0.0079	0.0079

Impact of the Memory Size M

We fist show how the cache size M impacts in the average backhaul transmission rate when LRFC caching scheme is implemented. We consider two different setups. In the first setup, all users are connected to exactly one hub, i.e. $\gamma_1 = 1$ and file popularity is uniformly distributed (i.e Zipf distributed with parameter $\alpha = 0$). A library size of $n = 100$ files is assumed, being each file is fragmented into $k = 10$ input symbols is assumed.

The number of LRFC coded symbols w_j cached at each hub was optimized, by solving the problem (6.41) for $q = 2$, $q = 4$ and $q = 128$.

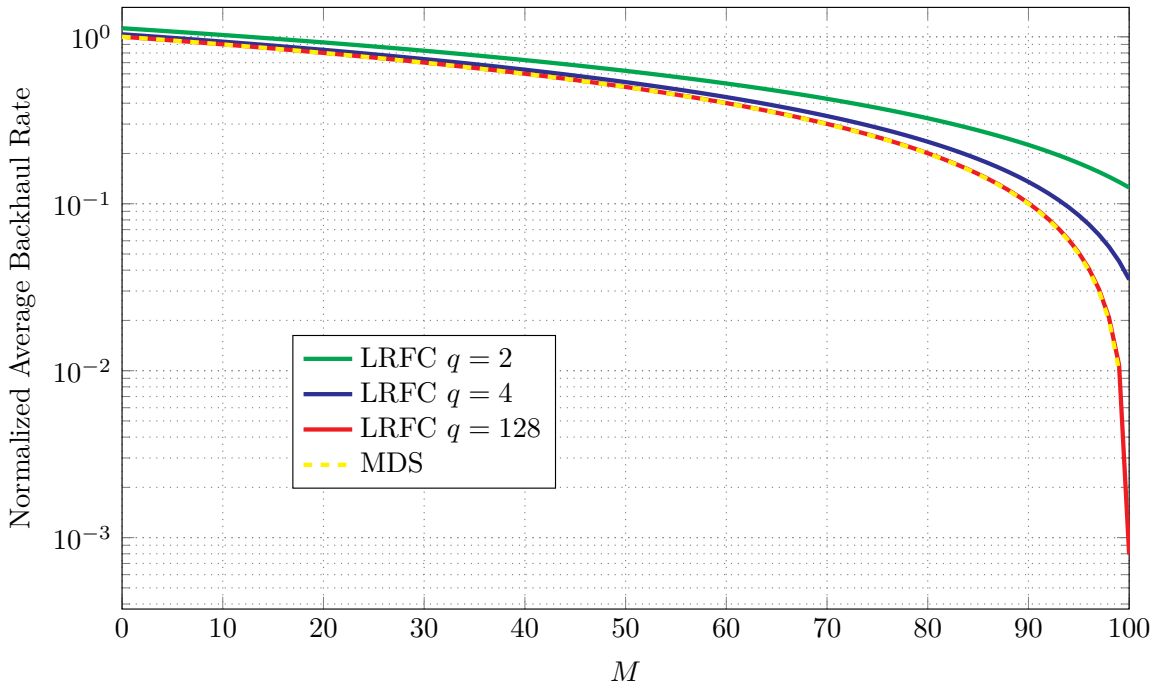


Fig. 6.3 Normalized average backhaul transmission rate as a function of memory size M for MDS and LRFC codes over $\mathbb{F}_q$ for $q = 2, 4, 128$ given $n = 100$, $k = 10$, $\alpha = 0$ and $\gamma_1 = 1$. Solid curves represent LRFC schemes while the dashed curve represents the MDS scheme.

The average backhaul transmission rate of the fountain coding caching scheme is obtained by numerically evaluating the expression (6.26). As a benchmark the MDS caching scheme from [31] was used. We evaluate the normalized backhaul transmission rate for different values of the memory size M and for different Galois field orders q.

In Fig. 6.3, the normalized average backhaul transmission rate as a function of the memory size M is shown. We can observe that the LRFC caching scheme follows the same trend of the MDS caching scheme. We further observe how the penalty on $\mathbb{E}[T_{\text{LRFC}}]/k$ for using LRFC with respect to an MDS code becomes smaller for increasing values of the q and already for $q = 128$ is almost negligible. When the cache size coincides with the library size (i.e. $M = 100$), the backhaul rate for the MDS scheme becomes zero while for the LRFC scheme the $\mathbb{E}[T_{\text{LRFC}}]/k$ coincides with the average overhead $\mathbb{E}[\Delta_{\text{LRFC}}]$.

In the second setup, we assume that users can be connected to multiple hubs according the connectivity distribution given by (6.43). The file popularity is distributed according the Zipf distribution with parameter $\alpha = 0.8$. We assume that the library

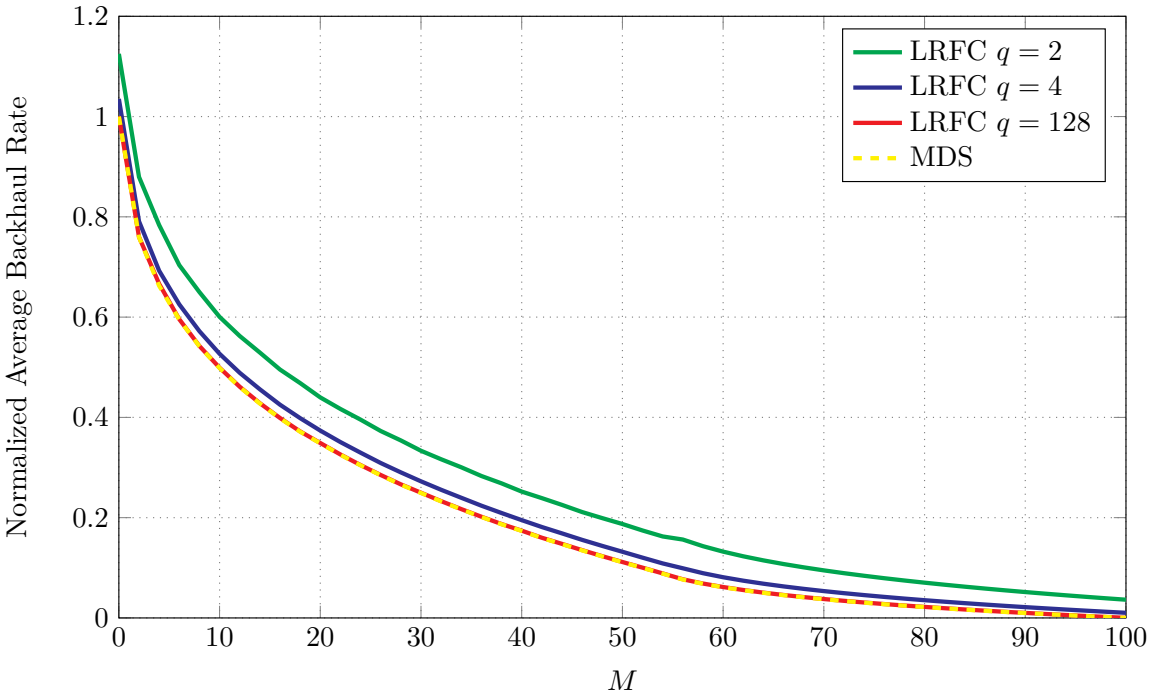


Fig. 6.4 Normalized average backhaul rate as a function of memory size M for MDS and LRFC codes over $\mathbb{F}_q$ for $q = 2, 4, 128$ given $n = 100$, $k = 10$, $\alpha = 0.8$ and $\gamma_1 = 0.2907$, $\gamma_2 = 0.6591$, $\gamma_3 = 0.0430$, $\gamma_4 = 0.0072$.

cardinality is $n = 100$ and each file is fragmented into $k = 10$ input symbols. The optimal cache placing is computed for each LRFC scheme as well as for the MDS.

