Quantize-and-Forward Schemes for the Orthogonal Multiple-Access Relay Channel

Georg Zeitler, *Student Member, IEEE*, Gerhard Bauch, *Senior Member, IEEE*, Jörg Widmer, *Senior Member, IEEE*

***Abstract*—The multiple-access relay channel with two sources, a single relay, and one destination is considered. Under the assumption of noisy source–relay links causing the relay to be unable to decode without error, we propose a framework for designing one- and two-dimensional quantizers for quantizing the soft information at the relay. These quantizers are mutual-information preserving. Simulation results show a) that mutual-information preserving quantization schemes outperform techniques in which the soft information is forwarded in an analog fashion to the destination, b) that two-dimensional quantization outperforms one-dimensional quantization for source–relay links of different quality, and c) that diversity order of two can be gained in block Rayleigh fading channels by having the relay adaptively select a two-dimensional quantizer from a fixed set of quantizers shared with the destination, depending on the channel state on the source–relay links.**

***Index Terms*—Quantization, cooperative systems, relays, iterative methods**

I. Introduction

DIVERSITY techniques have been widely studied as an effective means to combat multipath fading effects inherent in wireless communication channels. Multiple transmit and/or receive antennas can often provide a form of spatial diversity whenever the application of much simpler time or frequency diversity techniques is precluded due to delay or bandwidth constraints. However, due to size limitations on the mobile devices of, e.g., a cellular communication network, the placement of multiple antennas at such mobile terminals is not always a feasible option. Cooperative diversity, first proposed in [1], [2], introduces spatial diversity without interfering with the size limitations of the terminals by allowing nodes to cooperate in facilitating their transmissions. In some cases, cooperation is achieved by employing a relay node whose sole purpose is to facilitate the transmissions of other nodes. Besides the gain in reliability, relays are also envisioned to provide coverage extension for cell-edge users of cellular networks [3] at reasonable cost. However, under orthogonal transmission – an assumption often made in practice –, the extra resources allocated to the relay result in a loss of spectral efficiency, a loss which can be reduced by allowing several users to share one relay for joint processing. We therefore focus on the orthogonal multiple-access relay channel (MARC) [4] in the following, where two sources transmit independent information to a common destination via a single relay.

In related work, diversity achieving schemes are proposed for distributed antenna systems [5], and for the MARC using low-density parity-check codes [6] and distributed turbo codes [7], combined with network coding [8] by performing joint network channel coding at the relay. Since the relay performs some form of (joint) re-encoding of the source messages in such a decode-and-forward scheme [9], [10], it is common to that work that the relay node is required to decode the source messages perfectly. However, even if the relay fails at recovering the source messages fully, the information available at the relay can still be beneficial for decoding at the destination if forwarded properly. As such, amplify-and-forward [10] or soft-decode-and-forward schemes [11] avoid error propagation effects occurring if residual bit errors remain at the relay after a hard decision about the information sequence.

In more recent work [12], the authors combine soft decoding and analog forwarding of beliefs from the relay with network coding in the MARC, in that they form and transmit the beliefs about the network coded code bits of both users at the relay, thereby achieving notable gains in symmetric additive white Gaussian noise (AWGN) channels. However, the block of beliefs about the network coded code bits is sent to the destination in an analog manner; furthermore, forming the beliefs about the network coded code bits turns out to be suboptimal especially in situations where the source–relay channels are of different quality.

The contributions of this paper are two-fold. Firstly, building on the system description in [12], we propose a framework for designing mutual-information preserving one-dimensional quantizers for the soft information at the relay that take into account the rate constraint on the relay–destination link by formulating the optimization problem as a tradeoff between quantization rate and obtained mutual information. Quantization of log-likelihood ratios was also considered in [13] under the assumption of them being conditionally Gaussian distributed. Secondly, noting that forming the beliefs about the network coded bits is particularly disadvantageous if the soft information of the two users at the relay has different reliability, the proposed framework is extended to the design of two-dimensional quantizers operating directly on the soft

Georg Zeitler is with the Institute for Communications Engineering, Technische Universität München, D-80333 München, Germany (e-mail: georg.zeitler@tum.de).

Gerhard Bauch is with the Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, D-85577 Neubiberg, Germany, and with DOCOMO Communications Laboratories Europe GmbH, Landsbergerstrasse 312, D-80687 München, Germany (e-mail: gerhard.bauch@unibw.de).

Jörg Widmer is with Institute IMDEA Networks, Avenida del Mar Mediterraneo, 22, 28918 Leganes (Madrid), Spain (e-mail: joerg.widmer@imdea.org). This work was carried out while Jörg Widmer was with DOCOMO Communications Laboratories Europe GmbH.

The material in this paper was partly presented at the International Symposium on Turbo Codes & Related Topics, Lausanne, Switzerland, September 2008, at the IEEE International Conference on Communications, Dresden, Germany, June 2009, and at the Personal, Indoor and Mobile Radio Communications Symposium, Tokyo, Japan, September 2009.

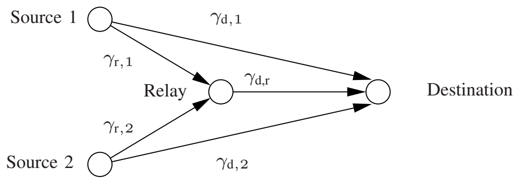


Fig. 1. The multiple-access relay channel. The relay only knows $\gamma_{r,1}$ and $\gamma_{r,2}$.

information of both users, without going through the intermediate step of computing the likelihoods of the network coded message. In doing so, the available rate on the relay–destination link is appropriately divided among the two users, according to the quality of the soft information that is at hand for each user at the relay. Finally, by employing quantization at the relay, digital transmission with its well-known advantages can be leveraged on the relay–destination link.

The paper is structured as follows. The system model is introduced in Section II, based on which we design mutual-information preserving quantizers in Section III. Numerical results for various channel models are shown in Section IV, before we end with concluding remarks in Section V. Throughout, vectors are written in bold font, while scalars appear in normal font. Random variables are printed in upper case letters, and their realizations appear in lower case characters. The entropy of a random variable X is $H(X)$, mutual information between two random variables X and Y is written as $I(X;Y)$, and expectation is denoted by $\mathrm{E}[\cdot]$. The term $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

II. System Model

A. Sources

At each source $i \in \{1,2\}$, a block of independent information bits $\boldsymbol{U}_i \in \{0,1\}^{k_i}$ is encoded with a channel code of rate k_i/n_i to a block of code bits $\boldsymbol{X}_i \in \{0,1\}^{n_i}$, which is then modulated to the channel symbols $\boldsymbol{S}_i \in \mathbb{M}_i^{m_i}$, where $\mathbb{M}_i$ is the modulation alphabet of size M_i at the i-th source. In the rest of the paper, we assume that the number of code bits satisfies $n = n_1 = n_2$.

B. Channel Model

The channel model is shown in Fig. 1. In our model, the transmissions from the sources and the relay are assumed to be orthogonalized either in frequency or in time. Despite the suboptimality of this constraint, the restriction to orthogonal channels eases practical implementation. Note that the restriction to orthogonal channels also includes a half-duplex constraint often imposed on the relay for implementation reasons, so that the relay cannot transmit and receive simultaneously in the same frequency band. Without loss of generality, we assume the orthogonality to be guaranteed by time division; consequently, a first slot is assigned to source 1, a second slot to source 2, and a third slot to the relay. Let $\boldsymbol{S}_\mathrm{r} \in \mathbb{M}_\mathrm{r}^{m_\mathrm{r}}$ be the transmitted vector from the modulation alphabet $\mathbb{M}_\mathrm{r}$ at the relay. For a path-loss coefficient p and distances $d_{\mathrm{r},i}$, $d_{\mathrm{d},i}$, and $d_{\mathrm{d,r}}$ between the terminals, the received signals at the relay and at the destination read

$$\boldsymbol{Y}_i^{(\mathrm{t})} = \frac{h_{\mathrm{t},i}}{\sqrt{d_{\mathrm{t},i}}^{p}} \boldsymbol{S}_i + \boldsymbol{N}_i^{(\mathrm{t})} \quad \mathrm{t} \in \{\mathrm{r},\mathrm{d}\}, i \in \{1,2\} \tag{1}$$

$$\boldsymbol{Y}_\mathrm{r}^{(\mathrm{d})} = \frac{h_{\mathrm{d,r}}}{\sqrt{d_{\mathrm{d,r}}}^{p}} \boldsymbol{S}_\mathrm{r} + \boldsymbol{N}_\mathrm{r}^{(\mathrm{d})}, \tag{2}$$

where $h_{\mathrm{r},i}$, $h_{\mathrm{d},i}$, and $h_{\mathrm{d,r}}$ are the complex channel fading coefficients satisfying $\mathrm{E}[|h_{\mathrm{r},i}|^2] = \mathrm{E}[|h_{\mathrm{d},i}|^2] = \mathrm{E}[|h_{\mathrm{d,r}}|^2] = 1$, and the additive noise variables are independent circularly symmetric complex Gaussian random variables with zero mean and variance normalized to unity. The average values of the signal-to-noise ratio (SNR) are given as $\rho_{\mathrm{r},i} = P_i/d_{\mathrm{r},i}^p$, $\rho_{\mathrm{d},i} = P_i/d_{\mathrm{d},i}^p$, and $\rho_{\mathrm{d,r}} = P_\mathrm{r}/d_{\mathrm{d,r}}^p$, where P_i and P_r are the powers of the sources and the relay. Throughout, we make the common assumption that the receivers know the instantaneous SNR values $\gamma_{\mathrm{r},i} = |h_{\mathrm{r},i}|^2\rho_{\mathrm{r},i}$, $\gamma_{\mathrm{d},i} = |h_{\mathrm{d},i}|^2\rho_{\mathrm{d},i}$, and $\gamma_{\mathrm{d,r}} = |h_{\mathrm{d,r}}|^2\rho_{\mathrm{d,r}}$ of their channels, and that the transmitters only possess knowledge about the average SNR. In particular, the relay is assumed to lack both average and instantaneous channel state information (CSI) of the source–destination channels due to their fading nature and limited signaling from the destination to the relay.

