International Zurich Seminar on Communications

March 3–5, 2010

Sorell Hotel Zürichberg, Zurich, Switzerland

Proceedings

Acknowledgment of Support

IEEE

The International Zurich Seminar on Communications (IZS) is organized jointly by the IEEE Switzerland Chapter on Digital Communication Systems and the Department of Information Technology and Electrical Engineering of ETH Zurich.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Financial support by Phonak is gratefully acknowledged.

PHONAK life is on

Conference Organization

General Co-Chairs

Amos Lapidoth and Helmut Bölcskei

Technical Program Committee

Ezio Biglieri
Pierre Chevillat
Bernard Fleury
Christina Fragouli
Martin Hänggi
Walter Hirt
Gerhard Kramer
Hans-Andrea Loeliger
Stefan Mangold
Stefan Moser
Nikolai Nefedov
Igal Sason
Jossy Sayir
Giorgio Taricco
Emre Telatar
Emanuele Viterbo
Pascal Vontobel
Roger Wattenhofer
Armin Wittneben

Organizers of Invited Sessions

Elza Erkip
David Gesbert
Yossi Steinberg
Pascal Vontobel

Local Organization

Conference Secretaries

Barbara Aellig
Sylvia Beringer
Rita Hildebrand
Claudia Zürcher

Web and Publications

Michael Lerjen

Table of Contents

Keynote Talks

Wed 09:00 – 10:00
Frans Willems, Technical University of Eindhoven
Source Codes, Semantic Codes, and Search Codes

Thu 09:00 – 10:00
Frank Kschischang, University of Toronto
An Algebraic Approach to Physical Layer Network Coding

Fri 09:00 – 10:00
Gregory Wornell, MIT
On the Sufficiency of Ignorance:
Recent Lessons in Communication System Architecture

Session 1 Wed 10:30 – 13:00

Invited session organizer: Yossi Steinberg, Technion

*Invited papers are marked by an asterisk.

Session 2 Wed 14:30 – 15:50

Session 3 Wed 16:20 – 17:40

Session 4 Thu 10:30 – 12:35

Invited session organizer: Elza Erkip, Polytechnic Institute of NYU

Session 5 Thu 14:00 – 15:40

Session 6 Thu 16:10 – 17:50

Session 7 Fri 10:30 – 12:35

Invited session organizer: Pascal Vontobel, HP Labs

Session 8 Fri 14:00 – 16:05

Invited session organizer: David Gesbert, Eurecom

On information-theoretic secrecy for wireless networks

Etienne Perron, Suhas Diggavi and Emre Telatar
EPFL, Lausanne, Switzerland
Email: {etienne.perron,suhas.diggavi,emre.telatar}@epfl.ch

Abstract—**In this paper we summarize our recent work on information-theoretically secure wireless relay network communication. In such communication, the goal is to send information between two special nodes ("source" and "destination") in a (memoryless) network with authenticated relays, where the secrecy is with respect to a class of eavesdroppers. We develop achievable secrecy rates when authenticated relays also help increase secrecy rate by inserting noise into the network.**

I. Introduction

The seminal paper of Wyner [20] on the degraded wiretap channel and its generalization in [5] laid the foundations for information-theoretic secrecy in broadcast channels. In the recent past, information-theoretic secrecy has been applied to wireless networks with results on secrecy for MIMO broadcast channels, multiple access channels, interference channels and relay channels (see [10] and references therein). Cooperative strategies in wireless networks has been an active area of research (see [8] and references therein). In [16], [17] cooperative secrecy for *arbitrary* wireless networks was studied[1]. This work was inspired by recent (approximate) characterizations of the wireless relay network [3].

In this paper we summarize the studies in [16], [17], [14]. We will state the results for layered relay networks[2]. The main results are as follows. We first develop a "separable" strategy to provide information-theoretic secrecy for wireless networks, which operates on the principle of providing end-to-end secrecy, while the network operates without presupposing the secrecy requirement. This is developed for (layered) deterministic, Gaussian and discrete memoryless networks. We also develop a noise insertion strategy that allows a subset of the nodes in the network to insert random noise to aid in secure communication. We state the achievable secrecy rates for such active relay strategies, again for deterministic, Gaussian and discrete memoryless networks. We also state a simple outer bound for secrecy of such networks.

The paper is organized as follows. In Section II, we set up the problem and the notation. Some basic results for information flow without secrecy constraints are also established. We summarize the main results in Section III. We end with a short discussion in Section IV.

II. Preliminaries

We consider transmission over a relay network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$, where $\mathcal{V}$ is the set of vertices representing the communication nodes in the relay network and $\mathcal{L}$ is the set of annotated channels between the nodes, which describe the signal interactions. Note that these channels are not point-to-point links, rather, they model how the transmitted signals are superimposed and received at the receiving nodes (*i.e.,* there is broadcast and interference). We consider a special node $S \in \mathcal{V}$ as the source of the message which wants to securely communicate to another special node $D \in \mathcal{V}$ (the destination) with the help of a set of (authenticated) relay nodes $\mathcal{A} \subset \mathcal{V}$ in the network. We assume that a subset $\mathcal{B} \subseteq \mathcal{A}$ of the relay nodes is allowed to generate and use independent random messages. These special relay nodes are called "noise inserting" nodes. The secrecy is with respect to a set of possible (passive) eavesdropper nodes $\mathcal{E} \subset \mathcal{V}$ where $\mathcal{E}$ is disjoint from $\mathcal{A} \cup \{S, D\}$. We want to keep all or part of the message secret if any one of the possible eavesdropper nodes $E \in \mathcal{E}$ listens to the transmissions in the relay network. Note that the class of eavesdroppers that we define is discrete, *i.e.*, we assume that all possible eavesdroppers and their channels can be enumerated. If there is a continuum of possible eavesdropper channels, our model can approximate this via "quantization" of this continuum.

A. Signal interaction models

The results in this paper are stated for layered networks formally defined as follows.

Definition 1: A relay network is *layered* if for every (i, j) such that $i \in \{S\} \cup \mathcal{B}$ and $j \in \mathcal{V}$, all the paths from i to j have the same length (the same number of hops in $\mathcal{L}$). A *non-layered* network is a network in which at least one node pair (i, j) does not have this property.

Using the time-expanded networks, as used in [3], we can extend the results for layered networks to non-layered networks. The network we consider is constituted by (layered) memoryless channel interactions, which include broadcast and multiple access interference [4] in the following ways.

Wireless interaction model: In this well-accepted model [19], transmitted signals get attenuated by (complex) gains to which independent (Gaussian) receiver noise is added. More formally, the received signal y_j at node $j \in \mathcal{V}$ at time t is given by,

$$y_j[t] = \sum_{i \in \mathcal{N}_j} h_{ij} x_i[t] + z_j[t], \qquad (1)$$

[1] The case when a *single* relay node present, as an extension of the classical relay channel to the secrecy problem was studied in [13], [9].

[2] A layered network (given more precisely in Definition 1, is loosely that all paths from source to destination are the same length. As in [3], we can extend the results for layered networks to non-layered networks using time-expansion on the network.

where h_{ij} is the complex channel gain between node i and j which is the annotation of the channels in $\mathcal{L}$, x_i is the signal transmitted by node i, and $\mathcal{N}_j$ are the set of nodes that have non-zero channel gains to j. We assume that the average transmit power constraints for all nodes is 1 and the additive receiver Gaussian noise is of unit variance. We use the terminology *Gaussian wireless network* when the signal interaction model is governed by (1).

Deterministic interaction model: In [1], a simpler deterministic model which captures the essence of wireless interaction was developed. The advantage of this model is its simplicity, which gives insight to strategies for the noisy wireless network model in (1). The linear deterministic model of [1] simplifies the wireless interaction model in (1) by eliminating the noise and discretizing the channel gains through a binary expansion of q bits. Therefore, the received signal $y_j^{(d)}$ which is a binary vector of size q is modeled as

$$y_j^{(d)}[t] = \sum_{i \in \mathcal{N}_j} \mathbf{G}_{ij} x_i^{(d)}[t], \tag{2}$$

where $\mathbf{G}_{ij}$ is a $q \times q$ binary matrix representing the (discretized) channel transformation between nodes i and j and $x_i^{(d)}$ is the (discretized) transmitted signal. All operations in (2) are done over the binary field. We use the terminology *linear deterministic network* when the signal interaction model is governed by (2).

Discrete memoryless interaction model: The received signal y_j at node $j \in \mathcal{V}$ in layer l of the network, at time t is related to the inputs at time t through a DMC specified by, $p(y_j[t] | \{x_i[t]\}_{i \in \mathcal{N}_{l-1}})$, where $\mathcal{N}_{l-1}$ are the nodes in layer $l-1$.

To simplify the comparison between different results, we group the most important definitions below.

Definition 2: For $\mathcal{I} \subseteq \mathcal{V}$ and $j \in \mathcal{V}$, define $\Lambda(\mathcal{I}; j)$ to be the set of all cuts (Ω, Ω^c) that separate set $\mathcal{I}$ from j. More precisely, $\Lambda(\mathcal{I}; j)$ is the set of all $\Omega \subset \mathcal{V}$ such that $\mathcal{I} \subseteq \Omega$ and $j \in \Omega^c$.

Definition 3: For a (layered) relay network the transmit distribution $p(\{x_i\}_{i \in \mathcal{V}})$ and quantizers $p(\hat{y}_i | y_i)$, belong to the class $\mathcal{P}$ if for all $p \in \mathcal{P}$, we have

$$p = \left[\prod_{i \in \mathcal{V}} p(x_i)\right] p(\{y_j\}_{j \in \mathcal{V}} | \{x_i\}_{i \in \mathcal{V}}) \prod_{i \in \mathcal{V}} p(\hat{y}_i | y_i). \tag{3}$$

For given $\mathcal{I} \subseteq \mathcal{V}$ and $j \in \mathcal{V}$, we define an achievable rate between between $\mathcal{I}$ and j as

$$\hat{R}_{\mathcal{I};j}(p) \triangleq \min_{\Omega \in \Lambda(\mathcal{I};j)} \left[I(X_\Omega; \hat{Y}_{\Omega^c} | X_{\Omega^c}) - \sum_{i \in \Omega} I(Y_i; \hat{Y}_i | X_{\mathcal{V}}) \right] \tag{4}$$

where $X_{\mathcal{V}}$ are channel inputs, $Y_{\mathcal{V}}$ correspond to the channel outputs, and $\hat{Y}_{\mathcal{V}}$ are the quantized variables, all governed by $p \in \mathcal{P}$.

Definition 4: For a given transmit and quantization distribution $p \in \mathcal{P}$, a subset $\psi \subseteq \mathcal{V}$, a node $j \in \mathcal{E} \cup \{D\}$, define $\mathcal{R}_{\psi;j}(p)$ to be the set of all tuples $B_\psi = (B_i)_{i \in \psi}$ such that the components of the tuple are non-negative and such that for any subset $\mathcal{I} \subseteq \psi$, $\sum_{i \in \mathcal{I}} B_i \leq R_{\mathcal{I};j}(p)$, where the quantity $R_{\mathcal{I};j}(p)$ is the information-theoretic min-cut defined below,

$$R_{\mathcal{I};j}(p) \triangleq \min_{\Omega \in \Lambda(\mathcal{I};j)} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}). \tag{5}$$

Note that there is a difference between $\hat{R}_{\mathcal{I};j}(p)$ given in (4) and $R_{\mathcal{I};j}(p)$ given in (5), since $\hat{R}_{\mathcal{I};j}(p)$ is the achievable rate induced by a given (quantize-map-forward) relay strategy, whereas $R_{\mathcal{I};j}(p)$ is related to a cut-value, both evaluated for $p \in \mathcal{P}$.

Definition 5: For a given input and quantization distribution $p \in \mathcal{P}$, a subset $\psi \subseteq \mathcal{V} \setminus \{S\}$, and a node $j \in \mathcal{E} \cup \{D\}$, define $\hat{\mathcal{R}}_{\psi;j}(p)$ to be the set of all tuples $(B', B_\psi) = (B', (B_i)_{i \in \psi})$ such that the components of the tuple are non-negative and such that for any subset $\mathcal{I} \subseteq \psi$,

$$B' + \sum_{i \in \mathcal{I}} B_i \leq \hat{R}_{\mathcal{I} \cup \{S\};j}(p).$$

Note that for a given $\psi \subseteq \mathcal{V} \setminus \{S\}$, $\hat{\mathcal{R}}_{\psi;j}(p)$ differs from $\mathcal{R}_{\psi \cup \{S\};j}(p)$ in two ways. First, $\mathcal{R}_{\psi \cup \{S\};j}(p)$ is related to information-theoretic cut-values, evaluated for a particular $p \in \mathcal{P}$, and $\hat{\mathcal{R}}_{\psi;j}(p)$ is related to the achievable rate for a particular (quantization) relay strategy. Secondly, $\mathcal{R}_{\psi \cup \{S\};j}(p)$ imposes constraints for all subsets of ψ including those that do not contain S, *i.e.*, like a MAC region. In Definition 5 for $\hat{\mathcal{R}}_{\psi;j}(p)$, all the rate-constraints involve S.

Secrecy requirements:: The notion of information-theoretic secrecy is defined through the *equivocation* rate R_e, which is the residual uncertainty about the message when the observation of the strongest eavesdropper is given. More formally, [20], [5], given a (T, ϵ)-code, the equivocation rate is $\frac{1}{T} \min_{E \in \mathcal{E}} H(W | \mathbf{Y}_E)$, where W is the uniformly distributed source message, $\mathbf{Y}_E$ is the sequence of observations at eavesdropper E and $H(\cdot|\cdot)$ denotes the (conditional) entropy [4]. The "perfect" (weak) secrecy capacity is the largest transmitted information rate R, such that $R = R_e$ is achievable. This notion can be strengthened to *strong* perfect secrecy, if the equivocation is defined in bits $\min_{E \in \mathcal{E}} H(W | \mathbf{Y}_E)$, instead of a rate [12]. Using the tools developed in [12], we can convert all the results to strong secrecy, once we have proved it for weak secrecy (see also [14]).

B. Information flow over layered networks

Here we summarize results about communication in layered networks that form an ingredient to our main results on secrecy over relay networks. With no secrecy requirements, the transmission scheme is the same as developed in [3], and is informally described below.

Network operation:: Each node in the network generates codes independently using a distribution $p(x_i)$. The source S chooses a random mapping from messages $w \in \{1, \ldots, 2^{RT}\}$ to its transmit typical set $\mathcal{T}_{x_S}$, and therefore we denote by $\mathbf{x}_S^{(w)}, w \in \{1, \ldots, 2^{TR}\}$ as the possible transmit sequences for each message. Each received sequence $\mathbf{y}_i$ at node i is quantized to $\hat{\mathbf{y}}_i$ and this quantized sequence is randomly mapped onto a transmit sequence $\mathbf{x}_i$ using a random function $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$, which is chosen such that each quantized sequence is mapped uniformly at random to a transmit sequence. This random mapping can be represented by the following construction. Generate 2^{TR_i} sequences $\mathbf{x}_i$ from the distribution $\prod_j p(\mathbf{x}_i[j])$, and generate 2^{TR_i} sequences $\hat{\mathbf{y}}_i$ using a product distribution $\prod_j p(\hat{\mathbf{y}}_i[j])$. We denote the 2^{TR_i} sequences of $\hat{\mathbf{y}}_i$ as $\hat{\mathbf{y}}_i^{(k_i)}, k_i \in \{1, \ldots, 2^{TR_i}\}$. Note that standard rate-distortion theory tells us that we need $R_i > I(Y_i; \hat{Y}_i)$ for this quantization to be successful. Note that since the uniformly at random mapping produces $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$, for a quantized value of index k_i, we will denote it by $\hat{\mathbf{y}}_i^{(k_i)}$ and the sequence it is mapped to by $\mathbf{x}_i^{(k_i)} = f_i(\hat{\mathbf{y}}_i^{(k_i)})$.

In [3], this scheme was analyzed for deterministic and Gaussian networks. It was established that for deterministic networks, all rates up to $\min_{\Omega \in \Lambda(S:D)} H(Y_{\Omega^c}|X_{\Omega^c})$ for any product distribution of the nodes can be achieved. For linear deterministic networks, (2), this coincides with the cut-set outer bound. For Gaussian networks, an approximate max-flow, min-cut bound was established, which showed that all rates up to $\min_{\Omega \in \Lambda(S:D)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) - \kappa$, was achievable, where κ was a universal constant, independent of SNR and channel parameters [3].

In the multisource problem, a set of sources $\mathcal{S} \subset \mathcal{V}$ wish to communicate independent messages to the destination D over the network. Each of the relay nodes operate as above, except if it is also a source, then the transmitted sequence is a (uniform random) mapping of both its message and its received (quantized) signal. This scheme, which is a simple extension of the scheme studied in [3], was studied for the deterministic and Gaussian interaction models in [17], [14]. Its simple extension to (layered) memoryless networks is stated below.

Theorem 1: For any memoryless layered network, from a set of sources $\mathcal{S}$ to a destination D, we can achieve any rate vector satisfying

$$\sum_{k \in \mathcal{I} \subseteq \mathcal{S}} R_k \leq \hat{R}_{\mathcal{I};j}(p)$$

for some distribution $p \in \mathcal{P}$ defined in (3), where $\hat{R}_{\mathcal{I};j}(p)$ is defined in (4).

III. Main Results

Broadly there are a sequence of three (increasing generality) ideas to the achievability scheme. (i) *Separable scheme:* The relay network is operated as described in Section II-B, but the secrecy is induced by an end-to-end scheme overlaid on this. (ii) *Noise insertion:* In addition to the above operation, a subset of the authenticated relays insert independent messages (noise) which are intended to disrupt the eavesdropper and are not required to be decoded. (iii) *Auxiliary variables:* In addition to the above, the source prefixes an artificial multiuser channel in order to allow multiple auxiliary variables.

We will state the results in increasing generality in order to clarify and interpret the results. For the simplest case, where relay nodes operate using the quantize-map-forward strategy described in Section II-B, without regard to secrecy requirements, the end-to-end separable scheme achieves the following secrecy region.

Theorem 2: For a given distribution $p \in \mathcal{P}$ defined in (3), the (strong) perfect secrecy rate between the source S and destination D with respect to a class of eavesdroppers $\mathcal{E}$, with $\hat{R}_{S;D}(p)$ given in (4), is lower bounded as

$$\bar{C}_s \quad \geq \quad \hat{R}_{S;D}(p) - \max_{E \in \mathcal{E}} \min_{\Omega \in \Lambda_E} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}),$$

where the second term is evaluated for $p \in \mathcal{P}$.

Special cases of this result for deterministic and Gaussian networks was shown in [16]. In the deterministic case, as in [2], the relays do not quantize the inputs, but map-and-forward it. Therefore, for deterministic networks the perfect secrecy rate is $\min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) - \max_{E \in \mathcal{E}} \min_{\Omega \in \Lambda_E} H(Y_{\Omega^c}|X_{\Omega^c})$. In the Gaussian case, by using a quantizer that gets distortion equal to the noise variance (see [3]), $I(Y_i; \hat{Y}_i|X_{\mathcal{V}})$ is a constant (depending on the noise variance and not the channels), for every relay node i.

We can improve and generalize the result in Theorem 2 by using noise insertion at an arbitrary subset $\mathcal{B} \subset \mathcal{V}$. These independent messages are not needed to be decoded anywhere, but can be used to "jam" the eavesdroppers.

Definition 6: For an input distribution p, we define the following function:

$$F(p) = \max_{B_{\mathcal{B}} \in \cap_{E \in \mathcal{E}} \mathcal{R}_{\mathcal{B};E}(p)} \Big[\max_x \{x : (x, B_{\mathcal{B}}) \in \hat{\mathcal{R}}_{\mathcal{B};D}(p)\} \\ - \max_x \{x : (x, B_{\mathcal{B}}) \in \cup_{E \in \mathcal{E}} \mathcal{R}_{\mathcal{B} \cup \{S\};E}(p)\} \Big].$$

Theorem 3: The (strong) perfect secrecy for any (layered) relay network is lower bounded as

$$\bar{C}_s \geq \max_{p \in \mathcal{P}} F(p),$$

Basically the idea in Theorem 3 is that the noise insertion effectively creates virtual MAC regions (for the eavesdroppers and the legitimate receiver). The projection of the difference of these regions onto the source message rate yields the secrecy rate[3]. That is, the noise insertion "fills" up the eavesdropper rate region with "junk" information, thereby protecting the information. This notion can actually be formalized, as seen in [14]. Also note that this is a way to think of the wiretap channel [20], where all the junk information is at the source. The noise insertion just distributes the origins of the junk all over the network. This strategy was analyzed for deterministic and Gaussian networks in [14], and the above result is its simple generalization for memoryless networks.

In order to introduce auxiliary variables, we prefix an artificial memoryless channel in the sources, thereby modifying the channel law for the networks. Since this does not change the basic arguments for Theorem 3 (or its special case of

[3] A related strategy was developed in [9] for the Gaussian (single) relay channel where the relay forwarded Gaussian noise along with decoded information.

Theorem 2), we do not restate the result. Note that in this case the form of the secrecy rate is the same, except that we can also optimize over the choice of the artificial channels. This essentially would be generalization of the approach taken in [15] for the wiretap channel, to the case of relay networks. Also, following the program of [12] one can focus on showing results for weak secrecy, but (as mentioned earlier), using the techniques of [12] we can obtain it for strong secrecy (see [14] for more details).

The next result is a simple upper bound on the perfect secrecy rate for an arbitrary number of noise-inserting nodes presented in [17].

Theorem 4: For a single eavesdropper E,

$$R_s \leq \max_{p(\{x_i\}_{i\in\mathcal{V}})} \min_{\Omega\in\Lambda(\mathcal{SB},D)} I(X_\Omega; Y_{\Omega^c}|Y_E, X_{\Omega^c}), \qquad (6)$$

where, in contrast to Theorems 2 and 3, the maximization is not only over product distributions but over all possible $p(\{x_i\}_{i\in\mathcal{V}})$.

The statement of Theorem 4 is valid for any type of signal interaction, including noisy channels.

IV. Discussion

In this paper we have summarized some of our studies on a communication scenario with secrecy requirement for wireless relay networks. We attempt to model the uncertainty in the eavesdropper's wireless channel, by developing the secrecy rates for a class of eavesdropper channels. It is possible to interpret the secret message generated as secret key generation, and therefore we can use the techniques outlined in this paper to generate an unconditionally (strongly) secure key. One of the important open questions is to develop characterizations of secrecy rates over networks. To obtain such a characterization we need a matching converse stating that no scheme can do better. The outer bound developed in Theorem 4 is quite simple, and we need more sophisticated outer bounding techniques. Another important issue to address is the relevance of these results for wireless networks. In order to make them more applicable, we need to ensure robustness of these results to uncertainties in (network) channel knowledge and eavesdroppers. An interesting approach to addressing this might be the use of feedback. In the seminal paper [11], Maurer showed the surprising result that feedback can allow information-theoretic secrecy, even when the eavesdropper channel dominates that of the legitimate receiver. The use of feedback for network secrecy is a scarcely explored topic and one we believe is worth pursuing. Some preliminary results in this direction were presented in [18]. The recent results of [6], [7] have established strategies also for key-agreement between a set of nodes in a single-hop network. Therefore, we believe that robustness using feedback, is another promising research direction.

References

[1] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Illinois, USA, Sep. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[2] ——, "Wireless network information flow," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Illinois, USA, Sep. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[3] ——, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inform. Theory*, 2009, submitted, http://arxiv.org/abs/0906.5394.

[4] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, May 1978.

[6] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Transactions on Information Theory*, vol. 50, pp. 3047–3061, December 2004.

[7] ——, "Secrecy capacities for multiterminal channel models," *IEEE Transactions on Information Theory*, vol. 54, pp. 2437–2452, June 2008.

[8] G. Kramer, I. Maric, and R. Yates, *Cooperative Communications*. Foundations and Trends in Networking, 2006.

[9] L. Lai and H. E. Gamal, "The Relay-Eavesdropper Channel: Cooperation for Secrecy," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep. 2008.

[10] Y. Liang, H. V. Poor, and S. Shamai, *Information Theoretic Security*. Foundations and Trends in Communications and Information Theory, 2009.

[11] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[12] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," *EUROCRYPT, LNCS 1807*, pp. 351–368, 2000, Springer.

[13] Y. Oohama, "Relay channels with confidential messages," *IEEE Trans. Inform. Theory*, Nov. 2006, submitted.

[14] E. Perron, "Information-theoretic secrecy for wireless networks," Ph.D. dissertation, School of Computer and Communication Sciences, EPFL., Lausanne, Switzerland, August 2009, available from http://library.epfl.ch/theses/?nr=4476.

[15] E. Perron, S. Diggavi, and E. Telatar, "A multiple access approach for the compound wiretap channel," in *Proc. of the IEEE Inform. Theory Workshop*, Taormina, Italy, Oct. 2009.

[16] ——, "On cooperative wireless network secrecy," in *Proc. of IEEE Infocom*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1935–1943.

[17] ——, "On noise insertion strategies for wireless network secrecy," in *Proc. of the Information Theory and Applications Workshop*, San Diego, USA, Feb. 2009, pp. 77–84.

[18] E. Perron, S. N. Diggavi, and I. E. Telatar, "Wireless network secrecy with public feedback," in *46th Annual Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, USA, September 2008, pp. 753–760.

[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[20] A. Wyner, "The wire-tap channel," *Bell System Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.

Refined Edge-Cut Bounds for Network Coding

Gerhard Kramer
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089 USA
gkramer@usc.edu

Sadegh Tabatabaei Yazdi
Dept. Electrical Engineering
Texas A&M University
College Station, TX
sadegh@neo.tamu.edu

Serap A. Savari
Dept. Electrical Engineering
Texas A&M University
College Station, TX
savari@ece.tamu.edu

***Abstract*—Existing edge-cut bounds for network coding are refined to include different rate weights.**

Edge-Cut Bounds and Rate Weights

Cut-set bounds that partition network *nodes* are a standard method for understanding network information flow [1]. Several recent improvements of these methods focus instead on network *edges* [2]–[5]. One limitation of the new bounds is that they treat source messages "equally" in the sense that each source rate R_s has the same weight. We here outline a simple refinement that introduces variable rate weights. We focus on the methods of [2], [5] but our approach also applies to the methods described in [3], [4], for example.

Consider a network graph $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a set of vertices (or nodes) and $\mathcal{E}$ is a set of edges. Consider the derivation in [5, Appendix] where the first few steps give

$$\sum_{k\in\mathcal{S}_d} R_k \le \frac{1}{N} I(W_{\mathcal{S}_d}; Y^N_{\mathcal{E}_d} \mid Z^N_{\mathcal{E}_d^C} W_{\mathcal{S}_d^C}) \tag{1}$$

where $\mathcal{S}_d$ is a set of source indexes, R_k is the rate of message W_k, N is the number of clock-ticks, $\mathcal{E}_d$ is a subset of $\mathcal{E}$, $Y^N_{\mathcal{E}_d} = \{Y^N_e : e \in \mathcal{E}_d\}$ is the set of output streams corresponding to edges in $\mathcal{E}_d$, $\mathcal{E}_d^C$ is the complement of $\mathcal{E}_d$ in $\mathcal{E}$, $Z_{\mathcal{E}_d^C} = \{Z^N_e : e \in \mathcal{E}_d^C\}$ is the noise corresponding to edges not in $\mathcal{E}_d$, and $W_{\mathcal{S}_d^C} = \{W_s : s \in \mathcal{S}_d^C\}$. The derivation in [5, Appendix] continues and gives a single-letter bound of the form

$$\sum_{k\in\mathcal{S}_d} R_k \le I(X_{\mathcal{V}(\mathcal{E}_d)}; Y_{\mathcal{E}_d} \mid X_{\bar{\mathcal{V}}(\mathcal{E}_d)\cap s(\mathcal{S}_d)^C}) \tag{2}$$

where $\bar{\mathcal{V}}(\mathcal{E}_d)$ is the set of vertices in which the edges of $\mathcal{E}_d$ terminate, and $s(\mathcal{S}_d)$ is the set of vertices corresponding to the sources with indexes in $\mathcal{S}_d$. The next step requires optimizing the joint probability distribution of the network inputs X_v, $v \in \mathcal{V}$. For classical networks, the best inputs X_v are independent and the right-hand side of (2) is a sum of edge capacities. Observe that (2) has rate-weights of unity.

However, consider the classical network shown in Fig. 1 where message W_1 is destined for both nodes 2 and 3 (the variable $\hat{W}_1(v)$ is the estimate of W_1 at vertex v). Message W_1 must clearly use *twice* the network resources to reach these nodes as compared to W_2 reaching node 4, and this fact should be reflected in a bound of the form

$$2R_1 + R_2 \le C_{1,2} + C_{1,3} \tag{3}$$

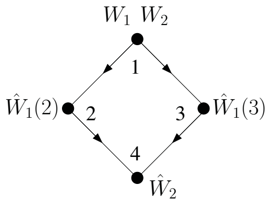


Fig. 1. Example network.

where C_e is the capacity of edge e. To prove that (3) is valid, consider $\mathcal{S}_d = \{1, 2\}$ and $\mathcal{E}_d = \{(1, 2), (1, 3)\}$, and follow the steps in [5] to arrive at the bound (1) which we expand as

$$\begin{aligned} R_1 + R_2 &\le \frac{1}{N} H(Y^N_{1,2} Y^N_{1,3}) \\ &= \frac{1}{N} \left[H(Y^N_{1,2}) + H(Y^N_{1,3}) - I(Y^N_{1,2}; Y^N_{1,3}) \right]. \end{aligned} \tag{4}$$

We now observe that W_1 must effectively be a function of both $Y^N_{1,2}$ and $Y^N_{1,3}$ and so it follows that $I(Y^N_{1,2}; Y^N_{1,3}) \ge NR_1$. Inserting this bound into (4) and using standard steps, we arrive at (3). Furthermore, this idea generalizes in several ways to strengthen the results of [2]–[5]. Some of these generalizations will be discussed in the presentation.

Acknowledgment

G. Kramer was supported by NSF Grant CCF-09-05235.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[2] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *J. Network and Systems Management*, vol. 14, no. 1, pp. 49–67, March 2006.

[3] N. J. A. Harvey, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2345–2364, June 2006.

[4] S. Thakor, A. Grant, and T. Chan, "Network coding capacity: a functional dependence bound," *Proc. 2009 IEEE Int. Symp. Inform. Theory*, Seoul, Korea, pp. 263–267, June 28-July 3, 2009.

[5] G. Kramer and S. A. Savari, "Capacity bounds for relay networks," *Proc. 2006 Workshop Inf. Theory Appl.,* UCSD Campus, La Jolla, CA, Feb. 6-10, 2006.

The Multiple Access Channel with Causal and Strictly Causal Side Information at the Encoders

Amos Lapidoth

Signal and Information Processing Laboratory
ETH Zürich
CH-8092 Zürich, Switzerland
amos.lapidoth@isi.ee.ethz.ch

Yossef Steinberg

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
ysteinbe@ee.technion.ac.il

***Abstract*—We study the state-dependent multiple access channel (MAC) with causal side information at the encoders. We consider two general models. In the first model, the state sequence is available at the two encoders in a strictly causal manner. We derive an achievable region, which is tight for the special case of a Gaussian MAC where the state sequence comprises the channel noise. Although the two senders do not have access to each other's massage and no feedback is present, the capacity for this case coincides with the capacity of the same MAC without side information, but with full cooperation between the users. A Schalkwijk-Kailath type algorithm is developed, which achieves this capacity with a double exponential decay of the maximal probability of error. In the second model we consider, the state sequence is available, as in Shannon's model, to the two encoders in a causal manner. A simple extension of the previous result, with the inclusion of Shannon strategies, yields an achievability result for this problem.**

***Index Terms*—Causal state information, feedback, multiple access channel, strictly-causal state-information.**

I. Introduction

The problem of coding for state dependent channels with state information at the encoder has been studied extensively in two main scenarios: causal state information and noncausal state information. The case where the state is available in a strictly-causal manner or with a given fixed delay, has not attracted much attention, possibly because in single-user channels, strictly-causal state-information (SI) does not increase capacity. However, like feedback, strictly-causal SI can be beneficial in multiple user channels. This can be seen using the examples of Dueck [1]. Specifically, Deuck constructs an additive noise broadcast channel (BC), where the noise is common to the two users. The input and additive noise are defined in a way that the resulting BC is not degraded. The encoder learns the channel noise via the feedback, and transmits it to the two users. Although valuable rate–that otherwise could be used to transmit user messages–is spent on the transmission of the noise, the net effect is an increase in channel capacity, due to the noise being common to both users. In Dueck's example, the noise is transmitted losslessly to the two users. However, based on his observations, it is straightforward to construct examples where only lossy transmission of the noise is possible, and yet the capacity region is increased by this use of feedback. There is only one encoder in the BC and thus, identifying the additive noise as channel state, feedback in Dueck's example is equivalent to knowledge of the state in a strictly causal manner.

In this paper we study the state-dependent multiple access channel (MAC) with common state information at the encoders. Two main models are considered: the strictly causal model, where at time i both encoders have access to a common state sequence up to time $i-1$ (or possibly with larger delay), and the causal model, in the spirit of Shannon [4], where at time i both encoders have access to a common state sequence up to (and including) time i.

As in the case of broadcast channels, strictly causal knowledge of the state increases the MAC's capacity. Since only past (or delayed) samples of the state are known, neither binning (as in Gel'fand and Pinsker's channel [2]) nor strategies [4] can be employed. Instead, we derive a general achievable region based on a block-Markov coding scheme. The encoders, having access to a common state sequence, compress and transmit it to the decoder. The users cannot establish cooperation in the transmission of the messages, but they do cooperate in the transmission of the compressed state, thus increasing the achievable rates. The resulting region is tight for the Gaussian MAC where the state comprises the channel noise. Specifically, it is shown that for this channel, a proper choice of the random variables in our achievable region yields the capacity region of the same MAC without side information but with full cooperation between the encoders. Since strictly causal state information does not increase the capacity of single user channels, it also cannot increase the capacity of the MAC with full cooperation. Consequently, full cooperation is the best that one can hope for, and thus the region must be tight. Although the users do not have access to each other's message and no feedback is available, a Schalkwijk-Kailath type algorithm can be devised for this channel, yielding the full cooperation region with a double exponential decay in the probability of error. The general achievability result, and the Schalkwijk-Kailath algorithm, make use of all the past samples of the channel noise. It turns out, however, that much less is needed to achieve the full cooperation region. Assume that, instead of having all the past noise samples, only S_1 and S_2 are known to the encoders, in a strictly causal manner (i.e., available at times 3,4,....). As demonstrated in Section II, althought only two noise samples are known, the full cooperation region can still be achieved.

The causal model is also treated with block Markov coding, but the transmission at time i can depend on the state S_i. All other ingredients of the coding scheme remain intact. The resulting achievable region contains the naive region, which uses Shannon strategies for the MAC without block Markov coding, with the inclusion being in some cases strict.