In Fig. 6.8 the normalized average backhaul transmission rate for different code caching schemes as a function of the memory size M, when users can be connected to multiple hubs, is shown.

Also in this setup, it is observed that the performance of the LRFC caching scheme approaches the optimal one for large values of q. Note that since an MDS code achieves the best possible performance, this result shows implicitly that solving the optimization problem in (6.28)-(6.30), yields a solution that is close to that of solving the optimization problem in (6.31)-(6.33). We further observe that the LRFC caching scheme in a system where hubs have storage capabilities equal to 10% of the library size can reduce the average backhaul transmission rate for at least 40% (when $q = 2$) with respect to a system with no caching ($M = 0$).

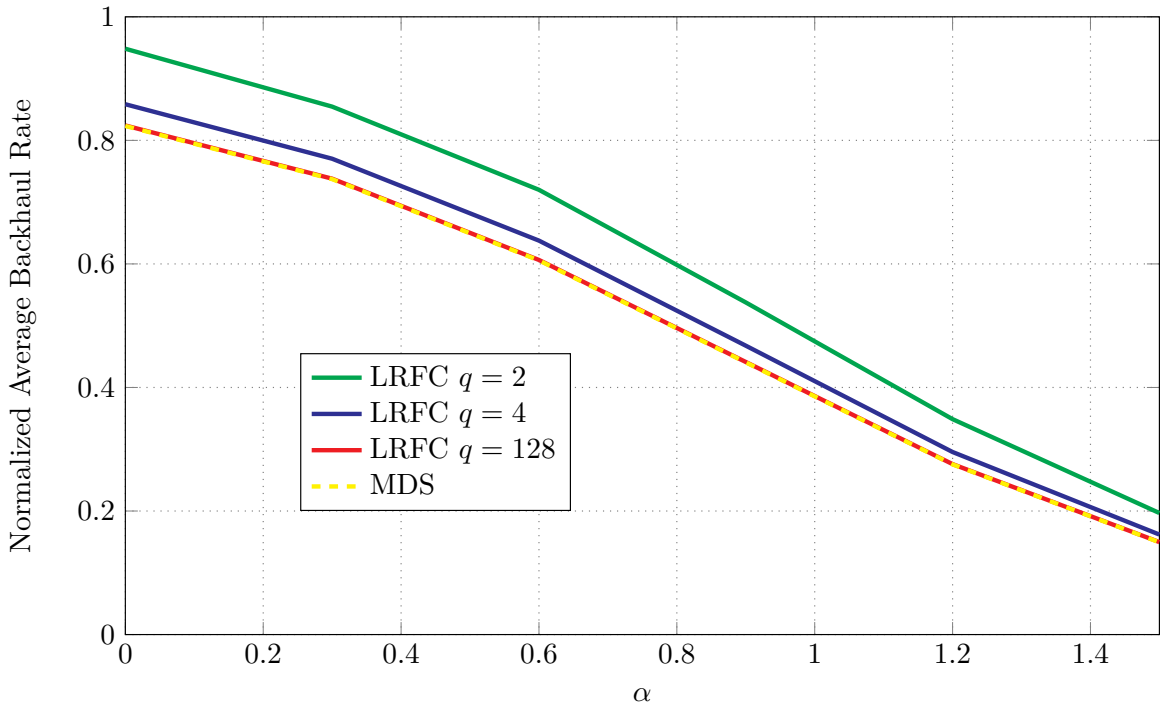


Fig. 6.5 Normalized average backhaul rate as a function of the file parameter distribution α for MDS and LRFC codes over $\mathbb{F}_q$ for $q = 2, 4, 128$ given $n = 100$, $k = 10$, $M = 10$ and $\gamma_1 = 0.2907$, $\gamma_2 = 0.6591$, $\gamma_3 = 0.0430$, $\gamma_4 = 0.0072$.

Impact of the File Distribution

The impact of the file parameter distribution α on the normalized average backhaul transmission rate in Fig. 6.5 is shown. We consider the connectivity distribution given in (6.43), a fixed memory size $M = 10$ and library cardinality $n = 100$. Because of the nature of the Zipf distribution, for increasing values of the parameter α all caching schemes become more efficient since most of requests are concentrated on a small number of files. The plot shows that for $\alpha = 0.2$, a LRFC in $\mathbb{F}_2$ requires roughly 12% more transmissions over the backhaul link than a LRFC in $\mathbb{F}_{128}$. For $\alpha = 1.5$ the LRFC of order $q = 2$ requires only 4.7% more than in $q = 128$.

Impact of the Library Size

In the following, we investigate the impact of the library size on the average backhaul transmission rate. In this setup, the Zipf distribution with parameter $\alpha = 0.8$, a fixed memory size of $M = 10$, $k = 10$ and the connectivity distribution given by (6.43) is considered. We evaluate the average backhaul transmission rate for

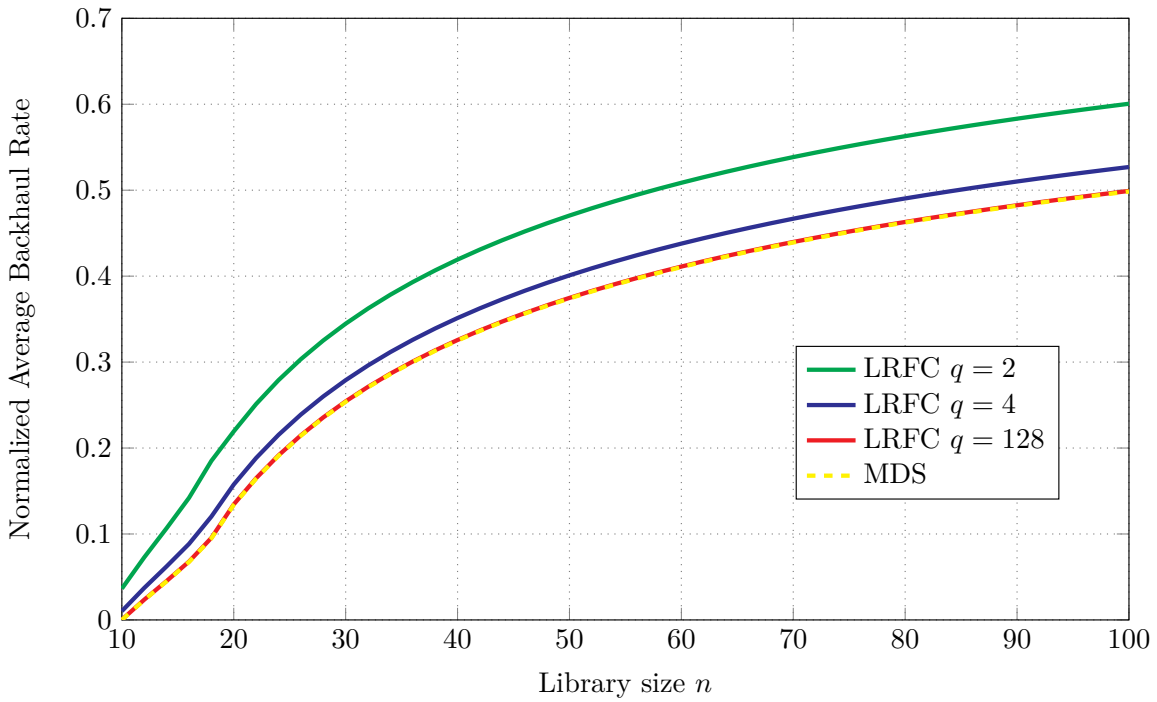


Fig. 6.6 Normalized average backhaul rate as a function of the library size n for MDS and LRFC codes over $\mathbb{F}_q$ for $q = 2, 4, 128$ given $\alpha = 0.8$, $k = 10$, $M = 10$ and $\gamma_1 = 0.2907$, $\gamma_2 = 0.6591$, $\gamma_3 = 0.0430$, $\gamma_4 = 0.0072$.

different cardinalities of the library. In Fig. 6.6 the normalized average backhaul rate is shown as a function of the library size. For a fixed memory size, the average backhaul transmission rate increases as the library size increases. As it can be observed, also in this case the proposed LRFC caching scheme performs similarly to a MDS scheme.