C. Relay Operations

The operations at the relay considered in this work are restricted to methods generating and transforming soft information about the coded bits of each user for transmission to the destination.

1) Generation of soft information: Upon reception of $\boldsymbol{y}_1^{(\mathrm{r})}$ and $\boldsymbol{y}_2^{(\mathrm{r})}$, the relay's first option is to invoke soft demappers to compute log-likelihood ratios (LLRs) $\boldsymbol{\ell}_i = \boldsymbol{\ell}_i^{(\mathrm{dem})} \in \mathbb{R}^n$, $i = 1,2$, about the coded bits, where, for $m = 1,2,\ldots,n$,

$$\ell_{i,m}^{(\mathrm{dem})} = \ln \frac{p\left(x_{i,m} = 0 | y_{i,j}^{(\mathrm{r})}\right)}{p\left(x_{i,m} = 1 | y_{i,j}^{(\mathrm{r})}\right)}, \quad j = \lceil m/\log_2(M_i) \rceil. \tag{3}$$

Alternatively, the relay performs soft decoding to calculate $\boldsymbol{\ell}_i = \boldsymbol{\ell}_i^{(\mathrm{dec})} \in \mathbb{R}^n$, $i = 1,2$, where

$$\ell_{i,m}^{(\mathrm{dec})} = \ln \frac{p\left(x_{i,m} = 0 | \boldsymbol{y}_i^{(\mathrm{r})}\right)}{p\left(x_{i,m} = 1 | \boldsymbol{y}_i^{(\mathrm{r})}\right)}, \quad m = 1,2,\ldots,n. \tag{4}$$

2) Processing of soft information: The first strategy for processing the soft information $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$ is the one of [12], where the relay computes soft information about the network coded code bits based on $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$. Specifically, the relay first interleaves $\boldsymbol{\ell}_2$ to avoid short cycles in the factor graph associated with the iterative decoder introduced in Section II-D, yielding the block $\boldsymbol{\ell}_2' \in \mathbb{R}^n$ carrying soft information about $\boldsymbol{x}_2'$ (the interleaved version of $\boldsymbol{x}_2$). Then, the relay forms the soft information $\boldsymbol{\ell} \in \mathbb{R}^n$ about $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2'$, where [14]

$$\begin{aligned}\ell_m = \ell_{1,m} \boxplus \ell_{2,m}' &\triangleq \ln\left(\frac{1 + e^{\ell_{1,m} + \ell_{2,m}'}}{e^{\ell_{1,m}} + e^{\ell_{2,m}'}}\right) \\ &\approx \mathrm{sign}(\ell_{1,m})\mathrm{sign}(\ell_{2,m}') \min\left\{|\ell_{1,m}|, |\ell_{2,m}'|\right\}.\end{aligned} \tag{5}$$

The second strategy combines the computation of soft information about the network coded code bits from [12] with scalar deterministic quantization of $\boldsymbol{\ell}$. More formally, the relay employs a quantizer with quantization rule $q(\ell)$, $q : \mathbb{R} \to \mathcal{Z}$, yielding the compressed version $\boldsymbol{z} \in \mathcal{Z}^n$, where $\mathcal{Z} = \{1, 2, \ldots, N\}$ refers to the quantizer index set. Since the quantizer is invariant for the entire block, the m-th component z_m of $\boldsymbol{z}$ is given by $z_m = q(\ell_m)$, $m = 1, 2, \ldots, n$. Transforming the quantization rule into a probability mass function $p(z|\ell) = \mathbb{1}_{\{q(\ell)=z\}}$, we have the mass function $p(z|x)$ associated with the quantization given as

$$p(z|x) = \int p(z|\ell) f(\ell|x) \mathrm{d}\ell, \qquad (6)$$

where $f(\ell|x)$ is the density of the soft information ℓ conditioned on x, which is assumed to be known or obtained from measurement (cf. Section III).

By investigating (5), we note that the reliability of the soft information $\boldsymbol{\ell}$ is dominated by $\min\left\{|\ell_{1,m}|, |\ell'_{2,m}|\right\}$, so that $|\ell_m|$ is limited by the weaker user at the relay. Such a scenario occurs, e.g., if the source–relay links have different SNR. Therefore, in the third proposed strategy for processing soft information at the relay, the XOR computation is omitted. Instead, the relay performs deterministic two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}'_2$, which is described by the quantization rule $q(\ell_1, \ell_2)$, $q : \mathbb{R}^2 \to \mathcal{Z}$, where again, $\mathcal{Z}$ is the index set of the quantizer. As before, the quantizer is assumed invariant for the entire block, so that the m-th element of $\boldsymbol{z} \in \mathcal{Z}^n$ is given by $z_m = q(\ell_{1,m}, \ell'_{2,m})$, $m = 1, 2, \ldots, n$. Defining $p(z|\ell_1, \ell_2) = \mathbb{1}_{\{q(\ell_1,\ell_2)=z\}}$ and writing $f(\ell_1, \ell_2|x_1, x_2)$ for the density of (ℓ_1, ℓ_2) conditioned on (x_1, x_2), the mass function $p(z|x_1, x_2)$ is obtained as

$$p(z|x_1, x_2) = \iint p(z|\ell_1, \ell_2) f(\ell_1, \ell_2|x_1, x_2) \mathrm{d}\ell_1 \mathrm{d}\ell_2, \qquad (7)$$

where the density $f(\ell_1, \ell_2|x_1, x_2)$ is known or approximated by measurement.

3) Transmission from the relay: Utilizing a quantizer at the relay does not necessarily result in equiprobable quantization indices $\{1, 2, \ldots, N\}$ of the quantizer index set $\mathcal{Z}$. Hence, the sequence $\boldsymbol{z}$ needs to be source encoded, yielding the block of bits $\boldsymbol{u}_\mathrm{r} \in \{0, 1\}^{k_\mathrm{r}}$, which is then channel encoded using a channel code of rate $R_\mathrm{r} = k_\mathrm{r}/n_\mathrm{r}$ to the code bits $\boldsymbol{x}_\mathrm{r} \in \{0, 1\}^{n_\mathrm{r}}$ before modulation to the symbols $\boldsymbol{s}_\mathrm{r} \in \mathbb{M}_\mathrm{r}^{m_\mathrm{r}}$.

D. Destination

The destination uses the iterative receiver [12] shown in Fig. 2. It contains two soft-in/soft-out (SISO) decoders using the received words $\boldsymbol{y}_1^{(\mathrm{d})}$ and $\boldsymbol{y}_2^{(\mathrm{d})}$ from the direct links. Furthermore, since the code bits of the two users are coupled by the joint processing at the relay, these two SISO decoders are connected by *relay check nodes* using $\boldsymbol{y}_\mathrm{r}^{(\mathrm{d})}$ from the relay, drawn as gray squares in Fig. 2. The relay check nodes allow exchange of soft information between the component SISO decoders, so that the overall decoder resembles a turbo decoder, in contrast to which, however, *code bits* of two *independent* sources are coupled.

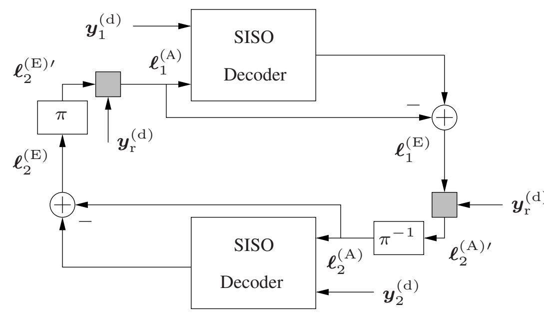


Fig. 2. Iterative decoder.

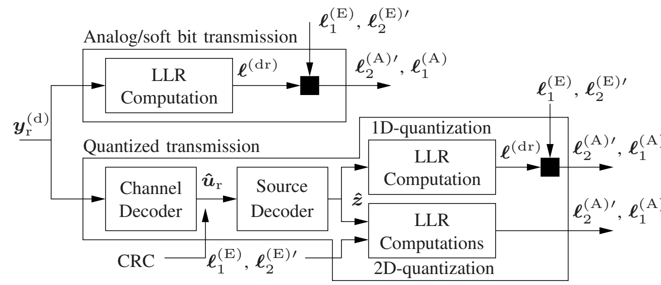


Fig. 3. Relay check node for analog, soft bit, and quantized transmission from the relay.

