Efficient Control Signaling for Resource Allocation in OFDMA Networks

A Thesis

Submitted to the Faculty

of

Drexel University

by

Gwanmo Ku

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

June 2014

Acknowledgements

This thesis would not be possible without the help of many people.

First and foremost, I would like to thank my advisor Prof. John MacLaren Walsh for his wonderful guidance over the last five years. He has been a profoundly positive influence on my graduate career as a researcher. His endless wise advice and friendly guidance throughout my graduate studies is much appreciated. I would also like to thank my committee members Prof. Steven Weber, Prof. Jaudelice de Oliveira, Prof. Len Cimini, and Prof. Javier Garcia-Frias for their help and patience.

I especially want to thank my family for their constant love and support while I completed this work.

I would like to thank all the members of the ASPITRG group.

I want to thank the National Science Foundation and the Air Force Office of Scientific Research for funding this research in part.

Table of Contents

List of Tables

List of Figures

Abstract

Efficient Control Signaling for Resource Allocation
in OFDMA Networks

Gwanmo Ku
Advisor: John MacLaren Walsh, Ph.D.

This thesis designs efficient control signaling for resource allocation in OFDMA networks, with special attention given to improving the resource controller in the LTE standard. We are interested in two aspects of resource controller design, the amount of control information a resource controller utilizes, and the performance, for instance the data spectral efficiency, it attains. Our overall aim is to understand the fundamental tradeoff between these two quantities, and to learn how to design resource controllers that approach this tradeoff.

To get a sense of the state of the art in resource controller design, we first investigate the resource controller in the LTE standard, evaluating the amount of control information it requires. We thoroughly catalog the physical layer signals related to resource allocation and link adaptation with a focus on the location and formatting of the pertinent reference and control signals, as well as the decisions they enable. This enables us to determine the fraction of the time frequency resource grid spent on control information. This control signaling overhead in LTE occupies a large percentage of the time-frequency footprint of wireless network traffic and is not efficiently encoded.

After this, we set about determining the fundamental tradeoff between the amount of control information and the spectral efficiency by modeling the problem using information theory. In order to design more efficient control signaling schemes, we will model the control signals as messages in a distributed lossy source code, for which the rate distortion function describes an optimum tradeoff between a rate,

reflecting the overhead, and a distortion, reflecting the performance. Although there is no closed form expression for the rate region, we derive a novel adaptation of the Blahut-Arimoto algorithm to the CEO model with independent sources, and use it to numerically calculate the rate distortion function. The developed algorithm is then utilized to calculate the rate distortion function for a series of simple resource allocation models.

Finally, for these physical layer resource allocation models, we design practical distributed quantizers that yield control signaling encodings for a resource controller that approaches the fundamental overhead performance tradeoff limit calculated.

1. Overview : Practical Resource Controller Design

Multiuser OFDMA (Orthogonal Frequency Division Multiple Access) systems have a growing tendency to use resource control technology such as resource allocation, adaptive modulation and coding (AMC), hybrid ARQ (HARQ), MIMO, and CQI reporting in order to increase system throughput and reliability. Because of the technologies used, a large amount of control information is required for data transmission. This control signaling overhead occupies a large percentage of the time-frequency footprint of wireless network traffic and is often not efficiently encoded. In order to design more efficient control signaling schemes, we will model the control signals as messages in a distributed lossy source code, for which the rate distortion function describes is an optimum tradeoff between a rate, reflecting the overhead, and a distortion, reflecting the performance. Therefore, the primary research issues are thus what minimum rate of control information is required to make resource decisions with a target performance and how we can design efficient practical resource control signaling schemes approaching the fundamental limit for this tradeoff.

We utilize distributed source coding theory to find the fundamental tradeoffs between collaboration information overhead and system performance for resource allocation in OFDMA systems, compare these tradeoffs with the amount of overhead and performance achieved by existing resource allocation designs in the 4G cellular standards, and design an improved resource controller which approaches the fundamental tradeoff. In particular, our work will establish fundamental limits explicitly describing the tradeoffs between the performance of a distributed cooperative resource controller and the amount of collaboration overhead information it must exchange. In Chapter 2, we will begin by reviewing the existing resource controller design employed in the LTE standard, focusing on the way in which control signals and salient

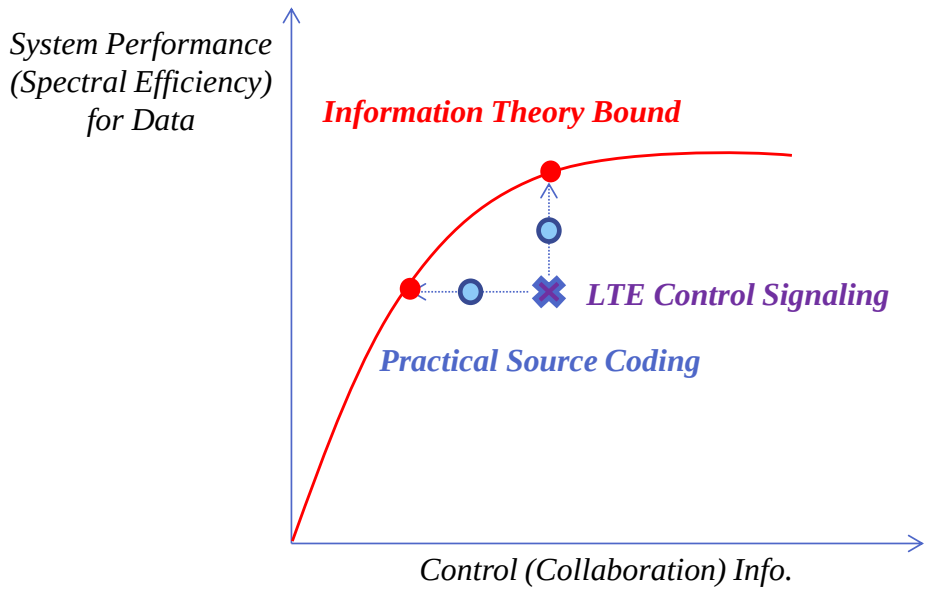


Figure 1.1: Overhead Performance Tradeoffs

feedback signals for resource allocation and link adaptation are encoded. As suggested in Figure 1.1, we will argue that LTE control signaling overhead, at about 32 % of downlink transmission, is still not near to the theoretical performance limit that could be achieved with the same amount of control information.

We then set about, in Chapters 3 and 4, determining the minimum, over all possible resource controller designs, amount of control information that is necessary to obtain a target resource allocation performance. In particular, we develop a simple model for a resource allocation problem by assuming that each user in the system knows their own channel state information, and must compress this information, then send the result to the basestation. The basestation, in turn, upon receipt of this compressed channel state information must make the best resource allocation and link adaptation decisions possible. In order to illustrate our ideas here, we take as our metric the spectral efficiency, but our approach could be modified to include other metrics reflecting fairness and quality of service as well.

A key idea utilized in this thesis is to view this overall resource controller design as an instance of a lossy source code for the CEO problem, in which the encoding nodes are the compressors of the channel state, and whose decoding node, the CEO, must estimate the resource allocation that maximizes the spectral efficiency. The loss measured is the reduction in spectral efficiency relative to an omniscient controller having access to all of the channel state in the network. A key issue in then how to compute the rate distortion function for the CEO problem with independent sources. In particular, we will show that, owing to independence of the channel states, there is an analytical optimization problem whose solution gives the rate distortion function describing the fundamental tradeoff between spectral efficiency and the number of bits utilized for the compressed channel state. This solution of this optimization must be computed numerically, and for some problems, including the one in question, the problem can be non-convex. In this vein, a major innovation we provide in Chapter 3 is a modification of the Blahut-Arimoto algorithm for computing the rate distortion function for the lossy compression of a single source to a hybrid alternating minimization (whose components are Blahut-Arimoto-like algorithms for each user) which can be shown to converge for the class of CEO problems under investigation. Because the original optimization problem can be non-convex, and the developed algorithm, while if initialized correctly can converge to the global optimum, may converge to suboptimal point, a key aspect of the method we develop is an initialization for this algorithm based on a method for calculating the global optimum for a particular, minimum distortion, point on the tradeoff. We then utilize this technique in Chapter 4 to calculate the fundamental limits for the resource allocation model we have introduced, utilizing a channel state distribution that is motivated by the LTE standard.

Having established fundamental performance limits for the simplified model for

resource allocation and link adaptation in Chapter 4, in Chapter 5 we set about designing practical resource control signaling schemes to approach these limits. Via both analysis and numerical calculation, this chapter shows how to design quantizers which approach the rate distortion function, and thus yield resource control signaling schemes close to the fundamental limit set out by the rate distortion function.

2. Evaluation of Resource Allocation in LTE and LTE Advanced

In this chapter,[1] we review the resource control signaling utilized in the LTE standard. Our aim is to both show that physical layer control signaling related to resource allocation and link adaptation occupies a large time frequency footprint in the wireless network, and also that this control signaling is not efficiently encoded relative to the minimum amount of control information necessary to attain a desired resource allocation and link adaptation performance. Resource allocation and link adaptation in LTE and LTE Advanced are discussed with a focus on the location and formatting of the pertinent reference and control signals, as well as the decisions they enable. The signals and encoding standardized for time-frequency resource allocation, rate control, channel feedback, and power control are all reviewed in this chapter. Finally, from the location of control channels in the LTE standard, we approximately estimate the portion of control information in downlink time-frequency resource grid.

2.1 Introduction and Organization

Mobile cellular communication standards have converged to the use of orthogonal frequency division multiplexing (OFDM) in order to achieve high data rates and spectral efficiencies. Modern wireless communication technologies must be capable of transmitting at the higher data rate required for conveying multimedia information such as images and video as well as voice, and this inadvertently necessitates the use of wireless signals occupying a large spectral bandwidth. Characteristically, such wide-band wireless signals suffer from time-dispersive and frequency-selective fading due to multi-path propagation. Markedly, OFDM's ability to translate a frequency-

[1]The contents of this chapter has been submitted to the *IEEE Communications Surveys and Tutorials* journal and has received a first round of positive reviews. The manuscript is under revision for a second round of review.

selective fading channel into multiple independent flat-fading channels makes it a powerful technique for mitigating this multi-path fading. This capability has led to its widespread adoption in current mobile cellular communication standards such as third generation partnership project (3GPP) LTE [1] and WiMAX (Worldwide Interoperability for Microwave Access) [2].

A key issue that such cellular systems must deal with is time-frequency dynamics in the channel status between the assigned mobile users and the base station in a cell. Generally speaking, in order to maximize system throughput, the usable time and frequency resources must be dynamically allocated according to channel dependent methods in both fast and slow fading environments. When channel dynamics are slow enough to enable the allocation of those frequency resources with favorable fading to a mobile station, localized resource allocation, in which these favorable resources are allocated to this mobile station, is preferable. Conversely, when channels vary so quickly that accurate feedback about which frequency resources have favorable fading is unavailable at a base station, it is preferable to utilize distributed resource allocation in which those frequencies allocated to a given mobile station are maximally spread out to enable the greatest diversity. Additionally, to guarantee a minimum transmission performance, as well as to provide varying quality of service, in the presence of these time frequency channel dynamics and interference, link adaptation techniques, including rate adaptation and power controls are necessary.

We review how these dual problems of resource allocation and link adaptation are enabled in the LTE and LTE Advanced standards under the broad heading of dynamic scheduling. Consistent with the type of information that is provided in the standards, we focus on the location, format, and encoding of the possible control decisions for resource allocation and link adaptation that the standards enable. Additionally, we review the location, format, and encoding of the feedback measurement information

that is salient for making these control decisions.

For those completely unfamiliar with LTE and LTE Advanced, we have included a high level introduction to the standard in Section 2.2 reviewing the multiple access and modulation techniques utilized, as well as the support for multiple antenna transmission. After this we pass to the main goal of the chapter, which is to review the parts of the standard involving resource allocation and rate adaptation. We begin by introducing in Section 2.3 the time frequency resource grid from which resource allocations are made to the various users. We then review in Section 2.4 the general problems of resource allocation and link adaptation, identifying the types of control information that are necessary to support these processes, as well as where these various types of control information are placed in the resource grid. Section 2.5 is then dedicated to describing the various resource allocation control decisions that these standards enable and how they are encoded, while Section 2.6 identifies the enabled decisions and salient information for link adaptation and how they are encoded in these standards.

2.2 Overview of LTE Standard

LTE is a 3GPP [1, 3] body of standards that evolves the 3G universal mobile telecommunications system (UMTS) radio-access technology to enable a high-data-rate, low-latency and packet-optimized radio-access technology.

While the previous enhancements to UMTS by 3GPP, including high speed downlink packet access (HSDPA) and enhanced Uplink, provided improvements that would keep the 3GPP radio access architecture competitive for several years, in order to ensure competitiveness a decade or more, 3GPP proposed a LTE of the 3GPP radio access technology. The primary design requirements for LTE included a higher user data rate (100/50 Mbps for downlink (DL) / uplink (UL)), a reduced latency (be-

low 10ms with shortened TTI (transmission timing interval) compared to the 3G services), a cost effective migration from the UMTS radio interface and architecture, an improved system capacity and coverage, and support for scalable bandwidth (up to 20MHz, wider and less than 5 MHz depending to the available system bandwidth) [4–10].

The LTE standard is specified in the technical specification (TS) 36 series (36.XXX) of documents in Release 8 from 3GPP and extended into Release 9 with small enhancements - such as location, emergency and broadcasting services [1,3,11,12]. The LTE Advanced standards are also specified in the TS 36 series in Release 10 and beyond [3, 13]. In order to provide developers with a stable development phase of implementation and to be able to easily update new features required by market, LTE and LTE Advanced use a system of parallel releases from Release 8 [14] onward.

A key feature introduced in LTE to meet the new requirements to increase system throughput and reliability was the support of multi-antenna transmission and reception employing multiple input multiple output (MIMO) communications. The DL transmission with multiple transmit and receive antennas (two or four in LTE [15,16], up to eight in LTE Advanced [17,18]) allows for multi-layer transmissions (up to four layers in LTE [15, 16], up to eight layers in LTE Advanced [17, 18]). The UL multi-antenna transmission is only supported in LTE Advanced with up to four transmit and receive antennas allowing for UL multi-layer transmission with up to four streams [17, 18]. This multi-antenna transmission will be discussed in greater detail in Section 2.2.3.

In addition to those features included in LTE, in Release 10, LTE Advanced supports carrier aggregation (CA) which allows wider transmission bandwidths by aggregating two or more spectrum bands [17, 19–23] as well as coordinated multi-point (CoMP) operation in Release 11 [18, 24], which allows up to five serving cells

to communicate with a single mobile in order to improve the coverage of high data rates, the cell-edge throughput, and the system throughput.

The overall architecture of the LTE network, consisting of the air interface referred to as the evolved universal terrestrial radio access network (E-UTRAN) and the evolved packet core (EPC), is illustrated in Figure 2.1. EPC provides a framework for an evolution or migration of the 3GPP system to a higher-data-rate, lower latency, packet-optimized system that supports multiple other radio access technologies. An important characteristic of EPC is that it no longer utilizes circuit switching, hence, for traditional voice service, EPC defines an interface with a mobile switching center to allow for the conversion between packet switched and circuit switched traffic. In keeping with this transition to a completely packet switched architecture, 3GPP supports all-IP network (AIPN) for LTE services over E-UTRAN and EPC [25]. This transition to an AIPN significantly reduces the network cost and helps cope with the rapid growth of Internet protocol (IP) data traffic. At the boundary between the EPC backbone of LTE and the radio "air interface", we encounter the evolved Node B (eNB), which is a logical node for radio transmission to/from the user equipment (UE), which is what LTE calls mobile phones and networking cards [12, 26–29]. Naturally the duties in administering the network are split between the eNB and EPC. eNB covers the radio resource management duties including the scheduling of the use of the time-frequency resources and the making of the necessary measurements for mobility and scheduling. EPC, on the other hand, handles the mobility control including roaming, authentication, handover, packet routing, per-user based packet filtering, UE IP address allocation, etc. As this chapter is focused on resource scheduling, most of the discussion will be focused on the enhanced universal terrestrial radio access (E-UTRA) interface between the eNB and UE parts of the standard rather than on EPC.

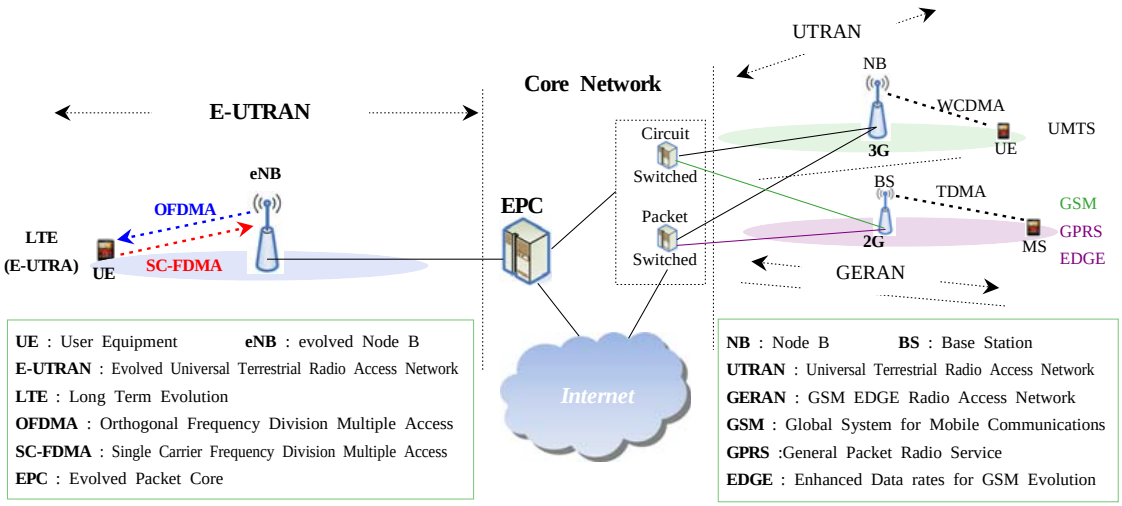


Figure 2.1: Overall Architecture of E-UTRAN and EPC

2.2.1 Radio Interface Protocol Around Physical Layer

In LTE, the E-UTRA radio interface between the UE and the eNB is composed of three layers as depicted in Figure 2.2 [15–18]. Layer 1, the physical layer (PHY), modulates physical symbols over the radio interface [11,30–32], while layer 2 achieves multiplexing, scheduling, priority handling, and error correction [12, 26–28, 33–36]. The medium access control (MAC) sub-layer of layer 2 controls the shared access to the radio interface across different UEs, and maps between logical channels utilized by upper layers and the transport channels utilized at the PHY [12, 26–28]. The transport channels between the PHY and MAC sub-layers are used for specifying how the information is transferred over the radio interface, while the logical channels between MAC and the upper layers are used for specifying what type of information is transferred [15–18]. Layer 3, named the radio resource control (RRC) layer, handles the radio configuration control and the radio resource management with the purpose of broadcasting system information, paging (seeking a UE), and establishment/maintenance/release of a connection between the UE and E-UTRAN.

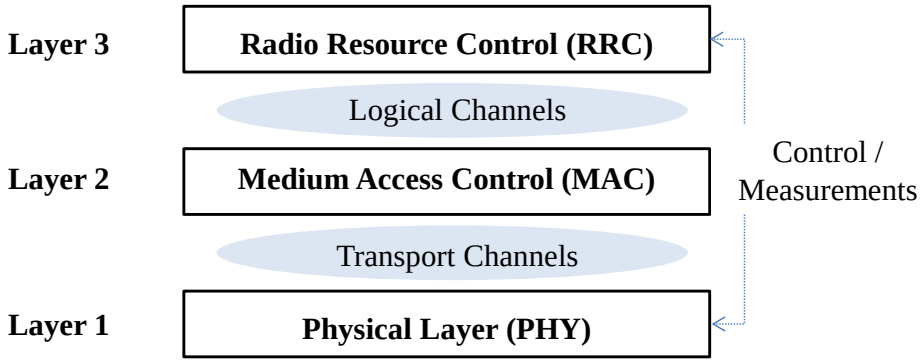


Figure 2.2: LTE Radio Interface Protocol Architecture around PHY

Additionally, its connection with the rest of the network is achieved through the IP interface with EPC.

2.2.2 Multiple Access

To share and schedule access to the time frequency radio resources across users, LTE supports two different multiple access schemes in the UL and DL [15–18]. For DL transmission from an eNB to UEs, conventional orthogonal frequency division multiple access (OFDMA) with a cyclic prefix (CP) is utilized. For UL transmission from UE to an eNB, DFT (Discrete Fourier Transformation) spread -OFDMA (also called as single carrier - frequency division multiple access (SC-FDMA)) with a CP is utilized as illustrated in Figure 2.1.

In order to support DL requirements such as flexible bandwidth and high data rates, LTE utilizes OFDMA which is a combined version of OFDM and FDMA (Frequency Division Multiple Access) [15,30,37–44]. Symbol block-wise transmission with a CP larger than the maximum delay spread in the multi-path fading channel removes inter symbol interference and inter carrier interference in OFDM. By allocating different subcarriers on different OFDM symbols to different users in OFDMA, the signals for different users in a given cell are orthogonal, and hence intra-cell interfer-

ence is avoided. However, inter-cell interference between several cells is still possible as frequency reuse between cells allows for user data between neighboring cells to be non-orthogonal. LTE suppresses this type of inter-cell interference by additionally spreading certain data with orthogonal cell-specific identity codes for different UEs [45].

Despite a number of advantages, OFDM has a drawback of high peak to average power ratio (PAPR), which has motivated the use of a modified version (SC-FDMA) for the UL part of the standard [40, 42–44, 46–50]. In particular, when excited with high PAPR signals, a power amplifier, especially in a battery powered hand held device, must operate inefficiently with a large back-off in order to prevent significant system performance degradation due to the power amplifier nonlinearity [51]. SC-FDMA overcomes the PAPR issue by modulating on each subcarrier not the quadrature amplitude modulation (QAM) data directly, but rather the output of another DFT. This creates a signal with a lower PAPR that resembles an oversampled quadrature amplitude modulation (QAM) signal, and thereby allows UE power amplifiers to operate in a more efficient regime [49, 52]. In addition to providing this low PAPR, SC-FDMA with a CP also achieves UL inter-user orthogonality by selecting different subsets of subcarriers to place the DFT-spread data on, and additionally enables efficient frequency domain equalization at the receiver side.

2.2.3 Multi-antenna Transmission

In order to provide robustness to channel fading and enable higher data rates, LTE allows the use of multi-antenna transmission utilizing a precoding process that supports both transmit diversity (enabling greater reliability) and spatial multiplexing (enabling greater data rates). Figure 2.3 shows the general physical channel processing in LTE standard [11, 30–32].

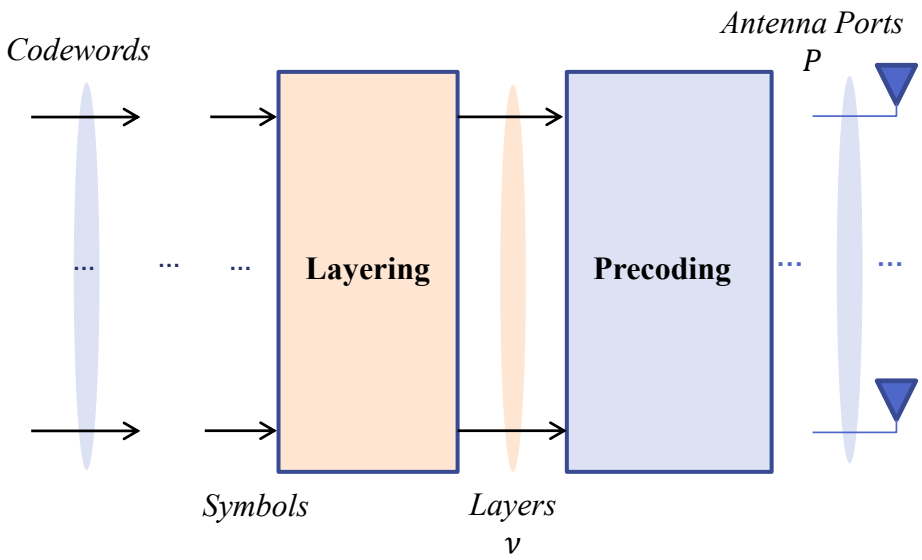


Figure 2.3: Overview of MIMO in LTE

First, blocks of complex symbols from up to two codewords (transport blocks) per UE, are broken up into disjoint parts and collected into layers or streams, which are then multiplied by a precoding matrix to obtain the symbols mapped to the antenna ports. LTE in Rel. 8-9 supports DL MIMO with two or four transmit antennas and two or four receive antennas, which allow for multi-layer transmissions with up to four streams [11, 15, 16, 30]. In Rel. 10-11, LTE Advanced adds support for up to 8 $\times$ 8 DL MIMO allowing 8 streams (layers) and 4 $\times$ 4 UL MIMO allowing 4 streams (layers). Additionally, by allocating different streams to different users, Multi-user MIMO is implemented in LTE Advanced in both UL and DL [17, 18, 31, 32].

The DL MIMO in LTE (Rel. 8-11) supports several different modes via the selection of the precoding matrix. These modes are the open loop transmit diversity mode, the open/closed loop spatial multiplexing modes, as well as a multiuser MIMO mode which is a special version of closed loop spatial multiplexing. In contrast, the UL MIMO supports only closed loop spatial multiplexing, and only in LTE Advanced [53, 54]. The distinction between "open loop" and "closed loop" modes is that closed

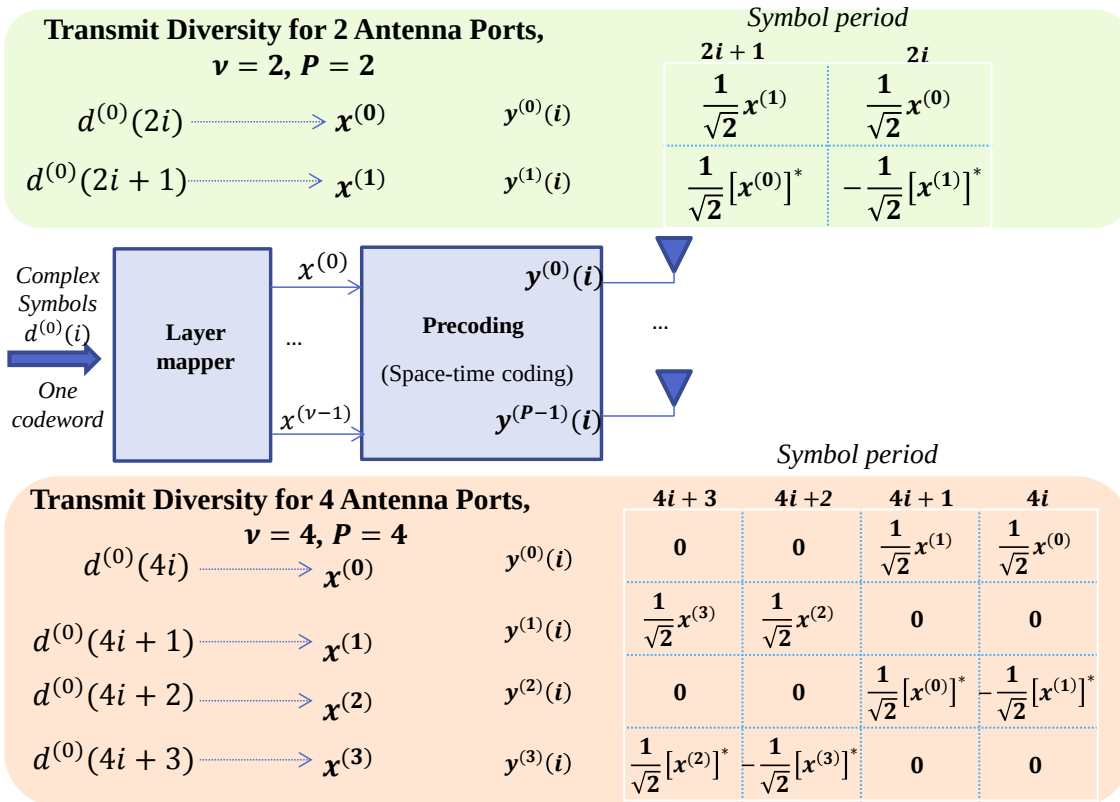


Figure 2.4: DL Transmit Diversity on Two or Four Antenna Ports

loop modes require feedback from the receiver to the transmitter, while the open loop modes do not.

When it is difficult to get the feedback from the UEs due to fast channel variation, an open loop MIMO technique can be used [55]. The layering and precoding operation for the open loop transmit diversity on the DL transmission on two or four antenna ports is illustrated in Figure 2.4 [11, 30–32, 56, 57]. On the other hand, when accurate channel feedback from the UEs is available at the eNB, a closed loop MIMO technique is preferable.

To increase the overall system throughput, LTE enables three types of spatial multiplexing: open loop spatial multiplexing, closed loop spatial multiplexing, and multi-user MIMO. The closed loop spatial multiplexing simply selects a precoding ma-

trix (via feedback from the UE to the eNB) from a predefined list. For four antenna ports, the precoding matrices are generated by a Householder reflection matrix offering competitive performance and potential UE complexity reduction [58]. The open loop spatial multiplexing is also based on additional matrices specifying large-delay cyclic delay diversity [11,30–32,59]. Finally, multi-user MIMO is a type of closed loop spatial multiplexing which allocates different layers to different users [53, 54, 60–62].

2.3 Time-Frequency Resource Grid

In this section, we will review the time frequency resource grid utilized in LTE, indicating the temporal frame structure, the time frequency units for resource allocation, and where in this time-frequency resource grid the control information for resource allocation and link adaptation is located.

2.3.1 Time Units: Temporal Frame Structure

LTE supports two methods of duplexing DL and UL transmission: frequency division duplexing (FDD) which requires a paired spectrum allocation, and time division duplexing (TDD) which is utilized when the devices must operate within an unpaired spectrum allocation [11, 15–18, 30–32, 44, 63]. In both instances, the temporal structure of the DL and UL transmissions is organized in units of a radio frame with a 10 ms frame duration $= 307200 \cdot T_s$ [11, 30–32, 44, 64]. Here, and in other parts of the LTE standard, temporal duration is measured with respect to the smallest time unit $T_s = 32.552ns$, which is the sample period for the base length 2048 fast Fourier transformation (FFT) utilized in the DL OFDM signal. Note that this implies that $T_s = \frac{1}{N_{FFT} \times \triangle f}$, where $N_{FFT} = 2048$ and $\triangle f = 15kHz$ are the FFT / inverse-FFT (IFFT) size and sub-carrier spacing, respectively. This frame duration was selected in order to provide a round trip time of less than 10ms, which in turn

enables the standard to achieve the same latency as the enhanced 3G services HSDPA and HSUPA (high speed uplink packet data access) [65].

While the temporal length of a frame is the same regardless of whether or not FDD or TDD is employed, within this frame LTE utilizes a different temporal structure for these two options, referred to as type 1 for FDD and type 2 for TDD.

Frame Structure Type 1: FDD

Radio frame structure type 1, utilized for FDD, is illustrated in Figures 2.5 and 2.6. Each frame consists of 20 equally sized "slots" with pairs of adjacent slots forming a "sub-frame". A slot (0.5ms = $15360 \cdot T_s$) is the unit in which resource allocations are given, while a sub-frame (1ms), consisting of two slots, is the shortest interval between the transmission of time-frequency scheduling control information. Each slot, in turn consists of either six or seven OFDM symbols, each of duration $2048 \cdot T_s \approx 66.667\mu s$, together with CPs, whose lengths differ depending on whether six or seven OFDM symbols are used. Seven OFDM symbols are usually used, and come with "normal" CP lengths of $160 \cdot T_s \approx 5.2083\mu s$ for the first symbol, and $144 \cdot T_s \approx 4.6875\mu s$ for the remaining six symbols [11, 30–32, 66]. Alternatively, the eNB may opt to use six OFDM symbols per slot, enabling proper operation in the presence of delay spreads exceeding $4.6875\mu s$, by selecting an "extended" CP length of $512 \cdot T_s \approx 16.667\mu s$ [11,30–32,66]. Which CP length is to be utilized is indicated by the eNB in the system information block component of the physical broadcast channel (PBCH) along with other vital system configuration information. We will discuss the location and format of the broadcast channel later in Section 2.4.1. These CP lengths were selected based on multipath fading models that utilized excess tap delay spans around $0.410\mu s$, $2.510\mu s$, and $5\mu s$ [19, 20, 67–72] for low, medium, and high delay spread environments, respectively. Furthermore, a key consideration for the FDD

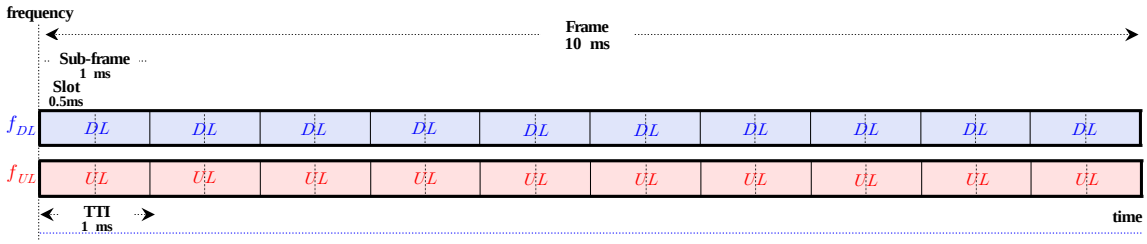


Figure 2.5: Radio Frame Structure Type 1 : Full Duplex FDD

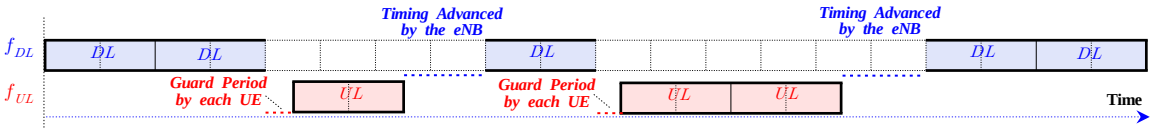


Figure 2.6: Half Duplex FDD Operation at Each UE

mode is whether or not both the UL and the DL may transmit simultaneously (so-called full-duplex operation), or whether one must be silent while the other is utilized (so-called half-duplex operation) only at each UE. The main purpose of a half duplex FDD operation is to reduce the UE complexity by removing the duplexer [73, 74]. A duplex filter is required when there are DL/UL transmissions simultaneously. LTE supports a mix of half duplex operation of each UE and of full duplex operation of an eNB. While an eNB must be able to transmit and receive simultaneously to schedule multiple UEs, under the half-duplex mode, at a given time instant each UE must either listen on the DL or transmit on the UL, i.e. UEs may not simultaneously transmit and receive. In half duplex operation, a UE determines whether or not it should be sending information on the UL or listening on the DL from the scheduling information in a DL control subframe. While both full-duplex FDD and half-duplex FDD operations use the frame structure type 1 [11, 30–32] described above, naturally the half duplex constraint introduces some temporal peculiarities when this mode is utilized. In particular, as illustrated in Figure 2.6, a guard period (GP) is inserted

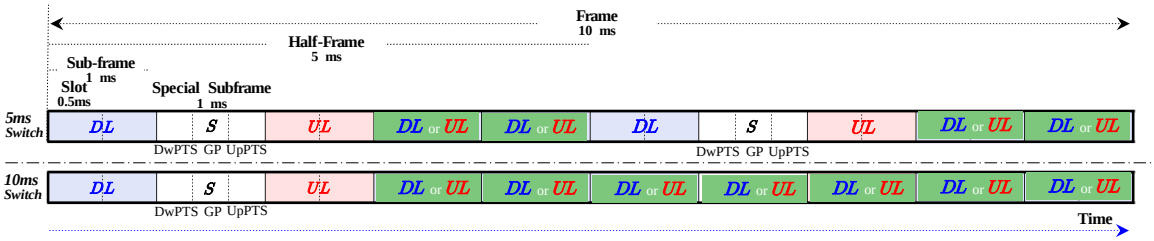


Figure 2.7: Radio Frame Structure Type 2 : TDD

between a DL and an UL transmission in order to allow the UE electronics to do the necessary switching, and a timing advance at the eNB is necessary when switching from UL to DL [75].

Frame Structure Type 2: TDD

The type 2 radio frame structure, which is utilized for TDD, is illustrated in Figure 2.7. Much of the temporal structure of type 1 is shared with type 2. Namely, every frame consists of 10 subframes each of 1ms duration, and each subframe allocated to the UL or DL consists of two slots, each of 0.5ms duration. Each slot, in turn, may utilize the same number of OFDM symbols (six or seven) and CP sizes as under type 1.

However, the differences between type 1 and type 2 frame structure stem from the nature of TDD. To enable TDD, frame structure type 2 must allow for the resource to be sometimes allocated to the DL and sometimes allocated to the UL. Additionally, as the participants are not perfectly synchronized, a GP is utilized when transitioning between DL and UL use. Hence, each subframe is dedicated to exactly one of DL, UL, or a "special subframe" transition between DL and UL use. The assignment of subframes to these three classes is indicated by eNB with control information, sent on the broadcast channel through the PBCH [76], selecting one of seven DL/UL "configurations" (i.e. transition patterns) with different DL/UL ratios as depicted in

UL-DL Configuration	DL : UL Allocation	DL to UL Switch Point Periodicity	Sub-frame Number									
			0	1	2	3	4	5	6	7	8	9
0	2 : 6	5 ms	D	S	U	U	U	D	S	U	U	U
1	4 : 4	5 ms	D	S	U	U	D	D	S	U	U	D
2	6 : 2	5 ms	D	S	U	D	D	D	S	U	D	D
3	6 : 3	10 ms	D	S	U	U	U	D	D	D	D	D
4	7 : 2	10 ms	D	S	U	U	D	D	D	D	D	D
5	8 : 1	10 ms	D	S	U	D	D	D	D	D	D	D
6	3 : 5	5 ms	D	S	U	U	U	D	S	U	U	D

Table 2.1: TDD UL-DL Configurations

Table 2.1.

The special subframe utilized to transition from DL to UL transmission, which is designed to ease synchronization between the UEs and eNB, always consists of three parts: a downlink pilot time slot (DwPTS), a guard period (GP), and an uplink pilot time slot (UpPTS) [77–87]. While the special subframe is always of 1ms duration, these three parts can vary in duration according to ten possibilities depicted in Table 2.2. In release 11, these were modified to only support special subframe configuration 9 for normal CP in DL and 7 for extended CP in DL for the co-existence to legacy TDD system [32,88,89]. The eNB selects one of these formats for the special subframe, and indicates which one is selected to all of the UEs in the cell as part of the PBCH, which will be discussed in section 2.4.1.

2.3.2 Time Frequency Units: Physical Resource Blocks

One of the key characteristics necessary for LTE to become a global standard was adaptability to a wide variety of government specified spectral bands. Government frequency allocation bandwidth sizes ranging over 1.4, 3, 5, 10, 15, and 20 MHz are

Special Subframe Configuration	Normal CP in DL					Extended CP in DL				
	DwPTS	Normal CP in UL		Extended CP in UL		DwPTS	Normal CP in UL		Extended CP in UL	
		GP	UpPTS	GP [ms]	UpPTS		GP [ms]	UpPTS	GP	UpPTS
0	3	10	1	0.702	1	3	0.679	1	8	1
1	9	4		0.273		8	0.262		3	
2	10	3		0.202		9	0.179		2	
3	11	2		0.131		10	0.095		1	
4	12	1		0.059		3	0.607	2	7	2
5	3	9	2	0.609	2	8	0.191		2	
6	9	3		0.190		9	0.107		1	
7	10	2		0.119		5*	0.441*		5*	
8	11	1		0.047						
9	6*	6*		0.405*						

Table 2.2: TDD Special Subframe Configurations (Number of OFDMA/SC-FDMA Symbols, *only supported in Rel. 11)

supported in Releases 8-11 [19, 20, 67–72]. Additionally, from Release 10 [17–20, 69, 72] and onward, in order to allow for larger combined bandwidths and even greater versatility for various government frequency allocations, CA is supported. Broadly speaking, CA, which was a major component of the improvement brought about by Release 10, refers to the combination of multiple disjoint frequency allocations to allow for a larger total bandwidth. At present, at most five such government frequency allocation bands can be aggregated together. Each of the bands that are being combined under CA themselves still need to be either 1.4, 3, 5, 10, 15, or 20 MHz, but their use is coordinated together by the standard to allow for larger effective total bandwidth, support for a larger number of users, and a higher peak data rate. CA is created by coordinating transmissions between multiple LTE eNBs operating in the different bands.

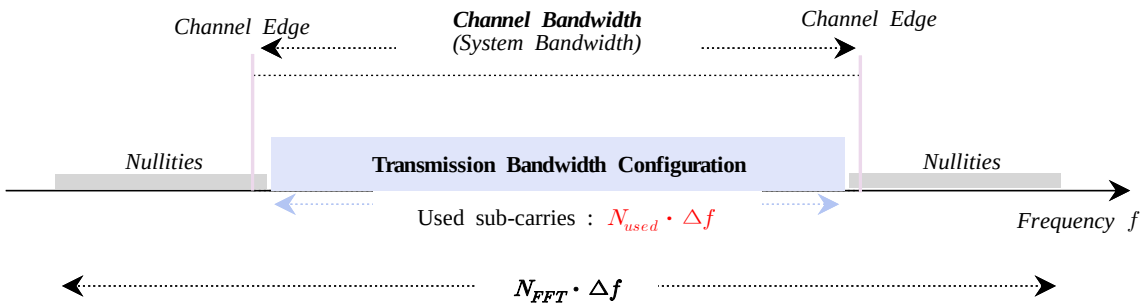


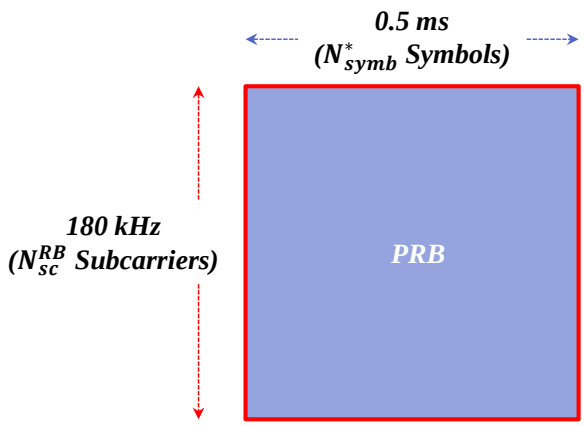
Figure 2.8: Supported Resources in LTE

System BW [MHz]		**1.4**	**3**	**5**	**10**	**15**	**20**
Bandwidth Configurations [Resource Blocks]		6	15	25	50	75	100
FFT/IFFT Size N		2048 for $\triangle f$=15kHz 4096 for $\triangle f$=7.5kHz					
Sub-carriers per RB		12 for $\triangle f$=15kHz 24 for $\triangle f$=7.5kHz					
Used Subcarriers	$\triangle f$=15kHz	72	180	300	600	900	1200
	$\triangle f$=7.5kHz	144	360	600	1200	1800	2400
Sampling Rate		30.72 MHz					

Table 2.3: Supported Resource Parameters in LTE

This bandwidth variety of each government frequency allocation is enabled, while maintaining the same FFT size (default 2048, but 4096 for multimedia broadcast and multicast service (MBMS)) [90, 91]) and sampling frequency $T_s^{-1} = 30.72MHz$, by scaling the number of "null" subcarriers on which zero is modulated, as pictorially represented in Figure 2.8 and Table 2.3 [19, 67–69, 92].

Regardless of the total amount of bandwidth used, LTE allocates time frequency resources in terms of the same resource assignment unit, called a physical resource block (PRB), which consists of 180 kHz below Figure 2.9 in the frequency domain and one slot (0.5 ms, CP associated 6 or 7 OFDMA/SC-FDMA symbols) in the time



CP Type \ Parameter		N^*_{symb}		N^{RB}_{sc}
		N^{DL}_{symb}	N^{UL}_{symb}	
Normal CP	$\Delta f = 15kHz$	**7**	**7**	**12**
Extended CP	$\Delta f = 15kHz$	**6**	**6**	
	$\Delta f = 7.5kHz$	**3**	**-**	**24**

Figure 2.9: Physical Resource Block (PRB) Parameters

domain, as shown in Figures 2.10 and 2.9 [11, 30–32, 93–95]. A PRB always consists of 180kHz in frequency, and by default this means 12 consecutive sub-carriers [96–98], but can alternatively be 24 consecutive sub-carriers when MBMS [15, 90, 91, 99] is in use owing to the higher FFT size used under MBMS (4096 instead of 2048 [90, 91]). As such, the number of available PRBs varies depending on the government frequency allocation bandwidth in use, as indicated in Table 2.3.

Several key factors influenced the selection of a PRB of this size. First of all, it is desirable that the time-frequency dynamics of the channel remain fixed over the duration of a PRB. One back-of-the-envelope handle on how quickly the channel may vary reasons that the Doppler spread encountered at peak mobility speeds associated with traveling on high speed trains (roughly 300-350 km/h), together with a carrier frequency of $\approx$ 2.7GHz, amounts to a Doppler frequency of around 750 Hz, which indicates that the fastest a channel changes is on a time scale of roughly 1.33ms,

which is conveniently larger than twice the 0.5ms duration of a PRB. Second of all, the worst case delay spread the standard is designed to handle is $512 \cdot T_s$ while the typical delay spread it is designed to handle is $144 \cdot T_s$, and as windows of this temporal size corresponds to a sinc with a main lobe width of roughly $120kHz$ and $384kHz$ in frequency, respectively, the channel should vary only a little or not at all, even in the longest delay spread conditions, over the different frequencies in a PRB.