6.6.2 LT Numerical Results

We consider the system described in Section 6.2 which is illustrated in Fig. 6.1. In such scenario, we evaluate the normalized backhaul transmission rate defined as $\mathbb{E}[T_{\mathrm{LT}}]/k$ for the LT caching scheme.

Given the connectivity γ_h and the requested θ_j probability distribution for each scenario, we optimize the number of fragments w_j to be cached during the placement phase by solving the optimization problem (6.40)-(6.42). Finally, given w_j we calculate $\mathbb{E}[T_{\mathrm{LT}}]/k$ from (6.35). In all setups, we consider that each file is fragmented in $k = 10000$ input symbols.

LT Probability of Decoding Failure

We assume that the GEO implements an LT code characterized by a RSD [55] (Appendix C) with parameter $c = 0.05$ and $d = 3$, where c and d have been chosen so that the average overhead is minimized. The probability of decoding failure $P_{\mathrm{LT}}(\delta)$ was derived as explained in Section 5.4.2.

We consider the probability of failure as a function of δ for RSD under iterative peeling decoding given in Fig. 5.5. From (6.34), we obtained that the overhead average is

$$\mathbb{E}[\Delta_{\mathrm{LT}}] = 431.95.$$

Impact of the Memory Size

We first show how the cache size M impacts in the average backhaul transmission rate when LT caching scheme is implemented. We consider two different setups.

In our first setup, all users are connected to exactly one hub, i.e. $\gamma_1 = 1$ and file popularity is uniformly distributed (i.e., Zipf distributed with parameter $\alpha = 0$). A library size of $n = 100$ where each file is fragmented into $k = 10000$ input symbols is assumed.

The average backhaul transmission rate of the LT caching scheme can be obtained by numerically evaluating the expression (6.35). As a benchmark the MDS caching scheme from [31] was used. We evaluate the normalized backhaul transmission rate for different values of the memory size M.

In Fig. 6.7, the normalized average backhaul transmission rate as a function of the memory size M is shown. We can observe that the caching scheme based on LT codes performs close to the caching scheme based on MDS. The cost in terms of average

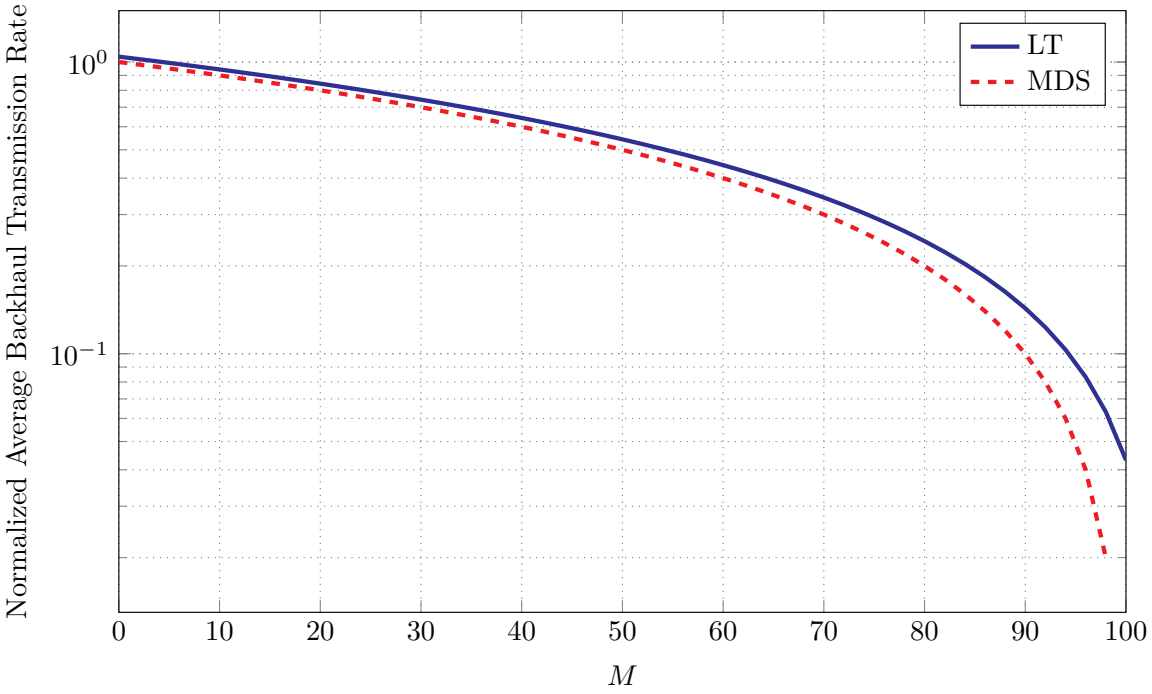


Fig. 6.7 Normalized average backhaul transmission rate as a function of memory size M for LT and MDS codes given $n = 100$, $k = 10000$, $\alpha = 0$ and $\gamma_1 = 1$.

transmission rate for using a LT code is only the 4.32%. This quantity coincides with the average overhead of the LT code needed for successfully decode the content.

In our second setup, we consider the connectivity distribution given by (6.43). We assume that files are requested according to a Zipf distribution with $\alpha = 0.8$.

In Fig. 6.8, the normalized average backhaul transmission rate is shown as a function of the cache size M. The plot shows that the LT caching scheme is comparable to the optimal caching scheme given by the MDS codes. We further observe that, as the cache size increases, the difference between the two approaches becomes negligible. Due to the low probability of being requested of the files that are not cached.

Impact of the File Distribution

In the following, we show how the probability file distribution impacts in the average backhaul transmission rate when the LT caching scheme is implemented.

We consider the connectivity given by (6.43). A fixed memory size $M = 10$ and library size $n = 100$ files is assumed. In Fig. 6.9, the normalized average backhaul transmission rate is shown as a function of the shape parameter α. As expected, when files are equiprobable ($\alpha = 0$), the MDS caching scheme outperforms the scheme based