II. Problem formulation and main results

A. Basic definitions

We are given a discrete memoryless state-dependent MAC $P_{Y|S,X_1,X_2}$ with state alphabet $\mathcal{S}$, state probability mass function (PMF) P_S, input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and output alphabet $\mathcal{Y}$. Sequences of letters from $\mathcal{S}$ are denoted by $s^n = (s_1, s_2, \dots, s_n)$ and $s_i^j = (s_i, s_{i+1} \dots, s_j)$. Similar notation holds for all alphabets, e.g. $x_1^n = (x_{1,1}, x_{1,2}, \dots, x_{1,n})$, $x_{2,i}^j = (x_{2,i}, x_{2,i+1}, \dots, x_{2,j})$. When there is no risk of ambiguity, n-sequences will sometimes be denoted by boldface letters, $\boldsymbol{y}$, $\boldsymbol{x}_1$, etc. The laws governing n sequences of state and output letters are given by

$$P^n_{Y|S,X_1,X_2}(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{x}_1, \boldsymbol{x}_2) = \prod_{i=1}^{n} P_{Y|S,X_1,X_2}(y_i|s_i, x_{1,i}, x_{2,i}),$$

$$P^n_S(\boldsymbol{s}) = \prod_{i=1}^{n} P_S(s_i).$$

For notational convenience, we henceforth omit the superscript n, and we denote the channel by P. Let $\phi_k \colon \mathcal{X}_k \to [0, \infty)$, $k = 1, 2$, be single letter cost functions. The cost associated with the transmission of sequence $\boldsymbol{x}_k$ at input k is defined as

$$\phi_k(\boldsymbol{x}_k) = \frac{1}{n} \sum_{i=1}^{n} \phi_k(x_{k,i}), \quad k \in \{1, 2\}.$$

B. The strictly causal model

Definition 1: Given positive integers M_1, M_2, let $\mathcal{M}_1$ be the set $\{1, 2, \dots, M_1\}$ and similarly, $\mathcal{M}_2$ the set $\{1, 2, \dots, M_2\}$. An $(n, M_1, M_2, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal side information at the encoders is a pair of sequences of encoder mappings

$$f_{k,i} \colon \mathcal{S}^{i-1} \times \mathcal{M}_k \to \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \dots, n \qquad (1)$$

and a decoding map

$$g \colon \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$$

such that the input cost costs are bounded by Γ_k

$$\phi_k(\boldsymbol{x}_k) \le \Gamma_k, \quad k = 1, 2,$$

and the average probability of error is bounded by ϵ

$$\begin{aligned} P_e &= 1 - \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \sum_{\boldsymbol{s}} P_S(\boldsymbol{s}) \cdot \\ &\quad P\left(g^{-1}(m_1, m_2)|\boldsymbol{s}, \boldsymbol{f}_1(\boldsymbol{s}, m_1), \boldsymbol{f}_2(\boldsymbol{s}, m_2)\right) \le \epsilon, \end{aligned}$$

where $g^{-1}(m_1, m_2) \subset \mathcal{Y}^n$ is the decoding set of the pair of messages (m_1, m_2), and

$$\boldsymbol{f}_k(\boldsymbol{s}, m_k) = (f_{k,1}(m_k), f_{k,2}(s_1, m_k), \dots, f_{k,n}(s^{n-1}, m_k)).$$

The rate pair (R_1, R_2) of the code is defined as

$$R_1 = \frac{1}{n} \log M_1, \qquad R_2 = \frac{1}{n} \log M_2.$$

A rate-cost quadruple $(R_1, R_2, \Gamma_1, \Gamma_2)$ is said to be achievable if for every $\epsilon > 0$ and sufficiently large n there exists an $(n, 2^{nR_1}, 2^{nR_2}, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal side information for the channel $P_{Y|S,X_1,X_2}$. The capacity-cost region of the channel with strictly causal SI is the closure of the set of all achievable quadruples $(R_1, R_2, \Gamma_1, \Gamma_2)$, and is denoted by $\mathcal{C}_{sc}$. For a given pair (Γ_1, Γ_2) of input costs, $\mathcal{C}_{sc}(\Gamma_1, \Gamma_2)$ stands for the section of $\mathcal{C}_{sc}$ at (Γ_1, Γ_2). Our interest is in characterizing $\mathcal{C}_{sc}(\Gamma_1, \Gamma_2)$.

Let $\mathcal{P}_{sc}$ be the collection of all random variables (U, V, X_1, X_2, S, Y) whose joint distribution satisfies

$$P_{U,V,X_1,X_2,S,Y} = P_S P_{X_1|U} P_{X_2|U} P_U P_{V|S} P_{Y|S,X_1,X_2}. \quad (2)$$

Note that (2) implies the Markov relations $X_1 \ominus U \ominus X_2$ and $V \ominus S \ominus Y$, and that the triplet (X_1, U, X_2) is independent of (V, S). Let $\mathcal{R}_{sc}$ be the convex hull of the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, U, V) && (3) \\ R_2 &\le I(X_2; Y|X_1, U, V) && (4) \\ R_1 + R_2 &\le I(X_1, X_2; Y|U, V) && (5) \\ R_1 + R_2 &\le I(X_1, X_2, V; Y) - I(V; S) && (6) \\ \Gamma_k &\ge \mathbb{E}\phi_k(X_k), \qquad k = 1, 2 \end{aligned}$$

for some $(U, V, X_1, X_2, S, Y) \in \mathcal{P}_{sc}$. Our main result for the strictly causal case is the following.

Theorem 1: $\mathcal{R}_{sc} \subseteq \mathcal{C}_{sc}$.

The proof is based on a scheme where a lossy version of the state is conveyed to the decoder using Wyner-Ziv compression [7] and block-Markov encoding for the MAC with common message [5]. The proof is omitted. In some cases, the region $\mathcal{R}_{cs}$ coincides with $\mathcal{C}_{cs}$. The next example is such a case. Although Theorem 1 is proved for the discrete memoryless case, we apply it here for the Gaussian model. Extension to continuous alphabets can be done as in [6].

Example 1: Consider the Gaussian MAC, with input power constraints $\mathbb{E}X_k^2 \le \Gamma_k$, $k = 1, 2$, where the state comprises the channel noise:

$$Y = X_1 + X_2 + S, \qquad S \sim N(0, \sigma_s^2). \qquad (7)$$

The capacity region of this channel, when S is known strictly causally at the two encoders, is the collection of all pairs (R_1, R_2) satisfying

$$R_1 + R_2 \le \frac{1}{2} \log \left(1 + \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma_s^2} \right). \qquad (8)$$

The region (8) is the capacity region of the same MAC when S is not known to any of the encoders, but with full cooperation between the users—a situation equivalent to a single user channel with a vector input constraint. Since strictly causal SI does not increase the capacity of a single user channel, we only have to show achievability in (8). We show it by properly choosing the random variables in (3)–(6). Let us first examine the maximal R_1. Set $U = X_2$, and let X_1, X_2, V, S be zero mean, jointly Gaussian, with X_1, X_2 independent of V, S. Then (3)-(6) reduce to the two bounds on R_1:

$$R_1 \leq \frac{1}{2}\log\left(\frac{\sigma^2_{x_1|x_2} + \sigma^2_{s|v}}{\sigma^2_{s|v}}\right) \tag{9}$$

$$R_1 \leq \frac{1}{2}\log\left(\frac{\Gamma_\Sigma + \sigma^2_s}{\sigma^2_s}\right) \tag{10}$$

where $\sigma^2_{x_1|x_2}$ is the variance of X_1 conditioned on X_2; $\sigma^2_{s|v}$ is analogously defined; and Γ_Σ is the power of the sum X_1+X_2. In full cooperation, $\Gamma_\Sigma = (\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2$, but then $\sigma^2_{x_1|x_2} = 0$, which nullifies the right hand side of (9). Note, however, that we can take the limit $\sigma^2_{s|v} \to 0$ without effecting (10). Thus we can approach the full cooperation rate as closely as desired, by first reducing $\sigma^2_{s|v}$, so that the right hand side of (9) is kept high, and then reducing $\sigma^2_{x_1|x_2}$. This proves that with $R_2 = 0$, the rate

$$R_1 = \frac{1}{2}\log\left(1 + \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma^2_s}\right) \tag{11}$$

is achievable. By symmetry and time sharing, (8) is achievable.

We next describe a Schalkwijk-Kailath type algorithm [3], that achieves the same rate, with double exponential decay in the maximal probability of error. As with the proof of (8), first the achievability of (11) is shown. The rest will follow by symmetry and time sharing. Split the interval $[0, 1]$ into M_1 equally spaced sub-intervals. Let θ_1 be the center of one of these sub-intervals, representing the message of user 1, as in [3]. At the first time instance, the users transmit

$$X_{1,1} = \theta_1, \quad X_{2,1} = 0. \tag{12}$$

The corresponding channel output is

$$Y_1 = \theta_1 + S_1. \tag{13}$$

Starting from time instance $i = 2$ and on, the noise sample S_1 is known at both encoders. Thus the two encoders now cooperate to transmit S_1 to the decoder. Since they now have a common message to transmit, knowing the states in a strictly causal manner is equivalent to feedback. Applying the same algorithm as in [3], after n iterations the receiver constructs a Maximum Likelihood estimate of S_1, denoted $\hat{S}_1^{(n)}$, whose error satisfies

$$\mathbb{E}((S_1 - \hat{S}_1^{(n)})^2|S_1) = \frac{\sigma^2_s}{(\alpha^2)^n} \tag{14}$$

with

$$\alpha^2 = 1 + \gamma^2, \quad \gamma^2 = \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma^2_s}. \tag{15}$$

Based on this estimate and on Y_1, the decoder can now construct an estimate $\hat{\theta}_1^{(n)}$ of θ_1

$$\hat{\theta}_1 = Y_1 - \hat{S}_1^{(n)} \tag{16}$$

whose error satisfies

$$\mathbb{E}((\theta_1 - \hat{\theta}_1^{(n)})^2|S_1) = \frac{\sigma^2_s}{(\alpha^2)^n}. \tag{17}$$

Therefore, choosing $M_1 = nr_1$, with $r_1 \leq \frac{1}{2}\log(1 + \gamma^2)$, the probability of error vanishes doubly exponentially as $n \to \infty$. This proves that (11) is achievable. By symmetry, and applying time sharing, this algorithm achieves the region (8).

We next show that the region (8) is achievable also when the states at only two time instances, say S_1 and S_2, are known from time $i = 3$ and on. It suffices to show achievability of (11) with only S_1 known, from time $i = 2$. First transmissions and output are given by (12), (13). At times $i = 2$ and on, both users know S_1. They cooperate in transmitting to the receiver a quantized version of S_1 via a regular code for the single user Gaussian channel. Specifically, fix $\epsilon > 0$ and choose β such that $P_S(|S_1| > \beta) \leq \epsilon$. Define $r_1 = \frac{1}{2}\log(1 + \gamma^2)$. We employ two partitions. First, partition the interval $[0, 1]$ into $M_1 = 2^{nr_1}/(4\beta)$ sub intervals, where the centers represent the messages of user 1. Let θ_1 be the center of one of these sub intervals. Partition the interval $[-\beta, \beta]$ into 2^{nr_1} sub intervals, and denote by $m_q, q = 1, 2, \ldots, 2^{nr_1}$ their center points. Define

$$S_q = \arg\min_{m_q} |S_1 - m_q| \tag{18}$$

The two users transmit S_q to the receiver, via a single user code. Denote by $\hat{S}_q$ the receiver's estimate of S_q. Clearly, for n large enough,

$$P\left(\left|S_1 - \hat{S}_q\right| \geq \frac{2\beta}{(1 + \gamma^2)^{n/2}}\right) \leq 2\epsilon \tag{19}$$

implying that the receiver can detect θ_1 with probability of error not exceeding 2ϵ. Note that M_1 provides the claimed rate.

C. The causal model

The definition of codes and achievable rates remain as in Section II-B, with the only difference being the definition of encoding maps: in the causal case (1) is replaced by

$$f_{k,i} \colon \mathcal{S}^i \times \mathcal{M}_k \to \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \ldots, n. \tag{20}$$

The capacity region and its section at (Γ_1, Γ_2) are denoted by $\mathcal{C}_c$ and $\mathcal{C}_c(\Gamma_1, \Gamma_2)$, respectively. Let $\mathcal{P}_c$ be the collection of all random variables $(U, U_1, U_2, V, X_1, X_2, S, Y)$ whose joint distribution can be written as

$$P_U P_{U_1|U} P_{U_2|U} P_{V|S} P_S P_{X_1|U,U_1,S} P_{X_2|U,U_2,S} P_{Y|S,X_1,X_2}. \tag{21}$$

Observe that (21) implies the Markov relations $U_1 \multimap U \multimap U_2$ and $V \multimap S \multimap Y$, and that the triple (U_1, U, U_2) is independent of (V,S). Let $\mathcal{R}_c$ be the convex hull of the collection of all $(R_1,R_2,\Gamma_1,\Gamma_2)$ satisfying

$$\begin{aligned} R_1 &\leq I(U_1;Y|U_2,U,V) && (22)\\ R_2 &\leq I(U_2;Y|U_1,U,V) && (23)\\ R_1+R_2 &\leq I(U_1,U_2;Y|U,V) && (24)\\ R_1+R_2 &\leq I(U_1,U_2,V;Y)-I(V;S) && (25)\\ \Gamma_k &\geq \mathbb{E}\phi_k(X_k), \qquad k=1,2 \end{aligned}$$

for some $(U,U_1,U_2,V,X_1,X_2,S,Y)\in\mathcal{P}_c$. Our main result for the causal case is the following.

Theorem 2: $\mathcal{R}_c \subseteq \mathcal{C}_c$.

The proof proceeds along the lines of the proof of Theorem 1, except that the inputs X_k, $k=1,2$, are allowed to depend on the state S, and that additional external random variables U_1 and U_2 that do not depend on S are introduced. This resembles the situation in coding for the single user channel with causal side information, where a random Shannon strategy can be represented by an external random variable independent of the state. The proposed scheme outperforms the naive approach of using strategies without block Markov encoding of the state. This latter naive approach leads to the region comprising all (R_1,R_2) satisfying

$$\begin{aligned} R_1 &\leq I(T_1;Y|T_2,Q)\\ R_2 &\leq I(T_2;Y|T_1,Q)\\ R_1+R_2 &\leq I(T_1,T_2;Y|Q) && (26) \end{aligned}$$

for some $P_Q P_{T_1|Q} P_{T_2|Q}$, where T_k are random Shannon strategies [4], whose realizations are mappings $t_k\colon \mathcal{S}\to\mathcal{X}_k$, $k=1,2$; Q is a time sharing random variable; and

$$P_{Y|T_1,T_2}(y|t_1,t_2)=\sum_{s\in\mathcal{S}} P_S(s)P_{Y|S,X_1,X_2}(y|s,t_1(s),t_2(s)).$$

Clearly $\mathcal{R}_c$ contains the region of the naive approach as we can choose V in (22)–(25) to be a null random variable. The next example demonstrates that the inclusion can be strict.

Example 2: Noiseless binary MAC, with input selector. Consider the noiseless binary MAC where $\mathcal{X}_1=\mathcal{X}_2=\mathcal{Y}=\{0,1\}$, $\mathcal{S}=\{1,2\}$ and $P_S(S=2)=p$ for some $p>0.5$. The state S determines which of the two inputs is connected to the output:

$$Y=X_S.$$

Block Markov Coding. Both users know the state and hence know, at each time, which user is connected to the output. Thus, they can compress the state using $H(S)=H_b(p)$ bits per channel use and transmit the state sequence to the decoder, via block Markov coding. If they do so, the decoder knows S, and the users can now share between them a clean channel. Since they already spent $H_b(p)$ bits in transmitting the state, the net rate remaining to share between them is

$$R_1+R_2=1-H_b(p). \qquad (27)$$

Note, however, that not all the line (27) is achievable. The users do not know each other's message. Thus, user 1 can transmit its own message only $(1-p)$ fraction of the time. We conclude that the following rate is achievable for user 1:

$$R_1=[1-H_b(p)](1-p) \quad \text{[bits]}. \qquad (28)$$

The Naive Approach. From the region (26) and the extreme points of the capacity region of the classical MAC, the maximal rate that user 1 can transmit is:

$$R_1=\max I(T_1;Y|T_2=t_2), \qquad (29)$$

where the maximum is over the distribution of T_1 and over all mappings $t_2\colon\mathcal{S}\to\mathcal{X}_2$. The strategy t_2 influences the output only when $S=2$, in which case it gives a certain input X_2, connected directly to the output. User 1 is then disconnected. Therefore, the exact value of t_2 is immaterial. Assume that $t_2(s=2)=0$.

Similarly, t_1 influences the output only when $S=1$, in which case it gives a certain input X_1 directly connected to the output. Since the strategies are chosen independently of S, the MAC reduces to a Z-channel from user 1:

$$P(Y=0|X=0)=1, \quad P(Y=0|X=1)=p. \qquad (30)$$

The capacity of this channel is given by

$$C(p)=\log_2\left(1+(1-p)p^{\frac{p}{1-p}}\right) \qquad \text{[bits]}. \qquad (31)$$

At the limit where p approaches 1, we have

$$C(p)\approx(1-p)e^{-1}\log_2 e\approx 0.53(1-p), \qquad (32)$$

which, at the limit $p\to 1$, is strictly less than (28).

References

[1] G. Dueck, “Partial feedback for two-way and broadcast channels,” *Inf. Contr.*, vol. 46, pp. 1–15, 1980.

[2] S. I. Gel'fand and M. S. Pinsker, “Coding for channel with random parameters,” *Probl. Inform. & Control*, vol. 9, no. 1, pp. 19–31, 1980.

[3] J. P. M. Schalkwijk, “A coding scheme for additive noise channels with feedback—Part II: band-limited signals,” *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, April 1966

[4] C. Shannon, “Channels with side information at the transmitter,” *IBM J. Res. Devel.*, vol. 2, pp.289–293, 1958.

[5] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell System technical Journal*, vol. 52, pp. 1037–1076, Sept. 1973.

[6] A. D. Wyner, “The rate-distortion function for source coding with side infomation at the decoder—II: general sources,” *Inf. Contr.*, vol. 38, pp. 60–80, 1978.

[7] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

On the Binary Symmetric Wiretap Channel

Yanling Chen
Q2S, Centre of Excellence
Norwegian University of Science and Technology
7491, Trondheim, Norway
Email: yaning.chen@q2s.ntnu.no

A. J. Han Vinck
Institute for Experimental Mathematics
Univeristy of Duisburg-Essen
45326, Essen, Germany
Email: vinck@iem.uni-due.de

***Abstract*—In this paper, we investigate the binary symmetric wiretap channel. We show that the *secrecy capacity* can be achieved by using *random linear codes*. Furthermore, we explore the *coset coding scheme* constructed by linear codes. As a result, we give an upper bound on the total *information loss*, which sheds light on the design of the applicable coset codes for the secure transmission with limited information leakage.**

I. Introduction

The concept of the wiretap channel was first introduced by Wyner [1]. His model is a form of degraded broadcast channel. Assume that the wiretapper knows the encoding scheme used at the transmitter and the decoding scheme used at the legitimate receiver. The objective is to maximize the rate of reliable communication from the source to the legitimate receiver, subject to the constraint that the wiretapper learns as little as possible about the source output. In fact, there is a maximum rate, above which secret communication between the transmitter and the legitimate receiver is impossible. Wyner [1] has determined this *secrecy capacity* when both main channel and the wiretap channel are discrete memoryless.

In this paper, we focus on the problem of developing a forward coding scheme for provably secure, reliable communication over a wiretap channel. Basic idea has been introduced by Wyner in [1] for the special case when the main channel is noiseless and the wiretap channel is a binary symmetric channel (BSC). Another example is given by Thangaraj et al. [3] for the case with a noiseless main channel and a binary erasure wiretap channel. In this paper, we consider the specific case when both the main channel and the wiretap channel are BSCs. Our main contribution is twofold. We start with a random coding scheme similar to the one proposed in [4]. We give a strict mathematical proof to show that the secrecy capacity can be achieved by using random linear codes. Furthermore, we address the coset code constructed by linear codes and analyze its information leakage. We derive an upper bound on the total information loss and show that under certain constraint one can construct a coset code to insure a secure transmission with limited information leakage.

II. Model description

We consider the communication model as shown in Fig. 1. Suppose that all alphabets of the source, the channel input and the channel output are equal to $\{0,1\}$. The main channel is a BSC with crossover probability p and we denote it by BSC(p). The wiretap channel is a BSC(p_w), where $0 \leq p < p_w \leq 1/2$. Note that a BSC(p_w) is equivalent to the concatenation of a BSC(p) and a BSC(p^*), where $p^* = (p_w - p)/(1-2p)$. Thus the channel model shown in Fig. 1 is equivalent to Wyner's model with a BSC(p) main channel and a BSC(p^*) wiretap channel. Its secrecy capacity due to [1] is $C_s = h(p_w) - h(p)$.

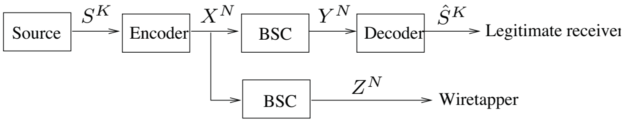


Fig. 1. Binary symmetric wiretap channel.

To transmit a K-bit secret message S^K, an N-bit codeword X^N is sent to the channel. The corresponding output at the legitimate receiver is Y^N, at the wiretapper is Z^N. Thus the error occurred over the main channel is $E^N = Y^N - X^N$, over the wiretap channel is $E_w^N = Z^N - X^N$. Assume that S^K is uniformly distributed. The *transmission rate* to the legitimate receiver is

$$R = K/N. \tag{1}$$

The *equivocation* of the wiretapper is defined to be

$$d = \frac{H(S^K|Z^N)}{H(S^K)} = \frac{H(S^K|Z^N)}{K}. \tag{2}$$

At the legitimate receiver, on receipt of Y^N, the decoder makes an estimate $\hat{S}^K$ of the message S^K. The *error probability* P_e of decoding is defined to be

$$P_e = \Pr\{\hat{S}^K \neq S^K\}. \tag{3}$$

We refer to the above as an encoder-decoder (K, N, d, P_e).

In this paper, when the dimension of a sequence is clear from the context, we will denote the sequences in boldface letters for simplicity. For example, $\mathbf{x}$ is the sequence x^N and $\mathbf{s}$ is s^K, etc. A similar convention applies to random variables, which are denoted by upper-case letters.

III. Secrecy capacity achieving codes

In this section, we perform a random linear code to establish the achievability of the secrecy capacity. For this aim, we need to construct an encoder-decoder (K, N, d, P_e) such that for arbitrary $\varepsilon, \zeta, \delta > 0$,

$$R \geq h(p_w) - h(p) - \varepsilon, \quad d \geq 1 - \zeta, \quad P_e \leq \delta. \tag{4}$$

A. Parameter settings

First, we set up the parameters for the encoder-decoder (K, N, d, P_e). Randomly choose a binary matrix H_1 with $N - K_1$ rows and N columns. Independently and randomly choose another binary matrix H with K rows and N columns. Assume that $K \leq K_1$ and let $K_2 = K_1 - K$. We construct

$$\mathrm{H}_2 = \left[\begin{array}{c} \mathrm{H}_1 \\ \mathrm{H} \end{array} \right]. \tag{5}$$

Then H_2 is a binary matrix with $N-K_2$ rows and N columns. For arbitrary small $\epsilon > 0$, we take

$$K_1 = \lfloor N[1 - h(p) - 2\epsilon] \rfloor;$$
$$K_2 = \lfloor N[1 - h(p_w) - 2\epsilon] \rfloor.$$

Here $\lfloor x \rfloor$ stands for the maximal integer $\leq x$. For given $\varepsilon > 0$, let $N_0 > 1/\varepsilon$. It is easy to verity that for $N > N_0$, we have

$$R = K/N \geq h(p_w) - h(p) - \varepsilon. \tag{6}$$

In what follows, we will assume that H_1, H and H_2 are of full rank. The reason is due to Lemma 6 in [2]. In order to send a secret message $\mathbf{s}$, a sequence $\mathbf{x}$ is chosen at random from the solution set of the following equation

$$\mathbf{x}{\mathrm{H}_2}^{\mathrm{T}} = \left[\begin{array}{cc} \mathbf{x}{\mathrm{H}_1}^{\mathrm{T}} & \mathbf{x}\mathrm{H}^{\mathrm{T}} \end{array} \right] = \left[\begin{array}{cc} \mathbf{0} & \mathbf{s} \end{array} \right], \tag{7}$$

where ${\mathrm{H}_2}^{\mathrm{T}}, {\mathrm{H}_1}^{\mathrm{T}}$ and H^{T} are the transposes of the matrices $\mathrm{H}_2, \mathrm{H}_1$ and H, respectively.

In the following, we will show that the secrecy capacity can be achieved by the random linear codes in two parts, the reliability: $P_e \to 0$ as $N \to \infty$; and the security: $d \to 1$ as $N \to \infty$.

B. Reliability proof

In this subsection, we will prove that $P_e \to 0$ as $N \to \infty$.

The legitimate receiver uses typical set decoder. The decoder examines the typical set $T_E^N(\epsilon)$, the set of error sequences $\mathbf{e}$ that satisfy

$$2^{-N[h(p)+\epsilon]} \leq \Pr(\mathbf{E} = \mathbf{e}) \leq 2^{-N[h(p)-\epsilon]}. \tag{8}$$

If exactly one sequence $\mathbf{e}$ satisfies $\mathbf{e}{\mathrm{H}_1}^{\mathrm{T}} = \mathbf{y}{\mathrm{H}_1}^{\mathrm{T}}$, the typical set decoder reports it as the hypothesized error sequence. Otherwise, the typical decoder reports an error.

The error probability of the typical set decoder at the legitimate receiver, can be written as follows,

$$P_e = P_T + P_{\mathrm{H}_1}, \tag{9}$$

where P_T is the probability that the true error sequence is itself not typical, and P_{H_1} is the probability that the true error sequence is typical and at least one other typical sequence clashes with it.

We first analyze P_T. For given $\epsilon, \delta > 0$, there exists N_1, such that $\Pr\{\mathbf{e} \in T_E^N(\epsilon)\} \geq 1 - \delta/2$ for $N \geq N_1$. Therefore, when $N \geq N_1$, $P_T = 1 - \Pr\{\mathbf{e} \in T_E^N(\epsilon)\} \leq \delta/2$.

Now we consider P_{H_1}. Suppose that the true error sequence is $\mathbf{e} \in T_E^N(\epsilon)$. If any of the typical error sequence $\mathbf{e}' \neq \mathbf{e}$, satisfies $(\mathbf{e}' - \mathbf{e}){\mathrm{H}_1}^{\mathrm{T}} = \mathbf{0}$, then we have an error. Let

$$T_{\mathbf{e}}(\epsilon) = \{\mathbf{e}' : \mathbf{e}' \in T_E^N(\epsilon), \mathbf{e}' \neq \mathbf{e}\}. \tag{10}$$

We have

$$P_{\mathrm{H}_1} \leq \sum_{\mathbf{e} \in T_E^N(\epsilon)} \Pr(\mathbf{E} = \mathbf{e}) \sum_{\mathbf{e}' \in T_{\mathbf{e}}(\epsilon)} \mathbf{1}[(\mathbf{e}' - \mathbf{e}){\mathrm{H}_1}^{\mathrm{T}} = \mathbf{0}],$$

where $\mathbf{1}[\cdot]$ is the truth function, whose value is 1 if the statement in the bracket is true and 0 otherwise.

Consider the average of P_{H_1}, $\bar{P}_{\mathrm{H}_1}$, over all possible H_1. Denote averaging over all possible H_1 by $\langle \cdot \rangle_{\mathrm{H}_1}$. We have

$$\bar{P}_{\mathrm{H}_1} \leq \sum_{\mathbf{e} \in T_E^N(\epsilon)} \Pr(E^N = \mathbf{e}) \sum_{\mathbf{e}' \in T_{\mathbf{e}}(\epsilon)} \langle \mathbf{1}[(\mathbf{e}' - \mathbf{e}){\mathrm{H}_1}^{\mathrm{T}} = \mathbf{0}] \rangle_{\mathrm{H}_1}.$$

Since for any non-zero binary sequence $\mathbf{v}$, the probability that $\mathbf{v}{\mathrm{H}_1}^{\mathrm{T}} = \mathbf{0}$, averaging over all possible H_1, is $2^{-(N-K_1)}$, so

$$\bar{P}_{\mathrm{H}_1} < |T_E^N(\epsilon)| 2^{-(N-K_1)} \leq 2^{-N(1-h(p)-\epsilon-K_1/N)}.$$

Note that $K_1/N < 1-h(p)-\epsilon$. For given $\epsilon, \delta > 0$, there exists an N_2, when $N \geq N_2$, $\bar{P}_{\mathrm{H}_1} \leq \delta/8$. By Markov inequality,

$$\Pr(P_{\mathrm{H}_1} > \delta/2) < \frac{\bar{P}_{\mathrm{H}_1}}{\delta/2} \leq \frac{\delta/8}{\delta/2} = \frac{1}{4}.$$

Thus we have $\Pr(P_{\mathrm{H}_1} \leq \delta/2) = 1 - \Pr(P_{\mathrm{H}_1} \geq \delta/2) > 3/4$.

So far we have shown that there are more than $3/4$ random choices from all possible H_1 such that, for given $\epsilon, \delta > 0$, when $N \geq \max\{N_1, N_2\}$, $P_e = P_T + P_{\mathrm{H}_1} \leq \delta/2 + \delta/2 = \delta$. This concludes the proof of reliability.

C. Security proof

In this subsection, we will prove that $d \to 1$ as $N \to \infty$.

Consider the wiretapper's equivocation in three steps:

1) show that $H(\mathbf{S}|\mathbf{Z}) \geq N[h(p_w) - h(p)] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z})$.
2) show that $H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \leq h(P_{ew}) + P_{ew}K_2$. Here P_{ew} means a wiretapper's error probability to decode $\mathbf{x}$ in the case that $\mathbf{s}$ is known to the wiretapper.
3) show that for arbitrary $0 < \lambda < 1/2$, $P_{ew} \leq \lambda$.

Combining the above steps, we obtain that $d \to 1$ as $N \to \infty$.

First we prove step 1 by considering

$$\begin{aligned} H(\mathbf{S}|\mathbf{Z}) &= H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{S}, \mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{X}|\mathbf{Z}) + H(\mathbf{S}|\mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \\ &\overset{(a)}{=} H(\mathbf{X}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \\ &\geq H(\mathbf{X}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{Y}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \\ &= I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \\ &= N[I(X; Y) - I(X; Z)] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \\ &= N[h(p_w) - h(p)] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}), \end{aligned}$$

where (a) follows from the fact that $H(\mathbf{S}|\mathbf{X}, \mathbf{Z}) = 0$;

Now we prove step 2. Suppose that $\mathbf{S}$ takes value $\mathbf{s}$. For given $\mathrm{H}_2, \mathbf{s}$, we consider the solution set of equation (7) as a codebook, $\mathbf{X}$ in the codebook as the input codeword, wiretapper's observation $\mathbf{Z}$ as the corresponding output of passing $\mathbf{X}$ through the wiretap channel. From $\mathbf{Z}$, the decoder estimates $\mathbf{X}$ as $\hat{\mathbf{X}} = g(\mathbf{Z})$. Define the probability of error

$$P_{ew} = \Pr(\hat{\mathbf{X}} \neq \mathbf{X}). \tag{11}$$

From Fano's inequality, we have $H(\mathbf{X}|\mathbf{s},\mathbf{Z}){\le}h(P_{ew}) + P_{ew}K_2$. Therefore, $H(\mathbf{X}|\mathbf{S},\mathbf{Z}) \le h(P_{ew}) + P_{ew}K_2$. Thus we complete the proof of step 2.

Now we proceed to step 3. Note that the estimate $g(\mathbf{Z})$ of the decoder can be arbitrary. Here we use the typical set decoder. With the knowledge of $\mathbf{s}$ and $\mathbf{z}$, the decoder tries to find the codeword $\mathbf{x}$ sent to the channel. The decoder examines the typical set $T^N_{E_w}(\epsilon)$, the set of error sequences $\mathbf{e}_w$ that satisfy

$$2^{-N[h(p_w)+\epsilon]} \le \Pr(\mathbf{E}_w = \mathbf{e}_w) \le 2^{-N[h(p_w)-\epsilon]}.$$

If exactly one sequence $\mathbf{e}_w$ satisfies $\mathbf{e}_w \mathrm{H_2}^{\mathrm{T}} = \mathbf{z}\mathrm{H_2}^{\mathrm{T}}$, the decoder reports it as the hypothesized error sequence. Otherwise, a decoding error is reported.

The error probability of the typical set decoder at the wiretapper can be written as follows,

$$P_{ew} = P_{T_w} + P_{\mathrm{H}_2}, \tag{12}$$

where P_{T_w} is the probability that the true error sequence is itself not typical, and P_{H_2} is the probability that the true error sequence is typical and at least one other typical sequence clashes with it.

We first analyze P_{T_w}. For given $\epsilon, \lambda > 0$, there exists N_3, such that $\Pr\{\mathbf{e}_w \in T^N_{E_w}(\epsilon)\} \ge 1 - \lambda/2$ for $N \ge N_3$. Therefore, when $N \ge N_3$, $P_{T_w} = 1 - \Pr\{\mathbf{e}_w \in T^N_{E_w}(\epsilon)\} \le \lambda/2$.