From the alternative statistical channel model standpoint, the size of the PRB was selected via the coherence bandwidth and the coherence time associated with the TU (Typical Urban) channel model with a $1.07\mu s$ delay spread [100–106] and (again) the Doppler spread associated with high speed train travel. Following Rappaport [107], the coherence bandwidth, formally defined as the point where the Fourier transform of the supposed exponential power delay profile ($\frac{1}{2\pi\tau_{rms}} e^{-t/\tau_{rms}} U(t)$) reaches half its maximum, can be approximated as $\frac{1}{5 \cdot \tau_{rms}} \approx 187kHz$. Hence, the very close PRB bandwidth should allow for flat fading across a single PRB in frequency. Selecting a coherence time (again following Rappaport [107]) associated with the Doppler frequency ($843kHz$) for a 2.6GHz carrier when moving at 350km/h we arrive at $\frac{0.423}{f_{Doppler}} \approx 0.5ms$, which is the slot time duration of the PRB. Selecting these values makes it likely that the channel will remain relatively fixed over a single PRB.

Because the smallest unit in a time frequency allocation is a PRB consisting of $N_{sc}^{RB} = 12$ or 24 subcarriers over N_{symb}^{DL} or N_{symb}^{UL} (6 or 7) OFDM symbols, the transmitted signal in each slot can be broken up into a resource grid as shown in Figure 2.11. Each resource element (subcarrier) is uniquely defined by a subcarrier(frequency k)-symbol(time l) index pair (k,l) in a grid, and this grid is allocated in the unit of PRBs to various UEs.

The fixed size of a PRB then dictates that, depending on the bandwidth of the government frequency allocation, the total number of PRBs along the frequency axis

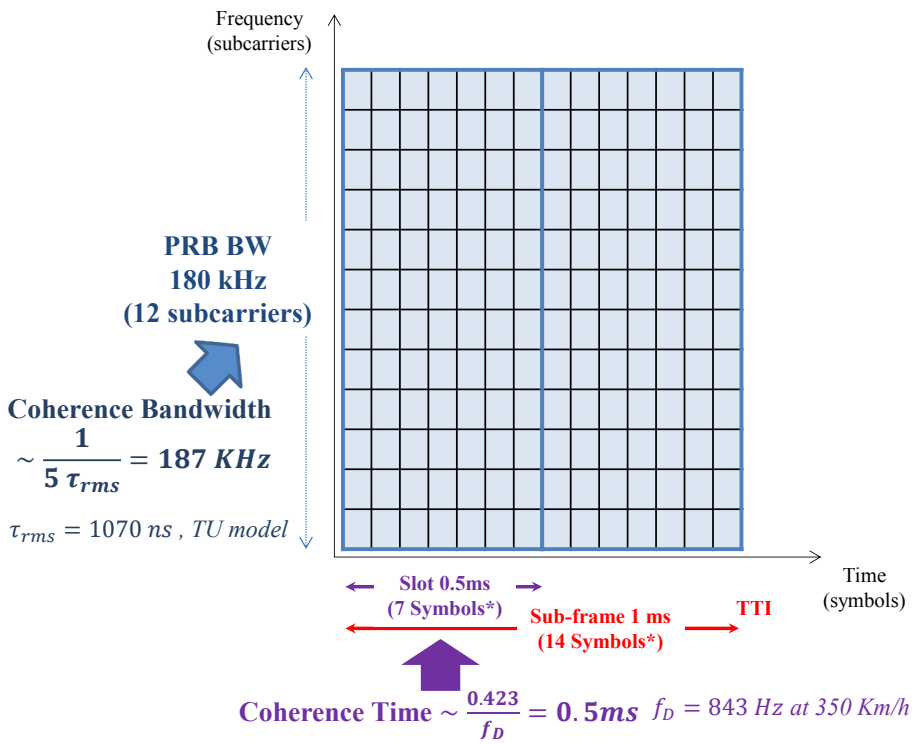


Figure 2.10: Physical Resource Block (PRB), Slot, and Sub-frame

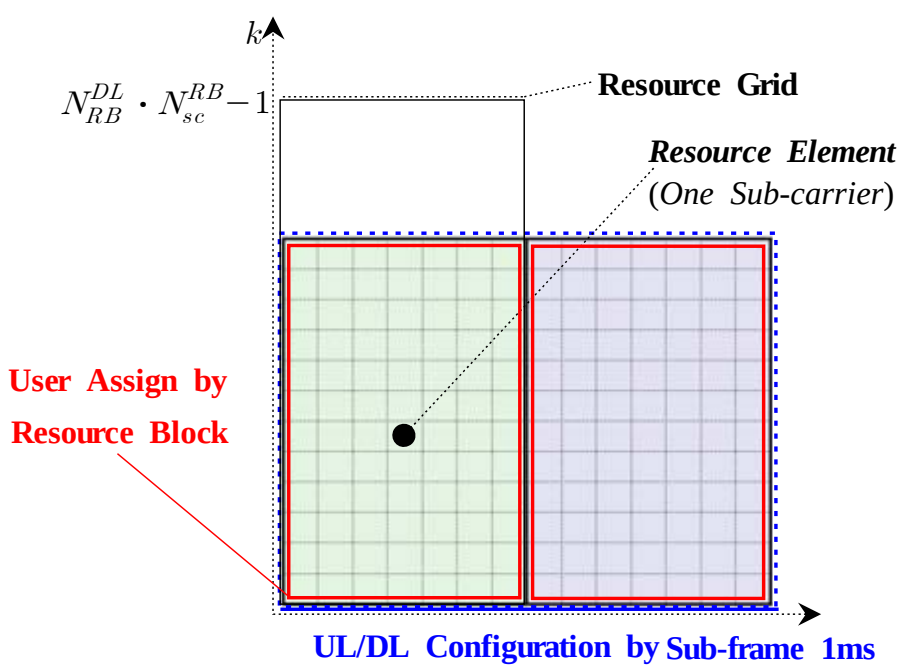


Figure 2.11: Resource Grid

in this resource grid, which is referred to as the "bandwidth configuration" in the standard, ranges from 6 PRBs for 1.4 MHz to 100 PRBs for 20MHz, as discussed in Sections 5.2.1 and 6.2.1 of [11, 30–32].

2.4 Types of Resource Control Information and their Location

Having reviewed the resource grid in Figure 2.11 within which scheduling time-frequency allocations are made to users, we presently pass to explaining where within the time-frequency grid the various types of control information regarding link adaptation and resource allocation are placed. In order to do this, we also review the various types of resource allocation and link adaptation control information that are exchanged in the standard. A more detailed discussion in Sections 2.5 and 2.6 will discuss that the various resource allocation and link adaptation decisions are available and how they are encoded within these defined regions.

2.4.1 Downlink Control Information (DCI) Format and Location (PCFICH, PHICH, PBCH, and (E)PDCCH)

The control information in the LTE standard that is pertinent for resource allocation and link adaptation in the DL can be grouped into four types of channels: physical broadcast channel (PBCH), (enhanced) physical downlink control channels ((E)PDCCHs), physical hybrid-ARQ indicator channels (PHICHs), and the physical control format indicator channel (PCFICH). The most essential system information of the cell is transmitted on the PBCH every 4 frames, i.e. every 40ms [12, 26–28, 108–112]. PDCCHs are the entity responsible for carrying UL and DL scheduling assignments and other link adaptation control information for a given eNB UE connection [11, 30–32, 113–115]. In addition to the PDCCH, LTE Advanced in Release 11 introduces an alternative to the PDCCH, the enhanced control chan-

nel region. Release 11 introduced the EPDCCH to increase control or data capacity for specific UEs in order to enable adaptive control algorithms like DL MIMO with beamforming, CoMP, MU-MIMO with large number of users, and CA [116–120]. In addition to the peak rate gains they enable, many of these enhancements were introduced in order to deal with interference control at the cell edges. Finally, each PHICH carries acknowledgments (ACKs)/negative acknowledgments (NACKs) for a collection of UEs in a hybrid automatic repeat request (HARQ) group, and in this way contributes to rate control.

Here we review the location of these control channels, while we go into greater detail regarding the control decisions they enable for resource allocation and link adaptation in Sections 2.5 and 2.6, respectively. The standard specifies that the locations of the control channels mentioned above will be mapped in the following order in a non-overlapping manner: 1) PCFICH 2) PHICH 3) PDCCH 4) PBCH 5) PDSCH. Thus, we will follow this order in our discussion of the location of these channels, which is depicted in Figure 2.12, and elaborated below.

PCFICH and Control Region Layout

Regardless of the frame structure, *every* DL subframe contains a portion of transmission dedicated to control and to user (data) information. In LTE, the ratio between control information (PDCCH) and UE data (PDSCH) within each of these subframes is adjustable, with the control information always preceding UE data in a subframe. The proportions between these two are set by the control format indicator (CFI) in the PCFICH, which is within the first OFDM symbol of *every DL subframe*. With quadruplets of 16 QPSK (Quadrature Phase Shift Keying) modulated symbols (32 bits) scrambled with a cell specific sequence, and located in four given frequency locations specified in [11, 30–32], the CFI in the PCFICH specifies the length of the

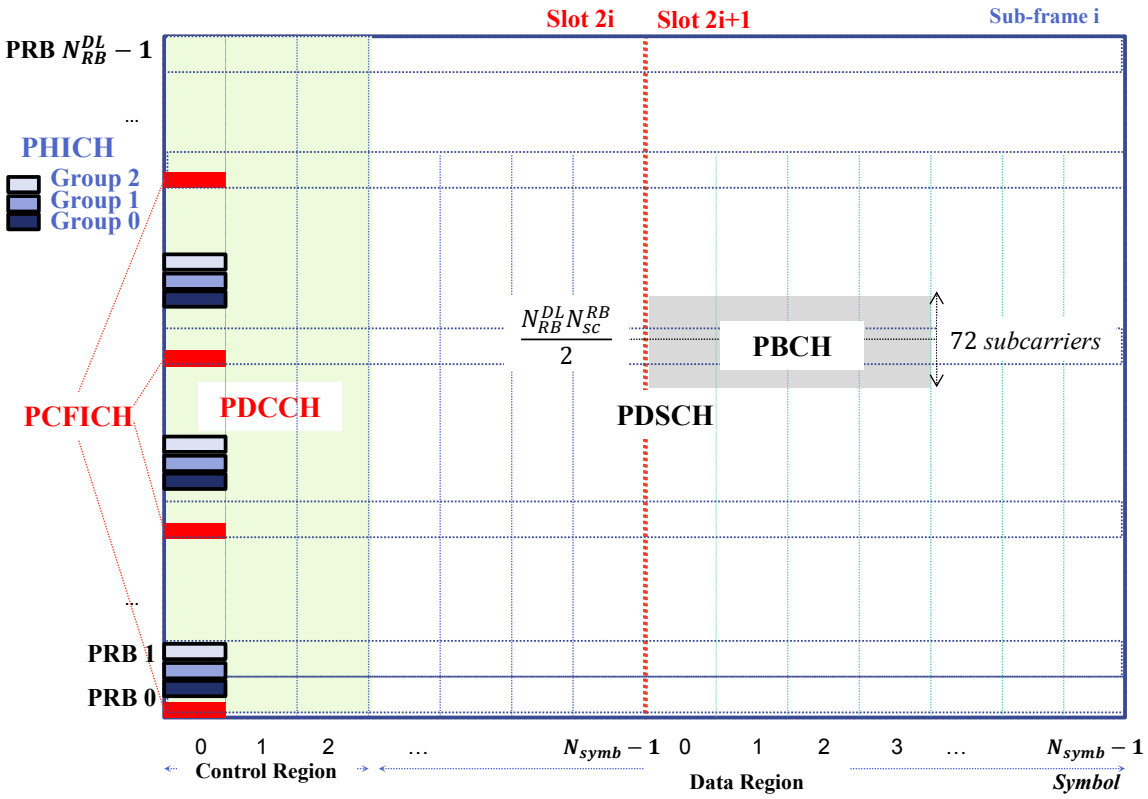


Figure 2.12: DL Control Information : PCFICH, PHICH, PBCH, PDCCH, PDSCH

control information in a sub-frame, as illustrated in Figure 2.12. The control information shared over the PBCH when a UE associates with an eNB tells the UE which subcarriers in the first OFDM symbol have modulated the data for PCFICH.

The CFI in the PCFICH indicates that the first X OFDM symbols in the subframe are to be dedicated to control information, where the allowable values for X depend on some other characteristics [11,30–32]. In particular, the allowable lengths (in number of OFDM symbols) of the DL control region (PDCCH) that an eNB can select in subframes 1 and 6 in TDD mode are 1 or 2 when the bandwidth configuration (N_{RB}^{DL}) is larger than 10, and 2 when the bandwidth configuration (N_{RB}^{DL}) is less than or equal to 10. Under the FDD mode, these lengths may be chosen among 1, 2, or 3 OFDM symbols when the number of DL PRBs per OFDM symbol (N_{RB}^{DL}) is larger than among 10, and 2 or 3 OFDM symbols when N_{RB}^{DL} is less than or equal to 10, as specified in 6.7 [11,30–32].

Within this time frequency region associated with the first $1, 2$ or 3 OFDM symbols of each subframe that the PCFICH indicates as dedicated to DL control information, the control information is mapped to subcarriers and symbols in the units of resource element groups (REGs) which are smaller that the PRB units for time frequency resource allocation for data that were introduced in Section 2.3. A REG consists of 6 consecutive subcarriers in the first OFDM symbol of a subframe, while in the second OFDM symbol they consist of 4 or 6 consecutive subcarriers, and in the third OFDM symbol they consist of 4 consecutive subcarriers. Whether the REGs are made up of 4 or 6 subcarriers in the second OFDM symbol of a subframe is determined by whether a normal CP or an extended length CP is utilized, as well as the number of cell specific reference signals in use based on the characteristics of the multi-antenna transmission [11,30–32]. This yields a REG grid within the control region as depicted in Figures 2.13 and 2.14, where we have assumed the PCFICH indicated a control

region length of three OFDM symbols and the REGs are indicated by the blue lined grid, within which the remaining control information will be mapped.

PHICH

The next control channel which is mapped is the physical hybrid ARQ indicator channel (PHICH). The PHICH carries DL HARQ information in response to UL transmissions, indicating that these transmissions were successful using a ACK or unsuccessful (retransmission required) using a NACK. LTE allows a group-wise HARQ process: each HARQ group consists of 8 or 4 different UEs (8 for Normal CP, 4 for Extended CP) separated through different orthogonal sequences in order to reduce overhead, forming spread signals which consist of quadruplets of 12 binary phase shift keying (BPSK) modulated symbols. A PHICH resource is identified by the PHICH group number and the orthogonal sequence index within the group. Each PHICH group is then mapped to three frequency locations specified in Section 6.9 of [30] and illustrated in Figure 2.12.

(E)PDCCH

In LTE, the physical downlink control channels (PDCCH) carry all of the DL and UL scheduling information, informing individual UEs as to where in time frequency in the DL / UL to look for/place their information, how it will be encoded as modulated, as well as of the UL transmission power control commands. In LTE Advanced, as of Release 11, the enhanced physical downlink control channel (EPDCCH) serves the same purposes and also supports some extra control information necessary for CA and CoMP, but it is mapped to a different portion of the time frequency plane.

Each PDCCH or EPDCCH is assigned to a UE (to whom the scheduling grant or power control command is given to) by including a 16 bit cyclic redundancy check

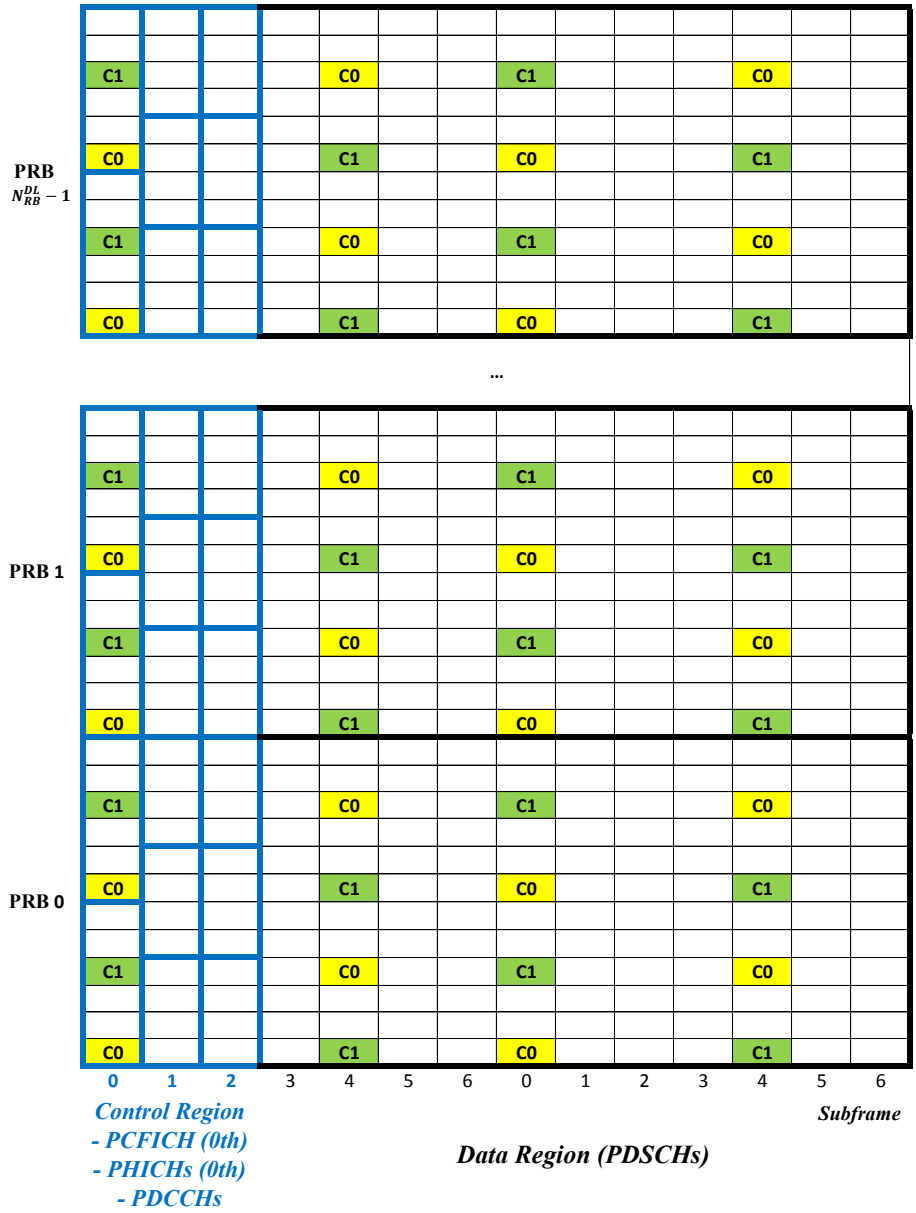


Figure 2.13: Resource Element Group (Two Cell-Specific RS and CFI=3)

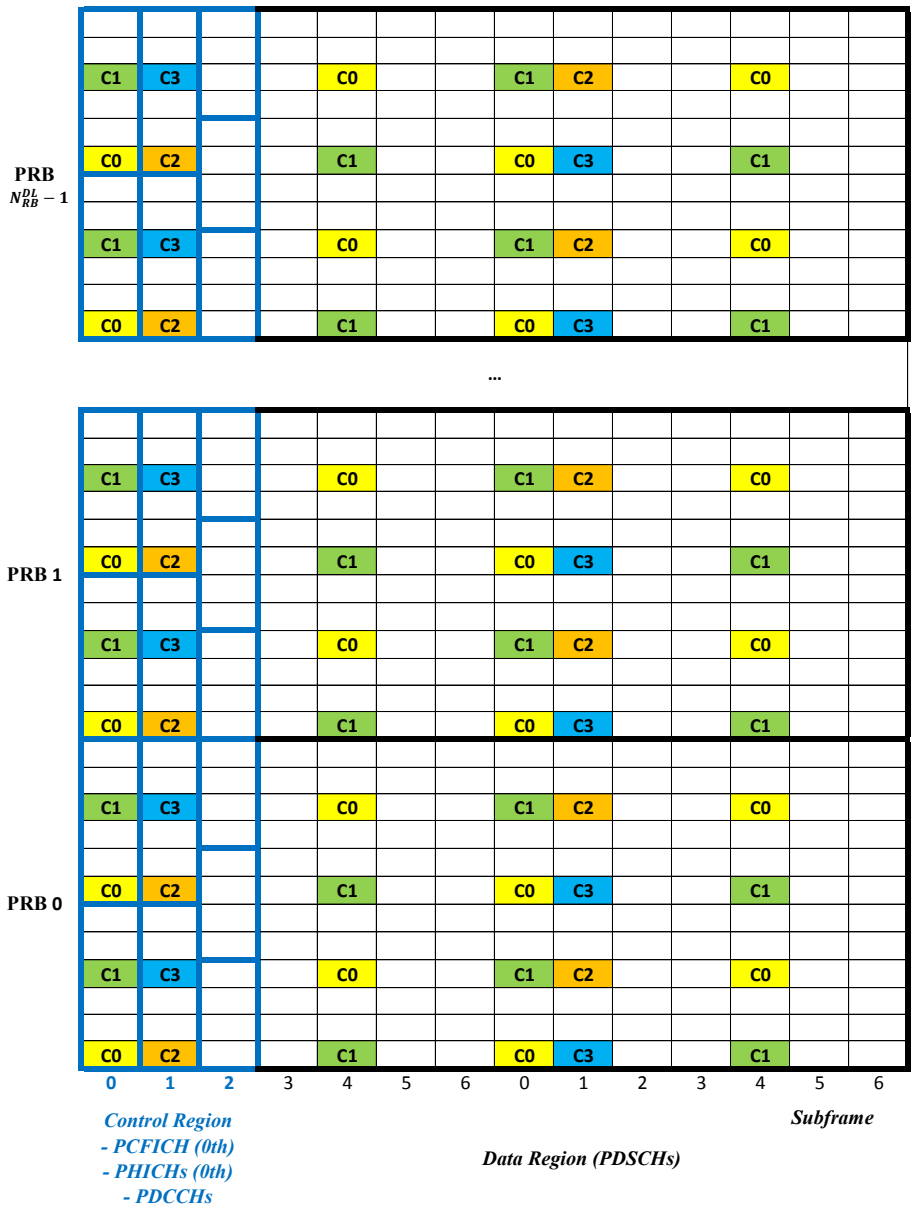


Figure 2.14: Resource Element Group (Four Cell-Specific RS and CFI=3)

PDCCH Format	Number of CCEs	Number of Resource Element Groups	# PDCCH bits
0	**1**	**9**	**72**
1	**2**	**18**	**144**
2	**4**	**36**	**288**
3	**8**	**72**	**576**

Table 2.4: PDCCH Format

(CRC) attachment which is subsequently scrambled, along with the other control bits, with the cell radio network temporary identifier (C-RNTI) of the UE [33–36, 121]. An UE determines which PDCCH or EPDCCH it is assigned to by descrambling with its C-RNTI, and then checking to see if the CRC checks correctly.

A single PDCCH is an aggregation of 1, 2, 4, or 8 Control Channel Elements (CCEs) with each CCE consisting of 72 bits as shown in Table 2.4 as specified in 6.8.1 of [30]. These control channel elements are mapped to a collection of REGs (6.2.4 [30]) that are adjacent in the control portion of the time frequency plane, with those REGs that are already associated with PCFICH or PHICH – both a part the first OFDM symbol of a subframe – being omitted, as depicted in Figure 2.15.

A key difference between PDCCH and EPDCCH is that EPDCCH is no longer confined to the first few OFDM symbols of a subframe indicated by PCFICH, but it can continue during the whole subframe. PRBs are set aside for EPDCCH use in temporally contiguous pairs, so that a PRB pair is a collection of contiguous subcarriers during a series of OFDM symbols associated with a full subframe. In particular, the time-frequency regions devoted to PDSCH and to EPDCCH are indicated by setting the starting symbol location of EPDCCH ($l_{EPDCCHStart}$) which can be selected from 1, 2, 3, or 4 in Release 11. After that, Release 11 allows the size of EPDCCH to be

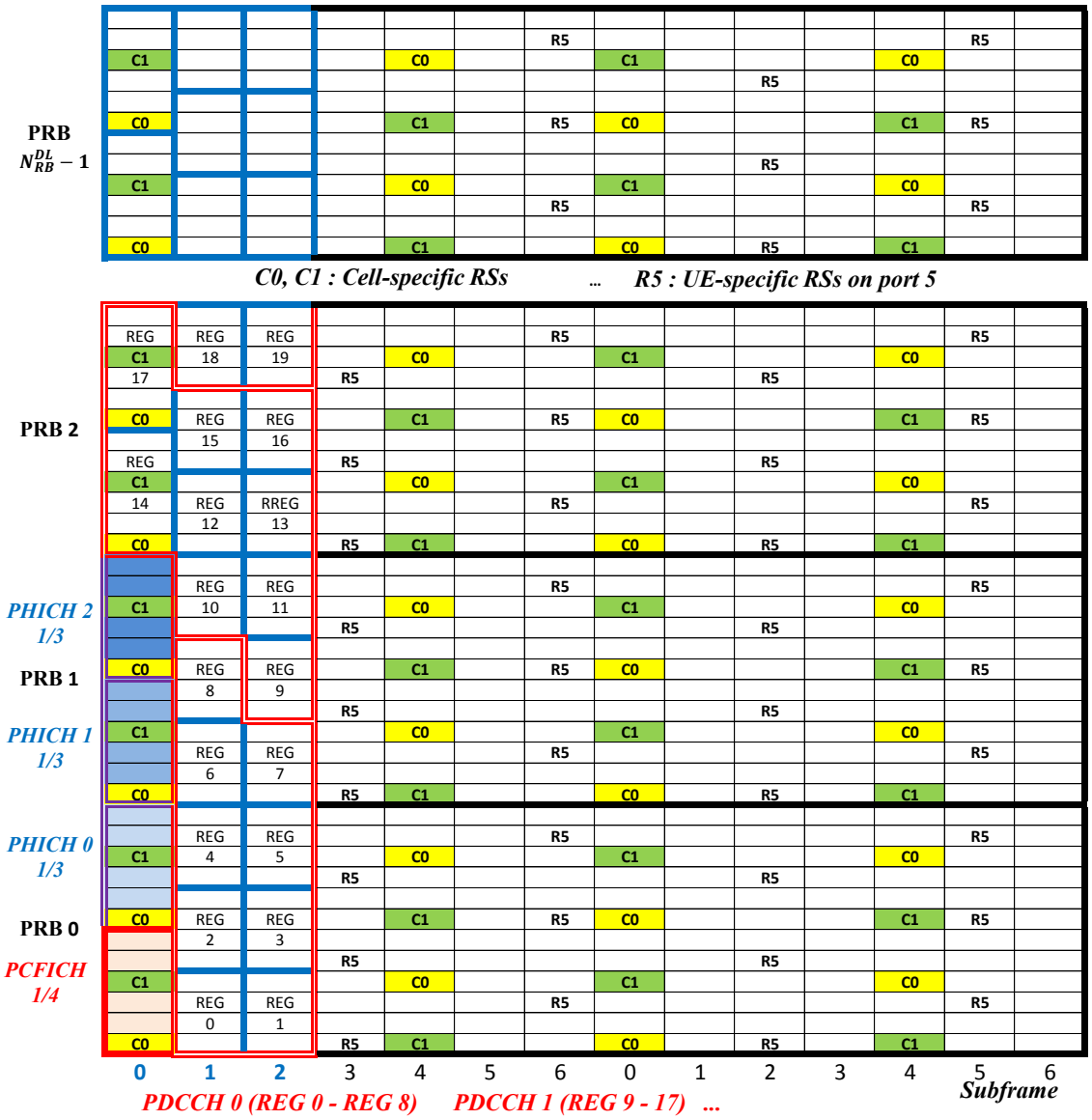


Figure 2.15: CCE Mapping

	Assume No Cell-specific RS					**EREG index 0-15**					*One Resource Block Pair*			
11	11	7	3	15	11	DM RS	DM RS	3	15	11	7	3	DM RS	DM RS
10	10	6	2	14	10	DM RS	DM RS	2	14	10	6	2	DM RS	DM RS
9	9	5	1	13	9	1	7	1	13	9	5	1	9	15
8	8	4	0	12	8	0	6	0	12	8	4	0	8	14
7	7	3	15	11	7	15	5	15	11	7	3	15	7	13
6	6	2	14	10	6	DM RS	DM RS	14	10	6	2	14	DM RS	DM RS
5	5	1	13	9	5	DM RS	DM RS	13	9	5	1	13	DM RS	DM RS
4	4	0	12	8	4	14	4	12	8	4	0	12	6	12
3	3	15	11	7	3	13	3	11	7	3	15	11	5	11
2	2	14	10	6	2	12	2	10	6	2	14	10	**4**	10
1	1	13	9	5	1	DM RS	DM RS	9	5	1	13	9	DM RS	DM RS
0	0	12	8	4	0	DM RS	DM RS	8	4	0	12	8	DM RS	DM RS
	0	1	2	3	4	5	6	0	1	2	3	4	5	6

Starting Symbol Location of EPDCCHs *Subframe*

Figure 2.16: Enhanced Resource Element Groups

selected among 2, 4, or 8 PRB pairs [32].

Each PRB pair assigned to EPDCCH is then divided up into 16 enhanced resource element groups [32] as illustrated in Figure 2.16. The enhanced resource element groups (EREGs) are then aggregated into enhanced CCEs (ECCE) as shown in the Table 2.5.

The supported EPDCCH formats depending on the required amount of the enhanced control information, measured in enhanced resource element groups, are shown in the Table 2.6. The eNB can select a EPDCCH format that utilizes frequency selective transmission or, alternatively, a EPDCCH format that is distributed in frequency to maximize frequency diversity. More details regarding EPDCCH are available in Sections 6.2.4A and 6.8A of [32].

Normal cyclic prefix			**Extended cyclic prefix**	
Normal subframe	Special subframe, configuration 3, 4, 8	Special subframe, configuration 1, 2, 6, 7, 9	Normal subframe	Special subframe, configuration 1, 2, 3, 5, 6
4		**8**		

Table 2.5: Number of EREGs per ECCE

EPDCCH format	**Number of ECCEs** for one EPDCCH, N_{EPDCCH}^{ECCE} **Case 1 :** DCI formats 2/2A/2B/2C/2D and $N_{RB}^{DL} \geq 25$ or any DCI format and $n_{EPDCCH} < 104$, normal CP in normal subframes or special subframes Conf. 3,4,8 **Case 2** : otherwise			
	Case 1		Case 2	
	Localized transmission	Distributed transmission	Localized transmission	Distributed transmission
0	2	2	1	1
1	4	4	2	2
2	8	8	4	4
3	16	16	8	8
4	-	32	-	16

Table 2.6: Supported EPDCCH formats

DCI Formats		Purpose
UL Scheduling (PUSCH)	**0**	UL Scheduling + TPC (PUSCH)
	4	UL Scheduling with **CLSM** + TPC (PUSCH) (Rel. 10-11)
DL Scheduling (PDSCH)	**1**	Scheduling, TPC (PUCCH)
	1A	**Compact** Scheduling with **TxD**, TPC (PUCCH)
	1B	**Compact** Scheduling with **CLSM**, TPC (PUCCH)
	1C	**Very Compact** Scheduling
	1D	**Compact** Scheduling with **MU-MIMO**, TPC (PUCCH)
	2	Scheduling with **CLSM or TxD**, TPC (PUCCH)
	2A	Scheduling with **Large CDD or TxD**, TPC (PUCCH)
	2B	Scheduling with **Dual Layer Transmission**, TPC (PUCCH) (Rel. 9-11)
	2C	**Up to 8 Layered Compact** Scheduling, TPC (PUCCH) (Rel. 10-11)
	2D	**Up to 8 Layered Compact** Scheduling for **CoMP**, TPC (PUCCH) (Rel. 11)
UL Power Control	**3**	TPC for PUCCH, PUSCH 2bit Power Adjustment
	3A	TPC for PUCCH, PUSCH 1bit Power Adjustment

TPC (Transmit Power Control) **CLSM** (Closed Loop Spatial Multiplexing) **TxD** (Transmit Diversity)
MU-MIMO (Multi-User MIMO) **CDD** (Cyclic Delay Diversity) **CoMP** (Coordinated Multi Point)

Table 2.7: Supported DCI Formats in Rel. 8-11

In order to implement resource allocation and link adaptation for each user, LTE and LTE Advanced specify a resource assignment field giving DCI within the PDCCH/EPDCCH assigned to that user. Altogether, LTE and LTE Advanced specify multiple types of DCI formats, each one indicating the organization of the bits within a PDCCH/EPDCCH, whose various purposes are specified in Table 2.7. Each DCI format that can be utilized for PDCCH can also be utilized with EPDCCH, with the exception that two additional fields are added in each case. These fields specify which of the carriers (under CA) to use and provide extra multiple HARQ-ACKs/NACKs to support CA and CoMP.

PBCH

The next control channel, which is allocated resources in the grid after PCFICH, PHICH, and PDCCH/EPDDCH, is PBCH, which carries essential system informa-

tion before normal data transmission. In particular, PBCH contains the master information block, which carries important parameters about the cell, including the number of PRBs, the DL/UL configuration (in TDD mode), and key parameters for UL power control. PBCH is transmitted every 40 ms, which is every four frames, and this timing interval can be blindly detected without explicit signaling [12, 26–28]. After scrambling with a cell-specific sequence, the BPSK modulated symbols associated with PBCH are mapped to predefined subcarrier locations as illustrated in Figure 2.12. Altogether, PBCH consists of 72 subcarriers over a duration of 4 symbols, not including subcarriers within the PBCH region that are dedicated to reference signals.

2.4.2 Uplink Control Information Format and Location

Just as (E)PDCCH carries the control information regarding resource allocation and link adaptation on the DL, on the UL this information is carried on the physical uplink control channel (PUCCH). In particular, PUCCH carries reports of channel quality information (CQI), UL scheduling requests, in which a UE asks for time-frequency resources on the uplink shared (data) channel UL-SCH (PUSCH), and HARQ ACK/NACK in response to a DL transmission. PUCCH is never transmitted simultaneously with the physical uplink shared channel (PUSCH) from the same UE. PUCCH supports multiple formats, as shown in Table 2.8.

PUCCH format 1 supports a positive scheduling request (SR) for PUSCH which indicates that the UE would like to be given some resources on the PUSCH shared uplink channel. PUCCH format 1a and 1b are used for reporting 1 or 2 bit HARQ ACK/NACK response of the DL transmission, respectively. PUCCH format 2 is used for the CQI reporting process. PUCCH formats 2a/2b are applicable for both the CQI reporting process and the HARQ ACK/NACK in the case of normal CP. The way that these formats are encoded is discussed in further detail in Section 5.4 of [30]. Finally,

Format	Modulation	# bits / Subframe	Purpose
1	N/A	N/A	Positive Scheduling Request (SR) for PUSCH
1a	BPSK	1 (1 Symbol)	1 bit HARQ-ACK
1b	QPSK	2 (1 Symbol)	2 bits HARQ-ACK
2	QPSK	20 (10 Symbol)	a CSI Report
2a	QPSK+BPSK	21 (10 + 1 Symbols)	a CSI and 1 bit HARQ-ACK (*Normal CP*)
2b	QPSK+QPSK	22 (10 + 1 Symbols)	a CSI and 2 bit HARQ-ACK (*Normal CP*)
* 3 (Rel.10-11)	QPSK	48 (24 Symbols)	Multiple CSI and HARQ-ACKs

Table 2.8: Supported PUCCH Formats (* From Rel. 10)

we note that from Release 10 and onward, in order to support more information for the channel state information (CSI) and a number of HARQ ACKs/NACKs under CA, PUCCH format 3 is introduced in [31] and [32].

The way to map PUCCHs into the UL resource grid is illustrated in Figure 2.17. In particular, a UE is mapped to a user index (m) based on its C-RNTI and some other parameters, in a manner defined in Section 5.4 of [11, 30–32] and Section 6.3.2 of [108–111]. It is important to note that, if, in response to a UL scheduling request it had sent previously on the PUCCH, a UE is granted use of a PUSCH UL data channel in a subframe, it will not utilize its PUCCH in the same subframe.

2.5 Time Frequency Resource Allocation

The principles of high rate wireless engineering dictate that in time-frequency scheduled systems, the usable time and frequency resources in the resource grid should be adaptively allocated in different manners in fast fading and slow fading environments for maximizing throughput [122–125]. In a slow fading environment, in order to maximize throughput, it is desirable to utilize those frequencies which the intended receiver has a good signal, yielding resource allocations which are localized

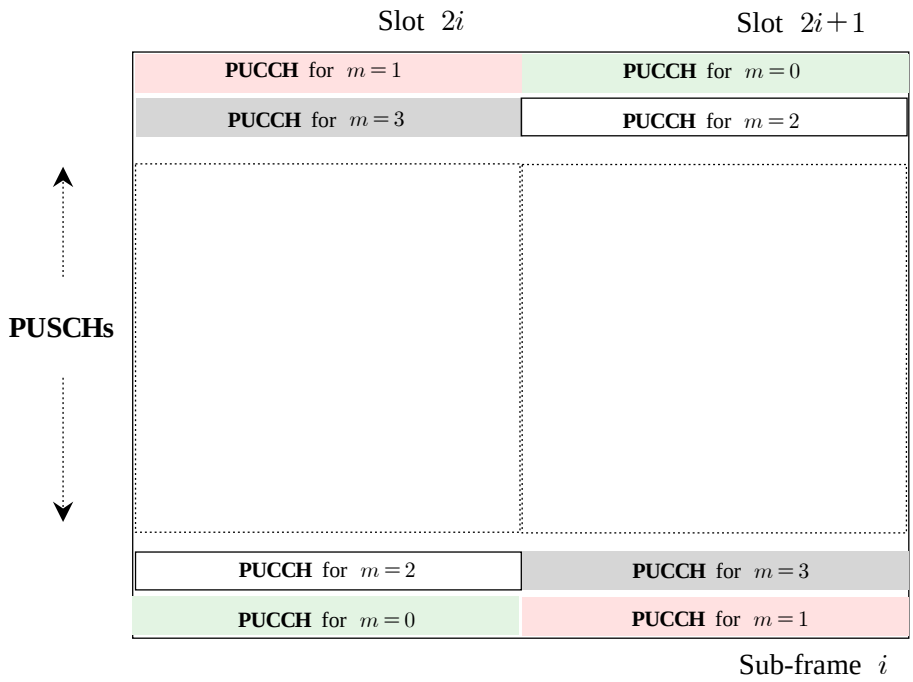


Figure 2.17: PUCCH Mapping

in frequency. Alternatively, in fast fading environments, in which the channel quality feedback is stale before it can be acted upon, distributed resource allocation, which diversifies the frequencies of transmission, is favored, as illustrated by Figure 2.18.

Another key general consideration in time frequency resource allocation signaling is how much freedom is allowed when allocating different collections of time frequency blocks to users. In general, a wider variety of resource assignment patterns may allow for higher throughput, but it also requires more bits to indicate which one is selected.

Finally, as in all cellular systems, there is an asymmetry between DL and UL when it comes to scheduling, and hence different resource allocation control decisions are allowed in these two cases.

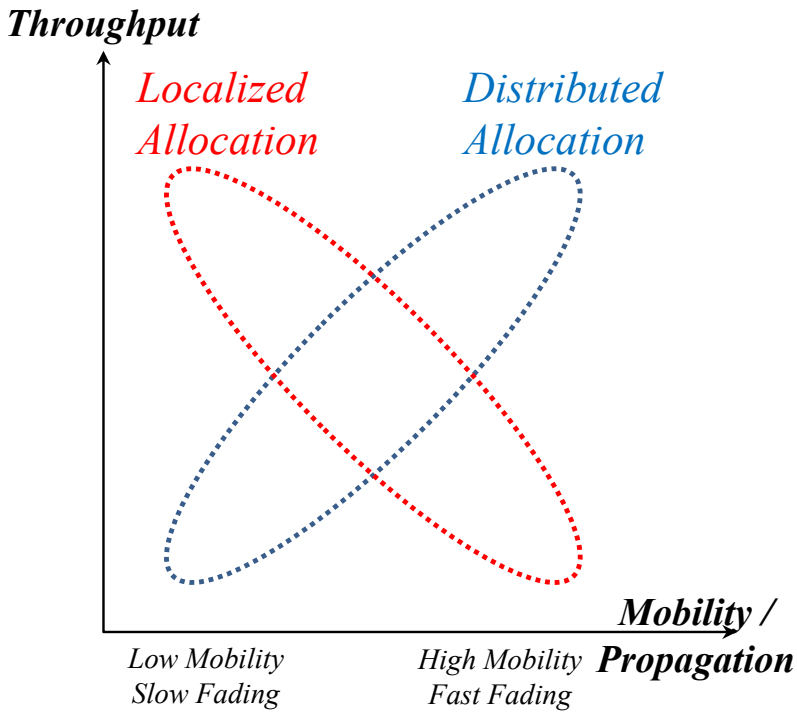


Figure 2.18: Channel Dependent Resource Allocations

2.5.1 DL Resource Allocation

In view of the considerations mentioned in the previous section (localized vs distributed and the amount of information necessary to encode an assignment), LTE supports three types of DL resource allocations: type 0, 1, and 2, which trade off these key concerns. Resource allocation types 0 and 1, which are encoded in DCI formats 1/2/2A/2B*(Release 9-11)/2C**(Release 10-11)/2D***(Release 11) as depicted in Table 2.7, specify which collection of time-frequency blocks are assigned to a user through a bit map: a binary value of 1 indicates an assignment to the user associated with this PDCCH, while a value of 0 indicates the lack thereof. The key difference between resource allocation type 0 and type 1 is that type 0 maps each bit to P consecutive PRBs, which are themselves already 12 subcarriers. Here, P is a parameter which depends on the overall system bandwidth as depicted in Figure

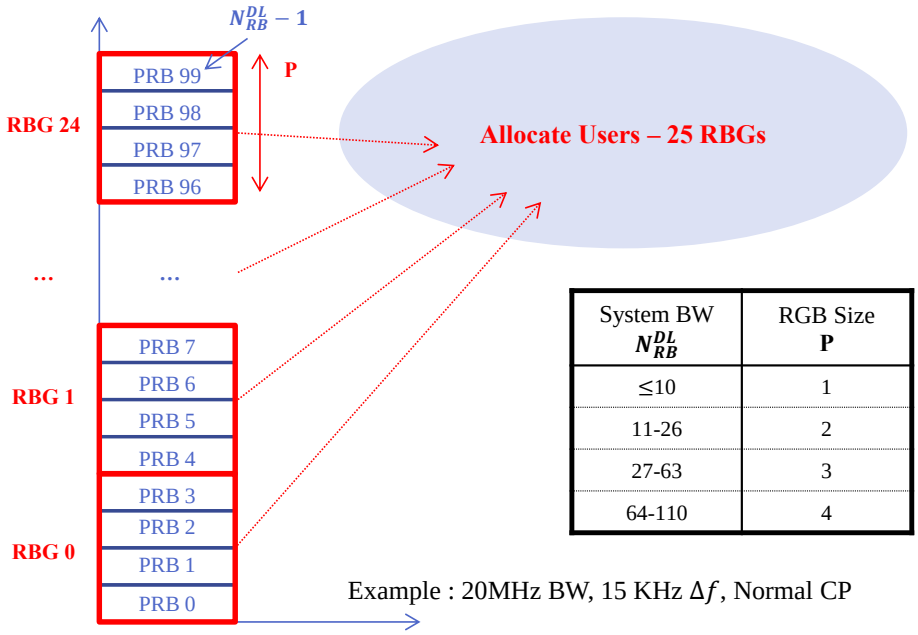
2.19. While this has the advantage that it requires only $\lceil \frac{N_{RB}^{DL}}{P} \rceil$ bits for the bitmap, it limits the total number of UEs which can be assigned resources in a single subframe to less than $\lceil \frac{N_{RB}^{DL}}{P} \rceil$.

The resource allocation type 1, on the other hand, allows for greater freedom when allocating PRBs to users. The resource blocks are still broken up into $\lceil \frac{N_{RB}^{DL}}{P} \rceil$ groups, however these groups are themselves collected into P evenly spaced subsets (via a standard modulo relationship). Resource allocation type 1 first selects one of the P subsets, then utilizes a bitmap within this subset, as depicted in Figure 2.20.

As both resource allocations type 0 and 1 utilize a bitmap component, they can require a relatively large number of bits. Additionally, while they can be utilized to obtain distributed resource allocations, they are both somewhat tailored to slow fading situations, as they are designed to allow for PRBs with high quality to be specifically selected for a user through the bitmap.

Resource allocation type 2, which is encoded in DCI formats 1A/1B/1C/1D, provides an alternative to type 0 and 1 that requires far fewer bits (only $\lceil \log_2 \left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2} \right) \rceil$) [33–36, 53, 54, 60, 61, 126], and still allows for both localized and distributed resource allocation. As depicted in Figure 2.21, it encodes a resource allocation of contiguous "virtual" resource blocks via an "offset" indicating a starting resource block and a length specifying the number of virtual resource blocks in the allocation. While these resource blocks have contiguous indices, distributed resource allocations are still enabled by allowing the virtual resource block indices to be mapped to PRBs via a permutation that introduces a gap between the PRBs mapped to by adjacent virtual resource blocks. The specific distributed pattern is defined in Section 6.2.3 of [30]. Alternatively, if a compact localized resource allocation is desired, one may map virtual resource blocks to PRBs of the same index.