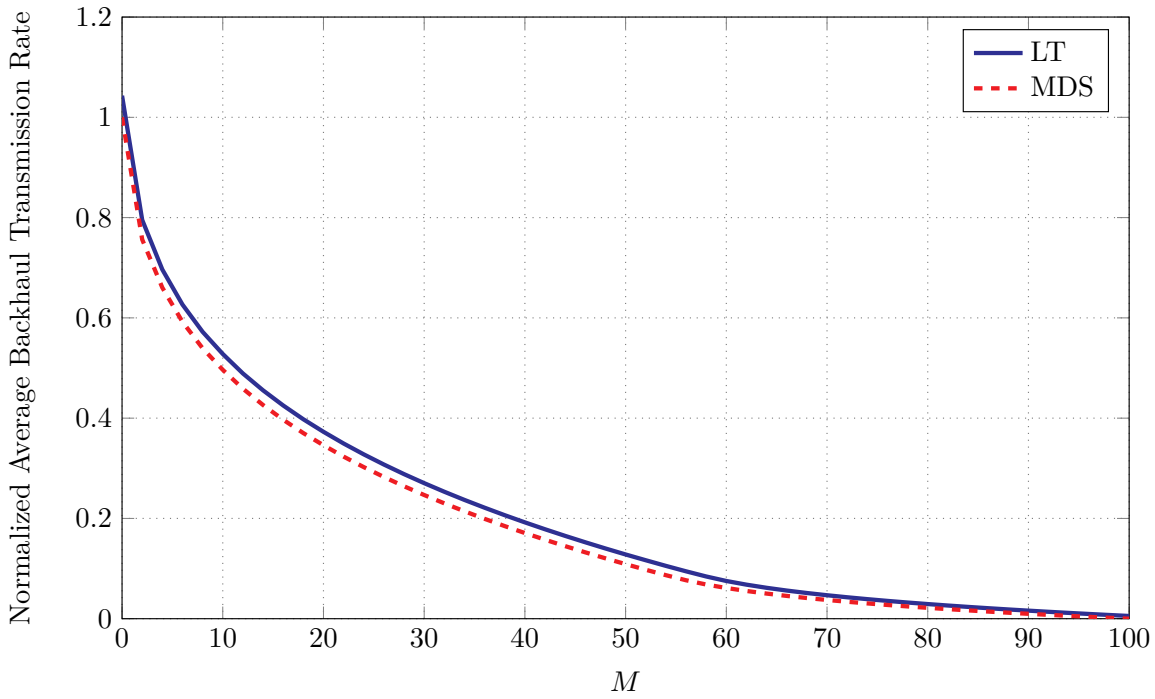


Fig. 6.8 Normalized average backhaul transmission rate as a function of memory size M for LT and MDS codes given $n = 100$, $k = 10000$, $\alpha = 0.8$ and $\gamma_1 = 0.2907, \gamma_2 = 0.6591, \gamma_3 = 0.0430$ and $\gamma_4 = 0.0072$.

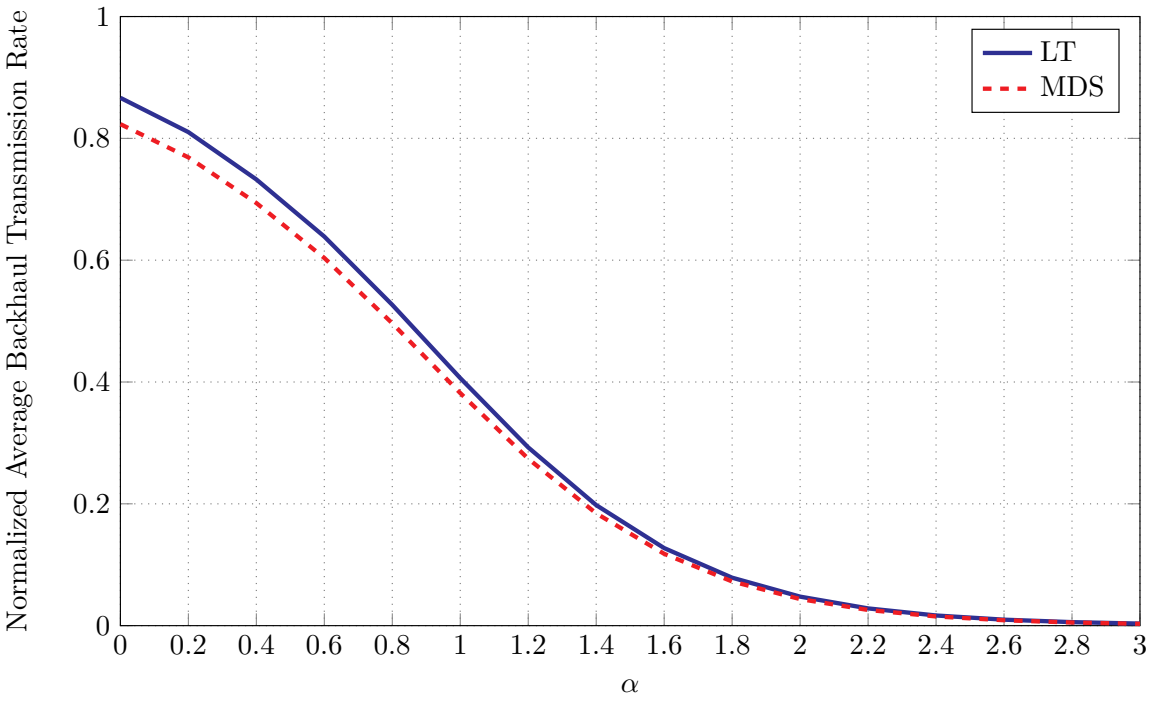


Fig. 6.9 Normalized average backhaul transmission rate as a function of the file parameter distribution α for LT and MDS given $n = 100$, $k = 10000$, $M = 10$ and $\gamma_1 = 0.2907$, $\gamma_2 = 0.6591$, $\gamma_3 = 0.0430$, $\gamma_4 = 0.0072$.

on LT codes. However as α increases, the performance of the LT scheme approaches that of the MDS scheme.

Impact of Library Size

In the following, we investigate the impact of the library size on the average backhaul rate. In this setup, we consider the Zipf distribution with parameter $\alpha = 0.8$, a fixed memory size of $M = 10$, $k = 10000$ and the connectivity distribution given by (6.43). We evaluate the average backhaul transmission rate for different cardinalities of the library.

In Fig. 6.10 the normalized average backhaul transmission rate is shown as a function of the library size. For a fixed memory size, the average backhaul transmission rate increases as the library size increases. As it can be observed, also in this case the proposed LT caching scheme performs similarly to a MDS scheme.

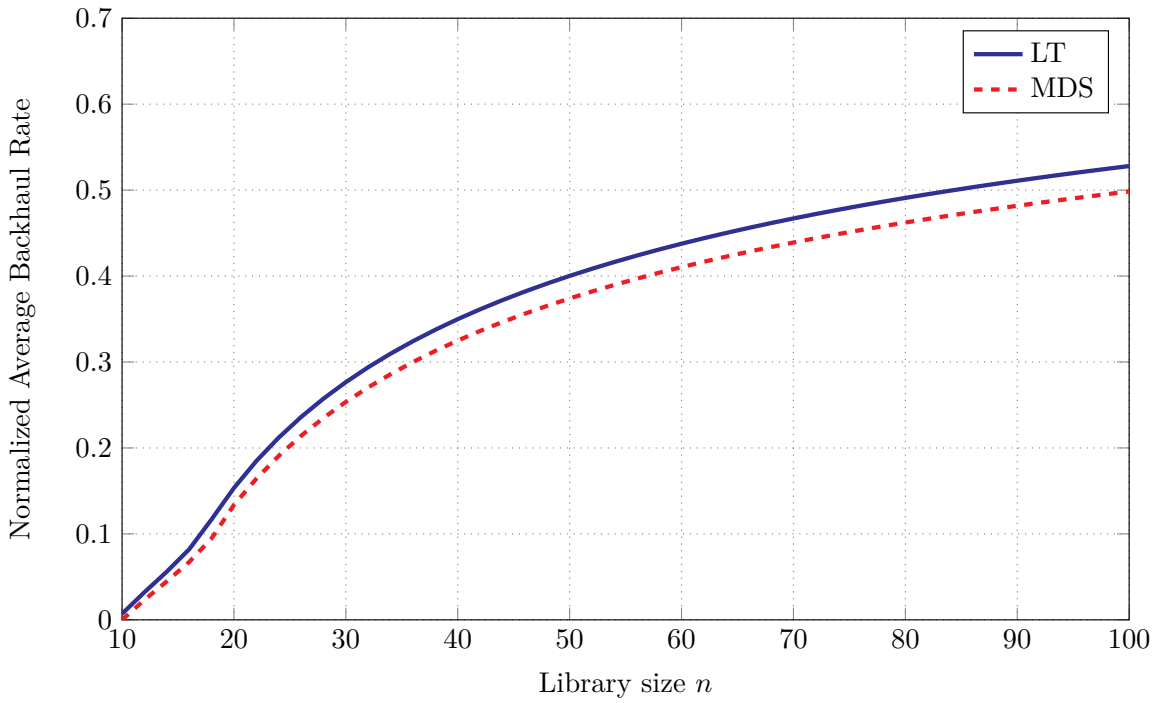


Fig. 6.10 Normalized average backhaul transmission rate as a function of of the library size n for LT and MDS given $\alpha = 0.8$, $k = 10000$, $M = 10$ and $\gamma_1 = 0.2907$, $\gamma_2 = 0.6591$, $\gamma_3 = 0.0430$, $\gamma_4 = 0.0072$.