Now we consider P_{H_2}. Suppose that the true error sequence is $\mathbf{e}_w \in T^N_{E_w}(\epsilon)$. If any of the typical error sequence $\mathbf{e}'_w \ne \mathbf{e}_w$, satisfies $(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_2}^{\mathrm{T}} = \mathbf{0}$, then we have an error. Let

$$T_{\mathbf{e}_w}(\epsilon) = \{\mathbf{e}'_w : \mathbf{e}'_w \in T^N_{E_w}(\epsilon), \mathbf{e}'_w \ne \mathbf{e}_w\}. \tag{13}$$

We have

$$P_{\mathrm{H}_2} \le \sum_{\mathbf{e}_w \in T^N_{E_w}(\epsilon)} \Pr(\mathbf{E}_w = \mathbf{e}_w) \sum_{\mathbf{e}'_w \in T_{\mathbf{e}_w}(\epsilon)} \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_2}^{\mathrm{T}} = \mathbf{0}].$$

Consider the average of P_{H_2}, $\bar{P}_{\mathrm{H}_2}$. We have

$$\bar{P}_{\mathrm{H}_2} \le \sum_{\mathbf{e}_w \in T^N_{E_w}(\epsilon)} \Pr(\mathbf{E}_w = \mathbf{e}_w) \sum_{\mathbf{e}'_w \in T_{\mathbf{e}_w}(\epsilon)} \langle \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_2}^{\mathrm{T}} = \mathbf{0}]\rangle_{\mathrm{H}_2}.$$

Note that for fixed $\mathbf{e}_w, \mathbf{e}'_w$,

$$\begin{aligned}
&\langle \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_2}^{\mathrm{T}} = \mathbf{0}]\rangle_{\mathrm{H}_2} \\
&= \langle\langle \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_1}^{\mathrm{T}} = \mathbf{0}] \cdot \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H}^{\mathrm{T}} = \mathbf{0}]\rangle_{\mathrm{H}_1}\rangle_{\mathrm{H}} \\
&= \langle \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H_1}^{\mathrm{T}} = \mathbf{0}]\rangle_{\mathrm{H}_1} \cdot \langle \mathbf{1}[(\mathbf{e}'_w - \mathbf{e}_w)\mathrm{H}^{\mathrm{T}} = \mathbf{0}]\rangle_{\mathrm{H}} \\
&= 2^{-(N-K_1)} \cdot 2^{-K} = 2^{-(N-K_2)}.
\end{aligned}$$

Therefore,

$$\bar{P}_{\mathrm{H}_2} < |T^N_{E_w}(\epsilon)| 2^{-(N-K_1+K)} \le 2^{-N(1-h(p_w)-\epsilon-K_2/N)}.$$

Note that $K_2/N < 1 - h(p_w) - \epsilon$. For given $\epsilon, \lambda > 0$, there exists N_4, when $N \ge N_4$, $\bar{P}_{\mathrm{H}_2} \le \lambda/8$. By Markov inequality,

$$\Pr(P_{\mathrm{H}_2} > \lambda/2) < \frac{\bar{P}_{\mathrm{H}_2}}{\lambda/2} \le \frac{\lambda/8}{\lambda/2} = \frac{1}{4}.$$

Thus we have $\Pr(P_{\mathrm{H}_2} \le \lambda/2) = 1 - \Pr(P_{\mathrm{H}_2} > \lambda/2) > 3/4$.

So far we have shown that there there are more than $3/4$ random choices from all possible H_1 and more than $3/4$ random choices from all possible H, such that, for given $\epsilon > 0$ and $\lambda > 0$, when $N \ge \max\{N_3, N_4\}$, $P_{ew} = P_{T_w} + P_{\mathrm{H}_2} \le \lambda/2 + \lambda/2 = \lambda$. This completes the proof of step 3.

As a conclusion of above discussion, for given $\varepsilon, \delta, \zeta, \epsilon > 0$, when $N \ge \max\{N_0, N_1, N_2, N_3, N_4\}$, there are more than $1/2\ (3/4 + 3/4 - 1)$ random choices of all possible H_1 and more than $3/4$ random choices from all possible H such that $P_e \le \delta$ and $P_{ew} \le \lambda$. In addition to (6), it is shown that there are H_1 and H that lead to a random linear code satisfying (4).

IV. Analysis of Information Leakage

In this section, we adopt the code structure of the random coding scheme but use normal linear codes in our construction for ease of implementation. We address the security of the coset coding scheme by analyzing its total information loss.

A. Coset coding scheme

Consider the communication model in Fig. 1. Note that in this section, H_1 and H_2 (thus H) are certainly of full rank. In particular, H_1, H_2 are parity check matrices of an (n, k_1) linear code C_1 and an (n, k_2) linear code C_2, respectively. Here $C_2 \subset C_1$ and $k = k_1 - k_2$. We use the same encoding strategy (equation (7)). The codebook in the encoding scheme is shown in Table I. At the legitimate receiver, the decoder uses syndrome decoding. It is easy to see that the coset code by C_1 and C_2 has error correcting capability beyond C_1.

TABLE I

The codebook in the encoding scheme

Space of input $\mathbf{x}$	Secret $\mathbf{s}$	Set of codewords w.r.t. secret $\mathbf{s}$
C_1	$\mathbf{s}(1)$	$\mathbf{x}(1) + C_2$
	$\mathbf{s}(2)$	$\mathbf{x}(2) + C_2$
	$\vdots$	$\vdots$
	$\mathbf{s}(2^k)$	$\mathbf{x}(2^k) + C_2$

B. Security analysis

The total information obtained by the wiretapper through his observation is $I(\mathbf{S};\mathbf{Z})$. We define it as the *information loss* (IL) of the scheme. First we have the following lemma.

Lemma 4.1: $H(\mathbf{Z}|\mathbf{S}) = H(\mathbf{Z}|\mathbf{S} = \mathbf{0})$.

Proof: For given $\mathbf{s}(i)$, $1 \le i \le 2^k$, we have

$$\begin{aligned}
p_{\mathbf{Z}|\mathbf{S}}(\mathbf{z}|\mathbf{s}(i)) &= \sum_{\mathbf{x}\in\mathbf{x}(i)+C_2} p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}(i)) p_{\mathbf{Z}|\mathbf{X},\mathbf{S}}(\mathbf{z}|\mathbf{x},\mathbf{s}(i)) \\
&\overset{(a)}{=} \frac{1}{2^{k_2}} \sum_{\mathbf{x}\in\mathbf{x}(i)+C_2} p_w^{w(\mathbf{x}+\mathbf{z})}(1-p_w)^{n-w(\mathbf{x}+\mathbf{z})} \\
&= \frac{1}{2^{k_2}} \sum_{\mathbf{v}\in\mathbf{z}+\mathbf{x}(i)+C_2} p_w^{w(\mathbf{v})}(1-p_w)^{n-w(\mathbf{v})},
\end{aligned} \tag{14}$$

where $w(\mathbf{v})$ is the Hamming weight of sequence $\mathbf{v}$ and (a) follows that $p_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}(i)) = 1/2^{k_2}$, $p(\mathbf{z}|\mathbf{x},\mathbf{s}(i)) = p(\mathbf{z}|\mathbf{x})$, and the fact that the wiretap channel is a BSC(p_w).

From (14), we see that $p_{\mathbf{Z}|\mathbf{S}}(\mathbf{z}|\mathbf{s}(i))$ is determined by the weight distribution of the coset $\mathbf{z} + \mathbf{x}(i) + C_2$. Note that for given $\mathbf{s}(i)$, $\{\mathbf{z} + \mathbf{x}(i) + C_2, \mathbf{z} \in \{0,1\}^n\}$ is a permutation of $\{\mathbf{z}+C_2, \mathbf{z} \in \{0,1\}^n\}$. As a straightforward consequence, $\{p_{\mathbf{Z}|\mathbf{S}}(\mathbf{z}|\mathbf{s}(i)), \mathbf{z} \in \{0,1\}^n\}$ is a permutation of $\{p_{\mathbf{Z}|\mathbf{S}}(\mathbf{z}|\mathbf{s} = \mathbf{0}), \mathbf{z} \in \{0,1\}^n\}$. Thus we have $H(\mathbf{Z}|\mathbf{S}) = H(\mathbf{Z}|\mathbf{S} = \mathbf{0})$. ■

Let C be a set of binary sequences of length n. We define

$$P_C(r) = \frac{1}{|C|} \sum_{\mathbf{v} \in C} r^{w(\mathbf{v})}(1-r)^{n-w(\mathbf{v})}. \tag{15}$$

where $0 \le r \le 1/2$ and $|C|$ is the cardinality of C. Note that the set of $\mathbf{x}$ corresponding to $\mathbf{s} = \mathbf{0}$ is C_2. We easily derive

$$p_{\mathbf{Z}|\mathbf{S}}(\mathbf{z}|\mathbf{s} = \mathbf{0}) = P_{\mathbf{z}+C_2}(p_w); \tag{16}$$

$$p_{\mathbf{Z}}(\mathbf{z}) = P_{\mathbf{z}+C_1}(p_w). \tag{17}$$

Then we have the following theorem.

Theorem 4.2: (An upper bound on IL)

$$\mathrm{IL} \le \log[2^n P_{C_2}(p_w)]. \tag{18}$$

Proof: The proof outline is as follows. By Lemma 4.1,

$$\mathrm{IL} = I(\mathbf{S};\mathbf{Z}) = H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{S} = \mathbf{0}). \tag{19}$$

We divide IL into two parts: $\mathrm{IL} = \mathrm{IL}_1 + \mathrm{IL}_2$, where

$$\mathrm{IL}_1 = \sum_{\mathbf{z}} P_{\mathbf{z}+C_2}(p_w) \log \frac{P_{\mathbf{z}+C_2}(p_w)}{P_{C_2}(p_w)}; \tag{20}$$

$$\mathrm{IL}_2 = \sum_{\mathbf{z}} P_{\mathbf{z}+C_1}(p_w) \log \frac{P_{C_2}(p_w)}{P_{\mathbf{z}+C_1}(p_w)}. \tag{21}$$

We can easily bound IL_1 by applying Theorem 1.19 in [5],

$$\mathrm{IL}_1 \le [\sum_{\mathbf{z} \notin C_2} P_{\mathbf{z}+C_2}(p_w)] \log \frac{1-(1-2p_w)^{k_2+1}}{1+(1-2p_w)^{k_2+1}} \le 0. \tag{22}$$

For IL_2, we apply the log-sum inequality and obtain

$$\mathrm{IL}_2 \le \log[2^n P_{C_2}(p_w)]. \tag{23}$$

Combining (22) and (23), we complete our proof. ■

Note that $2^n P_{C_2}(p_w)$ has a close relation with the probability of undetected error $P_{\mathrm{ue}}(C_2, p_w)$ defined in [5]. In fact,

$$\begin{aligned} 2^n P_{C_2}(p_w) &= 2^{n-k_2}[(1-p_w)^n + P_{\mathrm{ue}}(C_2, p_w)]; \\ &\overset{(a)}{=} 1 + [2(1-p_w)]^n P_{\mathrm{ue}}(C_2^{\perp}, (1-2p_w)/(2-2p_w)), \end{aligned}$$

where $C_2^{\perp}$ is the dual code of C_2 and (a) is due to the MacWilliam's identity and Theorem 2.1 in [5].

A binary (n,k) linear code C is called *good for error detection* if $P_{\mathrm{ue}}(C,r) \le 2^{-n}(2^k - 1)$, for all r, $0 \le r \le 1/2$. Let $R_2 = k_2/n$ and $\gamma = 2^{(1-R_2)}(1-p_w)$. Applying Theorem 2.43 and Theorem 2.51 in [5], we have

Lemma 4.3: $1 \le 2^n P_{C_2}(p_w) \le [2(1-p_w)]^{n-k_2}$.

Corollary 4.4: $\mathrm{IL} \le (n-k_2)[1 + \log(1-p_w)]$.

Lemma 4.5: If $C_2^{\perp}$ is good for error detection, then

$$1 \le 2^n P_{C_2}(p_w) < 1 + \gamma^n. \tag{24}$$

In the following, we consider $2^n P_{C_2}(p_w)$ as a random variable and investigate its first moment $\mathrm{E}_{\mathrm{H}_2}[2^n P_{C_2}(p_w)]$ and the second moment $\mathrm{E}_{\mathrm{H}_2}[(2^n P_{C_2}(p_w))^2]$ over all possible binary matrices H_2 of full rank. We will show that under certain constraint, our bound is asymptotically tight.

Lemma 4.6: (First and second moment of $2^n P_{C_2}(p_w)$)

$$\mathrm{E}_{\mathrm{H}_2}[2^n P_{C_2}(p_w)] = \gamma^n + \theta_1[1-(1-p_w)^n];$$

$$\mathrm{E}_{\mathrm{H}_2}[(2^n P_{C_2}(p_w))^2] = \theta_1\theta_2 + \gamma^{2n} + 2\theta_1\gamma^n + \theta_{t_1} + \theta_{t_2}.$$

Here $\theta_1 = (2^n - 2^{n-k_2})/(2^n - 1); \theta_2 = (2^n - 2^{n-k_2+1})/(2^n - 2);$
$\theta_{t_1} = \theta_1\gamma^n\{[(p_w^2 + (1-p_w)^2)/(1-p_w)]^n - 3(1-p_w)^n\};$
$\theta_{t_2} = -\theta_1\theta_2\{[p_w^2 + (1-p_w)^2]^n + 2(1-p_w)^n[1-(1-p_w)^n]\}.$

Lemma 4.7: If $R_2 > 1 + \log(1-p_w)$, then $\gamma < 1$ and

$$\lim_{n\to\infty} \mathrm{E}_{\mathrm{H}_2}[2^n P_{C_2}(p_w)] = 1. \tag{25}$$

Note that as $n \to \infty$, $\theta_1, \theta_2 \to 1$, $\theta_{t_2} \to 0$. If $R_2 > 1 + \log(1-p_w)$, then $\theta_{t_1} \to 0$ and thus the variance of $2^n P_{C_2}(p_w)$ approaches to 0 as $n \to \infty$. Based on this argument and Chebyshev's inequality, we have Theorem 4.8.

Theorem 4.8: If $R_2 > 1 + \log(1-p_w)$, for any $\epsilon > 0$,

$$\Pr\{2^n P_{C_2}(p_w) \le 2^{\epsilon}\} \to 1, \quad \text{as } n \to \infty. \tag{26}$$

As a conclusion of above discussion, C_2 plays a crucial role in insuring the secure transmission. For coset codes of short length, the code which minimizes $2^n P_{C_2}(p_w)$ might be a good candidate of C_2 by Theorem 4.2. Lighted by Lemma 4.5, codes, whose dual codes are good for error detection, can be good choices for C_2 especially when $R_2 > 1 + \log(1-p_w)$. If we allow n to grow, by Theorem 4.8 one can bound the information leakage arbitrarily small once we add enough randomness into the coding scheme via C_2. Furthermore, due to the constraint $R_2 > 1 + \log(1-p_w)$, the maximum secrecy rate in this case is $-\log(1-p_w) - h(p)$ instead of $h(p_w) - h(p)$.

V. Conclusion

In this paper, we investigate the binary symmetric wiretap channel. We give a strict mathematical proof that its secrecy capacity can be achieved by using random linear codes. Furthermore, we explore the coset coding scheme and give an upper bound on its total information loss. The bound implies the significance of C_2 in limiting the information leakage and gives hints on how to choose a satisfactory C_2. In particular, due to its close relation with the concept of undetected error probability, numerous results on codes for error detection can be applied to the design of applicable coset codes. We further show that the bound is asymptotically tight under certain constraint. The last but not least, we point out that the scheme has a sacrifice on efficiency and it is not very suitable for the case when $p < p_w \le 1 - 2^{-h(p)}$.

References

[1] A. D. Wyner, *The wire-tap channel.* Bell Sys. Tech. J., vol. 54, pp. 1355-1387, 1975.

[2] L. H. Ozarow and A. D. Wyner, *Wire-tap channel II.* Proc. Eurocrypt 84, A workshop on Advances in Cryptology: Theory and Application of Cryptographic Techniques. Paris, France, 1984.

[3] Andrew Thangaraj, Souvik Dihidar, A. R. Calderbank, Steven W. McLaughlin and Jean-Marc Merolla, *Applications of LDPC codes to the wiretap channel.* IEEE Trans. Info. Theory, vol. IT-53(8), pp. 2933-2945, 2007.

[4] Gérard Cohen and Gilles Zemor, *The wire-tap channel applied to biometrics.* Proc. Int. Symp. Inf. Theory & its Apps. Parma, Italy, 2004.

[5] Torleiv Kløve, *Codes for error detection.* World Scientific Publishing Co. Pte. Ltd., 2007.

Multi-Terminal Source Coding: Can Zero-rate Encoders Enlarge the Rate Region?

Badri N. Vellambi and Roy Timo
Institute for Telecommunications Research
University of South Australia
Email: {badri.vellambi, roy.timo}@unisa.edu.au

***Abstract*—Consider the multi-terminal source coding (MSC) problem wherein l discrete memoryless sources are compressed by l physically separate encoders, and a decoder is required to reconstruct the sources within certain distortion constraints. This paper focuses on the following question: Does the removal of a zero-rate link change the rate region? Though the question seems simple, its complication lies in the limiting nature of the rate region definition. Although intuition suggests that the answer should be no, resolving this question appears to be difficult for at least three reasons: (1) there is no known single-letter characterization of the MSC rate region; (2) there is no known elementary argument for rate-transfer from a zero-rate encoder to others; and (3) there is no known exponentially strong converse, whose existence would otherwise answer the question. In this paper, we answer the question for a number of special cases of the general MSC problem. Our proof techniques use a "super-code" style analysis along with new results from the helper problem. We note, however, that these techniques appear to fall short of answering the question in general.**

I. Introduction

One of the primary goals of information theory is the explicit characterization of rate regions for transmitting data over a network meeting certain requirements. The requirements are either lossless reconstructions of sources or lossy reconstructions certified by a prescribed distortion measure [1]. Such network rate regions are usually defined using sequences of block codes [2] and have a form as follows.

$$\mathcal{R}(\mathcal{D}) = \bigcap_{\varepsilon>0} \bigcup_{n\in\mathbb{N}} \mathcal{R}(\mathcal{D}, n, \varepsilon) = \lim_{\varepsilon\downarrow 0} \Bigg(\bigcup_{n\in\mathbb{N}} \mathcal{R}(\mathcal{D}, n, \varepsilon) \Bigg). \quad (1)$$

Here, $\mathcal{R}(\mathcal{D})$ represents the rate region for demands $\mathcal{D}$, and $\mathcal{R}(\mathcal{D}, n, \varepsilon)$ represents the set of rates at which there exists a block code of length n meeting the demands within a failure probability of ε. While properties such as convexity and closedness of the rate regions are straightforward to verify [1], continuity of rate regions w.r.t. the source statistics and the demands is harder to establish. Gu *et al*. have established the continuity of rate regions w.r.t. demands and source distribution for general classes of network problems [3]–[5]. Note that when a single-letter characterization of the rate region of a problem is known, it is almost trivial to ascertain the verity of such properties. However, multi-terminal information theory is fraught with simple problems such as the partial side-information (PSI) problem [6], multiple descriptions (MD) problem [7], and the multi-terminal source coding (MSC) problem [8] that remain unsolved.

In this work, we focus on one question: *Is the rate region of a network with zero rate on a link, the same as that of the network with that link deleted?* Though the question seems simple, its complication lies in the limiting nature of the definition of rate regions. When the sources in the network emit non-i.i.d. symbols, several examples can be designed to show that asymptotically zero-rate links can alter the region (see Example 1 of Sec. IV-B). However, when the sources emit i.i.d. symbols, and when the demands are lossy (within a required distortion) and/or lossless reconstructions, the answer to this question (in cases where it is known) has always affirmed that zero-rate links do not alter the region. In a majority of network cases where the answer is known, an explicit description of the rate region is also known. In some cases, even if the rate region is unknown, the existence of an exponentially strong converse suffices to answer this question [5], [9]. However, the existence of such suitably strong converses is a hard information-theoretic problem in itself. Here, we attempt to answer this question for the multi-terminal source coding problem that is formalized in Sec. III. Note that for this problem in its generality, neither is the rate region, nor is the existence of a strong converse known.

We have been able to use standard information-theoretic tools in a constructive fashion to show that under many settings of the MSC problem, the rate region with zero rate on a link is the same that when the link is absent. In specific, we establish that in both PSI and MSC problems with two discrete memoryless sources (DMSs), zero-rate links can be deleted without altering the rate region. However, for more than two correlated sources, this result is established only when a certain Markov property holds for the source joint distribution and for specific distortion requirements.

The remainder of the paper is organized as follows. Section II summarizes the notations employed throughout this paper. Section III presents the formulation of the PSI and MSC problems and various terminologies associated with the definition of the rate region. Section IV presents the results and proofs and Section VI concludes the paper.

II. Notations

Throughout the paper, the following notations are employed. For $n_1, n_2 \in \mathbb{N}$, $n_1 \leq n_2$, $[n_1] \triangleq \{1, \ldots, n_1\}$ and $[n_1 \sim n_2] \triangleq \{n_1, \ldots, n_2\}$. $\mathbf{0}_k$ represents the $1 \times k$ all-zero vector. Uppercase letters (e.g., X, Y) are reserved for random variables (RVs) and the respective script versions (e.g., $\mathscr{X}$, $\mathscr{Y}$) correspond to their alphabets. The realizations of RVs are usually denoted by lowercase letter (e.g., x, y). Subscripts are used for components of vectors, i.e., $x_{[n]}$ denotes a vector of length n and x_i represents the i^{th} component of $x_{[n]}$. We let $\mathcal{S}_n^{\varepsilon}(P)$ to denote the set of all ε-strongly P-typical sequences of length n [1]. When the underlying probability distribution is clear, H and I refers to the entropy and mutual information functionals. The Hamming distortion measure on a set $\mathscr{X}$ is denoted by $\partial_H^{\mathscr{X}}$, and lastly, $\mathbb{E}$ denotes the expectation operator.

III. Problem Definition

Given a DMS emitting $(X_i^{(1)}, \dots, X_i^{(l)})_{i \in \mathbb{N}}$ in an i.i.d. fashion with each symbol l-tuple having a joint distribution $p_{X^{(1)} \cdots X^{(l)}}$, the *multi-terminal source coding (MSC) problem* aims to identify rates at which encoders have to separately encode sequences $\{x_i^{(k)}\}_{i \in \mathbb{N}}$, $k \in [l]$, using l encoders so that l suitably distorted reconstructions can be constructed at the joint decoder (see Fig. 1).

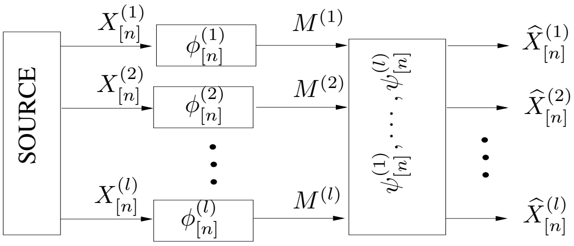


Fig. 1. The multi-terminal source coding problem

For $k \in [l]$, the reconstruction $(\widehat{X}_i^{(k)})_{i \in \mathbb{N}}$ is a sequence of elements from the reconstruction alphabet $\widehat{\mathscr{X}^{(k)}}$ and the acceptability of the reconstruction is evaluated by a distortion criterion using the distortion measure $\partial^{(k)} : \mathscr{X}^{(k)} \times \widehat{\mathscr{X}^{(k)}} \to \mathbb{R}^+$. A rate-distortion pair $(\mathbf{R}, \mathbf{\Delta}) \triangleq (R_1, \dots, R_l, \Delta_1, \dots, \Delta_l)$ is said to be achievable if for each $\varepsilon > 0$, there exists an ε-achievable block code $(\phi_{[n]}^{(1)}, \dots, \phi_{[n]}^{(l)}, \psi_{[n]}^{(1)}, \dots, \psi_{[n]}^{(l)})$. That is, $\forall \varepsilon > 0$, $\exists n \in \mathbb{N}$, s. t. $\forall k \in [l]$, there exist encoders $\phi_{[n]}^{(k)} : \mathscr{X}_{[n]}^{(k)} \to \mathscr{M}^{(k)}$ and decoders $\psi_{[n]}^{(k)} : \mathscr{M}^{(1)} \times \cdots \times \mathscr{M}^{(l)} \to \widehat{\mathscr{X}}_{[n]}^{(k)}$ satisfying:

A1. $\frac{1}{n} \sum_{i=1}^n \mathbb{E}\, \partial^{(k)}(X_i^{(k)}, \widehat{X}_i^{(k)}) \le \Delta_k + \varepsilon$, where $\widehat{X}_{[n]}^{(k)} \triangleq \psi_{[n]}^{(k)}\big(\phi_{[n]}^{(1)}(X_{[n]}^{(1)}), \dots, \phi_{[n]}^{(l)}(X_{[n]}^{(l)})\big)$, and

A2. $|\mathscr{M}^{(k)}| \le 2^{n(R_k + \varepsilon)}$.

Given $\mathbf{\Delta} \ge \mathbf{0}$, we say a rate vector $\mathbf{R}$ is achievable if $(\mathbf{R}, \mathbf{\Delta})$ is achievable in the aforementioned sense, and denote $\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})[p_{X^{(1)} \cdots X^{(l)}}]$ to be the set of achievable rate vectors. This set, known as the rate region, is convex and closed [1]. For each distortion measure, we let $\partial_{\max}^{(k)} \triangleq \min_{\widehat{x} \in \widehat{\mathscr{X}^{(k)}}} \sum_{x \in \mathscr{X}^{(k)}} p_{X^{(k)}}(x) \partial^{(k)}(x, \widehat{x})$. Note that when $\Delta_k \ge \partial_{\max}^{(k)}$, the k^{th} encoder can even operate at zero rate. However, any message from this encoder can *help* decoders to obtain less-distorted reconstructions of other sources. Given distortions $\mathbf{\Delta}$, we set $\mathcal{H}(\mathbf{\Delta}) \triangleq \{k \in [l] : \Delta_k \ge \partial_{\max}^{(k)}\}$ to be the set of *helper* sources.

As a special case, the MSC problem with $l = 2$ and $\Delta_2 \ge \partial_{\max}^{(2)}$ is called as the *partial side-information (PSI) problem*. In this case, the rate region is independent of the actual value of Δ_2 and is denoted by $\mathcal{R}^{\text{PSI}}(\Delta_1)[p_{X^{(1)}X^{(2)}}]$. Lastly, when clear, we drop the reference to the underlying source distribution in rate region notations.

IV. The Results

In this section, we present the results and proofs. First, the invariance of the rate region under the deletion of zero-rate links is established for the PSI problem. The invariance is then proved for the MSC problem with two sources followed by a direct extension to multiple sources. Although the invariance result for the MSC problem subsumes that of the PSI problem, the proof techniques for the two cases are very different. While the proof for the MSC problem exploits the knowledge of the rate region for the common helper problem (See Appendix A), that of the PSI problem is self-contained and constructive in nature. Finally, the invariance for the MSC problem when $l > 2$ is established for a class of sources that have certain Markovian property.

A. The Partial Side-information Problem

Theorem 1: Let $\mathsf{R}_{X^{(1)}}$ be the rate-distortion function for a DMS with distribution $p_{X^{(1)}}$ under the distortion measure $\partial^{(1)}$. Then,

$$\inf \left\{R : (R, 0) \in \mathcal{R}^{\text{PSI}}(\Delta_1)[p_{X^{(1)}X^{(2)}}]\right\} = \mathsf{R}_{X^{(1)}}(\Delta_1) \qquad (2)$$

Proof: Since $R \ge \mathsf{R}_{X^{(1)}}(\Delta_1) \Rightarrow (R, 0) \in \mathcal{R}^{\text{PSI}}(\Delta_1)$, we only need to show the reverse implication. Let $\varepsilon > 0$ and $(R, 0) \in \mathcal{R}^{\text{PSI}}(\Delta_1)$. Let $(\phi_{[n]}^{(1)}, \phi_{[n]}^{(2)}, \psi_{[n]}^{(1)}, \psi_{[n]}^{(2)})$ be an ε-achievable code for this rate-distortion tuple. Set $U \triangleq \phi_{[n]}^{(1)}(X_{[n]}^{(1)})$ and $V \triangleq \phi_{[n]}^{(2)}(X_{[n]}^{(2)})$ and let $\mathscr{U}, \mathscr{V}$ be their alphabets, respectively. Notating $q_n \triangleq p_{UX_{[n]}^{(1)}V}$, we have

$$\sum_{\substack{u \in \mathscr{U}, v \in \mathscr{V} \\ x_{[n]} \in \mathscr{X}_{[n]}^{(1)}}} \frac{q_n(u, x_{[n]}, v)}{n} \sum_{j=1}^n \partial^{(1)}((\psi_{[n]}^{(1)}(u, v))_j, x_j) \le \Delta_1 + \varepsilon.$$

Now, choose $m \in \mathbb{N}$ and a code C_m of $2^{m(I(UX_{[n]}^{(1)};V)+\varepsilon)}$ codewords from $\mathscr{V}_{[m]}$ (with each component of each codeword selected independently and identically using P_V) such that

B1. $\Pr[(X_{[n]})_{[m]} \in \mathcal{S}_m^{\varepsilon}(P_{X_{[n]}})] \ge 1 - \varepsilon$.

B2. $\Pr\left[\left\{(u_{[m]}, (x_{[n]})_{[m]}) \in \mathscr{U}_{[m]} \times (\mathscr{X}_{[n]})_{[m]} : \exists \mathbf{c} \in C_m \text{ s.t. } (u_{[m]}, (x_{[n]})_{[m]}, \mathbf{c}) \in \mathcal{S}_m^{\varepsilon}(q_n)\right\}\right] \ge 1 - \varepsilon$.

Consider the scheme where the $X^{(2)}$ encoder sends a constant message and $X^{(1)}$ encoder sends the index of a $\mathbf{v} \in C_m$ that is jointly typical with $(U_{[m]}, (X_{[n]}^{(1)})_{[m]})$ in addition to $U_{[m]} \triangleq (\phi_{[n]}^{(1)}(X_{[n]}^{(1)})), \dots, \phi_{[n]}^{(1)}(X_{[nm-n+1 \sim nm]}^{(1)}))$. Note that by B2, for almost all source realizations, at least one such $\mathbf{v}$ exists. If no such typical vector exists, the first codeword is transmitted by default. This scheme can be effected with a rate

$$\tilde{R}_1 = R_1 + \frac{1}{n} I(UX_{[n]}^{(1)}; V) + \varepsilon \overset{(a)}{\le} R_1 + 2\varepsilon, \qquad (3)$$

where (a) follows because the data processing inequality for $UX_{[n]}^{(1)} \ominus X_{[n]}^{(2)} \ominus V$ ensures $I(UX_{[n]}^{(1)}; V) \le I(X_{[n]}^{(2)}; V) \le \log_2 |\mathscr{V}| \le n\varepsilon$. The decoder uses the indices to generate $\widehat{X}_{[ln+1 \sim (l+1)n]}^{(1)} = \psi_{[n]}^{(1)}(U_{l+1}, V_{l+1})$, $l = 0, \dots, m-1$. By construction, we guarantee an average distortion of no more than $(\Delta_1+\varepsilon)(1+\varepsilon)(1-2\varepsilon)+2\varepsilon\partial^{(1)}_{\max} < \Delta_1+\varepsilon(1+2\partial^{(1)}_{\max})$. Since ε is arbitrary, we see that if $(R_1,0)\in\mathcal{R}^{\text{PSI}}(\Delta_1)$, then the rate R_1 suffices to construct a rate-distortion code for $p_{X^{(1)}}$ meeting a distortion of Δ_1 under $\partial^{(1)}$. ■

B. Multi-terminal Source Coding Problem with Two Sources

We first present an example that shows that the i.i.d. nature of the source is important for zero-rate links to play an insignificant role in shaping rate regions.

Example 1: Consider the MSC problem for the $\{X_i^{(1)},X_i^{(2)}\}_{i\in\mathbb{N}}$, where $\{X_i^{(1)}\}$ are i.i.d. with each index $X_i^{(1)}$ having a distribution $p_{X^{(1)}}$, and $X_i^{(2)}=(X_i^{(1)},A)$, where A is a discrete RV over $\mathscr{A}$ with distribution p_A that: (1) is statistically independent of each $X_{[s]}^{(1)}$ for $s\in\mathbb{N}$, and (2) meets $H(A)>0$. Then, for $\partial^{(1)}=\partial_H^{\mathscr{X}^{(1)}}$ and $\partial^{(2)}=\partial_H^{\mathscr{X}^{(2)}}$ and $\mathbf{\Delta}=\mathbf{0}_2$, $(H(X),0)$ is achievable, since one can use a good compression scheme for the $X^{(1)}$ side and convey A using $\lceil\log_2|\mathscr{A}|\rceil$ bits from the $X^{(2)}$ side. However, by deleting the link from the $X^{(2)}$ encoder, one cannot reconstruct the $X^{(2)}$ with zero distortion.

The following result shows such an event cannot occur for i.i.d. sources.

Theorem 2: Let $\mathcal{R}^{\text{H}}$, R^{PRD} denote the rate region for the common helper problem and the rate-distortion function for the partially-blind rate-distortion problem (see Appendix A), respectively. Then, the following are equivalent.

C1. $(R,0)\in\mathcal{R}^{\text{H}}(\Delta_1,\Delta_2)[p_{X^{(1)}X^{(2)}}]$.
C2. $(R,0)\in\mathcal{R}^{\text{MSC}}(\Delta_1,\Delta_2)[p_{X^{(1)}X^{(2)}}]$.
C3. $R\geq\mathsf{R}^{\text{PRD}}(\Delta_1,\Delta_2)[p_{X^{(1)}X^{(2)}}]$.

Proof: It is straightforward to see that C3 $\Rightarrow$ C2 and C2 $\Rightarrow$ C1. To show C1 $\Rightarrow$ C3, let $(R,0)\in\mathcal{R}^{\text{H}}(\Delta_1,\Delta_2)$. Then, from the rate region for the common helper problem (Appendix A), we have $p_{U^*V^*X^{(1)}X^{(2)}}\in\mathcal{P}^{\text{H}}(\Delta_1,\Delta_2)$, such that $R_2=I(X^{(1)}X^{(2)};V^*|U^*)=0$. This functional being zero in conjunction with the chain $U^*\ominus X^{(1)}\ominus X^{(2)}$ establish $V^*\ominus U^*\ominus X^{(1)}\ominus X^{(2)}$. Now, for $j=1,2$, let f_j denote functions that map $\mathscr{U}^*\times\mathscr{V}^*$ to the respective reconstruction alphabets $\mathscr{X}^{(j)}$, such that $f_j(U^*,V^*)$ meets required distortion constraint of Δ_j under $\partial^{(j)}$. Define for $j=1,2$, functions $h_j:\mathscr{U}^*\to\mathscr{V}^*$, $\tilde{f}_j:\mathscr{U}^*\to\widehat{\mathscr{X}^{(j)}}$ by

$$h_j(u)\triangleq\arg\min_{v\in\mathscr{V}^*}\sum_{x\in\mathscr{X}^{(j)}}p_{X^{(j)}|U^*}(x|u)\partial^{(j)}(x,f_j(u,v)). \quad (4)$$

$$\tilde{f}_j(u)\triangleq f_j(u,h_j(u)). \quad (5)$$

Observe that by construction, for $j=1,2$,

$$\mathbb{E}\,\partial^{(j)}(X^{(j)},\tilde{f}_j(U^*))\leq\mathbb{E}\,\partial_x(X^{(j)},f_j(U^*,V^*))\leq\Delta_j. \quad (6)$$

Thus, there exists a distribution $p_{U^*X^{(1)}X^{(2)}}$ with (1) $|\mathscr{U}^*|\leq|\mathscr{X}^{(1)}||\mathscr{X}^{(2)}|+4$; (2) $U^*\ominus X^{(1)}\ominus X^{(2)}$; and (3) functions $\tilde{f}_j$ that provide reconstructions $\widehat{X}^{(j)}$ from U^* meeting the required distortions. Therefore, $p_{U^*X^{(1)}X^{(2)}}\in\mathcal{P}^{\text{PRD}}(\Delta_1,\Delta_2)$ (possibly after altering the definition of R^{PRD} to include auxiliary RVs with alphabet sizes up to $|\mathscr{X}^{(1)}||\mathscr{X}^{(2)}|+4$, which does not alter the PRD rate region). Therefore, we have

$(R,0)\in\mathcal{R}^{\text{H}}(\Delta_1,\Delta_2)\Rightarrow R\geq\mathsf{R}^{\text{PRD}}(\Delta_1,\Delta_2)$. ■

At this point, we would like to remark that the inner bound $\mathcal{R}^{\text{MSC}}_{in}(\Delta_1,\Delta_2)$ by Berger and Tung [10] and the outer bound $\mathcal{R}^{\text{MSC}}_{out}(\Delta_1,\Delta_2)$ obtained from traditional converse techniques (that replaces the chain $U\ominus X^{(1)}\ominus X^{(2)}\ominus V$ in the inner bound with $U\ominus X^{(1)}\ominus X^{(2)}$ and $X^{(1)}\ominus X^{(2)}\ominus V$) also agree on the $R_2=0$ plane. That is,

$$\begin{aligned}(R,0)\in\mathcal{R}^{\text{MSC}}_{in}(\Delta_1,\Delta_2)\Leftrightarrow\ &(R,0)\in\mathcal{R}^{\text{MSC}}_{out}(\Delta_1,\Delta_2) &(7)\\ \Leftrightarrow\ &R\geq\mathsf{R}^{\text{PRD}}(\Delta_1,\Delta_2), &(8)\end{aligned}$$

thereby providing an alternate proof of the invariance result for the MSC problem when $l=2$. Further, Theorem 2 can be extended for the $l>2$ setting to show that zero-rate encoders cannot help when there is only one link carrying positive rate.