Having outlined the possibilities for the three types of DL resource allocation, we



System BW N_{RB}^{DL}	RGB Size P
≤10	1
11-26	2
27-63	3
64-110	4

Figure 2.19: DL Resource Allocation Type 0

briefly describe the parts of the DCI formats they are encoded into.

Encoding DL Resource Allocation Type 0

The encoding of resource allocation type 0 by using DCI format 1 is summarized in Table 2.9. The DCI format 1 is used for DL scheduling information with UL control channel power offset. The first bit, the resource allocation header, is set to 0 to indicate that this is resource allocation type 0 (localized resource allocation). The $\lceil \frac{N_{RB}^{DL}}{P} \rceil$ bits are used for the localized resource allocation bitmap, indicating which of the $\lceil \frac{N_{RB}^{DL}}{P} \rceil$ blocks of P PRBs are allocated in the assignment. The remaining bits specify the modulation and coding scheme utilized on the DL, the HARQ process number indicating where to feedback the ACK/NACK after receipt of this transmission, a new data indicator for use with the HARQ process, a redundancy version indicating how many HARQ retransmissions have occurred for this data, an UL transmit power

Bits	DCI Format 1
0 or 3	**Carrier Indicator (Rel. 10-11)**
1	**Resource Allocation Header**: Set to **0**
$\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$	**Resource Allocation Type 0** $\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$: *Resource Assignment*
5	***MCS***
3 or 4	**HARQ Process number** : 3 for FDD, 4 for TDD
1	**New Data Indicator**
2	**Redundancy Version**
2	**TPC command for PUCCH**
2	**Downlink Assignment Index** : TDD
2	**HARQ ACK resource offset for EPDCCH (Rel. 11)**

Table 2.9: DL Resource Allocation Type 0 with DCI format 1

control (TPC) command for PUCCH (where the HARQ ACK/NACK in response to the DL transmission will be carried), and a downlink assignment index. The last one of these bits, the downlink assignment index, is only used for the TDD mode indicating which DL subframe within this frame (organized via a TDD configuration between DL and UL transmission as described in Section 2.3.1, Figure 2.7, Table 2.1) the data will appear.

The encoding of resource allocation type 0 by using DCI format 2/2A/2B/2C/2D is summarized in Table 2.10, from which it can be discerned that much of the formatting is shared with DCI format 1, the exception being that these formats add support for a closed/open loop MIMO option, and carry the necessary information for it [33–36]. These options will be discussed in further detail in Section 2.6.

Bits	DCI format 2/2A/2B*/2C*/2D*
0 or 3	**Carrier Indicator in Rel. 10-11**
1	**Resource allocation header** : Set to **0**
$\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$	**Resource Allocation Type 0** $\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$: *Resource Assignment*
2	**UL Power Control** (PUCCH)
2	**Downlink Assignment Index** (TDD DL/UL Configuration 1-6)
3 or 4	**HARQ process number** : 3 bits (FDD), 4 bits (TDD)
1 or 3	**1 bit** : Codeword Swap Flag or Scrambling Identity **3 bits** : Antenna port(s), Scrambling identity, # layer in Rel. 10-11
0 or 1	**SRS Request** only for TDD mode in Rel. 10-11
8+8	**For transport block 1 & 2** : 5 bits (*MCS*) + 1 bit (New data indicator)+ 2 bits (Redundancy version)
0, 2, 3, 6	**DCI Format 2 Closed Loop MIMO** : 3 (# Ant. ports 2), 6 (# Ant. ports 4) **DCI Format 2A Open Loop MIMO** : 0 (# Ant. ports 2), 2 (# Ant. ports 4)
2	**DCI Format 2D*** : PDSCH RE Mapping and Quasi-Co-Location Indicator
2	**HARQ-ACK resource offset only for EPDCCH in Rel. 11**

Table 2.10: DL Resource Allocation Type 0 with DCI format 2/2A/2B/2C/2D

Encoding DL Resource Allocation Type 1

The resource allocation type 1 uses $\lceil \frac{N_{RB}^{DL}}{P} \rceil$ bits for resource block assignment consisting of three fields to indicate the selected resource block group (RBG) subset ($\lceil \log_2(P) \rceil$ bits), a shift of the resource allocation span (1 bit), and the bitmap information for PRBs ($\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$ bits) [33–36, 53, 54, 60, 61, 87, 127]. As discussed in the beginning of this section, the PRBs are broken up into P groups via modulo P counting, and the RBG subset assignment ($\lceil \log_2(P) \rceil$ bits) indicates which of these groups the bit map is associated with. Then, within the subset of resource blocks, the remaining $\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$ bits each indicate (via 1 or 0) whether the associated PRB (within this RBG) is assigned in this resource allocation. Sometimes, $\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$ bits are not enough to have one bit for every PRB, so the 1

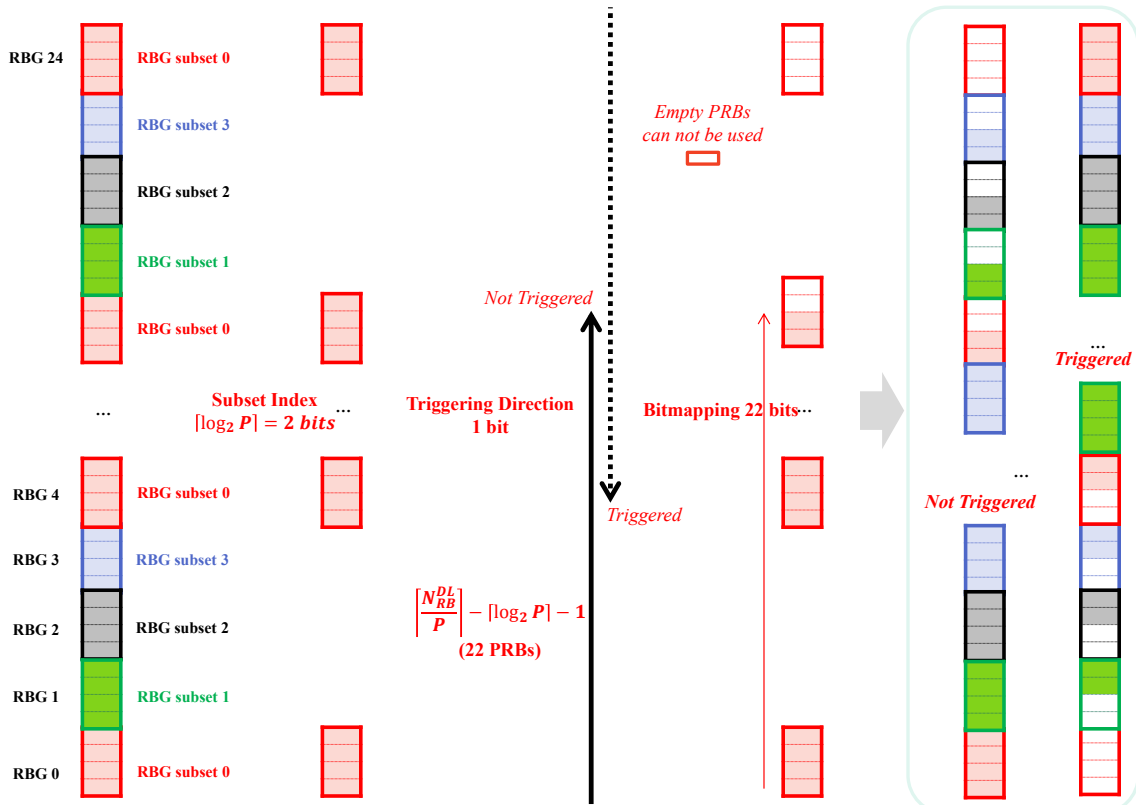


Figure 2.20: DL Resource Allocation Type 1 with 20 MHz System BW and 15 kHz Subcarrier Spacing

bit shift of the resource allocation span indicates whether the bits are to be mapped to the first $\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$ PRBs or to the last $\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$ PRBs.

For example, Figure 2.20 shows a type 1 distributed resource allocation with a 20 MHz system bandwidth, 15 kHz frequency spacing, and normal CP, and a group size $P = 4$. In this case the bitmap will consist of 22 bits ($\lceil \frac{N_{RB}^{DL}}{P} \rceil - \lceil \log_2(P) \rceil - 1$), but the PRB subset 0 has 28 PRBs, and the PRB subsets 1,2, and 3 each have 24 PRBs. As illustrated in Figure 2.20, the 1 bit "shift trigger" indicates in this instance that either the first 22 PRBs or the last 22 PRBs are allocated with the bit map.

The encoding of resource allocation type 1 by using DCI format 1/2/2A/2B/2C/2D is summarized in Tables 2.11 and 2.12. The contents of the other fields in this DCI

Bits	DCI Format 1
0 or 3	**Carrier Indicator (Rel. 10-11)**
1	**Resource Allocation Header**: Set to **1**
$\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$	**Resource Allocation Type 1** $\lceil \log_2(P) \rceil$ **bits** : *Selected Resource Block Subset* **1 bits** : a *shift of resource allocation span* $\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil - \lceil \log_2(P) \rceil - 1$ **bits** : *Resource Assignment*
5	***MCS***
3 or 4	**HARQ Process number** : 3 for FDD, 4 for TDD
1	**New Data Indicator**
2	**Redundancy Version**
2	**UL Power Control (PUCCH)**
2	**Downlink Assignment Index** : TDD
2	**HARQ ACK resource offset for EPDCCH (Rel. 11)**

Table 2.11: DL Resource Allocation Type 1 with DCI format 1

format will be discussed in Section 2.6.

Encoding DL Resource Allocation Type 2

Resource allocation type 2, whose encoding by DCI formats 1A/1B/1C/1D is summarized in Tables 2.13, 2.14, 2.15, and 2.16, aims to reduce the amount of control information necessary for the resource allocation by assigning a collection of resource blocks that are contiguous in index. Because the resource assignment is contiguous in index, only the starting index and the ending index or length need to be specified, thus significantly reducing the amount of control required relative to the bitmap based allocation. Either distributed or localized resource allocations can still be enabled in this case when using DCI formats 1A/1B/1D (collectively called "compact scheduling") via a selection of a one bit flag indicating whether the indices will be interpreted as virtual resource block (VRB) indices, which are then mapped in a

Bits	DCI format 2/2A/2B*/2C*/2D*
0 or 3	**Carrier Indicator in Rel. 10-11**
1	**Resource allocation header** : Set to **1**
$\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil$	**Resource Allocation Type 1** $\lceil \log_2(P) \rceil$ **bits** : *Selected Resource Block Subset* **1 bits** : a *shift of resource allocation span* $\left\lceil \frac{N_{RB}^{DL}}{P} \right\rceil - \lceil \log_2(P) \rceil - 1$ **bits** : *Resource Assignment*
2	**UL Power Control** (PUCCH)
2	**Downlink Assignment Index** (TDD DL/UL Configuration 1-6, Not FDD)
3 or 4	**HARQ process number** : 3 bits (FDD), 4 bits (TDD)
1 or 3	1 bit : codeword swap flag or Scrambling Identity **3 bits : DCI format 2C*/2D***: Antenna port(s), Scrambling identity, # layer
0 or 1	**SRS Request** only for TDD mode in Rel. 10-11
8+8	**For transport block 1 & 2** : 5 bits (*MCS)* + 1 bit (New data indicator) + 2 bits (Redundancy version)
0, 2, 3, 6	**DCI Format 2 Closed Loop MIMO** : 3 (#Ant. ports 2), 6 (#Ant. ports 4) **DCI Format 2A Open Loop MIMO** : 0 (#Ant. ports 2), 2 (#Ant. ports 4)
2	**DCI Format 2D*** : PDSCH RE Mapping and Quasi-Co-Location Indicator
2	**HARQ-ACK resource offset only for EPDCCH in Rel. 11**

Table 2.12: DL Resource Allocation Type 1 with DCI format 2/2A

permuted manner to PRBs, or the indices of the PRBs directly. In these compact scheduling modes, the encoding specifies a starting and ending resource block index using $\lceil \log_2 \left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2} \right) \rceil$ bits. DCI format 1C, which is named very compact scheduling, on the other hand, always utilizes VRB indices, and allocates the contiguous VRBs in groups of 2 or 4 (2 for 6-49 PRB system bandwidth, 4 for 50-110 PRB system bandwidth) contiguous ones at a time.

2.5.2 UL Resource Allocation

As discussed in Section 2.2, in the UL, LTE utilizes SC-FDMA in contrast to its use of OFDMA in the DL. This means that the UL resource allocations in LTE have significantly less flexibility than the DL resource allocations, as any allocation

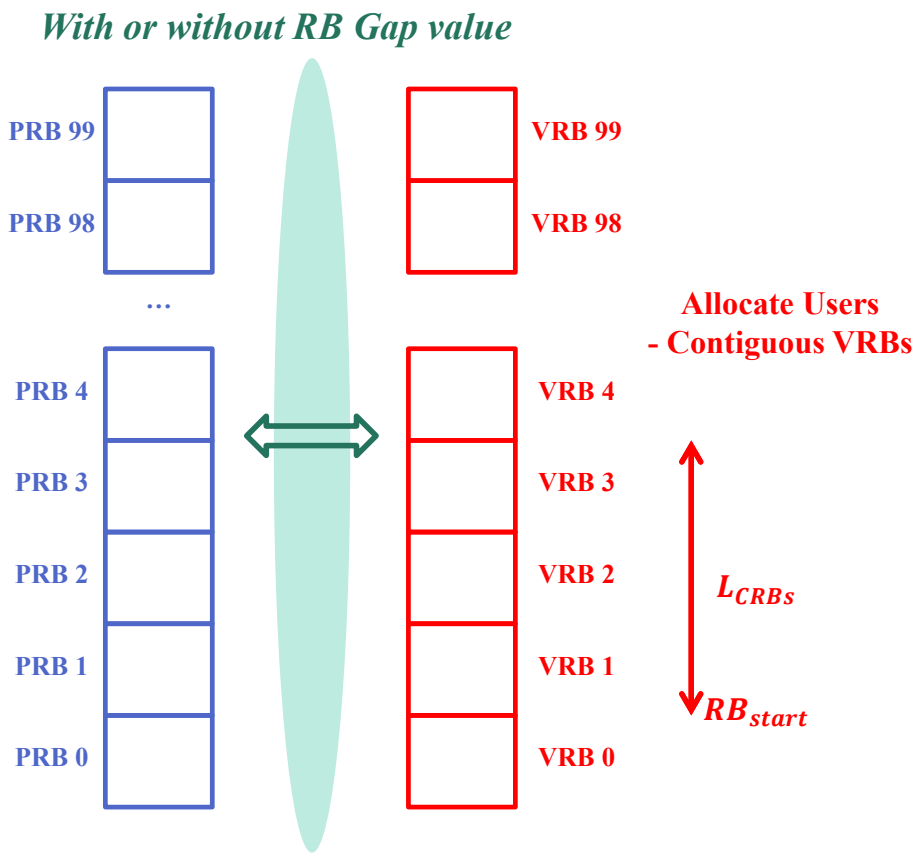


Figure 2.21: DL Resource Allocation Type 2 with 20 MHz System BW and 15 KHz Subcarrier Spacing

Bits	DCI Format 1A
0 or 3	**Carrier Indicator in Rel. 10-11**
1	Format Flag **: Set to 1**
-	**Random Access Initiated** - 1 bit : set to **0** (Localized VRB) - $\left\lceil \mathbf{log_2}\left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}\right)\right\rceil$ bits : Resource Assignment : **all bits set to 1** **Setting** : 6 bits (Preamble Index), 4 bits (PRACH Mask Index) **Otherwise** - 1 bit : **0** (Localized VRB) or **1** (Distributed VRB) - $\left\lceil \mathbf{log_2}\left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}\right)\right\rceil$ bits : Resource block assignment -- For distributed VRB and $N_{RB}^{DL} \geq 50$ and not associated with RA/P/SI-RNTI **1** bit for gap indicator + $\left\lceil \mathbf{log_2}\left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}\right)\right\rceil - \mathbf{1}$ bits - 5 bits : ***MCS*** - 3 or 4 bits : **HARQ process number** : 3 bits (FDD) , 4 bits (TDD) - 1 bit : **New data indicator** - 2 bits : **Redundancy version** - 2 bits : **UL Power Control** (PUCCH) - 2 bits **: Downlink Assignment Index** (TDD, not FDD) - **0 or 1 bit** : SRS Request in Rel. 10-11 - **2 bits** : HARQ ACK resource offset for EPDCCH in Rel. 11

Table 2.13: DL Resource Allocation Type 2 with DCI format 1A

of UL resources must be, at any given time, associated with a series of contiguous PRBs. This form of UL resource allocation is called UL resource allocation type 0 and continues to be supported in all of the releases of LTE (Rel. 8-11). LTE Advanced (Rel. 10-11) adds additional support for resource allocation type 1, which allows an allocation of two discontiguous "clusters" or blocks of contiguous PRBs to be allocated simultaneously continuing to utilize DFT-spread OFDM for the two clusters to keep the peak to average power ratio low. We presently discuss these two UL resource allocation types in more detail.

Bits	DCI Format 1B
0 or 3	**Carrier Indicator in Rel. 10-11**
1	**0** for Localized VRB or **1** for Distributed VRB
-	$\left\lceil \mathbf{log_2}\left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}\right)\right\rceil$ **bits : For distributed VRB and** $N_{RB}^{DL} \geq 50$ **1** bit for gap indicator + $\left\lceil \mathbf{log_2}\left(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}\right)\right\rceil - \mathbf{1}$ bits (Resource block assignment)
5	***MCS***
3 or 4	**HARQ process number** : 3 bits (FDD), 4 bits (TDD)
1	**New data indicator**
2	**Redundancy version**
2	**UL Power Control (PUCCH)**
2	**Downlink Assignment Index** : TDD
2 or 4	***MIMO (TPMI)*** : 2 (# Ant. 2), 4 (# Ant. 4)
1	***MIMO (PMI)***
2	**HARQ-ACK resource offset for EPDCCH in Rel. 11**

Table 2.14: DL Resource Allocation Type 2 with DCI format 1B

UL Resource Allocation Type 0

UL resource allocation type 0 is similar to DL resource allocation type 2 [33–36, 53, 54, 60, 61, 128]. Again, a resource allocation is specified as a collection of adjacent virtual resource blocks, encoded via offset and length indices. In UL resource allocation type 0, however, virtual resource blocks which are adjacent must map to PRBs which are adjacent, as SC-FDMA is the modulation used by the transmitters in the UL. Some diversity of frequency is still enabled, however, by shifting the virtual resource blocks up in frequency according to a "hopping" pattern, which is enabled with a flag in the associated DCI format 0. The specific hopping pattern is defined in Section 5.3.4 of [30]. Additionally, UL Power Control for PUSCH and CQI Request fields are included in DCI format 0 (these will be discussed in the next section). Although it does not support the hopping option, the DCI format 4 introduced in

Bits	DCI Format 1C
1	Gap value
-	***Resource Assignment*** : $\left\lceil \log_2\left(\left\lfloor \frac{N_{VRB,Gap}^{DL}}{N_{RB}^{Step}} \right\rfloor \left(\left\lfloor \frac{N_{VRB,Gap}^{DL}}{N_{RB}^{Step}} \right\rfloor + 1\right)/2\right) \right\rceil$ bits
5	***Transport Block Size in Rel. 8*** ***MCS in Rel. 9-11***

Table 2.15: DL Resource Allocation Type 2 with DCI format 1C

LTE Advanced can also be utilized to encode this localized UL resource allocation as specified in Table 2.18.

UL Resource Allocation Type 1 by DCI Format 4

As described in the introduction to this section, and depicted in Figure 2.23, UL resource allocation type 1 is a non-contiguous UL resource allocation within two clusters in LTE Advanced. The resource allocation information indicates to a scheduled UE two sets of PRBs with each set including one or more consecutive PRB groups of size P. In order to indicate two pairs of starting and ending points, one for each of the two clusters, 4 locations need to be specified in the resource allocation. Naturally, the starting location of the second cluster will not immediately follow the ending location of the first cluster, for this would be a normal single contiguous allocation and not a pair of clusters. Additionally, it is possible that one or both of the two clusters has only one RBG in it. If the first cluster has only one RBG in it, it indicates a phony ending location that is the resource block immediately preceding the starting RBG of the second cluster. If the second cluster has only one RBG in it, it indicates that its ending RBG is a phony RBG that is one past the last RBG. Given this manner of encoding, $\log_2\left(\binom{\lceil \frac{N_{RB}^{DL}}{P} \rceil + 1}{4}\right)$ bits are encoded to indicate the starting and ending locations of the two clusters.

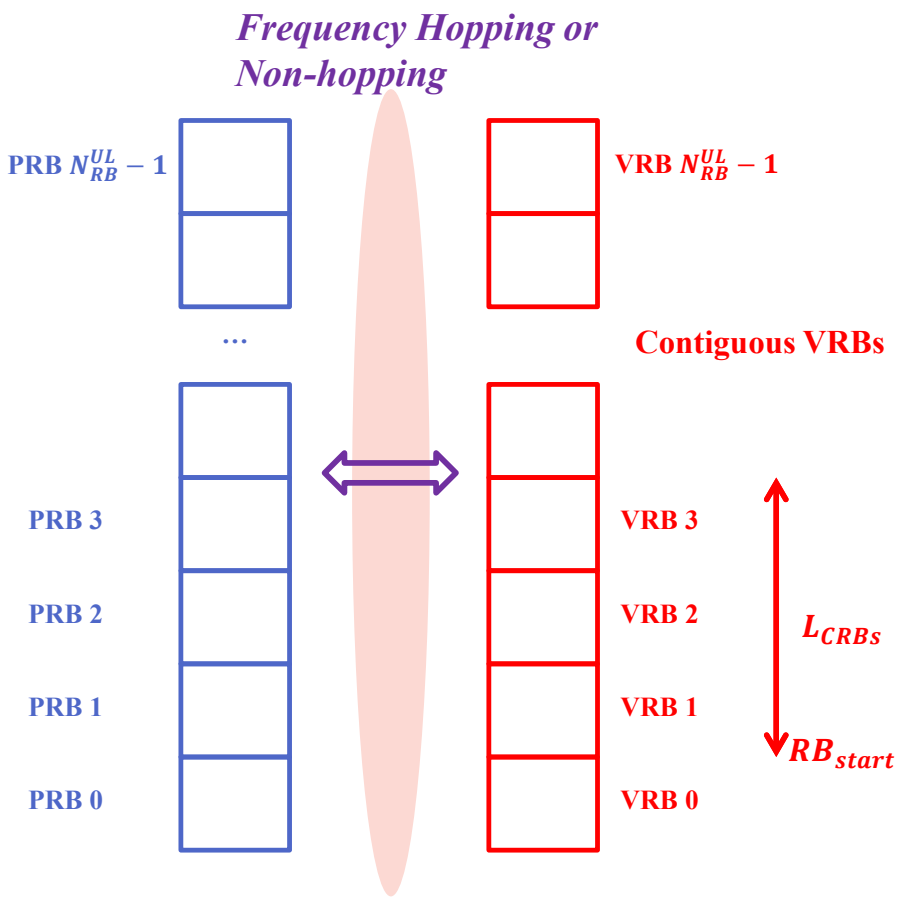


Figure 2.22: UL Resource Allocation Type 0

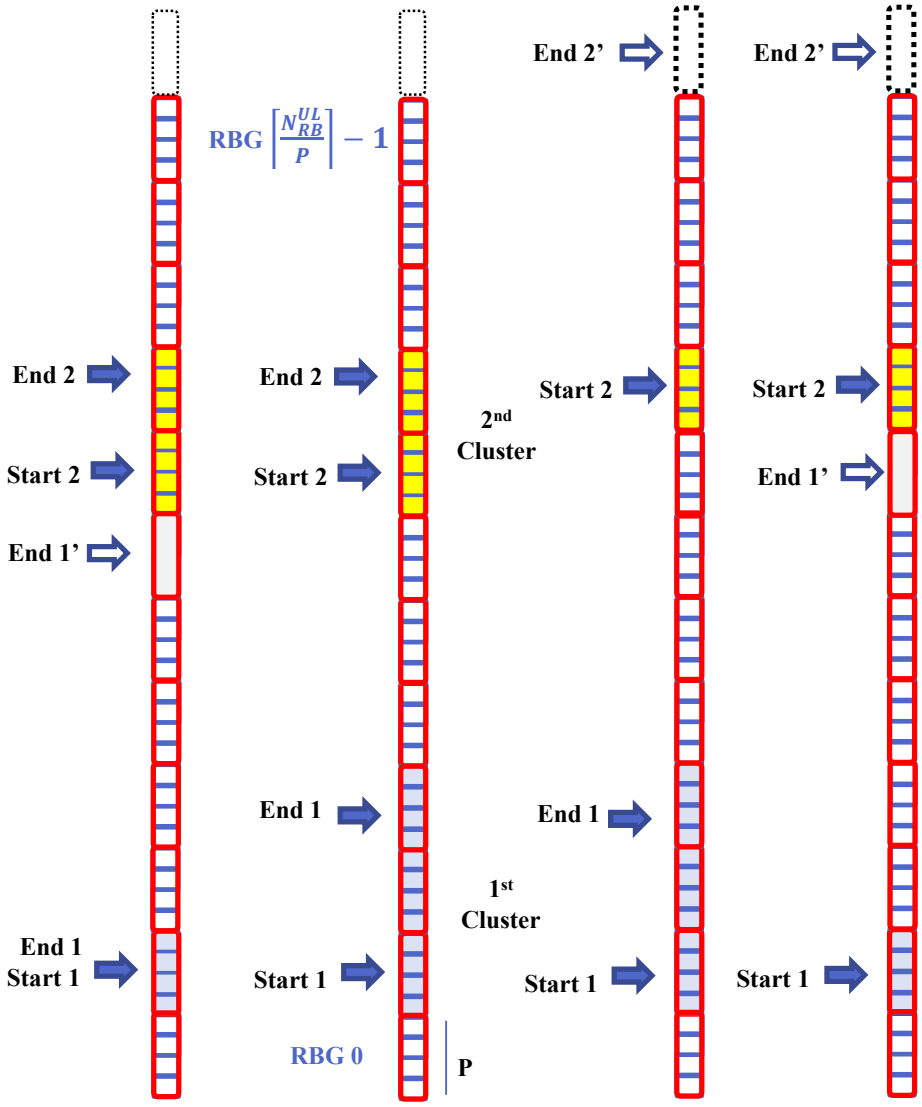


Figure 2.23: UL Resource Allocation Type 1

Bits	DCI Format 1D
0 or 3	**Carrier Indicator** in Rel. 10-11
1	**0** (Localized VRB) or **1** (Distributed VRB)
-	$\left\lceil \mathbf{log_2}(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}) \right\rceil$: **For distributed VRB and** $N_{RB}^{DL} \geq 50$ **1** bit for gap indicator + $\left\lceil \mathbf{log_2}(\frac{N_{RB}^{DL}(N_{RB}^{DL}+1)}{2}) \right\rceil - \mathbf{1}$ bits
5	***MCS***
3 or 4	**HARQ process number** : 3 bits (FDD), 4 bits (TDD)
1	**New data indicator**
2	**Redundancy version**
2	**UL Power Control (PUCCH)**
2	**Downlink Assignment Index** (TDD, not FDD)
2 or 4	***MIMO (TPMI)*** : 2 (# Ant. 2), 4 (# Ant. 4)
1	***Downlink power offset***
2	***HARQ-ACK resource offset only for EPDCCH in Rel. 11***

Table 2.16: DL Resource Allocation Type 2 with DCI format 1D

2.6 Link Adaptation

In order to enable a reliable transmission data rate and throughput over a radio link between an eNB and the UEs, LTE supports channel dependent link adaptation, which is mainly implemented with adaptive modulation and coding (AMC) and HARQ functionality [26, 129]. LTE DL and UL link adaptation are specified in sections 5.1.7.1 and 5.2.7.1 of the standard [26], respectively. Naturally, link adaptation is driven by CSI which is encoded into channel state indicators in LTE.

The CSI provided from the UEs to the eNB in LTE consists of three parts: a channel quality indicator (CQI), a precoding matrix indicator (PMI), and a rank indicator (RI). CQI is an index specifying the highest achievable modulation order (QPSK, 16QAM, and 64QAM) and code rate that can provide a packet error rate of

Bits	DCI Format 0
0 or 3	**Carrier Indicator in Rel. 10-11**
1	**Format Flag** : Set to **0**
1	**Frequency Hopping Flag** : Non-hopping (0) or Hopping (**1**)
-	$\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil$ bits **- Hopping** : $N_{UL,hop}$ bits : 1 ($6 \le N_{RB}^{UL} \le 49$) or 2 bits ($50 \le N_{RB}^{UL} \le 110$), Hopping Location Information $\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil - N_{UL,hop}$ **bits** : Resource Assignment in the first slot **- Non Hopping** : $\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil$ **bits** : Resource Assignment
5	***MCS and Redundancy Version***
1	**New Data Indicator**
2	**TPC Command of Scheduled PUSCH**
3	**Cyclic Shift for DM RS**
2	**UL Index : TDD**
2	**Downlink Assignment Index (DAI) : TDD**
1 or 2	1 bit of CQI Request in Rel. 8-9, 1 or 2 bits of CSI Request in Rel. 10-11
0 or 1	**SRS request in Rel. 10-11**
1	**Resource Allocation Type in Rel. 10-11**

Table 2.17: UL Resource Allocation Type 0 with DCI format 0

0.1 or less. In addition to CQI, to support MIMO communications, LTE allows UEs to feedback a predefined PMI. Additionally, LTE uses a RI to specify the number of "streams", or QAM symbols, transmitted for every channel use, for MIMO operation, thereby indicating which point to select in the tradeoff between spatial multiplexing and transmit diversity.

Based on this feedback from the UE regarding its channel qualities, eNB can select a transmission power level, and a modulation and coding scheme (MCS), as well as the resources to allocate the UE [130, 131]. The part of this process selecting the MCS index is called AMC, and relies on accurate CQI reporting to specify the link's current condition.

Bits	DCI Format 4 in Rel. 10-11
0 or 3	**Carrier Indicator in Rel. 10-11**
-	**Max(**$\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil$, $\left\lceil \log_2\left(\left(\left\lceil \frac{N_{RB}^{UL}}{P}+1\right\rceil\right)\right)\right\rceil$**) bits** **- Resource Allocation Type 0** $\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil$: Resource Assignment
2	**TPC Command of Scheduled PUSCH**
3	**Cyclic Shift for DM RS and OCC Index**
2	**UL Index : TDD**
2	**Downlink Assignment Index (DAI) : TDD**
1 or 2	**CSI Request**
2	**SRS request**
1	**Resource Allocation Type : Set to 0**
6+6	**For Transport Block 1 & 2** 5 bits (**MCS**) + 1 bit (**New Data Indicator**)
3 or 6	**Precoding Information and # layers** : 3 (# Ant. 2) or 6 (# Ant. 4)

Table 2.18: UL Resource Allocation Type 0 with DCI format 4

In addition to AMC, in order to reduce interference, LTE also implements TPC for UL transmission and adaptive transmission bandwidth via resource allocation [26]. In order to maintain the same SINR (Signal to Interference plus Noise Ratio) at the receiver guaranteeing fairness, adaptive transmission power should be considered to compensate for the link path loss, but this method can lead to severe interferences at the cell edge. In order to control the achievable data rate, the adaptive transmission bandwidth (resource assignment) of each UE is also useful for link adaptation.

In order to approach a high achievable data rate, LTE implements rate control through a rate matching process in which the AMC selects a coding rate and modulation, and HARQ information which selects a collection of bits in a circular buffer (CB) to enable extra redundant information transmissions when the initial transmissions are unsuccessful. In this section, we discuss these rate and power control processes

Bits	DCI Format 4 in Rel. 10-11
0 or 3	**Carrier Indicator**
-	**Max(**$\left\lceil \log_2\left(\frac{N_{RB}^{UL}(N_{RB}^{UL}+1)}{2}\right)\right\rceil$, $\left\lceil \log_2\left(\left(\left\lceil \frac{N_{RB}^{UL}}{P_4}+1\right\rceil\right)\right)\right\rceil$**) bits** **- Resource Allocation Type 1** $\left\lceil \log_2\left(\left(\left\lceil \frac{N_{RB}^{UL}}{P_4}+1\right\rceil\right)\right)\right\rceil$ bits
2	**TPC Command of Scheduled PUSCH**
3	**Cyclic Shift for DM RS and OCC Index**
2	**UL Index : TDD**
2	**Downlink Assignment Index (DAI) : TDD**
1 or 2	**CSI Request**
2	**SRS request**
1	**Resource Allocation Type : Set to 1**
6+6	**For Transport Block 1 & 2** 5 bits (**MCS**) + 1 bit (**New Data Indicator**)
3 or 6	**Precoding Information and # layers** : 3 (# Ant. 2) or 6 (# Ant. 4)

Table 2.19: UL Resource Allocation Type 1 with DCI format 4

in detail focusing on the salient CSI feedback, the allowable decisions, and how they are encoded.

2.6.1 Rate Control via AMC

As stated in the introduction, LTE provides two mechanisms for rate adaptation. The first of these, AMC [132–134], changes the modulation order among QPSK, 16-QAM, and 64-QAM, and adjusts the rate of the channel code. The channel code rate is adjusted at the eNB by modifying the transport block size (TBS) index, which in turn, dictates the subset of parity check and systematic information, generated by a rate 1/3 turbo code, that is transmitted. In the DL, both the modulation order and the transport block size index utilized for a given transmission are specified by the MCS index, which is transmitted in the PDCCH specifying the resource assignment

MCS	CQI	Modulation	Code rate	Spectral Efficiency	MCS	CQI	Modulation	Code rate	Spectral Efficiency
-	0	Out of range			15	9	16QAM	0.6015	2.4063
-	1	QPSK	0.0761	0.1523	16			0.6425	2.5684
0	2		0.1171	0.2344	17		64 QAM	0.4277	2.5684
1			0.1533	0.3057	18	10		0.4550	2.7305
2	3		0.1884	0.3770	19			0.5048	3.0264
3			0.2451	0.4893	20	11		0.5537	3.3223
4	4		0.3007	0.6016	21			0.6015	3.6123
5			0.3701	0.7393	22	12		0.6503	3.9023
6	5		0.4384	0.8770	23			0.7021	4.2128
7			0.5136	1.0264	24	13		0.7539	4.5234
8	6		0.5878	1.1758	25			0.8027	4.8193
9			0.6630	1.3262	26	14		0.8525	5.1152
10		16 QAM	0.3320	1.3262	27			0.8886	5.3349
11	7		0.3691	1.4766	28	15		0.9257	5.5547
12			0.4238	1.6953	29		QPSK	Reserved	Reserved
13	8		0.4785	1.9141	30		16QAM		
14			0.5400	2.1602	31		64QAM		

Table 2.20: Supported MCS Index Associated CQI

dictating the location of the transmission. The MCS index consists of five bits encoding the 28 different possibilities indicated in Table 2.20 adapted from Section 7.1 and 8.6 of [60]. The location of the MCS index within each DCI format is indicated in Tables 2.9–2.16. By decoding the DCI in the PDCCH associated with it, a UE learns not only which time frequency resource blocks are carrying its data, but also the modulation order and code rate with which to decode them. By combining the TBS index and the assigned resource bandwidth, each UE can determine the associated total transport block size according to 7.1.7.2 [60].

2.6.2 Rate Control via HARQ

HARQ enables the reliable delivery of information by combining error correction with error detection. In the DL, LTE uses a form of HARQ known as HARQ with

incremental redundancy [33–36, 135–137]. In this transmission technique, data is segmented into transmission blocks and coded both with a CRC error detection code, and an error correction code, in the case of LTE, a turbo code. The transmitter originally transmits the data segment with only a small amount of redundant parity information from the error correction code, and the receiver checks to see if the transmission, together with the inner error correction decoder, lead to no errors with the CRC error detection code. If no error is detected, the block is deemed to have been successfully transmitted, and an acknowledgment ACK is fed back to the transmitter. Alternatively, if an error is detected, the receiver feeds back a negative acknowledgment NACK to the transmitter, and the transmitter sends additional parity check bits as well as potentially retransmits some of the information bits that were already sent. The receiver combines the information received in the two rounds, again decodes with the outer error correction turbo code, and checks to see if there were any errors with the CRC error detection code. Again, an ACK is sent if the second transmission resulted in error free decoding, while an NACK is sent if it did not. Again, the NACK leads to (re)transmission of extra redundant information from the block. As it is implemented in LTE, up to three rounds of retransmissions are possible in the DL.

In the remainder of this section, we will provide more details regarding how the segments and retransmission are formed.

As explained above, the HARQ process in LTE consists of three steps: CRC insertion for error detection, Turbo channel coding for data error correction, and bit selection from a stored buffer as shown in Figure 2.24. At the upper layers, data for transmission is broken up into *transport blocks*, whose size ($16 - 75376$ [53, 54, 60, 61]) in bits can be calculated based on the MCS index and the number of assigned PRBs accompanying the assignment of resources to this data block. To each transport block

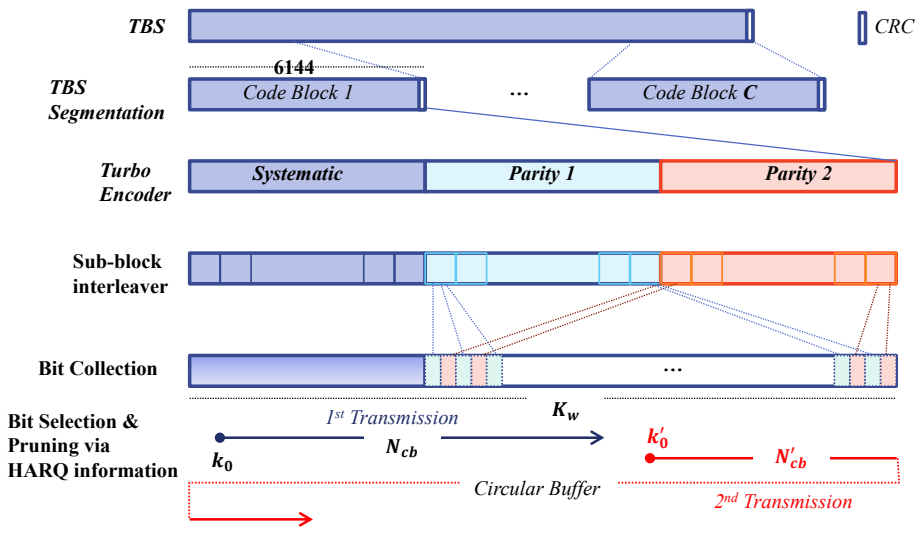


Figure 2.24: DL Rate Matching for Data by Circular Buffer

a CRC code of 24 bits is appended for error detection. However, for the purposes of HARQ, if the transport blocks are large (larger than 6144 bits), they can be broken into smaller pieces called *code blocks*, each consisting of 6120 bits and including an additional appended 24 bit CRC [138–140]. The use of these extra CRCs allows for code blocks to be processed in parallel [136].

In the second step, in order to provide error correction functionally of segmented code blocks, LTE uses turbo channel coding for data with a constant rate 1/3. The outputs of the turbo encoder including systematic, parity 1, and parity 2 are sub-block interleaved separately based on a predefined permutation with a sub-block length 32 to improve error performance. The last step is bit selection from the turbo coded and sub-block interleaved bits. Systematic bits are stored into a contiguous region in the CB, and sub-block interleaved parity bits are interlaced into the remaining portion of CB (even location for parity 1 and odd location for parity 2) to guarantee un-biased selection between two parities [135, 136, 141]. By the setting of a starting point (k_0) of information within the CB and bit selections (N_{cb}), LTE supports the rate control

and retransmission via the HARQ NACK. The two parameters have an important role for the rate control and retransmission process. The rate control can be achieved by the bit selection (N_{cb}) in the CB. The effective coding rate of a coded block is determined by the input bits to the turbo encoder and the bit selection (N_{cb}). In DL transmission, the bit selection ($N_{cb} = \min\{\lfloor \frac{N_{IR}}{C} \rfloor, K_w\}$) is based on a soft buffer size (N_{IR}) for the transport block of each user, the number of code blocks (C), and the CB size (K_w), but the bit selection in the UL transmission ($N_{cb} = K_w$) uses a fixed length of CB (K_w). The soft buffer size (N_{IR}) in DL can be defined as a function of the total number of soft channel bits according to the UE category [19, 67–69], the MIMO option [11, 30–32, 53, 54, 60, 61], and the maximum number of DL HARQ process [53, 54, 60, 61, 142].

The starting point of the bit selections (N_{cb}) within CB is indicated by the 2 bit HARQ redundancy version (0,1,2,3) within the DCI specifying a retransmission information. When the retransmission is required by the HARQ NACK which should be encoded by the repetition code (000 for NACK and 111 for ACK) via either PHICH for the response of UL transmission or PUCCH/PUSCH for the response of DL transmission, the transmitter needs to retransmit the coded bits that have not yet been transmitted in CB with another different starting point of information (k_0'). Because the redundancy information for UL transmission is always set to 0, the starting point in UL is a fixed point. These specific parameters are defined in Section 5.1.4 of [33].

LTE uses DL asynchronous adaptive HARQ and UL synchronous HARQ, where asynchronous and synchronous refer to the times associated with retransmissions [12, 26–28, 143–149]. The synchronous HARQ operation is restricted to retransmit at predefined time-instances relative to the original transmission, but the asynchronous HARQ operation may occur at any time [52, 150, 151]. For the design of DL/UL

HARQ, the delay of the retransmission process, the control signaling, and flexibility in scheduling retransmission are important considerations for the choice between synchronous and asynchronous operation [151]. The asynchronous HARQ operation provides flexible retransmission times, but requires more control information for scheduling the retransmissions, while the synchronous HARQ operation utilizes less control signaling than that of the asynchronous HARQ operation [151].

For each transmission on the DL, UE reports the UL HARQ ACK/NACKs in response to the DL transmission via PUSCH or PUCCH. After receiving these ACK / NACKs from UE, the eNB sends the HARQ process number and redundancy information when the retransmission is scheduled via the PDCCH. LTE supports HARQ soft-combining based on incremental redundancy. Under this incremental redundancy, if the UE receives a new transmission (redundancy version in PDCCH is 0) from the eNB, the UE should replace the data currently in the soft buffer for this TB with the received data, otherwise the UE combines the received data with the data currently in the soft buffer for this TB [143]. If the data in the soft buffer is successfully decoded for this TB, the UE sends a HARQ ACK to the eNB, otherwise the UE requests a retransmission with a HARQ NACK. In summary, based on the HARQ process number (3 bits in FDD, 4 bits in TDD) and redundancy version (2 bits, 0-3) in DCI formats, LTE achieves a DL rate matching process via HARQ.

As explained above, the HARQ responses to transmissions on the UL are synchronous. When the eNB receives a UL stream from a UE, the eNB synchronously reports the associated HARQ ACK/NACKs on the DL via the PHICH. When the UE receives a PDCCH (DCI format 0) regardless of the content of PHICH information (HARQ ACK or NACK), the UE should do either a new transmission or a retransmission. When the UE receives a HARQ ACK via PHICH with no PDCCH, the UE must keep its data in its HARQ buffer. To resume retransmissions, a PDCCH should

be received. When the UE receives HARQ NACK via PHICH, the UE should perform a non-adaptive retransmission based on the previous UL resource assignment.

As it would be expected, CSI plays a important role in the scheduling operation and in the HARQ operation. In the following subsection, we will deal with how CSI is reported in the LTE standard.

2.6.3 Channel State Information (CSI) Feedback

The CSI that can be reported by a UE in the LTE standard is broken up into three types of information: a channel quality indicator (CQI), a precoding matrix indicator (PMI), and a rank indication (RI) [60,61,152]. For single antenna transmission only the CQI feedback is necessary. The purpose of CQI reporting is to enable power control, channel dependent scheduling, and AMC. As it is defined in the LTE standard, the CQI index indicates the highest MCS which could be sustained under the present channel conditions, while simultaneously maintaining a transport block error probability not exceeding 0.1 [153–155]. In order to determine this, a UE measures the channel using the down reference signals which are modulated using QPSK as illustrated in Figure 2.25. The CQI index is chosen from a 4-bit CQI table which consists of 16 leveled indices with a 1.892 dB SINR granularity [156]. This table is mapped from the 5 bit MCS table, with two adjacent MCS values associated with every CQI value as shown in Table 2.20.