6.7 Conclusion

In this chapter, we considered caching schemes based on rate-less code for caching content at the edge in a satellite heterogeneous network. We evaluated the performance of the schemes in terms of average backhaul transmission rate. Thus, with the aim of minimizing the number of transmission from the core of the network during the delivery phase, we formulated the optimization placement problem for the LRFC caching scheme as well for the LT caching scheme. We derived for each scheme the analytical expression of the average backhaul transmission rate and its tight upper bound.

First, we presented simulation results where we compared the performance of LRFC scheme with an MDS scheme. Our results indicate that the performance of the simple LRFC caching scheme built over a finite field of moderate order approaches that of the optimal caching scheme base on MDS. The LRFC caching schemes are characterized by the possibility to generate an unlimited amount of encoded output symbols, but they are still impractical because of its high decoding complexity.

Thus, we analyzed the LT caching scheme and we evaluated the LT caching scheme under iterative decoding. We compared the performance of the LT caching scheme with that of an optimal MDS scheme. Our simulation results indicate that the performance of LT codes approaches that of the optimal scheme, but exhibiting a much lower encoding and decoding complexity.

Chapter 7

Final Conclusions

In this dissertation we have investigated caching techniques for heterogeneous networks from two different perspectives.

On the one hand, we have investigated the cooperative techniques to deliver cached content over Gaussian interference channels under quality of service constraints in heterogeneous networks. As a first step, we have studied and analyzed cooperative and non-cooperative delivery techniques in Gaussian interference channels. We formulated the caching problem aiming at minimizing the average power consumption when each content has a different probability to be requested and, from the physical layer point of view, when each content has a different transmission rate to satisfy and one-shot delivery phase is considered. Three different approaches were investigated: (i) no cooperation, (ii) limited cooperation and (iii) full cooperation. We proposed an algorithm to find the best cache allocation and a low-complexity sub-optimal algorithm as well. Our simulation results indicate that memorizing content with higher probabilities is not always the most convenient strategy, but rate transmission constraints should be taken into account.

On the other hand, we have investigated caching schemes based on rate-less codes in heterogeneous networks. In particular, we first proposed to cache and deliver encoded content using LRFCs. We formulated the placement optimization problem with the objective of minimizing the backhaul transmissions. Our simulation results show that the LRFC caching scheme performance approaches that of the optimal caching scheme based on MDS codes. However, the application in real world scenario remains questionable for both schemes, mainly because of their high encoding and decoding complexity. A more practical caching scheme based on LT codes was analyzed next. The design as well the optimization placement problem of the LT codes caching scheme

were derived. Our results show that the caching scheme based on LT can achieve performance very close to the optimal caching scheme with much lower encoding and decoding complexity.

References

[1] V. Wong, R. Schober, D. Kwan Ng, and L. Wang, *Key techonologies for 5G wireless systems*. Cambridge, U.K.: Cambridge Univ. Press., 2017.

[2] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *2012 Proceedings IEEE INFOCOM*, March 2012, pp. 1107–1115.

[3] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, September 2008.

[4] C.-L. I, L. J. Greenstein, and R. D. Gitlin, "A microcell/macrocell cellular architecture for low- and high-mobility wireless users," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 885–891, Aug 1993.

[5] V. Chandrasekhar and J. G. Andrews, "Uplink capacity and interference avoidance for two-tier femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3498–3509, July 2009.

[6] L. T. W. Ho and H. Claussen, "Effects of user-deployed, co-channel femtocells on the call drop probability in a residential scenario," in *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2007, pp. 1–5.

[7] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 136–143, December 2014.

[8] M. Gregori, J. Gómez-Vilardebò, J. Matamoros, and D. Gündüz, "Joint transmission and caching policy design for energy minimization in the wireless backhaul link," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1004–1008.

[9] J. Liu and S. Sun, "Energy efficiency analysis of dense small cell networks with caching at base stations," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016, pp. 2944–2948.

[10] A. C. Güngör and D. Gündüz, "Proactive wireless caching at mobile user devices for energy efficiency," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 186–190.

[11] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Exploiting caching and multicast for 5g wireless networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2995–3007, April 2016.

[12] F. Gabry, V. Bioglio, and I. Land, "On energy-efficient edge caching in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3288–3298, Dec 2016.

[13] M. M. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 171–175.

[14] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[15] E. Bastug, J. Guénégo, and M. Debbah, "Proactive small cell networks," in *ICT 2013*, May 2013, pp. 1–5.

[16] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug 2014.

[17] N. Zhao, F. Cheng, F. R. Yu, J. Tang, Y. Chen, G. Gui, and H. Sari, "Caching uav assisted secure transmission in hyper-dense networks based on interference alignment," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2281–2294, May 2018.

[18] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1046–1061, May 2017.

[19] H. Wu, J. Li, H. Lu, and P. Hong, "A two-layer caching model for content delivery services in satellite-terrestrial networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[20] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[21] E. Recayte and G. Cocco, "Caching in heterogeneous networks with per-file rate constraints," *accepted to: IEEE Transactions on Communications*, Oct 2018. [Online]. Available: https://arxiv.org/pdf/1810.06475.pdf

[22] E. Recayte, F. Lázaro, and G. Liva, "Caching at the edge with fountain codes," in *Proc. 9th Advanced Satellite Mobile Systems Conference*, Berlin, Sept 2018. [Online]. Available: https://arxiv.org/pdf/1807.05619.pdf

[23] E. Recayte, G. Cocco, and A. Vanelli-Coralli, "Caching in gaussian interference channel with qos constraints," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[24] E. Recayte, F. Lázaro, and G. Liva, "Caching at the edge with LT codes," in *Proc. International Symposyum on Turbo Codes*, Hong Kong, Dec 2018. [Online]. Available: https://arxiv.org/pdf/1810.07081.pdf

[25] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 809–813.

[26] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 3092–3107, May 2017.

[27] J. S. P. Roig, D. Gündüz, and F. Tosato, "Interference networks with caches at both ends," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[28] Y. Cao, F. Xu, K. Liu, and M. Tao, "A storage-latency tradeoff study for cache-aided mimo interference networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[29] J. Hachem, U. Niesen, and S. N. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5359–5380, July 2018.

[30] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, Dec 2013.

[31] V. Bioglio, F. Gabry, and I. Land, "Optimizing MDS codes for caching at the edge," in *Proc. IEEE Globecomm*, San Diego, U.S.A., Dec. 2015.

[32] J. Liao, K. Wong, Y. Zhang, Z. Zheng, and K. Yang, "Mds coded cooperative caching for heterogeneous small cell networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–7.

[33] F. Zhang, Y. Sun, Z. Hao, P. Si, R. Yang, Y. Zhang, M. Yu, and M. Yu, "A mds-based caching strategy for heterogeneous network," in *2018 10th International Conference on Communication Software and Networks (ICCSN)*, July 2018, pp. 184–188.

[34] E. Ozfatura and D. Gündüz, "Mobility and popularity-aware coded small-cell caching," *IEEE Communications Letters*, vol. 22, no. 2, pp. 288–291, Feb 2018.