Theorem 3: For $l>2$ and $\mathbf{\Delta}\geq 0$

$$(R,\mathbf{0}_{l-1})\in\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}]\Leftrightarrow R\geq\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}].$$

Proof: Note that

$$(R,\mathbf{0}_{l-1})\in\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}]\Rightarrow(R,0)\in\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})[p_{X^{(1)}Y}],$$

where $Y=(X^{(2)}\cdots X^{(l)})$. Notice here that the distortions $\partial^{(k)}$ for $k>1$ can be equivalently seen as distortion measures for the Y-source. However, from Theorem 2, we notice that

$$(R,0)\in\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})[p_{X^{(1)}Y}]\Rightarrow R\geq\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}Y}]. \quad (9)$$

However, since $\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}Y}]=\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}]$, the proof is complete because $R\geq\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}]$ is achievable for the MSC problem. ■

C. Multi-terminal Source Coding Problem for $l>2$ sources

Here, we show that for a class of sources and under certain distortions $\mathbf{\Delta}$, the MSC rate region with zero rates on certain links is the same as that of the MSC problem with the same constraints and with the zero-rate links deleted.

Theorem 4: Suppose $\exists\, S\subset[l], i\in[l]\setminus S$, such that $X^{(S)}\ominus X^{(i)}\ominus X^{((S\cup\{i\})^c)}$. Additionally, if $S\subseteq\mathcal{H}(\mathbf{\Delta})$, then all rate vectors in $\mathcal{R}^{\text{MSC}}(\mathbf{\Delta})\cap\{R_j=0:j\in S\}$ are achievable even if the encoders encoding $X^{(j)}$, $j\in S$ send a constant message.

Proof: Since the proof is a simple multi-source adaptation of that of Theorem 1 that establishes a rate-transfer argument, we present only an outline of the proof. Given an ε-achievable code C with l encoders, construct a block supercode C' with a bigger block length, wherein the encoders corresponding to the indices of S transmit constant messages, and the i^{th} encoder constructs a codebook that will transmit along with its usual message, additional message that corresponds to a typical realization of the messages that would be originally sent over the $|S|$ zero-rate links, i.e., from the encoders encoding $X^{(j)}$, $j\in S$. In doing so, the rate from the encoders encoding $\{X_j:j\in S\}$ is transferred to that of i. Note that this additional rate incurred is bounded above by $|S|\varepsilon$. The proof is complete by noting that ε is arbitrary. ■

Note that the above result is different from that of $l=2$ case, since for $l>2$, the rate regions of suitable sub-networks may be unknown. For example, consider the MSC problem for a DMS with distribution $p_{X_1X_2X_3}$ s. t. $X_1 \ominus X_2 \ominus X_3$. Theorem 4 guarantees

$$\mathcal{R}^{\text{MSC}}(\Delta_1, \Delta_2, \partial^{(3)}_{\max}) \cap \{R_3 = 0\} \cong \mathcal{R}^{\text{PSI}}(\Delta_1, \Delta_2)[p_{X^{(1)}X^{(2)}}],$$
$$\mathcal{R}^{\text{MSC}}(\partial^{(1)}_{\max}, \Delta_2, \Delta_3) \cap \{R_1 = 0\} \cong \mathcal{R}^{\text{PSI}}(\Delta_2, \Delta_3)[p_{X^{(2)}X^{(3)}}],$$

where $\cong$ signifies that the right-hand region is the appropriate projection of the one on the left. Note that this result is previously unknown, since the rate region for the PSI problem remains open. Additionally, $\forall\, \mathbf{\Delta} \geq \mathbf{0}_3$, Theorem 3 guarantees

$$(0, R_2, 0) \in \mathcal{R}^{\text{MSC}}(\mathbf{\Delta}) \Rightarrow R_2 \geq \mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}X^{(2)}X^{(3)}}].$$

V. Acknowledgements

This work has been support by the Australian Research Council under the ARC Discovery Grant DP0880223.

VI. Conclusions

The invariance of the MSC rate region under the deletion of zero-rate links was studied. Though the question of invariance remains open in general, it was shown that the rate region remains unaltered if zero-rate links are deleted from the PSI and MSC problems with two correlated DMSs. When more than two correlated DMSs are present, it was established that the deletion of zero-rate links from some helper encoders do not alter the MSC rate region provided the source distribution has a certain Markov structure.

Appendix A
Allied Problems and their rate regions

Problem 1: (*The partially-blind rate-distortion problem*) Given a discrete source emitting $(X_i^{(1)}, \ldots, X_i^{(l)})_{i\in\mathbb{N}}$ in an i.i.d. fashion with each symbol l-tuple having the joint distribution $p_{X^{(1)}\cdots X^{(l)}}$. The problem aims to identify the rates at which the $X^{(1)}$ sequence can be encoded so that suitably "noisy" reconstruction $(\widehat{X}_i^k)_{i\in\mathbb{N}}$ for each $k \in [l]$ is constructed by the block decoder. The acceptability of the reconstructions are determined by distortion criteria using distortion measures $\partial^{(k)} : \mathscr{X}^{(k)} \times \widehat{\mathscr{X}^{(k)}} \to \mathbb{R}^+$. A pair $(R_1, \mathbf{\Delta})$ is said to be achievable if for each $\varepsilon > 0$, $\exists\, n \in \mathbb{N}$, $\phi^{(1)}_{[n]} : \mathscr{X}_{[n]} \to \mathscr{M}^{(1)}$ and $\psi_{[n]} : \mathscr{M}^{(1)} \to \widehat{\mathscr{X}}^{(1)}_{[n]} \times \cdots \times \widehat{\mathscr{X}}^{(l)}_{[n]}$ s. t.:

D1. $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \partial^{(k)}(X_i^{(k)}, \widehat{X}_i^{(k)}) \leq \Delta_k + \varepsilon$, $\forall k \in [l]$, and
D2. $|\mathscr{M}^{(1)}| \leq 2^{n(R_1+\varepsilon)}$.

The infimum of achievable rates $\mathsf{R}^{\text{PRD}}(\mathbf{\Delta})[p_{X^{(1)}\cdots X^{(l)}}]$ can be shown to be as follows.

$$\mathsf{R}^{\text{PRD}}(\mathbf{\Delta}) = \inf_{p_{UX^{(1)}\cdots X^{(l)}} \in \mathcal{P}^{\text{PRD}}(\mathbf{\Delta})} I(X^{(1)}; U),$$

where $\mathcal{P}^{\text{PRD}}(\mathbf{\Delta})$ is the set of distributions $p_{UX^{(1)}\cdots X^{(l)}}$ s. t.:

E1. $U \ominus X^{(1)} \ominus (X^{(2)} \cdots X^{(l)})$, $|\mathscr{U}| \leq |\mathscr{X}^{(1)}| + l$, and
E2. $\forall\, k \in [l]$, $\exists\, f^{(k)} : \mathscr{U} \to \widehat{\mathscr{X}^{(k)}}$, $\mathbb{E}\, \partial^{(k)}(X^{(k)}, f^{(k)}(U)) \leq \Delta_k$.

Problem 2: (*The common helper problem*) Given a discrete source emitting $(X_i^{(1)}, X_i^{(2)})_{i\in\mathbb{N}}$ in an i.i.d. fashion with each pair having the joint distribution $p_{X^{(1)}X^{(2)}}$. The problem aims to identify the rates at which information must be sent by encoders so that suitably "noisy" versions $(\widehat{X}_i^{(1)})_{i\in\mathbb{N}}$ and $(\widehat{X}_i^{(2)})_{i\in\mathbb{N}}$ are constructed by a joint block decoder. Here, the first encoder (the helper encoder) has access to the $X^{(1)}$-sequence, whereas the second one has access to both $X^{(1)}$- and $X^{(2)}$-sequences. As before. the acceptability of reconstructions are evaluated by distortion criteria using distortion measures $\partial^{(1)} : \mathscr{X}^{(1)} \times \widehat{\mathscr{X}^{(1)}} \to \mathbb{R}^+$ and $\partial^{(2)} : \mathscr{X}^{(2)} \times \widehat{\mathscr{X}^{(2)}} \to \mathbb{R}^+$. A quadruplet $(R_1, R_{12}, \Delta_1, \Delta_2)$ is said to be achievable if for each $\varepsilon > 0$, $\exists\, n \in \mathbb{N}$, $\phi^{(1)}_{[n]} : \mathscr{X}^{(1)}_{[n]} \to \mathscr{M}^{(1)}$, $\phi^{(12)}_{[n]} : \mathscr{X}^{(1)}_{[n]} \times \mathscr{X}^{(2)}_{[n]} \to \mathscr{M}^{(12)}$, $\psi^{(1)}_{[n]} : \mathscr{M}^{(1)} \times \mathscr{M}^{(12)} \to \widehat{\mathscr{X}}^{(1)}_{[n]}$ and $\psi^{(2)}_{[n]} : \mathscr{M}^{(1)} \times \mathscr{M}^{(12)} \to \widehat{\mathscr{X}}^{(2)}_{[n]}$ s. t.

F1. $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \partial^{(j)}\Big(X_i^{(j)}, \widehat{X}_i^{(j)}\Big) \leq \Delta_j + \varepsilon$, $j = 1, 2$, where $\widehat{X}^{(j)}_{[n]} \triangleq \psi^{(j)}_{[n]}\big(\phi^{(1)}_{[n]}(X^{(1)}_{[n]}), \phi^{(12)}_{[n]}(X^{(1)}_{[n]}, X^{(2)}_{[n]})\big)$, and
F2. $|\mathscr{M}_t| \leq 2^{n(R_t+\varepsilon)}$, $t = 1, 12$.

Even though the problem defines two separate encoders, allowing the $X^{(1)}$ encoder to send its encoded message to the $X^{(1)}X^{(2)}$ encoder does not alter the rate region. This setting is the more readily seen as the common helper setup [11]. The set $\mathcal{R}^{\text{H}}(\Delta_1, \Delta_2)$ of achievable rates is given by

$$\mathcal{R}^{\text{H}}(\mathbf{\Delta}) = \left\{ \begin{matrix} R_1 \geq I(X^{(1)}; U) \\ R_{12} \geq I(X^{(1)}X^{(2)}; V|U) \end{matrix} : p_{X^{(1)}X^{(2)}UV} \in \mathcal{P}^{\text{H}}(\mathbf{\Delta}) \right\},$$

where $\mathcal{P}^{\text{H}}(\mathbf{\Delta})$ is the set of distributions $P_{UVX^{(1)}X^{(2)}}$ s. t.:

G1. $U \ominus X^{(1)} \ominus X^{(2)}$, $|\mathscr{U}| \leq |\mathscr{X}^{(1)}||\mathscr{X}^{(2)}| + 4$.
G2. $|\mathscr{V}| \leq (|\mathscr{X}^{(1)}||\mathscr{X}^{(2)}| + 2)^2 - 2$.
G3. $\exists\, f_1 : \mathscr{U} \times \mathscr{V} \to \widehat{\mathscr{X}^{(1)}}$, $\mathbb{E}\, \partial^{(1)}(X^{(1)}, f_1(U, V)) \leq \Delta_1$.
G4. $\exists\, f_2 : \mathscr{U} \times \mathscr{V} \to \widehat{\mathscr{X}^{(2)}}$, $\mathbb{E}\, \partial^{(2)}(X^{(2)}, f_2(U, V)) \leq \Delta_2$.

References

[1] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, USA: Academic Press, 1982.

[3] T. Han and K. Kobayashi, "A dichotomy of functions f(x, y) of correlated sources (x, y)," *Information Theory, IEEE Transactions on*, vol. 33, no. 1, pp. 69–76, Jan 1987.

[4] W.-H. Gu and M. Effros, "On the continuity of achievable rate regions for source coding over networks," in *Information Theory Workshop, IEEE*, Sept. 2007, pp. 632–637.

[5] W.-H. Gu, M. Effros, and M. Bakshi, "A continuity theory for lossless source coding over networks," in *Communication, Control, and Computing, 46th Annual Allerton Conference on*, Sept. 2008, pp. 1527–1534.

[6] T. Berger, K. Housewright, J. Omura, S. Yung, and J. Wolfowitz, "An upper bound on the rate distortion function for source coding with partial side information at the decoder," *Information Theory, IEEE Transactions on*, vol. 25, no. 6, pp. 664–666, Nov 1979.

[7] L. Zhao, P. Cuff, and H. Permuter, "Consolidating achievable regions of multiple descriptions," in *Information Theory, IEEE International Symposium on*, 28 June-July 3 2009, pp. 51–54.

[8] T. Berger and R. Yeung, "Multiterminal source encoding with encoder breakdown," *Information Theory, IEEE Transactions on*, vol. 35, no. 2, pp. 237–244, Mar 1989.

[9] W.-H. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *Information Theory, IEEE International Symposium on*, 28 June-July 3 2009, pp. 2316–2320.

[10] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, Nov 1978.

[11] H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a common helper," in *Information Theory, IEEE International Symposium on*, 28 June-July 3 2009, pp. 1473–1477.

Mismatched Decoding for the Relay Channel

Charlotte Hucher and Parastoo Sadeghi
The Australian National University, Canberra, ACT, 0200, Australia
Email: {charlotte.hucher,parastoo.sadeghi}@anu.edu.au

***Abstract*— We consider a discrete memoryless relay channel, where both relay and destination may have an incorrect estimation of the channel. This estimation error is modeled with mismatched decoders. In this paper, we provide a lower-bound on the mismatch capacity of the relay channel. Moreover, we prove that this lower-bound is indeed the exact capacity for the degraded relay channel when random coding is used.**

I. Introduction

The decoding method that minimizes the error probability is the maximum-likelihood (ML) decoder. However, it cannot always be implemented in practice because of some channel estimation errors or hardware limitations. An alternative decoder can then be a mismatched one, based on a different metric. The theoretical performance of mismatched decoding has been studied since the 1980's when Csiszàr and Körner in [1], and Hui in [2] both provided a lower-bound on the achievable capacity in a point-to-point communication channel. In [3], the authors proved that this lower-bound is the exact capacity when random coding is used. The mismatch capacity of multiple-access channels has also been characterized in [4].

There is increasing evidence that future wireless communications will be based not on point-to-point transmission anymore, but on cooperation between the nodes in a network (see [5],[6]). The simplest model of a cooperative network is the relay channel for which capacity bounds have been derived in 1979 by Cover and El Gamal in [7].

In this paper, we consider a discrete memoryless relay channel with mismatched decoders at both receivers (the relay and the destination). We provide a lower-bound on the mismatch capacity of such a channel and prove that it is indeed the exact capacity of the mismatched degraded relay channel when random coding is used.

II. The relay channel and mismatched decoder

We consider a discrete memoryless relay channel consisting of one source, one relay and one destination. We use the same setup as in [7]: The source broadcasts a signal $x_1 \in \mathcal{X}_1$ which is received by both the relay and the destination. The relay transmits a signal $x_2 \in \mathcal{X}_2$ which is received by the destination. Received signals at relay and destination are denoted by $y_1 \in \mathcal{Y}_1$ and $y \in \mathcal{Y}$ respectively (see Figure 1).

The channel is modeled by a set of probability distributions $p(y_1, y|x_1, x_2)$. We consider three mismatched decoders using the metrics $q_{sr}(x_1, x_2, y_1)$, $q_{rd}(x_2, y)$ and $q_{sd}(x_1, x_2, y)$, where the subscripts sr, rd and sd stand for the source-relay, relay-destination and source-destination links, respectively.

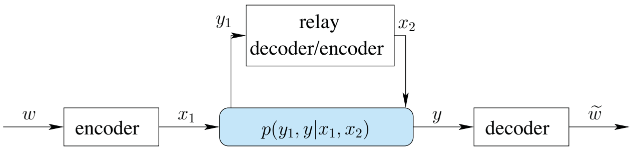


Fig. 1. Relay channel

The following setup is used to prove the achievability of the lower-bound on the mismatch capacity derived in this paper.

We consider the transmission of B blocks of length n. In each block $i \in \{1, \ldots, B\}$, a message $w_i \in \{1, \ldots, 2^{nR}\}$ is transmitted from the source. Let us partition the set $\{1, \ldots, 2^{nR}\}$ into 2^{nR_0} independent subsets denoted by $S_s, s \in \{1, \ldots, 2^{nR_0}\}$, such that any $w_i \in \{1, \ldots, 2^{nR}\}$ belongs to a unique subset S_{s_i}. The message is then coded as $\mathbf{x}_1(w_i|s_i) \in \mathcal{X}_1^n$ and $\mathbf{x}_2(s_i) \in \mathcal{X}_2^n$ at the source and relay, respectively.

Random coding: The choice of the set $\mathcal{C} = \{x_1(.|.), x_2(.)\}$ of codewords is random:

- 2^{nR_0} iid codewords in $\mathcal{X}_2^n$ are first generated according to the probability distribution $p(\mathbf{x}_2) = \prod_{i=1}^n p(x_{2i})$ and indexed by $s \in \{1, \ldots, 2^{nR_0}\}$: $\mathbf{x}_2(s)$;
- for each $\mathbf{x}_2(s)$, 2^{nR} iid codewords in $\mathcal{X}_1^n$ are generated according to the probability distribution $p(\mathbf{x}_1|\mathbf{x}_2(s))$ and indexed by $w \in \{1, \ldots, 2^{nR}\}$: $\mathbf{x}_1(w|s)$.

Two transmission steps: Let us assume that $(i-2)$ blocks have already been sent. Thus the relay has already decoded w_{i-2} and s_{i-1}, the destination has decoded w_{i-3} and s_{i-2}. In order to derive a lower bound on the capacity of the mismatched relay channel, we choose to use threshold decoders as in [2], [4]. Indeed, it can be proven that the decoding error probability of an ML mismatched decoder (which is implemented in practice) is upper-bounded by the decoding error probability of the considered threshold decoder.

In block $(i-1)$, the source and relay transmit $\mathbf{x}_1(w_{i-1}|s_{i-1})$ and $\mathbf{x}_2(s_{i-1})$, respectively; the relay and destination receive $\mathbf{y}_1(i-1)$ and $\mathbf{y}(i-1)$. The relay is able to detect w_{i-1} as the unique w such that $(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}_1(i-1))$ is jointly typical and $q_{sr}(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}_1(i-1))$ is larger than a threshold to be defined later. The relay is thus able to determine s_i such that $w_{i-1} \in S_{s_i}$. The source is obviously also aware of s_i.

In block i, the source and relay transmit $\mathbf{x}_1(w_i|s_i)$ and $\mathbf{x}_2(s_i)$, respectively; the relay and destination receive $\mathbf{y}_1(i)$ and $\mathbf{y}(i)$. The destination can detect s_i as the unique s such that $(\mathbf{x}_2(s_i), \mathbf{y}(i))$ is jointly typical and $q_{rd}(\mathbf{x}_2(s_i), \mathbf{y}(i))$ is larger than some threshold. It is then able to detect

w_{i-1} as the unique w such that $w \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$, where $\mathcal{L}(\mathbf{y}(i-1))$ is the set of all $w \in \{1, \dots, 2^{nR}\}$ such that $(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1))$ is jointly typical and $q_{sd}(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1))$ is larger than some threshold.

Notation: Let $E_p(q(x))$ denote the expected value of $q(x)$ w.r.t. the probability distribution $p(x)$. Let $I_f(X;Y)$ denote the usual mutual information between X and Y w.r.t. the probability distribution $f(x,y)$.

For a probability distribution $p(x)$ on a finite set $\mathcal{X}$ and a constant $\delta > 0$, let $N^{\delta}_{p(x)}$ denote the set of all probability distributions on $\mathcal{X}$ that are within δ of $p(x)$: $N^{\delta}_{p(x)} = \{f \in \mathcal{P}(\mathcal{X}) : \forall x \in \mathcal{X}, |f(x) - p(x)| \leq \delta\}$. Let $T^{\delta}_{p(x)}$ be the set of all sequences in $\mathcal{X}^n$ whose type is in $N^{\delta}_{p(x)}$: $T^{\delta}_{p(x)} = \left\{\mathbf{x} \in \mathcal{X}^n : \mathbf{f}_x \in N^{\delta}_{p(x)}\right\}$, where $\mathbf{f}_x(x)$ is the number of elements of the sequence $\mathbf{x}$ that are equal to x, normalized by the sequence length n. In the following, we drop the subscript arguments when the context is clear enough and write $f(x) \in N^{\delta}_{p}$ and $(\mathbf{x}, \mathbf{y}) \in T^{\delta}_{p}$.

III. Main results

Theorem 1: (Achievability) Consider a discrete memoryless relay channel described by the probability distribution $p(y_1, y|x_1, x_2)$. The capacity C_M obtained using the mismatched decoders $q_{sr}(x_1, x_2, y_1)$, $q_{rd}(x_2, y)$ and $q_{sd}(x_1, x_2, y)$ is lower-bounded by:

$$C_{LM} = \max_{p(x_1,x_2)} I_{LM}(p(x_1, x_2)) \tag{1}$$

$$\begin{aligned} I_{LM}(p(x_1, x_2)) = \min\{&I_{sr}(p(x_1, x_2)), I_{rd}(p(x_2)) \\ &+ I_{sd}(p(x_1, x_2))\} \end{aligned} \tag{2}$$

where

$$I_{sr}(p(x_1, x_2)) = \min_{f \in \mathcal{D}_{sr}} I_f(X_1; Y_1|X_2) + I_f(X_1; X_2) \tag{3}$$

$$I_{rd}(p(x_2)) = \min_{f \in \mathcal{D}_{rd}} I_f(X_2; Y) \tag{4}$$

$$I_{sd}(p(x_1, x_2)) = \min_{f \in \mathcal{D}_{sd}} I_f(X_1; Y|X_2) + I_f(X_1; X_2) \tag{5}$$

and

$$\begin{aligned} \mathcal{D}_{sr} =\{&f(x_1, x_2, y_1) : f(x_1) = p(x_1), f(x_2, y_1) = p(x_2, y_1), \\ &E_f(q_{sr}(x_1, x_2, y_1)) \geq E_p(q_{sr}(x_1, x_2, y_1))\} \end{aligned} \tag{6}$$

$$\begin{aligned} \mathcal{D}_{rd} =\{&f(x_2, y) : f(x_2) = p(x_2), f(y) = p(y), \\ &E_f(q_{rd}(x_2, y)) \geq E_p(q_{rd}(x_2, y))\} \end{aligned} \tag{7}$$

$$\begin{aligned} \mathcal{D}_{sd} =\{&f(x_1, x_2, y) : f(x_1) = p(x_1), f(x_2, y) = p(x_2, y), \\ &E_f(q_{sd}(x_1, x_2, y)) \geq E_p(q_{sd}(x_1, x_2, y))\}. \end{aligned} \tag{8}$$

Let now suppose that there exist three probability distributions $\hat{f}_i$, $i \in \{sr, rd, sd\}$ such that $E_{\hat{f}_i}(q_i) > E_p(q_i)$ with strict inequality.

Theorem 2: (Converse) With the above assumption and for a degraded relay channel, if for some input distribution $p(x_1, x_2)$, the rate $R > C_{LM}$, then the average probability of error, averaged over all random codebooks drawn according to $p(x_1, x_2)$, approaches one as the block length tends to infinity.

IV. Proof of Theorem 1

A. Upper-bounding the error probability

In order to prove the achievability of this lower-bound, we consider the four possible error events described in [7] adapted to the use of a threshold decoder. For each block i, these possible error events are:

- E_{0i}: $(\mathbf{x}_1(w_i|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i), \mathbf{y}(i))$ is not jointly typical;
- E_{1i}: there exists $\tilde{w} \neq w_i$ such that $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is jointly typical and $q_{sr}(\mathbf{x}_1(\tilde{w}|s_i),\ \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is larger than some threshold;
- E_{2i}: there exists $\tilde{s} \neq s_i$ such that $(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i))$ is jointly typical and $q_{rd}(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i))$ is larger than some threshold;
- $E_{3i} = E'_{3i} \cup E''_{3i}$
  - E'_{3i}: $w_{i-1} \notin S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$
  - E''_{3i}: there exists $\tilde{w} \neq w_{i-1}$ such that $\tilde{w} \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$.

Let F_i be the decoding error event in block i. Let us assume that no error has occurred till block $i-1$. Thus, the decoding error probability in block i is given by:

$$p_e(i) = \sum_{k=0}^{3} \Pr\left\{E_{ki} \cap F^c_{i-1} \bigcap_{l=0}^{k-1} E^c_{li}\right\} \triangleq \sum_{k=0}^{3} p_{ek}(i).$$

1) Probability of error event E_{0i}: By Sanov's theorem [8, Theorem 11.4.1], this probability is exponentially small in n. There exists $\psi > 0$ such that $p_{e0}(i) < 2^{-n\psi}$.

2) Probability of error event E_{1i}: An error occurs if there exists $\tilde{w} \neq w_i$ such that the metric $q_{sr}(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is greater than the threshold

$$\Upsilon^{\delta}_{sr} = \min_{\tilde{p} \in N^{\delta}_{p}} \sum_{(x_1,x_2,y_1) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1} \tilde{p}(x_1, x_2, y_1) q_{sr}(x_1, x_2, y_1), \tag{9}$$

where δ is a small positive number.

The probability of error event E_{1i} is thus given by

$$\begin{aligned} p_{e1}(i) = \Pr\{&\exists \tilde{w} \neq w_i, \mathbf{x}_1(\tilde{w}|s_i) \in T^{\delta}_{p}, \\ &q_{sr}(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i)) \geq \Upsilon^{\delta}_{sr}|F^c_{i-1}\}. \end{aligned}$$

Using Sanov's theorem, we obtain the upper-bound

$$\begin{aligned} p_{e1}(i) &\leq (2^{nR} - 1)(n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}_1|} 2^{-n\tilde{R}^{\delta}_{sr}} \\ &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}_1|} 2^{-n(\tilde{R}^{\delta}_{sr} - R)}, \end{aligned}$$

where

$$\tilde{R}^{\delta}_{sr} = \min_{f(x_1,x_2,y) \in \mathcal{D}^{\delta}_{sr}} D(f(x_1, x_2, y_1) \| p(x_1) p(x_2, y_1)) \tag{10}$$

$$\begin{aligned} \mathcal{D}^{\delta}_{sr} = \{&f(x_1, x_2, y_1) : f(x_1) \in N^{\delta}_{p}, f(x_2, y_1) \in N^{\delta}_{p}, \\ &E_f(q_{sr}(x_1, x_2, y_1)) \geq \Upsilon^{\delta}_{sr}\}, \end{aligned} \tag{11}$$

with $D(.\|.)$ denoting the KL divergence [8, equation (2.26)].

Thus, if $R < \tilde{R}^{\delta}_{sr}$, the probability of error event E_{1i} is exponentially small in n: $p_{e1}(i) < 2^{-n\psi}$ for some $\psi > 0$.

3) Probability of error event E_{2i}: Using the threshold

$$\Upsilon_{rd}^{\delta} = \min_{\tilde{p} \in N_p^{\delta}} \sum_{(x_2,y) \in \mathcal{X}_2 \times \mathcal{Y}} \tilde{p}(x_2,y) q_{rd}(x_2,y), \tag{12}$$

we can write the probability of error event E_{2i} as

$$\begin{aligned} p_{e2}(i) &= \Pr\left\{\exists \tilde{s} \neq s_i, \mathbf{x}_2(\tilde{s}) \in T_p^{\delta}, q_{rd}(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i)) \geq \Upsilon_{rd}^{\delta} | F_{i-1}^c\right\} \\ &\leq (n+1)^{|\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{rd}^{\delta} - R_0)}, \end{aligned}$$

where the upper-bound is obtained using Sanov's theorem with

$$\tilde{R}_{rd}^{\delta} = \min_{f(x_2,y) \in \mathcal{D}_{rd}^{\delta}} D(f(x_2,y) \| p(x_2)p(y)) \tag{13}$$

$$\begin{aligned} \mathcal{D}_{rd}^{\delta} = \{ &f(x_2,y) : f(x_2) \in N_p^{\delta}, f(y) \in N_p^{\delta}, \\ &E_f(q_{rd}(x_2,y)) \geq \Upsilon_{rd}^{\delta} \}. \end{aligned} \tag{14}$$

If $R_0 < \tilde{R}_{rd}^{\delta}$, then $p_{e2}(i) < 2^{-n\psi}$ for some $\psi > 0$.

4) Probability of error event E_{3i}: Error event E_{3i} can be decomposed into two different events E'_{3i} and E''_{3i}.

If we assume that the previous transmission was correctly received at destination, then $w_{i-1} \in \mathcal{L}(\mathbf{y}(i-1))$. Moreover, the fact that the error event E_{2i} does not occur implies that $w_{i-1} \in S_{\hat{s}_i} = S_{s_i}$. Thus the first term of the decomposition has a probability zero and we only need to consider E''_{3i}.

$$\begin{aligned} p_{e3}(i) &= \Pr\left\{\exists \tilde{w} \neq w_{i-1}, \tilde{w} \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1)) | F_{i-1}^c\right\} \\ &\leq E\left\{ \sum_{\tilde{w} \neq w_{i-1}, \tilde{w} \in \mathcal{L}(\mathbf{y}(i-1))} \Pr\left\{\tilde{w} \in S_{s_i}\right\} | F_{i-1}^c \right\} \\ &\leq E\left\{ \|\mathcal{L}(\mathbf{y}(i-1))\| 2^{-nR_0} | F_{i-1}^c \right\} \end{aligned}$$

where $\|.\|$ denotes the cardinality of the considered set.

Let

$$\varphi(\tilde{w}|\mathbf{y}) = \begin{cases} 1, & \mathbf{x}_1(\tilde{w}|s_{i-1}) \in T_p^{\delta}, \\ & q_{sd}(\mathbf{x}_1(\tilde{w}|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1)) \geq \Upsilon_{sd}^{\delta} \\ 0, & \text{otherwise.} \end{cases}$$

where the threshold is defined by

$$\Upsilon_{sd}^{\delta} = \min_{\tilde{p} \in N_p^{\delta}} \sum_{(x_1,x_2,y) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}} \tilde{p}(x_1,x_2,y) q_{sd}(x_1,x_2,y). \tag{15}$$

Using Sanov's theorem, we can upper-bound the expected cardinality of $\mathcal{L}(\mathbf{y}(i-1))$ given that $\tilde{w} \neq w_{i-1}$

$$\begin{aligned} &E\left\{ \|\mathcal{L}(\mathbf{y}(i-1)) | \tilde{w} \neq w_{i-1}\| \big| F_{i-1}^c \right\} \\ &\quad = E\left\{ \sum_{\tilde{w} \neq w_{i-1}} \varphi(\tilde{w}|\mathbf{y}) \big| F_{i-1}^c \right\} \\ &\quad \leq (2^{nR} - 1)(n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n\tilde{R}_{sd}^{\delta}} \\ &\quad \leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} - R)}, \end{aligned}$$

where

$$\tilde{R}_{sd}^{\delta} = \min_{f(x_1,x_2,y) \in \mathcal{D}_{sd}^{\delta}} D(f(x_1,x_2,y) \| p(x_1)p(x_2,y)) \tag{16}$$

$$\begin{aligned} \mathcal{D}_{sd}^{\delta} = \{ &f(x_1,x_2,y) : f(x_1) \in N_p^{\delta}, f(x_2,y) \in N_p^{\delta}, \\ &E_f(q_{sd}(x_1,x_2,y)) \geq \Upsilon_{sd}^{\delta} \}. \end{aligned} \tag{17}$$

The error probability is then upper-bounded by

$$\begin{aligned} p_{e3}(i) &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} - R)} 2^{-nR_0} \\ &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} + R_0 - R)}. \end{aligned}$$

Replacing R_0 by the constraint previously computed

$$p_{e3}(i) \leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} + \tilde{R}_{rd}^{\delta} - R)}.$$

If $R < \tilde{R}_{sd}^{\delta} + \tilde{R}_{rd}^{\delta}$, then $p_{e3}(i) < 2^{-n\psi}$ for some $\psi > 0$.

B. Existence of a random code

If $R < \tilde{R}_{sr}^{\delta}$ and $R < \tilde{R}_{rd}^{\delta} + \tilde{R}_{sd}^{\delta}$, then the total error probability is exponentially small in n: $p_e < 4B \times 2^{-n\psi}$. Thus, as n tends to infinity, the probability of finding a set of codewords $\mathcal{C}$ respecting $p_e(\mathcal{C}) < 4B \times 2^{-n\psi}$ tends to one.

Let $\mathcal{C}$ be such a set of codewords of length n. Its average error probability is lower than $4B \times 2^{-n\psi}$. Throwing away the worst half of the codewords, we end up with a set of codewords $\mathcal{C}^*$ of length $\frac{n}{2}$ whose maximum error probability is lower than $2 \times 4B \times 2^{-n\psi}$, which tends to zero, and whose rate is $R - \frac{1}{n}$ which tends to R.