The operations of the relay check nodes of course depend on how the soft information at the relay is processed and transmitted. Fig. 3 shows a summary, where the black square is a standard check node whose function in the factor graph is the indicator function $\mathbb{1}_{\{x_m = x_{1,m} \oplus x'_{2,m}\}}$, so that we have $\boldsymbol{\ell}_2^{(\mathrm{A})\prime} = \boldsymbol{\ell}_1^{(\mathrm{E})} \boxplus \boldsymbol{\ell}^{(\mathrm{dr})}$ and $\boldsymbol{\ell}_1^{(\mathrm{A})} = \boldsymbol{\ell}_2^{(\mathrm{E})\prime} \boxplus \boldsymbol{\ell}^{(\mathrm{dr})}$. In case of scalar quantization at the relay, the destination first needs to recover an estimate $\hat{\boldsymbol{z}}$ of the quantizer output at the relay depending on the success of decoding $\boldsymbol{u}_\mathrm{r}$. At this point, we presume the existence of a cyclic redundancy check (CRC) in $\boldsymbol{u}_\mathrm{r}$, which we assume to be perfect in the sequel. Then, in order to avoid catastrophic error propagation through the source decoder in case of residual errors in $\hat{\boldsymbol{u}}_\mathrm{r}$, the entire transmission from the relay is discarded in that case, so that there is no exchange of soft information between the component decoders. Otherwise, we can assume the source decoder output $\hat{\boldsymbol{z}}$ to correspond to the quantization indices obtained at the relay node, i.e., $\hat{\boldsymbol{z}} = \boldsymbol{z}$. For one-dimensional quantization of $\boldsymbol{\ell}$ at the relay, the sequence $\boldsymbol{z}$ specifies the probability distribution $p(x_m|z_m)$, $m = 1, 2, \ldots, n$, from which we obtain

$$\ell_m^{(\mathrm{dr})} = \ln\left(\frac{p(x_m = 0|z_m)}{p(x_m = 1|z_m)}\right), \quad m = 1, 2, \ldots, n, \qquad (8)$$

which are the input to a check node with indicator function $\mathbb{1}_{\{x_m = x_{1,m} \oplus x'_{2,m}\}}$.

For two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}'_2$ at the relay, the situation is different in that the soft information about $\boldsymbol{x}$ is not formed at the relay. Nevertheless, the quantization at the relay specifies the distribution $p(z_m|x_{1,m}, x'_{2,m})$, $m = 1, 2, \ldots, n$, which is available given perfect reconstruction of $\boldsymbol{z}$ at the destination. Consequently, the coupling of the two component decoders in the factor graph of the iterative decoder at the destination occurs through the function nodes specified by

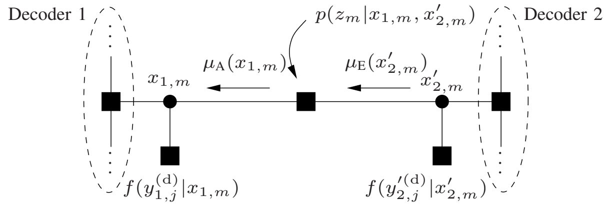


Fig. 4. One section of the factor graph at the destination with messages $\mu_{\mathrm{A}}(x_{1,m})$ and $\mu_{\mathrm{E}}(x'_{2,m})$, where $j = \lceil m/\log_2(M_i)\rceil$.

$p(z_m|x_{1,m}, x'_{2,m})$, a section of which is shown in Fig. 4. For convenience, we describe the operations of that function node in terms of likelihood ratios. To that end, define for $\xi \in \{0, 1\}$

$$\ell(x_{1,m}, x'_{2,m} = \xi|z_m) \triangleq \ln\left(\frac{p(z_m|x_{1,m} = 0, x'_{2,m} = \xi)}{p(z_m|x_{1,m} = 1, x'_{2,m} = \xi)}\right) \quad (9)$$

$$\ell(x_{1,m} = \xi, x'_{2,m}|z_m) \triangleq \ln\left(\frac{p(z_m|x_{1,m} = \xi, x'_{2,m} = 0)}{p(z_m|x_{1,m} = \xi, x'_{2,m} = 1)}\right) \quad (10)$$

$$\ell(x_{1,m}, x'_{2,m}|z_m) \triangleq \ln\left(\frac{p(z_m|x_{1,m} = 0, x'_{2,m} = 1)}{p(z_m|x_{1,m} = 1, x'_{2,m} = 0)}\right). \quad (11)$$

Then, for $m = 1, 2, \dots, n$, $\ell^{(\mathrm{A})}_{1,m}$ and $\ell^{(\mathrm{A})\prime}_{2,m}$ are given by

$$\ell^{(\mathrm{A})}_{1,m} = \ln\left(\frac{1 + e^{\ell^{(\mathrm{E})\prime}_{2,m}} e^{\ell(x_{1,m}=0, x'_{2,m}|z_m)}}{e^{\ell^{(\mathrm{E})\prime}_{2,m}} e^{-\ell(x_{1,m}, x'_{2,m}|z_m)} + e^{-\ell(x_{1,m}, x'_{2,m}=1|z_m)}}\right) \quad (12)$$

$$\ell^{(\mathrm{A})\prime}_{2,m} = \ln\left(\frac{1 + e^{\ell^{(\mathrm{E})}_{1,m}} e^{\ell(x_{1,m}, x'_{2,m}=0|z_m)}}{e^{\ell^{(\mathrm{E})}_{1,m}} e^{\ell(x_{1,m}, x'_{2,m}|z_m)} + e^{-\ell(x_{1,m}=1, x'_{2,m}|z_m)}}\right). \quad (13)$$

Appendix A gives a proof of (12) and (13).

E. Reference Schemes

In addition to point-to-point links without the use of the relay, we also consider analog transmission of $\boldsymbol{\ell}$ [12] from the relay and transmission as soft bit [15] as reference schemes. For the sake of completeness, Fig. 3 includes the corresponding operations at the destination.

III. Quantizer Design

In this section, we study the design of mutual-information preserving quantizers for application at the relay to allow maximal exchange of soft information between the component decoders at the destination. Throughout, we restrict the design framework to the case where the quantizer output can be perfectly recovered at the destination.

A. One-dimensional Quantizers

The most common distortion measure for the design of quantizers and for rate-distortion problems involving real-valued random variables is the squared-error distortion for its simplicity and convenience in analysis, and despite its lack of perceptual meaningfulness for some problems [16]. In general, finding the right distortion measure for a particular problem can be a difficult and controversial task [16, Chapter 2.4]. Given that fact, it seems to be equally hard to choose a distortion function for the problem we consider here, namely quantizing the soft information at the relay node. Therefore, we follow the approach taken by Tishby *et al.* in [17], where they deal with the rate distortion problem in a different way using the notion of *relevance through another variable*. Instead of putting the constraint on the average distortion for some distortion measure chosen *a-priori*, the constraint is that the quantization $q(L)$ should contain some minimum level of information about a third variable, the *relevant* variable, which, in our case, is the random variable $X = X_1 \oplus X_2$. Given a random variable L representing the soft information at the relay, the goal is to find a quantized version $q(L)$ that contains as much *relevant* information as possible, which is information about X. That is, instead of forcing, e.g., the squared error between L and $q(L)$ to be small, the goal is to preserve as much information as possible in $q(L)$ about X.

We now formalize these ideas. In order to design a quantization function q with N quantization regions, we aim at solving the optimization problem

$$I^* = \sup_{q:\mathbb{R}\to\mathcal{Z}} I(X; q(L)) \quad \text{s.t. } |\mathcal{Z}| = N. \quad (14)$$

In order to allow computation of a good mutual-information preserving quantizer in the following, we make a number of simplifying assumptions. First, finely quantize the range of the continuous random variable L with density $f(\ell)$, yielding a random variable $\bar{L}$ with finite range and mass function $p(\ell)$, so that $\bar{L}$ is from a finite set $\mathcal{L}$. The optimization problem at hand now is

$$\bar{q}^* = \operatorname*{argmax}_{\bar{q}:\mathcal{L}\to\mathcal{Z}} I(X; \bar{q}(\bar{L})) \quad \text{s.t. } |\mathcal{Z}| = N. \quad (15)$$

The second step comprises a transformation of the mapping $\bar{q}$, similar to before, into a conditional mass function $p(z|\ell) = \mathbb{1}_{\{\bar{q}(\ell)=z\}}$ allowing us to rewrite (15) as

$$p^*(z|\ell) = \operatorname*{argmax}_{p(z|\ell)\in\mathcal{P}_1} I(X; Z), \quad (16)$$

where the constraint set

$$\mathcal{P}_1 = \Big\{p(z|\ell) : p(z|\ell) \in \{0, 1\} \ \forall\ z \in \mathcal{Z}, \ell \in \mathcal{L}, \sum_z p(z|\ell) = 1 \ \forall\ \ell \in \mathcal{L}, |\mathcal{Z}| = N\Big\} \quad (17)$$

ensures that the mapping $p(z|\ell)$ is a valid conditional mass function, and that it represents a scalar deterministic quantizer with N quantization regions.

Before proceeding, consider the related problem

$$\max_{p(z|\ell)\in\mathcal{P}'_1} I(X; Z), \quad (18)$$

with

$$\mathcal{P}'_1 = \Big\{p(z|\ell) : p(z|\ell) \geq 0\ \forall z \in \mathcal{Z}, \ell \in \mathcal{L}, \sum_z p(z|\ell) = 1\ \forall \ell \in \mathcal{L}\Big\}.$$

The set $\mathcal{P}'_1$ is a polyhedron, and therefore convex [18]; also, $\mathcal{P}'_1$ is bounded and closed, and therefore compact [18, Section 2.2]. Further, $I(X; Z)$ is convex in the distribution $p(z|\ell)$, for fixed $p(x)$ and $p(\ell|x)$ [19]. Consequently, Problem (18) is a convex *maximization* over a compact polyhedral set, whose maximum is therefore attained at an extreme point of $\mathcal{P}_1'$ [18, Theorem 3.4.7]. However, to solve (18) with global optimality, one still needs to search all extreme points of $\mathcal{P}_1'$, which is prohibitively complex since there are $N^{|\mathcal{L}|}$ of those. Since all the extreme points of $\mathcal{P}_1'$ correspond to a distribution $p(z|\ell) \in \{0,1\}$, this also shows that scalar *deterministic* quantization (as considered in (16)) is optimal, i.e., the mutual information $I(X;Z)$ cannot be improved by allowing randomized quantization, so that (18) and (16) have the same maximizer.

In the following, we are restricting ourselves to finding a locally optimal solution to (16).