[35] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York: Wiley, 2006, chapter 15.

[36] A. Carleial, "Interference channels," *Information Theory, IEEE Transactions on*, vol. 24, no. 1, pp. 60–70, jan 1978.

[37] S. T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *Information Theory, IEEE Transactions on*, vol. 27, no. 1, pp. 49–60, jun 1981.

[38] T. Cover, "Broadcast channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 2–14, jan 1972.

[39] C. E. Shannon, "Two-way communication channels," in *Berkeley Symp. on Mathematical Statistics and Probability*, 1961.

[40] H. Sato, "Two-user communication channel," *Information Theory, IEEE Transactions on*, vol. 23, no. 3, pp. 295–304, may 1977.

[41] J. Jian, Y. Xin, and G. H., "Interference channels with common information," *Information Theory, IEEE Transactions on*, vol. 54, no. 1, pp. 171–187, jan 2008.

[42] M. Costa, "Writing on dirty paper," *Information Theory, IEEE Transactions on*, vol. 29, no. 3, pp. 439–441, may 1983.

[43] H. Weingarten, Y. Steinberg, and H. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *Information Theory, IEEE Transactions on*, vol. 9, no. 52, pp. 3936–3964, sept 2006.

[44] D. Tse and P. Viswanath, *Fundamentals of Wireless Comm.*, 1st ed. UK: Cambridge, 2005.

[45] S. Boyd, *Convex Optimization*. Eds. Cambridge, UK: Cambridge Univ. Press., 2004, vol. 5.

[46] J. Oh, S. Kim, R. Narasimhan, and J. Cioffi, "Transmit power optimization for Gaussian vector broadcast channels," in *Proc. 2005 IEEE International Conf. on Commun.*, Seoul, South Korea, Aug. 2005.

[47] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658 – 2668, Oct. 2003.

[48] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan 2004.

[49] M. M. Azari, F. Rosas, K. C. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity„" *Information Theory, IEEE Transactions on*, vol. 66, no. 1, pp. 330–344, jan 2018.

[50] C. F. Lo, "The sum and difference of two lognormal random variables," *Journal of the Society for Industrial & Applied Mathematics*, vol. 2012, pp. 1–13, 2012.

[51] G. Lansberg, "Uber eine anzahlbestimmung und eine damit zusammenhangende reihe," *Journal fur die reine und angewandte Mathematic*, vol. 3, pp. 87–88, 1893.

[52] R. Lidl and H. Niederreiter, *Finite fields*. Cambridge Univ. Press, 1997.

[53] R. Singleton, "Maximum distance q-nary codes," vol. 10, no. 2, pp. 116–118, 1964.

[54] G. Liva, E. Paolini, and M. Chiani, "Performance versus overhead for fountain codes over $\mathbb{F}_q$," *IEEE Comm. Letters.*, vol. 14, no. 2, pp. 178–180, 2010.

[55] M. Luby, "LT codes," in *Proc. 43rd Annual IEEE Symp. on Foundations of Computer Science*, Vancouver, Canada, Nov. 2002, pp. 271–282.

[56] R. Karp, M. Luby, and A. Shokrollahi, "Finite length analysis of LT codes," in *Proc. 2004 IEEE International Symp. on Inf. Theory*, Chicago, Illinois, US, Jun. 2004.

[57] A. Shokrollahi, "Theory and applications of Raptor codes," *Mathknow*, vol. 3, pp. 59–89, 2009.

[58] L. A. Adamic and B. Huberman, "Zipf's law and the internet," vol. 3, 11 2001.

[59] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," vol. 17, Oct 2009.

Appendix A

KKT conditions for GIC-C power minimizaton

The optimization problem can be rewritten as follows

$$\underset{P_i,P_j}{\text{minimize}} \ \frac{\tilde{P}_i}{a_{11}} + \frac{\tilde{P}_j}{a_{22}}, \tag{A.1}$$

$$\text{subject to } -\sqrt{2^{2R_i}-1} + \sqrt{\tilde{P}_i} + \sqrt{c_{12}\cdot\tilde{P}_j} \geq 0, \tag{A.2}$$

$$-\sqrt{2^{2R_i}-1} + \sqrt{\tilde{P}_j} + \sqrt{c_{21}\cdot\tilde{P}_i} \geq 0, \tag{A.3}$$

$$\tilde{P}_i, \tilde{P}_j \geq 0. \tag{A.4}$$

The KKT conditions to be satisfied are:

$$\begin{aligned} \nabla_{\mathbf{P}} L(\mathbf{P}^*, \lambda^*) &= 0, \\ g_m(\mathbf{P}^*) &\geq 0, \quad m = 1, ..., n \\ \lambda_m^* &\geq 0, \quad m = 1, ..., n \\ \lambda_m^* \cdot g_m(\mathbf{P}^*) &= 0, \quad m = 1, ..., n. \end{aligned}$$

Let us indicate with $\mathbf{P}^* = [P_i^*, P_j^*]$ the solution and let $L(\mathbf{P}^*, \lambda^*)$ be the Lagrange function associated with the optimization problem

$$\begin{aligned} L(\mathbf{P}^*, \lambda) = -\frac{\tilde{P}_i^*}{a_{11}} - \frac{\tilde{P}_j^*}{a_{22}} + \lambda_1 \left(-\sqrt{2^{2R_i}-1} + \sqrt{\tilde{P}_i^*} + \sqrt{c_{12}\cdot\tilde{P}_j^*}\right) + \\ \lambda_2 \left(-\sqrt{2^{2R_i}-1} + \sqrt{\tilde{P}_j} + \sqrt{c_{21}\cdot\tilde{P}_i}\right) + \lambda_3 \cdot \tilde{P}_i^* + \lambda_4 \cdot \tilde{P}_j^*. \end{aligned} \tag{A.5}$$

Let us consider the points where functions $g_1, ..., g_m$ are differentiable. Applying the KKT conditions we have

$$\frac{\delta L(\mathbf{P}^*, \lambda)}{\delta \tilde{P}_i} = -\frac{1}{a_{11}} + \frac{\lambda_1^*}{2}\frac{1}{\sqrt{\tilde{P}_i^*}} + \sqrt{c_{21}}\frac{\lambda_2^*}{2}\frac{1}{\sqrt{\tilde{P}_i^*}} + \lambda_3 = 0, \tag{A.6}$$

$$\frac{\delta L(\mathbf{P}^*, \lambda)}{\delta \tilde{P}_j} = -\frac{1}{a_{22}} + \sqrt{c_{12}}\frac{\lambda_1^*}{2}\frac{1}{\sqrt{\tilde{P}_i^*}} + \frac{\lambda_2^*}{2}\frac{1}{\sqrt{\tilde{P}_j^*}} + \lambda_4 = 0, \tag{A.7}$$

$$g_1(\tilde{P}_i^*, \tilde{P}_j^*) = -\sqrt{2^{2R_i} - 1} + \sqrt{\tilde{P}_j^*} + \sqrt{c_{12} \cdot \tilde{P}_i^*} \geq 0, \tag{A.8}$$

$$g_2(\tilde{P}_i^*, \tilde{P}_j^*) = -\sqrt{2^{2R_i} - 1} + \sqrt{\tilde{P}_i^*} + \sqrt{c_{21} \cdot \tilde{P}_j^*} \geq 0, \tag{A.9}$$