The CQI reported should be optimized differently for localized resource allocation and distributed resource allocation, because in localized allocation it is desirable to know each sub-band CQI information, while in distributed resource allocation it is sufficient to use only the average CQI. In particular, both wideband (for distributed resource allocation) and sub-band (for localized resource allocation) feedback modes are supported, as detailed in Figure 2.26 [157–160]. For distributed resource alloca-

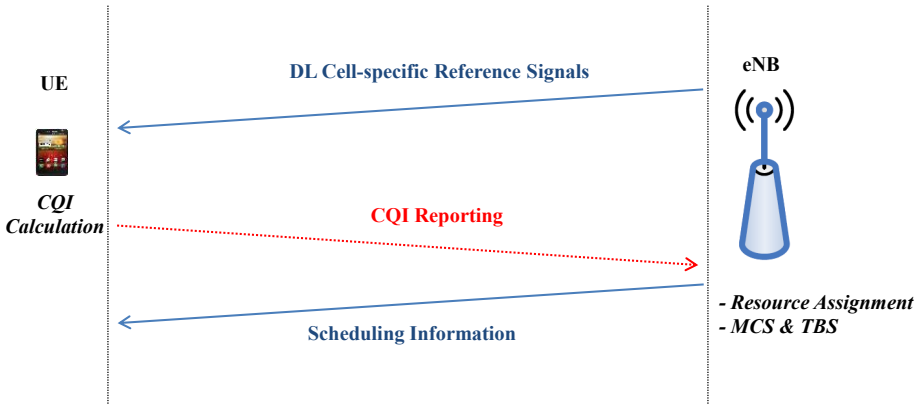


Figure 2.25: CQI Reporting Procedure

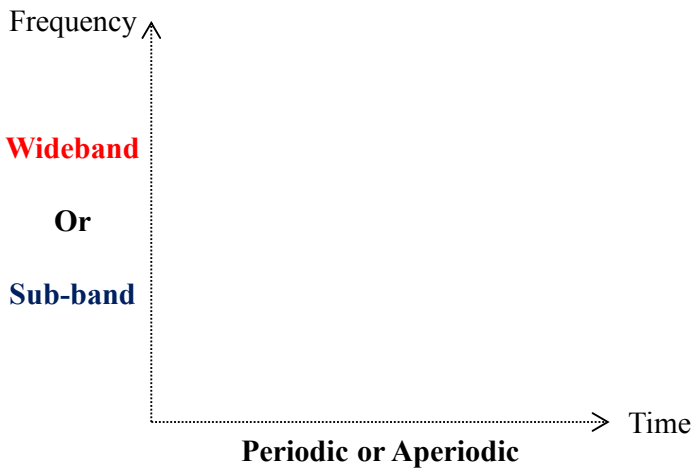


Figure 2.26: CQI Reporting Modes

tion, wide-band reporting selects a median CQI of an entire system bandwidth in the cell. For localized resource allocation, sub-band feedback reports the differential CQIs of the sub-bands with a wideband CQI. eNB selects a suitable CQI reporting mode for a UE based on channel variation.

Additionally, CQI reporting must be carefully designed in order to efficiently handle the tradeoff between control signaling overhead and link-adaptation performance. A key parameter involved in this careful design is the CQI reporting interval indicat-

System Bandwidth (N_{RB}^{DL})	Sub-band Size (k)
6-7	NA
8-10	4
11-26	4
27-63	6
64-110	8

Table 2.21: Supported Sub-band Size

ing how often the CQI information must be sent on the UL. The minimum reporting interval is the duration of a subframe (1 ms). However, LTE supports both periodic and aperiodic CQI reporting as it is illustrated in Figure 2.26 [53, 54, 60, 61, 161, 162]; these modes are selected for each UE by the eNB.

Aperiodic CQI Reporting via PUSCH

Aperiodic CQI reporting for a UE is triggered by a CQI request using DCI format 0, or a random access response grant from the MAC sub-layer. When these events occur, after waiting a given number of subframes, the UE reports its CQIs together with UL data on PUSCH as specified in Section 7.2.1 of [60]. Three types of reports are possible: wide-band CQI, higher layer configured sub-band CQI, and UE-specific sub-band CQI [53, 54, 60, 61, 163]. Here, a sub-band simply refers to a collection of k contiguous PRBs, where k may take values as shown in Table 2.21, implying that there are a $N = \lceil \frac{N_{RB}^{DL}}{k} \rceil$ sub-bands.

As depicted in Figure 2.27, in wideband CQI feedback mode, each UE reports a

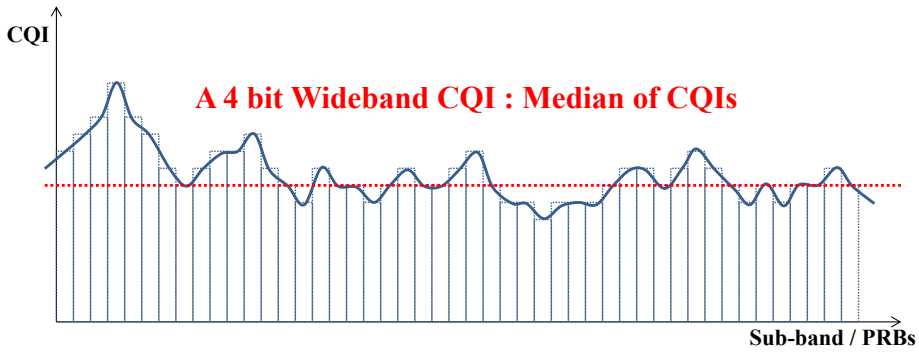


Figure 2.27: Wide-band Feedback

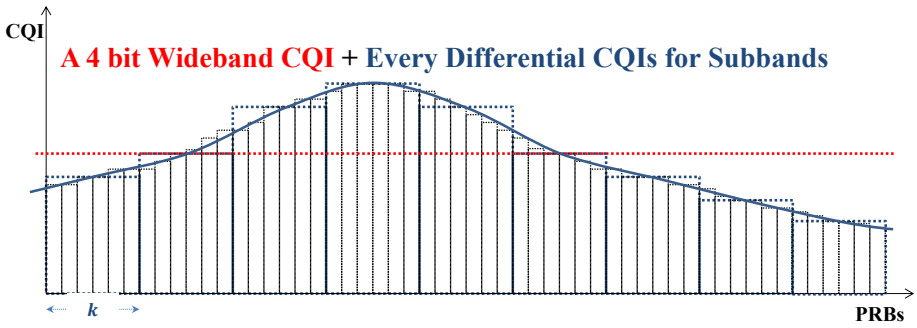


Figure 2.28: Higher Layer Configured Sub-band Feedback

single 4 bit CQI index corresponding to the median of the CQIs in each sub-band. Additionally, under MIMO operation, in wideband CQI mode, the precoding matrix indicator (PMI) is fed back for each of the N sub-bands, with a 2-bit PMI under the 2 antenna port operation and a 4-bit PMI under the 4 antenna port operation.

In the higher layer configured (sub-band CQI) feedback mode, the UEs report both the 4-bit wideband CQI, and the differential CQIs of each one of the N sub-bands, as shown in Figure 2.28 and Table 2.21 [53, 54, 60, 61, 164]. For MIMO operation, in addition to CQIs, under this mode a single PMI is feedback that is to be utilized across all of the N subbands.

Finally, in the UE-specific sub-band CQI reporting mode, the UEs report a 4 bit

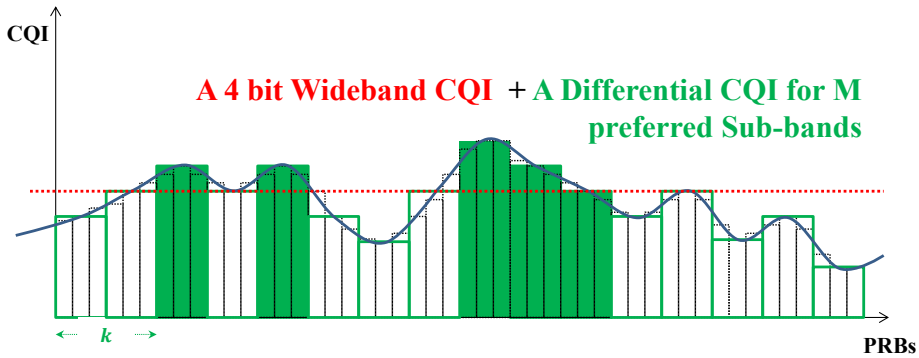


Figure 2.29: UE Selected Subband CQI Feedback

System Bandwidth (N_{RB}^{DL})	Sub-band Size (k)	Number of Sub-bands (M)
6-7	NA	NA
8-10	2	1
11-26	2	3
27-63	3	5
64-110	4	6

Table 2.22: Supported Selected Sub-bands

wideband CQI and the differential CQI for the M preferred sub-bands [53, 54, 60, 61, 165] which it selects as shown in Table 2.22 and Figure 2.29. In order to specify the sub-bands it selected to report the differential CQI for, the UE must feedback $L = \lceil \log_2 \binom{N}{M} \rceil$ bits in additional to the 4-bit wideband CQI and the 2-bit differential CQI. In the MIMO operation, a PMI can be fed back for each of the selected sub-bands.

Periodic CQI Reporting via PUCCH

In the periodic CQI reporting mode, CQI reports must fit within the 20 bits available in the PUCCH, as outlined in Table 2.8. This transmission occurs periodically every T sub-frames, where T can range over 2, 5, 10, 20, 32, 40, 64, 80, 128, and 160 sub-frames for the FDD mode. In the TDD mode, the possible reporting periods are a function of the selected UL/DL configuration type, but example values of possible T are 1, 5, 10, 20, 40, 80, and 160 sub-frames [53, 54, 60, 61, 166, 167].

There are two types of PUCCH CQI feedback, wideband and sub-band CQI reporting; which one is utilized is indicated via the CQI format indicator field. Aside from the periodicity of reporting, and the fact that the reports are put on PUCCH instead of PUSCH, the periodic wideband CQI is identical to the aperiodic wideband CQI reporting. A single 4-bit CQI that is the median CQI over the N sub-bands is fed back.

Under the periodic UE-selected sub-band CQI reporting mode, the sub-bands are further partitioned into J "bandwidth parts", and the UE selects a bandwidth part to report CQI. A bandwidth part j, $0 \leq j \leq J-1$, has consecutive frequencies and consists of N_j sub-bands as shown in Table 2.23. The number of sub-bands associated with J is either $\lceil \frac{N_{RB}^{DL}}{k \cdot J} \rceil$ or $\lceil \frac{N_{RB}^{DL}}{k \cdot J} \rceil - 1$ depending on N_{RB}^{DL}, k, and J. As such, in addition to the single 4-bit CQI of the selected bandwidth part, the selected part is indicated using $L = \lceil \log_2 \lceil \frac{N_{RB}^{DL}}{k \cdot J} \rceil \rceil$ bits.

2.6.4 Power Control

A key component of link adaptation is power control. In LTE, power control is performed through an open loop procedure in the DL, and though closed loop in the UL. Power control is important because although different sub-carriers within a given cell are orthogonal, the UE (UE* in Figure 2.31) still suffers from inter cell interference

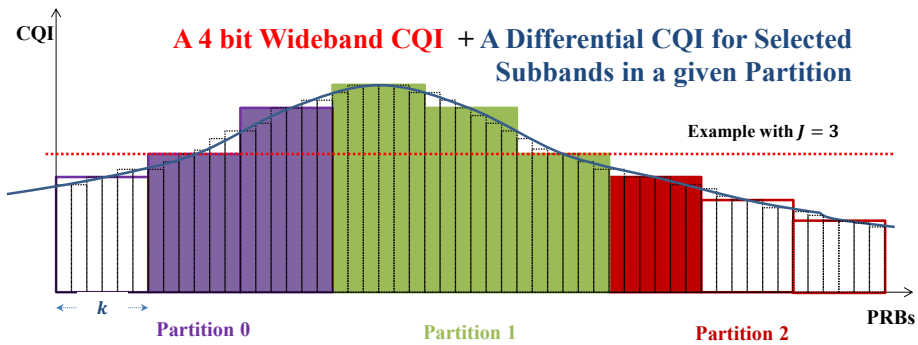


Figure 2.30: UE Selected Subband CQI Feedback

System Bandwidth (N_{RB}^{DL})	Sub-band Size (k)	Bandwidth Parts (J)
6-7	NA	NA
8-10	4	1
11-26	4	2
27-63	6	3
64-110	8	4

Table 2.23: Sub-band Size and Bandwidth Parts for Periodic Feedback

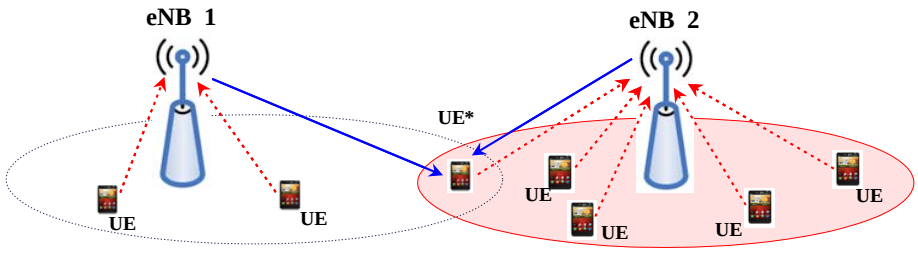


Figure 2.31: Interferences in DL and UL

between a UE and neighbor cells at the cell edge in the DL transmission and the eNB also suffers from interferences due to the errors from the RF hardware impairments (un-synchronization and power amplifier nonlinearity) of each UE as illustrated in Figure 2.31. LTE specifies a DL open loop power allocation based on the average UE energy and the UL closed loop power control based on average power over a SC-FDMA symbol [53, 54, 60, 61]. In the following subsections, we will investigate how to adjust the link status based on power control in the UL and the DL transmissions.

UL Power Control

UL power control is necessary in order to reduce the interference at the eNB caused by unsynchronized and RF-hardware impaired UE signals. It is performed in the LTE standard via a closed loop control algorithm which monitors the received power of the SC-FDMA signals at the base station, and then feeds back commands controlling the average transmit power ($P_{PUSCH,c}(i)$, $P_{PUCCH,c}(i)$, $P_{SRS,c}(i)$, and $P_{PRACH,c}(i)$) of the different UL physical channels [60] as shown in Figure 2.32. The received power is estimated at the UEs with the reference signal received power (RSRP), which is defined as the linear average power of the cell specific reference signal within the considered measurement frequency bandwidth [168–175].

The power control for the PUSCH, as depicted in Figure 2.33, is set via the

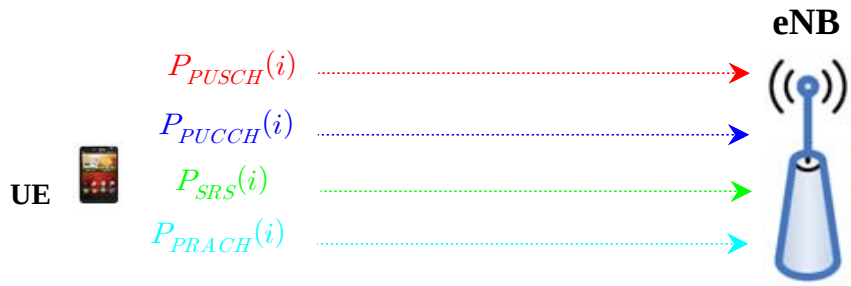


Figure 2.32: UL Power Control: Different UL physical channels have different Power Controls

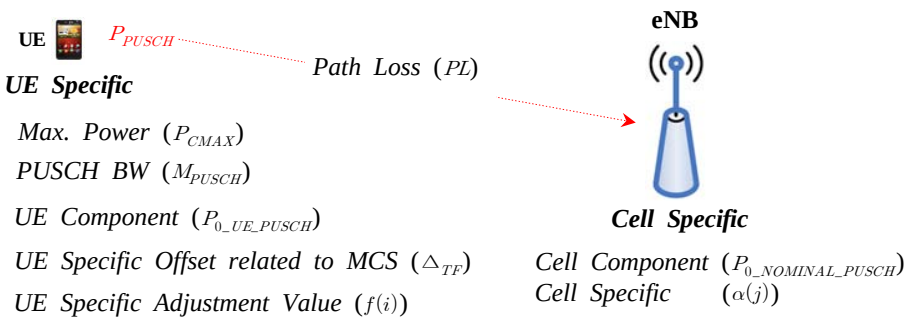


Figure 2.33: PUSCH Power Control

equation

$$P_{PUSCH,c}(i) = \min\{P_{CMAX,c}, \\ 10\log_{10}(M_{PUSCH,c}(i)) + \\ P_{O\_PUSCH,c}(j) + \alpha_c(j) \cdot PL_c + \\ \Delta_{TF,c}(i) + f_c(i)\} \quad [dBm] \tag{2.1}$$

as specified in Section 5.1.1 of [53, 54, 60, 61] and [173, 176, 177]. Here, $P_{CMAX,c}$ is the maximum transmit power of the UE class [67, 177]. Each UE class identifies a group of requirements for UE manufacturers to meet; the maximum transmit power is one of these requirements. Furthermore, formula (2.1) depends on the number of PRBs

($M_{PUSCH}(i)$) assigned to the UE, so that the transmit power of the PUSCH scales with the bandwidth [178]. Another key figure in (2.1) is the noise power adjustment ($P_{O\_PUSCH}(J)$) which reflects the compensation for noise at both the UE and the eNB, and is indexed according to a j selected by higher layers [60,172,177,179,180]. Perhaps the most important term in the equation is the partial path loss (PL) compensation, additionally adjusted with the compensation factor α between 0 and 1 to allow for flexible UL interference control. Additionally, the term $\Delta_{TF}(i)$ acts as a UE-specific power offset [175, 181–183]. This term $\Delta_{TF}(i)$ includes both power adjustments from derived TPC commands (provided a DCI format of 0/4/3 or 3A was used), as well as an adjustment based on the current MCS scheme employed. The power adjustment based on the current MCS scheme employed is selected via a table that is transmitted in the master information block on the PBCH.

Finally, in order to directly adjust the power during the closed loop feedback, LTE uses a UE specific correction value (either accumulated or absolute) ($f(i)$) that is communicated in the UL grant via the adjustment power level (δ_{PUSCH}) in the TPC command under DCI formats for UL grant (0/4) or power control (3/3A) as specified in Table 2.24 [60, 172, 172, 178].

In order to enable the eNB to perform the transmit power control algorithm at the eNB, the UEs also report the PHR (Power Headroom Report) to the eNB. This report indicates the unused transmit power of the UE, which is the difference between the configured transmit power and the required transmit power control at the UE [53, 54, 60, 61, 167, 184, 185].

The process of UL power control via the PUCCH bears a great resemblance to the process of power control via the PUSCH, and is depicted in Figure 2.34. It defines the various quantities involved in the power control process and outlines which parameters are cell-specific and UE-specific. The required transmit power of PUCCH

DCI Format	TPC Command	δ_{PUSCH} [dB]	
		Accumulated	Absolute (Only 0/4)
0/4/3	0	-1	-4
	1	0	-1
	2	1	1
	3	3	4
3A	0	-1	-
	1	1	-

Table 2.24: TPC commands and DCI formats for PUSCH

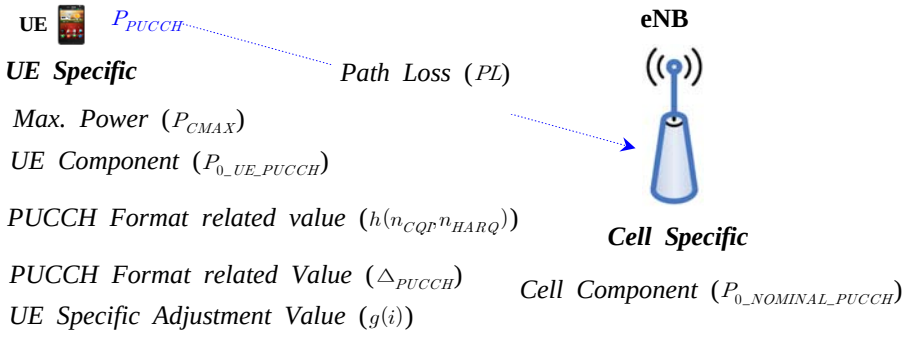


Figure 2.34: PUCCH Power Control

DCI Format	TPC Command	δ_{PUCCH} [dB]
1A/1B/1D/1/ 2A/2B/2C/2D/2/ 3	0	-1
	1	0
	2	1
	3	3
3A	0	-1
	1	1

Table 2.25: TPC commands and DCI formats for PUCCH

in a subframe (index i) is defined by (2.2) below, as specified in Section 5.1.2 of [60].

$$\begin{aligned} P_{PUCCH}(i) = \min\{&P_{CMAX}, h(n_{CQI} + n_{HARQ}) + \\ &P_{O\_PUCCH} + PL + \qquad [dBm] \\ &\Delta_{PUCCH}(F) + g(i)\}. \end{aligned} \tag{2.2}$$

An important element to recognize in this power control equation is that the power is a function of the total number of encoded bits, which, because this is the PUCCH, include both n_{CQI} bits for CQI information and n_{HARQ} bits for HARQ ACK/NACK information [53,54,60,61]. Additionally, a key difference between the PUCCH and the PUSCH is that there is no partial path loss compensation for the PUCCH, only full path loss PL compensation is allowed. The element of this equation most implicated in the closed loop power control process is $\Delta_{PUCCH}(F)$, which incorporates a MCS offset and a UE specific closed loop adjustment [60]. As it is indicated in greater detail in Table 2.25, this closed loop adjustment is specified in the TPC command for the PUCCH, which is in every DCI format except DCI from formats 0/4.

The standard has notably less to say about the power control for the physical ran-

dom access channel (PRACH) and the sounding reference signal (SRS) [60]. However, one should note that the transmit power setting for the SRS (P_{SRS}), is determined via the combination of the PUSCH power control information, the SRS transmission bandwidth, and other UE-specific parameters. Finally, the power control for the PRACH, which sets the transmit power PRACH (P_{PRACH}), is simply a function of the configured UE transmit power and the estimated UL path loss.

DL Power Allocation

The primary goal of DL power allocation in LTE is to mitigate inter-cell interference, especially at the cell edge, while provisioning fairly UE data rates. The DL power control is adjusted by allocating different power levels to different UEs by controlling the transmit energy per resource element (EPRE), which is defined as the average energy take over all constellation points for each UE [53, 54, 60, 61].

Although there is no specification about the key factors to determine the DL power level in the standard itself, a notable proposal from a standard group meeting is to derive DL power control via CQI based open-loop transmission power control [186]. The proposed method specifies how to decide DL transmission power for a UE based on its reported CQI. Based on the CQI feedback from each UE, the eNB allocates the transmission power level depending on the channel status, giving a higher power level for bad channel quality, and a lower power level for good channel qualities. This open-loop method does not require any additional control information, and may adjust the UEs' power levels under fast fading without using the history of the past transmission power setting.

In the DL power allocation control information specified in Release 8-11, each UE is given a pair of power ratios ρ_A and ρ_B by the eNB. These power ratios reflect a ratio of what PDSCH EPRE is used to the cell-specific reference signal's EPRE.

# Ant. Ports	**Type A** : ρ_A (Only PDSCH)		**Type B** : ρ_B (PDSCH + Cell-specific RS)	
	Normal CP	Extended CP	Normal CP	Extended CP
1 or 2	1, 2, 3, 5, 6	1, 2, 4, 5	0, 4	0, 3
4	2, 3, 5, 6	2, 4, 5	0, 1, 4	0, 1, 3

Table 2.26: Types of PDSCH REs corresponding to OFDMA symbol index within a non-MBSFN subframe

By measuring the RS EPRE, and utilizing this ratio, the UE can in turn calculate the PDSCH EPRE. There are two power ratios ρ_A and ρ_B because different OFDM symbol indices within a subframe are allocated different powers as shown in Table 2.26 [53, 54, 60, 61, 154, 187].

2.7 Conclusion

This chapter reviewed resource allocation and link adaptation in the LTE and LTE Advanced standards. We focused our discussion of the location, format, and encoding of the control decisions that the standards enable for resource allocation and link adaptation, as well as the location, format, and encoding of the feedback information that is salient for making these decisions. From the location of control channels in the LTE standard, we approximately estimate the portion of control information including reference signals in resource grid to be about 1/3 of DL transmission from Section 2.4.1. This is a surprisingly large amount of overhead, especially considering the high cost that many providers pay for LTE spectrum, which has been in the billions of US dollars in the United States. Hence, an important consideration is how to efficiently compress these control information. Bearing this in mind, we will develop a framework for optimizing control signal encoding to trade overhead for performance in the most efficient manner in the next chapter.

3. The CEO Problem with Independent Sources

As we discussed in the previous chapter, control information overhead supporting resource allocation and link adaptation utilizes a surprisingly large fraction of the time frequency resources in modern OFDMA systems, currently standing, for instance, at 1/3 of DL transmission in the LTE standard. Much of this DL control information tells users scheduling and adaptation decisions, including where to look in the time frequency resource grid for their data, and what coding rate and modulation it will be encoded with. To make the channel and user dependent scheduling decisions, on the UL control information is also used to support CQI reporting, enable scheduling requests, and to feed back HARQ information in the response to DL transmission. MIMO-OFDMA systems utilize control signaling in this manner with the aim of increasing the spectral efficiency on the fraction of time frequency resources spent on data transmission.

Time frequency resources spent on sending exclusively reference, measurement, and control signals can not be utilized for sending user data, hence it is desirable to encode them in a manner that minimizes the amount of control information dedicated resources necessary to achieve a target spectral efficiency performance. Here we aim to provide a framework for guiding how control signaling schemes can be designed to trade the amount of information exchanged for performance in the most efficient manner possible. In particular, we model the control signal design problem as an instance of a CEO lossy distributed source coding problem, the structure of which is depicted in Fig. 3.1. In this problem, a series of nodes encode local observations into messages which are forwarded to a fusion center, the central estimation officer (CEO), which uses the message to estimate an underlying sequence statistically related to the local observations.

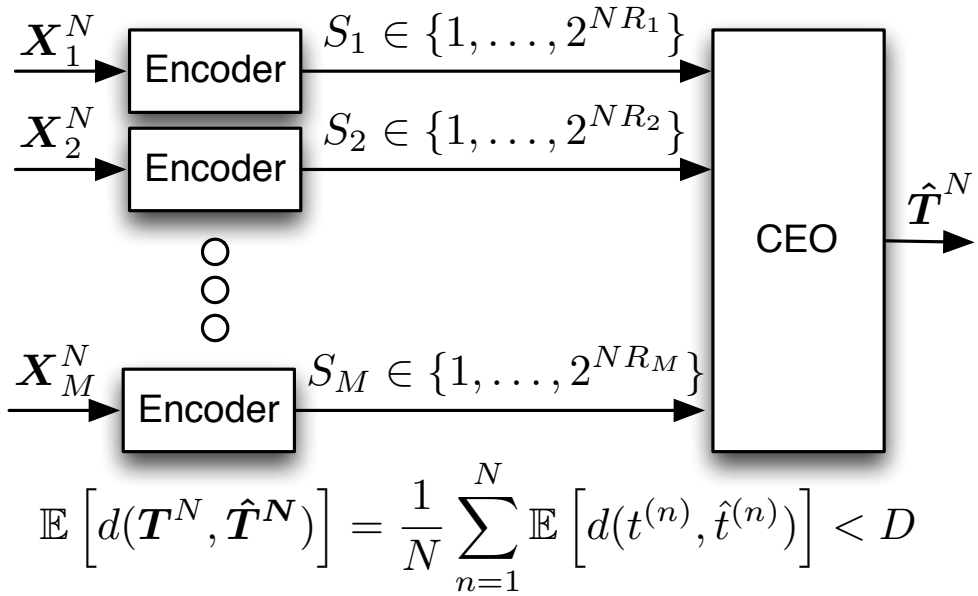


Figure 3.1: The Central Estimation Officer Problem.

To see why this model is relevant to resource allocation and link adaptation control signaling, observe that control signaling is necessary in a network primarily because the state of the network, embodied by channel and queue states, is not available in aggregate in any one place in the network (e.g. at the base station in a cellular system) for making decisions. Rather, each user knows the states of its own channels and queues, and in a cellular network, exchanges this information with a controller, the base station, to enable it to make decisions, which it must then forward back to the users. This local state information at the UEs/users can be treated as local observations at the nodes (each a UE/users) in a CEO problem, and the controller is modeled as the CEO. The channel qualities and queue states are encoded at each node into messages which are sent to the CEO, which must then lossily estimate the best control policy from them. The CEO forms its estimates by minimizing a distortion, which we model as the expected difference between the performance

that an omniscient controller, which had all of the channel and queue state in the network at its disposal, would achieve and the performance yielded by the CEOs resource decisions, conditioned on the information provided by the messages from the network nodes. Hence, the rate of the messages represents the control information overhead required for the control scheme, and the distortion the gap to the omniscient controller's performance.

In this chapter[1], with this motivation for the relevance of the CEO problem to the design of control signaling schemes, we will first review the expression for the rate distortion region for the CEO problem with independent sources. Next, we will describe a new method for computing this expression to obtain the CEO rate distortion region for particular sources and distortion metrics. The convergence and initialization of this method will then be discussed, with several examples illustrating the ideas. We then will illustrate how this method can be used to calculate fundamental overhead performance tradeoffs in wireless networks in the next Chapter 4.

3.1 Introduction and Motivation

As depicted in Fig. 3.1, in the central estimation officer (CEO) problem [188, 189] a series of sensors $m \in \mathscr{M} := \{1, \ldots, M\}$ make local observations $\boldsymbol{X}_m^N := \left[X_m^{(n)} \,|n \in \{1, \ldots, N\}\right]$, then encode these observations into rate limited messages, $S_m = f_m(\boldsymbol{X}^N) \in \{1, \ldots, 2^{NR_m}\}$, $m \in \mathscr{M}$, from which the central estimation officer, a fusion center, must form estimates of some statistically related sequence $\boldsymbol{T}^N = \left[T^{(n)} \,|n \in \{1, \ldots, N\}\right]$, jointly distributed with the observations $\boldsymbol{X}_{\mathscr{M}}^N = \left[\boldsymbol{X}_m^N \,|m \in \mathscr{M}\right]$ via the probability mass function

$$p_{\boldsymbol{T}^N \boldsymbol{X}_{\mathscr{M}}^N}(\boldsymbol{t}^N, \boldsymbol{x}_{\mathscr{M}}^N) = \prod_{n=1}^{N} p_{T,\boldsymbol{X}_{\mathscr{M}}}(t^{(n)}, \boldsymbol{x}_{\mathscr{M}}^{(n)}). \tag{3.1}$$

[1]The sections of this chapter will be submitted as a journal article to the *IEEE Transactions on Signal Processing*.

The estimates $\hat{\boldsymbol{T}}$ must yield a distortion $\mathbb{E}[d(\boldsymbol{T}, \hat{\boldsymbol{T}})]$, playing the role of a cost in a repeated Bayesian estimation, which does not exceed a desired value D, and the goal is to determine the fundamental tradeoff, over all possible codes, between the rates $\boldsymbol{R} = [R_m | m \in \mathscr{M}]$ and the distortion bound D.

A single-letter information theoretic expression for the rate distortion region for the generic CEO problem [188] is unavailable [189], with the exception of several special cases including the Quadratic Gaussian CEO problem [190, 191]. However, in the case when all of the sources are independent, so that (3.1) further simplifies to

$$p_{\boldsymbol{T}^N \boldsymbol{X}_{\mathscr{M}}^N}(\boldsymbol{t}^N, \boldsymbol{x}_{\mathscr{M}}^N) = \prod_{n=1}^{N} p_{T|\boldsymbol{X}}(t^{(n)} | \boldsymbol{x}^{(n)}) \prod_{m \in \mathscr{M}} p_{X_m}(x_m^{(n)}),$$

the rate distortion region simplifies considerably, and is equal to the closure of the convex hull of (3.2) below. Note that we will assume w.l.o.g. that $p_{X_i}(x_i) > 0$ for all $x_i \in \mathscr{X}_i$.

$$\left\{ (\boldsymbol{R}, D) \;\middle|\; \begin{array}{c} \forall m \in \mathscr{M} \quad R_m \geq I(X_m; U_m), \\ U_m \leftrightarrow X_m \leftrightarrow \boldsymbol{X}_{\mathscr{M}\setminus\{m\}}, \boldsymbol{U}_{\mathscr{M}\setminus m}, T, \\ \hat{T} \leftrightarrow \boldsymbol{U}_{\mathscr{M}} \leftrightarrow T, \boldsymbol{X}_{\mathscr{M}}, \\ \mathbb{E}[d(T, \hat{T})] \leq D \end{array} \right\} \tag{3.2}$$

Thm. 1: The closure of the convex hull of the rate region (3.2) is the rate distortion region for the given CEO problem with independent sources.

Proof: Let $f_m : \mathscr{X}_m^N \to \{1, \ldots, 2^{NR_m}\}$ and $g : \prod_{m \in \mathscr{M}} \{1, \ldots, 2^{NR_m}\} \to \hat{\mathscr{T}}^N$ be a

series of encoders and decoders, respectively, achieving the expected distortion D.

$$\begin{aligned} NR_m &\geq H(f_m(\boldsymbol{X}_m^N)) \geq I(f_m(\boldsymbol{X}_m^N); \boldsymbol{X}_m^N) \\ &= H(\boldsymbol{X}_m^N) - H(\boldsymbol{X}_m^N | f_m(\boldsymbol{X}_m^N)) \\ &= \sum_{n=1}^{N} H(X_n) - H(X_n | \boldsymbol{X}^{n-1}, f_m(\boldsymbol{X}_m^N)) \\ &\geq \sum_{n=1}^{N} I(X_n; f_m(\boldsymbol{X}_m^N)) = \sum_{n=1}^{N} I(X_n; U_m^{(n)}) \end{aligned} \tag{3.3}$$

where $U_m^{(n)} = f_m(\boldsymbol{X}_m^N)$. Furthermore, defining $\hat{T}_n = g^{(n)}(f_1(\boldsymbol{X}_1^N), \ldots, f_M(\boldsymbol{X}_M^N))$, the distortion obtained is

$$D \geq \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[d(T_n, \hat{T}_n)] \tag{3.4}$$

But this shows that R_m and D can be expressed a convex combination of a series of mutual informations and distortions for the random variables $U_m^{(n)}, \hat{T}_n$, which as defined also obey the Markov conditions in (3.2).

To determine the cardinality bounds necessary for the random variables U_i, $i \in \mathscr{M}$, we will show using the Carathéodory-Fenchel theorem [192, 193] that no matter what the cardinality of U_i is, we can achieve the same rates and distortion with a U_i of cardinality $|\mathscr{X}_i|+1$. Let $\mathscr{X}_i = \{a_1, \ldots, a_{|\mathscr{X}_i|}\}$, and define the following series of $|\mathscr{X}_i|-1$ functions mapping $\mathscr{P}(\mathscr{X}_i)$ (the set of probability distributions on $\mathscr{X}_i$) to $\mathbb{R}$

$$F_k(q_{X_i}) := q_{X_i}(a_i), \ \forall k \in \{1, \ldots, |\mathscr{X}_i - 1\}, \ \forall q_{X_i} \in \mathscr{P}(\mathscr{X}_i) \tag{3.5}$$

also define the two additional functions

$$F_{|\mathscr{X}_i|}(q_{X_i}) := \sum_{x_i \in \mathscr{X}_i} -q_{X_i}(x_i) \log_2(q_{X_i}(x_i)), \ \forall q_{X_i} \in \mathscr{P}(\mathscr{X}_i) \tag{3.6}$$

and $F_{|\mathscr{X}_i|+1}(q_{X_i}) :=$

$$\min_{\hat{t}} \sum_{\boldsymbol{x}} d(t, \hat{t}) p_{T|\boldsymbol{X}}(t|\boldsymbol{x}) q_{X_i}(x_i) \prod_{j \neq i} p_{U_j|X_j}(u_j|x_j) p_{X_j}(x_j) \tag{3.7}$$

Let $p_{\tilde{U}_m|X_m}(\cdot|\cdot)$ be any conditional distribution, where $\tilde{U}_m$ is in a set $\tilde{\mathscr{U}_i}$ of arbitrary cardinality. It, together with the source distribution $p_{X_i}(\cdot)$, induces a series of $|\tilde{\mathscr{U}_i}|$ distributions $\{p_{X_i|\tilde{U}_i}(\cdot|\tilde{u}_i)|\tilde{u}_i \in \tilde{\mathscr{U}_i}\}$ each in $\mathscr{P}(\mathscr{X}_i)$ and a $p_{\tilde{U}_i}(\tilde{u}_i)$ such that $\forall k \in \{1, \ldots, |\mathscr{X}_i| - 1\}$

$$\sum_{\tilde{u}_j \in \tilde{\mathscr{U}_j}} p_{\tilde{U}_i}(\tilde{u}_i) F_k(p_{X_i|\tilde{U}_i}(\cdot|\tilde{u}_i)) = p_{X_i}(a_k), \text{ and} \tag{3.8}$$

$$\sum_{\tilde{u}_j \in \tilde{\mathscr{U}_j}} p_{\tilde{U}_i}(\tilde{u}_i) F_{|\mathscr{X}_i|}(p_{X_i|\tilde{U}_i}(\cdot|\tilde{u}_i)) = H(X_i|\tilde{\mathscr{U}_i}) \tag{3.9}$$

$$\sum_{\tilde{u}_j \in \tilde{\mathscr{U}_j}} p_{\tilde{U}_i}(\tilde{u}_i) F_{|\mathscr{X}_i|+1}(p_{X_i|\tilde{U}_i}(\cdot|\tilde{u}_i)) = \min_{\hat{t}(\cdot)} \mathbb{E}[d(T, \hat{t}(\tilde{\boldsymbol{U}}_{\mathscr{M}}))] \tag{3.10}$$

Stacking the F_k's into a vector function $\boldsymbol{F} : \mathscr{P}(\mathscr{X}_i) \to \mathbb{R}^{|\mathscr{X}_i|+1}$, we observe that (3.8), (3.9), and (3.10) can be viewed as a convex combination of points in $\boldsymbol{F}(\mathscr{P}(\mathscr{X}_i))$ with coefficients $p_{\tilde{U}_i}(\tilde{u}_i)$. Because $\boldsymbol{F}(\mathscr{P}(\mathscr{X}_i))$ is a bounded connected set in $\mathscr{R}^{|\mathscr{X}_i|+1}$, the Carathéodory-Fenchel theorem implies that any such convex combination of points in $\boldsymbol{F}(\mathscr{P}(\mathscr{X}_i))$ can be represented as a convex combination of at most $|\mathscr{X}_i| + 1$ points in $\boldsymbol{F}(\mathscr{P}(\mathscr{X}_i))$. Let these points be $\boldsymbol{F}(p_{X_i|U_i}(\cdot|u_i)), u_i \in \mathscr{U}_i = \{1, \ldots, \mathscr{X}_i + 1\}$, and let their convex combination be $p_{U_i}(u_i)$. Then, these $p_{X_i|U_i}(\cdot|u_i)$ and $p_{U_i}(u_i)$ yield a $p_{U_i|X_i} = p_{X_i|U_i} p_{U_i} / p_{X_i}$ that give the same point in the rate distortion region as $p_{\tilde{U}_i|X_i}$, and which is defined over the support $\mathscr{U}_i = \{1, \ldots, |\mathscr{X}_i| + 1\}$. This shows that it suffices to take $|\mathscr{U}_i| = |\mathscr{X}_i| + 1$.

Next, consider any distribution obeying the conditions in (3.2) together with R_i obeying the inequalities therein. Draw 2^{NR_i} length N sequences I.I.D. from each

marginal distribution p_{U_i}. To encode an observed sequence $\boldsymbol{X}_m^N$, select one of the 2^{NR_i} sequences that is ϵ jointly strongly typical with $\boldsymbol{X}_m^N$. Such a sequence will exist with arbitrarily high probability as $\epsilon \to \infty$ owing to the fact that $R_i > I(X_i; U_i)$. Send the index of this sequence to the decoder. At the decoder, draw the $\hat{t}^{(n)}$ from the selected conditional distribution

$$\prod_{n=1}^{N} p_{\hat{T}|\boldsymbol{X}}(\hat{t}^{(n)}|\boldsymbol{u}^{(n)}) \tag{3.11}$$

where $\boldsymbol{u}^{(n)}$ is the nth element of each of the M selected sequences from the encoders. By the Markov lemma, this sequence will be strongly jointly typical with the observed sequences $\boldsymbol{X}^N$, and hence the distortion obtained will can be made as close to the expected distortion as desired through selection of ϵ and N, hence obeying the distortion bound D. ■

This expression has not been mentioned in the literature primarily because much of the interest in multi-terminal source coding is in exploiting dependent sources, with the seminal work [188, 194, 195] angling to understand that class of problems. Additionally, as the proof above shows, from a pure information theory perspective, it is quite direct to derive (3.2) from elementary arguments from even introductory rate distortion theory [196].

However, as we outline in Section 3.5, there are in fact a wide variety of real world problems which can be modeled CEO problems with independent sources, and for which the rate distortion region is of fundamental interest. For instance, the whole class of independent lossy function computation problems, in which the independent sources want to send the minimum amount of information necessary for the CEO to compute a function of them in a lossy manner, falls into this class of problems.

While the exact information theoretic expression is available for this class of problems, given a particular distribution for the sources and a particular distortion metric,

the rate distortion region must still be computed. This is in general a difficult task if one sets about attempting to solve the associated optimization directly. In the classic single source direct case, finding the rate distortion function is equivalent to solving a convex optimization problem, which can be solved via the globally convergent [197–199] iterative Blahut Arimoto algorithm [200, 201]. The natural question then if there is some extension of the Blahut Arimoto algorithm that can be utilized to calculate the rate distortion function in the CEO with independent sources context.

In this chapter, we give in Section 3.2 an answer to this question in the form of an alternating minimization algorithm for calculating the rate region for the CEO problem with independent sources, which, for a fixed collection of Lagrange multipliers dictating the tradeoffs between the rates and the distortion, is shown in Section 3.3 to be guaranteed to converge to the minimum of the Lagrangian when it is convex. The minimization steps in the alternating minimization are themselves ordinary Blahut-Arimoto algorithms, albeit with iteration dependent exotic distortion measures. We also show that a key property of the original Blahut-Arimoto algorithm, namely that it always converges to the *global* optimum of the Lagrangian, is in fact lost after the generalization, owing to the possibility that the distortion and Lagrangian are not convex. An example in Section 3.5 demonstrates that the value of the Lagrangian that is reached depends on the initialization of the algorithm in these cases. As such, careful, or at least repeated random, initialization of the algorithm is necessary to obtain superior local optima for problems yielding non-convex distortions and Lagrangians. We provide in Section 3.3 a method of careful initialization based on an explicit solution for the global optimum at the minimum distortion point. By initializing nearby Lagrange multipliers at this point, we can remain in the region of attraction for their global optima, and in this manner trace out the rate distortion region. The remaining examples in Section 3.5 serve to demonstrate the applicability

of the independent CEO model to a variety of real world problems, and the utility of the proposed computational method in calculating rate distortion tradeoffs in these scenarios.

3.2 Computing the Rate Distortion Region

The problem of computing the rate distortion region (3.2) is equivalent, up to a following convex hull, to selecting $p_{U_i|X_i} \forall i \in \mathscr{M}$ and $p_{\hat{T}|\boldsymbol{U}_{\mathscr{M}}}$ to minimize

$$\sum_{i \in \mathscr{M}} \lambda_i I(X_i; U_i) + \mu \mathbb{E}[d(T, \hat{T})] \tag{3.12}$$

where the Lagrange multipliers $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\mu \geq 0$ are parameters that will be utilized to trace out the boundary of rate distortion region. We will discuss how to select the series of Lagrange multipliers, then convert the results from minimizing the Lagrangian with respect to them in Section 3.4. Here, we provide an algorithm, which, for a particular collection of Lagrange multiplies $\boldsymbol{\lambda}, \mu \geq 0$ will converge to the minimum of the Lagrangian.

Aiming to generalize the Blahut-Arimoto algorithm [200, 201], we begin with the standard relaxation that introduces a separate variable $Q_i(u_i|x_i)$ and $q_i(u_i)$ for $p_{U_i|X_i}$ and for p_{U_i}, respectively. This yields the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda) =$

$$\sum_{i \in \mathscr{M}} \lambda_i \sum_{(x_i, u_i)} p_{X_i}(x_i) Q_i(u_i|x_i) \log \left(\frac{Q_i(u_i|x_i)}{q_i(u_i)} \right) + \\ \mu \sum_{\boldsymbol{u}, \boldsymbol{x}, t, \hat{t}} d(t, \hat{t}) p_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \Lambda(\hat{t}|\boldsymbol{u}) \prod_{i \in \mathscr{M}} Q_i(u_i|x_i) p_{X_i}(x_i)$$

where we have introduced the vectorized notation $\boldsymbol{Q} = [Q_i(u_i|x_i) \,|u_i \in \mathscr{U}_i, x_i \in \mathscr{X}_i, i \in \mathscr{M}]$, $\boldsymbol{q} = [q_i(u_i) \,|u_i \in \mathscr{U}_i, i \in \mathscr{M}]$, and $\boldsymbol{\Lambda} = \left[\Lambda(\hat{t}|\boldsymbol{u}) \,\middle|\, \hat{t} \in \hat{\mathscr{T}}, \boldsymbol{u} \in \mathscr{U}_1 \times \cdots \times \mathscr{U}_M\right]$. Next, we observe that this Lagrangian can be

minimized with respect to one user's conditional distribution Q_i or marginal distribution q_i at a time in a manner identical to the single user Blahut-Arimoto algorithm. Here we will use the additional notation

$\boldsymbol{Q}_{\setminus m} = [Q_i(u_i|x_i)\,|u_i \in \mathscr{U}_i, x_i \in \mathscr{X}_i, i \in \mathscr{M} \setminus m]$ and

$\boldsymbol{q}_{\setminus m} = [q_i(u_i)\,|u_i \in \mathscr{U}_i, i \in \mathscr{M} \setminus \{m\}]$.