Proposition 1: Solving Problem (16) is equivalent to solving

$$p^*(z|\ell) = \underset{p(z|\ell)\in\mathcal{P}_1}{\operatorname{argmin}} \operatorname{E}\left[D\left(p(x|\bar{L})||p(x|Z)\right)\right]. \tag{19}$$

Proof: By the chain rule for mutual information [19], we have

$$I(X;\bar{L},Z) = I(X;Z) + I(X;\bar{L}|Z) = I(X;\bar{L}) + I(X;Z|\bar{L}).$$

Since $X \leftrightarrow \bar{L} \leftrightarrow Z$ form a Markov chain, we have $p(x|\ell,z) = p(x|\ell)$ and $I(X;Z|\bar{L}) = 0$, so that one obtains

$$I(X;Z) = I(X;\bar{L}) + \underbrace{I(X;Z|\bar{L})}_{=0} - I(X;\bar{L}|Z) \tag{20}$$

$$= I(X;\bar{L}) - \sum_{x,\ell,z} p(x,\ell,z) \log_2\left(\frac{p(x,\ell|z)}{p(x|z)p(\ell|z)}\right) \tag{21}$$

$$= I(X;\bar{L}) - \sum_{\ell,z} p(\ell,z) \sum_x p(x|\ell) \log_2\left(\frac{p(x|\ell)}{p(x|z)}\right) \tag{22}$$

$$= I(X;\bar{L}) - \sum_{\ell,z} p(\ell,z) D\big(p(x|\ell)||p(x|z)\big) \tag{23}$$

$$= I(X;\bar{L}) - \operatorname{E}\left[D\big(p(x|\bar{L})||p(x|Z)\big)\right]. \tag{24}$$

For $I(X;\bar{L})$ is fixed for a given $p(x,\ell)$, the maximization in (16) is equivalent to minimizing the expected relative entropy $\operatorname{E}[D(p(x|\bar{L})||p(x|Z))]$. ■

Hence, the relative entropy between $p(x|\ell)$ and $p(x|z)$ can be seen as the distortion measure[1]

$$d(\ell,z) = D(p(x|\ell)||p(x|z)) = \sum_x p(x|\ell) \log_2\left(\frac{p(x|\ell)}{p(x|z)}\right) \tag{25}$$

for the problem at hand. Note that relative entropy *emerged* as the right distortion measure for the quantizer design problem by posing it as an optimization of relevant information.

In Problem (19), the probability distribution $p(x,\ell)$ is fixed and known, and has to be obtained numerically. Hence, $p(z|\ell)$ is the only free variable. This is due to the fact that $p(x,\ell)$ is fixed, and $p(z) = \sum_\ell p(\ell)p(z|\ell)$ and $p(x|z) = (1/p(z))\sum_\ell p(x,\ell)p(z|\ell)$ are fully determined by $p(z|\ell)$. Although the optimal distribution $p(z|\ell)$ cannot be obtained in closed form, we propose an iterative optimization algorithm that can be shown to converge to a Karush-Kuhn-Tucker (KKT) point [18, Section 4.3] of (18). We summarize the algorithm in Fig. 5. Convergence of the algorithm follows since the update of the mapping $p(z|\ell)$ in line 9 of the algorithm is chosen such that the average distortion of the new mapping is no worse than that of the previous mapping, and from the concavity of $I(X;\bar{L}|Z)$ with respect to $p(z|\ell)$. In essence, this algorithm is reminiscent of the Lloyd algorithm [20], where however, in our algorithm, the distortion function $d(\ell,z)$ is given by the relative entropy (25), and therefore depends on the mapping $p(z|\ell)$. This is reflected in the update of line 12 of the algorithm. The algorithm in Fig. 5 is also related to the iterative information bottleneck algorithm [17], where our algorithm can be recovered by choosing the Lagrange parameter β of [17] to be $\beta \gg 0$ to ensure that the mapping $p(z|\ell) \in \{0,1\}$ corresponds to a deterministic quantizer. The iterative algorithm of [17] in turn is reminiscent of the celebrated Blahut–Arimoto algorithm [21] for computing channel capacities and rate distortion functions, with the main difference that the algorithm for computing the mapping in the information bottleneck setting updates the distribution $p(x|z)$ as well to incorporate the dependency of the distortion $d(\ell,z)$ on the mapping $p(z|\ell)$ to be optimized. Relative entropy as a distortion measure for vector quantizer design was also employed in [22]; however, the algorithm in [22] is explicitly formulated for quantizing LLRs, whereas the algorithm in Fig. 5 can be used to design a quantizer for maximum mutual information irrespective of whether $\bar{L}$ is an LLR or not, as long as the joint probability mass function $p(x,\ell)$ or an estimate thereof is available. Further, from the algorithm in Fig. 5, the connection to the information bottleneck principle [17] is evident, which is a general framework for the tradeoff between rate and relevant mutual information.

1: **Input:** $p(x,\ell), N = |\mathcal{Z}|, \epsilon > 0$
2: **Initialization:** randomly choose a valid mapping $p(z|\ell) \in \{0,1\}$, $D^{(\text{old})} \leftarrow I(X;\bar{L})$
3: $p(z) \leftarrow \sum_\ell p(\ell)p(z|\ell)$
4: $p(x|z) \leftarrow (1/p(z))\sum_\ell p(x,\ell)p(z|\ell)$
5: $d(\ell,z) \leftarrow D(p(x|\ell)||p(x|z))$
6: $D^{(\text{new})} \leftarrow \operatorname{E}[d(\bar{L},Z)]$ (Compute average distortion)
7: **while** $D^{(\text{old})} - D^{(\text{new})} \geq \epsilon$ **do**
8: $\quad D^{(\text{old})} \leftarrow D^{(\text{new})}$
9: $\quad$ find, for each ℓ, $z_\ell^* = \operatorname{argmin}_z d(\ell,z)$, and set $p(z|\ell) \leftarrow \mathbb{1}_{z=z_\ell^*}$ (New mapping)
10: $\quad p(z) \leftarrow \sum_\ell p(\ell)p(z|\ell)$ (Update of probabilities)
11: $\quad p(x|z) \leftarrow (1/p(z))\sum_\ell p(x,\ell)p(z|\ell)$
12: $\quad d(\ell,z) \leftarrow D(p(x|\ell)||p(x|z))$ (Update distortion function)
13: $\quad D^{(\text{new})} \leftarrow \operatorname{E}[d(\bar{L},Z)]$ (Update average distortion)
14: **end while**

Fig. 5. Algorithm to compute $p(z|\ell)$.

Remark 1: Since the resulting mapping $p(z|\ell) \in \{0,1\}$ represents a scalar quantizer, the mutual information $I(\bar{L};Z) = H(Z)$, which is also the rate of the resulting quantizer. Fixing N, the rate of the quantizer is therefore upper bounded by $\log_2(N)$. Since Problem (16) is a convex *maximization* problem, the algorithm in Fig. 5 is only guaranteed to find a solution satisfying the necessary conditions for local optimality. In our attempt to find a good mutual-information preserving quantizer, the algorithm is therefore repeatedly carried out with random starting conditions until a satisfactory solution

[1]Strictly speaking, the relative entropy $D(p||q)$ is not a distortion measure, since it does not satisfy the triangle inequality and is not symmetric, i.e., $D(p||q) \neq D(q||p)$ in general.

is obtained.

B. Two-dimensional Quantizers

As for one-dimensional quantization of $\boldsymbol{\ell}$, the framework for designing a two-dimensional quantizer for $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2'$ will be established using mutual information as a figure of merit, but with a different expression as relevant information, whose choice will be motivated by the following two arguments.

1) Inspection of the iterative decoding process at the destination: During that process, cf. Fig. 2, the component decoders produce random variables $\boldsymbol{L}_1^{(\mathrm{E})}$ and $\boldsymbol{L}_2^{(\mathrm{E})}$ with mutual information $I(X_1; L_1^{(\mathrm{E})})$ and $I(X_2; L_2^{(\mathrm{E})})$, which is the input information to the corresponding relay check node. At this point, again assuming error-free transmission of the quantizer output $\boldsymbol{Z}$, we note that the relay check node in the receiver processes $\boldsymbol{Z}$ and $\boldsymbol{L}_i^{(\mathrm{E})}$ to produce *a-priori* information for the corresponding component decoder. Therefore, we would like the quantizer at the relay to be such that $I(X_i; Z, L_j^{(\mathrm{E})})$, $i, j \in \{1, 2\}$, $i \neq j$, is maximal, both for the information exchange from decoder 1 to decoder 2, and vice versa. Since

$$I(X_i; Z, L_j^{(\mathrm{E})}) = I(X_i; L_j^{(\mathrm{E})}) + I(X_i; Z|L_j^{(\mathrm{E})}) \quad (26)$$
$$= I(X_i; Z|L_j^{(\mathrm{E})}), \quad (27)$$

where it is assumed that $I(X_i; L_j^{(\mathrm{E})})$, $i \neq j$, vanishes, we are left with maximizing $I(X_i; Z|L_j^{(\mathrm{E})})$, which is, however, hard to maximize due to its dependence on the variable $L_j^{(\mathrm{E})}$ that changes its statistics with increasing number of iterations. We therefore propose the following. Observing that the extrinsic information $L_j^{(\mathrm{E})}$ from component decoder j being perfectly reliable corresponds to X_j being given, we optimize the mutual information $I(X_i; Z|X_j)$ instead of $I(X_i; Z|L_j^{(\mathrm{E})})$, thereby removing the dependency on the changing statistics of $L_j^{(\mathrm{E})}$. Consequently, to allow maximal information transfer from decoder 1 to decoder 2, $I(X_2; Z|X_1)$ should be maximized, and analogously, for decoder 1 to receive maximal information from decoder 2, $I(X_1; Z|X_2)$ should be as large as possible. Various combinations of these information expressions can be taken to form the relevant information term. We propose to take the average of $I(X_1; Z|X_2)$ and $I(X_2; Z|X_1)$ as the relevant information term, i.e., $I_{\mathrm{rel}} \triangleq I(X_1; Z|X_2) + I(X_2; Z|X_1)$.