$$g_3(\tilde{P}_i^*, \tilde{P}_j^*) = \tilde{P}_i^* \geq 0, \tag{A.10}$$

$$g_4(\tilde{P}_i^*, \tilde{P}_j^*) = \tilde{P}_j^* \geq 0, \tag{A.11}$$

$$\lambda_1^* \geq 0, \tag{A.12}$$

$$\lambda_2^* \geq 0, \tag{A.13}$$

$$\lambda_3^* \geq 0, \tag{A.14}$$

$$\lambda_4^* \geq 0, \tag{A.15}$$

$$\lambda_1^* \cdot g_1(\tilde{P}_i^*, \tilde{P}_j^*) = \lambda_1^* \left(-\sqrt{2^{2R_i} - 1} + \sqrt{\tilde{P}_j^*} + \sqrt{c_{12} \cdot \tilde{P}_i^*}\right) = 0, \tag{A.16}$$

$$\lambda_2^* \cdot g_2(\tilde{P}_i^*, \tilde{P}_j^*) = \lambda_2^* \left(-\sqrt{2^{2R_i} - 1} + \sqrt{\tilde{P}_i^*} + \sqrt{c_{21} \cdot \tilde{P}_j^*}\right) = 0, \tag{A.17}$$

$$\lambda_3^* \cdot g_3(\tilde{P}_i^*, \tilde{P}_j^*) = \lambda_3^* \cdot \tilde{P}_i^* = 0, \tag{A.18}$$

$$\lambda_4^* \cdot g_4(\tilde{P}_i^*, \tilde{P}_j^*) = \lambda_4^* \cdot \tilde{P}_j^* = 0. \tag{A.19}$$

From eqs. (A.6), (A.7), (A.18) and (A.19) follows that $\lambda_3 = 0$ and $\lambda_4 = 0$ then from eqs. (A.6) and (A.7) we find

$$\tilde{P}_i^* = \frac{1}{4}\left(\lambda_1^* \cdot a_{11} + \lambda_2^* \cdot a_{11} \cdot \sqrt{c_{21}}\right)^2, \tag{A.20}$$

$$\tilde{P}_j^* = \frac{1}{4}\left(\lambda_1^* \cdot a_{22} \cdot \sqrt{c_{12}} + \lambda_2^* \cdot a_{22}\right)^2. \tag{A.21}$$

By plugging in eqs. (A.16) and (A.17) the results of eqs. (A.20) and (A.21), we get the following system of equations

$$\lambda_1^*\left[\lambda_1^*(a_{11} + a_{22}c_{12}) + \lambda_2^*(a_{11}\sqrt{c_{21}} + a_{22}\sqrt{c_{12}}) - 2\sqrt{2^{2R_i} - 1}\right] = 0, \tag{A.22}$$

$$\lambda_2^*\left[\lambda_1^*(a_{22}\sqrt{c_{12}} + a_{11}\sqrt{c_{21}}) - 2\sqrt{2^{2R_i} - 1} + \lambda_2^*(a_{22} + a_{11}c_{21})\right] = 0. \tag{A.23}$$

Now we need to find the quadruples (λ_1, λ_2, λ_3, λ_4) which solve the system composed by equations eqs. (A.6), (A.7), (A.18), (A.19), (A.22) and (A.23). We obtain the following three solutions

$$S_1 = \left\{ \frac{2\sqrt{2^{2R_i}-1} - \lambda_2(a_{11}\sqrt{c_{21}} + a_{22}\sqrt{c_{12}})}{a_{11} + c_{12}a_{22}}, \frac{2\sqrt{2^{2R_i}-1}(1 - \frac{\sqrt{c_{12}}a_{22} + \sqrt{a_{21}}a_{11}}{a_{11}+a_{12}a_{22}})}{a_{22} + c_{21}a_{11} - \beta}, 0, 0 \right\}, \tag{A.24}$$

$$S_2 = \left\{ \frac{2\sqrt{2^{2R_i}-1}}{a_{11} + a_{22}c_{12}}, 0, 0, 0 \right\}, \tag{A.25}$$

$$S_3 = \left\{ 0, \frac{2\sqrt{2^{2R_i}-1}}{a_{22} + c_{21}a_{11}}, 0, 0 \right\}, \tag{A.26}$$

where

$$\beta = \frac{(a_{11}\sqrt{c_{21}} + a_{22}\sqrt{c_{12}})(\sqrt{c_{12}}a_{22} + \sqrt{c_{21}}a_{11})}{a_{11} + c_{12}a_{22}}. \tag{A.27}$$

By plugging S_1, S_2 or S_3 in eqs. (A.8) to (A.15) we obtain at most three feasible points (P_i^*, P_j^*). We also need to consider pairs of values $(\tilde{P}_i, \tilde{P}_j)$ where the constraint functions are not differentiable. Such pairs are

$$A = (0, \tilde{P}_j), \tag{A.28}$$

$$B = (\tilde{P}_i, 0), \tag{A.29}$$

$$C = (0, 0). \tag{A.30}$$

Note that the point $C = (0, 0)$ cannot give a feasible solution due to the fact that in eq. (A.1) we would have $\sqrt{2^{2 \cdot R_i} - 1} < 0$.
Let us calculate the value of $\tilde{P}_i$ and $\tilde{P}_j$ in A and B. In point A and with $\tilde{P}_i = 0$ we have

$$\begin{cases} \frac{1}{2}\log_2\left(1 + c_{12} \cdot \tilde{P}_j\right) & \geq R_i \\ \frac{1}{2}\log_2\left(1 + \tilde{P}_j\right) & \geq R_i, \end{cases} \tag{A.31}$$

then

$$A = \left(0, \max\left\{2^{2R_i} - 1, \frac{2^{2R_i} - 1}{c_{12}}\right\}\right) \tag{A.32}$$

and $\tilde{P}_j = 0$ for point B

$$B = \left(\max \left\{ 2^{2R_i} - 1, \frac{2^{2R_i} - 1}{c_{21}} \right\}, 0 \right). \tag{A.33}$$

Thus, the minimum power will be among A, B and the feasible points $(\tilde{P}_i^*, \tilde{P}_j^*)$ obtained by inserting S_1, S_2 and S_3 in eqs. (A.8) to (A.15).
The cost of this channel is

$$c_{GIC-C}(i,j) = \tilde{P}_i + \tilde{P}_j,$$

where $\tilde{P}_i + \tilde{P}_j$ are derived from the previous results.

A.0.0.1 BC Power Minimization

The power optimization problem in the BC can be written as follows

$$\underset{P_-, P_+}{\text{minimize}} \; P_- + P_+, \tag{A.34}$$

$$\text{subject to } \frac{1}{2} \log_2 \left(1 + a_+ \cdot P_+\right) \geq R_+, \tag{A.35}$$

$$\frac{1}{2} \log_2 \left(1 + \frac{a_- \cdot P_-}{1 + a_- \cdot P_+} \right) \geq R_-, \tag{A.36}$$

$$P_-, P_+ \geq 0, \tag{A.37}$$

The overall problem can be solved analytically as follows

$$P_+ = \frac{2^{2R_+} - 1}{a_{+0}}, \tag{A.38}$$

$$P_- = \frac{(2^{2R_-} - 1)(1 + a_{-0} \cdot P_+)}{a_{-0}}. \tag{A.39}$$

Plugging P_+ in the expression of P_-, we obtain

$$P_- = \frac{1}{a_-}(2^{2R_-} - 1) \left(1 + \frac{a_-}{a_+}(2^{2R_+} - 1) \right). \tag{A.40}$$

The overall cost of this channel $c_{BC}(i,j)$ is

$$c_{BC}(i,j) = \frac{2^{2R_+} - 1}{a_+} + \frac{(2^{2R_-} - 1)(1 + a_- P_+)}{a_-}. \tag{A.41}$$

Appendix B

MIMO-BC

In Section 4.2.1.5, we derived that minimizing the power in the MIMO-BC correspond to solve the following problem [46]

$$\text{find } \mathbf{S}_i^M \tag{B.1}$$

$$\text{subject to } \mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^T + \mathbf{N}_i = \tilde{\lambda}_i^{KKT}\mathbf{P}_i \tag{B.2}$$

$$\frac{1}{2}\log_2 \frac{|\mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^T + \mathbf{N}_i|}{|\mathbf{N}_i|} = R_i \tag{B.3}$$