Thm. 2: **(i)** For a fixed $\boldsymbol{Q}, \boldsymbol{q}_{\setminus m}, \boldsymbol{\Lambda}$ the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ is minimized by

$$q_m(u_m) = \sum_{x_m} p_{X_m}(x_m) Q_m(u_m|x_m), \tag{3.13}$$

while **(ii)** for a fixed $\boldsymbol{Q}_{\setminus m}, \boldsymbol{q}, \boldsymbol{\Lambda}$, the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ is minimized by

$$Q_i(u_i|x_i) = \frac{q_i(u_i) 2^{-\frac{\mu}{\lambda_i} d_i(x_i, u_i)}}{\sum_{u_i'} q_i(u_i') 2^{-\frac{\mu}{\lambda_i} d_i(x_i, u_i')}}. \tag{3.14}$$

where

$$d_i(x_i, u_i) = \sum_{(t, \hat{t}, \boldsymbol{x}_{\setminus i}, \boldsymbol{u}_{\setminus i})} d(t, \hat{t}) p_{T, \boldsymbol{X}}(t, \boldsymbol{x}) \Lambda(\hat{t}|\boldsymbol{u}) Q_{\setminus i}(\boldsymbol{u}_{\setminus i}|\boldsymbol{x}_{\setminus i}), \tag{3.15}$$

with

$$Q_{\setminus i}(\boldsymbol{u}_{\setminus i}|\boldsymbol{x}_{\setminus i}) = \prod_{m \in \mathscr{M} \setminus i} Q_i(u_i|x_i), \tag{3.16}$$

and **(iii)** for a fixed $\boldsymbol{Q}, \boldsymbol{q}$, the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ is minimized by the Bayes detector

$$\Lambda(\hat{t}|\boldsymbol{u}) = \begin{cases} \alpha(\boldsymbol{u}) & \hat{t} \in \arg\min_{\hat{t} \in \hat{\mathscr{T}}} \mathbb{E}[d(T, \hat{t})|\boldsymbol{U} = \boldsymbol{u}] \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

where $\alpha(\boldsymbol{u})$ is selected to ensure

$$\sum_{\hat{t}} \Lambda(\hat{t}|\boldsymbol{u}) = 1 \tag{3.18}$$

and $\mathbb{E}[d(T,\hat{t})|\boldsymbol{U}=\boldsymbol{u}]$ denotes

$$\sum_{(t,\hat{t},\boldsymbol{x}_{\setminus i},\boldsymbol{u}_{\setminus i})} d(t,\hat{t})p_{T,\boldsymbol{X}}(t,\boldsymbol{x})\prod_{i\in\mathscr{M}} Q_i(u_i|x_i). \tag{3.19}$$

Proof: (**i**) and (**ii**) follow by recognizing that with $\boldsymbol{Q}_{\setminus},\boldsymbol{q}_{\setminus}$ fixed, the part of the Lagrangian associated with Q_i and q_i is identical to that used in the classical Blahut-Arimoto algorithm with slope $s=-\frac{\mu}{\lambda_i}$. (**iii**) Follows from the fact that the only term in the Lagrangian which depends on Λ is the mean distortion, which is minimized by a Bayes detector. ■

From these facts, we can attempt to minimize the Lagrangian through alternating minimization. Namely, updating q_m by (3.13), Q_m by (3.14), or Λ by (3.17) will always yield a non-increase of the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q},\boldsymbol{q},\Lambda)$ since they achieve a minimum with the remaining parameters fixed. As such, any infinite sequence of such updates will yield a monotone non-increasing sequence of Lagrangian values, which, since the Lagrangian is bounded below by its global minimum, must converge, albeit not necessarily to a global minimum. One especially attractive sequence of such updates is as follows.

Independent CEO Blahut-Arimoto Algorithm:

For a given $i\in\mathscr{M}$, alternately update q_i and Q_i according to (3.13) and (3.14) until convergence is reached, iterating the following over iteration index k_i

$$q_i^{(\ell,k_i)}(u_i)=\sum_{x_i\in\mathscr{X}_i} Q_i^{(\ell,k_i)}(u_i|x_i)p_{X_i}(x_i). \tag{3.20}$$

$$Q_i^{(\ell,k_i+1)}(u_i|x_i)=\frac{q_i^{(\ell,k_i)}(u_i)2^{-\frac{\mu}{\lambda_i}d_i^{(\ell)}(x_i,u_i)}}{\sum_{u_i'\in\mathscr{U}_i} q_i^{(\ell,k_i)}(u_i')2^{-\frac{\mu}{\lambda_i}d_i^{(\ell)}(x_i,u_i')}} \tag{3.21}$$

until the Lagrangian no longer changes over iterations, and the associated Q_i is

$$Q_i^{(\ell,*)}(u_i|x_i) = \lim_{k_i\to\infty} Q_i^{(\ell,k_i)}(u_i|x_i) =: Q_i^{(\ell+1,1)} \tag{3.22}$$

where the distortion $d_i^{(\ell)}(x_i, u_i) :=$

$$\sum_{(t,\hat{t},\boldsymbol{x}_{\setminus i},\boldsymbol{u}_{\setminus i})} d(t,\hat{t}) p_{T,\boldsymbol{X}}(t,\boldsymbol{x}) \Lambda^{(\ell,i)}(\hat{t}|\boldsymbol{u}) Q_{\setminus i}^{(\ell,i)}(\boldsymbol{u}_{\setminus i}|\boldsymbol{x}_{\setminus i}), \tag{3.23}$$

with

$$Q_{\setminus i}^{(\ell,i)}(\boldsymbol{u}_{\setminus i}|\boldsymbol{x}_{\setminus i}) = \prod_{j<i} Q_j^{(\ell,*)}(u_j|x_j) \prod_{j'>i} Q_{j'}^{(\ell-1,*)}(u_{j'}|x_{j'}).$$

Next, update the Bayes detector according to (3.17), i.e.

$$\Lambda^{(\ell',i')}(\hat{t}|\boldsymbol{u}) := \begin{cases} \alpha^{(\ell',i')}(\boldsymbol{u}) & \hat{t} \in \arg\min_{\hat{t}\in\hat{\mathscr{T}}} J^{(\ell,i)}(\hat{t},\boldsymbol{u}) \\ 0 & \text{otherwise} \end{cases}$$

with $i' = i+1$ and $\ell' = \ell$ if $i < M$ and $i' = 1$ and $\ell' = \ell+1$ if $i = M$, and where $\alpha^{(\ell',i')}(\boldsymbol{u})$ is selected to ensure that

$$\sum_{\hat{t}\in\hat{\mathscr{T}}} \Lambda^{(\ell',i')}(\hat{t}|\boldsymbol{u}) = 1 \quad \forall \boldsymbol{u} \in \mathscr{U}, \quad \text{and}$$

$$J^{(\ell,i)}(\hat{t},\boldsymbol{u}) = \sum_{t,\boldsymbol{x}} d(t,\hat{t}) p_{\boldsymbol{X},T}(\boldsymbol{x},t) Q_{\setminus i}^{(\ell,i)}(\boldsymbol{u}_{\setminus i}|\boldsymbol{x}_{\setminus i}) Q_i^{(\ell,*)}(u_i|x_i)$$

Next, move on to the next user ($i \leftarrow i+1$ if $i < M$ and $i \leftarrow 1, \ell \leftarrow \ell+1$ if $i = M$), and repeat the process.

The reason why this particular alternating minimization sequence is attractive is that (3.20) (3.21) (3.22) are the ordinary Blahut-Arimoto algorithm for the exotic distortion (3.23), and this is guaranteed to give a convergent $Q_i^{(\ell,*)}$ and $q_i^{(\ell,*)}$ achieving

the minimum of the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ with $\boldsymbol{Q}_{\backslash i}, \boldsymbol{q}_{\backslash i}, \boldsymbol{\Lambda}$ fixed. As such, the subsequence of iterates $Q_i^{(\ell,*)}$,$q_i^{(\ell,*)}$, and $\Lambda^{\ell,i}$ themselves consist of an cyclic minimization in both of the arguments Q_i, q_i then $\Lambda^{\ell,i}$ for each user i.

3.3 Initialization and Convergence Analysis

Recall that since it is an alternating minimization, the algorithm we presented in Section 3.2 always yields a monotone non-increasing sequence of Lagrangians which is bounded below and hence converges. The only remaining interesting issue is whether or not the convergent Lagrangian value is indeed the global minimum.

We begin by observing that the algorithm we have presented is in some instances globally convergent.

Thm. 3: If the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ is a convex function of $(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ the independent CEO Blahut-Arimoto algorithm will converge to the global minimum of the Lagrangian from any initialization.

Proof: Note that the global minimum of the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ with respect to all of its variables $\boldsymbol{Q}, \boldsymbol{q}, \Lambda$ is obviously a fixed point of this algorithm, or any other alternating minimization sequence. Also, as the algorithm forms an alternating minimization, yielding Lagrangian values that are monotone non-increasing and bounded below, it must yield a sequence of Lagrangians which converges. Furthermore, any fixed point of this algorithm will zero the gradient of the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ with respect to all the variables $\boldsymbol{Q}, \boldsymbol{q}$. Hence, when the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ is a convex function of all of its variables $\boldsymbol{Q}, \boldsymbol{q}, \Lambda$, the algorithm above will converge to the global minimum of the Lagrangian from any initialization. ■

As the rate part of the Lagrangian is convex, and the sum of convex functions is

also convex, the Lagrangian will be convex if the expected distortion

$$\sum_{\boldsymbol{u},\boldsymbol{x},t,\hat{t}} d(t,\hat{t}) p_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \Lambda(\hat{t}|\boldsymbol{u}) \prod_{i\in\mathscr{M}} Q_i(u_i|x_i) p_{X_i}(x_i) \tag{3.24}$$

is a convex function of $\boldsymbol{Q}, \boldsymbol{\Lambda}$. As we explore in Section 3.5, some interesting problems exhibit this convex and global convergence behavior.

However, as we shall also see in Section 3.5, there are interesting problems for which neither the expected distortion (3.24) nor the Lagrangian $\mathscr{L}_{\boldsymbol{\lambda}\mu}(\boldsymbol{Q}, \boldsymbol{q}, \Lambda)$ are convex. This is a major difference between the classical Blahut-Arimoto algorithm and the one necessary for the multi-user CEO case, and it is due to the product between variables in the optimization that occurs in the expected distortion (3.24). Indeed, in the classical Blahut-Arimoto case, there is no such product, and the expected distortion is a linear function and hence convex.

In the non-convex case, where the algorithm converges is heavily dependent on its initialization, and hence a good initialization is necessary. Fortunately, there are key points in the rate distortion region for which closed form solutions to the non-convex optimization are available from the theory of distributed function computation.

In particular, the points associated with the minimum possible distortion

$$D_{min} = \sum_{t\in\mathscr{T},\boldsymbol{x}\in\mathscr{X}} p_{\boldsymbol{X}}(\boldsymbol{x}) \min_{\hat{t}\in\hat{\mathscr{T}}} \mathbb{E}[d(T,\hat{t})|\boldsymbol{X}=\boldsymbol{x}] \tag{3.25}$$

can have a distributed function computation interpretation. Indeed, if for every $\boldsymbol{x} \in \mathscr{X}$ there exists a *unique* $\hat{t}(\boldsymbol{x})$ such that

$$\mathbb{E}[d(T,\hat{t}(\boldsymbol{x}))|\boldsymbol{X}=\boldsymbol{x}] = \min_{\hat{t}\in\hat{\mathscr{T}}} \mathbb{E}[d(T,\hat{t})|\boldsymbol{X}=\boldsymbol{x}] \tag{3.26}$$

Then the problem of finding the minimum sum rate $\sum_{i\in\mathscr{M}} R_i$ yielding expected dis-

tortion D_{min} is equivalent to finding the minimum rate necessary to compute function $\hat{t}(\boldsymbol{X})$ at the CEO in the Shannon lossless manner. This minimum rate lossless distributed function computation problem is solved in the independent sources case via the framework of graph entropy [202, 203]. In this regard, for each $i \in \mathscr{M}$ define the *characteristic graph* $\mathscr{G}_i$ to be a graph with vertices $\mathscr{X}_i$, and for which x_i and x_i' are adjacent if and only if there exists $\boldsymbol{x}_{\setminus i} \in \boldsymbol{X}_{\setminus i}$ such that $\hat{t}(x_i, \boldsymbol{x}_{\setminus i}) \neq \hat{t}(x_i, \boldsymbol{x}_{\setminus i})$.

Thm. 4 (Lossless Distributed Function Computation [202, 203])**:** If there exists a unique function $\hat{t}(\boldsymbol{x})$ satisfying (3.26), then the minimum sum rate $R_{\sum}$ achievable for the minimum distortion D_{min} from (3.25) is achieved by the $\boldsymbol{Q}$ achieving a rate equal to the solution of the convex optimization

$$\sum_{i=1}^{M} \min_{Q_i(\cdot|\cdot) \in \Gamma(\mathscr{G}_i)} I(X_i; U_i) \tag{3.27}$$

where the constraint $Q_i(\cdot|\cdot) \in \Gamma(\mathscr{G}_i)$ requires that $Q_i(\cdot|x_i)$ has support on those u_i ranging over the maximal independent sets of $\mathscr{G}_i$ containing x_i.

Proof: For any distributed function computation problem with independent sources, the independency together with the assumption WLOG that each sources' distribution is positive on it's support guarantees: (1) the Zig-zag condition from [203], and (2) the unique existence of the achievable minimum sum-rate: (see Corollary 17 and the discussion following Theorem 18 in [203])

$$\min \sum_{i=1}^{M} R_i = \sum_{i=1}^{M} H_{G_i}(X_i) \tag{3.28}$$

where $H_{G_i}(X_i) = \min_{p(u_i|x_i) \in \Gamma(\mathscr{G}_i)} I(U_i; X_i)$ represents the graph entropy of the characteristic graph of source X_i. ■

Since the minimum sum rate with D_{min} is associated with the Lagrange multipliers

$\lambda_i = 0$ for all $i \in \mathscr{M}$ and $\mu \neq 0$, this initialization is especially useful for large values of $\frac{\mu}{\lambda_i}$ $\forall i \in \mathscr{M}$, as it can still be in the region of convergence for the global optimum, provided the globally optimal $\boldsymbol{Q}$ moves continuously in the Lagrange multipliers $\boldsymbol{\lambda}, \mu$. The convergent $\boldsymbol{Q}$s for these Lagrange multipliers can next be utilized as multiple candidate initializations, possibly along with the optimal $\boldsymbol{Q}$ from (3.27), for smaller values of $\frac{\mu}{\lambda_i}$ and so on. This yields an initialization strategy which can, at least in principle for a dense enough grid of $\frac{\mu}{\lambda_i}$, give the rate distortion curve for those problems for which there are globally optimal $\boldsymbol{Q}$s that are continuous in the Lagrange multipliers $\boldsymbol{\lambda}, \mu$ that have a region of attraction under the algorithm that contains an open ball in the set of possible initializations.

Note that for those problems for which the solution $\hat{t}(\boldsymbol{x})$ obeying (3.26) is not unique, we can combinatorially search over all possible candidate functions $\hat{t}(\boldsymbol{x})$ to find the one with the minimum (3.27). In some cases, such as the arg max case we present in Section 3.5, this minimum can be proven to again achieve the minimum sum rate with distortion D_{min} [204].

Finally, we observe that even when the problem is non-convex, and does neither exhibit the continuity in $\boldsymbol{\lambda}, \mu$, nor obey the uniqueness in (3.26), nor the tightness of the minimum over all possible solutions to (3.26), or when the grid of Lagrange multipliers has not been selected to be fine enough, these initializations, together with a large collection of random initializations, followed by the selection of the minimum among all of the converged Lagrangians (over different initializations) for a given $\boldsymbol{\lambda}, \mu$ can still yield very good information theoretically optimized inner bounds to the rate distortion region to which practical quantization schemes can be compared.

While the graph entropy based initialization has the merits described above, sometimes it is also of interest to develop inner bounds based on random initializations as well, for instance to investigate the probability, under a random initialization strategy,

of encountering the global minimum. With regard to Monte Carlo random initialization, we note that it is helpful to observe that the optimization is invariant to any permutations of the sets $\mathscr{U}_i$, and thus it suffices to sample uniformly from a polytope of initializations removing this symmetry, for instance by selecting only those $Q_i(u_i|x_i)$ such that the marginal $q_i(u_i)$s are sorted in decreasing order. This can be done using any method of sampling uniformly from a polytope such as the hit and run method, which, starting at an initial point in the polytope, selects a direction on the unit sphere to move in at random, finds the maximum and minimum distance it can move in this direction, then selects an amount to move in this direction uniformly distributed between this minimum and maximum.

3.4 Selecting the Lagrange Multipliers and Calculating the Convex Hull

The algorithm in Section 3.2 provides a method for minimizing the Lagrangian for a given value of the multipliers $\boldsymbol{\lambda}, \mu$, but in practice we will either want to calculate the whole rate distortion region or at least the whole sum rate distortion function. To do this, one selects a grid of multipliers $\{\boldsymbol{\lambda}_k, \mu_k\}$, which can WLOG satisfy $\mu_k + \sum_{i \in \mathscr{M}} \lambda_{k,i} = 1,\ \forall k$, then runs the algorithm from Section 3.2 with the initialization strategy from Section 3.3 for each $\boldsymbol{\lambda}_k, \mu_k$ recording the rate vector

$$r_i^{(k)} = \sum_{u_i, x_i} p_{X_i}(x_i) Q_i^{k,*}(u_i|x_i) \log_2 \left(\frac{Q_i^{k,*}(u_i|x_i)}{q_i^{\boldsymbol{\lambda}_k, \mu_k, *}(u_i)} \right)$$

and the distortion

$$d^{(k)} = \sum_{t, \hat{t}, \boldsymbol{x}, \boldsymbol{u}} d(t, \hat{t}) p_{T, \boldsymbol{X}}(t, \boldsymbol{x}) \Lambda^{\boldsymbol{\lambda}_k, \mu_k, *}(\hat{t}|\boldsymbol{u}) \prod_i Q_i^{k,*}(u_i|x_i)$$

associated with the convergent $Q_i^{k,*}, q_i^{k,*}, \Lambda^{k,*}$ attaining the best minimum found (over possibly different initializations) for this $\boldsymbol{\lambda}_k, \mu_k$.

Next, in order to keep only those points on the Pareto frontier of the rate distortion region, one can form the homogenization and polar, which creates a cone in one higher dimension whose inequality description is $\{\mathbf{r} \,|\mathbf{H}\mathbf{r} \geq \mathbf{0}\}$ with

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}_{M+1\times 1} & \mathbb{I}_{M+1} \\ \mathbf{1}_{K\times 1} & \mathbf{V} \end{bmatrix}, \ \mathbf{V} = \begin{bmatrix} r_1^{(1)} & \cdots & r_M^{(1)} & d^{(1)} \\ r_1^{(2)} & \cdots & r_M^{(2)} & d^{(2)} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Here, the first $M+1$ rows of $\mathbf{H}$ are acting as extreme rays along the rate and distortion axes so that we can get the polyhedron associated with all points element-wise greater than a point on the Pareto frontier. Taking the convex hull of the rows of this matrix is equivalent to finding those rows that are redundant when the matrix is viewed as a series of linear inequalities, e.g. sequentially passing through the rows, the row $\boldsymbol{h}_k$ can be removed if [205]

$$\min_{\mathbf{r}\left|\mathbf{H}_{\setminus k}\mathbf{r}\geq 0\right.} \boldsymbol{h}_k\mathbf{r} \geq 0 \tag{3.29}$$

where $\mathbf{H}_{\setminus k}$ are all the rows currently remaining in $\mathbf{H}_k$ except for k. Numerically, this convex hull can be taken with any number of algorithms such as Quickhull [206]. Next, it is often desirable to not just have a collection of points on the Pareto frontier of the rate distortion region, but also the inequality description of a polyhedral inner bound to it. For this, one performs a representation conversion between the extremal representation of the polyhedron and its inequality representation using, for instance, the double descriptions method [205]. This inequality description can allow one to evaluate the distortion bound on an evenly spaced grid of rate points, which is useful for obtaining plots such as those in Section 3.5.

3.5 Example Rate Distortion Regions

Here we present some illustrative example rate distortion regions calculated by our software implementation of the algorithm described in this chapter in C. This software, which is capable of making use of parallelism across multiple cores and multiple computers is available with documentation at [207].

Let's begin with an example in which $M = 2$ users observe independent fair bits, and the CEO wishes to compute the binary sum of these two bits under the Hamming distortion metric. A contour plot of the rate distortion region for this case is provided in Fig. 3.2, and a surface plot with the convergent rate distortion vectors associated with the chosen grid of Lagrange multipliers is shown in Fig. 3.3. The part of the contour plot corresponding with one of the two users sending its bit without compression corresponds correctly with the series of values $(R_1, R_2, D) = (1, 1 - H_2(D), D)$ that it should, as is evidenced by its correspondence with the x markers on the axes of the plot which were calculated with this function. For this problem, all of the 100000 random initializations uniformly sampled from the symmetry removed set of initializations converge to the same optimum, suggesting that the Lagrangian is convex.

Next we consider an example exhibiting local convergence behavior. In this example, two users observe numbers uniformly distributed over $\{1, 2, 3\}$ and CEO wishes to known the user index $T \in \{1, 2\}$ of a user whose observation is equal to $\max\{X_1, X_2\}$. The distortion is the difference between this maximum and the value of the user selected

$$\mathbb{E}[d(T, \hat{t})|X_1, X_2] = \max\{X_1, X_2\} - X_{\hat{t}} \tag{3.30}$$

For this experiment, as shown in Fig. 3.6, for each collection of $\boldsymbol{\lambda}, \mu$ with a small enough $\frac{\mu}{\lambda_i} \forall i$, nearly all of the 1,000 random initializations converge to the same min-

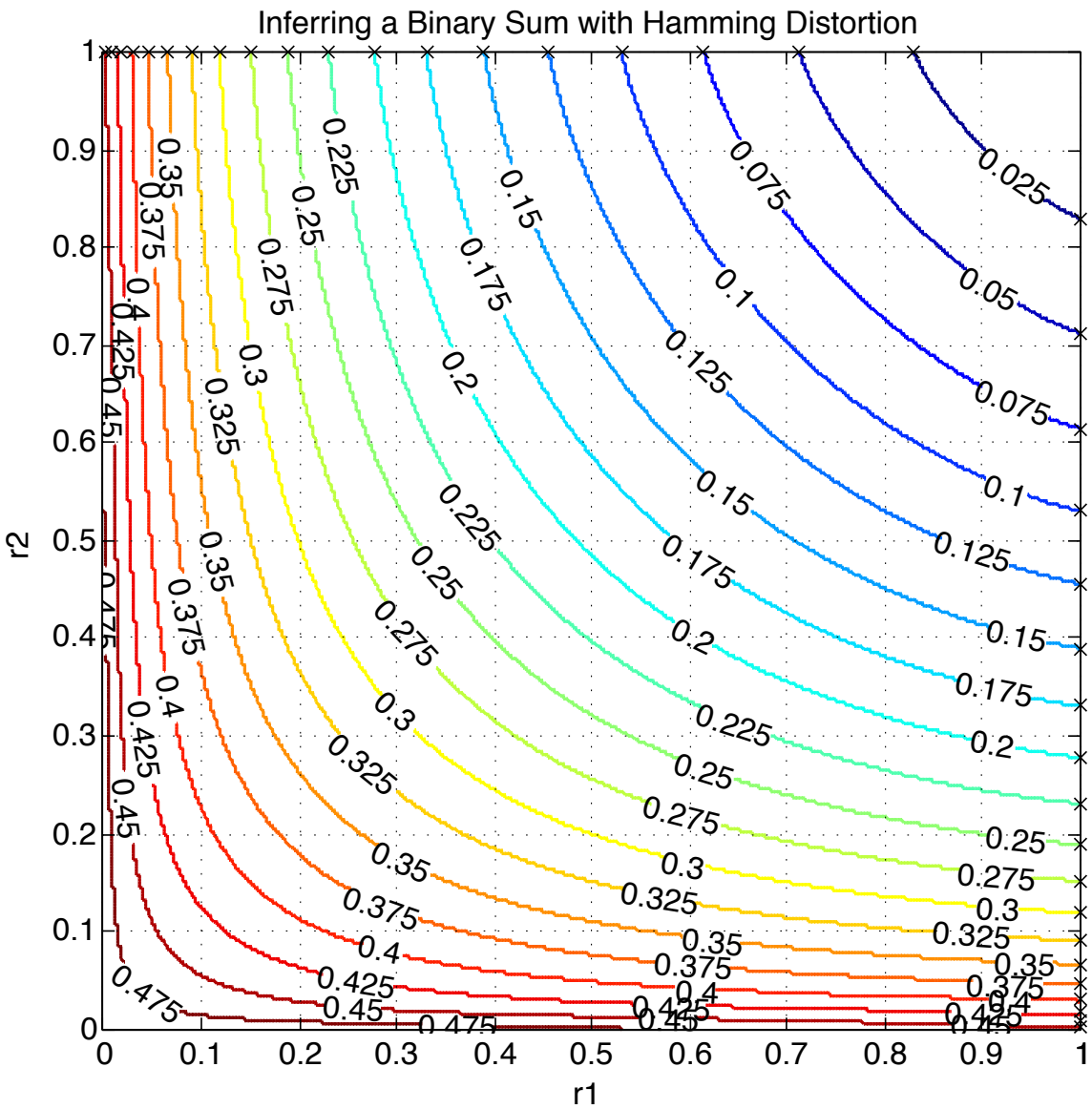


Figure 3.2: Contour plot of the rate distortion region where two users observe independent bits and the CEO wishes to know the sum of these two bits with a bound on the Hamming distortion given by the labels on the contours.

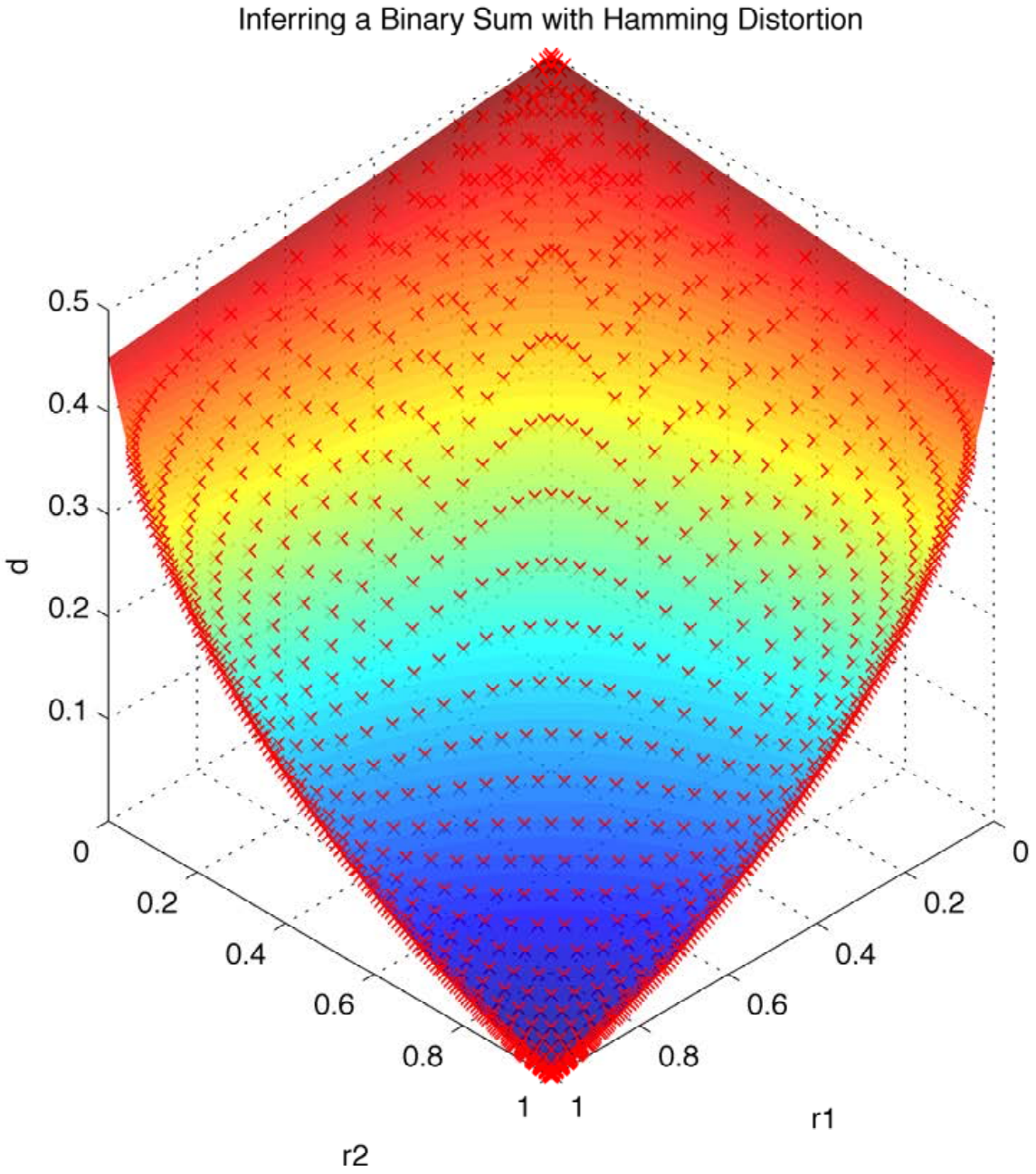


Figure 3.3: Surface plot for the problem from Fig. 3.2 with the convergent values of the Lagrangian from the various Lagrange multipliers superposed as red xs.

imum answer, while for large $\frac{\mu}{\lambda_i} \forall i$ multiple minima are encountered by the random initializations, and a far smaller fraction converge to the same minimum that the graph entropy based initialization converges to. This effect was observed to become more exaggerated as the cardinality of the random variables for whom to compute the arg max of grew, as shown in Fig. 3.7. A contour plot for the resulting achievable rate distortion region after keeping only the best minima and taking the convex hull is shown in Fig. 3.4. Additionally a surface plot for the resulting achievable rate region with the convergent rate distortion points superposed is shown in Fig. 3.5. For this curve, there is a sharp transition between regions that the rate distortion points converge to as a function of the Lagrange multipliers, owing to an affine part of the rate distortion region due to the convex hull nature of the Pareto frontier of the rate distortion region. This example demonstrates the importance of including the graph entropy based initialization, as well as the local convergence behavior of the algorithm when the Lagrangian is not convex.

3.6 Conclusions

This chapter introduced a generalization of the Blahut-Arimoto algorithm capable of calculating the rate distortion region for the CEO problem with independent sources. When the distortion is convex, the algorithm is globally convergent to the minimum of a Lagrangian, while when it is non-convex, a solution for a global minimum at the minimum distortion point from the theory of decentralized function computation is provided as a useful initialization. A series of examples showed that the utility of the rate region computation method as well as the applicability of the CEO with independent sources model to practical problems, especially those involving lossy decentralized function computation.

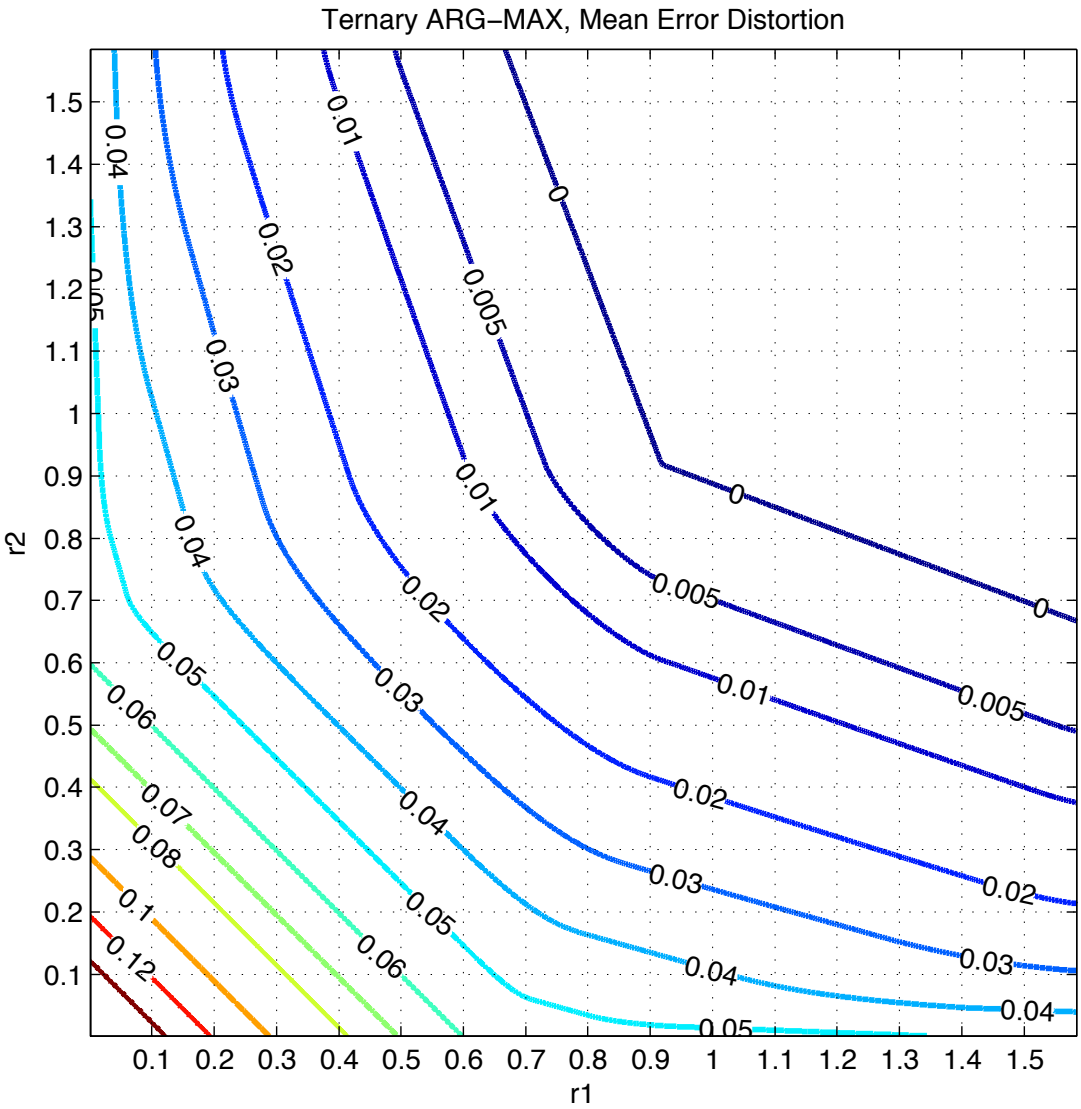


Figure 3.4: Contour plot of the rate distortion region where two users make independent observations uniformly distributed over the set $\{1, 2, 3\}$ and the CEO wishes to determine a user having the largest observation.

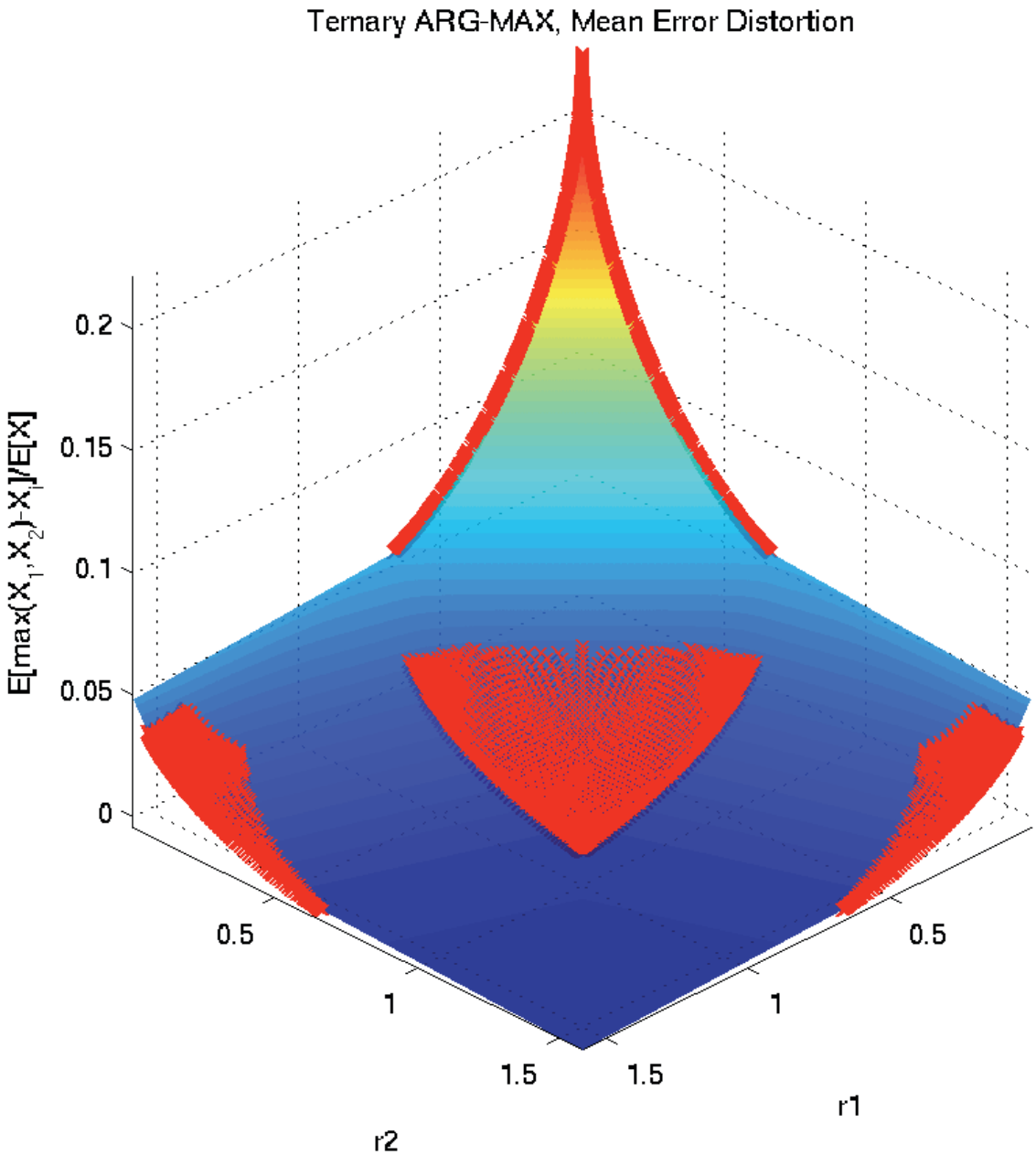


Figure 3.5: Surface plot of the rate distortion region where two users make independent observations uniformly distributed over the set $\{1, 2, 3\}$ and the CEO wishes to determine a user having the largest observation. The red xs mark locations where the algorithm converged to for different Lagrange multipliers. There is a sharp transition between these locations owing to affine regions in the rate distortion region in this problem.

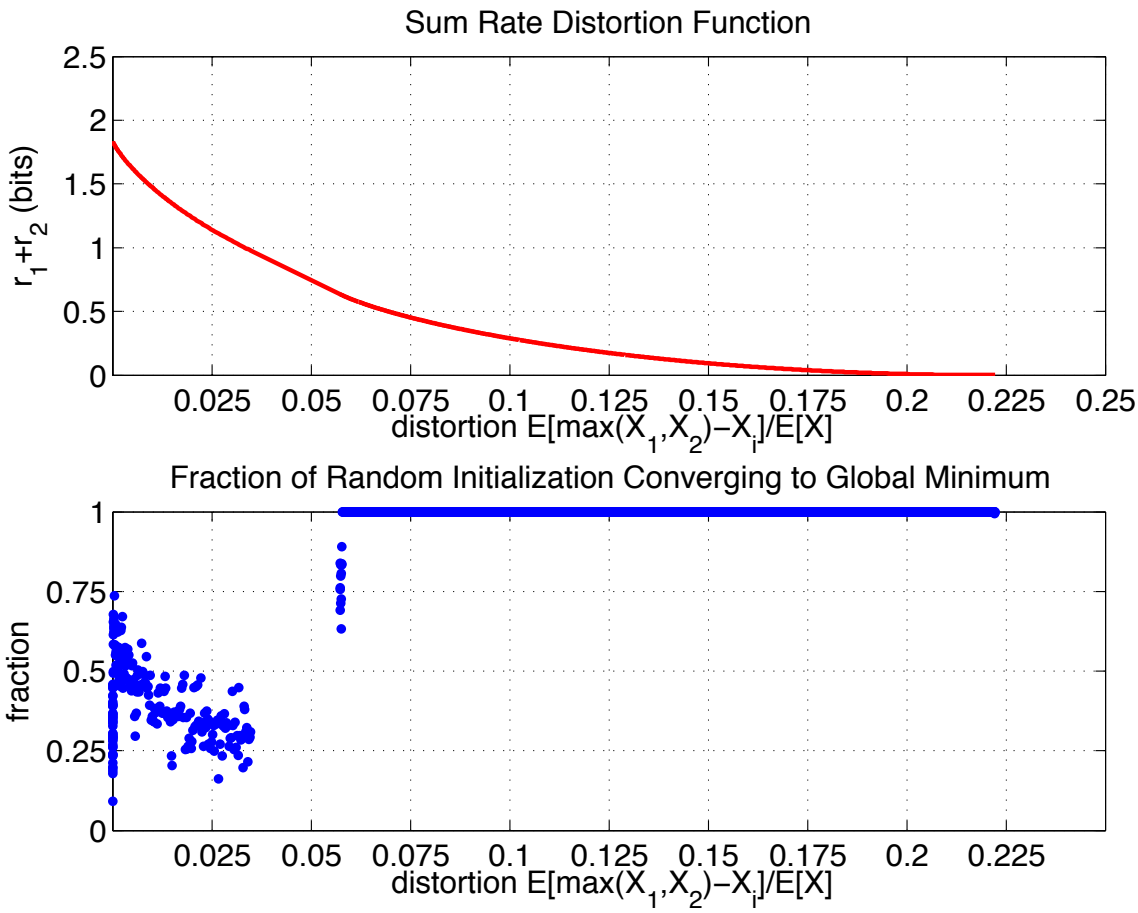


Figure 3.6: Top: Sum rate distortion function for the problem where two users make independent observations uniformly distributed over the set $\{1, 2, 3\}$ and the CEO wishes to determine a user having the largest observation. Bottom: The fraction of random initializations that converge to the global minimum. For low distortions, this fraction becomes smaller, and this effect becomes more exaggerated as the cardinality of the variables X_i grows, making the use of the graph entropy based initialization more important.

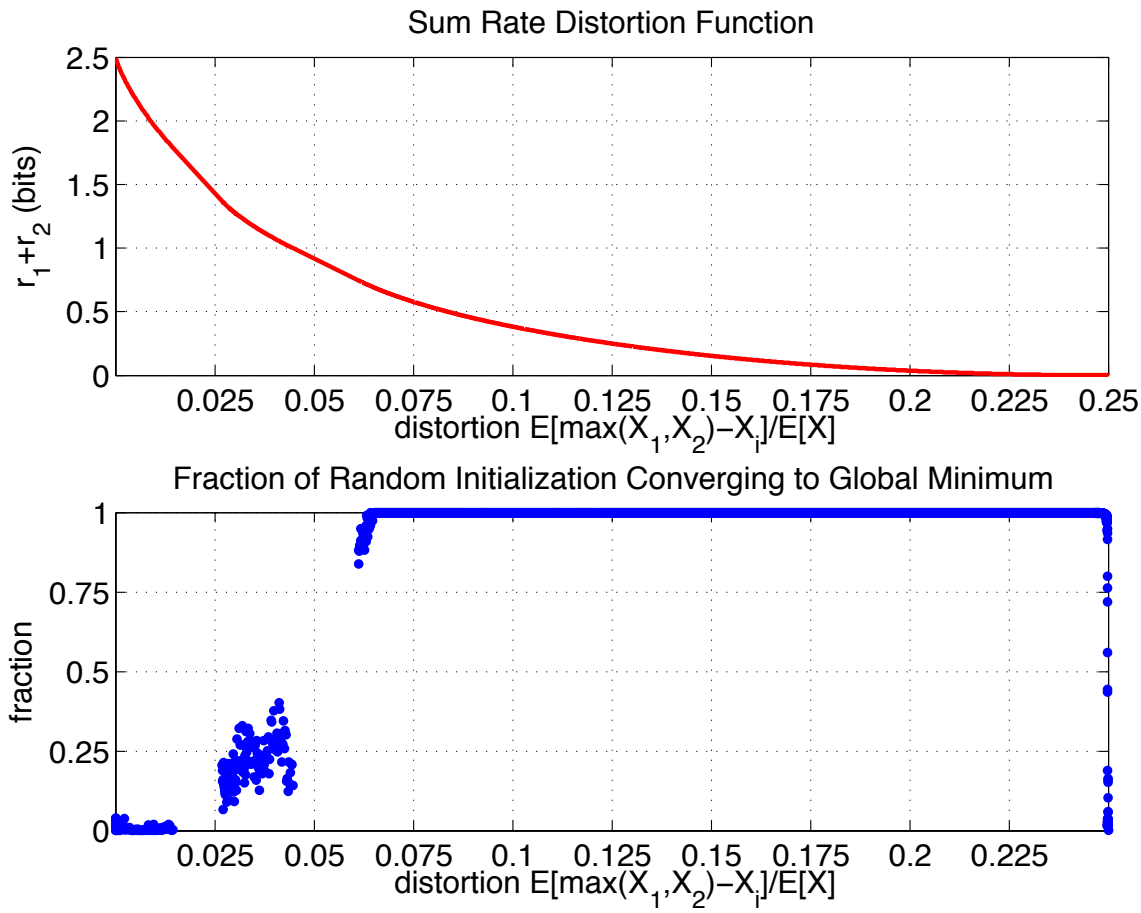


Figure 3.7: Same as Fig. 3.6, except for sources uniformly distributed over $\{1, 2, 3, 4\}$. Observe that a smaller fraction of low distortion random initializations converge.