2) Connection with one-dimensional quantization of soft information about XORed code bits: In addition to the motivation above, we also establish a connection between the proposed cost function I_{rel} for two-dimensional quantization with the cost function employed for the design of one-dimensional quantizers in the following proposition.

Proposition 2: Let L be the soft information about $X = X_1 \oplus X_2$ of two independent and equally likely bits X_1 and X_2, and let q_1 be the quantization function of a quantizer processing L, so that $Z = q_1(L)$. Further, assume that $f(\ell_i|x_i)$ satisfies the symmetry condition $f(\ell_i|0) = f(-\ell_i|1)$. Then, $I(X; q_1(L)) = I(X_1, X_2; q_1(L))$ and $I(X_1; q_1(L)) = I(X_2; q_1(L)) = 0$.

The proof is relegated to Appendix B. From Proposition 2 we see that maximizing $I(X; q_1(L))$ corresponds to maximizing $I(X_1, X_2; q_1(L))$, subject to $I(X_1; q_1(L)) = I(X_2; q_1(L)) = 0$. For two-dimensional quantization using a quantization function $q_2 : \mathbb{R} \times \mathbb{R} \rightarrow \mathcal{Z}$, we relax the condition $I(X_1; q_2(L_1, L_2)) = I(X_2; q_2(L_1, L_2)) = 0$, but seek to choose the quantization function q_2 such that $q_2(L_1, L_2)$ carries as much information about the pair (X_1, X_2) as possible, while carrying little information about X_1 and X_2 alone. This is reflected in the choice of

$$I_{\mathrm{rel}} = I(X_1; q_2(L_1, L_2)|X_2) + I(X_2; q_2(L_1, L_2)|X_1) \quad (28)$$
$$= 2I(X_1, X_2; q_2(L_1, L_2)) - I(X_1; q_2(L_1, L_2)) \quad (29)$$
$$- I(X_2; q_2(L_1, L_2)).$$

To compute a two-dimensional mutual information preserving quantizer, we finely quantize the ranges of the continuous random variables L_1 and L_2 with densities $f(\ell_1)$ and $f(\ell_2)$ to obtain discrete variables $\bar{L}_1 \in \mathcal{L}_1$ and $\bar{L}_2 \in \mathcal{L}_2$ with probability mass functions $p(\ell_1)$ and $p(\ell_2)$, where both $\mathcal{L}_1$ and $\mathcal{L}_2$ are finite sets. Writing the mapping $\bar{q} : \mathcal{L}_1 \times \mathcal{L}_2 \rightarrow \mathcal{Z}$ as $p(z|\ell_1, \ell_2) = \mathbb{1}_{\{\bar{q}(\ell_1, \ell_2) = z\}}$, we pose the optimization problem we wish to solve as

$$p^*(z|\ell_1, \ell_2) = \underset{p(z|\ell_1, \ell_2) \in \mathcal{P}_2}{\operatorname{argmax}} \; I_{\mathrm{rel}}, \quad (30)$$

where

$$\mathcal{P}_2 = \Bigg\{ p(z|\ell_1, \ell_2) : p(z|\ell_1, \ell_2) \in \{0,1\}, \forall (z, \ell_1, \ell_2) \in \mathcal{Z} \times \mathcal{L}_1 \times \mathcal{L}_2$$
$$\sum_{z \in \mathcal{Z}} p(z|\ell_1, \ell_2) = 1, \forall (\ell_1, \ell_2) \in \mathcal{L}_1 \times \mathcal{L}_2, |\mathcal{Z}| = N \Bigg\}.$$

Proposition 3: Solving Problem (30) is equivalent to solving

$$\underset{p(z|\ell_1, \ell_2) \in \mathcal{P}_2}{\operatorname{argmin}} \Big\{ \mathrm{E}\left[2D\big(p(x_1, x_2|\bar{L}_1, \bar{L}_2) || p(x_1, x_2|Z)\big) \right] \quad (31)$$
$$- \mathrm{E}\left[D\big(p(x_1|\bar{L}_1) || p(x_1|Z)\big) \right] - \mathrm{E}\left[D\big(p(x_2|\bar{L}_2) || p(x_2|Z)\big) \right] \Big\}.$$

Proof: Similar to the proof of Proposition 1 we use that $(X_1, X_2) \leftrightarrow (\bar{L}_1, \bar{L}_2) \leftrightarrow Z$ forms a Markov chain, and that the term $I(X_1, X_2; \bar{L}_1, \bar{L}_2)$ is fixed, so that maximizing I_{rel} is equivalent to minimizing $2I(X_1, X_2; \bar{L}_1, \bar{L}_2|Z) - I(X_1; \bar{L}_1|Z) - I(X_2; \bar{L}_2|Z)$, which can be expressed in terms of relative entropies as in (31). ∎

Remark 2: To compute an approximate solution to (31), we can use an appropriately modified version of the iterative algorithm in Fig. 5. The algorithm for the two-dimensional quantizer design is given in Fig. 6. Similarly to Section III-A, the mass functions $p(x_1, \ell_1)$ and $p(x_2, \ell_2)$ are obtained numerically, and we run the algorithm in Fig. 6 with different starting conditions to ensure that a good two-dimensional quantizer is found. For a deterministic quantizer with quantization rule $q(\ell_1, \ell_2)$ we then have $I(\bar{L}_1, \bar{L}_2; Z) = H(Z)$ as the rate of the resulting quantizer.

C. Examples of Quantizers

In this section, we present some examples for quantizers obtained with the algorithms in Figs. 5 and 6. In all cases, the

TABLE I
COMPARISON OF QUANTIZER CHARACTERISTICS.

Quantizer Type	γ_r	$\lvert\mathcal{Z}\rvert$	Decision Border(s)	$\ln\big(p(x=0\vert z)/p(x=1\vert z)\big)$	$I(X;\bar{L})$	$I(X;Z)$	$H(Z)$
Mutual information	0 dB	2	[0]	[-5.61, 5.61]	0.985	0.964	1
Mutual information	-2 dB	3	[-1.45, 1.45]	[-3.55, 0, 3.55]	0.762	0.714	1.425
Lloyd–Max	-2 dB	3	[-3.25, 3.25]	[-4.91, 0, 4.91]	0.762	0.634	1.585
Uniform	-2 dB	3	[-4.75, 4.75]	[-6.05, 0, 6.05]	0.762	0.484	1.495
Mutual information	-3 dB	5	[-2.30, -0.71, 0.71, 2.30]	[-3.35, -1.30, 0, 1.30, 3.35]	0.505	0.485	2.286

```
Input: p(x1, x2, ℓ1, ℓ2), N = |Z|, ε > 0
Initialization: randomly choose a valid mapping
        p(z|ℓ1, ℓ2) ∈ {0,1}, D^(old) ← I(X1, X2; L̄1, L̄2)
p(z) ← Σ_{ℓ1,ℓ2} p(ℓ1, ℓ2) p(z|ℓ1, ℓ2)
p(x1, x2|z) ← (1/p(z)) Σ_{ℓ1,ℓ2} p(x1, x2, ℓ1, ℓ2) p(z|ℓ1, ℓ2)
d(ℓ1, ℓ2, z) ← 2D(p(x1, x2|ℓ1, ℓ2)||p(x1, x2|z))
        −D(p(x1|ℓ1)||p(x1|z)) − D(p(x2|ℓ2)||p(x2|z))
D^(new) ← E[d(L̄1, L̄2, Z)]          (Compute average distortion)
while D^(old) − D^(new) ≥ ε do
    D^(old) ← D^(new)
    find z*_{ℓ1,ℓ2} = argmin_z d(ℓ1, ℓ2, z),          (New mapping)
    and set p(z|ℓ1, ℓ2) ← 1_{z=z*_{ℓ1,ℓ2}}
    p(z) ← Σ_{ℓ1,ℓ2} p(ℓ1, ℓ2) p(z|ℓ1, ℓ2)  (Update of probabilities)
    p(x1, x2|z) ← (1/p(z)) Σ_{ℓ1,ℓ2} p(x1, x2, ℓ1, ℓ2) p(z|ℓ1, ℓ2)
    d(ℓ1, ℓ2, z) ← 2D(p(x1, x2|ℓ1, ℓ2)||p(x1, x2|z))
        −D(p(x1|ℓ1)||p(x1|z)) − D(p(x2|ℓ2)||p(x2|z))
                                    (Update distortion function)
    D^(new) ← E[d(L̄1, L̄2, Z)]      (Update average distortion)
end while
```

Fig. 6. Algorithm to compute $p(z|\ell_1, \ell_2)$.

underlying channel codes are recursive convolutional codes with generator

$$G(D) = \left(1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right) \tag{32}$$

and information blocklength $k = k_1 = k_2 = 1996$. BPSK modulation is employed at the sources.