C. Letting δ tend to zero

We note that $\lim_{\delta \to 0} \Upsilon_{sr}^{\delta} = E_p(q_{sr}(x_1,x_2,y_1))$. Thus, the set $\mathcal{D}_{sr}^{\delta}$ becomes

$$\begin{aligned} \mathcal{D}_{sr} = \{ &f(x_1,x_2,y_1) : f(x_1) = p(x_1), f(x_2,y_1) = p(x_2,y_1), \\ &E_f(q_{sr}(x_1,x_2,y_1)) \geq E_p(q_{sr}(x_1,x_2,y_1)) \} \end{aligned}$$

and with these new constraints on the probability distribution

$$D(f(x_1,x_2,y_1) \| p(x_1)p(x_2,y_1)) = I_f(X_1;Y_1|X_2) + I_f(X_1;X_2).$$

Thus the first constraint of the rate becomes

$$I_{sr}(p(x_1,x_2)) \triangleq \min_{f \in \mathcal{D}_{sr}} I_f(X_1;Y_1|X_2) + I_f(X_1;X_2). \tag{18}$$

In the same way, we find the final expressions of $I_{rd}(p(x_1,x_2))$ and $I_{sd}(p(x_1,x_2))$.

V. Proof of Theorem 2

The proof of Theorem 2 is in essence similar to the one of [4, Theorem 3].

A. Decoding at relay

Let us assume that

$$R > \min_{f \in \mathcal{D}_{sr}} I_f(X_1;Y_1|X_2) + I_f(X_1;X_2). \tag{19}$$

Let f^* be the probability distribution that achieves (19). Let $\tilde{f} = (1-\epsilon)f^* + \epsilon \hat{f}_{sr}$. We recall that $\hat{f}_{sr}$ is a probability distribution that respects $E_{\hat{f}_{sr}}(q_{sr}) > E_p(q_{sr})$. Then, for sufficiently small ϵ,

$$R > I_{\tilde{f}}(X_1;Y_1|X_2) + I_{\tilde{f}}(X_1;X_2) \tag{20}$$

$$E_{\tilde{f}}(q_{sr}) > E_p(q_{sr}). \tag{21}$$

Using (20), (21) and the continuity of the divergence, we can find $\Delta > 0$, $\epsilon > 0$ and a neighborhood U of $\tilde{f}(x_1, x_2, y_1)$ such that for all $f \in U$ and $p'(x_2, y_1) \in N_p^\epsilon$, we have

$$\begin{aligned} R &> D(f(x_1|x_2, y_1)\|p(x_1)|p'(x_2, y_1)) + \Delta && (22)\\ E_f(q_{sr}) &> E_p(q_{sr}) + \Delta, && (23) \end{aligned}$$

where $D(.\|.|.)$ is defined in [4].

Let V be a sufficiently small neighborhood of $p(x_1, x_2, y_1)$, such that for every $p'(x_1, x_2, y_1) \in V$, then $p'(x_2, y_1) \in N_p^\epsilon$ and

$$E_{p'}(q_{sr}) < E_p(q_{sr}) + \Delta. \qquad (24)$$

Let us assume that the true message in block i is w_i and that the triple $(\mathbf{x}_1(w_i|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ has empirical type in V. If there exists another message $\tilde{w} \neq w_i$ such that the empirical type of $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is in U, then a decoding error occurs.

Let $W(\tilde{w})$ take the value 1 if the empirical type of $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is in U and 0 otherwise.

The expectation of $W = \sum_{\tilde{w} \neq w_i} W(\tilde{w})$ given $\mathbf{y}_1$ is

$$E(W|\mathbf{y}_1) = (2^{nR} - 1)\pi_0 \doteq 2^{nR}\pi_0, \qquad (25)$$

where $\doteq$ denotes the behavior of the expression when $n \to \infty$ and $\pi_0 = E(W(\tilde{w}^*)|\mathbf{y}_1) = \Pr\{W(\tilde{w}^*) = 1|\mathbf{y}_1\}$ with $\tilde{w}^*$ being a random message different from w_i.

For two different messages $\tilde{w}$ and $\tilde{w}'$, the events $W(\tilde{w})$ and $W(\tilde{w}')$ are independent. Thus the variance of W given $\mathbf{y}_1$ is

$$\text{Var}(W|\mathbf{y}_1) = \sum_{\tilde{w} \neq w_i} \text{Var}(W(\tilde{w})|\mathbf{y}_1).$$

Since $W(\tilde{w})$ can only take the values 0 and 1, we can upper-bound $\text{Var}(W(\tilde{w})|\mathbf{y}_1) \leq E(W(\tilde{w})|\mathbf{y}_1)$. Thus

$$\text{Var}(W|\mathbf{y}_1) \leq \sum_{\tilde{w} \neq w_i} E(W(\tilde{w})|\mathbf{y}_1) \doteq 2^{nR}\pi_0. \qquad (26)$$

Using (25), (26) and the fact that for any random variable X, $\Pr(X = 0) \leq \frac{\text{Var}(X)}{E(X)^2}$, we can write

$$\Pr(W = 0|\mathbf{y}_1) \dot{\leq} \frac{2^{nR}\pi_0}{(2^{nR}\pi_0)^2} = 2^{-nR}\frac{1}{\pi_0}. \qquad (27)$$

Using the second part of Sanov's theorem, we obtain the asymptotic behavior $\pi_0 \doteq 2^{-n\tilde{R}_{sr}}$, where $\tilde{R}_{sr} = \min_{f \in \mathcal{U}} D(f(x_1|x_2, y_1)\|p(x_1)|p'(x_2, y_1))$. Using (22), we can then lower-bound $\pi_0 \dot{\geq} 2^{-n(R-\Delta)}$ and conclude that the probability of no decoding error tends to 0 when $n \to \infty$:

$$\Pr(W = 0|\mathbf{y}_1) \dot{\leq} 2^{-nR}2^{n(R-\Delta)} = 2^{-n\Delta}. \qquad (28)$$

B. Decoding at destination

The second inequality can be dealt in two separate parts. Indeed, we have shown in the direct part that

$$\begin{aligned} R_0 &\leq \min_{f \in \mathcal{D}_{rd}} I_f(X_2; Y) && (29)\\ R &\leq R_0 + \min_{f \in \mathcal{D}_{sd}} I_f(X_1; Y|X_2) + I_f(X_1; X_2). && (30) \end{aligned}$$

We thus have to show that if one of these inequalities is reversed, an error occurs with asymptotic probability one. This can be done using the same reasoning as in previous subsection.

VI. Matched decoding case

In the matched decoding case, i.e. $q_{sr}(x_1, x_2, y_1) = \log p(y_1|x_1, x_2)$, $q_{rd}(x_2, y) = \log p(y|x_2)$ and $q_{sd}(x_1, x_2, y) = \log p(y|x_1, x_2)$, the capacity coincides with the one of degraded relay computed by Cover and El Gamal in [7]. Indeed, for any distribution $f \in \mathcal{D}_{sr}$, we have

$$\begin{aligned} &I_f(X_1; Y_1|X_2) + I_f(X_1; X_2)\\ &\geq I_f(X_1; Y_1|X_2)\\ &= H(Y_1|X_2) - H_f(Y_1|X_1, X_2) && (31)\\ &\geq H(Y_1|X_2) + \sum_{x_1, x_2, y_1} f(x_1, x_2, y_1) \log p(y_1|x_1, x_2) && (32)\\ &\geq H(Y_1|X_2) + \sum_{x_1, x_2, y_1} p(x_1, x_2, y_1) \log p(y_1|x_1, x_2) && (33)\\ &= I(X_1; Y_1|X_2), \end{aligned}$$

where (31) holds because $f(x_2, y_1) = p(x_2, y_1)$, (32) follows from the non-negativity of the divergence and (33) is obtained using $E_f(\log p(y_1|x_1, x_2)) \geq E(\log p(y_1|x_1, x_2))$.

Moreover, by choosing $f(x_1, x_2, y_1) = p(x_1)p(x_2)p(y_1|x_1, x_2) \in \mathcal{D}_{sr}$, $I_f(X_1; X_2) = 0$ and $I_f(X_1; Y_1|X_2) + I_f(X_1; X_2) = I(X_1; Y_1|X_2)$. The bound is achievable, so $I_{sr}(p(x_1, x_2)) = I(X_1; Y_1|X_2)$.

In the same way, we can prove that $I_{rd}(p(x_1, x_2)) = I(X_2; Y)$ and $I_{sd}(p(x_1, x_2)) = I(X_1; Y|X_2)$.

Finally, in the matched case, the following rate is achievable

$$\begin{aligned} R &= \min\{I(X_1; Y_1|X_2), I(X_2; Y) + I(X_1; Y|X_2)\} && (34)\\ &= \min\{I(X_1; Y_1|X_2), I(X_1, X_2; Y)\}. && (35) \end{aligned}$$

Acknowledgement

This work was supported under Australian Research Council's Discovery Projects funding scheme (project no. DP0984950). The authors would also like to thank Prof. Lapidoth for useful discussions.

References

[1] I. Csiszàr and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[2] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, M.I.T., 1983, chap. IV.

[3] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[4] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity. Part I and II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[6] A. Nosratinia and A. Hedayat, "Cooperative Communication in Wireless Networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[7] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

Polar Codes for the m-User MAC and Matroids

Emmanuel Abbe, Emre Telatar
Information Processing group, EPFL
Lausanne 1015, Switzerland
Email: {emmanuel.abbe,emre.telatar}@epfl.ch

***Abstract*—In this paper, a polar code for the m-user multiple access channel (MAC) with binary inputs is constructed. In particular, Arıkan's polarization technique applied individually to each user will polarize any m-user binary input MAC into a finite collection of extremal MACs. The extremal MACs have a number of desirable properties: (i) the 'uniform sum rate'[1] of the original channel is not lost, (ii) the extremal MACs have rate regions that are not only polymatroids but matroids and thus (iii) their uniform sum rate can be reached by each user transmitting either uncoded or fixed bits; in this sense they are easy to communicate over. Provided that the convergence to the extremal MACs is fast enough, the preceding leads to a low complexity communication scheme that is capable of achieving the uniform sum rate of an arbitrary binary input MAC. We show that this is indeed the case for arbitrary values of m.**

I. Introduction

In [2], Arıkan shows that a single-user binary input channel can be "polarized" by a simple process that coverts n independent uses of this channel into n successive uses of "extremal" channels. These extremal channels are binary input and either almost perfect or very noisy, i.e., having a uniform mutual information either close to 1 or close to 0. Furthermore, the fraction of almost perfect channels is close to the uniform mutual information of the original channel. For a 2-user binary input MAC, by applying Arikan's construction to each user's input separately, [6] shows that a similar phenomenon appears: the n independent uses of the MAC are converted into n successive uses of "extremal" binary inputs MACs. These extremal MACs are of four kinds: (1) each users sees a very noisy channel, (2) one of the user sees a very noisy channel and the other sees an almost perfect channel, (3) both users see an almost perfect channel, (4) a pure contention channel: a channel whose uniform rate region is the triangle with vertices (0,0), (0,1), (1,0). Moreover [6] shows that the uniform sum rate of the original MAC is preserved during the polarization process, and that the polarization to the extremal MACs occurs fast enough. This allows the construction of a polar code to achieve reliable communication at uniform sum rate.

In this paper, we investigate the case where m is arbitrary. In the two user case, the extremal MACs are not just MACs for which each users sees either a perfect or pure noise channel, as there is also the pure contention channel. However, the uniform rate region of the 2-user extremal MACs are all polyhedron with integer valued constraints. We will see in this paper that the approach used for the 2-user case faces a new phenomenon when the number of users reaches 4, and the extremal MACs are no longer in a one to one correspondence with the polyhedron having integer valued constraints. To characterize the extremal MACs, we first show how an unusual relationship between random variables defined in terms of mutual information falls precisely within the independence notion of the matroid theory. This relationship is used to show that the extremal MACs are equivalent to linear deterministic channels, which is then used to conclude the construction of a polar code ensuring reliable communication for arbitrary values of m. Finally, the problem of considering m arbitrary large is of interest for a polarization of the additive white Gaussian noise channel.

II. The Polarization Construction

We consider a m-user multiple access channel with binary input alphabets (BMAC) and arbitrary output alphabet $\mathcal{Y}$. The channel is specified by the conditional probabilities

$$P(y|\bar{x}), \quad \text{for all } y \in \mathcal{Y} \text{ and } \bar{x} = (x[1], \dots, x[m]) \in \mathbb{F}_2^m.$$

Let $E_m := \{1, \dots, m\}$ and let $X[1], \dots, X[m]$ be mutually independent and uniformly distributed binary random variables. Let $\bar{X} := (X[1], \dots, X[m])$. We denote by Y the output of $\bar{X}$ through the MAC P. For $J \subseteq E_m$, we define

$$\begin{aligned} X[J] &:= \{X[i] : i \in J\}, \\ I[J](P) &:= I(X[J]; Y X[J^c]), \end{aligned}$$

where J^c denotes the complement set of J in E_m. Note that

$$\mathcal{I}(P) := \{(R_1, \dots, R_m) : 0 \le \sum_{i \in J} R_i \le I[J](P), \forall J \subseteq E_m\}$$

is an inner bound to the capacity region of the MAC P. We refer to $\mathcal{I}(P)$ as the uniform rate region and to $I[E_m](P)$ as the uniform sum rate. We now consider two independent uses of such a MAC. We define $\bar{X}_1 := (X_1[1], \dots, X_1[m])$, $\bar{X}_2 := (X_2[1], \dots, X_2[m])$, where $X_1[i], X_2[i]$, with $i \in E_m$, are mutually independent and uniformly distributed binary random variables. We denote by Y_1 and Y_2 the respective outputs of $\bar{X}_1$ and $\bar{X}_2$ through two independent uses of the MAC P:

$$\bar{X}_1 \xrightarrow{P} Y_1, \quad \bar{X}_2 \xrightarrow{P} Y_2. \tag{1}$$

We define two additional binary random vectors $\bar{U}_1 := (U_1[1], \dots, U_1[m])$, $\bar{U}_2 := (U_2[1], \dots, U_2[m])$ with mutually

[1] In this paper all mutual informations are computed when the inputs of a MAC are distributed independently and uniformly. The resulting rate regions, sum rates, etc., are prefixed by 'uniform' to distinguish them from the capacity region, sum capacity, etc.

independent and uniformly distributed components, and we put $\bar{X}_1$ and $\bar{X}_2$ in one to one correspondence with $\bar{U}_1$ and $\bar{U}_2$ with $\bar{X}_1 = \bar{U}_1 + \bar{U}_2$ and $\bar{X}_2 = \bar{U}_2$, where the addition is the modulo 2 component wise addition.

Definition 1. Let $P : \mathbb{F}^m \to \mathcal{Y}$ be a m-user BMAC. We define two new m-user BMACs, $P^- : \mathbb{F}_2^m \to \mathcal{Y}^2$ and $P^+ : \mathbb{F}_2^m \to \mathcal{Y}^2 \times \mathbb{F}_2^m$, by $P^-(y_1, y_2|\bar{u}_1) := \sum_{\bar{u}_2 \in \mathbb{F}_2^m} \frac{1}{2^m} P(y_1|\bar{u}_1 + \bar{u}_2)P(y_2|\bar{u}_2)$ and $P^+(y_1, y_2, \bar{u}_1|\bar{u}_2) := \frac{1}{2^m} P(y_1|\bar{u}_1 + \bar{u}_2)P(y_2|\bar{u}_2)$ for all $\bar{u}_i \in \mathbb{F}_2^m$, $y_i \in \mathcal{Y}$, $i = 1, 2$.

That is, we have now two new m-user BMACs with extended output alphabets:

$$\bar{U}_1 \overset{P^-}{\to} (Y_1, Y_2), \quad \bar{U}_2 \overset{P^+}{\to} (Y_1, Y_2, \bar{U}_1) \tag{2}$$

which also defines $I[J](P^-)$ and $I[J](P^+)$, $\forall J \subseteq E_m$.

This construction is the natural extension of the construction for $m = 1, 2$ in [2], [6]. Here again, we are comparing two independent uses of the same channel P (cf. (1)) with two successive uses of the channels P^- and P^+ (cf. (2)). Note that $I[J](P^-) \leq I[J](P) \leq I[J](P^+)$, $\forall J \subseteq E_m$.

Definition 2. Let $\{B_n\}_{n \geq 1}$ be i.i.d. uniform random variables valued in $\{-, +\}$. Let the random processes $\{P_n,\ n \geq 0\}$ and $\{I_n[J],\ n \geq 0\}$, for $J \subseteq E_m$, be defined by $P_0 := P$,

$$P_{n+1} := P_n^{B_{n+1}}, \quad I_n[J] := I[J](P_n), \quad \forall n \geq 0.$$

III. Results

Summary: In Section III-A, we show that $\{I_n[J], J \subseteq E_m\}$ tends a.s. to sequence of number which defines a matroid (cf. Definition 5). We then see in Section III-B that the extreme points of a uniform rate region with matroid constraints can be achieved by each user sending uncoded or frozen bits; in particular the uniform sum rate can be achieved by such strategies. We then show in Section III-D, that for arbitrary m, $\{I_n[J], J \subseteq E_m\}$ does not tend to an arbitrary matroid but to a binary matroid (cf. Definition 6). This is used to show that the convergence to the extremal MACs happens fast enough, and that the construction of previous section leads to a polar code having a low encoding and decoding complexity and achieving the uniform sum rate on any binary MAC.

A. The extremal MACs

Lemma 1. *$\{I_n[J], n \geq 0\}$ is a bounded super-martingale when $J \subsetneq E_m$ and a bounded martingale when $J = E_m$.*

Proof: For any $J \subseteq E_m$, $I_n[J] \leq m$ and

$$\begin{aligned}
2I[J](P) &= I(X_1[J]X_2[J]; Y_1Y_2X_1[J^c]X_2[J^c]) \\
&= I(U_1[J]U_2[J]; Y_1Y_2U_1[J^c]U_2[J^c]) \\
&= I(U_1[J]; Y_1Y_2U_1[J^c]U_2[J^c]) \\
&\quad + I(U_2[J]; Y_1Y_2U_1[J^c]U_2[J^c]U_1[J]) \\
&\geq I(U_1[J]; Y_1Y_2U_1[J^c]) \\
&\quad + I(U_2[J]; Y_1Y_2\bar{U}_1U_2[J^c]) \\
&= I[J](P^-) + I[J](P^+),
\end{aligned} \tag{3}$$

where equality holds above, if $J^c = \emptyset$, i.e., if $J = E_m$. ■

Note that the inequality in the above are only due to the bounds on the mutual informations of the P^- channel. Because of the equality when $J = E_m$, our construction preserves the uniform sum rate. As a corollary of previous Lemma, we have the following result.

Theorem 1. *The process $\{I_n[J], J \subseteq E_m\}$ converges a.s..*

Note that for a fixed n, $\{I_n[J], J \subseteq E_m\}$ denotes the collection of the 2^m random variables $I_n[J]$, for $J \subseteq E_m$. When the convergence takes place (a.s.), let us define $I_\infty[J] := \lim_{n\to\infty} I_n[J]$. From previous theorem, $I_\infty[J]$ is a random variable valued in $[0, |J|]$. We will now further characterize these random variables.

Lemma 2. *For any $\varepsilon > 0$ and any m-user BMAC P, there exists $\delta > 0$, such that for any $J \subseteq E_m$, if $I[J](P^+) - I[J](P) < \delta$, we have $I[J](P) - I[J \setminus i] \in [0, \varepsilon) \cup (1 - \varepsilon, 1]$, $\forall i \in J$, where $I[\emptyset] = 0$.*

Lemma 3. *With probability one, $I_\infty[i] \in \{0, 1\}$ and $I_\infty[J] - I_\infty[J \setminus i] \in \{0, 1\}$, for every $i \in E_m$ and $J \subseteq E_m$.*

Note that Lemma 3 implies in particular that $\{I_\infty[J], J \subseteq E_m\}$ is a.s. a discrete random vector.

Definition 3. We denote by $\mathcal{A}_m$ the support of $\{I_\infty[J], J \subseteq E_m\}$ (when the convergence takes place, i.e., a.s.). This is a subset of $\{0, \ldots, m\}^{2^m}$.

We have already seen that not every element in $\{0, \ldots, m\}^{2^m}$ can belong to $\mathcal{A}_m$. We will now further characterize the set $\mathcal{A}_m$.

Definition 4. A polymatroid is a set E_m, called the ground set, equipped with a function $f : 2^m \to \mathbb{R}$, called a rank function, which satisfies

$$\begin{aligned}
&f(\emptyset) = 0 \\
&f[J] \leq f[K], \quad \forall J \subseteq K \subseteq E_m, \\
&f[J \cup K] + f[J \cap K] \leq f[J] + f[K], \quad \forall J, K \subseteq E_m.
\end{aligned}$$

Theorem 2. *For any MAC and any distribution of the inputs $X[E]$, we have that $\rho(S) = I(X[S]; YX[S^c])$ is a rank function on E, where we denote by Y the output of the MAC with input $X[E]$. Hence, (E, ρ) is a polymatroid.*

(A proof of this result can be found in [7].) Therefore, any realization of $\{I_n[J], J \subseteq E_m\}$ defines a rank function and the elements of $\mathcal{A}_m$ define polymatroids.

Definition 5. A matroid is a polymatroid whose rank function is integer valued and satisfies $f(J) \leq |J|$, $\forall J \subseteq E_m$. We denote by MAT_m the set of all matroids with ground state E_m. We also define a basis of a matroid by the collection of maximal subsets of E_m for which $f(J) = |J|$. One can show that a matroid is equivalently defined from its bases.

Using Lemma 3 and the definition of a matroid, we have the following result.

Theorem 3. *For every $m \geq 1$, $\mathcal{A}_m \subseteq \mathrm{MAT}_m$.*

We will see that the inclusion is strict for $m \geq 4$.

B. Communicating On Matroids

We have shown that, when n tends to infinity, the MACs that we create with the polarization construction of Section II are particular MACs: the mutual informations $I_\infty[J]$ are integer valued (and satisfy the other matroid properties). A well-known result of matroid theory (cf. Theorem 22 of [4]) says that the vertices of a polymatroid given by a rank function f are the vectors of the following form:

$$\begin{aligned} &x_{j(1)} = f(A_1), \\ &x_{j(i)} = f(A_i) - f(A_{i-1}), \quad \forall 2 \leq i \leq k \\ &x_{j(i)} = 0, \quad \forall k < i \leq m, \end{aligned}$$

for some $k \leq m$, $j(1), j(2), \dots, j(m)$ distinct in E_m and $A_i = \{j(1), j(2), \dots, j(i)\}$. He hence have the following.

Corollary 1. *The uniform rate regions of the MACs defined by $\mathcal{A}_m$ have vertices on the hypercube $\{0,1\}^m$. In particular, when operating at a vertex each user sees either a perfect or pure noise channel.*

C. Convergence Speed and Representation of Matroids

Convention: for a given m, we write the collection $\{I_\infty[J], J \subseteq E_m\}$ by skipping the empty set (since $I_\infty[\emptyset] = 0$) and as follows: when $m = 2$, we order the sequence as $(I_\infty[1], I_\infty[2], I_\infty[1,2])$, and when $m = 3$, as $(I_\infty[1], I_\infty[2], I_\infty[3], I_\infty[1,2], I_\infty[1,3], I_\infty[2,3], I_\infty[1,2,3])$, etc.

When $m = 2$, [6] shows that $\{I_\infty[J], J \subseteq E_m\}$ belongs a.s. to $\{(0,0,0), (0,1,1), (1,0,1), (1,1,1), (1,1,2)\}$. These are precisely all the matroids with two elements. The speed of convergence to these matroids is shown to be fast in [6] through the following steps. The main idea is to deduce the convergence speed of $I_n[J]$ from the convergence speed obtained in the single user setting, which we know is fast enough, namely as $o(2^{-n^\beta})$, for any $\beta < 1/2$, cf. [3]. We do not need to check the speed convergence for $(0,0,0)$. For $(1,0,1)$, the speed convergence can be deduced from the $m = 1$ speed convergence result as follows. First note that $I(X[1]; Y) \leq I[1](P) = I(X[1]; Y X[2])$. Then, it is shown that, if $I[1](P_n)$ tends to 1, it must be that along those paths of the B_n process, $\hat{I}[1](P_n)$ tends to 1 as well, where $\hat{I}[i](P) = I(X[i]; Y)$. Now, since $\hat{I}[1](P_n)$ tends to 1, it must tend fast from the single user result. A similar treatment can be done for $(0,1,1)$ and $(1,1,2)$. However, for $(1,1,1)$, another step is required. Indeed, for this case, $\hat{I}[1](P_n)$ and $\hat{I}[2](P_n)$ tend to zero. Hence, $\hat{I}[1,2](P) = I(X[1] + X[2]; Y)$ is introduced and it is shown that $\hat{I}[1,2](P_n)$ tends to 1. Moreover, if we denote by Q the single user channel between $X[1] + X[2]$ and Y, we have that $\hat{I}[1,2](P) = I(Q)$, $\hat{I}[1,2](P^-) = I(Q^-)$ and $\hat{I}[1,2](P^+) = I(U_2[1] + U_2[2]; Y_1 Y_2 U_1[1] U_1[2]) \geq I(U_2[1] + U_2[2]; Y_1 Y_2 U_1[1] + U_1[2]) = I(Q^+)$. Hence, using the single user channel result, $\hat{I}[1,2](P_n)$ tends to 1 fast. Note that a property of the matroids $\{(0,0,0), (0,1,1), (1,0,1), (1,1,1), (1,1,2)\}$ is that we can express any of them as the uniform rate region of a deterministic linear channel: $(1,0,1)$ is in particular the uniform rate region of the MAC whose output is $Y = X[1]$, $(0,1,1)$ corresponds to $Y = X[2]$, $(1,1,1)$ to $Y = X[1] + X[2]$ and $(1,1,2)$ to $(Y_1, Y_2) = (X[1], X[2])$.

Now, when $m = 3$, all matroids are also in a one to one correspondence with linear forms and a similar treatment to the $m = 2$ case is possible. This is related to the fact that any matroid on 2 or 3 elements can be represented in the binary field. We now introduce the definition of binary matroids.

Definition 6. *Linear matroids:* let A be a $k \times m$ matrix over a field. Let E_m be the index set of the columns in A. The rank of $J \subseteq E_m$ is defined by the rank of the sub-matrix with columns indexed by J.
Binary matroids: A matroid is binary if it is a linear matroid over the binary field. We denote by BMAT_m the set of binary matroids with m elements.

1) The $m = 4$ Case: We have that MAT_4 contains 17 unlabeled matroids (68 labeled ones). However, there are only 16 unlabeled binary matroids with ground state 4. Hence, there must be a matroid which does not have a binary representation. This matroid is given by $(1,1,1,1,2,2,2,2,2,2,2,2,2,2,2)$ (one can easily check that this is not a binary matroid). It is denoted $U_{2,4}$ and is the uniform matroid of rank 2 with 4 elements (for which any 2 elements set is a basis). Luckily, one can show that there is no MAC leading to $U_{2,4}$ and the following holds.

Lemma 4. $\mathcal{A}_4 \subset \mathrm{BMAT}_4 \subsetneq \mathrm{MAT}_4$.

Hence, the $m = 4$ case can be treated in a similar manner as the previous cases. We conclude this section by proving the following result, which implies Lemma 4.

Lemma 5. *$U_{2,4}$ cannot be the uniform rate region of any MAC with four users and binary uniform inputs.*

Proof: Assume that $U_{2,4}$ is the uniform rate region of a MAC. We then have

$$I(X[i,j]; Y) = 0, \tag{4}$$
$$I(X[i,j]; Y X[k,l]) = 2, \tag{5}$$

for all i, j, k, l distinct in $\{1,2,3,4\}$. Let y_0 be in the support of Y. For $x \in \mathbb{F}_2^4$, define $\mathbb{P}(x|y_0) = W(y_0|x) / \sum_{z \in \mathbb{F}_2^4} W(y_0|z)$. Then from (5), $\mathbb{P}(0,0,*,*|y_0) = 0$ for any choice of $*,*$ which is not $0,0$ and $\mathbb{P}(0,1,*,*|y_0) = 0$ for any choice of $*,*$ which is not $1,1$. On the other hand, from (4), $\mathbb{P}(0,1,1,1|y_0)$ must be equal to p_0. However, we have form (5) that $\mathbb{P}(1,0,*,*|y_0) = 0$ for any choice of $*,*$ (even for $1,1$ since we now have $\mathbb{P}(0,1,1,1|y_0) > 0$). At the same time, this implies that the average of $\mathbb{P}(1,0,*,*|y_0)$ over $*,*$ is zero. This brings a contradiction, since from (4), this average must equal to p_0. ■

Moreover, a similar argument can be used to prove a stronger version of Lemma 5 to show that no sequence of MACs can have a uniform rate region that converges to $U_{2,4}$.

2) Arbitrary values of m: We have seen in previous section that for $m = 2, 3, 4$, the extremal MACs are not any matroids but binary matroids. This allows us to conclude that $\{I_n[J], J \subseteq E_m\}$ must tend fast enough to $\{I_\infty[J], J \subseteq E_m\}$. Indeed, by working with the linear deterministic representation of the MACs, the problem of showing that the convergence speed is fast in the MAC setting is a consequence of the single-user setting result shown in [2]. We now show that this approach can be used for any values of m.

Definition 7. A matroid is BUMAC if its rank function r can be expressed as $r(J) = I(X[J]; YX[J^c])$, $J \subseteq E_m$, where $X[E]$ has independent and binary uniformly distributed components, and Y is the output of a binary input MAC.

Theorem 4. *A matroid is BUMAC if and only if it is binary.*

The converse of this theorem is easily proved and the direct part uses the following steps, which are detailed in [1]. First the following theorem on the representation of binary matroids due to Tutte, whose proof can be found in [5].

Theorem 5 (Tutte). *A matroid is binary if and only if it has no minor that is $U_{2,4}$.*

A minor of matroid is a matroid which is either a restriction or a contraction of the original matroid to a subset of the ground set. A contraction can be defined as a restriction on the dual matroid, which is another matroid whose bases are the complement set of the bases of the original matroid. Using Lemma 4, we already know that $U_{2,4}$ is not a restriction of any BUMAC matroid. To show that a BUMAC matroid cannot have $U_{2,4}$ as a contraction, Lemma 4 can be used in a dual manner, since one can show that the rank function of the dual of a BUMAC matroid is given by $r^*(J) = |J| - I(X[J]; Y)$.

Theorem 6. *Let $X[E]$ have independent and binary uniformly distributed components. Let Y be the output of a MAC with input $X[E]$ and for which $f(J) = I(X[J]; YX[J^c])$ is integer valued, for any $J \subseteq E_m$. We know from previous theorem that $f(\cdot)$ is also the rank function of a binary matroid, so let A be a matrix representation of this binary matroid. We then have*

$$I(AX[E]; Y) = \mathrm{rank} A = f(E_m).$$

The proof of previous theorem, with further investigations on this subject can be found in [1]. Moreover, one can show a stronger version of these theorems for MACs having a uniform rate region which tends to a matroid. Now, this result tells us that the extremal MACs are equivalent to linear deterministic channels. This suggests that we could have started from the beginning by working with $S[J](P) := I(\sum_{i \in J} X_i; Y)$ instead of $I[J](P) = I(X[J]; YX[J^c])$ to analyze the polarization of a MAC. The second measure is the natural one to study a MAC, since it characterizes the rate region. However, we have just shown that it is sufficient to work with the first measure for the purpose of the polarization problem considered here. Indeed, one can show that $S[J](P_n)$ tends either to 0 or 1 and Eren Şaşoğlu has provided a direct argument showing that these measures are fully characterizing the extremal MACs. Moreover, the process of identifying which matroids can have a rank function derived from an information theoretic measure, such as the entropy, has been investigated in different works, cf. [8] and references therein. In the next section, we summarize our polar code construction for the MAC.

D. Polar code construction for MACs

Let $n = 2^l$ for some $l \in \mathbb{Z}_+$ and let $G_n = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\otimes l}$ denote the lth Kronecker power of the given matrix. Let $U[k]^n := (U_1[k], \dots, U_n[k])$ and

$$X[k]^n = U[k]^n G_n, \quad k \in E_m.$$

When $X[E_m]^n$ is transmitted over n independent uses of P to receive Y^n, define the channel $P_{(i)} : \mathbb{F}_2^m \to \mathcal{Y}^n \times \mathbb{F}_2^{m(i-1)}$ to be the channel whose inputs and outputs are $U_i[E_m] \to Y^n U^{i-1}[E_m]$. Let $\varepsilon > 0$ and let $A[k] \subset \{1, \dots, n\}$ denote the sets of indices where information bits are transmitted by user k, which are chosen as follows: for a fixed $i \in \{1, \dots, n\}$, if $\|\{I[J](P_{(i)}) : J \subseteq E_m\} - \mathbb{B}\| < \varepsilon$ for some binary matroid $\mathbb{B}$ (where the distance above refers to the euclidean distance between the corresponding 2^m dimensional vectors), then pick a basis of $\mathbb{B}$ and include i in $A[k]$ if k belongs to that basis. If no such binary matroid exists, do not include i in $A[k]$ for all $k \in E_m$. Choose the bits indexed by $A[k]^c$, for all k, independently and uniformly at at random, and reveal their values to the transmitter and receiver.

For an output sequence Y^n, the receiver can then decode successively $U_1[E_m]$, then $U_2[E_m]$, etc., till $U_n[E_m]$. Moreover, since $I[E_m](P)$ is preserved through the polarization process (cf. the equality in (3)), we guarantee that for every $\delta > 0$, there exists a n_0 such that $\sum_{k=1}^m |A[k]| > n(I[E_m](P) - \delta)$, for $n \geq n_0$. Using the results of previous section, we can then show the following theorem, which ensures that the code described above allows reliable communication at sum rate.

Theorem 7. *For any $\beta < 1/2$, the block error probability of the code described above, under successive cancellation decoding, is $o(2^{-n^\beta})$.*

Moreover, this codes has an encoding and decoding complexity of $O(n \log n)$, from [2].

References

[1] E. Abbe, *Information, Matroid and Extremal Channels.* Preprint.

[2] E. Arıkan, *Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,* IEEE Trans. Inform. Theory, vol. IT-55, pp. 3051–3073, July 2009.

[3] E. Arıkan and E. Telatar, *On the rate of channel polarization,* in Proc. 2009 IEEE Int. Symp. Inform. Theory, Seoul, pp. 1493–1495, 2009.

[4] J. Edmonds, *Submodular functions, matroids and certain polyhedra,* Lecture Notes in Computer Science, Springer, 2003.

[5] J. Oxley, *Matroid Theory,* Oxford Science Publications, New York, 1992.

[6] E. Şaşoğlu, E. Telatar, E. Yeh, *Quasi-polarization for the two user binary input multiple access channel.* Preprint.

[7] D. Tse and S. Hanly, *Multi-access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities,* IEEE Trans. Inform. Theory, vol. IT-44, no. 7, pp. 2796-2815, November 1998.