4. Fundamental Overhead Performance Tradeoffs in Wireless Resource Allocation: CEO Based Examples

Recall that we motivated our study of methods of computing the rate distortion region for the CEO problem with independent sources in Chapter 3 by the fact that it could be potentially be used to derive fundamental limits for the tradeoff between control information overhead and performance for resource allocation and link adaptation. There we argued that local network state could be viewed as the observations at the nodes, and that the CEO could play the role of the network controller, which must estimate the control policy based on the encodings of the local network state send by the network nodes. We suggested that the distortion metric that the CEO could choose its policy to minimize could be the expected difference between the performance of an omniscient controller, having access to all the state information in the network, and the performance of the decisions it selects, given the messages it received.

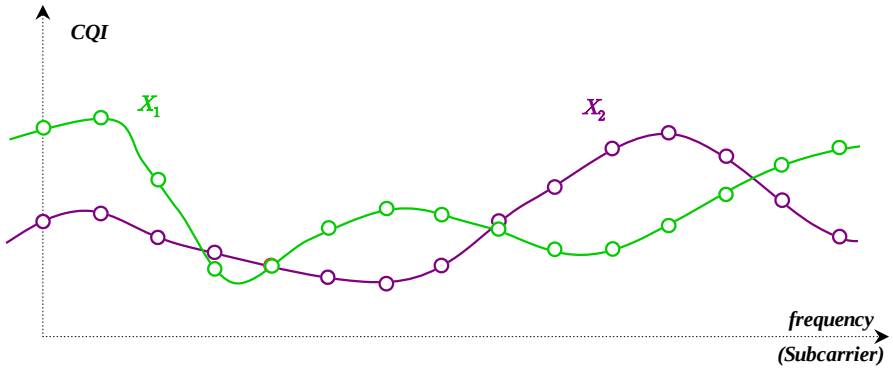


Figure 4.1: Channel Quality

To solidify these rather abstract ideas, while they can be applied to more efficiently encode control information in a wide variety of contexts at different wireless network layers, in this chapter we will focus on problems in physical layer channel based resource allocation and link adaptation in cellular networks for simplicity. The input to each encoder will be the sequence of channel qualities of different physical resource blocks, representing a collection of frequencies, at the user associated with this encoder, as depicted in Fig. 4.1.

The CEO is aiming to make channel scheduling decisions which will enable it to maximize the data spectral efficiency, minimizing the gap between the spectral efficiency it attains, and the one that an omniscient controller, having perfect knowledge of all the channel states in the network, would achieve. We will consider simple models for physical layers employing downlink communications with three different coding capabilities: broadcast communications, rateless coded communications, and adaptive modulation and coding[1].

4.1 Broadcast Communication Based Resource Allocation Model

In the broadcast model, the base station must encode a common broadcast message for every user to successfully decode on each channel, and hence wants to know the lowest (over all users) CQI on each channel. In order to be the most consistent with the following two examples, we will treat this as the problem of learning the maximum (over all users) negative CQI for each channel, i.e. the lossy computation of the function $Z = \max(X_1, ..., X_N)$. As depicted in Fig. 4.2, the distortion in

[1]The material in Chapters 4 and 5 is included as part of a large collaborative journal submission [204] on overhead performance tradeoffs in resource allocation to the *IEEE Transactions on Information Theory*.

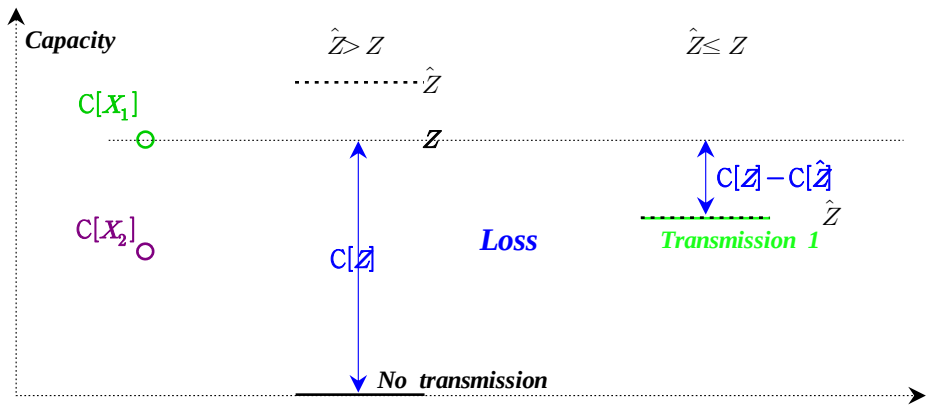


Figure 4.2: Distortion Measure for Function $Z = \max(X_1, X_2)$

broadcast model will be

$$d(Z, \hat{Z}) = \begin{cases} Z - \hat{Z} & \text{if } \hat{Z} \leq Z \\ Z & \text{otherwise} \end{cases} \tag{4.1}$$

This is because if the base station over estimates the maximum negative CQI, its message will not get through, and the gap to omniscient spectral efficiency will be the entire maximum. However, if it underestimates the maximum negative CQI, its message will be successfully decoded, and the gap to optimal spectral efficiency will be the difference between the maximum and its underestimate.

4.2 Rateless Communication Based Resource Allocation Model

In the rateless model, the base station is sending private messages, and employs a near-perfect rateless code [208–210], enabling it to communicate at close to the capacity of a selected channel without explicitly requiring the channel's state at the transmitter. This is achieved by letting the block length of the encoding adapt (with the rate decreasing as the block length increases) based on the channel quality at the receiver, with the receiver indicating when it has successfully decoded. Well designed

HARQ schemes can be viewed as a form of rateless code in this manner. In this situation, the base station needs only to know a user attaining the maximum (over all users) CQI on each channel whose private message to schedule, i.e. an element $Z \in \arg\max(X_1, ..., X_N)$ for each channel, as the subsequent rateless encoder will approach the capacity without having to know what it was at the basestation. As depicted in Fig. 4.3, the distortion in this case will be the difference between the maximum (over users) capacity of each channel and the capacity of the user that the base station selects for that channel.

$$d(Z, \hat{Z}) = X_Z - X_{\hat{Z}} \tag{4.2}$$

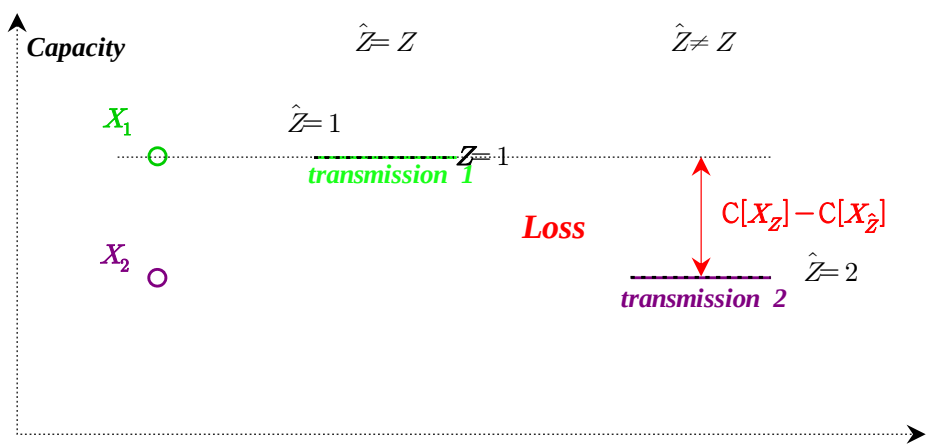


Figure 4.3: Distortion Measure for Function $Z = \arg\max(X_1, X_2)$

4.3 Adaptive Modulation and Coding Communication Based Resource Allocation Model

In the adaptive modulation and coding model, the base station is sending private messages, and wishes to select the modulation and coding rate to match the capacity of the channel selected. Hence, in this situation the base station wishes to know a user attaining the maximum (over all users) CQI, and what this maximum value is (i.e. both the $Z = \max(\cdot)$ and an element $i \in \arg\max(\cdot)$)). In this case, as depicted in Fig. 4.4, the distortion will be the difference between the true max and the estimated max if the estimated max is less than the capacity of the selected user, and will be the entire true max otherwise, since the message will not be successfully decoded.

$$d((Z,i),(\hat{Z},\hat{i})) = \begin{cases} Z - \hat{Z} & \text{if } \hat{Z} \leq X_{\hat{i}} \\ Z & \text{otherwise} \end{cases} \tag{4.3}$$

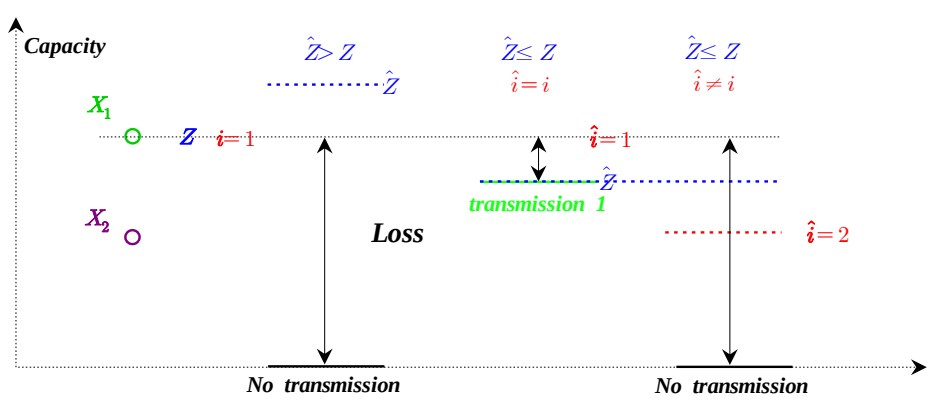


Figure 4.4: Distortion Measure for Function $Z = \max(X_1, X_2)$ and $i = \arg\max(X_1, X_2)$

4.4 Calculated Overhead Performance Tradeoffs

Here, we will utilize the algorithm described in Chapter 3 to calculate the rate distortion for the distortions inspired by the three different resource allocation models from Section 4.1, Section 4.2, and Section 4.3 in which we wish to compute the max, $\arg\max$, and both the $\arg\max$ and max in a lossy manner.

We begin by selecting the channel gains to be uniformly distributed over capacities associated with the 16 CQI levels in LTE standard, as specified in Table 2.20 [53, 54, 60, 61] in Section 2.6.1. The numerical calculation of the rate distortion function in the max, $\arg\max$, and max and $\arg\max$ problems in the simplest case of two users is shown in the Fig. 4.5-4.7.

From these figures, we begin by observing that a reduction in the amount of control information can be achieved by recognizing that the controller only needs to compute a function of the channel qualities rather than knowing them entirely. If we wish to compute these functions over two users losslessly in the sense of achieving the same spectral efficiency as a controller observing the channel qualities perfectly, we are interested in the zero distortion point of these plots, which is associated with the graph entropy as discussed in Section 3.3. For instance, if we are interested in losslessly computing the max as in the broadcast case from 4.1, we can achieve this with 3.875 bits per user instead of 4 bits, a savings of 3.125%. If we wish to determine the $\arg\max$ as well as the max in a way that guarantees the same spectral efficiency as a controller observing both CQIs perfectly, we obtain a smaller reduction of only 1.5625%. Greater savings can be achieved if only the $\arg\max$ is needed, which yields a 23.4% rate saving at a zero distortion point. However, all of these reductions per user can be shown to decrease as the function must be computed over more than two users.

More importantly, we observe from these figures that a substantial reduction in

control information can be achieved with a negligible effect on the data spectral efficiency. For instance, in the max case in Fig. 4.5, the control overhead spent on reporting CQI can be reduced by 25.9% at a loss of only a reduction of 1.16% spectral efficiency performance relative to an omniscient controller. If we only need to determine the arg max as in the rateless case, Fig. 4.6 shows that we could reduce the control information by 48.65% at a loss of only 1.16% in spectral efficiency. Even if we need to know both the arg max and max, we can still reduce the control information overhead spent on CQI reporting by 10% at only a 0.15% reduction in data spectral efficiency.

In each of these instances, the extra time frequency resources freed by sending less control information can be utilized on user data, enhancing the overall spectral efficiency, since an order of magnitude more resources are freed than the reduction in the spectral efficiency. This effect is further amplified when one considers that control information has to be channel coded at a very low rate to ensure that everyone can decode it.

These plots were for only two users, however, and so it is necessary to study the effect of increasing the number of users on the rate distortion curves. In this vein, the plots in Fig. 4.8–4.10 show the rate required per user as a function of the fractional gap to omniscient spectral efficiency as a function of multiple users, with uniformly distributed ternary capacities $X_i \in \{1, 3, 5\}$. At least in this simplified context, the per user savings in overhead rate achieved by allowing a small loss in spectral efficiency remains relatively constant as the number of users are increased.

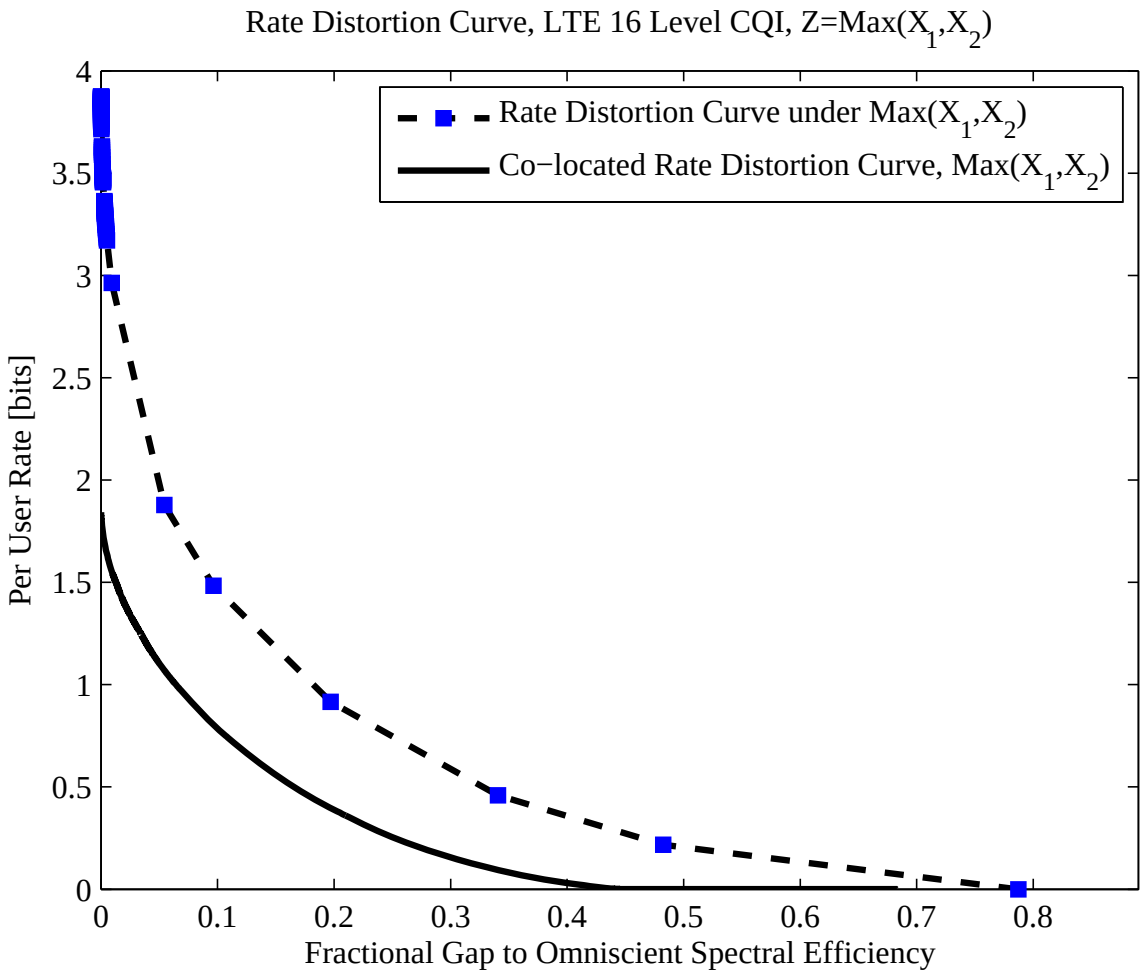


Figure 4.5: Per User Rate Distortion Curve under Max in LTE

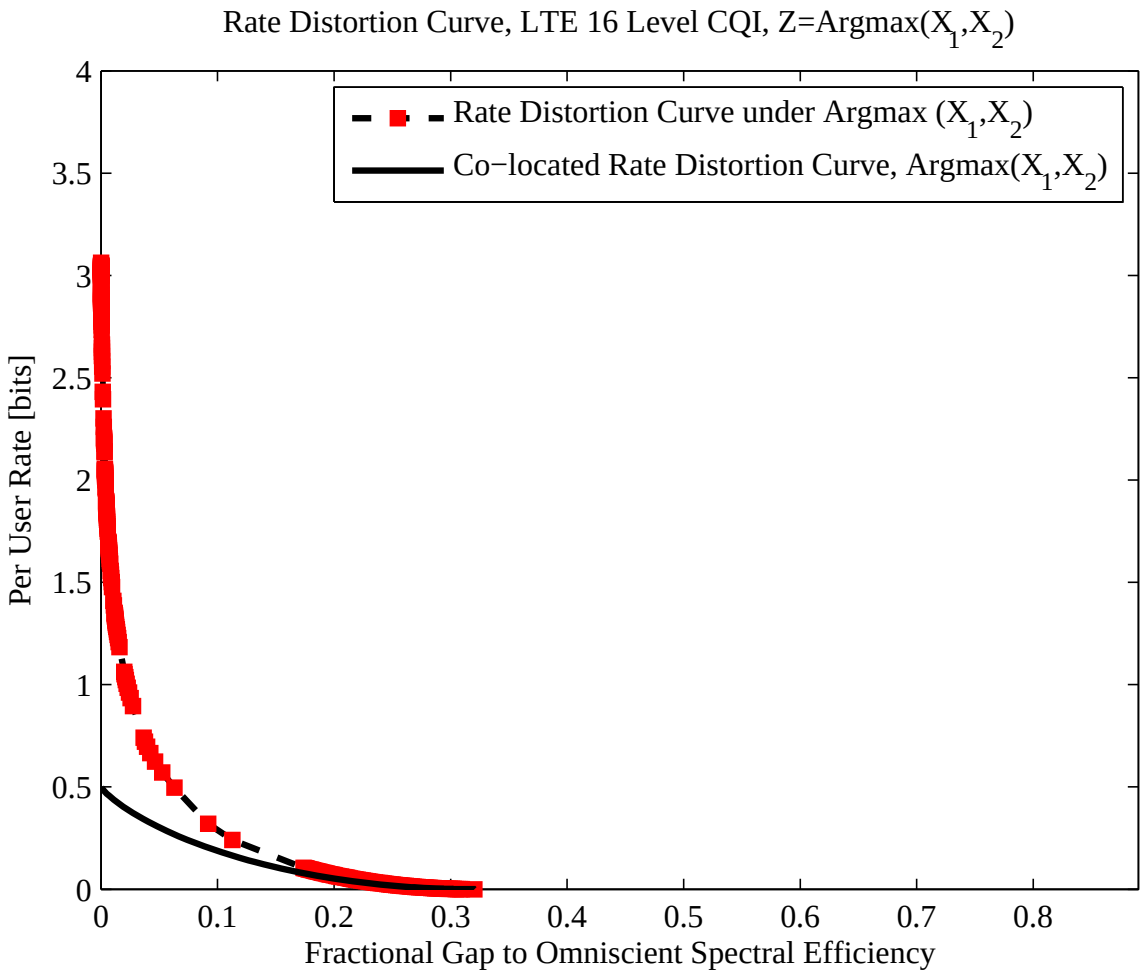


Figure 4.6: Per User Rate Distortion Curve under Argmax in LTE

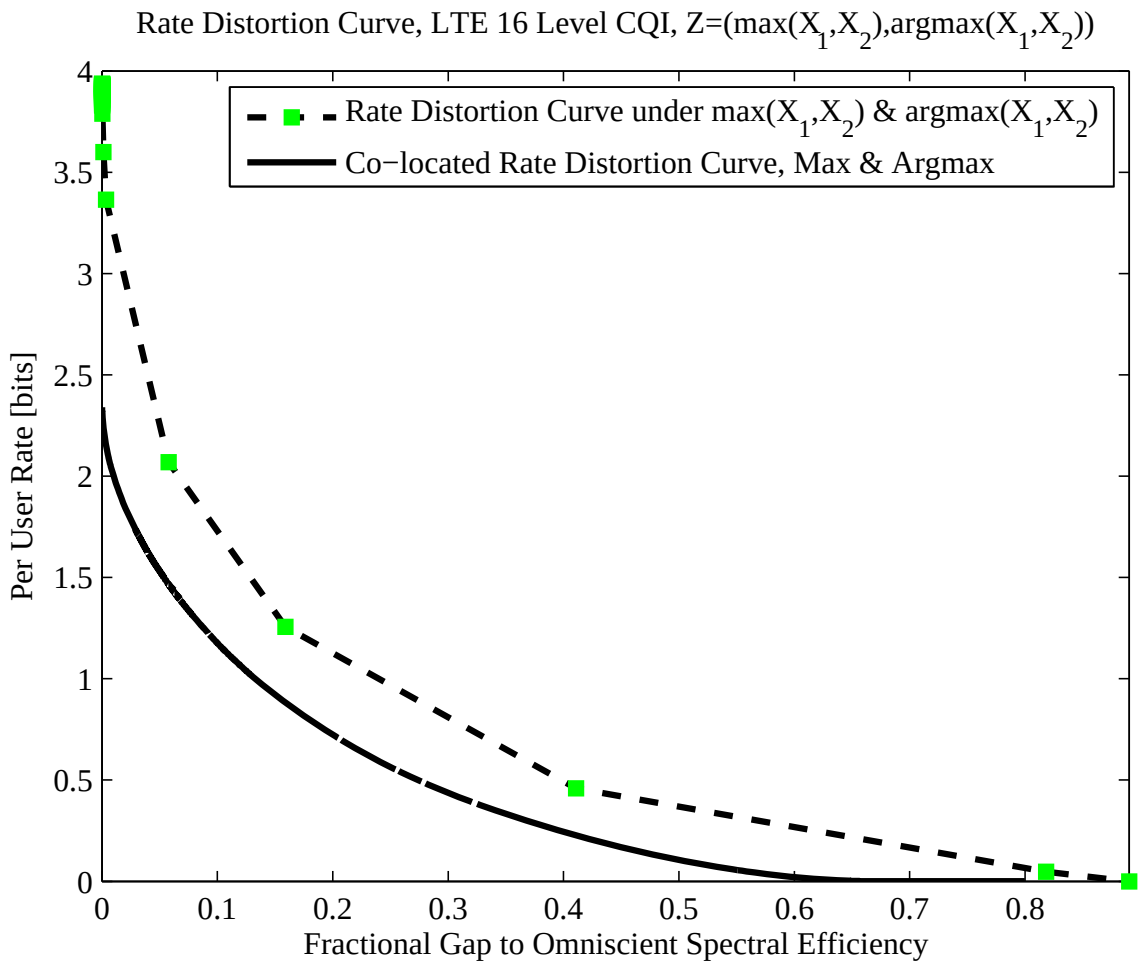


Figure 4.7: Per User Rate Distortion Curve under Max and Argmax in LTE

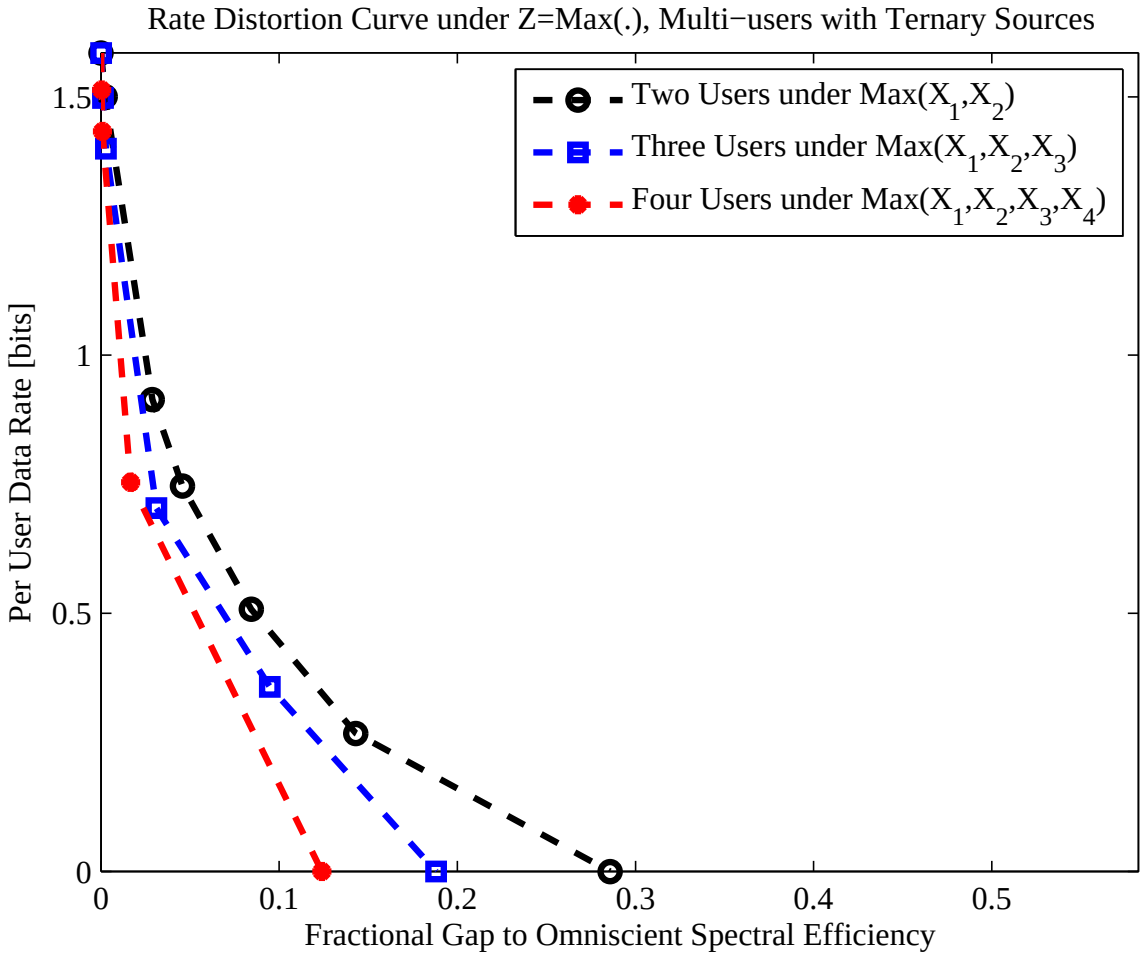


Figure 4.8: Per User Rate Distortion Curve under Max with Ternary Sources

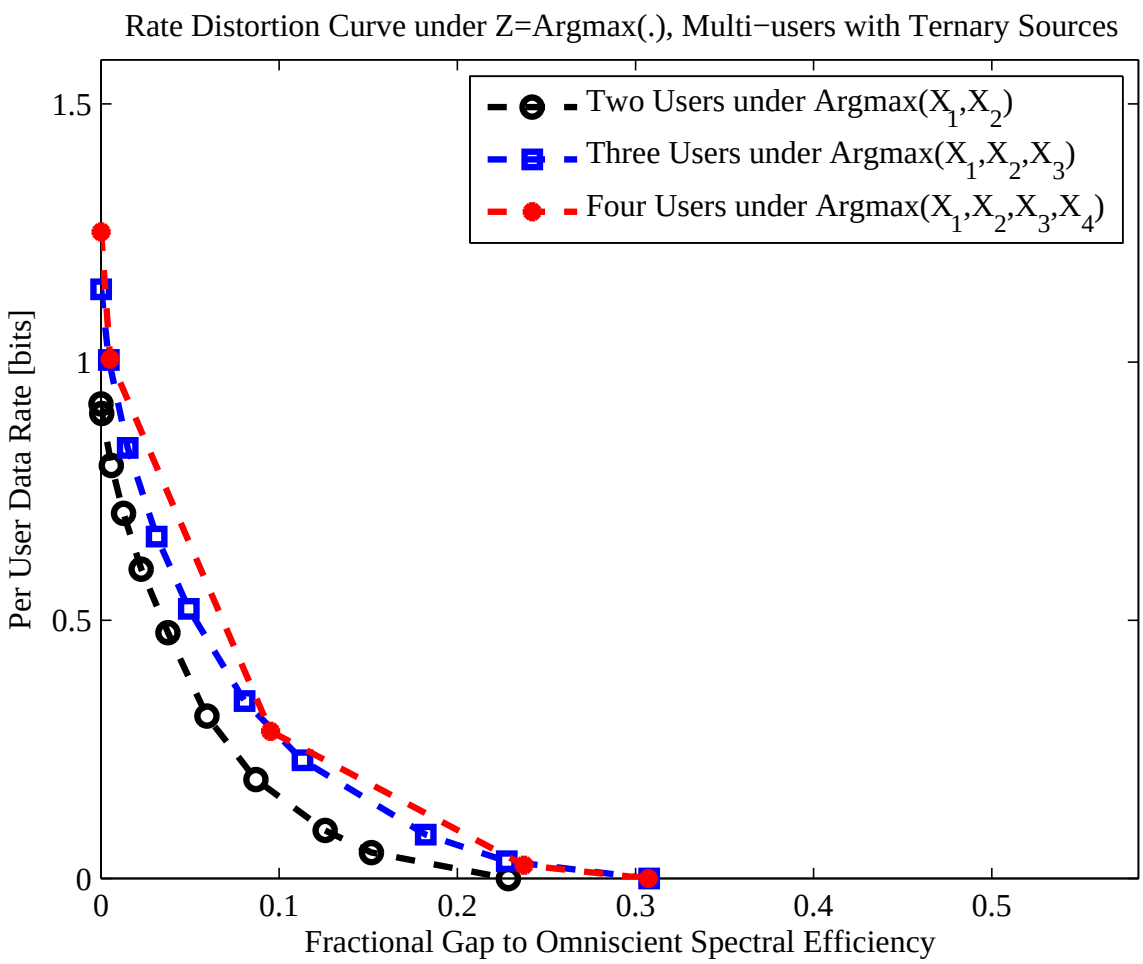


Figure 4.9: Per User Rate Distortion Curve under Argmax with Ternary Sources

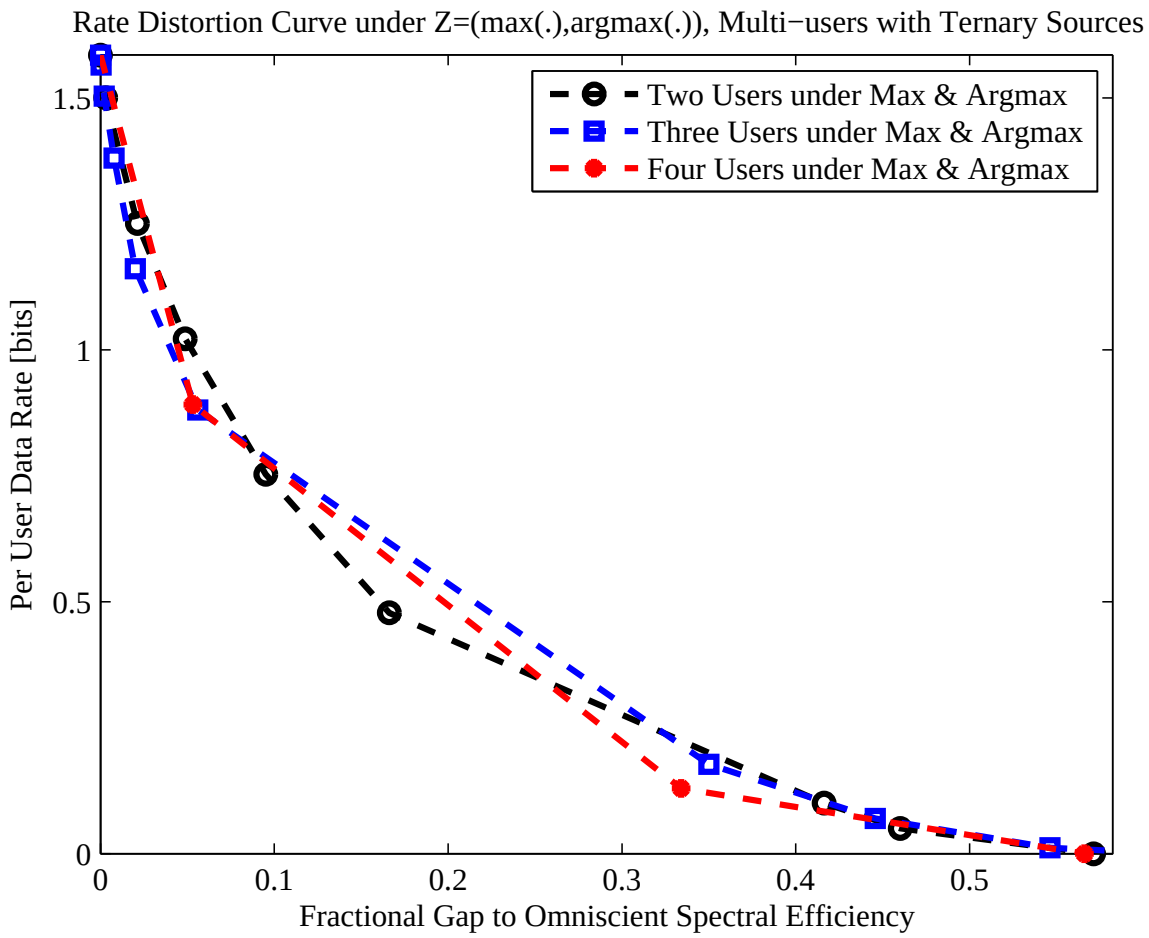


Figure 4.10: Per User Rate Distortion Curve under Max and Argmax with Ternary Sources

5. Practical Distributed Functional Scalar Quantizer Design

In Chapter 4 we demonstrated how one could use the CEO problem with independent sources, along with the algorithm from Chapter 3, to calculate the fundamental tradeoff between control information overhead and the performance for some simple models for cellular wireless resource allocation. While this method is capable of calculating limits over all possible non-interactive control signaling schemes, it does not directly yield a practical design which achieves these limits. Hence, in this chapter, we would like to consider practical control signaling schemes which approach performance overhead tradeoffs calculated with the CEO problem in the manner demonstrated in Chapter 4.

In particular, continuing our focus on the problem of channel quality information related signaling, and the associated distortion metrics presented in Chapter 4 inspired by several different physical layer capabilities, we will design the best encoders for independent channel qualities to minimize these distortion metrics. Here, while we studied channel qualities as discrete valued random variables in Chapter 4, we will find it both more realistic and more analytically convenient to treat them as continuous random variables that must be quantized. The analytical convenience comes from the fact that the resulting optimizations associated with building the control signals involve continuous optimization problems when continuously distributed channel qualities are considered, as opposed to combinatorial optimization problems when discrete random variable observations are considered. These continuous optimizations are more tractable, especially as the problem scales. Additionally, in any real system, the channel qualities will be continuous, and modeling them as discrete reflects that quantization has already occurred, but the problem under consideration effectively involves the optimal design of that quantizer.

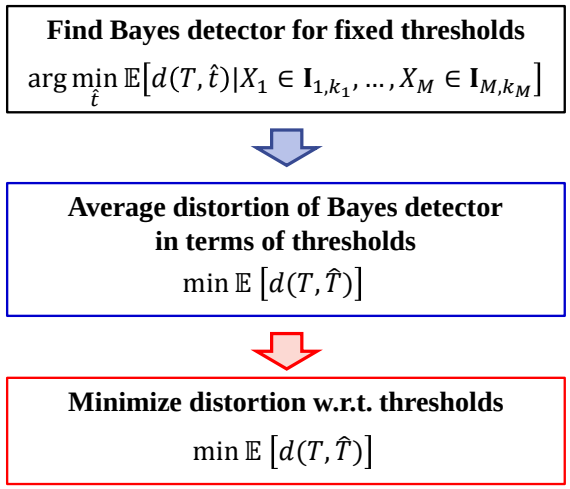


Figure 5.1: Method utilized to design the scalar quantizers.

We will still utilize the methods from Chapters 3 and 4 to calculate fundamental limits by approximating continuous sources uniformly distributed over $[0,1]$ with limits of discrete sources uniformly distributed over $\left\{\frac{1}{K}, \ldots, \frac{K-1}{K}\right\}$ as $K \to \infty$. This is possible because if a quantization scheme exists for a $\mathscr{U}([0,1])$ source it can be used as a quantization scheme for the $\mathscr{U}\left\{\frac{1}{K}, \ldots, \frac{K-1}{K}\right\}$ source, and will achieve a difference in distortion that will $\to 0$ as $K \to \infty$. Similarly, a quantization scheme for the $\mathscr{U}\left\{\frac{1}{K}, \ldots, \frac{K-1}{K}\right\}$ source can be used as a quantization scheme for the $\mathscr{U}([0,1])$ source by first selecting the nearest element in the set $\left\{\frac{1}{K}, \ldots, \frac{K-1}{K}\right\}$ to the continuous source $\mathscr{U}([0,1])$, then further discrete quantizing it, yielding a change in distortion which also $\to 0$ as $K \to \infty$.

Bearing this method for approximating the fundamental limits in mind, we set about the problem of designing scalar quantizers for $\mathscr{U}([0,1])$ sources under the three distortion metrics from Chapter 4 to approach these fundamental limits.

5.1 Scalar Quantizer Design

The process we will follow to design optimal scalar quantizers is depicted in Fig. 5.1. Let $X_i^{(n)}$ represent the channel capacity of user i on physical resource block n, $i \in \{1, ..., M\}$. We will assume that the random variables $X_i^{(n)}$ are i.i.d. with cumulative distribution function (CDF) $F_{X_i}(x_i)$ and probability density function (PDF) $f_{X_i}(x_i)$ defined on the support $\mathscr{X}_i \subseteq \mathbb{R}$. The quantizer for user i breaks its support $\mathscr{X}_i$ into K intervals indexed by $k_i \in \{0, ..., K\}$

$$\mathscr{X}_i = \bigcup_{k_i=1}^{K} \mathbf{I}_{i,k_i}, \quad \mathbf{I}_{i,k_i} = [l_{i,k_i-1}, l_{i,k_i}]. \tag{5.1}$$

First, the Bayes' detector/estimator selecting the estimate $\hat{t}(\boldsymbol{k}, \boldsymbol{\ell})$ which minimizes the expected distortion given the quantization indices $\boldsymbol{k} = [k_1, \ldots, k_M]$ is found by solving

$$\hat{t}(\boldsymbol{k}, \boldsymbol{\ell}) = \arg\min_{\hat{t}} \mathbb{E}\left[d(T, \hat{t}) \mid X_1 \in \mathbf{I}_{k_1}, \ldots, \ X_M \in \mathbf{I}_{k_M}\right] \tag{5.2}$$

as a function of the thresholds $\boldsymbol{\ell} = [l_{i,k} \,|k \in \{1, \ldots, K\}]$ between the quantization levels and the distribution $F_{X_i}(x_i), i \in \{1, \ldots, M\}$ of the sources. Next, the expected distortion of the Bayes' detector/estimator is expressed in terms of exclusively the thresholds between the quantization levels and the distribution of the source.

$$D(\boldsymbol{\ell}) = \sum_{\boldsymbol{k} \in \{1,\ldots,K\}^M} \mathbb{E}\left[d(T, \hat{t}\,(\boldsymbol{k}, \boldsymbol{\ell})) \,|X_1 \in \mathbf{I}_{k_1}, \ldots, \ X_M \in \mathbf{I}_{k_M}\right] \prod_{i=1}^{M} \mathbb{P}[X_i \in \mathbf{I}_{k_i}] \tag{5.3}$$

Finally, this expression is numerically optimized to yield the minimum distortion quantizer for each number of quantization levels K.

$$D_K = \min_{\boldsymbol{\ell}} D(\boldsymbol{\ell}). \tag{5.4}$$

The resulting sum-rate distortion pair for each K is then (R_K, D_K) where

$$R_K = -\sum_{i=1}^{M}\sum_{k_1=1}^{K} \mathbb{P}\left[X_i \in \mathbf{I}_{k_i}\right] \log_2\left(\mathbb{P}\left[X_i \in \mathbf{I}_{k_i}\right]\right) \tag{5.5}$$

which assumes that the quantization indices will be block Huffman coded so as to approach a rate equal to their entropy. Note that we do not consider the more complicated case of entropy constrained scalar quantization as the simpler minimum distortion quantizers already require calculations that are somewhat dense, and also, for the sources of interest, yield rate distortion tradeoffs close to the fundamental limits.

In the remainder of this chapter, we perform the calculations in (5.2) and (5.3) in detail, then numerically perform the optimization (5.4) for the case of a uniform source $\mathscr{U}([0,1])$.

5.2 Two User Max Scalar Quantizer for Broadcasting

For broadcasting, we consider a resource allocation function $T = \max(X_1, X_2)$. In order to find a maximum value between users' sources, the base station want to estimate the maximum value based on the observations from each user. The associated distortion measure is

$$d(T,\hat{T}) = \begin{cases} T - \hat{T} & \text{if } \hat{T} \leq T \\ T & \text{otherwise} \end{cases} = T - \hat{T}\mathbb{1}_{\hat{T}\leq T}$$

Thm. 5: The optimal Bayes detector for $T = \max(X_1, X_2)$ is given by

$$\hat{t}^* = \begin{cases} \hat{t}_1^* & \text{If } l_{1,k_1-1} \geq l_{2,k_2} \\ \hat{t}_2^* & \text{If } l_{2,k_2-1} \geq l_{1,k_1} \\ \hat{t}_{01}^* = \arg\max_{\hat{t}_{11}^*, \hat{t}_{12}^*} \left[w_{11}(\hat{t}_{11}^*), w_{12}(\hat{t}_{12}^*) \right] & \text{If } \max(l_{2,k_2-1}, l_{1,k_1-1}) < l_{2,k_2} \leq l_{1,k_1} \\ \hat{t}_{02}^* = \arg\max_{\hat{t}_{21}^*, \hat{t}_{22}^*} \left[w_{21}(\hat{t}_{21}^*), w_{22}(\hat{t}_{22}^*) \right] & \text{If } \max(l_{1,k_1-1}, l_{2,k_2-1}) < l_{1,k_1} \leq l_{2,k_2} \end{cases} \tag{5.6}$$

where

$$\hat{t}_1^* = \begin{cases} \text{sol} \left\{ \hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t} f_{X_1}(\hat{t}), -2 f_{X_1}(\hat{t}) - \hat{t} f'_{X_1}(\hat{t}) \leq 0 \right\} & \hat{t} \in \mathbf{I}_{1,k_1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases} \tag{5.7}$$

$$\hat{t}_2^* = \begin{cases} \text{sol} \left\{ \hat{t} | F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t} f_{X_2}(\hat{t}), -2 f_{X_2}(\hat{t}) - \hat{t} f'_{X_2}(\hat{t}) \leq 0 \right\} & \hat{t} \in \mathbf{I}_{2,k_2} \\ l_{2,k_2-1} & \text{otherwise} \end{cases} \tag{5.8}$$

$$w_{11}(\hat{t}) \doteq \hat{t} \left[1 - F_{X_1|X_1 \in \mathbf{I}_{1,k_1}}(\hat{t}) F_{X_2|X_2 \in \mathbf{I}_{2,k_2}}(\hat{t}) \right], \qquad w_{12}(\hat{t}) \doteq \hat{t} \left[1 - F_{X_1|X_1 \in \mathbf{I}_{1,k_1}}(\hat{t}) \right]$$

$$\hat{t}_{11}^* = \begin{cases} \text{sol} \left\{ \hat{t} | w'_{11}(\hat{t}) = 0, w''_{11}(\hat{t}) \leq 0 \right\} & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{2,k_2} \\ \max(l_{1,k_1-1}, l_{2,k_2-1}) & \text{otherwise} \end{cases} \tag{5.9}$$

$$\hat{t}_{12}^* = \begin{cases} \text{sol} \left\{ \hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t} f_{X_1}(\hat{t}), -2 f_{X_1}(\hat{t}) - \hat{t} f'_{X_1}(\hat{t}) \leq 0 \right\} & l_{2,k_2} \leq \hat{t} \leq l_{1,k_1} \\ l_{2,k_2} & \text{otherwise} \end{cases} \tag{5.10}$$

$$w_{21}(\hat{t}) \doteq \hat{t} \left[1 - F_{X_1|X_1 \in \mathbf{I}_{1,k_1}}(\hat{t}) F_{X_2|X_2 \in \mathbf{I}_{2,k_2}}(\hat{t}) \right], \qquad w_{22}(\hat{t}) \doteq \hat{t} \left[1 - F_{X_2|X_2 \in \mathbf{I}_{2,k_2}}(\hat{t}) \right]$$

$$\hat{t}_{21}^{*} = \begin{cases} \text{sol}\left\{\hat{t} | w_{21}'(\hat{t}) = 0, w_{21}''(\hat{t}) \leq 0\right\} & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{1,k_1} \\ \max(l_{1,k_1-1}, l_{2,k_2-1}) & \text{otherwise} \end{cases} \tag{5.11}$$

$$\hat{t}_{22}^{*} = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t} f_{X_2}(\hat{t}), -2f_{X_2}(\hat{t}) - \hat{t} f_{X_2}'(\hat{t}) \leq 0\right\} & l_{1,k_1} \leq \hat{t} \leq l_{2,k_2} \\ l_{1,k_1} & \text{otherwise} \end{cases} \tag{5.12}$$

Furthermore, the expected distortion when using the optimal Bayes detector is given by

$$\begin{aligned}
&\mathbb{E}\left[d(T, \hat{t})\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \left\{ \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 f_{X_1}(x_1) dx_1 - \hat{t}_1^{*} \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t}_1^{*})\right] \right\} \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_2} \left\{ \int_{l_{2,k_2-1}}^{l_{2,k_2}} x_2 f_{X_2}(x_2) dx_2 - \hat{t}_2^{*} \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(\hat{t}_2^{*})\right] \right\} \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{2,k_2}} t \left\{ f_{X_1}(t)[F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})] + [F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})] f_{X_2}(t) \right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{l_{2,k_2}}^{l_{1,k_1}} t f_{X_1}(t) dt [F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \max_{\hat{t}_{11}^{*}, \hat{t}_{12}^{*}} \left[w_{11}(\hat{t}_{11}^{*}), w_{12}(\hat{t}_{12}^{*})\right] [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{1,k_1}} t \left\{ f_{X_2}(t)[F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})] + [F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})] f_{X_1}(t) \right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{l_{1,k_1}}^{l_{2,k_2}} t f_{X_2}(t) dt [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \max_{\hat{t}_{21}^{*}, \hat{t}_{22}^{*}} \left[w_{21}(\hat{t}_{21}^{*}), w_{22}(\hat{t}_{22}^{*})\right] [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})]
\end{aligned}$$

Proof : For a Max Quantizer, an average distortion can be expressed as follows

$$\begin{aligned}
\mathbb{E}\left[d(T,\hat{T})\right] &= \mathbb{E}\left[\mathbb{E}\left[d(T,\hat{T})|U_1,U_2\right]\right] \\
&= \sum_{k_1,k_2} \mathbb{E}\left[d(T,\hat{T})|U_1=k_1,U_2=k_2\right] \mathbf{P}\left[U_1=k_1,U_2=k_2\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \mathbb{E}\left[d(X_1,\hat{T})|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_2} \mathbb{E}\left[d(X_2,\hat{T})|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_0} \mathbb{E}\left[d(T,\hat{T})|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]
\end{aligned}$$

Where $\mathbf{T}_1$ is the region satisfying $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$ $(l_{2,k_2} \leq l_{1,k_1-1})$, $\mathbf{T}_2$ is the region satisfying $\mathbf{I}_{1,k_1} < \mathbf{I}_{2,k_2}$ $(l_{1,k_1} \leq l_{2,k_2-1})$, and $\mathbf{T}_0$ is the shared region between $\mathbf{I}_1$ and $\mathbf{I}_2$ $(\max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \min(l_{1,k_1}, l_{2,k_2}) \leq \max(l_{1,k_1}, l_{2,k_2}))$. In order to find a quantizer minimizing an average distortion, we need to evaluate a minimum distortion term as follows,

$$\begin{aligned}
&\arg\min_{\hat{t}} \mathbb{E}\left[d(T,\hat{t})|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&\quad= \arg\min_{\hat{t}} \mathbb{E}\left[T-\hat{t}\mathbb{1}_{T>\hat{t}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&\quad= \arg\min_{\hat{t}} \left\{\mathbb{E}\left[T|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]-\hat{t}\mathbf{P}\left[T>\hat{t}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]\right\}
\end{aligned}$$

For region $\mathbf{T}_1$, when the maximum value $T = \max(X_1, X_2)$ is determined by X_1, the region $\mathbf{T}_1$ can be expressed as $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$ $(l_{2,k_2} < l_{1,k_1-1})$.

$$\begin{aligned}
\hat{t}_1^* &= \arg\min_{\hat{t}} \left\{\mathbb{E}\left[T|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]-\hat{t}\mathbf{P}\left[T>\hat{t}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]\right\} \\
&= \mathbb{E}\left[X_1|X_1\in\mathbf{I}_{1,k_1}\right]-\arg\min_{\hat{t}} \hat{t}\mathbf{P}\left[X_1>\hat{t}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&= \arg\max_{\hat{t}} \hat{t}\mathbf{P}\left[X_1>\hat{t}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]
\end{aligned}$$

In order to find an optimum estimation $\hat{t}_1^*$ minimizing the average distortion in region $\mathbf{T}_1$, the necessary and sufficient condition is to find $\hat{t}$ maximizing $\hat{t}\mathbf{P}\left[X_1 > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]$.