The first set of examples involves some illustration on how the designed quantizers look like for different values of the source–relay SNR $\gamma_{\mathrm{r},i}$ and alphabet sizes $|\mathcal{Z}|$, in case of one-dimensional quantizers for $\boldsymbol{\ell}$ and BCJR [23] soft decoders at the relay. Here, we assume symmetric source–relay channels, i.e., $\gamma_\mathrm{r} = \gamma_{\mathrm{r},1} = \gamma_{\mathrm{r},2}$. Numerical characteristics of the designed quantizers are summarized in Table I. As the source–relay SNR decreases, the number of quantization regions required to achieve a mutual information $I(X;Z)$ *close to the limit* $I(X;\bar{L})$ increases, leading to an increase in rate on the relay–destination link. To highlight the effectiveness of the proposed quantizer design framework using mutual information as a figure of merit compared to other well-known quantization methods, we also exemplarily show the parameters of the Lloyd–Max [20] and uniform quantizer in Table I for $\gamma_\mathrm{r} = -2$ dB and $|\mathcal{Z}| = 3$. Evidently, both the Lloyd–Max and the uniform quantizer require a higher rate than the quantizer designed with the proposed algorithm, while preserving less relevant mutual information.

In the second set of examples, the relay performs soft demapping only. Fig. 7 depicts the partitioning of the (ℓ_1, ℓ_2)-plane into quantization regions as obtained by the iterative optimization algorithm. Each of the resulting regions is gray-level coded, with each gray level corresponding to one symbol of the quantizer alphabet $\mathcal{Z}$. Using $|\mathcal{Z}| = 3$ regions, the quantizer is shown in Fig. 7(a) for symmetric source–relay links at $\gamma_{\mathrm{r},1} = \gamma_{\mathrm{r},2} = 4$ dB. Note that this partition effectively mimics one-dimensional quantization of the soft information $\boldsymbol{\ell}$ about the XOR-coded bits. In contrast, if channel conditions on the source–relay links are profoundly different, then the relay should preferably allocate more of the rate available on the relay–destination channel to the stronger user, and this is exactly achieved with the two-dimensional formulation of the quantization problem at the relay, as shown in Fig. 7(b) for again $|\mathcal{Z}| = 3$ regions, where $\gamma_{\mathrm{r},1} = -8$ dB and $\gamma_{\mathrm{r},2} = -1$ dB. Note that in this rather extreme case, all the quantization rate is allocated to the second user, whose soft information at the relay is vastly more reliable than the one of the first user. Finally, Fig. 7(c) displays a typical quantization mapping obtained for $\gamma_{\mathrm{r},1} = -4$ dB, $\gamma_{\mathrm{r},2} = 1$ dB, and $|\mathcal{Z}| = 5$ regions.

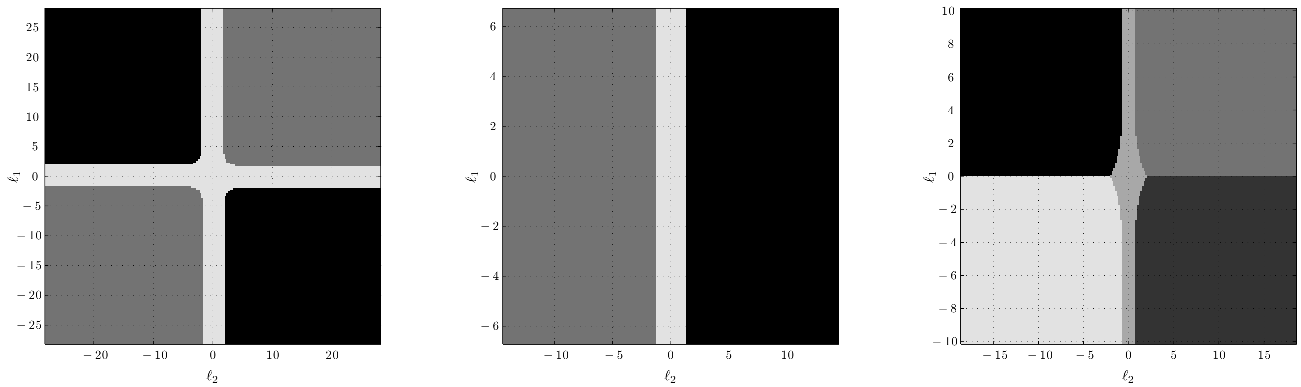
IV. SIMULATION RESULTS

A. Additive White Gaussian Noise Channels

We first show bit error rate (BER) results for AWGN channels, i.e., $h_{\mathrm{r},i} = h_{\mathrm{d},i} = h_{\mathrm{d,r}} = 1$, and symmetric links, for which $\gamma_\mathrm{d} = \gamma_{\mathrm{d},1} = \gamma_{\mathrm{d},2}$ as well as $\gamma_\mathrm{r} = \gamma_{\mathrm{r},1} = \gamma_{\mathrm{r},2}$. Both sources employ BPSK modulation. In the reference system without the aid of the relay, a recursive convolutional code with generator [24, Chapter 9.1]

$$G(D) = \left(1, \frac{1+D+D^3+D^4}{1+D^2+D^4}, \frac{1+D+D^2+D^3+D^4}{1+D^2+D^4}\right) \tag{33}$$

is used at the sources with $k = k_1 = k_2 = 996$ and $n = 3000$, yielding 6000 total uses of the channel. In the system with the relay, the sources use the recursive convolutional code with generator given in (32), again with $k = k_1 = k_2 = 996$ information bits and $n = 2000$, so that a fair comparison with the reference system is guaranteed. Throughout, $\gamma_\mathrm{r} = 4$ dB, and the relay employs soft demodulation to obtain the soft information. The scalar quantizer used at the relay is one with $N = 3$ quantization regions, for which $H(Z) = 1.257 < \log_2(3)$. Source coding at the relay is therefore beneficial to exploit the additional redundancy in $\boldsymbol{Z}$, and is performed with an arithmetic code [25]. We also employ the corresponding two-dimensional quantizer (shown in Fig. 7(a)) for comparison. Taking $\gamma_{\mathrm{d,r}} = 3.5$ dB ensures reliable transmission of $\boldsymbol{z}$ with a turbo code of appropriate rate as specified in the Universal Mobile Telecommunication System (UMTS) [26] standard, and 8-phase shift keying (8-PSK) modulation at the



(a) $\gamma_{\mathrm{r},1} = \gamma_{\mathrm{r},2} = 4\,\mathrm{dB}, |\mathcal{Z}| = 3$. (b) $\gamma_{\mathrm{r},1} = -8\,\mathrm{dB}, \gamma_{\mathrm{r},2} = -1\,\mathrm{dB}, |\mathcal{Z}| = 3$. (c) $\gamma_{\mathrm{r},1} = -4\,\mathrm{dB}, \gamma_{\mathrm{r},2} = 1\,\mathrm{dB}, |\mathcal{Z}| = 5$.

Fig. 7. Quantizers obtained for two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2'$ at the relay. The quantizer in (c) is also suitable for $\gamma_{\mathrm{r},1} = -1\,\mathrm{dB}, \gamma_{\mathrm{r},2} = 4\,\mathrm{dB}$, and QPSK modulation with soft demapping.

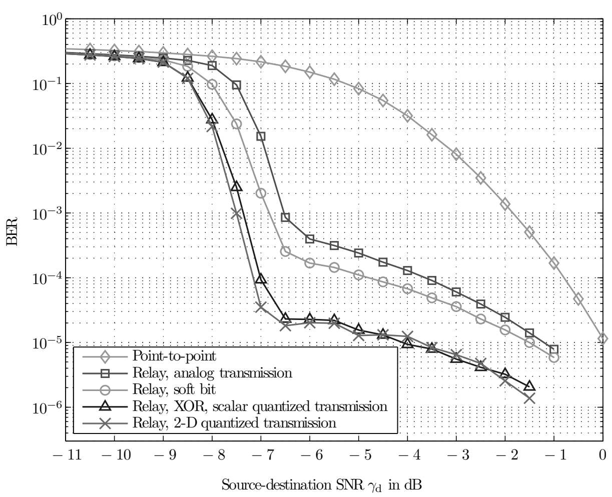


Fig. 8. BERs for symmetric source–relay channels, $\gamma_\mathrm{r} = 4\,\mathrm{dB}$, $|\mathcal{Z}| = 3$.

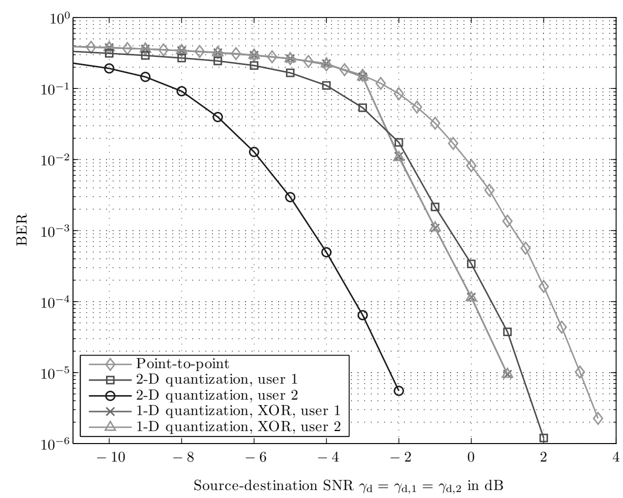


Fig. 9. BERs for asymmetric source–relay channels, $\gamma_{\mathrm{r},1} = \gamma_\mathrm{d} + 3\,\mathrm{dB}$, $\gamma_{\mathrm{r},2} = \gamma_\mathrm{d} + 8\,\mathrm{dB}$, $|\mathcal{Z}| = 5$.

relay. Note that $\gamma_{\mathrm{d,r}} = 3.5\,\mathrm{dB}$ is kept constant for analog and soft bit transmission as well. The corresponding BER curves are shown in Fig. 8, from which we observe that the schemes with the relay considerably outperform the reference point-to-point link; furthermore, quantized transmission provides a gain of roughly $1\,\mathrm{dB}$ over analog transmission in the waterfall region of the BER curve. In this symmetric scenario, the gain of two-dimensional quantization over scalar quantization is marginal; however, the picture changes for asymmetric source–relay links.