[8] Z. Zhang and R. Yeung, *On characterization of entropy function via information inequalities,* IEEE Trans. on Information Theory, vol. 44, no. 4, pp. 1140-1452, 1998.

Capacity of a Modulo-Sum Simple Relay Network

Youvaraj T. Sagar, Hyuck M. Kwon, and Yanwu Ding
Electrical Engineering and Computer Science, Wichita State University
Wichita, Kansas 67260, USA, {ytsagar; hyuck.kwon; yanwu.ding}@wichita.edu

***Abstract*[1] –This paper presents the capacity of a modulo-sum simple relay network. In previous work related to this paper, capacity was characterized for the case where the noise was transmitted to the relay. And the closed-form capacity was derived only for the noise with a Bernoulli-$(1/2)$ distribution. However, in this paper, the source is transmitted to the relay, and a more general case of noise with an arbitrary Bernoulli-(p) distribution, $p \in [0, 0.5]$, is considered. The relay observes a corrupted version of the source, uses a quantize-and-forward strategy, and transmits the encoded codeword through a separate dedicated channel to the destination. The destination receives both from the relay and source. This paper assumes that the channel is discrete and memoryless. After deriving the achievable capacity theorem (i.e., the forward theorem) for the binary symmetric simple relay network, this paper proves that the capacity is strictly below the cut-set bound. In addition, this paper presents the proof of the converse theorem. Finally, this paper extends the capacity of the binary symmetric simple relay network to that of an *m*-ary modulo-sum relay network.**

Index Terms – Channel capacity; relay network; modulo-sum channel; quantize-and-forward; single-input single-output; cut-set bound.

I. INTRODUCTION

The relay network is a channel that has one sender and one receiver, with a number of intermediate nodes acting as relays to assist with the communications between sender and receiver. This paper exchanges the terminology of the relay channel in [1] with the relay network frequently because here a network is defined as a system consisting of more than two nodes [2], whereas a channel is for communication between two nodes. The simplest relay network or channel has one sender, one receiver, and one relay node. Fig. 1 shows this type of relay network, which is called a "simple" relay network.

The first original model of a relay network was introduced by van der Meulen in 1971 [3]. After that, extensive research was done to find the upper bounds, cut-set bounds, and exact capacity for this network. In 1979, Cover and El Gamal obtained the capacity for a special class of channels called physically degraded relay channels [4]. In that paper, they discussed the capacity of the relay channel with feedback and found an upper bound for a simple relay network, which is shown in Fig. 1. Later, El Gamal and Aref found the capacity for a special class of relay channels called "semideterministic relay channels" in [5]. Then, Kim found the capacity for a class of deterministic relay channels in [6], where he modeled the simple relay network as a noiseless channel between the relay and the destination. Also, van der Meulen corrected his previous upper bound on the capacity of the simple relay network with and without delay in a paper [7].

Using Kim's results in [6], Aleksic et al. modeled the channel between the relay and the destination as a modular sum noise channel in [8]. Binary side information or channel state information is transmitted to the relay in [8]. He mentioned that the capacity of the simple relay network is not yet known. Recently, Tandon and Ulukus found a new upper bound for the simple relay network with general noise, obtained the capacity for symmetric binary erasure relay channel, and compared them with the cut-set bound in [9].

Aleksic et al. in [8] introduced a corrupting variable to the noiseless channel in [6], whereby the noise in the direct channel between the source and the destination is transmitted to the relay. The relay observes a corrupted version of the noise and has a separate dedicated channel to the destination. For this case, the capacity was characterized in [8]. However, the closed-form capacity was derived only for the noise with a Bernoulli-$(p = 1/2)$. distribution.

The objective of this paper is to find the capacity of the simple modular sum relay network and show that its capacity is strictly below the cut-set bound [4]. This paper also presents a closed-form capacity for a general case, such as for any p where the source is transmitted to both the relay and the destination.

This paper considers all noisy channels, i.e., from the source to the destination, from the source to the relay, and from the relay to the destination, as shown in Fig.

[1] This work was partly sponsored by the Army Research Office under DEPSCoR ARO Grant W911NF-08-1-0256, and by NASA under EPSCoR CAN Grant NNX08AV84A.

1, where all noisy channels are binary symmetric channels (BSCs) with a certain crossover probability, e.g., p. This paper also derives the capacity for this class of relay channels. In other words, the capacity of a modulo-sum simple relay network is presented here. The capacity proof for the binary symmetric simple relay network and the proof for the converse depend crucially on the input distribution.

For the BSC, a uniform input distribution at the source is assumed because this distribution maximizes the entropy of the output (or the capacity) regardless of additive noise. Furthermore, because of the uniform input distribution, the output of a binary symmetric relay network is independent of additive noise. After presenting the proof for the capacity of a binary symmetric simple relay network, this paper proves that the capacity obtained is strictly below the cut-set bound by using the results in [4]. Finally, this paper shows the converse theorem for this class of networks.

Section II describes the system model and presents the capacity of the binary symmetric simple relay network. Section III discusses the cut-set bound for the binary symmetric simple relay network and presents the numerical analysis results. Section IV extends the capacity to the m-ary modular additive case. Finally, Section V concludes the paper.

II. SYSTEM MODEL AND NETWORK CAPACITY

Fig. 2 shows a realistic binary phase-shift keying (BPSK) system under additive white Gaussian noise (AWGN), where X and Y are the binary input and output signal, respectively. Here, Y is obtained with a hard decision on the demodulated signal.

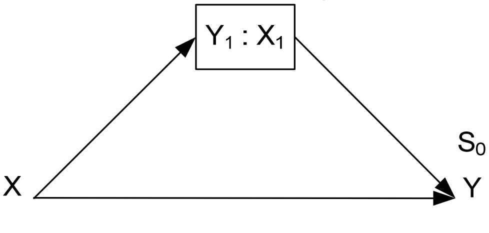


Fig. 1. Simple relay network.

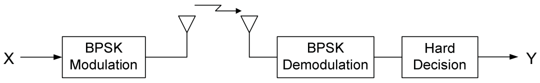


Fig. 2. Realistic BPSK communication system under AWGN.

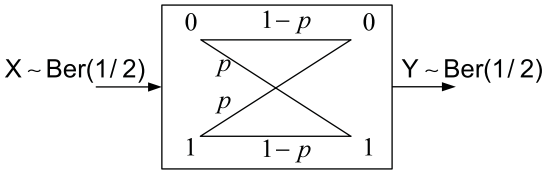


Fig. 3. BSC equivalent to Fig. 2.

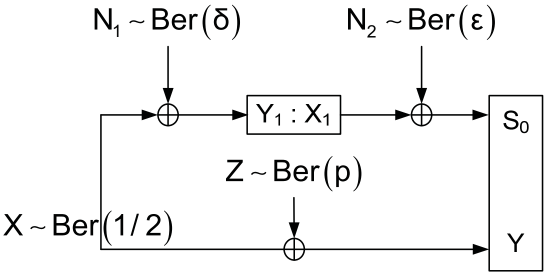


Fig. 4. Binary symmetric simple relay network.

Fig. 3 shows a BSC with the crossover probability p equivalent to the realistic communication system in Fig. 2. Here, the crossover probability p is equal to $Q\left[\sqrt{2E_b/N_0}\,\right]$, where $Q(\alpha) = \int_\alpha^\infty (1/\sqrt{2\pi})\, e^{-t^2/2}dt$, and E_b and N_0 denote the bit energy and the one-side AWGN power spectral density, respectively.

This paper models a channel between any adjacent nodes in Fig. 1 as a BSC that has one sender, one receiver (or destination), and one relay node [1]. The random variable Y represents the received signal through the direct channel and is written as $Y = X \oplus Z$, where X and Z denote the transmitted and noise random variable with distribution $Ber(1/2)$ and $Ber(p)$, respectively, and $\oplus$ denotes the binary modulo-sum, i.e., $Z = 1$ with probability p, and $Z = 0$ with probability $(1 - p)$.

The simple relay network in Fig. 1 can be redrawn as Fig. 4. Here, the relay node has an input Y_1 and an output X_1. The relay node observes the corrupted version of X, i.e., $Y_1 = X \oplus N_1$, encodes it using a codebook $\mathcal{U}^n$ of jointly typical strong sequences [1], and transmits the code symbol X_1 through another separate BSC to the destination node, where $\mathcal{U}$, n, and N_1 denote the alphabet of code symbols, the codeword length, and the noise random variable at the relay with distribution $Ber(\delta)$, respectively. The destination receives both Y, through the direct channel, and $S_0 = X_1 \oplus N_2$, through the relay node, where $N_2{\sim}Ber(\varepsilon)$ represents the noise at the destination for the relay network.

Note that the binary modulo-sum and the BSC can be extended to an m–ary modulo-sum and an m–ary symmetric channel (MSC).

To the authors' knowledge, there is no network capacity expression in the literature, even for the simple relay network shown in Fig. 4. Only the capacity of a deterministic relay channel, i.e., the case of $N_2 = 0$ in Fig. 4, is presented in [6]. The capacity of a relay network by replacing X with Z, i.e., the case where the relay observes a corrupted version of the direct channel noise Z, is presented in [8]. This paper presents the capacity of the simple relay network shown in Fig. 4 in the following theorem.

Theorem 1: The capacity C of the binary symmetric simple relay network shown in Fig. 4 is

$$C = \max_{p(u|y_1):I(U;Y_1)\le R_0} \{1 + H(Y|U) - H(Z) - H(X|U)\}, \quad (1)$$

where the maximization is over the U's conditional probability density function (p.d.f.) given Y_1; the cardinality of the alphabet $\mathcal{U}$, is bounded by $|\mathcal{U}| \le |\mathcal{Y}_1| + 2$; and R_0 is the capacity for the channel between X_1 and S_0, which can be written as

$$R_0 = \max_{p(x_1)} I(X_1; S_0). \quad (2)$$

The closed-form network capacity for the simple relay network shown in Fig. 4 can be written as

$$C = 1 + \mathcal{H}(\{\varepsilon * \delta\} * p) - \mathcal{H}(p) - \mathcal{H}(\varepsilon * \delta). \quad (3)$$

Here, $H(X)$ and $I(X;Y)$ are the entropy of X and the mutual information between X and Y, respectively [1]; $\mathcal{H}(\alpha)$ is the binary entropy function written as $\mathcal{H}(\alpha) = -\alpha\log_2\alpha - (1-\alpha)\log_2(1-\alpha)$; and $\alpha * \beta = \alpha(1-\beta) + (1-\alpha)\,\beta$ [10].

Proofs of the converse and achievability for this theorem are provided in appendices A and B of [12]. Proofs of other theorems in this paper are also in [12].

Note that if the direct channel noise Z is transmitted through the relay rather than X, then (1) becomes (3) of [8] or (4), written as

$$C = \max_{p(u|y_1):I(U;Y_1)\le R_0} 1 - H(Z|U). \quad (4)$$

This is because $H(Y|U)$ and $H(Z)$ in (1) will become 1 and $H(X) = 1$, respectively.

III. CUTSET BOUND AND ANALYTICAL RESULTS

This section shows that the capacity of the binary symmetric simple relay network in Fig. 4 is strictly below the cut-set bound, except for the two trivial points at $R_0 = 0$ and $R_0 = 1$ when $p = 0.5$. The capacity in (1) can be upper-bounded by the cut-set bound as

$$C \le \max_{p(x,x_1)} \min\{I(X,X_1;Y,S_0),\ I(X;Y,Y_1)\} \quad (5)$$

where the Ford-Fulkerson theorem [11], [4] is applied to the simple relay network in Fig. 4. Using (5), Theorem 2 can be established.

Theorem 2: The cut-set bound for the capacity of the binary simple relay network shown in Fig. 4 can be written as

$$C \le \min\{1 - H(Z) + R_0, 1 - H(Z) + 1 - H(N_1)\}$$
$$= \min\{1 - \mathcal{H}(p) + R_0,\ 1 - \mathcal{H}(p) + 1 - \mathcal{H}(\delta)\}. \quad (6)$$

Figs. 5(a) and 5(b) show the capacity in bits per transmission versus R_0 bits for $\delta = 0.1$, when $p = 0.1$ and $p = 0.5$, respectively. If $p = 0.5$, then the results are the same as those in [8]. Only the closed form of the capacity for the special case of $p = 0.5$ was analyzed and presented in [8], where the capacity C of the binary simple relay network was obtained by replacing X with Z at the relay input shown in Fig. 4 and written as [8]

$$C = 1 - \mathcal{H}(\mathcal{H}^{-1}\{1 - R_0\} * \delta). \quad (7)$$

Here $\mathcal{H}^{-1}(\cdot)$ is the inverse of $\mathcal{H}(p)$ in the domain $p \in [0,0.5]$. Note that the capacity in (3) of this paper is valid for a general p between 0 and 0.5, whereas the one in (34) of [8] or (7) is valid for only $p = 0.5$.

Note that the capacity in (3) is strictly below the cut-set bound in (6). Refer to Figure 5(b).

IV. CAPACITY FOR M-ARY MODULO-SUM RELAY NETWORK

This section extends the capacity derived for the binary symmetric simple relay network to the m-ary modular additive relay network. The received signal at the destination node can be written as $Y = X + Z\ (mod\ m)$. The relay observes the corrupted version of X, i.e., $Y_1 = X + N_1\ (mod\ m)$, and the relay also has a separate channel to the destination: $S_0 = X_1 + N_2\ (mod\ m)$ with a capacity $R_0 = \max_{p(x_1)} I(X_1; S_0)$. Therefore, (1) becomes (8) in Theorem 3.

Theorem 3: The capacity C of the symmetric m-ary modulo-sum simple relay network is

$$C = \max_{p(u|y_1):I(U;Y_1)\le R_0} \{m + H(Y\,|\,U) - H(Z) - H(X\,|\,U)\} \quad (8)$$

where maximization is over the conditional U's p.d.f. given Y_1 with $|\mathcal{U}| \le |\mathcal{Y}_1| + 2$, and R_0 is defined in (2).

Proof: The achievability for Theorem 3 follows the same steps as Theorem 1 by changing the binary to the m-ary case. Also, the uniform input distribution at the source maximizes the entropy of the output, regardless of the additive noise. Furthermore, because of the uniform input distribution, the output of an m-ary modulo-sum relay network is independent of the additive noise. Therefore, (8) holds true. The converse for Theorem 3 also holds true using the same steps of Theorem 1 by changing the binary modulo-sum to the m-ary modulo-sum.

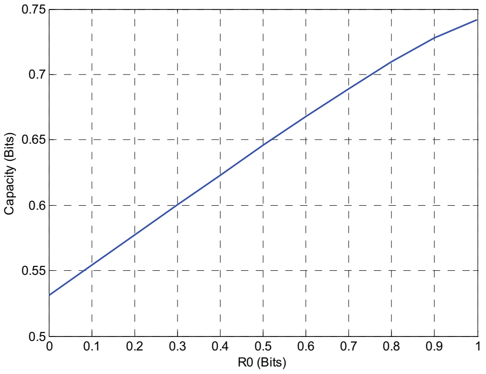


Fig. 5(a). Capacity of a binary symmetric simple relay network shown in Fig. 4 for $\delta = 0.1$ and $p = 0.1$.

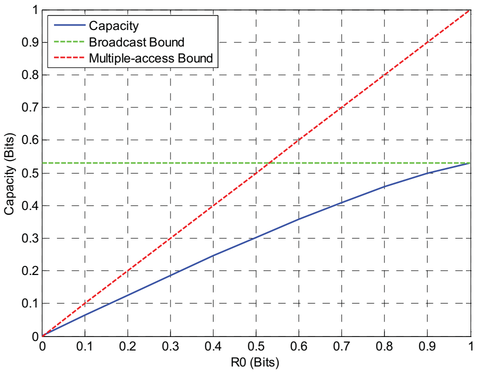


Fig. 5(b). Capacity and cutest bounds of a binary symmetric simple relay network shown in Fig. 4 for $\delta = 0.1$ and $p = 0.5$.

V. CONCLUSIONS

It has been an open problem to find the capacity of the simple relay network. This paper presented the closed form capacity of the binary symmetric simple relay network. Also, this paper extended the capacity for the binary to the m-ary modulo-sum symmetric simple relay network. Two conditions are necessary for the derivation of this capacity: (1) a uniform Bernoulli-$(1/2)$ input distribution, and (2) a modular additive channel between the two adjacent nodes. Using these conditions, both proofs for the achievability and the converse of the capacity theorem were presented. Furthermore, this paper derived the cut-set bound and presented the numerical results for this network. Finally, this paper claimed that the capacity is strictly below the cut-set bound and achievable using a quantize-and-forward strategy at the relay.

REFERENCES

1. T. M. Cover and J. A. Thomas, “Elements of Information Theory,” second edition, New York: Wiley, 2006.
2. M. Schwartz, “Telecommunication Networks: Protocols, Modeling and Analysis,” Menlo Park, CA: Addison-Wesely Publising Co., 1988.
3. E. C. Van Der Meulen. “Three-Terminal Communication Channels,” *Adv. Appl. Prob*., vol. 3, pp. 120-154, 1971.
4. T. M. Cover and El Gamal, “Capacity Theorems for the Relay Channel,” *IEEE Trans. Inf. Theory,* vol. IT-25, no. 5, pp. 572-584, Sept. 1979.
5. A. El Gamal and M. Aref, “The Capacity of the Semideterministic Relay Channel,” *IEEE Trans. on Inf. Theory,* vol. IT-28, no. 3, p. 536, May 1982.
6. Y. H. Kim, “Capacity of a Class of Deterministic Relay Channels,” *IEEE Trans. on Inf. Theory,* vol. IT-54, no. 3, pp. 1328-1329, Mar. 2008.
7. E. C. Van Der Meulen and P. Vanroose, “The Capacity of a Relay Channel, Both with and without Delay,” *IEEE Trans. on Inf. Theory,* vol. 53, no.10, pp. 3774-3776, Oct. 2007.
8. M. Aleksic, P. Razaghi, and W. Yu, “Capacity of a Class of Modulo-Sum Relay Channels,” *IEEE Trans. on Inf. Theory*, vol. 55, no. 3, pp. 921-930, Mar. 2009.
9. R. Tandon and S. Ulukus, “A New Upper Bound on the Capacity of a Class of Primitive Relay Channels,” *Communication, Control, and Computing, 2008 46th Annual Allerton Conference,* pp. 1562-1569, Sept. 2008.
10. A. D. Wyner and J. Ziv, “A Theorem on the Entropy of Certain Binary Sequences and Applications” *IEEE Trans. Inf. Theory,* vol. IT-19, no. 6, pp. 769-777, Nov. 1973.
11. L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton, NJ: Princeton University Press, 1962.
12. Youvaraj T. Sagar, Hyuck M. Kwon, and Yanwu Ding, "Capacity of a Modulo-Sum Simple Relay Network," submitted to *IEEE Transactions on Information Theory* in September 2009, and is available at the website: http://webs.wichita.edu/?u=ECE&p=/Wireless/Publications/

Improved Related-Key Impossible Differential Attacks on 8-Round AES-256

Hadi Soleimany, Alireza Sharifi, Mohammadreza Aref
Information Systems and Security Lab (ISSL)
EE Department, Sharif University of Technology, Tehran, Iran
E-mail: hadi.soleimany@gmail.com, asharifi@alum.sharif.edu, aref@sharif.edu

***Abstract*—In this paper, we propose two new related-key impossible differential attacks on 8-round AES-256, following the work of Zhang, et al. First, we propose a carefully chosen relation between the related keys, which can be extended to 8-round subkey differences. Then, we construct a 5.5-round related-key impossible differential. Using the differential, we present two new attacks on the 8-round AES-256 with 32 and 64 bit structures. Our 8-round AES-256 attacks leads to the best known attack on AES-256 with 2 related keys. The time complexity of the proposed related-key impossible differential attacks on 8-round AES-256 is $2^{102.5}$ and its data complexity is $2^{103.5}$.**

keywords: AES-256, related-key differentials cryptanalysis, impossible differential

I. Introduction

Rijndael [1] is an iterated block cipher with variable key and block lengths of 128 to 256 bits in steps of 32 bits. Rijndael versions with a block length of 128 bits, and key lengths of 128,192 and 256 bits have been adopted as the Advanced Encryption Standard (AES). Differential cryptanalysis [2] analyzes the evolvement of the difference between a pair of plaintexts in the following round outputs (differentials) in an iterated block cipher. The basic idea of impossible differential attack is to look for differentials that hold with probability 0 (or impossible differentials) to eliminate wrong keys and keep the right key. Related-key attacks [3], concentrate on the information which can be obtained from two encryptions using related (but unknown) keys. Related-key impossible differential attack [4] combines related-key attack and impossible differential cryptanalysis to make the attack more efficient.

The first impossible differential attack against AES was applied to 5 rounds of the AES-128 by Biham and Keller [5]. In [4], the first related-key impossible differential attack on 192-bit variants was proposed. Zhang, et. al applied three new related-key impossible differential attacks on 8-round AES-192 [6] and AES-256 [7] and concluded AES-256 has better resistance than AES-192 using the same cryptanalytic approach [7]. In this paper, we show that 8 round AES-256 can be attacked more efficient than 8 round AES-192 from overall complexity. We present 2 related-key impossible differential attacks on 8-round AES-256 with 2 related keys. Our 8-round AES-256 attacks leads to the best known attack on 8-round AES-256 with 2 related keys.

The paper is organized as follows: In Section II we briefly describe the AES algorithm. A new related-key impossible differential property of the AES-256 is introduced in Section III. In Section IV, using 64-bit structures, we propose a related-key impossible differential attack on the 8-round AES-256. In Section V we compare the performance of our attacks with the previous ones.

II. A brief description of AES

In AES [1] a 128-bit plaintext is represented by a 4×4 matrix of bytes, where each byte represents a value in $GF(2^8)$. An AES round is composed of four operations: SubBytes (SB), ShiftRows (SR), MixColumns (MC) and AddRoundKey (AK). The MixColumns operation is omitted in the last round and an initial key addition is performed before the first round for whitening. We also assume that the MixColumns operation is omitted in the last round of the reduced-round variants. The number of rounds is variable depending on the key length, 10 rounds for 128-bit key, 12 for 192-bit key and 14 for 256-bit key.

A. Notations

In this paper we use the following notations: X_i^I denotes the input block of round i, while X_i^S,X_i^R,X_i^M and X_i^O denotes intermediate values after applying SubBytes, ShiftRows, MixColumns and AddRoundKey operations of round i, respectively. Obviously, $X_{i-1}^O = X_i^I$ holds for $i \geq 2$. We denote the subkey of the i-th round by k_i and the initial whitening subkey by k_0. In some cases, we are interested in interchanging the order of the MixColumns operation and the Subkey Addition. As these operations are linear, they can be interchanged, first XORing the data with an equivalent key and then applying the MixColumns operation. We denote equivalent subkey for the modified version by w_i, i.e. $w_i = MC^{-1}(k_i)$, and X_i^W denotes the intermediate value after applying AddRoundKey with equivalent subkey. Let $X_{i,col(j)}$ denotes the j-th column of x_i where $j \in \{0,1,2,3\}$. We also denote the byte in the m-th row and n-th column of X_i by $X_{i,m,n}$ where $m,n \in \{0,1,2,3\}$. Another notation for bytes of x_i is an enumeration $\{0,1,2,...,15\}$ where the byte $X_{i,m,n}$ corresponds to byte $4n+m$ of X_i.

[0] This work was partially supported by Iran Telecommunications Research Center and the cryptography chair of the Iranian NSF.

III. 5.5-Round Related-Key Impossible Differential Property of AES-256

In this paper, using a property of MixColumns operation, we propose a new 5.5-round related-key impossible differential property which our attack is based on. First of all we use the following definitions: A byte which has different values (nonzero difference) in a pair is called an active byte while passive byte is a byte with zero difference in a pair. Now we state and prove the MixColumns property:

Theorem 3.1: A pair of columns at the input of MixColumns operation which contains two passive bytes cannot lead to two passive bytes and one or two active bytes within the output column.

Proof: Suppose that $\Delta X = (\Delta X_1, \Delta X_2, \Delta X_3, \Delta X_4)$ is the difference of input column and $\Delta Y = (\Delta Y_1, \Delta Y_2, \Delta Y_3, \Delta Y_4)$ is the corresponding output difference. Using MixColumns operation we have:

$$\begin{aligned}
\Delta Y_1 &= 02 \bullet \Delta X_1 \oplus 03 \bullet \Delta X_2 \oplus 01 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 \\
\Delta Y_2 &= 01 \bullet \Delta X_1 \oplus 02 \bullet \Delta X_2 \oplus 03 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 \\
\Delta Y_3 &= 01 \bullet \Delta X_1 \oplus 01 \bullet \Delta X_2 \oplus 02 \bullet \Delta X_3 \oplus 03 \bullet \Delta X_4 \\
\Delta Y_4 &= 03 \bullet \Delta X_1 \oplus 01 \bullet \Delta X_2 \oplus 01 \bullet \Delta X_3 \oplus 02 \bullet \Delta X_4
\end{aligned}$$

where "$\bullet$" is modular multiplication of Rijndael [1]. Without loss of generality, suppose X_1 and X_2 are two passive bytes, i.e. $\Delta X_1 = \Delta X_2 = 0$, we would have:

$$\begin{aligned}
\Delta Y_1 &= 01 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 \\
\Delta Y_2 &= 03 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 \\
\Delta Y_3 &= 02 \bullet \Delta X_3 \oplus 03 \bullet \Delta X_4 \\
\Delta Y_4 &= 01 \bullet \Delta X_3 \oplus 02 \bullet \Delta X_4
\end{aligned}$$

So if two bytes of output column, for example Y_1 and Y_2 have zero difference, i.e. $\Delta Y_1 = \Delta Y_2 = 0$, we will have the following system of equations:

$$\begin{aligned}
01 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 &= 0 \\
03 \bullet \Delta X_3 \oplus 01 \bullet \Delta X_4 &= 0
\end{aligned}$$

It is obvious that the only solution of the above system is $\Delta X_3 = \Delta X_4 = 0$ and consequently $\Delta Y_3 = \Delta Y_4 = 0$, i.e. the output column cannot have one or two active bytes. ■

Consider the difference between two related keys as follows: $\Delta K = K_1 \oplus K_2 = [(a,0,0,0),(a,0,0,0),(a,0,0,0),(a,0,0,0),(0,0,0,0),(0,0,0,0),(0,0,0,0),(0,0,0,0)]$.

Such a difference results in the round subkey differences as shown in Table I.

Using the above subkey differences and Theorem 3.1, we build a 5.5-round related-key impossible differential with probability equal to 1. The 5.5-round related-key impossible differential is:

$\Delta X_1^M = ((0,?,0,?),(?,0,?,0),(0,?,0,?),(?,0,?,0)) \nrightarrow$
$\Delta X_6^O = ((a,0,0,0),(0,0,0,0),(0,0,0,0),(0,0,0,0))$

where '***a***' is a known nonzero value and '**?**' denotes any value. Let $\Delta X_1^M = ((0,?,0,?),(?,0,?,0),(0,?,0,?),(?,0,?,0))$. From Table 1, Δk_1 is zero and it results in $\Delta X_2^I = \Delta X_1^O = \Delta X_1^M \oplus \Delta k_1 = ((0,?,0,?),(?,0,?,0),(0,?,0,?),(?,0,?,0))$ which leads to $\Delta X_2^R = ((0,0,0,0),(?,?,?,?),(0,0,0,0),(?,?,?,?))$ and as a result $\Delta X_2^M = ((0,0,0,0),(?,?,?,?),(0,0,0,0),(?,?,?,?))$. After adding the Δk_2 we have $\Delta X_3^I = \Delta X_2^O = ((a,0,0,0),(?,?,?,?),(a,0,0,0),(?,?,?,?))$ and after SubBytes and ShiftRows, we get $\Delta X_3^R = ((N,?,0,?),(?,0,?,0),(N,?,0,?),(?,0,?,0))$ where '***N***' denotes nonzero difference (possibly distinct). The second 3.5-round differential in the reverse direction is built as follows: The output difference $\Delta X_6^O = ((a,0,0,0),(0,0,0,0),(0,0,0,0),(0,0,0,0))$ is canceled by the subkey difference of the sixth round, i.e. $\Delta X_6^M = \Delta k_6 \oplus \Delta X_6^O = 0$. The zero difference ΔX_6^O is preserved through all the operations until the AddRoundKey operation of the fourth round, because the subkey difference of the fifth round is zero. Thus we have $\Delta X_4^M = \Delta k_4 \oplus \Delta X_4^O = ((a,0,0,0),(a,0,0,0),(0,0,0,0),(0,0,0,0))$ and consequently from Theorem 3.1 $\Delta X_4^R = ((N,N,N,N),(N,N,N,N),(0,0,0,0),(0,0,0,0))$. When rolling back the ΔX_4^R through the ShiftRows and SubBytes operations in the fourth round, we get the $\Delta X_3^O = \Delta X_4^I = ((N,0,0,N),(N,N,0,0),(0,N,N,0),(0,0,N,N))$. Finally after applying the AddRoundKey operation of the third round which has a zero difference, we can get $\Delta X_3^M = ((N,0,0,N),(N,N,0,0),(0,N,N,0),(0,0,N,N))$. It is obvious that $\Delta X_4^M = MC(\Delta X_4^R)$, but according to the Theorem 3.1, this is impossible, because ΔX_4^R has two passive bytes ΔX_4^M has two active bytes and two passive bytes.

Table I
Subkey Differences Required for the 5.5-Round Impossible Differential

Round (i)	$\Delta k_{i,col(0)}$	$\Delta k_{i,col(1)}$	$\Delta k_{i,col(2)}$	$\Delta k_{i,col(3)}$
0	$(a,0,0,0)$	$(a,0,0,0)$	$(a,0,0,0)$	$(a,0,0,0)$
1	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
2	$(a,0,0,0)$	$(0,0,0,0)$	$(a,0,0,0)$	$(0,0,0,0)$
3	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
4	$(a,0,0,0)$	$(a,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
5	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
6	$(a,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
7	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$
8	$(a,0,0,0)$	$(a,0,0,0)$	$(a,0,0,0)$	$(a,0,0,0)$

IV. Related-Key Impossible Differential Attack on 8-Round AES-256 using 64-bit structures

Using the above related-key impossible differential, we can attack an 8-round variant of AES-256.

A. The Attack Procedure

In order to make the attack faster, we first perform a precomputation. For all possible pairs of values of $x_{1,col(0)}^M$ and $x_{1,col(3)}^M$ which have the difference $\Delta x_{1,col(0)}^M = (a,?,?,0)$ and $\Delta x_{1,col(3)}^M = (?,?,0,0)$, compute the values of (0,1,5,6,10,11,12,15) for x_1^I. Store the pairs of 8-byte values in a hash table H_p indexed by the XOR difference in these

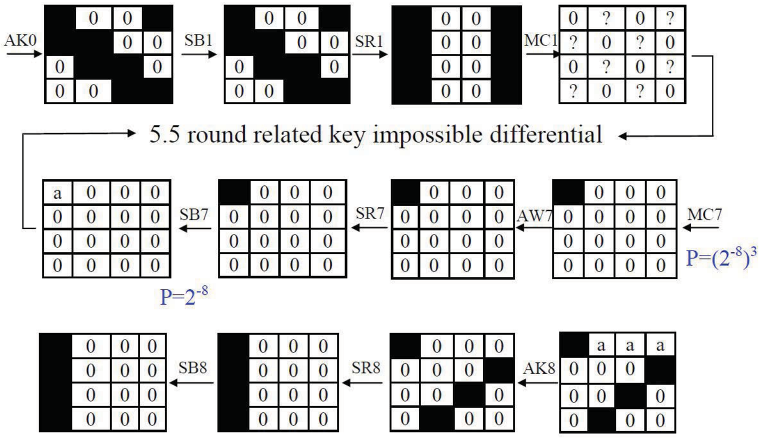


Figure 1. 8-round Impossible Differential Attack

bytes. There are 2^{64} possible values for the XOR difference in 8 bytes and $(2^{16})^4 \times (2^8)^4 = 2^{96}$ possible pairs of values of $x^M_{1,col(0)}$ and $x^M_{1,col(3)}$ with above condition. So H_p have 2^{64} bins and on average there are $\frac{2^{96}}{2^{64}} = 2^{32}$ pairs in each bin. The algorithm is as follows:

1. Generate two pools S_1 and S_2 of m plaintexts each, such that for each plaintext pair $P_1 \in S_1$ and $P2 \in S_2$, $P_1 \oplus P_2 = ((?, ?, 0, 0), (a, ?, ?, 0), (a, 0, ?, ?), (?, 0, 0, ?))$, where '**?**' denotes any byte value. Here we define a structure as a set of 2×2^{64} plaintexts which are selected from S_1 and S_2. Such a structure proposes $2^{64} \times 2^{64} = 2^{128}$ pairs of plaintexts.

2. Ask for the encryption of the pool S_1 under K_1, and of the pool S_2 under K_2. Denote the ciphertexts of the pool S_1 by T_1, and the encrypted ciphertexts of the pool S_2 by T_2.Such a structure proposes $2^{64} \times 2^{64} = 2^{128}$ pairs of plaintexts.

3. For all ciphertexts $C_2 \in T_2$, compute $C_2^* = C_2 \oplus ((0, 0, 0, 0), (a, 0, 0, 0), (a, 0, 0, 0), (a, 0, 0, 0))$.

4. Insert all the ciphertexts $C_1 \in T_1$ and the values $\{C_2^* | C_2 \in T_2\}$ into a hash table indexed by bytes 1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 14 and 15.

5. For each bin of the hash table with more than one ciphertext, select every pair (C_1, C_2). Note that every pair (C_1, C_2^*) in each bin of this hash table have zero difference in bytes 1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 14 and 15, so the pairs (C_1, C_2) have zero difference in bytes 1, 2, 3, 5, 6, 9, 11, 14 and 15, and difference 'a' in bytes 4,8 and 12. After these steps we expect to have $2^n \times 2^{128} \times (2^{-8})^{12} = 2^{n+32}$ plaintext pairs, where 2^n is the number of structures, whose corresponding ciphertext pairs are equal in bytes 1, 2, 3, 5, 6, 9, 11, 14 and 15, and difference 'a' in bytes 4,8 and 12.

6. Guess the 32-bit value at bytes 0, 7, 10 and 13 for the k_8. Decrypt partially these bytes in the last round, i.e. compute $x^O_{7,Col(0)} = SB^{-1} \circ SR^{-1}(x^O_8(0, 7, 10, 13) \oplus k_8(0, 7, 10, 13))$. Choose pairs whose difference $\Delta x^W_{7,col(0)} = MC^{-1}(\Delta x^O_{7,col(0)})$ are nonzero at byte ${}^{W}_{7,0,0}$ and zero at other three bytes. The probability of such a difference is $(2^{-8})^3 = 2^{-24}$.