Let's evaluate $\hat{t}\mathbf{P}\left[X_1 > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]$. Depending on the location of $\hat{t}$, we can determine the expected value as follows,

$$\hat{t}\mathbf{P}\left[X_1 > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] = \begin{cases} \hat{t} & \hat{t} \leq l_{1,k_1-1} \\ \hat{t}\frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t})}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} & \hat{t} \in I_{1,k_1} \\ 0 & \hat{t} \geq l_{1,k_1} \end{cases}$$

Since the maximum of each region can be included on the boundary from $\mathbf{I}_{1,k_1}$, it is suffices to evaluate as follows,

$$\hat{t}_1^* = \arg\max_{\hat{t} \in \mathbf{I}_{1,k_1}} \hat{t}\frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t})}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} \tag{5.13}$$

If it is possible to take the first and second derivative of the evaluation with respect to $\hat{t}$, the estimation $\hat{t}_1^*$ can be determined by the solution as follows,

$$\hat{t}_1^* = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t}f_{X_1}(\hat{t}), -2f_{X_1}(\hat{t}) - \hat{t}f'_{X_1}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{1,k_1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases} \tag{5.14}$$

In order to find an average distortion in the region $\mathbf{T}_1$, find $\mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right]$ as follows

$$\mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right] = \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 \frac{f_{X_1}(x_1)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} dx_1 \tag{5.15}$$

The average conditional minimum distortion in the region $\mathbf{T}_1$ given $X_1 \in \mathbf{I}_{1,k_1}$ and

$X_2 \in \mathbf{I}_{2,k_2}$ is

$$\begin{aligned}
&\mathbb{E}[X_1|X_1 \in \mathbf{I}_{1,k_1}] - \hat{t}_1^* \mathbf{P}[X_1 > \hat{t}_1^*|X_1 \in \mathbf{I}_{1,k_1}] \\
&\quad = \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 \frac{f_{X_1}(x_1)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} dx_1 - \hat{t}_1^* \frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t}_1^*)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})}
\end{aligned}$$

Therefore, the average distortion in $\mathbf{T}_1$,

$$\begin{aligned}
&\sum_{(k_1,k_2)\in\mathbf{T}_1} \mathbb{E}\left[d(X_1,\hat{t})|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \left\{ \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 f_{X_1}(x_1) dx_1 - \hat{t}_1^* \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t}_1^*)\right] \right\} \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})\right]
\end{aligned}$$

For region $\mathbf{T}_2$, when the maximum value T is determined by X_2, the region $\mathbf{T}_2$ can be expressed as $\mathbf{I}_{1,k_1} < \mathbf{I}_{2,k_2}$ or $l_{1,k_1} < l_{2,k_2-1}$.

$$\begin{aligned}
\hat{t}_2^* &= \mathbb{E}\left[X_2|X_2 \in \mathbf{I}_{2,k_2}\right] - \arg\min_{\hat{t}} \hat{t}\mathbf{P}\left[X_2 > \hat{t}|X_2 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&= \arg\max_{\hat{t}} \hat{t}\mathbf{P}\left[X_2 > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]
\end{aligned}$$

Similar to the region $\mathbf{T}_1$, the estimation $\hat{t}_2^*$ can be determined by the solution as follows,

$$\hat{t}_2^* = \begin{cases} \text{sol}\left\{\hat{t}|F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t}f_{X_2}(\hat{t}), -2f_{X_2}(\hat{t}) - \hat{t}f'_{X_2}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{2,k_2} \\ l_{2,k_2-1} & \text{otherwise} \end{cases} \tag{5.16}$$

The average conditional minimum distortion in the region $\mathbf{T}_2$ given $X_1 \in \mathbf{I}_{1,k_1}$ and

$X_2 \in \mathbf{I}_{2,k_2}$ is

$$\begin{aligned}
&\mathbb{E}[X_2|X_2 \in \mathbf{I}_{2,k_2}] - \hat{t}_2^* \mathbf{P}[X_2 > \hat{t}_2^*|X_2 \in \mathbf{I}_{2,k_2}] \\
&\quad = \int_{l_{2,k_2-1}}^{l_{2,k_2}} x_2 \frac{f_{X_2}(x_2)}{F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})} dx_2 - \hat{t}_2^* \frac{F_{X_2}(l_{2,k_2}) - F_{X_2}(\hat{t}_2^*)}{F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})}
\end{aligned}$$

Therefore, the average distortion in Region $\mathbf{T}_2$ is

$$\begin{aligned}
&\sum_{(k_1,k_2)\in\mathbf{T}_2} \mathbb{E}\left[d(X_2,\hat{t})|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_2} \left\{ \int_{l_{2,k_2-1}}^{l_{2,k_2}} x_2 f_{X_2}(x_2) dx_2 - \hat{t}_2^* \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(\hat{t}_2^*)\right] \right\} \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})\right]
\end{aligned}$$

For region $\mathbf{T}_0$, when the maximum value $T = \max(X_1, X_2)$ is not determined by either X_1 or X_2, the region $\mathbf{T}_0$ need to estimate $\hat{t}$ minimizing an average distortion under distribution of T as follows,

$$\begin{aligned}
\hat{t}_0^* &= \arg\min_{\hat{t}} \left\{\mathbb{E}\left[T|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] - \hat{t}\mathbf{P}\left[T > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]\right\} \\
&= \mathbb{E}\left[T|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] - \arg\min_{\hat{t}} \hat{t}\mathbf{P}\left[T > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&= \arg\max_{\hat{t}} \hat{t}\mathbf{P}\left[T > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]
\end{aligned}$$

For the case of the region by $\max(l_{2,k_2-1}, l_{1,k_1-1}) < l_{2,k_2} \leq l_{1,k_1}$, the CDF of T

given $X_1 \in \mathbf{I}_{1,k_1}$ and $X_2 \in \mathbf{I}_{2,k_2}$ is

$$
\begin{aligned}
F_{T|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}}(t) &= F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(t)F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(t) \\
&= \begin{cases} 0 & t \le \max(l_{1,k_1-1}, l_{2,k_2-1}) \\ F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(t)F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(t) & \max(l_{1,k_1-1}, l_{2,k_2-1}) \le t \le l_{2,k_2} \\ F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(t) & l_{2,k_2} \le t \le l_{1,k_1} \\ 1 & t \ge l_{1,k_1} \end{cases}
\end{aligned}
$$

then, $\hat{t}\mathbf{P}\left[T > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]$ is

$$
\begin{aligned}
&\hat{t}\mathbf{P}[T > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}] \\
&= \begin{cases} \hat{t} & \hat{t} \le \max(l_{1,k_1-1}, l_{2,k_2-1}) \\ \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & \max(l_{1,k_1-1}, l_{2,k_2-1}) \le \hat{t} \le l_{2,k_2} \\ \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})\right] & l_{2,k_2} \le \hat{t} \le l_{1,k_1} \\ 0 & \hat{t} \ge l_{1,k_1} \end{cases}
\end{aligned}
$$

Since the maximum of each region can be also included on the boundary from $\max(l_{1,k_1-1}, l_{2,k_2-1}) \le \hat{t} \le l_{1,k_1}$, it is suffices to evaluate as follows,

$$
\hat{t}^*_{01} = \arg\max_{\hat{t}} \begin{cases} \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & \max(l_{1,k_1-1}, l_{2,k_2-1}) \le \hat{t} \le l_{2,k_2} \\ \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})\right] & l_{2,k_2} \le \hat{t} \le l_{1,k_1} \end{cases}
$$

Let's suppose two functions as follows,

$$
w_{11}(\hat{t}) \doteq \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right], \qquad w_{12}(\hat{t}) \doteq \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t})\right] \tag{5.17}
$$

For the region within $\max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{2,k_2}$, a condition such that $w'_{11}(\hat{t}) = 0$ is

$$\begin{aligned}[-F_{X_1}(l_{1,k_1})+F_{X_1}(\hat{t}) + \hat{t}f_{X_1}(\hat{t})]F_{X_2}(l_{2,k_2-1}) + F_{X_1}(l_{1,k_1-1}) \left[-F_{X_2}(l_{2,k_2}) + F_{X_2}(\hat{t}) + \hat{t}f_{X_2}(\hat{t})\right] \\ = -F_{X_1}(l_{1,k_1})F_{X_2}(l_{2,k_2}) + F_{X_1}(\hat{t})F_{X_2}(\hat{t}) + \hat{t}\left[f_{X_1}(\hat{t})F_{X_2}(\hat{t}) + F_{X_1}(\hat{t})f_{X_2}(\hat{t})\right]\end{aligned}$$

Also, the second order condition ($w''_{11}(\hat{t}) \leq 0$) is

$$\begin{aligned}\left[2f_{X_1}(\hat{t}) + \hat{t}f'_{X_1}(\hat{t})\right]\left[F_{X_2}(l_{2,k_2-1}) - F_{X_2}(\hat{t})\right] + \left[F_{X_1}(l_{1,k_1-1}) - F_{X_1}(\hat{t})\right]\left[2f_{X_2}(\hat{t}) + \hat{t}f'_{X_2}(\hat{t})\right] \\ \leq 2\hat{t}f_{X_1}(\hat{t})f_{X_2}(\hat{t})\end{aligned}$$

Based on the first and second derivative of the evaluation with respect to $\hat{t}$, the estimation $\hat{t}^*_{11}$ can be determined by the solution as follows,

$$\hat{t}^*_{11} = \begin{cases} \text{sol}\left\{\hat{t} | w'_{11}(\hat{t}) = 0, w''_{11}(\hat{t}) \leq 0\right\} & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{2,k_2} \\ \max(l_{1,k_1-1}, l_{2,k_2-1}) & \text{otherwise} \end{cases} \tag{5.18}$$

For the region within $l_{2,k_2} \leq \hat{t} \leq l_{1,k_1}$, based on the first and second derivative of the function $w_{12}(\hat{t})$ with respect to $\hat{t}$, the estimation $\hat{t}^*_{12}$ can be determined by the solution as follows,

$$\hat{t}^*_{12} = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t}f_{X_1}(\hat{t}), -2f_{X_1}(\hat{t}) - \hat{t}f'_{X_1}(\hat{t}) \leq 0\right\} & l_{2,k_2} \leq \hat{t} \leq l_{1,k_1} \\ l_{2,k_2} & \text{otherwise} \end{cases} \tag{5.19}$$

Therefore, the optimum value of $\hat{t}$ within $\max(l_{2,k_2-1}, l_{1,k_1-1}) < l_{2,k_2} \leq l_{1,k_1}$ is determined by

$$\hat{t}^*_{01} = \arg\max_{\hat{t}^*_{11}, \hat{t}^*_{12}} \left[w_{11}(\hat{t}^*_{11}), w_{12}(\hat{t}^*_{12})\right]$$

For the case of the region by $\max(l_{1,k_1-1}, l_{2,k_2-1}) < l_{1,k_1} \leq l_{2,k_2}$,
$\hat{t}\mathbf{P}\left[T > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]$ is

$$\hat{t}\mathbf{P}[T > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}]$$
$$= \begin{cases} \hat{t} & \hat{t} \leq \max(l_{1,k_1-1}, l_{2,k_2-1}) \\ \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t}) F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{1,k_1} \\ \hat{t}\left[1 - F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & l_{1,k_1} \leq \hat{t} \leq l_{2,k_2} \\ 0 & \hat{t} \geq l_{2,k_2} \end{cases}$$

Since the maximum of each region can be also included on the boundary from $\max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{2,k_2}$, it is suffices to evaluate as follows,

$$\hat{t}^*_{02} = \arg\max_{\hat{t}} \begin{cases} \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t}) F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{1,k_1} \\ \hat{t}\left[1 - F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] & l_{1,k_1} \leq \hat{t} \leq l_{2,k_2} \end{cases}$$

Let's suppose two functions

$$w_{21}(\hat{t}) \doteq \hat{t}\left[1 - F_{X_1|X_1\in\mathbf{I}_{1,k_1}}(\hat{t}) F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right], \qquad w_{22}(\hat{t}) \doteq \hat{t}\left[1 - F_{X_2|X_2\in\mathbf{I}_{2,k_2}}(\hat{t})\right] \tag{5.20}$$

Similar to the case of $\max(l_{1,k_1-1}, l_{2,k_2-1}) < l_{2,k_2} \leq l_{1,k_1}$, the estimation $\hat{t}^*_{21}$ can be determined by the solution as follows,

$$\hat{t}^*_{21} = \begin{cases} \text{sol}\left\{\hat{t} | w'_{21}(\hat{t}) = 0, w''_{21}(\hat{t}) \leq 0\right\} & \max(l_{1,k_1-1}, l_{2,k_2-1}) \leq \hat{t} \leq l_{1,k_1} \\ \max(l_{1,k_1-1}, l_{2,k_2-1}) & \text{otherwise} \end{cases} \tag{5.21}$$

Also, the estimation $\hat{t}^*_{22}$ can be determined by the solution as follows,

$$\hat{t}^*_{22} = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t} f_{X_2}(\hat{t}), -2f_{X_2}(\hat{t}) - \hat{t} f'_{X_2}(\hat{t}) \leq 0\right\} & l_{1,k_1} \leq \hat{t} \leq l_{2,k_2} \\ l_{1,k_1} & \text{otherwise} \end{cases} \tag{5.22}$$

Therefore, the optimum value of $\hat{t}$ within $\max(l_{2,k_2-1}, l_{1,k_1-1}) < l_{1,k_1} \leq l_{2,k_2}$ is determined by

$$\hat{t}^*_{02} = \arg\max_{\hat{t}^*_{21}, \hat{t}^*_{22}} \left[w_{21}(\hat{t}^*_{21}), w_{22}(\hat{t}^*_{22})\right]$$

■

From the expression of average distortion as function of parameters of decision boundaries, an expected quantizer minimizing the expected distortion need to solve an optimization problem.

Example : Two user 2-level Distributed Homogeneous Scalar Quantizer with Uniform $(0, 1)$

For 2-level scalar quantizer, we set up parameters as follows,

$$\mathbf{I}_{1,1} = [l_{1,0}, l_{1,1}], \quad \mathbf{I}_{1,2} = [l_{1,1}, l_{1,2}]$$
$$\mathbf{I}_{2,1} = [l_{2,0}, l_{2,1}], \quad \mathbf{I}_{2,2} = [l_{2,1}, l_{2,2}]$$

where $l_{1,0} = l_{2,0} = 0$ and $l_{1,2} = l_{2,2} = 1$.

We want to find $l_{1,1}$ and $l_{2,1}$ minimizing an average distortion. For convenience, we analyze a homogeneous scalar quantizer which has same parameters between users. In this case, we set $l_{1,1} = l_{2,1} = l$. First, we need to find the expression of minimum average distortion as a function of l, then we solve an optimization problem. Based on the analysis in the previous part, we calculate an average distortion in a given

region by k_1 and k_2 as follows

$$\begin{cases} \frac{2}{3}l^3 - \frac{2}{3\sqrt{3}}l^3 & \text{If } \ k_1 = 1, k_2 = 1 \ \rightarrow \mathbf{T}_{01} \\ \frac{1}{4}(1-2l^2)l\mathbf{1}_{0.5\geq l} + \frac{1}{2}(1-l)^2 l\mathbf{1}_{0.5<l} & \text{If } \ k_1 = 2, k_2 = 1 \ \rightarrow \mathbf{T}_{1} \\ \frac{1}{4}l(1-2l^2)\mathbf{1}_{0.5\geq l} + \frac{1}{2}l(1-l)^2\mathbf{1}_{0.5<l} & \text{If } \ k_1 = 1, k_2 = 2 \ \rightarrow \mathbf{T}_{2} \\ -\frac{7}{27}l^3 + \frac{4}{3}l^2 - \frac{5}{3}l + \frac{2}{3} - \frac{2}{27}(4l^2 - 6l + 3)^{\frac{3}{2}} & \text{If } \ k_1 = 2, k_2 = 2 \ \rightarrow \mathbf{T}_{01} \end{cases} \tag{5.23}$$

In $0 < l \leq 0.5$, an average distortion is

$$\mathbb{E}\left[d(T,\hat{t})\right] = \frac{2}{3}l^3 - \frac{2}{3\sqrt{3}}l^3 + \frac{1}{2}(1-2l^2)l - \frac{7}{27}l^3 + \frac{4}{3}l^2 - \frac{5}{3}l + \frac{2}{3} - \frac{2}{27}(4l^2 - 6l + 3)^{\frac{3}{2}}$$

and minimum 0.2204 at $l = 0.5$.

In $0.5 \leq l < 1$, an average distortion is

$$\mathbb{E}\left[d(T,\hat{t})\right] = \frac{2}{3}l^3 - \frac{2}{3\sqrt{3}}l^3 + (1-l)^2 l - \frac{7}{27}l^3 + \frac{4}{3}l^2 - \frac{5}{3}l + \frac{2}{3} - \frac{2}{27}(4l^2 - 6l + 3)^{\frac{3}{2}}$$

and minimum 0.1742 at $l = 0.7257$.

Therefore, we need to choose a parameter $l_{1,1} = l_{2,1} = 0.7257$ with minimum distortion 0.1742 for homogeneous scalar quantizer.

Example : Two user K-level Distributed Scalar Quantizer with Uniform $(0,1)$

When two users' source is Uniform $(0,1)$, an expected minimum distortion is

$$
\begin{aligned}
&\mathbb{E}\left[d(T,\hat{t})\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \begin{cases} \frac{1}{4}(l_{1,k_1}^2 - 2l_{1,k_1-1}^2)(l_{2,k_2} - l_{2,k_2-1}) & l_{1,k_1} \geq 2l_{1,k_1-1} \\ \frac{1}{2}(l_{1,k_1} - l_{1,k_1-1})^2(l_{2,k_2} - l_{2,k_2-1}) & l_{1,k_1} < 2l_{1,k_1-1} \end{cases} \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_2} \begin{cases} \frac{1}{4}(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2}^2 - 2l_{2,k_2-1}^2) & l_{2,k_2} \geq 2l_{2,k_2-1} \\ \frac{1}{2}(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1})^2 & l_{2,k_2} < 2l_{2,k_2-1} \end{cases} \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \left\{ \left[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2\right] \Big|_{\max(l_{1,k_1-1},l_{2,k_2-1})}^{l_{2,k_2}} + \frac{1}{2}(l_{2,k_2} - l_{2,k_2-1})t^2 \Big|_{l_{2,k_2}}^{l_{1,k_1}} \right\} \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \{(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(w_{11}(t_{11}^*), w_{12}(t_{12}^*)\right)\} \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \left\{ \left[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2\right] \Big|_{\max(l_{1,k_1-1},l_{2,k_2-1})}^{l_{1,k_1}} + \frac{1}{2}(l_{1,k_1} - l_{1,k_1-1})t^2 \Big|_{l_{1,k_1}}^{l_{2,k_2}} \right\} \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \{(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(w_{21}(t_{21}^*), w_{22}(t_{22}^*)\right)\}
\end{aligned}
$$

where functions w and solutions t^* are determined by Equations 5.17, 5.20, 5.18, 5.19, 5.21, and 5.22.

Fig. 5.2 shows the per user rate and fractional gap to omniscient spectral efficiency with distributed functional max scalar quantizer under uniform distribution $U(0,1)$ parameterized by the solution. The homogeneous and heterogeneous quantizers resulting from numerical optimization yield performance very similar to one another. Representing the continuous source with a uniformly quantized discrete source, calculating the rate distortion region for this discrete source, then letting the number of quantization levels go to infinity, we obtain a bound for the rate distortion function of the continuous source. The fundamental limit for the overhead performance tradeoff for cardinality 16, 32, and 64 discrete approximations of the continuous uniform

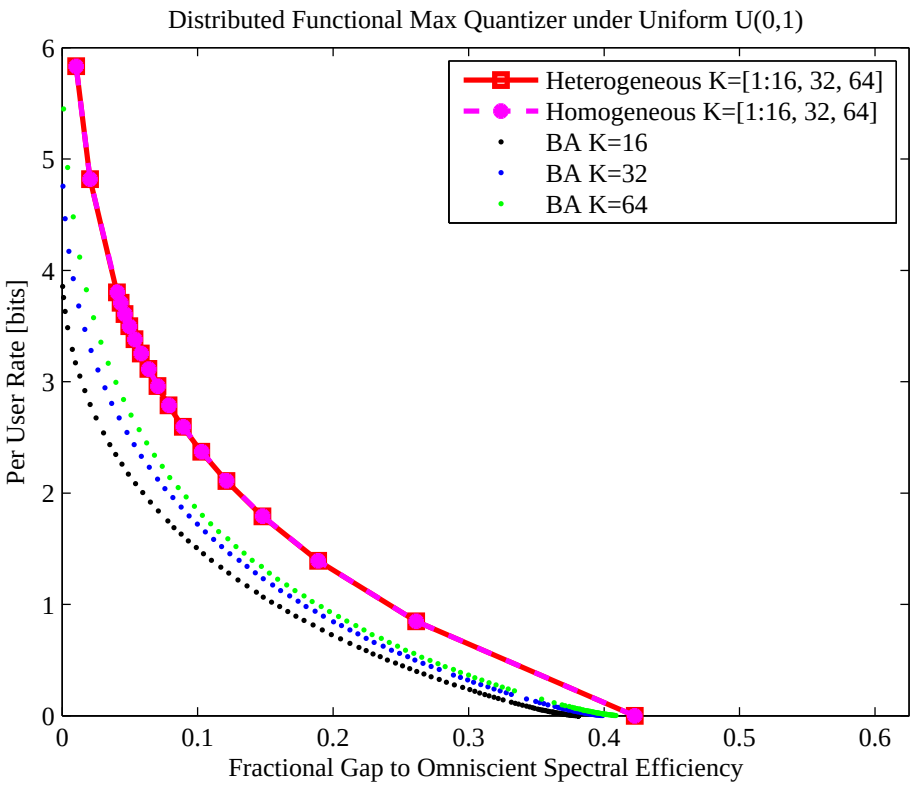


Figure 5.2: Distributed Functional max Scalar Quantizer under Uniform U(0,1)

source are depicted, and appear to be converging. The achievable scalar quantization schemes are not particularly far from the fundamental limits, leaving only a small gain possible from a better designed scalar of vector quantizer.

5.3 Argmax Scalar Quantizer

For rateless communication, we consider resource allocation functions $T = \arg\max(X_1, X_2)$. The associated distortion measure is

$$D_A(T, \hat{T}) = X_T - X_{\hat{T}} \tag{5.24}$$

Thm. 6: The optimal Bayes detector for $T = \arg\max(X_1, X_2)$ is given by

$$\hat{t}^* = \begin{cases} 1 & \text{If } l_{1,k_1-1} \geq l_{2,k_2} \\ 2 & \text{If } l_{2,k_2-1} \geq l_{1,k_1} \\ \hat{t}_0^* & \text{If } \max(l_{2,k_2-1}, l_{1,k_1-1}) \leq \min(l_{1,k_1}, l_{2,k_2}) \leq \max(l_{1,k_1}, l_{2,k_2}) \end{cases} \tag{5.25}$$

where

$$\hat{t}_0^* = \begin{cases} 1 & \text{If } \mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right] \geq \mathbb{E}\left[X_2 | X_2 \in \mathbf{I}_{2,k_2}\right] \\ 2 & \text{If } \mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right] < \mathbb{E}\left[X_2 | X_2 \in \mathbf{I}_{2,k_2}\right] \end{cases} \tag{5.26}$$

Furthermore, the expected distortion when using the optimal Bayes detector is given by

$$\begin{aligned}
&\mathbb{E}\left[d(T, \hat{t})\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{2,k_2}} t\left\{f_{X_1}(t)[F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})] + [F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})]f_{X_2}(t)\right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{l_{2,k_2}}^{l_{1,k_1}} t f_{X_1}(t) dt [F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \max\left(\mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right], \mathbb{E}\left[X_2 | X_2 \in \mathbf{I}_{2,k_2}\right]\right) \\
&\quad \cdot [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{1,k_1}} t\left\{f_{X_2}(t)[F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})] + [F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})]f_{X_1}(t)\right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{l_{1,k_1}}^{l_{2,k_2}} t f_{X_2}(t) dt [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \max\left(\mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right], \mathbb{E}\left[X_2 | X_2 \in \mathbf{I}_{2,k_2}\right]\right) \\
&\quad \cdot [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})]
\end{aligned}$$

Proof : For Argmax Quantizer, an average distortion is

$$\begin{aligned}
\mathbb{E}\left[D\left[T,\hat{T}\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[D(T,\hat{T})|U_1,U_2\right]\right] \\
&= \sum_{k_1,k_2} \mathbb{E}\left[X_T - X_{\hat{T}}|U_1=k_1,U_2=k_2\right]\mathbf{P}\left[U_1=k_1,U_2=k_2\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \mathbb{E}\left[X_1 - X_{\hat{T}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]\mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_2} \mathbb{E}\left[X_2 - X_{\hat{T}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]\mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_0} \mathbb{E}\left[X_T - X_{\hat{T}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]\mathbf{P}\left[X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]
\end{aligned}$$

Where $\mathbf{T}_1$ is the region satisfying $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$, $\mathbf{T}_2$ is the region satisfying $\mathbf{I}_{1,k_1} < \mathbf{I}_{2,k_2}$, and $\mathbf{T}_0$ is the shared region between $\mathbf{I}_1$ and $\mathbf{I}_2$. Similar to the case of Max quantizer, to find a quantizer minimizing an average distortion, we need to evaluate a minimum distortion term as follows,

$$\begin{aligned}
&\arg\min_{\hat{t}} \mathbb{E}\left[D(T,\hat{t})|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&\quad= \arg\min_{\hat{t}} \mathbb{E}\left[X_T - X_{\hat{t}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&\quad= \mathbb{E}\left[X_T|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] - \arg\min_{\hat{t}} \mathbb{E}\left[X_{\hat{t}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right] \\
&\quad= \arg\max_{\hat{t}} \mathbb{E}\left[X_{\hat{t}}|X_1\in\mathbf{I}_{1,k_1},X_2\in\mathbf{I}_{2,k_2}\right]
\end{aligned}$$

For region $\mathbf{T}_1$, when the user index having maximum value is determined by X_1, the region $\mathbf{T}_1$ can be expressed as $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$ or $l_{2,k_2} < l_{1,k_1-1}$. Because of $T = 1$, this means $\hat{t} = 1$. An expected distortion is 0.

For region $\mathbf{T}_2$, similar to the region $\mathbf{T}_1$, the $\hat{t} = 2$ and an expected distortion is also 0.

For region $\mathbf{T}_0$, when the user index having maximum value is not determined by

either X_1 or X_2, it suffices to compare $\mathbb{E}\left[X_1|X_1 \in \mathbf{I}_{1,k_1}\right]$ with $\mathbb{E}\left[X_2|X_2 \in \mathbf{I}_{2,k_2}\right]$. Then,

$$\hat{t} = \begin{cases} 1 & \text{If } \mathbb{E}\left[X_1|X_1 \in \mathbf{I}_{1,k_1}\right] \geq \mathbb{E}\left[X_2|X_2 \in \mathbf{I}_{2,k_2}\right] \\ 2 & \text{otherwise} \end{cases} \tag{5.27}$$

To get an average distortion in the region $\mathbf{T}_0$, we need to calculate following term

$$\mathbb{E}\left[X_{\hat{t}}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] = \int_{l_{\hat{t},k_{\hat{t}}-1}}^{l_{\hat{t},k_{\hat{t}}}} \hat{t} f_{\hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}}(\hat{t}) d\hat{t} \tag{5.28}$$

■

Example : Argmax Quantizer with Uniform$(0,1)$

When two users' source is Uniform$(0,1)$, an expected minimum distortion in region $(k_1, k_2) \in \mathbf{T}_1$ is 0 and $\hat{t} = 1$.

For $(k_1, k_2) \in \mathbf{T}_2$, an average distortion is also 0 and $\hat{t} = 2$.

For $(k_1, k_2) \in \mathbf{T}_{01}$, an average distortion is

$$\begin{aligned} &[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2]\Big|_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{2,k_2}} + \frac{1}{2}(l_{2,k_2} - l_{2,k_2-1})t^2\Big|_{l_{2,k_2}}^{l_{1,k_1}} \\ &\quad - (l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(\frac{1}{2}(l_{1,k_1} + l_{1,k_1-1}), \frac{1}{2}(l_{2,k_2} + l_{2,k_2-1})\right) \end{aligned}$$

For $(k_1, k_2) \in \mathbf{T}_{02}$, an average distortion is

$$\begin{aligned} &[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2]\Big|_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{1,k_1}} + \frac{1}{2}(l_{2,k_2} - l_{2,k_2-1})t^2\Big|_{l_{1,k_1}}^{l_{2,k_2}} \\ &\quad - (l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(\frac{1}{2}(l_{1,k_1} + l_{1,k_1-1}), \frac{1}{2}(l_{2,k_2} + l_{2,k_2-1})\right) \end{aligned}$$

Here, we will derive an expression for the distortion that a Bayes detector achieves as a function of quantization levels that are heterogeneous across users, then numerically optimize this expression to obtain the optimal quantization levels. When a

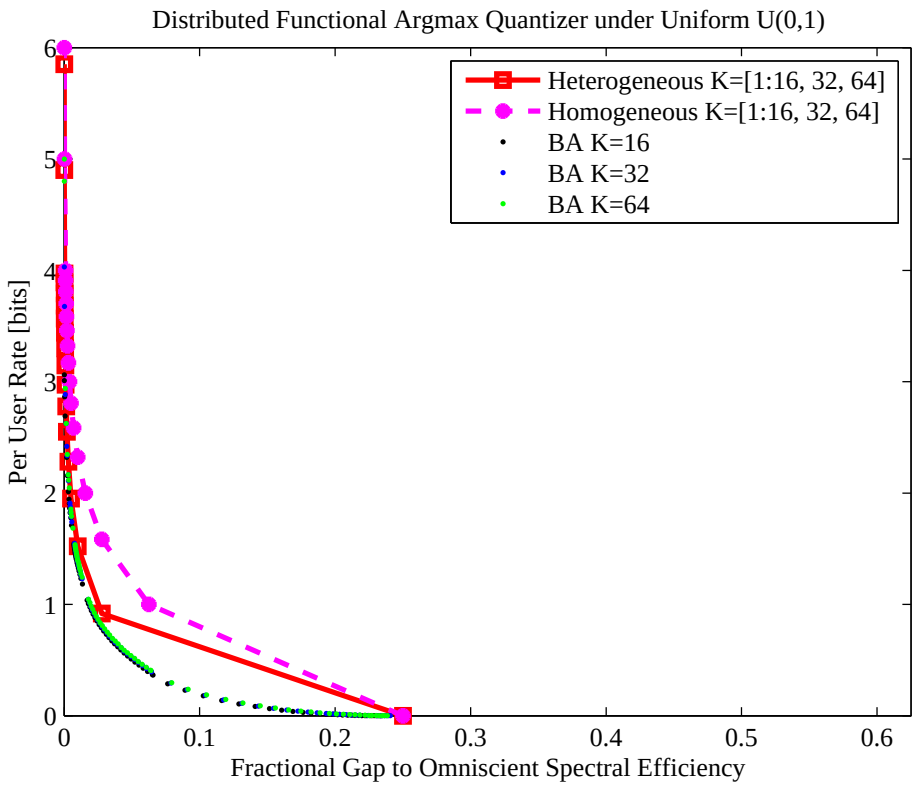


Figure 5.3: Distributed Functional arg max Scalar Quantizer under Uniform U(0,1)

homogeneous constraint is added, the results correctly recover the results in our collaborators work [211], which have derived an analytical expression for the homogeneous case. Interestingly, while [211] creates a, potentially suboptimal, heterogeneous quantizer from the homogeneous quantizer, we observe here that the optimal heterogeneous quantizer we have derived yields nearly identical distortions, at least for the two user case under investigation.

Fig. 5.3 shows the per user rate and fractional gap to omniscient spectral efficiency curve with distributed functional arg max scalar quantizer under uniform distribution $U(0,1)$ parameterized by the solution. Here, a large improvement is achieved by passing between the heterogeneous and homogeneous quantizer case. Additionally, all of the 16, 32, and 64 level fundamental limits are right on top of one another and

have already effectively converged to the continuous limit. Finally, we observe that the designed practical scalar scheme is right up against the fundamental overhead performance tradeoff.

5.4 Max and Argmax Scalar Quantizer

For AMC, we consider resource allocation functions $T_M = \max(X_1, X_2)$ and $T_A = \arg\max(X_1, X_2)$. The associated distortion measure is as follows,

$$D_{M,A}((T_M, T_A), (\hat{T}_M, \hat{T}_A)) = \begin{cases} T_M - \hat{T}_M & \text{if } \hat{T}_M \leq X_{\hat{T}_A} \\ T_M & \text{otherwise} \end{cases} \tag{5.29}$$

Thm. 7: The optimal Bayes detector for $(T_M, T_A) = (\max(X_1, X_2), \arg\max(X_1, X_2))$ is given by

$$\hat{t}^* = \begin{cases} \hat{t}_1^* & \text{If } l_{1,k_1-1} \geq l_{2,k_2} \\ \hat{t}_2^* & \text{If } l_{2,k_2-1} \geq l_{1,k_1} \\ \hat{t}_0^* & \text{If } \max(l_{2,k_2-1}, l_{1,k_1-1}) \leq \min(l_{1,k_1}, l_{2,k_2}) \leq \max(l_{1,k_1}, l_{2,k_2}) \end{cases} \tag{5.30}$$

where

$$\hat{t}_1^* = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t} f_{X_1}(\hat{t}), -2f_{X_1}(\hat{t}) - \hat{t} f'_{X_1}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{1,k_1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases} \tag{5.31}$$

$$\hat{t}_2^* = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t} f_{X_2}(\hat{t}), -2f_{X_2}(\hat{t}) - \hat{t} f'_{X_2}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{2,k_2} \\ l_{2,k_2-1} & \text{otherwise} \end{cases} \tag{5.32}$$

$$\hat{t}_0^* = \arg\max_{\hat{t}_1^*, \hat{t}_2^*} \left[w(\hat{t}_1^*), w(\hat{t}_2^*)\right] \tag{5.33}$$

$$w(\hat{t}_i) = \hat{t}_i \frac{F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}}) - F_{X_{\hat{i}}}(\hat{t})}{F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}}) - F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}-1})} \tag{5.34}$$

Furthermore, the expected distortion when using the optimal Bayes detector is given by

$$\begin{aligned}
&\mathbb{E}\left[d(T,\hat{t})\right] \\
&= \sum_{(k_1,k_2)\in\mathbf{T}_1} \left\{ \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 f_{X_1}(x_1)dx_1 - \hat{t}_1^* \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t}_1^*)\right] \right\} \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_2} \left\{ \int_{l_{2,k_2-1}}^{l_{2,k_2}} x_2 f_{X_2}(x_2)dx_2 - \hat{t}_2^* \left[F_{X_2}(l_{2,k_2}) - F_{X_2}(\hat{t}_2^*)\right] \right\} \left[F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})\right] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{\max(l_{1,k_1-1},l_{2,k_2-1})}^{l_{2,k_2}} t\left\{f_{X_1}(t)[F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})] + [F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})]f_{X_2}(t)\right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \int_{l_{2,k_2}}^{l_{1,k_1}} t f_{X_1}(t)dt[F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{01}} \max_{\hat{t}_1^*,\hat{t}_2^*}\left[w(\hat{t}_1^*), w(\hat{t}_2^*)\right] [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})] \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{\max(l_{1,k_1-1},l_{2,k_2-1})}^{l_{1,k_1}} t\left\{f_{X_2}(t)[F_{X_1}(t) - F_{X_1}(l_{1,k_1-1})] + [F_{X_2}(t) - F_{X_2}(l_{2,k_2-1})]f_{X_1}(t)\right\} dt \\
&+ \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \int_{l_{1,k_1}}^{l_{2,k_2}} t f_{X_2}(t)dt[F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})] \\
&- \sum_{(k_1,k_2)\in\mathbf{T}_{02}} \max_{\hat{t}_1^*,\hat{t}_2^*}\left[w(\hat{t}_1^*), w(\hat{t}_2^*)\right] [F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})][F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})]
\end{aligned}$$

Proof : For a Max and Argmax Quantizer, an average distortion is

$$
\begin{aligned}
&\mathbb{E}[D_{M,A}((T_M,T_A),(\hat{T}_M,\hat{T}_A))] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[D_{M,A}((T_M,T_A),(\hat{T}_M,\hat{T}_A))|U_1,U_2\right]\right] \\
&\quad = \sum_{k_1,k_2} \mathbb{E}\left[T_M - \hat{T}_M \mathbb{1}_{\hat{T}_M \le X_{\hat{T}_A}} | U_1 = k_1, U_2 = k_2\right] \mathbf{P}\left[U_1 = k_1, U_2 = k_2\right] \\
&\quad = \sum_{(k_1,k_2)\in \mathbf{T}_1} \mathbb{E}\left[X_1 - \hat{T}_M \mathbb{1}_{\hat{T}_M \le X_{\hat{T}_A}} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&\quad + \sum_{(k_1,k_2)\in \mathbf{T}_2} \mathbb{E}\left[X_2 - \hat{T}_M \mathbb{1}_{\hat{T}_M \le X_{\hat{T}_A}} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&\quad + \sum_{(k_1,k_2)\in \mathbf{T}_0} \mathbb{E}\left[T_M - \hat{T}_M \mathbb{1}_{\hat{T}_M \le X_{\hat{T}_A}} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \mathbf{P}\left[X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]
\end{aligned}
$$

Where $\mathbf{T}_1$ is the region satisfying $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$, $\mathbf{T}_2$ is the region satisfying $\mathbf{I}_{1,k_1} < \mathbf{I}_{2,k_2}$, and $\mathbf{T}_0$ is the shared region between $\mathbf{I}_1$ and $\mathbf{I}_2$. Similar to the case of Max quantizer, to find a quantizer minimizing an average distortion, we need to evaluate a minimum distortion term as follows,

$$
\begin{aligned}
&\arg\min_{\hat{t}_M,\hat{t}_A} \ \mathbb{E}\left[D_{M,A}((T_M,T_A),(\hat{t}_M,\hat{t}_A))|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&\qquad = \arg\min_{\hat{t}_M,\hat{t}_A} \ \mathbb{E}\left[T_M - \hat{t}_M \mathbb{1}_{\hat{t}_M \le X_{\hat{t}_A}} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&\qquad = \mathbb{E}\left[T_M | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&\qquad\quad - \arg\min_{\hat{t}_M,\hat{t}_A} \ \left\{\hat{t}_M \mathbf{P}\left[\hat{t}_M \le X_{\hat{t}_A} | X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]\right\}
\end{aligned}
$$

For region $\mathbf{T}_1$, when the maximum value $T_M = \max(X_1, X_2)$ is determined by X_1, the region $\mathbf{T}_1$ can be expressed as $\mathbf{I}_{2,k_2} < \mathbf{I}_{1,k_1}$ or $l_{2,k_2} < l_{1,k_1-1}$. Because of $T_M = X_1$,

we need to determine $T_A = 1$.

$$\begin{aligned}\hat{t}_1^* &= \mathbb{E}\left[X_1 | X_1 \in \mathbf{I}_{1,k_1}\right] - \arg\min_{\hat{t}} \left\{\hat{t}\mathbf{P}\left[\hat{t} \leq X_1 | X_1 \in \mathbf{I}_{1,k_1}\right]\right\} \\ &= \arg\min_{\hat{t}} \hat{t}\mathbf{P}\left[\hat{t} \leq X_1 | X_1 \in \mathbf{I}_{1,k_1}\right]\end{aligned}$$

In order to find an optimum estimation $\hat{t}_1^*$ minimizing the average distortion in region $\mathbf{T}_1$, the necessary and sufficient condition is to find $\hat{t}$ maximizing $\hat{t}\mathbf{P}\left[\hat{t} \leq X_1 | X_1 \in \mathbf{I}_{1,k_1}\right]$.