Fig. 9 shows results for asymmetric source–relay channel conditions. Again, $k = k_1 = k_2 = 996$, and the sources employ the recursive convolutional code with generator in (32) and QPSK modulation, yielding $m_1 = m_2 = 1000$. We assume that $\gamma_{\mathrm{r},1} = \gamma_\mathrm{d} + 3\,\mathrm{dB}$ and $\gamma_{\mathrm{r},2} = \gamma_\mathrm{d} + 8\,\mathrm{dB}$, and that $\gamma_{\mathrm{d,r}} = 18\,\mathrm{dB}$. The relay performs soft demapping of its received signals followed by quantization with $N = 5$ regions and an arithmetic encoder for source coding. Although some of the quantizers used in this scheme have $H(Z)$ very close to the limit of $\log_2(5)$ and hence, the redundancy in $\boldsymbol{Z}$ is small, source coding is used here to obtain a binary representation of $\boldsymbol{z}$. On the relay–destination link, we use the UMTS turbo code of appropriate rate and 256-Quadrature Amplitude Modulation (QAM) with $m_\mathrm{r} = 1000$. Note that for $\gamma_\mathrm{d} = -4\,\mathrm{dB}$, the quantizer used at the relay is given in Fig. 7(c). In the point-to-point link, the sources have $k = k_1 = k_2 = 996$ information bits, and use the convolutional code with generator in (33) with QPSK modulation, so that $m_1 = m_2 = 1500$. As expected, two-dimensional quantization considerably outperforms one-dimensional quantization of the soft information about the XOR-coded bits where the source–relay links have different SNR. The gains compared to the point-to-point link are most pronounced for small source–destination SNR, with user 2 as the stronger user at the relay doing clearly better than user 1.

B. Block Fading Channels

We now turn our attention to the performance of the proposed schemes in Rayleigh block fading channels, where we assume that the fading variables $h_{\mathrm{r},i}$, $h_{\mathrm{d},i}$, and $h_{\mathrm{d,r}}$ are mutually independent and each distributed according to $\mathcal{CN}(0,1)$. Throughout, we assume a symmetric network, i.e., $d_\mathrm{r} = d_{\mathrm{r},1} = d_{\mathrm{r},2}$ and $d_\mathrm{d} = d_{\mathrm{d},1} = d_{\mathrm{d},2}$. Further, the relay is placed closer to the destination than to the sources. In particular, we set $p = 3.52$ and consider two cases:

- Case 1: the relay is placed between the sources and the destination, with $d_\mathrm{r} = (9/10)d_\mathrm{d}$. Consequently, $\rho_\mathrm{r} = \rho_\mathrm{d} + 1.61\,\mathrm{dB}$, so that the source–relay SNR is only a little larger than the source–destination SNR.
- Case 2: the relay is placed behind the destination, with $d_\mathrm{r} = (3/2)d_\mathrm{d}$, so that $\rho_\mathrm{r} = \rho_\mathrm{d} - 6.20\,\mathrm{dB}$. The relaying scheme turns out to be useful even in this case in which the source–relay SNR is smaller than the source–destination SNR. This is in contrast to decode-and-forward schemes, in which the requirement that the relay can decode reliably requires a fairly high SNR on the source–relay links [7].

Due to the proximity of the destination and the relay, we take $\rho_{\mathrm{r,d}} = \rho_\mathrm{d} + 15\,\mathrm{dB}$.

If the relay is present, the sources have $k = k_1 = k_2 = 2000$ information bits to transmit, but now use the UMTS turbo code [26] of rate $1/2$ and BPSK modulation, so that $m_1 = m_2 = 4000$. The relay and the destination share a set $\mathcal{Q}$ of two-dimensional quantizers designed with the framework introduced in Section III-B. Given the realizations of the received sequences $(\boldsymbol{y}_1^{(\mathrm{r})}, \boldsymbol{y}_2^{(\mathrm{r})})$ and of $\gamma_{\mathrm{r},1}$ and $\gamma_{\mathrm{r},2}$, the relay selects the proper quantizer in $\mathcal{Q}$, computes the sequence $\boldsymbol{z}$, source encodes that sequence with an arithmetic encoder, and channel encodes using the UMTS turbo code of appropriate rate, yielding the sequence $\boldsymbol{s}_\mathrm{r} \in \mathbb{M}_\mathrm{r}^{m_\mathrm{r}}$ with $m_\mathrm{r} = 4000$, where the modulation alphabet $\mathbb{M}_\mathrm{r}$ at the relay is chosen to be 16-QAM. In this example, the entropy coding step is useful both to exploit the additional redundancy in the quantized sequence (depending on the actual choice of the quantizer), and to perform the mapping to a binary string efficiently. We compare two sets of quantizers. The first set $\mathcal{Q}_1$ contains one quantizer with $|\mathcal{Z}| = 5$ quantization regions for every pair of instantaneous SNR values $\gamma_{\mathrm{r},1}$ and $\gamma_{\mathrm{r},2}$ in the set $\mathcal{S}_1 = \{-9\,\mathrm{dB}, -8\,\mathrm{dB}, -7\,\mathrm{dB}, \ldots, 7\,\mathrm{dB}\}$. For this choice of $\mathcal{S}_1$, there are $|\mathcal{Q}_1| = |\mathcal{S}_1|^2 = 289$ quantizers available at the relay. Consequently, signaling the relay's quantizer choice to the destination requires at most 9 bits; we assume this signaling to be perfect in the sequel. The other set $\mathcal{Q}_2$ of quantizers consists of one quantizer with $|\mathcal{Z}| = 5$ regions for every pair of $\gamma_{\mathrm{r},1}$ and $\gamma_{\mathrm{r},2}$ in the set $\mathcal{S}_2 = \{-9\,\mathrm{dB}, 0\,\mathrm{dB}, 6\,\mathrm{dB}\}$, so that $|\mathcal{Q}_2| = 9$, resulting in 4 bits to signal the quantizer choice.

In the reference point-to-point system, the sources employ the UMTS turbo code [26] of rate $1/3$ and $k = k_1 = k_2 = 2000$ with BPSK modulation, yielding $m_1 = m_2 = 6000$. Our second reference system includes the relay, so that $k = k_1 = k_2 = 2000$ and $m_1 = m_2 = m_\mathrm{r} = 4000$; here, the relay does perform one-dimensional quantization of $\boldsymbol{\ell}$, the soft information about $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2'$. In particular, $\mathcal{S}_{\mathrm{XOR}} = \mathcal{S}_1$, so that the quantizer set shared by the relay and the destination contains $|\mathcal{Q}_{\mathrm{XOR}}| = 289$ quantizers with $|\mathcal{Z}| = 5$ quantization regions each. The third reference system under consideration includes the relay as well, so that $k = k_1 = k_2 = 2000$ and $m_1 = m_2 = m_\mathrm{r} = 4000$. However, instead of two-dimensional quantization, the relay performs one-dimensional quantization of the soft information $\boldsymbol{\ell}_j$ of the user j with the stronger

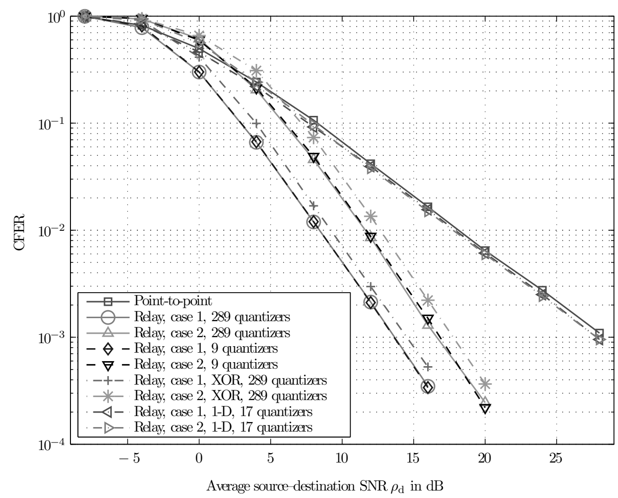


Fig. 10. Frame error rates for the MARC and the point-to-point link.

source–relay channel, while excluding the weaker user in the cooperation. For the stronger user j, there is one quantizer with $|\mathcal{Z}| = 5$ for each instantaneous SNR value $\gamma_{\mathrm{r},j}$ in $\mathcal{S}_1$. Including the bit required to signal the index of the stronger user at the relay, at most 6 signaling bits are needed. Note that the destination performs maximum-ratio combining for that user included in the cooperation.

The simulation results in Fig. 10 show the common frame error rate (CFER) of the reference systems and the system with the relay, for both network geometries. Based on these curves we observe that the schemes involving the relay achieve second order diversity for both case 1 and case 2, since the CFER decays proportional to ρ_d^{-2}. However, if the relay processes $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2')$ directly without going through the intermediate step of computing the likelihood ratios of the XOR of the coded vectors, considerably better performance is obtained, which is due to the fact that the reliability of the XOR at the relay is undesirably dominated by the weaker source–relay channel – a disadvantage avoided by joint quantization of the soft information at the relay. Particularly, the system with the proposed two-dimensional quantizers at the relay gains more than $10\,\mathrm{dB}$ compared to the point-to-point link at relevant CFERs of 10^{-3}. More importantly, simple one-dimensional quantization of the stronger user at the relay is not sufficient to achieve second order diversity. It is also important to note that the scheme involving joint quantization at the relay does not require the set of available quantizers at the relay to be prohibitively large. In fact, we observe that the system with 9 quantizers shared at the relay performs only marginally poorer than the one with a set of 289 quantizers, at considerable lower signaling overhead.