The problem in B.1-B.3 has not a closed-form solution. However, an iterative method can be applied for solving it based on the water-filling algorithm.
Let us recall the following SVD

$$\mathbf{N}_i = \mathbf{Q}_i\mathbf{\Delta}_i\mathbf{Q}_i^T$$
$$\hat{\mathbf{H}}_i = \mathbf{\Delta}_i^{-\frac{1}{2}}\mathbf{Q}_i^T\mathbf{H}_i$$

then the constraints B.2-B.3 are equivalent to

$$\hat{\mathbf{H}}_i\mathbf{S}_i\hat{\mathbf{H}}_i^T = \tilde{\lambda}_i^{KKT}\hat{\mathbf{P}}_i \tag{B.4}$$

$$\frac{1}{2}\log_2 |\hat{\mathbf{H}}_i\mathbf{S}_i\hat{\mathbf{H}}_i^T + \mathbf{I}| = R_i \tag{B.5}$$

where

$$\hat{\mathbf{P}}_i = (\mathbf{Q}_i \Delta_i^{\frac{1}{2}})^T \mathbf{P}_i (\mathbf{Q}_i \boldsymbol{\Delta}_i^{\frac{1}{2}}).$$

We can decompose $\hat{\mathbf{P}}_i$ as

$$\hat{\mathbf{P}}_i = \mathbf{F}_i \boldsymbol{\Sigma}_i \mathbf{F}_i^T$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with eigenvalues σ_{i1} and σ_{i2}. The aim is to find $\mathbf{S}_i$ and $\tilde{\lambda}_i^{KKT}$ which simultaneously satisfy conditions B.4 and B.5.

Water-filling technique can be applied in order to obtain $\tilde{\lambda}_i^{KKT}$ under constraint R_i. The water-filling level $\tilde{\lambda}_i^{KKT}$ and the rate vector $[r_{i1}, r_{i2}]$ such that

$$\begin{cases} \frac{2^{2r_{ij}}}{\sigma_{ij}} = \tilde{\lambda}_i^{KKT} & \text{if } \frac{1}{\sigma_{ij}} < \tilde{\lambda}_i^{KKT}, \\ r_{ij} = 0 & \text{if } \frac{1}{\sigma_{ij}} \geq \tilde{\lambda}_i^{KKT}, \end{cases}$$

where $r_{i1} + r_{i2} = R_i$ and if we define

$$\hat{\mathbf{S}}_i^M = \begin{bmatrix} 2^{2r_{i1}} & 0 \\ \\ 0 & 2^{2r_{i2}} \end{bmatrix}$$

the optimum covariance is given by

$$\mathbf{S}_i^M = \hat{\mathbf{H}}_i^{-1} (\mathbf{F}_i \hat{\mathbf{S}}_i^M \mathbf{F}_i^{-1} - I)(\hat{\mathbf{H}}_i^{-1})^T.$$

Appendix C

Degree Distribution for LT codes

C.1 Ideal Soliton Distribution

The degree distribution for the ISD is given by

$$\Omega_d^{ISD} = \begin{cases} \frac{1}{k} & d = 1 \\ \frac{1}{d \cdot (d-1)} & 1 < d \leq k \end{cases} \tag{C.1}$$

This distribution is obtained by imposing an expected ripple size of one across the whole decoding process. However, this degree distribution presents poor performance because variations in the expected behavior make most likely to not have a one degree output symbol and block the decoding process.

C.2 Robust Soliton Distribution

The robust soliton distribution tackle the issue of the ideal soliton. In fact, based on the design of the ISD, the RSD aims at having the expected ripple size relatively large at each step of the decoding process with high probability. This property guarantees that the decoding process not get block before its ending.

In the RSD, two parameters c and d are introduced for ensuring R output symbols with degree one where

$$R = c \cdot \sqrt{k} \log(\frac{k}{d}).$$

The parameter d is a bound on the fail decoding probability while c is a free parameter. The degree distribution probability is given by

$$\Omega_d^{RSD} = \frac{\Omega_d^{ISD} + \tau(d)}{\beta}$$

where

$$\tau(d) = \begin{cases} \frac{R}{d \cdot k} & 1 \leq d < \frac{k}{R-1} \\ \frac{R}{k} \log\left(\frac{R}{c}\right) & d = \frac{k}{R-1} \\ 0 & d > \frac{k}{R-1} \end{cases} \tag{C.2}$$

and

$$\beta = \sum_{d=1}^{k} \Omega_d^{ISD} + \tau(d).$$

Appendix D

Connectivity Distribution

The connectivity distribution γ_h is the probability that a user is connected to h hubs. Considering the system model presented in Section 6.2 where hubs are arranged to a uniform two dimensional grid with spacing d and each hub has a radius of coverage r.

The connectivity distribution can be derived with the Monte Carlo method or geometrically. Note that the distribution can be derived by only considering the area between four hubs, as shown in Figure D.1. Each color describe a different probability, in particular

- γ_1: yellow
- γ_2: green
- γ_3: lilac
- γ_4: light-blue

D.1 Monte-Carlo Method

We present the algorithm for calculating γ_h with Monte-Carlo the method. Note that deriving γ_h correspond to derive which percentage of the total area is filled by each color. Let us normalize the distance between the hubs and let us assume that the coordinates of h is $c_h = (x_h, y_h)$ such that

$$c_1 = (0,0), \quad c_2 = (1,0), \quad c_3 = (1,1), \quad c_4 = (0,1)$$

.

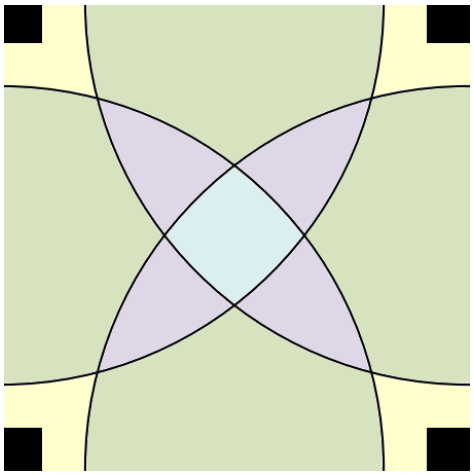


Fig. D.1 Effective area for deriving connectivity distribution

The algorithm for calculating the distribution given the radius coverage and the distance between two hubs is given in Algorithm 4. The algorithms consists on uniformly distributing a large number of points within the effective area and counting how many points are in the radius of coverage of one, two, three and four hubs.

Algorithm 4 $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ =calculate_MonteCarlo_distribution$(r, d, c_1, c_2, c_3, c_4)$

$n = 10^7$

$\mathbf{x\_unif} = rand(n)$ ▷ the x coordinate of n points are distributed uniformly between [0,1]

$\mathbf{y\_unif} = rand(n)$

$\gamma_1 = count\Big((\mathbf{x\_unif} - x_1)^2 + (\mathbf{y\_unif} - y_1)^2 \leq (r/d)^2\Big)/n$

$\gamma_2 = count\Big((\mathbf{x\_unif} - x_2)^2 + (\mathbf{y\_unif} - y_2)^2 \leq (r/d)^2\Big)/n$

$\gamma_3 = count\Big((\mathbf{x\_unif} - x_3)^2 + (\mathbf{y\_unif} - y_3)^2 \leq (r/d)^2\Big)/n$

$\gamma_4 = count\Big((\mathbf{x\_unif} - x_4)^2 + (\mathbf{y\_unif} - y_4)^2 \leq (r/d)^2\Big)/n$