7. Guess the value of subkey byte $w_{7,0,0}$ and compute $x^O_{6,0,0} = SB^{-1} \circ SR^{-1}(x^W_{7,0,0} \oplus w_{7,0,0})$ for all remaining pairs and choose pairs whose difference $\Delta x^O_{6,0,0}$ are '***a***'. The probability of such a difference is 2^{-8}. Thus, at the end of this step, we can get $2^n \times 2^{128} \times (2^{-8})^{-12} \times 2^{-24} \times 2^{-8} = 2^n$ pairs which have zero difference in all bytes except the first byte which have the difference '***a***'.

8. In this step, we eliminate wrong 64-bit values at (0,1,5,6,10,11,12,15) for the k_0 by showing that the impossible differential property holds, if these keys were used. We use the hash table H_p which has made in the precomputation stage. The algorithm of this step is as follows:

- Initialize a list A of the 2^{64} possible values at (0,1,5,6,10,11,12,15) of k_0.
- For each remaining pairs (P_1, P_2), compute $P' = P_1 \oplus P_2$ in the eight bytes (0,1,5,6,10,11,12,15).
- Access the bin P' in H_p, and for each pair (x,y) in that bin, $P_1 \oplus x$ remove from the list A the value, where P_1 is restricted to eight bytes.
- If A is not empty, output the values in A.

Note that there are 2^{32} pairs in each bin of H_p on average, so in the third part of this step, we eliminate about 2^{32} wrong keys for each plaintext pair (P_1, P_2). The probability of a wrong 64-bit value at bytes (0,1,5,6,10,11,12,15) for k_0 is $(1 - 2^{-64})$, so after analyzing all 2^n pairs, we expect only $2^{64} \times (1 - 2^{-64})^{2^{n+32}}$ wrong values of the eight bytes of k_0 remain. For $n = 38.5$, the expected number is about $2^{64} \times (1 - 2^{-64})^{2^{64} \times 2^{6.5}} \approx 2^{64} \times (e^{-1})^{2^{6.5}} \approx 2^{-67}$ and we can expect that only the right subkey remains. Unless the initial guess of the 32-bit value of the last round key k_8 or the 8-bit value of the key w_7 is correct, it is expected that we can eliminate the whole 64-bit value of k_0 in this step, i.e. the list A will be empty at the end of this step. Since the wrong values for k_8, w_7, k_0 occur with the small probability of $(2^8)^4 \times 2^8 \times 2^{-67} = 2^{-27}$. Hence if the list A is not empty, we can assume that the guessed 32-bit value for k_8 and 8-bit value for w_7 are correct.

B. The Attack Complexity

The data complexity of the attack is $2 \times 2^{n+64} = 2^{103.5}$ chosen plaintexts. The time complexity of the attack is consisted of three parts:

Step 6 requires $2 \times 2^{32} \times 2^{n+32} \times \frac{4}{16} = 2^{n+63}$ one round encryptions, because we must guess 2^{32} keys in this step, compute $X^{W}_{7,Col(0)}$ for each 2^{n+32} remained pairs from last steps.

Step 7 requires $2 \times 2^8 \times 2^{32} \times 2^{n+8} \times \frac{1}{16} = 2^{n+45}$ one round encryptions, because for all of guessed 2^{32} keys, we must guess 2^8 for k_8 and compute $X^{O}_{6,0,0)}$ for each 2^{n+8} remained pairs from last steps.

In step 8, 2^{n-64} pairs are analyzed. For each pair we need 2^{32} memory accesses to H_p and 2^{32} memory accesses to list A on average. This step is repeated 2^{40} times (for the guess of w_7 and k_8). Therefore the time complexity is $2^{40} \times 2^n \times (2^{32} + 2^{32}) = 2^{n+73}$ memory accesses, which are equivalent to about 2^{n+67} one round encryption (according to the implementations of NESSIE primitives [11]). Consequently for $n = 38.5$ the overall time complexity of the attack on 8-round AES-256 is about $\frac{2^{101.5}+2^{83.5}+2^{105.5}}{8} \approx 2^{102.5}$. The precomputation stage requires about $\frac{2 \times 2^{96}}{8} = 2^{94}$ encryptions and the required memory is about 2^{100} bytes. Meanwhile, $\frac{2^{64+8+32}}{2^3} = 2^{101}$ bytes of memory are needed to store the list of deleted key values k_8, w_7, k_0 for the attack.

To achieving an attack with lower time complexity which is decreased by the factor 2^{20}, at the cost of increasing data complexity by the factor $2^{15.5}$, we can use 32-bit structures instead of 64-bit structures. Like using 64-bit structures, we first perform a precomputation. For all possible pairs of values of $x^{M}_{1,col(0)}$ which has the difference $\Delta x^{M}_{1,col(0)} = (a, ?, ?, 0)$, compute the values of (0,5,10,15) for x^{I}_{1}. Store the pairs of 4-byte values in a hash table H_p indexed by the XOR difference in these bytes. There are 2^{32} possible values for the XOR difference in 4 bytes and $(2^{16})^2 \times (2^8)^2 = 2^{48}$ possible pairs of values of $x^{M}_{1,col(0)}$ with above condition. So H_p have 2^{32} bins and on average there are $\frac{2^{48}}{2^{32}} = 2^{16}$ pairs in each bin. The rest of the attack procedure is similar to 64-bit structure attack which we explain in this section.

V. Results and Discussion

In this paper, we proposed two new related-key impossible differential attacks on 8-round AES-256. Results in this paper are summarized in Table 2 and are compared with the previous attacks on 8-round AES-256. Attack on 8-round AES-256 with 64 bit structure leads to the best known attack on AES-256 with 2 related keys and both attacks are better than the previous one from overall complexity. Best related-key impossible differential attack on 8-round AES-192 in [6] has time complexity 2^{136}. So we can see that AES-256 does not have better resistance than AES-192 using the same cryptanalytic approach.

Table II
Summary of the Attacks to 8 Rounds of AES-256

Type	Data	Workload	Keys	Reference
RK Imp. Diff.	2^{53}	2^{215}	2	[7]
RK Imp. Diff.	2^{64}	2^{191}	2	[7]
RK Imp. Diff.	2^{88}	2^{167}	2	[7]
RK Imp. Diff.	2^{112}	2^{143}	2	[7]
Partial Sums	$2^{128} - 2^{119}$	2^{240}	1	[8]
Imp. Diff.	$2^{111.1}$	$2^{227.8}$	1	[9]
Imp. Diff.	$2^{89.1}$	$2^{229.7}$	1	[9]
Meet in the middle	2^{32}	2^{209}	1	[10]
RK Imp. Diff.	$2^{103.5}$	$2^{102.5}$	2	This paper
RK Imp. Diff.	2^{119}	2^{85}	2	This paper

VI. Conclusion

In this paper, we have proposed two new related-key impossible differential attacks against 8-round AES-256 using 64-bit and 32-bit structures. The dominant complexity of these attacks are lower than the previous related-key impossible differential attacks. Another important factor which made our attack more efficient is careful selection of two related keys difference, such that there is no unknown bytes in the subkey differences, which results in lower computational complexity.

References

[1] J. Daemen and V. Rijmen. “The Design of Rijndael:AES-the Advanced Encryption Standard”, Springer Verlag, 2002.

[2] E. Biham and A. Shamir. “Differential cryptanalysis of DES-like cryptosystems”, Journal of Cryptology, 4(1), pp. 3-72, 1991.

[3] E. Biham. “New Types of Cryptanalytic Attacks Using Related Keys”. Journal of Cryptology, 7(4), pp. 229-246, 1994.

[4] G. Jakimoski and Y. Desmedt. “Related-Key Differential Cryptanalysis of 192-bit Key AES Variants”. Selected Areas in Cryptography 2003, LNCS(3006), Springer-Verlag, pp. 208-221, 2004.

[5] E. Biham and N. Keller. “Cryptanalysis of Reduced Variants of Rijndael”. 3rd AES Conference, 2000.

[6] W. Zhang, W. Wu, L. Zhang and D. Feng. “Improved Related-Key Differential Attacks on Reduced-Round AES-192”. Selected Areas in Cryptography 2006, LNCS(4356), Springer-Verlag, pp. 15-20, 2006.

[7] W. Zhang, W. Wu, L. Zhang. “Related-Key Differential Attacks on Reduced-Round AES-256”. https://www.lois.cn/LOIS-AES/data/AES-256.pdf

[8] N. Ferguson, J. Kelsey, B. Schneier, M. Stay, D. Wagner and D. Whiting. “Improved Cryptanalysis of Rijndael”. FSE 2000, LNCS(1978), pp. 213-230, 2001.

[9] J. Lu, O. Dunkelman, N. Keller and J. Kim. “New Impossible Differential Attacks on AES”. INDOCRYPT 2008, LNCS(5365), Springer-Verlag, pp. 279-293, 2008.

[10] H. Demirci and A.A. Selcuk. “A Meet-in-the-Middle Attack on 8-Round AES”. FSE 2008, (LNCS-5806), Springer-Verlag, pp. 116-126, 2008.

[11] NESSIE - New European Schemes for Signatures, Integrity and Encryption, “Performance of Optimized Implementations of the NESSIE Primitives, version 2.0”. https://www.cosic.esat.kuleuven.be/nessie/deliverables/D21-v2.pdf

Bit Error Rate is Convex at High SNR

Sergey Loyka
SITE
University of Ottawa
Ottawa, K1N 6N5, Canada
e-mail: sergey.loyka@ieee.org

Victoria Kostina
Department of Electrical Engineering
Princeton University
Princeton, NJ, 08544, USA
e-mail: vkostina@princeton.edu.

Francois Gagnon
Department of Electrical Engineering
Ecole de Technologie Superieure
Montreal, H3C 1K3, Canada
e-mail: francois.gagnon@etsmtl.ca

Abstract— Motivated by a wide-spread use of convex optimization techniques, convexity properties of bit error rate of the maximum likelihood detector operating in the AWGN channel are studied for arbitrary constellations and bit mappings, which may also include coding under maximum-likelihood decoding. Under this generic setting, the pairwise probability of error and bit error rate are shown to be convex functions of the SNR in the high SNR regime with explicitly-determined boundary. The bit error rate is also shown to be a convex function of the noise power in the low noise/high SNR regime.

I. INTRODUCTION

Optimization problems of various kinds simplify significantly if the goal and constraint functions involved are convex. Indeed, a convex optimization problem has a unique global solution, which can be found either analytically or, with a reasonable effort, by several efficient numerical methods; its numerical complexity grows only moderately with the problem dimensionality and required accuracy; convergence rates and required step size can be estimated in advance; there are powerful analytical tools that can be used to attack a problem and that provide insights into such problems even if solutions, either analytical or numerical, are not found yet [1][2]. In a sense, convex problems are as easy as linear ones. Contrary to this, not only generic nonlinear optimization problems do not possess these features, they are not solvable numerically, i.e. their complexity grows prohibitively fast with problem dimensionality and required accuracy [2]. Thus, there is a great advantage in formulating or at least in approximating an optimization problem as a convex one.

In the world of digital communications, one of the major performance measures is either symbol or bit error rate (SER or BER). Consequently, when an optimization of a communication system is performed, either SER or BER often appears as goal or constraint functions. Examples include optimum power/rate allocation in spatial multiplexing systems (BLAST) [3], optimum power/time sharing for a transmitter and a jammer [4], rate allocation or precoding in multicarrier (OFDM) systems [5], optimum equalization [6], optimum multiuser detection [7], and joint Tx-Rx beamforming (precoding-decoding) in MIMO systems [8]. Symbol and bit error rates of the maximum likelihood (ML) detector have been extensively studied and a large number of exact or approximate analytical results are available for various modulation formats, for both non-fading and fading AWGN channels [9][10]. On the other hand, convexity properties of error rates are not understood so well, especially for constellations of complicated geometry, large dimensionality or when coding is used. Results in this area are scarce. Many known closed-form error rate expressions can be verified by differentiation to be convex, but this approach does not lead anywhere in general. Convexity properties for binary modulations have been studied in-depth in [4], including applications to transmitter and jammer optimizations, and were later extended to arbitrary multidimensional constellations in [11][12] in terms of the SER under maximum-likelihood detection. A log-concavity property of the SER as a function of the SNR [dB] for the uniform square-grid constellations has been established by Conti et al [13].

Unfortunately, convexity of SER does not say anything in general about convexity of the BER, since the latter depends on pairwise probabilities of error (PEP) and not on the SER [14]. Since the BER is an important performance indicator and thus appears as an objective in many optimization problems, we study its convexity in the present paper using a generic geometrical framework developed in [11][12]. Our setting is generic enough so that the results apply to constellations of arbitrary order, shape and dimensionality, which may also include coding under maximum likelihood decoding.

First, we establish convexity properties of the PEP as a function of SNR: it is convex at high SNR regime for any constellation/coding. Its low-SNR behavior depends on constellation dimensionality: it is concave in dimensions 1 and 2 with an odd number of inflection points at intermediate SNR, and it is convex in higher dimensions with an even number of inflection points. Based on this, convexity of the BER at high SNR is established for arbitrary constellation, bit mapping and coding. Thus, this property is a consequence of Gaussian noise density and maximum likelihood detection rather than particular constellation, bit mapping or coding technique. We also show that the BER is a convex function of the noise power in the small noise/high SNR mode.

II. SYSTEM MODEL

The standard baseband discrete-time system model with an AWGN channel, which includes matched filtering and sampling, is

$$\mathbf{r} = \mathbf{s} + \boldsymbol{\xi} \tag{1}$$

where $\mathbf{s}$ and $\mathbf{r}$ are n-dimensional vectors representing the Tx and Rx symbols respectively, $\mathbf{s} \in \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M\}$, a set of M constellation points, $\boldsymbol{\xi}$ is the additive white Gaussian noise (AWGN), $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I})$, whose probability density function (PDF) is

$$p_{\xi}(\mathbf{x}) = \left(2\pi\sigma_0^2\right)^{-n/2} e^{-|\mathbf{x}|^2/2\sigma_0^2} \tag{2}$$

where σ_0^2 is the noise variance per dimension, and n is the constellation dimensionality; lower case bold letters denote vectors, bold capitals denote matrices, x_i denotes i-th component of $\mathbf{x}$, $|\mathbf{x}|$ denotes L$_2$ norm of $\mathbf{x}$, $|\mathbf{x}| = \sqrt{\mathbf{x}^T\mathbf{x}}$, where the superscript T denotes transpose, $\mathbf{x}_i$ denotes i-th vector. The average (over the constellation points) SNR is defined as $\gamma = 1/\sigma_0^2$, which implies the appropriate normalization, $\frac{1}{M}\sum_{i=1}^{M}|\mathbf{s}_i|^2 = 1$.

Consider the maximum likelihood detector, which is equivalent to the minimum distance one in the AWGN channel, $\hat{\mathbf{s}} = \arg\min_{\mathbf{s}_i} |\mathbf{r} - \mathbf{s}_i|$. The probability of symbol error P_{ei} given that $\mathbf{s} = \mathbf{s}_i$ was transmitted is $P_{ei} = \Pr[\hat{\mathbf{s}} \neq \mathbf{s}_i | \mathbf{s} = \mathbf{s}_i] = 1 - P_{ci}$, where P_{ci} is the probability of correct decision. The SER averaged over all constellation points is $P_e = \sum_{i=1}^{M} P_{ei} \Pr[\mathbf{s} = \mathbf{s}_i] = 1 - P_c$. P_{ei} can be expressed as

$$P_{ei} = 1 - \int_{\Omega_i} p_{\xi}(\mathbf{x}) d\mathbf{x} \tag{3}$$

where Ω_i is the decision region (Voronoi region), and $\mathbf{s}_i$ corresponds to $\mathbf{x} = 0$, i.e. the origin is shifted for convenience to the constellation point $\mathbf{s}_i$. Ω_i can be expressed as a convex polyhedron [1],

$$\Omega_i = \{\mathbf{x} | \mathbf{A}\mathbf{x} \le \mathbf{b}\},\ \mathbf{a}_j^T = \frac{(\mathbf{s}_j - \mathbf{s}_i)}{|\mathbf{s}_j - \mathbf{s}_i|},\ b_j = \frac{1}{2}|\mathbf{s}_j - \mathbf{s}_i| \tag{4}$$

where $\mathbf{a}_j^T$ denotes j-th row of $\mathbf{A}$, and the inequality in (4) is applied component-wise. Clearly, P_{ei} and P_{ci} posses the opposite convexity properties.

Another important performance indicator is the pairwise probability of error (PEP) i.e. a probability $\Pr\{\mathbf{s}_i \to \mathbf{s}_j\} = \Pr[\hat{\mathbf{s}} = \mathbf{s}_j | \mathbf{s} = \mathbf{s}_i]$ to decide in favor of $\mathbf{s}_j$ given that $\mathbf{s}_i$, $i \neq j$, was transmitted, which can be expressed as

$$\Pr\{\mathbf{s}_i \to \mathbf{s}_j\} = \int_{\Omega_j} p_{\xi}(\mathbf{x}) d\mathbf{x} \tag{5}$$

where Ω_j is the decision region for $\mathbf{s}_j$ when the reference frame is centered at $\mathbf{s}_i$. The SER can now be expressed as

$$P_{ei} = \sum_{j \neq i} \Pr\{\mathbf{s}_i \to \mathbf{s}_j\} \tag{6}$$

and the BER can be expressed as a positive linear combination of PEPs [14]

$$\text{BER} = \sum_{i=1}^{M} \sum_{j \neq i} \frac{h_{ij}}{\log_2 M} \Pr\{\mathbf{s} = \mathbf{s}_i\} \Pr\{\mathbf{s}_i \to \mathbf{s}_j\} \tag{7}$$

where h_{ij} is the Hamming distance between binary sequences representing $\mathbf{s}_i$ and $\mathbf{s}_j$.

Note that the model and error rate expressions we are using are generic enough to apply to arbitrary constellations, which may also include coding under maximum-likelihood decoding (codewords are considered as points of an extended constellation). We now proceed to establish convexity properties of error rates in this generic setting.

III. Convexity of Symbol Error Rates

Convexity properties of symbol error rates of the ML detector in the SNR and noise power have been established in [11][12] for arbitrary constellation/coding (under ML decoding) and are summarized below for completeness and comparison purpose.

Theorem 1 (Theorem 1 and 2 in [11]): The SER P_e is a convex function of the SNR γ for any constellation/coding (under ML decoding) if $n \le 2$,

$$d^2 P_e / d\gamma^2 = P''_{e|\gamma} > 0 \tag{8}$$

For $n > 2$, the following convexity properties hold:

- P_{ei} is convex in the large SNR mode,
$$\gamma \ge \left(n + \sqrt{2n}\right) / d_{\min,i}^2 \tag{9}$$
where $d_{\min,i}$ is the minimum distance from $\mathbf{s}_i$ to its decision region boundary,
- P_{ei} is concave in the small SNR mode,
$$\gamma \le \left(n - \sqrt{2n}\right) / d_{\max,i}^2 \tag{10}$$
where $d_{\max,i}$ is the maximum distance from $\mathbf{s}_i$ to its decision region boundary,
- there are an odd number of inflection points, $P_{ci|\gamma}'' = P_{ei|\gamma}'' = 0$, in the intermediate SNR mode,
$$\left(n - \sqrt{2n}\right) / d_{\max,i}^2 \le \gamma \le \left(n + \sqrt{2n}\right) / d_{\min,i}^2 \tag{11}$$
- the SER P_e is convex at high SNR,
$$\gamma \ge \left(n + \sqrt{2n}\right) / d_{\min}^2 \tag{12}$$
where $d_{\min} = \min_i \{d_{\min,i}\}$ is the minimum distance to decision region boundary in the constellation.

Theorem 2 (Theorem 4 in [11]): Symbol error rates have the following convexity properties in the noise power $P_N = \sigma_0^2$, for any constellation/coding,

- P_{ei} is concave in the large noise mode,
$$P_N \ge d_{\max,i}^2 \left(n + 2 - \sqrt{2(n+2)}\right)^{-1} \tag{13}$$
- P_{ei} is convex in the small noise mode,
$$P_N \le d_{\min,i}^2 \left(n + 2 + \sqrt{2(n+2)}\right)^{-1} \tag{14}$$
- there are an odd number of inflection points for intermediate noise power,
$$d_{\min,i}^2 \left(n + 2 + \sqrt{2(n+2)}\right)^{-1} \le P_N \le d_{\max,i}^2 \left(n + 2 - \sqrt{2(n+2)}\right)^{-1} \tag{15}$$
- the SER P_e is convex in the small noise/high SNR mode,
$$P_N \le d_{\min}^2 \left(n + 2 + \sqrt{2(n+2)}\right)^{-1} \tag{16}$$

While the convexity properties above are important for many optimization problems, they do not lend any conclusions about convexity of the BER, since the latter is not directly related to P_e or P_{ei} in general. While, in some cases, the BER can be expressed as linear combination of P_{ei}, there are positive and negative terms so that no conclusion about convexity can be made in this case either. On the other hand, the BER can be expressed as a positive linear combination of pairwise probabilities of error so that the convexity of the latter implies the convexity of the former. Thus, we study below the convexity property of the PEP, from which the convexity property of the BER will follow.

IV. Convexity of Pairwise Probability of Error

In many cases, it is a pairwise error probability that is a key point in the analysis (e.g. for constructing a union bound and other performance metrics). Furthermore, it is also a basic building block for the BER in (7), so that we establish its convexity property first.

Theorem 3:

a) The pairwise error probability $\Pr\{\mathbf{s}_i \to \mathbf{s}_j\}$ is a convex function of the SNR at the high SNR region, $\gamma \geq (n+\sqrt{2n})/d_{\min,i}^2$, for any *n*;

b) for $n=1,2$, it is concave at the low SNR region, $\gamma \leq (n+\sqrt{2n})/(d_{ij}+d_{\max,j})^2$, where $d_{ij}=|\mathbf{s}_i-\mathbf{s}_j|$ is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, and there is an odd number of inflection points, $\Pr\{\mathbf{s}_i \to \mathbf{s}_j\}''=0$, in the intermediate SNR mode,

$$(n+\sqrt{2n})/(d_{ij}+d_{\max,j})^2 \leq \gamma \leq (n+\sqrt{2n})/d_{\min,i}^2 \qquad (17)$$

c) for $n>2$, the PEP is convex at the low SNR region, $\gamma \leq (n-\sqrt{2n})/(d_{ij}+d_{\max,j})^2$, and there is an even number of inflection points in-between,

$$(n-\sqrt{2n})/(d_{ij}+d_{\max,j})^2 \leq \gamma \leq (n+\sqrt{2n})/d_{\min,i}^2$$

Proof: See Appendix.

We note that Theorem 3(a) is stronger than Theorem 1 at the high SNR region since the latter follows from the former but the opposite is not always true (as the other SNR ranges in Theorem 3 above indicate). Unlike the SER, the pairwise error probability can be concave at low SNR even for $n=1,2$.

Since Theorem 3 holds for any constellation and bit mapping, it follows that the convexity property of the PEP at high SNR is a consequence of Gaussian noise density rather than particular modulation/coding used, where the latter determines only the SNR threshold.

V. Convexity of The BER at High SNR

We are now in a position to establish the main result of this paper.

Theorem 4: The BER is a convex function of the SNR, for any constellation and bit mapping, which may also include coding under maximum-likelihood decoding, at the high SNR regime,

$$\gamma \geq (n+\sqrt{2n})/d_{\min}^2, \qquad (18)$$

where $d_{\min}=\min_i\{d_{\min,i}\}$ is the minimum distance to the boundary in the constellation.

Proof: Using the relationship between the BER and the pairwise error probabilities in (7) and observing that a positive linear combination of convex functions is convex. Q.E.D.

We remark that the condition in (18) guarantees the convexity of all PEP, BER and SER. In some cases (Gray encoding and when nearest neighbor errors dominate), the BER is approximated as $\mathrm{SER}/\log_2 M$, so that it inherits the same convexity properties as in Theorems 1 and 2 above.

VI. Convexity of the PEP and BER in Noise Power

In a jammer optimization problem, it is convexity properties in noise power that are important [4]. Motivated by this fact, we study below convexity of the PEP and BER in the noise power.

Theorem 5: The PEP $\Pr\{\mathbf{s}_i \to \mathbf{s}_j\}$ is a convex function of the noise power $P_N=\sigma_0^2$, for any *n*, in the small noise/high SNR mode,

$$P_N \leq d_{\min,i}^2\left(n+2+\sqrt{2(n+2)}\right)^{-1} \qquad (19)$$

and in the large noise/low SNR mode,

$$P_N \geq (d_{ij}+d_{\max,j})^2\left(n+2-\sqrt{2(n+2)}\right)^{-1} \qquad (20)$$

Proof: See Appendix.

Based on this Theorem, the following convexity property of the BER is established.

Corollary 5.1: For any constellation and bit mapping, which may also include coding under ML decoding, the BER is a convex function of the noise power in the small noise/high SNR mode:

$$P_N \leq d_{\min}^2\left(n+2+\sqrt{2(n+2)}\right)^{-1} \qquad (21)$$

where specifics of the constellation/code determine only the upper bound in (21).

VII. References

[1] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[2] A. Ben-Tal, A. Nemirovski, Lectrures on Modern Convex Optimization, MPS-SIAM Series on Optimization, Philadelphia, 2001.

[3] V. Kostina, S. Loyka, On Optimum Power Allocation for the V-BLAST, IEEE Transactions on Communications, v. 56, N. 6, pp. 999-1012, June 2008.

[4] M. Azizoglu, Convexity Properties in Binary Detection Problems, IEEE Trans. Inform. Theory, v. 42, N. 4, pp. 1316-1321, July 1996.

[5] Y.-P. Lin, S.-M. Phoong, BER Minimized OFDM Systems With Channel Independent Precoders, IEEE Trans. Signal Processing, v.51, N.9, pp. 2369-2380, Sep. 2003.

[6] C.C. Yeh, J.R. Barry, Adaptive Minimum Bit-Error Rate Equalization for Binary Signaling, IEEE Trans. Communications, v.48, N.7, pp. 1226-1235, Jul. 2000.

[7] X. Wang, W.S. Lu, A. Antoniou, Constrained Minimum-BER Multiuser Detection, IEEE Trans. Signal Processing, v.48, N.10, pp. 2903-2909, Oct. 2000.

[8] D.P. Palomar, J.M. Cioffi, M.A. Lagunas, Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization, IEEE Trans. Signal Processing, v.51, N.9, pp. 2381-2401, Sep. 2003.

[9] J.M. Wozencraft, I.M. Jacobs, Principles of Communication Engineering, Wiley, 1965.

[10] J.R. Barry, E.A. Lee, D.G. Messerschmitt, Digital Coomunications (3rd Ed.), Kluwer, Boston, 2004.

[11] S. Loyka. V. Kostina, F. Gagnon, Symbol Error Rates of Maximum-Likelihood Detector: Convex/Concave Behavior and Applications, IEEE International Symposium on Information Theory (ISIT'07), June 2007, Nice, France.

[12] S. Loyka, V. Kostina, F. Gagnon, Error Rates of the Maximum-Likelihood Detector for Arbitrary Constellations: Convex/Concave Behavior and Applications, IEEE Transactions on Information Theory, accepted, 2009.

[13] A. Conti et al., Log-Concavity Property of the Error Probability with Application to Local Bounds for Wireless Communications, IEEE Trans. Information Theory, June 2009.

[14] J. Lassing et al, Computation of the Exact Bit-Error Rate of Coherent M-ary PSK with gray Code Bit Mapping, IEEE Trans. Communications, v. 51, N. 11, pp. 1758-1760, Nov. 2003.

VIII. Appendix

Proof of Theorem 3: The pairwise probability of error $P_{ij} = \Pr\{\mathbf{s}_i \to \mathbf{s}_j\}$ can be presented as

$$P_{ij} = \int_{\Omega_j} p_\xi(\mathbf{x}) d\mathbf{x} \tag{22}$$

where Ω_j is the decision region for $\mathbf{s}_j$ when the reference frame is centered at $\mathbf{s}_i$. Its second derivative in the SNR is

$$P''_{ij} = \int_{\Omega_j} \frac{d^2 p_\xi(\mathbf{x})}{d\gamma^2} d\mathbf{x} \tag{23}$$

where the derivative is

$$\frac{d^2 p_\xi(\mathbf{x})}{d\gamma^2} = \frac{1}{4}\left(\frac{\gamma}{2\pi}\right)^{n/2} e^{-\gamma|\mathbf{x}|^2/2} f\left(|\mathbf{x}|^2\right) \tag{24}$$

and $f(t) = (t - \alpha_1/\gamma)(t - \alpha_2/\gamma)$, $\alpha_1 = n + \sqrt{2n} > 0$, $\alpha_2 = n - \sqrt{2n} < \alpha_1$. Consider three different cases.

(i) If $d_{\min,i}^2 \geq \alpha_1/\gamma$, where $d_{\min,i} = \min_j(b_j)$ is the minimum distance from the origin to the boundary of Ω_i, then $f(|\mathbf{x}|^2) \geq 0$ $\forall \mathbf{x} \in \Omega_j$ so that the integral in (23) is clearly positive since the integrand is non-negative everywhere in the integration region and positive in some parts of it. Fig. 1 illustrates this case. This is a high SNR mode since $\gamma \geq \alpha_1 / d_{\min,i}^2$.

(ii) If $(d_{ij} + d_{\max,j})^2 \leq \alpha_1/\gamma$ and $n = 1,2$, where $d_{\max,j}$ is the maximum distance from the center of Ω_j to its boundary, then $f(|\mathbf{x}|^2) \leq 0$ $\forall \mathbf{x} \in \Omega_j$ so that the integral in (23) is clearly negative and the result follows. Fig. 2 illustrates this case. This is a low-SNR mode since $\gamma \leq \alpha_1/(d_{ij} + d_{\max,j})^2$. An odd number of inflection points in Theorem 3(b) follows from the continuity argument (P''_{ij} is a continuous function of the SNR).

(iii) Part (c) follows from the same argument as in (ii). *Q.E.D.*

Proof of Theorem 5: follows the same geometric technique as for Theorem 3. 2nd derivative of the PEP in the noise power can be expressed as

$$\frac{d^2 P_{ij}}{dP_N^2} = \int_{\Omega_j} \frac{d^2 p_\xi(\mathbf{x})}{P_N^2} d\mathbf{x} \tag{25}$$

where

$$\begin{aligned}
&\frac{d^2 p_\xi(\mathbf{x})}{dP_N^2} = \frac{1}{4P_N^4}\left(\frac{1}{2\pi P_N}\right)^{\frac{n}{2}} e^{-\frac{|\mathbf{x}|^2}{2P_N}} f^*\left(|\mathbf{x}|^2\right)\\
&f^*(t) = (t - \beta_1 P_N)(t - \beta_2 P_N),\\
&\beta_1 = n + 2 + \sqrt{2(n+2)},\ \ \beta_2 = n + 2 - \sqrt{2(n+2)}
\end{aligned} \tag{26}$$

and $\beta_1 > \beta_2 > 0$. Since $f^*(t)$ has the same structure as $f(t)$ in (24), the proof follows the same steps. In particular, if $d_{\min,i}^2 \geq \beta_1 P_N$, then $d^2 p_\xi / dP_N^2 > 0 \forall \mathbf{x} \in \Omega_j$ so that the integral in (25) is clearly positive. The other case is proved in a similar way. *Q.E.D.*

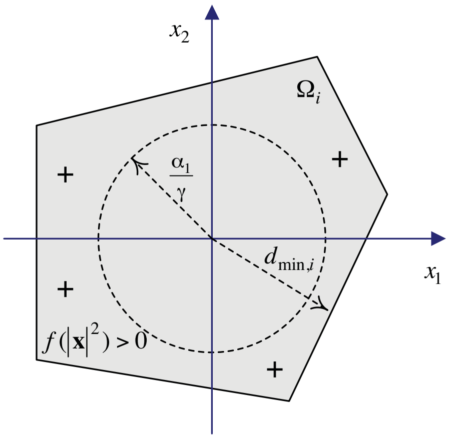


Fig. 1. Two-dimensional illustration of the problem geometry for Case 1. The decision region Ω_i is shaded. $f(|\mathbf{x}|^2)$ has a sign as indicated by "+" and "-".

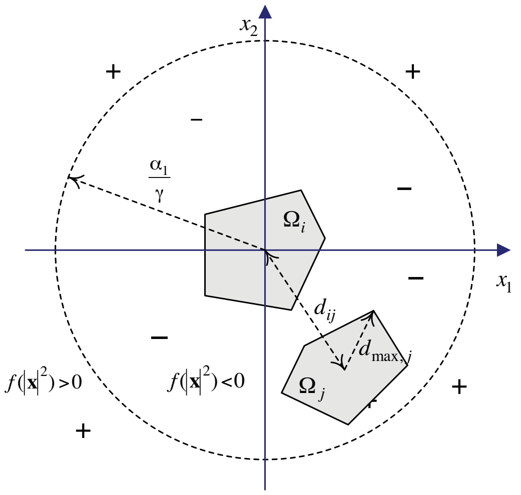


Fig. 2. Two-dimentional illustration of the problem geometry for Case 2.

Multi-Stream Information Transmission over Interference Channels

Dmitri Truhachev
Department of Computing Science, University of Alberta, Canada
email: dmitryt@ualberta.ca

***Abstract*—Communication over Gaussian interference channels is considered where each transmitter modulates its data in the form of multiple redundant data streams and each receiver performs parallel joint detection of the data streams followed by the individual error control decoding. A sufficient condition for decoding success in terms of data stream density is given for general interference channels. It is also demonstrated that the presented technique allows to achieve or closely approach generalized degrees of freedom for two-user symmetric interference channel.**

I. Introduction

Interference networks in general and classic interference channels in particular have drawn significant attention in the last decade due to the increasing popularity of wireless networks where interference resolution has started to play a major role. The results derived in the 70s and 80s were recently complemented by a number of new contributions [1], [2], [3], [4], [5], [6], [7], which stem from several novel approaches, such as interference alignment and coding utilizing number theory and additive combinatorics. While two-user interference and X-channels have been characterized in terms of capacity and generalized degrees of freedom, results for general K-user interference networks are still largely unknown.

While in varying channels or multiple input multiple output channels, interference can be efficiently combated by smart precoding [6], [7], this is not the case for static channels, which, besides some recent developments in [9], [10], [3], mostly remain untouched. It has been observed long ago that successive interference cancellation is sufficient [8] to achieve capacity in strong or weak interference regimes. Following this way of thinking, most of the approaches to communication over interference channels resort to splitting of the messages into parts, which are either jointly decoded or decoded successively by treating weaker messages as noise. For many user interference channels, this methodology results in the very complex organization of transmission since multiple message splittings are required to guarantee a desirable rate constellation at every receiver. In addition to the sophistication of this type of transmission and reception, the fragile arrangement of cancellation steps can be destroyed by realistic channel effects.