V. Conclusions

In this paper, we studied the MARC with two users and noisy source–relay links preventing successful decoding at the relay, so that the operations at the relay are limited to schemes generating and processing soft information. One- and two-dimensional deterministic quantizers were designed for the soft information at the relay based on the notion of

$$\ell_{1,m}^{(\mathrm{A})} = \ln\left(\frac{p(z_m|x_{1,m}=0,x'_{2,m}=0)\mu_{\mathrm{E}}(x'_{2,m}=0)+p(z_m|x_{1,m}=0,x'_{2,m}=1)\mu_{\mathrm{E}}(x'_{2,m}=1)}{p(z_m|x_{1,m}=1,x'_{2,m}=0)\mu_{\mathrm{E}}(x'_{2,m}=0)+p(z_m|x_{1,m}=1,x'_{2,m}=1)\mu_{\mathrm{E}}(x'_{2,m}=1)}\right) \tag{34}$$

$$= \ln\left(\frac{1+\frac{p(z_m|x_{1,m}=0,x'_{2,m}=0)}{p(z_m|x_{1,m}=0,x'_{2,m}=1)}\frac{\mu_{\mathrm{E}}(x'_{2,m}=0)}{\mu_{\mathrm{E}}(x'_{2,m}=1)}}{\frac{p(z_m|x_{1,m}=1,x'_{2,m}=0)}{p(z_m|x_{1,m}=0,x'_{2,m}=1)}\frac{\mu_{\mathrm{E}}(x'_{2,m}=0)}{\mu_{\mathrm{E}}(x'_{2,m}=1)}+\frac{p(z_m|x_{1,m}=1,x'_{2,m}=1)}{p(z_m|x_{1,m}=0,x'_{2,m}=1)}}\right). \tag{35}$$

relevant information, leading to an improvement over analog transmission methods from the relay. Simulation results further suggest that two-dimensional quantization at the relay outperforms schemes based on network coding the soft values in case of unequal channel quality on the source–relay channels. To perform the quantization, the relay does not require CSI about the source–destination links, a fact especially advantageous in wireless fading channels where this information may not always be available at the relay. In a Rayleigh block fading environment, the relay chooses a suitable quantizer from a fixed set based on the SNR on the incoming links, and forwards its compressed estimate of the received sequences to destination. We observe from numerical results that full diversity order of two can be gained with this scheme. The scheme incurs small delay, since no (soft) decoding is required at the relay node to achieve these gains. Further, we remark that the only overhead created through cooperation is due to the signaling of the quantizer choice at the relay to the destination, since the choice of the quantizer depends on the source–relay SNRs, which are assumed to be unknown at the destination. An efficient low-complexity implementation of two-dimensional quantization might be to approximate the boundaries of the two-dimensional quantizers by hyperplanes, so that the quantization can be found by comparing the vector of likelihoods $(\ell_{1,m}, \ell'_{2,m})$ to be quantized with a number of hyperplanes.

Acknowledgment

The authors gratefully acknowledge the reviewers' constructive comments. They also acknowledge the contribution of Ralf Kötter to the work presented in this paper. Moreover, Georg Zeitler would like to thank Gerhard Kramer for helpful comments and remarks.

Appendix A

To derive the message passing rules given in (12) and (13), consider Fig. 4. The message passing rules for function nodes are applied in the following to compute $\ell_{1,m}^{(\mathrm{A})} = \ln((\mu_{\mathrm{A}}(x_{1,m}=0))/(\mu_{\mathrm{A}}(x_{1,m}=1)))$. Since $\mu_{\mathrm{A}}(x_{1,m}) = \sum_{x_2\in\{0,1\}} p(z_m|x_{1,m},x_2)\mu_{\mathrm{E}}(x_2)$, one obtains (35) at the top of this page, which with the definitions in (9)-(11) yields (12). Along the same lines, one can also verify (13).

Appendix B

The proof of Proposition 2 consists of two parts. First, observe that

$$I(X_1,X_2;q_1(L)) = H(q_1(L)) - H(q_1(L)|X_1,X_2,X) \tag{36}$$

$$\geq H(q_1(L)) - H(q_1(L)|X) \tag{37}$$

$$= I(X;q_1(L)). \tag{38}$$

where (36) holds since $X = X_1 \oplus X_2$ is a deterministic function of X_1 and X_2, and (37) holds since conditioning does not increase entropy. Next, we show that $(X_1,X_2) \leftrightarrow X \leftrightarrow L$ form a Markov chain, i.e., that L is independent of (X_1,X_2) given X. To that end, define

$$g(\lambda_1,\lambda_2) \triangleq \ln\left(\frac{1+e^{\lambda_1+\lambda_2}}{e^{\lambda_1}+e^{\lambda_2}}\right), \qquad \lambda_1\in\mathbb{R}, \lambda_2\in\mathbb{R}, \tag{39}$$

and observe that $g(-\lambda_1,-\lambda_2) = g(\lambda_1,\lambda_2)$. Using the independence of X_1 and X_2 and the symmetry assumption on $f(\ell_i|x_i)$, and substituting $\lambda'_i = -\lambda_i$, we obtain for the conditional cumulative distribution function (CDF) of L that

$$F_{L|X_1,X_2}(\ell|0,0)$$

$$= \Pr\{L \leq \ell | X_1=0, X_2=0\} \tag{40}$$

$$= \iint_{(\lambda_1,\lambda_2):g(\lambda_1,\lambda_2)\leq\ell} f_{L_1,L_2|X_1,X_2}(\lambda_1,\lambda_2|0,0)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 \tag{41}$$

$$= \iint_{(\lambda_1,\lambda_2):g(\lambda_1,\lambda_2)\leq\ell} f_{L_1,L_2|X_1,X_2}(-\lambda_1,-\lambda_2|1,1)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 \tag{42}$$

$$= \iint_{(\lambda'_1,\lambda'_2):g(\lambda'_1,\lambda'_2)\leq\ell} f_{L_1,L_2|X_1,X_2}(\lambda'_1,\lambda'_2|1,1)\mathrm{d}\lambda'_1\mathrm{d}\lambda'_2 \tag{43}$$

$$= F_{L|X_1,X_2}(\ell|1,1). \tag{44}$$

Along the same lines, one can show that $F_{L|X_1,X_2}(\ell|0,1) = F_{L|X_1,X_2}(\ell|1,0)$, so that L is independent of (X_1,X_2) given $X = X_1 \oplus X_2$. Since $(X_1,X_2) \leftrightarrow X \leftrightarrow L$ form a Markov chain, also $(X_1,X_2) \leftrightarrow X \leftrightarrow L \leftrightarrow q_1(L)$ form a Markov chain; consequently, $I(X;q_1(L)) \geq I(X_1,X_2;q_1(L))$ by the data processing inequality, which together with (38) gives $I(X;q_1(L)) = I(X_1,X_2;q_1(L))$. For the second part of the proof, we write for $i\in\{1,2\}$

$$I(X_i;q_1(L)) \leq H(X_i) - H(X_i|X,q_1(L)) \tag{45}$$

$$= H(X_i) - H(X_i|X) = 0, \tag{46}$$

which, together with the nonnegativity of mutual information, gives $I(X_i;q_1(L)) = 0$.

References

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[2] ——, "User cooperation diversity, part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[3] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 100–105, Oct. 2009.

[4] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. IEEE Int. Symp. on Information Theory*, 2000, p. 40.

[5] Y. Chen, S. Kishore, and J. Li, "Wireless diversity through network coding," in *Proc. IEEE Wireless Communications and Networking Conf.*, 2006, pp. 1681–1686.

[6] C. Hausl, F. Schreckenbach, I. Oikonomidis, and G. Bauch, "Iterative network and channel decoding on a Tanner graph," in *Proc. 43rd Ann. Allerton Conf. on Communications, Control, and Computing*, 2005.

[7] C. Hausl, "Joint network-channel coding for the multiple-access relay channel based on turbo codes," *Europ. Trans. Telecomm.*, vol. 20, no. 2, pp. 175–181, Jan. 2009.

[8] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Apr. 2000.

[9] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 572–584, Sep. 1979.

[10] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[11] H. H. Sneessens and L. Vandendorpe, "Soft decode and forward improves cooperative communications," in *Proc. of the 6th IEEE Int. Conf. on 3G and Beyond*, 2005, pp. 1–4.

[12] S. Yang and R. Koetter, "Network coding over a noisy relay: A belief propagation approach," in *Proc. IEEE Int. Symp. on Information Theory*, 2007, pp. 801–804.

[13] W. Rave, "Quantization of log-likelihood ratios to maximize mutual information," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 283–286, Apr. 2009.

[14] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[15] Y. Li, B. Vucetic, T. F. Wong, and M. Dohler, "Distributed turbo coding with soft information relaying in multihop relay networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2040–2050, Nov. 2006.

[16] R. M. Gray, *Source Coding Theory*. Kluwer Academic Publishers, 1990.

[17] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Ann. Allerton Conf. on Communications, Control, and Computing*, 1999, pp. 368–377.

[18] M. S. Bazaara, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*. John Wiley and Sons, Inc., 2006.

[19] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, Inc., 2006.

[20] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 192–137, March 1982.

[21] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[22] M. Danieli, S. Forchhammer, J. Andersen, L. Christensen, and S. Christensen, "Maximum mutual information vector quantization of log-likelihood ratios for memory efficient HARQ implementations," in *Proc. Data Compression Conference (DCC)*, 2010, pp. 30–39.

[23] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

[24] R. E. Blahut, *Algebraic Codes for Data Transmission*. Cambridge University Press, 2003.

[25] J. J. Rissanen, "Generalized Kraft inequality and arithmetic coding," *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 198–203, May 1976.

[26] *Universal Mobile Telecommunications System (UMTS): Multiplexing and channel coding (FDD)*, European Telecommunications Standards Institute TS 125.212 V4.3.0, 2001.