Let's evaluate $\hat{t}\mathbf{P}\left[\hat{t} \leq X_1 | X_1 \in \mathbf{I}_{1,k_1}\right]$. Depending on the location of $\hat{t}$, we can determine the expected value as follows,

$$\hat{t}\mathbf{P}\left[X_1 > \hat{t} | X_1 \in \mathbf{I}_{1,k_1}\right] = \begin{cases} \hat{t} & \hat{t} \leq l_{1,k_1-1} \\ \hat{t}\frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t})}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} & \hat{t} \in I_{1,k_1} \\ 0 & \hat{t} \geq l_{1,k_1} \end{cases}$$

Since the maximum of each region can be included on the boundary from $\mathbf{I}_{1,k_1}$, it is suffices to evaluate as follows,

$$\hat{t}_1^* = \arg\max_{\hat{t} \in \mathbf{I}_{1,k_1}} \hat{t}\frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t})}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} \tag{5.35}$$

Based on the first and second derivative of the evaluation with respect to $\hat{t}$, the estimation $\hat{t}_1^*$ can be determined by the solution as follows,

$$\hat{t}_1^* = \begin{cases} \text{sol}\left\{\hat{t} | F_{X_1}(l_{1,k_1}) = F_{X_1}(\hat{t}) + \hat{t}f_{X_1}(\hat{t}), -2f_{X_1}(\hat{t}) - \hat{t}f'_{X_1}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{1,k_1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases} \tag{5.36}$$

In order to find an average distortion in the region $\mathbf{T}_1$, find $\mathbb{E}\left[X_1|X_1 \in \mathbf{I}_{1,k_1}\right]$ as follows

$$\mathbb{E}\left[X_1|X_1 \in \mathbf{I}_{1,k_1}\right] = \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 \frac{f_{X_1}(x_1)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} dx_1 \tag{5.37}$$

Therefore, the average minimum distortion in the region $\mathbf{T}_1$ given $X_1 \in \mathbf{I}_{1,k_1}$ and $X_2 \in \mathbf{I}_{2,k_2}$ is

$$\begin{aligned}
&\mathbb{E}[X_1|X_1 \in \mathbf{I}_{1,k_1}] - \hat{t}_1^* \mathbf{P}[X_1 > \hat{t}_1^*|X_1 \in \mathbf{I}_{1,k_1}] \\
&\quad = \int_{l_{1,k_1-1}}^{l_{1,k_1}} x_1 \frac{f_{X_1}(x_1)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})} dx_1 - \hat{t}_1^* \frac{F_{X_1}(l_{1,k_1}) - F_{X_1}(\hat{t}_1^*)}{F_{X_1}(l_{1,k_1}) - F_{X_1}(l_{1,k_1-1})}
\end{aligned}$$

For region $\mathbf{T}_2$, when the maximum value $T_M = \max(X_1, X_2)$ is determined by X_2, the region $\mathbf{T}_2$ can be expressed as $\mathbf{I}_{1,k_1} < \mathbf{I}_{2,k_2}$ or $l_{1,k_1} < l_{2,k_2-1}$. Because of $T_M = X_2$, we need to determine $T_A = 2$.

$$\begin{aligned}
\hat{t}_2^* &= \mathbb{E}\left[X_2|X_2 \in \mathbf{I}_{2,k_2}\right] - \arg\min_{\hat{t}} \hat{t}\mathbf{P}\left[X_2 > \hat{t}|X_2 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] \\
&= \arg\max_{\hat{t}} \hat{t}\mathbf{P}\left[X_2 > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]
\end{aligned}$$

Similar to the region $\mathbf{T}_1$, the estimation $\hat{t}_2^*$ can be determined by the solution as follows,

$$\hat{t}_2^* = \begin{cases} \text{sol}\left\{\hat{t}|F_{X_2}(l_{2,k_2}) = F_{X_2}(\hat{t}) + \hat{t}f_{X_2}(\hat{t}), -2f_{X_2}(\hat{t}) - \hat{t}f'_{X_2}(\hat{t}) \leq 0\right\} & \hat{t} \in \mathbf{I}_{2,k_2} \\ l_{2,k_2-1} & \text{otherwise} \end{cases} \tag{5.38}$$

Therefore, the average minimum distortion in the region $\mathbf{T}_2$ given $X_1 \in \mathbf{I}_{1,k_1}$ and

$X_2 \in \mathbf{I}_{2,k_2}$ is

$$\begin{aligned}&\mathbb{E}[X_2|X_2 \in \mathbf{I}_{2,k_2}] - \hat{t}_2^*\mathbf{P}\left[X_2 > \hat{t}_2^*|X_2 \in \mathbf{I}_{2,k_2}\right] \\ &\quad = \int_{l_{2,k_2-1}}^{l_{2,k_2}} x_2 \frac{f_{X_2}(x_2)}{F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})} dx_2 - \hat{t}_2^* \frac{F_{X_2}(l_{2,k_2}) - F_{X_2}(\hat{t}_2^*)}{F_{X_2}(l_{2,k_2}) - F_{X_2}(l_{2,k_2-1})}\end{aligned}$$

For region $\mathbf{T}_0$, when the maximum value $T_M = \max(X_1, X_2)$ is not determined by either X_1 or X_2, it suffices to compare $\hat{t}\mathbf{P}\left[T_M > \hat{t}|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right]$ under $T_A = 1$ and $T_A = 2$.

We need to find $\hat{t}_M^* = \hat{t}^*$ and $\hat{t}_A^* = \hat{i}^*$ as follows,

$$\arg\max_{\hat{t},\hat{i}} \hat{t} \frac{F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}}) - F_{X_{\hat{i}}}(\hat{t})}{F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}}) - F_{X_{\hat{i}}}(l_{\hat{i},k_{\hat{i}}-1})} \tag{5.39}$$

To get an average distortion in the region $\mathbf{t}_0$, we need to calculate following term

$$\mathbb{E}\left[T_M|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}\right] = \int_{l_{\hat{i}^*,k_{\hat{i}^*}-1}}^{l_{\hat{i}^*,k_{\hat{i}^*}}} t f_{T_M|X_1 \in \mathbf{I}_{1,k_1}, X_2 \in \mathbf{I}_{2,k_2}}(t)dt \tag{5.40}$$

■

Example : Max and Argmax Quantizer with Uniform$(0,1)$

When two users' source is Uniform$(0,1)$, an expected minimum distortion in region $(k_1, k_2) \in \mathbf{T}_1$ is

$$\begin{cases} \frac{1}{4}(l_{1,k_1}^2 - 2l_{1,k_1-1}^2)(l_{2,k_2} - l_{2,k_2-1}) & \text{If } \frac{l_{1,k_1}}{2} \geq l_{1,k_1-1} \\ \frac{1}{2}(l_{1,k_1} - l_{1,k_1-1})^2(l_{2,k_2} - l_{2,k_2-1}) & \text{otherwise} \end{cases}$$

For $(k_1, k_2) \in \mathbf{T}_2$, an average distortion is

$$\begin{cases} \frac{1}{4}(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2}^2 - 2l_{2,k_2-1}^2) & \text{If } \frac{l_{2,k_2}}{2} \geq l_{2,k_2-1} \\ \frac{1}{2}(l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1})^2 & \text{otherwise} \end{cases}$$

For $(k_1, k_2) \in \mathbf{T}_{01}$, an average distortion is

$$\begin{aligned} &[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2]\Big|_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{2,k_2}} + \frac{1}{2}(l_{2,k_2} - l_{2,k_2-1})t^2\Big|_{l_{2,k_2}}^{l_{1,k_1}} \\ &\quad - (l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(t_{11}^* \frac{l_{1,k_1} - t_{11}^*}{l_{1,k_1} - l_{1,k_1-1}}, t_{12}^* \frac{l_{2,k_2} - t_{12}^*}{l_{2,k_2} - l_{2,k_2-1}}\right) \end{aligned}$$

where t_{11}^* and t_{12}^* are

$$t_{11}^* = \begin{cases} \frac{l_{1,k_1}}{2} & \text{If } l_{1,k_1} \geq 2l_{1,k_1-1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases}, \qquad t_{12}^* = \begin{cases} \frac{l_{2,k_2}}{2} & \text{If } l_{2,k_2} \geq 2l_{2,k_2-1} \\ l_{2,k_2-1} & \text{otherwise} \end{cases}$$

For $(k_1, k_2) \in \mathbf{T}_{02}$, an average distortion is

$$\begin{aligned} &[\frac{2}{3}t^3 - \frac{1}{2}(l_{1,k_1-1} + l_{2,k_2-1})t^2]\Big|_{\max(l_{1,k_1-1}, l_{2,k_2-1})}^{l_{1,k_1}} + \frac{1}{2}(l_{2,k_2} - l_{2,k_2-1})t^2\Big|_{l_{1,k_1}}^{l_{2,k_2}} \\ &\quad - (l_{1,k_1} - l_{1,k_1-1})(l_{2,k_2} - l_{2,k_2-1}) \max\left(t_{21}^* \frac{l_{1,k_1} - t_{21}^*}{l_{1,k_1} - l_{1,k_1-1}}, t_{22}^* \frac{l_{2,k_2} - t_{22}^*}{l_{2,k_2} - l_{2,k_2-1}}\right) \end{aligned}$$

where t_{21}^* and t_{22}^* are

$$t_{21}^* = \begin{cases} \frac{l_{1,k_1}}{2} & \text{If } l_{1,k_1} \geq 2l_{1,k_1-1} \\ l_{1,k_1-1} & \text{otherwise} \end{cases}, \qquad t_{22}^* = \begin{cases} \frac{l_{2,k_2}}{2} & \text{If } l_{2,k_2} \geq 2l_{2,k_2-1} \\ l_{2,k_2-1} & \text{otherwise} \end{cases}$$

Fig. 5.4 shows the per user rate and fractional max and arg max scalar quantizer under uniform distribution $U(0, 1)$ parameterized by the solution. Here we have used

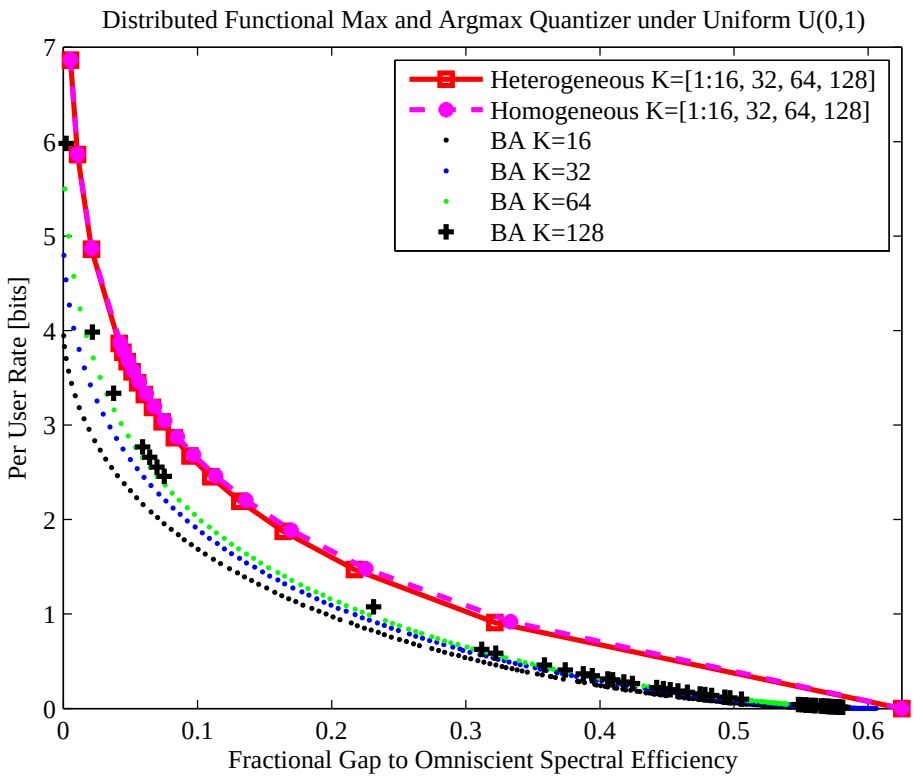


Figure 5.4: Distributed Functional max and arg max Scalar Quantizer under Uniform U(0,1)

16, 32, 64, and 128 quantization levels, and it is apparent that for non-zero distortions these are already close to the limit that they will converge to. Comparing this fundamental limit with the overhead performance tradeoff provided by the practical scalar quantization scheme, we observe that the two are pretty close, without a whole lot of room for improvement. Such improvement could be potentially obtained with a vector quantizer, but the scalar quantizer is already fairly close to the fundamental overhead performance tradeoff.

6. Conclusions and Future Works

This thesis explored the tradeoff between the amount of overhead control signaling and the performance of wireless resource allocation and link adaptation in cellular wireless networks with both an academic and industrial perspective. We began by reviewing the existing resource control signaling overhead in the LTE standard. We observed that control information regarding resource allocation and link adaptation is encoded in a direct and simple manner, and could be more efficiently encoded. Additionally, from the location of the reference signals and the physical control channels that convey most control information about resource allocation and link adaptation, we found the control signaling overhead currently occupies nearly 1/3 of the downlink time frequency resources.

Inspired by this pragmatic problem, we aimed to gain an understanding of the fundamental tradeoff, over all possible control signaling schemes, between control information overhead and the performance of a wireless resource allocation. We argued that in some situations this tradeoff could be modeled as a rate distortion function of the central estimation officer (CEO) problem, with the rate interpreted as the control overhead, and the distortion interpreted as the difference between the performance obtained with knowledge of perfect network state and the performance obtained by an imperfect controller.

In order to enable us to compute the fundamental overhead performance tradeoff for several simple physical layer inspired resource allocation models, then, we devised a method of computing rate distortion region for the CEO problem with independent sources via an novel adaptation of the Blahut-Arimoto algorithm. The resulting algorithm is globally convergent if the distortion is convex. When the distortion is not convex, we can initialize the algorithm at a global minimum for the minimum distor-

tion based a graph entropy limit for decentralized function computation, obtaining the rate distortion function global minimum under some extra conditions.

To illustrate how the CEO problem can be used to model limits for the relationship between control signaling information and the performance of a resource controller, we used the created algorithm to compute the overhead performance tradeoffs for a series of simple physical layer models for resource allocation. Finally, we derived practical control signaling schemes for these resource allocation models based on scalar quantization that approached the fundamental overhead performance tradeoff. In this manner, we illustrated how the CEO problem could be used to evaluate and design better control signaling in wireless networks.

Although the results presented in this thesis have outlined an efficient resource controller design to handle overhead-performance tradeoffs for some simple physical layer models, they can be extended with further work in a number of different directions. First of all, resource allocation models which consider higher layer performance metrics, and incorporate information about queues, could be studied within the rate distortion framework. Additionally, a better study could be made of how the quantizers and limits scale as the number of users increase. Finally, either entropy coded scalar quantization, or vector quantization could be utilized for the max and $(\max, \arg\max)$ problems, potentially yielding performance even closer to the fundamental limits.

Bibliography

[1] (2013) The 3GPP LTE website. [Online]. Available: http://www.3gpp.org/LTE

[2] (2013) The WiMAX forum website. [Online]. Available: http://www.wimaxforum.org/

[3] (2013) Specification numbering on the 3GPP LTE website. [Online]. Available: http://www.3gpp.org/specification-numbering

[4] *Revised Draft Report of the 26th 3GPP TSG RAN meeting*, 3GPP Technical Specification Group Meeting Document RP-050 003, Mar. 2005.

[5] *Proposed Study Item on Evolved UTRA and UTRAN*, 3GPP Technical Specification Group Meeting Document RP-040 461, Dec. 2004.

[6] *Report of the Joint WGs meeting on UTRA/UTRAN Long Term Evolution (Tokyo, Japan, 7-8 March 2005)*, 3GPP Technical Specification Group Meeting Document REV-05 090, May 2005.

[7] *Revised Draft Report of the Joint WGs meeting on UTRA/UTRAN Long Term Evolution (Quebec, Canada 30-31 May 2005)*, 3GPP Technical Specification Group Meeting Document REV-05 109, Sep. 2005.

[8] *Draft Report of the Joint WGs meeting on UTRA/UTRAN Long Term Evolution (Tallinn, 19-20 September 2005)*, 3GPP Technical Specification Group Meeting Document RP-050 626, Sep. 2005.

[9] *Report of the 3GPP TSG RAN Long Term Evolution Work Shop 2-3 November 2004, Toronto, Canada*, 3GPP (ETSI MCC) Technical Specification Group Meeting Document RP-040 496, Dec. 2004.

[10] *Rel-08_description_20130316 : Overview of 3GPP Release 8*, 3GPP Information Release 8, Rev. 029, Jan. 2013.

[11] *E-UTRA; Physical Channels and Modulation (Release 9)*, 3GPP Technical Specification TS 36.211, Rev. 910, Mar. 2010.

[12] *E-UTRA(N); Overall description; Stage 2 (Release 9)*, 3GPP Technical Specification TS 36.300, Rev. 9a0, Dec. 2012.

[13] (2013) The 3GPP LTE Advanced website. [Online]. Available: http://www.3gpp.org/LTE-Advanced

[14] (2013) The 3GPP LTE website : Releases. [Online]. Available: http://www.3gpp.org/Releases

[15] *E-UTRA; LTE Physical Layer - General Description (Release 8)*, 3GPP Technical Specification TS 36.201, Rev. 830, Mar. 2009.

[16] *E-UTRA; LTE Physical Layer - General Description (Release 9)*, 3GPP Technical Specification TS 36.201, Rev. 910, Mar. 2010.

[17] *E-UTRA; LTE Physical Layer - General Description (Release 10)*, 3GPP Technical Specification TS 36.201, Rev. a00, Dec. 2010.

[18] *E-UTRA; LTE Physical Layer - General Description (Release 11)*, 3GPP Technical Specification TS 36.201, Rev. b10, Dec. 2012.

[19] *E-UTRA; User Equipment (UE) radio transmission and reception (Release 10)*, 3GPP Technical Specification TS 36.101, Rev. a90, Dec. 2012.

[20] *E-UTRA; Base Station (BS) radio transmission and reception (Release 10)*, 3GPP Technical Specification TS 36.104, Rev. a90, Dec. 2012.

[21] *E-UTRA; Carrier Aggregation; Base Station (BS) radio transmission and reception (Release 10)*, 3GPP Technical Report TR 36.808, Rev. a00, Jun. 2012.

[22] *Way forward on carrier aggregation deployment scenarios and multiple timing advance*, 3GPP (NTT DOCOMO, INC.) Technical Specification Group Meeting Document RP-100 380, Mar. 2010.

[23] *Way forward on scenario combinations for CA WID*, 3GPP Technical Specification Group Meeting Document RP-100 390, Mar. 2010.

[24] *Coordinated multi-point operation for LTE physical layer aspects (Release 11)*, 3GPP Technical Report TR 36.819, Rev. b10, Dec. 2011.

[25] *Overall 3GPP System Evolution*, 3GPP (NEC) Technical Specification Group Meeting Document REV-05 022, Mar. 2005.

[26] *E-UTRA(N); Overall description; Stage 2 (Release 8)*, 3GPP Technical Specification TS 36.300, Rev. 8c0, Mar. 2010.

[27] *E-UTRA(N); Overall description; Stage 2 (Release 10)*, 3GPP Technical Specification TS 36.300, Rev. a90, Dec. 2012.

[28] *E-UTRA(N); Overall description; Stage 2 (Release 11)*, 3GPP Technical Specification TS 36.300, Rev. b40, Dec. 2012.

[29] *Architecture aspects of Home NodeB and Home eNodeB (Release 9)*, 3GPP Technical Report TR 23.830, Rev. 900, Sep. 2009.

[30] *E-UTRA; Physical Channels and Modulation (Release 8)*, 3GPP Technical Specification TS 36.211, Rev. 890, Dec. 2009.

[31] *E-UTRA; Physical Channels and Modulation (Release 10)*, 3GPP Technical Specification TS 36.211, Rev. a60, Dec. 2012.

[32] *E-UTRA; Physical Channels and Modulation (Release 11)*, 3GPP Technical Specification TS 36.211, Rev. b10, Dec. 2012.

[33] *E-UTRA; Multiplexing and channel coding (Release 8)*, 3GPP Technical Specification TS 36.212, Rev. 880, Dec. 2009.

[34] *E-UTRA; Multiplexing and channel coding (Release 9)*, 3GPP Technical Specification TS 36.212, Rev. 940, Sep. 2011.

[35] *E-UTRA; Multiplexing and channel coding (Release 10)*, 3GPP Technical Specification TS 36.212, Rev. a70, Dec. 2012.

[36] *E-UTRA; Multiplexing and channel coding (Release 11)*, 3GPP Technical Specification TS 36.212, Rev. b10, Dec. 2012.

[37] *Downlink Multiple Access for EUTRA*, 3GPP (Motorola) Technical Specification Group Meeting Document R1-050 246, Apr. 2005.

[38] *Key components of the EUTRA downlink multiple access scheme*, 3GPP (Huawei) Technical Specification Group Meeting Document R1-050 273, Apr. 2005.

[39] *Basic principles for the Evolved UTRA radio-access concept*, 3GPP Technical Specification Group Meeting Document R1-050 448, May 2005.

[40] *Approved Report of 3GPP TSG RAN WG1 #40bis in Beijing (Beijing, China, 04-08 April, 2005)*, 3GPP Technical Specification Group Meeting Document R1-050 574, May 2005.

[41] *Approved Report of 3GPP TSG RAN WG1 #41 in Athens (Athens, Greece, 09-13 May, 2005)*, 3GPP Technical Specification Group Meeting Document R1-050 995, Aug. 2005.

[42] *Approved Report (v.1.0.0) of 3GPP TSG RAN WG1 LTE Ad Hoc meeting in Helsinki (Helsinki, Finland, 23-25 January,2006)*, 3GPP Technical Specification Group Meeting Document R1-060 756, Feb. 2006.

[43] *LS on UTRAN LTE Multiple Access Selection*, 3GPP Technical Specification Group Meeting Document RP-050 758, Nov. 2005.

[44] *Revised draft report of the 30th 3GPP TSG RAN meeting (St Julian, Malta, 29 November-2 December 2005)*, 3GPP Technical Specification Group Meeting Document RP-060 002, Mar. 2006.

[45] *Downlink multiple access scheme for LTE*, 3GPP (LG Electronics) Technical Specification Group Meeting Document R1-050 274, Apr. 2005.

[46] *Uplink Considerations for UTRAN LTE*, 3GPP (Nokia) Technical Specification Group Meeting Document R1-050 251, Apr. 2005.

[47] *PAPR and Cubic Metric from a System Point of View for E-UTRA Uplink*, 3GPP (Motorola) Technical Specification Group Meeting Document R1-051 237, Oct. 2005.

[48] *Uplink Transmission Scheme for E-UTRA*, 3GPP Technical Specification Group Meeting Document R1-051 443, Nov. 2005.

[49] *Cubic metric performances of optimum spectrum-shaping functions for PAPR reduction*, 3GPP (Huawei) Technical Specification Group Meeting Document R1-060 329, Feb. 2006.

[50] *Approved Report of 3GPP TSG RAN WG1 #44 meeting*, 3GPP Technical Specification Group Meeting Document R1-061 099, Mar. 2006.

[51] G. Ku and J. Walsh, "Power amplifier nonlinearity effects on OFDM subcarrier transmit beamforming," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, Mar. 2011, pp. 1954–1959.

[52] *Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7)*, 3GPP Technical Report TR 25.814, Rev. 710, Sep. 2006.

[53] *E-UTRA; Physical layer procedures (Release 10)*, 3GPP Technical Specification TS 36.213, Rev. a80, Dec. 2012.

[54] *E-UTRA; Physical layer procedures (Release 11)*, 3GPP Technical Specification TS 36.213, Rev. b10, Dec. 2012.

[55] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Upper Saddle River, NJ: Prentice Hall, 2011.

[56] *MIMO AH Summary*, 3GPP Technical Specification Group Meeting Document R1-072 620, May 2007.

[57] *Draft Report of 3GPP TSG RAN WG1 #49 v0.4.0 (Kobe, Japan, 7-11 May, 2007)*, 3GPP Technical Specification Group Meeting Document R1-072 646, Jun. 2007.

[58] *Way Forward on 4-Tx Antenna Codebook for SU-MIMO*, 3GPP Technical Specification Group Meeting Document R1-071 799, Mar. 2007.

[59] *Text Proposal for TS 36.211 Regarding CDD Design*, 3GPP Technical Specification Group Meeting Document R1-073 182, Jun. 2007.

[60] *E-UTRA; Physical layer procedures (Release 8)*, 3GPP Technical Specification TS 36.213, Rev. 880, Sep. 2009.

[61] *E-UTRA; Physical layer procedures (Release 9)*, 3GPP Technical Specification TS 36.213, Rev. 930, Sep. 2010.

[62] *MIMO AH Summary*, 3GPP Technical Specification Group Meeting Document R1-073 225, Jun. 2007.

[63] *Support of operation in paired and unpaired spectrum*, 3GPP Technical Specification Group Meeting Document R1-050 731, Aug. 2005.

[64] *E-UTRA TDD Numerology and Frame structure*, 3GPP (Motorola) Technical Specification Group Meeting Document R1-071 853, Apr. 2007.

[65] *EUTRA Downlink Numerology*, 3GPP (Motorola) Technical Specification Group Meeting Document R1-050 520, May 2005.

[66] *TP- A modification of frame structure for CP length*, 3GPP Technical Specification Group Meeting Document R1-051 558, Nov. 2005.

[67] *E-UTRA; User Equipment (UE) radio transmission and reception (Release 8)*, 3GPP Technical Specification TS 36.101, Rev. 8k0, Dec. 2012.

[68] *E-UTRA; User Equipment (UE) radio transmission and reception (Release 9)*, 3GPP Technical Specification TS 36.101, Rev. 9e0, Dec. 2012.

[69] *E-UTRA; User Equipment (UE) radio transmission and reception (Release 11)*, 3GPP Technical Specification TS 36.101, Rev. b30, Dec. 2012.

[70] *E-UTRA; Base Station (BS) radio transmission and reception (Release 8)*, 3GPP Technical Specification TS 36.104, Rev. 8d0, Jul. 2012.

[71] *E-UTRA; Base Station (BS) radio transmission and reception (Release 9)*, 3GPP Technical Specification TS 36.104, Rev. 9d0, Sep. 2012.

[72] *E-UTRA; Base Station (BS) radio transmission and reception (Release 11)*, 3GPP Technical Specification TS 36.104, Rev. b31, Jan. 2013.

[73] *Support of half duplex FDD*, 3GPP (Vodafone Group) Technical Specification Group Meeting Document R1-070 533, Jan. 2007.

[74] *Proposed way forward on Half Duplex FDD operation in LTE*, 3GPP (NTT DoCoMo) Technical Specification Group Meeting Document R4-080 492, Feb. 2008.

[75] *Way Forward for Half Duplex*, 3GPP Technical Specification Group Meeting Document R1-080 598, Jan. 2008.

[76] *Physical-layer parameters to be configured by RRC*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-082 199, May 2008.

[77] *Frame structure design and analyze of OFDM TDD for LTE*, 3GPP Technical Specification Group Meeting Document R1-051 178, Oct. 2005.

[78] *Approved Report of 3GPP TSG RAN WG1 #42bis in San Diego (San Diego, USA, 10-14 October, 2005)*, 3GPP Technical Specification Group Meeting Document R1-051 568, Nov. 2005.

[79] *Consideration on LTE TDD Alternative Frame Structure*, 3GPP (Nokia) Technical Specification Group Meeting Document R1-063 528, Nov. 2006.

[80] *On Guard Interval of LTE TDD FS2*, 3GPP Technical Specification Group Meeting Document R1-073 851, Aug. 2007.

[81] *Draft Report of 3GPP TSG RAN WG1 #50 v0.2.0 (Athens, Greece, 20-24 August, 2007)*, 3GPP Technical Specification Group Meeting Document R1-073 896, Oct. 2007.

[82] *Way Forward on LTE TDD Frame Structure*, 3GPP Technical Specification Group Meeting Document R1-075 020, Nov. 2007.

[83] *Way forward on duration and usage of UpPTS*, 3GPP Technical Specification Group Meeting Document R1-080 535, Jan. 2008.

[84] *Way forward on Special sub-frame patterns for FS2*, 3GPP Technical Specification Group Meeting Document R1-080 602, Jan. 2008.

[85] *Way forward for TDD HARQ process*, 3GPP Technical Specification Group Meeting Document R1-081 124, Feb. 2008.

[86] *Way Forward on DwPTS design for LTE TDD*, 3GPP Technical Specification Group Meeting Document R1-081 135, Feb. 2008.

[87] *Report of 3GPP TSG RAN WG1 #52 v1.0.0 (Sorrento, Italy, 11-15 February, 2008)*, 3GPP Technical Specification Group Meeting Document R1-081 166, Mar. 2008.

[88] *Discussion on additional special subframe configuration*, 3GPP Technical Specification Group Meeting Document R1-121 712, Mar. 2012.

[89] *Final Report of 3GPP TSG RAN WG1 #68bis v1.0.0 (Jeju, South Korea, 26th-30th March, 2012)*, 3GPP Technical Specification Group Meeting Document R1-121 926, May 2012.

[90] *LS on 7.5 kHz carrier spacing*, 3GPP Technical Specification Group Meeting Document R1-070 628, Jan. 2007.

[91] *Approved Report of 3GPP TSG RAN WG1 #47bis v2.0.0 (Sorrento, Italy, 15-19 January, 2007)*, 3GPP Technical Specification Group Meeting Document R1-071 245, Feb. 2007.

[92] *On E-UTRA bandwidth option definition*, 3GPP (Ericsson, Nokia Siemens Networks) Technical Specification Group Meeting Document R4-071 107, Jun. 2007.

[93] *Discussion on TTI for E-UTRA*, 3GPP Technical Specification Group Meeting Document R1-062 036, Sep. 2006.

[94] *E-UTRA TTI, way forward*, 3GPP Technical Specification Group Meeting Document R1-062 441, Sep. 2006.

[95] *Approved minutes of 3GPP TSG RAN WG1 #46 in Tallinn (Tallinn, Estonia, 28 August-01 September, 2006)*, 3GPP Technical Specification Group Meeting Document R1-063 013, Oct. 2006.

[96] *Further Consideration on the Size of Physical Resource Block in EUTRA*, 3GPP (ZTE) Technical Specification Group Meeting Document R1-062 206, Sep. 2006.

[97] *Resource-block size*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-062 270, Sep. 2006.

[98] *Issues for FDD/TDD aligned Numerology*, 3GPP (Qualcomm Europe) Technical Specification Group Meeting Document R1-062 687, Oct. 2006.

[99] *Approved Report of 3GPP TSG RAN WG1 #47 (Riga, Latvia, 06-10 November, 2006)*, 3GPP Technical Specification Group Meeting Document R1-070 633, Jan. 2007.

[100] *Downlink OFDM Numerology*, 3GPP (Texas Instruments) Technical Specification Group Meeting Document R1-050 667, Jun. 2005.

[101] *Views on OFDM Parameter Set for Evolved UTRA Downlink*, 3GPP Technical Specification Group Meeting Document R1-050 386, May 2005.

[102] *EUTRA Downlink Numerology and Design*, 3GPP (Motorola) Technical Specification Group Meeting Document R1-050 583, Jun. 2005.

[103] *On E-UTRA DL numerology*, 3GPP (Nortel Networks) Technical Specification Group Meeting Document R1-050 645, Jun. 2005.

[104] *Bandwidth of resource blocks for DL OFDMA*, 3GPP (Siemens) Technical Specification Group Meeting Document R1-050 737, Aug. 2005.

[105] *Approved Report of 3GPP TSG RAN WG1 Ad Hoc on LTE in Sophia Antipolis (Sophia Antipolis, France, 20-21 June, 2005)*, 3GPP Technical Specification Group Meeting Document R1-050 996, Aug. 2005.

[106] *Consideration on Resource Block Size*, 3GPP Technical Specification Group Meeting Document R1-061 667, Jun. 2006.

[107] T. S. Rappaport, *Wireless communications : principles and practice*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[108] *E-UTRA; Radio Resource Control (RRC); Protocol specification (Release 8)*, 3GPP Technical Specification TS 36.331, Rev. 8i0, Dec. 2012.

[109] *E-UTRA; Radio Resource Control (RRC); Protocol specification (Release 9)*, 3GPP Technical Specification TS 36.331, Rev. 9d0, Dec. 2012.

[110] *E-UTRA; Radio Resource Control (RRC); Protocol specification (Release 10)*, 3GPP Technical Specification TS 36.331, Rev. a80, Dec. 2012.

[111] *E-UTRA; Radio Resource Control (RRC); Protocol specification (Release 11)*, 3GPP Technical Specification TS 36.331, Rev. b20, Dec. 2012.

[112] *Physical-layer parameters to be configured by RRC*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-080 621, Jan. 2008.

[113] *Outcome of Offline Discussion on DL Control Signaling*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-074 499, Oct. 2007.

[114] *Summary of email discussion on DL control signaling*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-080 869, Feb. 2008.

[115] *Summary of email discussion on DL control signaling*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-081 522, Mar. 2008.

[116] *On enhanced downlink control signalling for Rel-11*, 3GPP Technical Specification Group Meeting Document R1-112 530, Aug. 2011.

[117] *Application scenario for introducing enhanced PDCCH in Rel-11*, 3GPP Technical Specification Group Meeting Document R1-112 592, Aug. 2011.

[118] *Final Report of 3GPP TSG RAN WG1 #66 v1.0.0*, 3GPP Technical Specification Group Meeting Document R1-112 886, Oct. 2011.

[119] *Motivations and scenarios for ePDCCH*, 3GPP (Nokia) Technical Specification Group Meeting Document R1-113 155, Oct. 2011.

[120] *On the Structure and Usage Scenarios of ePDCCH*, 3GPP (Apple Inc.) Technical Specification Group Meeting Document R1-114 300, Nov. 2011.

[121] *LS on UE ID for resource assignment*, 3GPP Technical Specification Group Meeting Document R1-062 446, Sep. 2006.

[122] *Downlink resource allocation for E-UTRA*, 3GPP (LG Electronics) Technical Specification Group Meeting Document R1-051 049, Oct. 2005.

[123] *E-URTA Resource Allocation Design for Downlink in LTE*, 3GPP (ZTE) Technical Specification Group Meeting Document R1-060 251, Jan. 2006.

[124] *Requirements for Efficient OFDM Downlink Scheduling*, 3GPP (Alcatel) Technical Specification Group Meeting Document R2-060 003, Jan. 2006.

[125] *Adaptive UE Feedback Channel for Downlink Scheduling Decision in LTE*, 3GPP (Alcatel) Technical Specification Group Meeting Document R2-060 004, Jan. 2006.

[126] *DL Unicast Resource Allocation Signalling using L1L2 control channels*, 3GPP (NEC Group) Technical Specification Group Meeting Document R1-072 832, Jun. 2007.

[127] *Resource Allocation type 1 definition details*, 3GPP Technical Specification Group Meeting Document R1-081 140, Feb. 2008.

[128] *Uplink Resource Allocation for E-UTRA*, 3GPP (NEC Group, NTT DoCoMo) Technical Specification Group Meeting Document R1-070 881, Feb. 2007.

[129] *E-UTRA; Radio Frequency (RF) system scenarios (Release 8)*, 3GPP Technical Report TR 36.942, Rev. 840, Jun. 2012.

[130] *Link Adaptation Method in Evolved UTRA Uplink*, 3GPP (NTT DoCoMo) Technical Specification Group Meeting Document R1-051 392, Nov. 2005.

[131] *Link Adaptation for E-UTRA Downlink*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-060 101, Jan. 2006.

[132] *AMC and HARQ Using Frequency Domain Channel-dependent Scheduling in MIMO Channel Transmission*, 3GPP Technical Specification Group Meeting Document R1-050 942, Aug. 2005.

[133] *Adaptive Modulation and Channel Coding Rate Control for Frequency Domain Scheduling in Evolved UTRA Downlink*, 3GPP Technical Specification Group Meeting Document R1-051 149, Oct. 2005.

[134] *Adaptive Modulation and Channel Coding Rate Control for Single-antenna Transmission in Frequency Domain Scheduling in E-UTRA Downlink*, 3GPP Technical Specification Group Meeting Document R1-060 039, Jan. 2006.

[135] *Rate matching and interleaver for HARQ*, 3GPP (Samsung) Technical Specification Group Meeting Document R1-071 578, Mar. 2007.

[136] *Way forward on HARQ rate matching for LTE*, 3GPP Technical Specification Group Meeting Document R1-072 604, May 2007.

[137] *On the number of hybrid ARQ processes in LTE*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-073 027, Jun. 2007.

[138] *Summary of the E-mail discussion on channel coding*, 3GPP (NTT DoCoMo) Technical Specification Group Meeting Document R1-070 752, Feb. 2007.

[139] *Approved Report of 3GPP TSG RAN WG1 #48 v0.2.0 (St. Louis, USA, 12-16 February, 2007)*, 3GPP Technical Specification Group Meeting Document R1-071 768, Mar. 2007.

[140] *Text change to code block segmentation*, 3GPP Technical Specification Group Meeting Document R1-073 863, Aug. 2007.

[141] *Way Forward on LTE Rate Matching*, 3GPP (Nokia Siemens Networks, Nokia) Technical Specification Group Meeting Document R1-072 273, May 2007.

[142] *Maximum number of hybrid ARQ processes*, 3GPP Technical Specification Group Meeting Document R1-073 871, Aug. 2007.

[143] *E-UTRA; Medium Access Control (MAC) protocol specification (Release 8)*, 3GPP Technical Specification TS 36.321, Rev. 8c0, Mar. 2012.

[144] *E-UTRA; Medium Access Control (MAC) protocol specification (Release 9)*, 3GPP Technical Specification TS 36.321, Rev. 960, Mar. 2012.

[145] *E-UTRA; Medium Access Control (MAC) protocol specification (Release 10)*, 3GPP Technical Specification TS 36.321, Rev. a70, Dec. 2012.

[146] *E-UTRA; Medium Access Control (MAC) protocol specification (Release 11)*, 3GPP Technical Specification TS 36.321, Rev. b10, Dec. 2012.

[147] *Hybrid ARQ Scheme for E-UTRA Downlink*, 3GPP Technical Specification Group Meeting Document R2-060 840, Mar. 2006.

[148] *HARQ operation*, 3GPP (Samsung) Technical Specification Group Meeting Document R2-061 000, Mar. 2006.

[149] *HARQ Conclusions*, 3GPP Technical Specification Group Meeting Document R2-061 078, Mar. 2006.

[150] *Hybrid ARQ Scheme in E-UTRA Downlink*, 3GPP (NTT DoCoMo) Technical Specification Group Meeting Document R1-060 211, Jan. 2006.

[151] *Synchronous vs. Asynchronous HARQ*, 3GPP (Ericsson) Technical Specification Group Meeting Document R2-060 869, Mar. 2006.

[152] *Proposed way forward for CQI Feedback Control and Content in E-UTRA*, 3GPP Technical Specification Group Meeting Document R1-072 590, May 2007.

[153] *Summary of AH session on AI 6.4.5 UE Procedures for downlink shared channel*, 3GPP Technical Specification Group Meeting Document R1-075 069, Nov. 2007.

[154] *Draft Report of 3GPP TSG RAN WG1 #51 v1.0.0 (Jeju, South Korea, 5-9 November, 2007)*, 3GPP Technical Specification Group Meeting Document R1-080 002, Jan. 2008.

[155] *LS on CQI definition*, 3GPP Technical Specification Group Meeting Document R1-082 632, Jun. 2008.

[156] *Summary of email discussion on CQI reporting*, 3GPP (Samsung) Technical Specification Group Meeting Document R1-080 049, Jan. 2008.

[157] *Reply LS on CQI feedback*, 3GPP Technical Specification Group Meeting Document R1-072 639, May 2007.

[158] *CQI reporting principles to support LTE scheduling*, 3GPP (Alcatel, Samsung) Technical Specification Group Meeting Document R2-061 908, Jun. 2006.

[159] *Overview on CQI reporting*, 3GPP (NTT DoCoMo, Inc.) Technical Specification Group Meeting Document R2-061 983, Jun. 2006.

[160] *Considerations on CQI reference period*, 3GPP (Nokia, Nokia Siemens Networks) Technical Specification Group Meeting Document R4-081 123, May 2008.

[161] *Summary of CQI Ad Hoc*, 3GPP Technical Specification Group Meeting Document R1-074 501, Oct. 2007.

[162] *CQI Reporting Procedure for E-UTRA*, 3GPP Technical Specification Group Meeting Document R1-080 207, Jan. 2008.

[163] *All Sub-band Differential CQI Feedback for E-UTRA*, 3GPP (Texas Instruments) Technical Specification Group Meeting Document R1-074 694, Nov. 2007.

[164] *Summary of email discussion on CQI reporting*, 3GPP (Samsung) Technical Specification Group Meeting Document R1-081 234, Mar. 2008.

[165] *Summary of e-mail discussion on CQI reports*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-073 054, Jun. 2007.

[166] *Physical Layer Parameters for CQI Reporting*, 3GPP (Texas Instruments) Technical Specification Group Meeting Document R1-081 693, Mar. 2008.

[167] *Final Report of 3GPP TSG RAN WG1 #52bis v1.0.0 (Shenzhen, China, 31 March-4 April, 2008)*, 3GPP Technical Specification Group Meeting Document R1-081 711, May 2008.

[168] *E-UTRA; Physical layer - Measurements (Release 8)*, 3GPP Technical Specification TS 36.214, Rev. 870, Sep. 2009.

[169] *E-UTRA; Physical layer - Measurements (Release 9)*, 3GPP Technical Specification TS 36.214, Rev. 920, Jun. 2010.

[170] *E-UTRA; Physical layer - Measurements (Release 10)*, 3GPP Technical Specification TS 36.214, Rev. a10, Mar. 2011.

[171] *E-UTRA; Physical layer - Measurements (Release 11)*, 3GPP Technical Specification TS 36.214, Rev. b10, Dec. 2012.

[172] *Summary of Power Control Email Discussion*, 3GPP (Nokia Siemens Networks) Technical Specification Group Meeting Document R1-072 973, Jun. 2007.

[173] *LS on E-UTRA UL Power Control*, 3GPP Technical Specification Group Meeting Document R1-080 616, Jan. 2008.

[174] *Response LS on E-UTRA UL Power Control*, 3GPP Technical Specification Group Meeting Document R1-090 008, Jan. 2009.

[175] *PUSCH Power Control for Carrier Aggregation*, 3GPP Technical Specification Group Meeting Document R1-093 770, Oct. 2009.

[176] *Way Forward on Power Control of PUSCH*, 3GPP Technical Specification Group Meeting Document R1-073 224, Jun. 2007.

[177] *Summary of Power Control E-mail Discussion*, 3GPP (Nokia Siemens Networks) Technical Specification Group Meeting Document R1-073 677, Aug. 2007.

[178] *Uplink Power Control for E-UTRA - Comments on Open Issues*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-074 378, Oct. 2007.

[179] *LS on UE specific power control parameter*, 3GPP Technical Specification Group Meeting Document R1-073 888, Aug. 2007.

[180] *Signaling of Po in PUSCH Power Control Rule*, 3GPP (Alcatel-Lucent) Technical Specification Group Meeting Document R1-074 994, Nov. 2007.

[181] *Range and representation of delta TF PUCCH*, 3GPP (Ericsson) Technical Specification Group Meeting Document R1-082 581, Jun. 2008.

[182] *Delta tf values for PUCCH PC formula*, 3GPP (Nokia Siemens Networks, Nokia Corporation) Technical Specification Group Meeting Document R1-082 605, Jun. 2008.

[183] *Delta value for PUCCH power control*, 3GPP (Huawei) Technical Specification Group Meeting Document R1-083 046, Aug. 2008.

[184] *Power control headroom reports for EUTRAN uplink*, 3GPP (Nokia Siemens Networks, Nokia) Technical Specification Group Meeting Document R1-074 348, Oct. 2007.

[185] *Way forward for power headroom reporting*, 3GPP Technical Specification Group Meeting Document R1-081 701, Mar. 2008.

[186] *CQI-based Transmission Power Control for Control Channel in Evolved UTRA Uplink*, 3GPP Technical Specification Group Meeting Document R1-051 145, Oct. 2005.

[187] *Way-forward on Data Power Setting for PDSCH across OFDM Symbols*, 3GPP Technical Specification Group Meeting Document R1-075 077, Nov. 2007.

[188] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO Problem," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[189] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.

[190] Yasutada Oohama, "Rate-distortion theory for Gaussian multiterminal source coding system with several side informations at the decoder," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.

[191] V. Prabhakaran and D. Tse and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," *Proceedings of International Symposium on Information Theory*, p. 117, June 2004.

[192] I. Csiszár and J. K. örner, *Information Theory: Coding Theorems for Discrete Memoryless Sources, 2nd Ed.* Cambridge University Press, 2011.

[193] A. E. Gamal and Y.-H. Kim, *Network Information Theory.* The Edinburgh Building, Cambridge CB2 8RU, UK: Cambridge University Press, 2011.

[194] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, January 1976.

[195] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, no. 3, pp. 471–480, March 1973.

[196] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc, 2006.

[197] P. Boukris, "An upper bound on the speed of convergence of the blahut algorithm for computing rate-distortion functions (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 708–709, 1973.

[198] I. Csiszar, "On the computation of rate-distortion functions (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 122–124, 1974.

[199] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *In Statistics & Decisions*, no. 1, pp. 205–237, 1984.

[200] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[201] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[202] V. Doshi, D. Shah, M. Médard, and S. Jaggi, "Graph coloring and conditional graph entropy," in *40th Asilomar Conf. on Signals, Systems, and Computers*, 2006, pp. 2137–2141.

[203] V. Doshi, D. Shah, M. Médard, and M. Effros, "Functional compression through graph coloring," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3901–3917, August 2010.

[204] J. Ren, G. Ku, B. D. Boyle, S. Weber, and J. M. Walsh, "Overhead Performance Tradeoffs – A Resource Allocation Perspective," *IEEE Transactions on Information Theory*, 2014, to be submitted.

[205] M. Padberg, *Linear Optimization and Extensions, 2nd Ed.* Springer, 1999.

[206] Barber, C.B., Dobkin, D.P., and Huhdanpaa, H.T., "The quickhull algorithm for convex hulls," *ACM Trans. on Mathematical Software*, vol. 22, no. 4, pp. 469–483, dec 1996. [Online]. Available: http://www.qhull.org

[207] J. M. Walsh, "Drexel university adaptive signal processing and information theory research group software repository." [Online]. Available: http://www.ece.drexel.edu/walsh/aspitrg/software.html

[208] S. Verdu and S. Shamai, "Variable-rate channel capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2651–2667, June 2010.

[209] M. Luby, "Lt codes," in *The 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002.

[210] U. Erez, M. D. Trott, and G. W. Wornell, "Rateless coding for gaussian channels," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 530–547, February 2012.

[211] B. D. Boyle, J. M. Walsh, and S. Weber, "Distributed scalar quantizers for subband allocation," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, March 2014, pp. 1–6.