In this paper, we focus on a different approach based on transmitting information in the form of concurrent redundant data streams. A message in this case is composed of a multitude of individually coded data streams distinguished by random signature sequences [11], [12]. Parallel interference cancellation is used for joint detection of the streams at the receiver and is followed by individual error control decoding. This method of data modulation has already proven successful in multiuser coding for communication over multiple access channels. Particularly, it has been shown that multiple access channel capacity can be approached within one bit per dimension for a specific power distribution [11]. Furthermore, often, the power distribution does not need to be shaped precisely. Rather, maintaining the average number of data streams per power level below some critical threshold is sufficient [13].

We propose multiple redundant data stream transmission over interference channels. Each transmitter selects the number of data streams and their power levels according to the total transmit power, possibly taking into account the envisioned power levels of the interference data streams. Each receiver performs data stream separation followed by error correction. We formulate a sufficient condition for the decoding success in terms of the maximum number of received data streams per power level observed at each receiver. We apply the technique to two-user symmetric interference channels and reproduce the generalized degrees of freedom results [2] for almost all channel parameters. The important aspect of the presented method is the parallel joint detection, which is less sensitive to real world channel effects and has the potential for implementation friendly architectures.

II. System Model

A. Signalling Strategy

We start with an overview of the transmission format that was also described in [11], [12]. Assume that the signals are encoded using N signalling dimensions, which can, for example, be time or frequency slots, or both. A message X is composed of an arbitrary number, say J, of independent data streams. Each data stream $j = 1, 2, \ldots, J$ is individually error control encoded, binary phase shift keying (BPSK) modulated, and then modulated using an N-dimensional signature vector $\mathbf{s}_j$, which is power normalized to unity, i.e., $||\mathbf{s}_j|| = 1$. The signature vectors $\mathbf{s}_j$ are chosen randomly, independently of each other, so $\mathbf{E}(\mathbf{s}_i\mathbf{s}_j)^2 = 1/N$. Once data bits $\{u_n^{(j)}\}_{n=1}^{L}$ of the stream j are encoded by an error control encoder to produce $\{v_n^{(j)}\}_{n=1}^{L_1}$, each of the bits $v_n^{(j)}$ is multiplied by an N-dimensional vector $\mathbf{s}_j$ and then partitioned into M subsections

$$v_n^{(j)}\mathbf{s}_{j1}, v_n^{(j)}\mathbf{s}_{j2}, \ldots, v_n^{(j)}\mathbf{s}_{jM} \ , \qquad (1)$$

where $\mathbf{s}_j = (\mathbf{s}_{j1}, \mathbf{s}_{j2}, \ldots, \mathbf{s}_{jM})$. These subsections (1) are obtained for every bit $v_n^{(j)}$ and then permuted over the entire block of $L_1 M$ subsections using a permutor π_j specific to the stream j. Finally, each stream is given power P_j, and they are simultaneously transmitted over the channel. The total power is, therefore, $P = \sum_{j=1}^{J} P_j$, and the total information rate

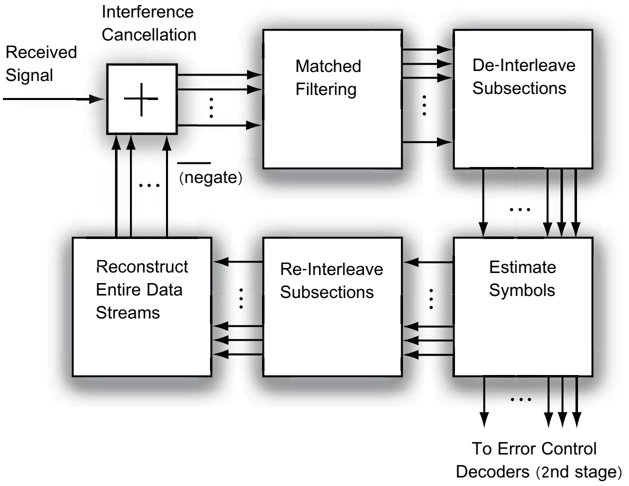


Fig. 1. Receiver architecture, first stage.

is $R = \sum_{j=1}^{J} R_j$, where individual rates $R_j \in [0, 1]$ (due to BPSK modulation per stream). First, we consider transmission over an additive white Gaussian noise (AWGN) (real-valued) channel and describe the receiver processing below.

B. Receiver Architecture and Decoding Dynamics

The receiver operates as the two-stage decoder described in [11]. The first stage is an iterative joint detector that functions in the following way (see Fig. 1). The received signal passes through a bank of matched filters ($\mathbf{s}_{jm}^*$) to distinguish parts of the transmitted message modulated by signature waveform subsections $\mathbf{s}_{jm}$. These received subsections are used to produce soft estimates $\hat{v}_n^{(j)}$ of the transmitted bits $v_n^{(j)}$. The soft bit estimates are then used to approximately reconstruct the transmitted signals and subtract the effect of interference (interference cancellation). This process is repeated a number of times and works similar to belief propagation decoding of low-density parity-check codes. Finally, the resulting data streams are passed to the individual error control decoders. The second stage of the decoding is the error correction performed for all data streams in parallel.

To study the decoding dynamics, we consider large values of J and approximate powers P_j, $j = 1, 2, \ldots, J$ by a continuous function $P(x)$, $x \in [0, J]$ where $P(x) = P_j$, for $x = j$. Without loss of generality, we can assume that $P(x)$ is a nondecreasing function of x. Normalizing by N we define $T(u) = P(uN)$, where $u \in [0, J/N]$. Finally, denoting $\beta = J/N$ and assuming a large number of subsections M, we can obtain the following equation (see [11]) for the residual noise and interference variance per data bit (here, σ^2 is the AWGN noise variance) at each detection iteration i

$$\sigma_{i+1}^2 = \int_0^{\beta} T(u) g\left(\frac{T(u)}{\sigma_i^2}\right) du + \sigma^2; \quad i = 1, 2, \ldots, \qquad (2)$$

$$g(s) = \mathbf{E}\left[\left(1 - \tanh\left(s + \xi\sqrt{s}\right)\right)^2\right] \;, \quad \xi \sim \mathcal{N}(0, 1) \;. \qquad (3)$$

The signal-to-noise ratio (SNR) of the data stream u after I iterations equals $T(u)/\sigma_I^2$. If $R(u)$ is the information rate of data stream u, then $R(u) \le C_{\text{BIAWGN}}(T(u)/\sigma_I^2)$ should be satisfied for error free decoding of the second stage. Here, by $C_{\text{BIAWGN}}(s)$, we denote the capacity of the binary input AWGN channel with SNR equal to s.

For power distribution

$$T(u) = \sigma^2 e^{u 2 \ln 2 + 1} \;, \qquad (4)$$

the total system's spectral efficiency per dimension C_{eff} is within one bit from the AWGN channel capacity (see the proof in [11])

$$C_{\text{AWGN}} - 1 \le C_{\text{eff}} = \int_0^{\beta} R(u) du \;, \qquad (5)$$

where $R(u)$, the information rate of stream u, is derived assuming the use of BPSK capacity achieving error control codes, i.e., $R(u) = C_{\text{BIAWGN}}(T(u)/\sigma_\infty^2)$.

Consider now arbitrary positive nondecreasing $T(u)$, let $F(u) = \ln T(u)$, and define function

$$f(x) = \left.\frac{dF^{-1}(t)}{dt}\right|_{t=x}$$

for $x \in [F(0), F(\beta)]$. We can think of $f(x)$ as density of data streams per power level. If $f(x)$ has support $[p_0, p_1]$, the total message power P and the total information rate C_{eff} (per dimension) are expressed as

$$P = \int_{p_0}^{p_1} e^x f(x) dx \quad \text{and} \quad C_{\text{eff}} = \int_{p_0}^{p_1} f(x) dx \;.$$

According to the definitions above, AWGN channel capacity can be achieved for constant density

$$f(x) = \frac{1}{2\ln 2}, \qquad \text{for} \qquad x \in [1 + \ln\sigma^2, 2\beta\ln 2 + 1 + \ln\sigma^2] \;,$$

(see (4)). In this case, the variance σ_i^2 converges as $i \to \infty$ to $\sigma_\infty^2 < 2\sigma^2$. We will say that convergence for given $f(x)$ and σ^2 happens if $\sigma_\infty^2 \le 2\sigma^2$.

Theorem 1: If $f(x) \le \frac{1}{2\ln 2}$ for $x \ge 1 + \ln\sigma^2$, and $f(x) = 0$ for $x \in [\ln\sigma^2, 1 + \ln\sigma^2)$ then convergence is guaranteed.

Proof: See Appendix. ■

We notice, however, that decoding convergence resulting from arbitrary $f(x)$ does not always guarantee the achievement of capacity.

C. Transmission over Interference Channels

Consider now a K-user interference channel with real channel coefficients h_{ij}, $i, j \in \{1, 2, \ldots, K\}$. Transmitter i needs to transmit its own message to its intended Receiver i as is customary. Additive white Gaussian noise with variance σ^2 is added independently at each receiver. Assume that transmitter i uses density function $f_i(x)$ to encode J_i data streams, $i = 1, 2, \ldots, K$. The following corollary from Theorem 1 is a sufficient condition for successful decoding

Corollary 2: If $y(x) = \sum_{i=1}^{K} f_i(x + 2\ln h_{ij}) \le \frac{1}{2\ln 2}$ for any x and any j, then convergence is guaranteed, and (per user per dimension) spectral efficiency of $\sum_{i=1}^{K} \frac{J_i}{NK}$ is achieved.

Fig. 2 illustrates the corollary. Four messages with densities $f_1(x), f_2(x), f_3(x)$ and $f_4(x)$ produce density $y(x)$ at a receiver. The density falls below critical level indicated by the dashed line. Interfering data streams which fall below the noise floor are suppressed and slightly increase the noise. Notice that the condition above is a sufficient condition, since it assumes the decoding of every data stream which arrives at the receiver with power sufficient to stick out from the noise floor. While it is sometimes an efficient and robust strategy, it is clearly suboptimal in many cases, since the receiver does not benefit from decoding interferers individually. Several works have demonstrated that decoding interference as a whole is a better strategy. For example, [10] presents a theoretical technique for the very special case of channel gains where coding employing number theory is used and the sum of the interferers is decoded. The interference cancellation is performed in [10] by decoding least significant and most significant bits of the signals in turn, i.e., from two directions (up and down).

Here, the possibility of decoding and cancelling from both sides, i.e., from the strongest power down and from weakest power up, is not available to us. However, we can show that the same effect can be accomplished using joint decoding only.

D. Stream Alignment

To decode the sum of two interferers as a whole for some interval $[a, b]$ of the power density, consider encoding two data steams (belonging to separate transmitters) with the same power level $x \in [a, b]$, same signatures, and same interleavers. Two data streams encoded this way "glue" (align) together in the channel so that instead of BPSK constellations $\{-1, 1\}$, 3-level constellations $\{-2, 0, 2\}$ arise. Each 3-level constellation can be decoded as a whole in the first stage of the decoding process. It does not need to be decoded in the second stage because it belongs to interference. Assume that $f_2(x)$ is the composite density of BPSK constellations and $f_3(x)$ is the density of the 3-level constellations observed at the receiver.

Theorem 3: If for all x

$$2 \ln 2 f_2(x) + 2.07 f_3(x) \leq 1$$

the decoding convergence is guaranteed.

Proof: See Appendix. ■

E. Stream Duplication

At times repeating (duplicating) some data streams at several power levels is a good idea. If the data at higher power level is not decoded, since it was aligned with other streams to produce 3-level constellations, it can be decoded at lower power levels. However, if (in the other subchannel) data is decoded at higher levels, this does not pose additional obstacle for convergence, since the knowledge from higher power level decoding propagates to lower power levels and eliminates the corresponding variance terms.

Theorem 4: If one stream of power P_1 is duplicated at lower power P_2, it is equivalent to an increase of the density at the higher power level by $\epsilon = P_2/P_1$.

Proof: The proof is omitted due to space limitations. ■

Armed with the above theorems, we will now construct achievability schemes for two-user interference channels.

III. Results for Two-User Symmetric Interference Channel

Consider a two-user symmetric Gaussian interference channel with real coefficients, i.e., we assume that $h_{11} = h_{22} = 1$ and $h_{12} = h_{21} = a$, where a is a parameter. Consider a symmetric rate point, i.e., both users transmit with the same power P. We will present coding in terms of the density $f(x)$ utilized by the transmitters. The support of the function $f(x)$ is $[\delta, \Delta]$, which is determined by the total transmit power P. The smallest power e^δ is determined by the noise variance σ^2 (δ is roughly equal to $\ln(\sigma^2) + 1$), such that the lowest data stream gets a rate close to 1. Let us define $\bar{a} = |2 \ln a|$ and consider generalized degrees of freedom (GDOF) as defined in [2]. Consider the interference level

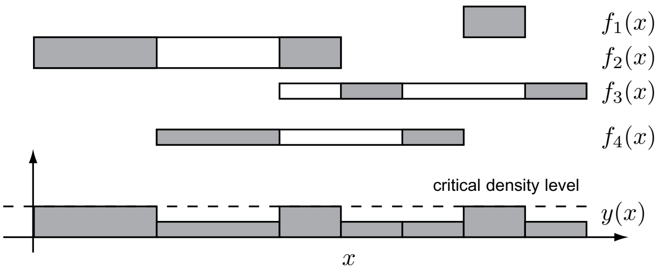


Fig. 2. An example of stream density which could be observed at a receiver in 4-user interference channel.

$$\alpha = \frac{\ln \mathrm{INR}}{\ln \mathrm{SNR}} = \frac{2 \ln a + \ln P - 2 \ln \sigma}{\ln P - 2 \ln \sigma} .$$

and our acheivable spectral efficiency relative to capacity of the AWGN channel with same power (σ^2 is kept constant)

$$\bar{d}(\alpha) = \lim_{P \to \infty, \frac{\ln \mathrm{INR}}{\ln \mathrm{SNR}} = \alpha} \frac{C_{\mathrm{eff}}}{C_{\mathrm{AWGN}}} .$$

We will compare $\bar{d}(\alpha)$ to optimal GDOF $d(\alpha)$ derived in [2].

A. Case $\alpha \in [0, 1/2]$

As previously observed, the weak interference case is relatively simple. Let $f(x) = \frac{1}{2 \ln 2} \mathbf{1}_{[\Delta - \bar{a}, \Delta]}(x)$ where $\mathbf{1}_{[a,b]}(\cdot)$ denotes the indicator function of the interval $[a, b]$. Each of the receivers observes a density

$$y(x) = f(x) + f(x - \bar{a}) = \frac{1}{2 \ln 2} \mathbf{1}_{[\Delta - 2\bar{a}, \Delta]}(x) \leq \frac{1}{2 \ln 2}$$

and convergence is guaranteed. The power of one message is $P = e^\Delta - e^{\Delta - \bar{a}}$ and the achievable GDOF is

$$\bar{d}(\alpha) = \frac{\bar{a}}{\Delta} = 1 - \alpha = d(\alpha) .$$

B. Case $\alpha \in [2, \infty]$

The case of very strong interference is also straightforward with $f(x) = \frac{1}{2 \ln 2} \mathbf{1}_{[0, \Delta]}(x)$,

$$\alpha = \frac{\Delta + \bar{a}}{\Delta} \quad \text{and} \quad \bar{d}(\alpha) = \frac{\Delta}{\Delta} = 1 = d(\alpha) .$$

C. Case $\alpha \in [1/2, 2/3]$

In this case, $f(x) = \frac{1}{2 \ln 2} \mathbf{1}_{[\Delta - 2\bar{a}, \Delta]}(x) + \frac{1}{2 \ln 2} \mathbf{1}_{[\delta, \bar{a}]}(x)$, and

$$\alpha = \frac{\Delta - \bar{a}}{\Delta} \quad \text{and} \quad \bar{d}(\alpha) = \frac{\Delta - \bar{a}}{\Delta} = \alpha = d(\alpha) .$$

D. Case $\alpha \in [3/2, 2]$

In this case the densities are given by

$$f(x) = \frac{1}{2 \ln 2} (\mathbf{1}_{[\delta, \Delta - \bar{a}]}(x) + \mathbf{1}_{[\Delta - \bar{a}, \bar{a}]}(x)) + \frac{3}{8 \ln 2} \mathbf{1}_{[\bar{a}, \Delta]}(x) ,$$

however, some streams need to be aligned and duplicated.

The situation is graphically depicted in Fig. 3 (with shift corresponding to Receiver 1). Block A_1 is a duplicate of A_2, and B_1 a duplicate of B_2, Block C is aligned with B_2 and block F is aligned with A_2. Blocks A_1, A_2, B_1, B_2, C and F have density $1/(4\ln 2)$ and E and D density $1/(8\ln 2)$. During the iterative detection process at the Receiver 1 B_2 will be detected first and, therefore, B_1 will disappear (Theorem 4). Moreover, A_2 will align with F, but its information will be decoded later when the process comes to A_1. Similar situation happens at Receiver 2. The densities of E and D are calculated according to Theorem 3 since alignment produces 3-level constellations. The resulting achievable GDOF is

$$\bar{d}(\alpha) = 0.75\alpha - 0.5 < \frac{\alpha}{2} = d(\alpha)$$

E. Case $\alpha \in [2/3, 3/2]$

Finding optimum stream density is a difficult task in this case and we use the approaches of Cases $\alpha \in [1/2, 2/3]$ and $\alpha \in [3/2, 2]$ to construct achievability schemes. Fig. 4 illustrates the results for $\bar{d}(\alpha)$ for all cases (blue curve) and compares them to known optimum GDOF (red curve).

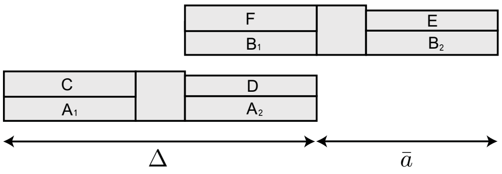


Fig. 3. Stream densities at Receiver 1 for the case $\alpha \in [3/2, 2]$.

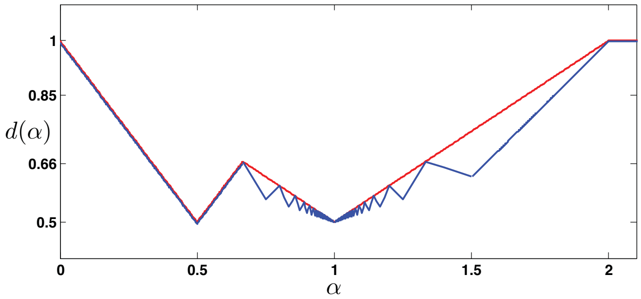


Fig. 4. Optimal GDOF (red curve) vs. GDOF achievable with Multi-stream signalling (blue curve).

IV. Conclusions

We discussed an application of multi-stream information transmission to interference channels. The reception involves parallel joint detection of the data streams followed by the individual error control decoding. The performance of the technique can be analyzed through the density of the data streams relative to the power at the transmitters and receivers. Generalized degrees of freedom can be achieved in several cases for two-user interference channels. It is also demonstrated that decoding sum of interferers as a whole as well as successive cancellation type decoding can be accomplished with parallel joint detection.

References

[1] G. Kramer, "Outer Bounds on the Capacity of Gaussian Interference Channels", *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 581 - 586, March 2004.

[2] R. Etkin, D. Tse, and Wang "Gaussian Interference Channel Capacity to Within One Bit", *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, December 2008.

[3] R. Etkin and E. Ordentlich "On the Degrees-of-Freedom of the K-User Gaussian Interference Channel", available in *ArXiv*.

[4] M. Maddah-Li, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3457-3470, August 2008.

[5] C. Huang, V. R. Cadambe, S. A. Jafar "Interference Alignment and the Generalized Degrees of Freedom of the X Channel", available in *ArXiv*.

[6] V. R. Cadambe and S. A. Jafar "Interference Alignment and Spatial Degrees of Freedom for the K User Interference Channel", *arXiv:0707.0323*.

[7] V. R. Cadambe, S. A. Jafar, and S. Shamai, "Interference Alignment on the Deterministic Channel and Application to Fully Connected Gaussian Interference Networks", *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 269–274., January 2009.

[8] A. B. Carleial, "A Case Where Interference Does Not Reduce Capacity", *IEEE Trans. on Inform. Theory*, vol. 21, pp. 569–570, Sep. 1975.

[9] A. Host-Madsen and A. Nosratinia, "The multiplexing gain of wireless networks," in Proceedings of *International Symposium on Information Theory*, Adelaide, July 2005.

[10] S. A. Jafar and S. Vishwanath, "Generalized Degrees of Freedom of the Symmetric Gaussian K-User Interferenvce Channels," available in *ArXiv*.

[11] D. Truhachev, C. Schlegel and L. Krzymien, "A Two-Stage Capacity-Achieving Demodulation/Decoding Method for Random Matrix Channels", *IEEE Tran. on Inform. Theory*, vol. 55, pp. 136–146, Jan. 2009.

[12] C. Schlegel, "CDMA with Partitioned Spreading", *IEEE Communications Letters*, vol. 11, no. 12, pp. 913–915, December 2007.

[13] D. Truhachev, S. Nagaraj, and C. Schlegel, "Throughput-Reliability Tradeoffs in Spread Spectrum Multi-Hop Ad Hoc Wireless Networks with Multi-Packet Detection," *IEEE International Conf. on Communications*, Dresden, Germany, June 2009.

Appendix

Since the full proofs cannot be included due to space limitations, we will only mention the most important steps.

Proof of Theorem 1: The convergence condition [11] is

$$\Phi(v) < 0 \quad \text{for all} \quad 2\sigma^2 \le v \le \int_0^\beta T(u)du + \sigma^2 \qquad (6)$$

$$\text{where } \Phi(v) = \int_0^\beta \frac{T(u)}{v} g\left(\frac{T(u)}{v}\right) du + \frac{\sigma^2}{v} - 1 \ . \qquad (7)$$

For $T(u) = \sigma^2 e^{u2\ln 2+1}$, an even stronger condition

$$-\frac{1}{2\ln 2}\int_0^{1/v} g(x)dx + \frac{\sigma^2}{v} < 0 \quad \text{for all} \quad v \in [2\sigma^2, \infty) \quad (8)$$

is satisfied and implies (6). Now consider arbitrary $T(u)$ defined by density $f(x)$ satisfying the condition of Theorem 1. Without loss of generality, assume its support to be $[0, \beta]$. By definition, $f(x) = \frac{1}{F'(u)}$ where $x = F(u)$. This implies $T(u) = \exp(\int_0^u \frac{1}{f(F(t))}dt)$, which we substitute into (7) together with a variable exchange $y = T(u)/v$. We obtain

$$\Phi(v) = \int_{1/v}^{T(\beta)/v} f(F(u(y)))g\left(y\right)dy + \frac{\sigma^2}{v} - 1 \ . \qquad (9)$$

Finally, since $f(\cdot) \le 1/(2\ln 2)$, the condition (8) follows.

Proof of Theorem 3:
The convergence condition in this case is

$$\Phi(v) = \int_{1/v}^{T(\beta)/v} f_2(F(u(y)))g\left(y\right)dy + \int_{1/v}^{T(\beta)/v} f_3(F(u(y)))g_3\left(y\right)dy + \frac{\sigma^2}{v} - 1 \ < 0. \qquad (10)$$

where $g_3(\cdot)$ is the average squared error for the 3-level constellation. Since $\int_0^\infty g_3(x)dx \approx 2.07$, (10) implies the condition of Theorem 2, which is sufficient for convergence.

Polynomial-Time Resource Allocation in Large Multiflow Wireless Networks with Cooperative Links

IZS: Invited Paper

Gareth Middleton and Behnaam Aazhang
Rice University
Houston, Texas
Email: {gbmidd, aaz}@rice.edu

***Abstract*—An integrated multiflow network model synthesizing physical layer rate control and link layer access control is presented, permitting the study of resource allocation in large networks while allowing multi-terminal cooperation at the physical layer. We discuss how to incorporate the cooperative unit into the broader scheduling and routing framework as a "metalink," a general notion capable of representing a variety of multiterminal physical layer topologies. Simulation results show the benefits of employing cooperative structures where needed.**

I. Introduction

Increasing throughput demands have recently begun forcing synergies between previously isolated areas of the communications stack, requiring that advanced physical layer techniques–such as signal-scale cooperation–be viewed in the context of link layer radio resource management. However, the complexity associated with the merging of models results in a system unsuitable for any form of analysis. In the past, these complexities have been managed in the physical layer by ignoring all but the smallest of topologies, and in the network layer by applying an abstract view of information flows between terminals. These simplifications have enabled a large body of work to develop in the two communities, though to date complexity issues have presented a formidable barrier to the joint study of physical-layer cooperation and link-layer scheduling and routing, particularly in the context of multiflow networks.

In this paper, we will formulate an integrated model, within which we will study the scheduling and routing problem while permitting information-theoretic *rate control* on links between terminals. As we will show, this enables studying cooperative technology in the broader resource-allocation paradigm of large networks. We will discuss how this can be accomplished within our framework, and we will present results showing how cooperative resource allocation in large networks differs from the allocation in the context of conventional point-to-point links.

II. System Model and Problem Statement

We consider an ad-hoc half-duplex TDMA network in which all terminals share the available bandwidth. Several source-destination pairs of terminals are chosen, each demanding maximum throughput. All other terminals may participate in the network, forwarding data in a multihop manner or in a cooperative mode, to be discussed in more detail below. We assume that all terminals have a maximum transmission power P_{Tx}, and that there is no average power constraint.

Most network-layer studies consider data to be composed of packets, which are of a predetermined size. Success of transmission is then binary: a sufficient signal-to-interference-and-noise (SINR) ratio or good proximity to the receiver guarantees that the entire packet is received. In this work, we suspend that assumption, introducing the *rate control* element to our model by describing the size of a packet using Shannon's upper-bounding equation [1]

$$R = \log_2(1 + \mathrm{SINR}) \tag{1}$$

for point-to-point links. For clarity of exposition but without loss of generality, we consider timeslots of 1 second and a bandwith of 1 Hz, allowing us to view Shannon's rate as the size in bits of a packet transmitted on a point-to-point link.

III. Approach and Example

Here we discuss our approach for allocating resources in multiflow networks with rate control, and show how this framework is amenable to the incorporation of cooperative technologies.

A. NFIC

As presented in [2], [3], we have developed a framework for performing resource allocation in large networks with multiple flows in $O(N^3)$ time. Due to space considerations, we refer the reader to our references for the details. Fundamentally, the approach hinges on a novel data structure called the *Network-Flow Interaction Chart*, which specifies the detailed interactions of data and terminals at all time instances in the network. The terminals are represented as nodes in the chart, replicated in the x direction to represent the time-slotted communication. Edges are drawn between nodes to represent transmission between two terminals at a specific timeslot. This data structure is well-suited to dynamic programming techniques, which enable us to schedule and route data for maximum throughput on a per-packet basis in polynomial time. Edges are assigned weights corresponding to the rate

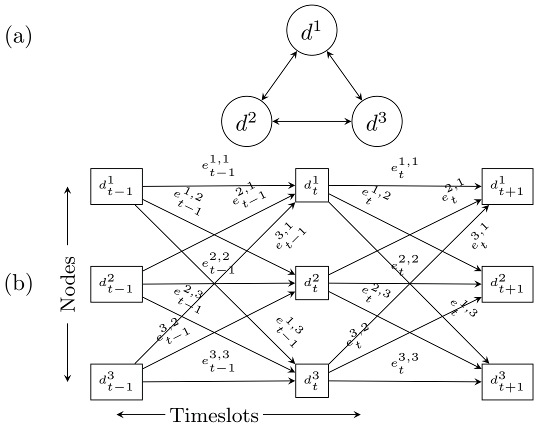


Fig. 1. An example topology (a) and the NFIC (b). Note that all nodes interfere with all others, and so the NFIC is fully connected.

achievable between the anchoring terminals at time t, and these weights are propagated through the chart by the dynamic programming algorithms. Edges are then “zeroed out” after a packet has been allocated to make them unavailable, thereby enforcing duplexing constraints. Similarly, terminals not transmitting or receiving an allocated packet have their edge values scaled down to prevent interference from affecting the existing packets. A network and corresponding NFIC is shown in Figure 1.

In order to incorporate cooperative links into the framework, we must represent the notion of data flowing to *two destinations* in one timeslot (the cooperative BC phase), and *constructively combining* in a subsequent timeslot (the cooperative MA phase) [4]. To accomplish this, we propose adding a new type of node to the NFIC framework, the metanode.

B. Metanodes

Nodes in the NFIC have, until now, represented exactly one terminal in the network. As a new class of nodes, the metanodes represent a coordinated combination of terminals, in this case the cooperative unit of a particular source, relay and destination. An edge *entering* a metanode corresponds in the schedule to data being transmitted from the source terminal to the intended destination and also the associated relay. An edge *exiting* a metanode corresponds in the schedule to the coordinated transmission of both the relay and source, where signals constructively combine at the destination. The weight on *both* the entering and departing metanode edges is set as the overall rate the cooperative unit can achieve, which (although not known explicitly) can be bounded in the time-shared relay case as [5], [6], [7]

$$\min\left\{\begin{array}{l}\log_2\left(1+P_s^{(1)}h_{sr}^{(1)}\right)+\log_2\left(1+P_s^{(2)}h_{sd}^{(2)}\right)\\ \log_2\left(1+P_s^{(1)}h_{sd}^{(1)}\right)+\log_2\left(1+P_s^{(2)}h_{sd}^{(2)}+P_r^{(2)}h_{rd}^{(2)}\right)\end{array}\right. \tag{2}$$

where P_s and P_r are transmission powers at the source and relay respectively, with channels h between all terminals being similarly labeled. We require that cooperation explicitly aid the transmission, i.e. that the channel is degraded: $\min\{|h_{sr}|^2, |h_{rd}|^2\} \geq |h_{sd}|^2$. Note that we assume zero correlation between the relay and source transmissions, and that the relay channel is in the BC and MA phases for equal amounts of time.

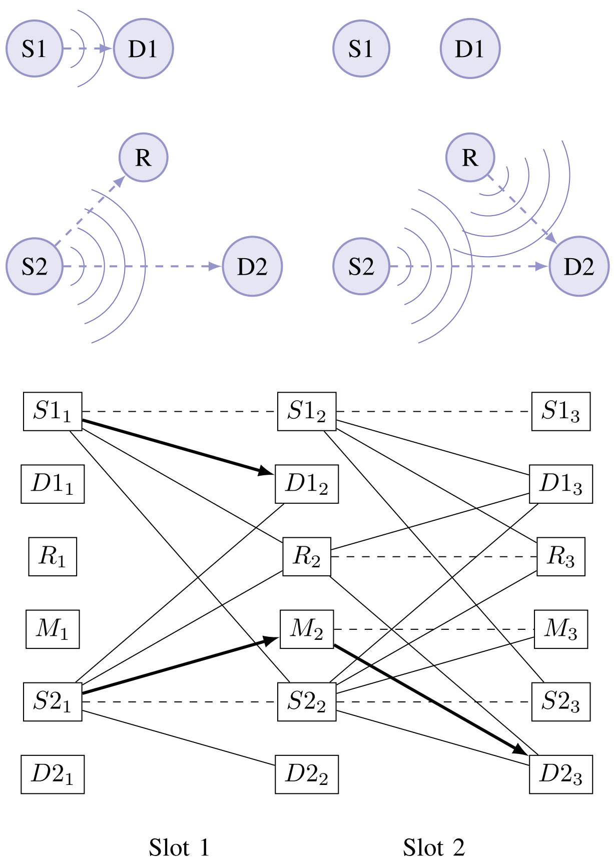


Fig. 2. Top: Network activity in timeslots 1 and 2. Both sources transmit in the first slot, their intended destinations shown by the broken lines. In timeslot 2, $S2$ and R transmit together in the MA stage of cooperation. Bottom: the NFIC corresponding to this network and its activity. Note that how the metanode M describes the cooperative behavior shown above.

C. Example: Metanodes in the NFIC

To illuminate the concept of a metanode in the NFIC, consider the small network of five nodes, shown in the upper pane of Figure 2. The channels are dominated by pathloss.

Here, two sources $S1$ and $S2$ are attempting to communicate with two matching destinations $D1$ and $D2$. The NFIC corresponding to this network is shown in the lower pane of Figure 2, where we have added the metanode M to capture the notion of cooperation occurring between $S2$, R, and $D2$. In particular, the edge between $S2_1$ and M_2 corresponds to the second source transmitting, but with that transmission being received by *both* the relay node and the intended destination. The edge from M_2 to $D2_3$ then corresponds to *both* the relay and source transmitting in the second timeslot.

The NFIC routing and corresponding network activity is shown in Figure 2. Here we preserve clarity by not drawing all edges in the NFIC, rather including only those which are relevant as data emanate from the sources. Note that in the second timeslot, although only *one* edge in the NFIC is active, *two* transmissions are occurring in the network!

In this way, we are able to use the *same* polynomial-time routing and scheduling technologies as described in [2], [3] in networks where cooperation is used.

D. Choice of Metanodes: Memory Complexity

The choice of best relay terminal for a given packet route is known to be NP-hard in the general fading enviroment [**?**], which when considered jointly with the NP-hard routing and scheduling problem, results a doubly complex selection problem. In our case, choosing the metanode, the source-relay-destination triple, can be simplified with the use of geographical information. The pathloss-dominant fading environment allows us to consider only nearby terminals as potential relays. Thus, for each terminal which may act as a source, we may locally select relays and corresponding destinations. These form our metanodes, which may be incorporated in the NFIC resource allocation decisions.

Our algorithms are $O(N^3)$ in the number of nodes in the NFIC, which means that adding metanodes to each terminal does not increase the order of the complexity. However, it does increase memory requirements, as new edges and terminals are introduced to the NFIC with weights which must be stored. Table I shows memory data for the NFIC with a varying number of metanodes defined *per* terminal. Memory requirements increase considerably over the multi-hop NFIC, though since all but two edges into and out of metanodes are zero, the sparsity of the resulting NFIC is considerable. This means that even for large numbers of defined metanodes, memory complexity does not overwhelm the algorithm.

TABLE I
PERCENTAGE INCREASE IN MEMORY REQUIREMENTS OVER MULTIHOP NFIC AND SPARSITY OF NFIC REPRESENTATION

Metanodes per Terminal	Raw Storage Increase (%)	Sparsity (%)
1	300	75
3	1500	93.75
5	3500	97.22
7	6300	98.44

IV. BROADER NETWORK ALLOCATION

The use of metanodes in the resource allocation algorithm as illustrated above can be applied to much larger networks, where the polynomial-time nature of our solutions permits allocations in networks of near-arbitrary size.

A. Example

We choose to study networks in which pathloss is the dominant channel effect, but this assumption is not required for our techniques to apply. We make this choice since it helps to illustrate where cooperative gains exist in the network. An example of resource allocation with cooperation is shown in Figure 3. Here, two flows compete for network resources, and both are able to leverage cooperative links. Since the network is dominated by pathloss, cooperation will assist the overall throughput of the flows only on the longest hop in the route, which is the throughput bottleneck.

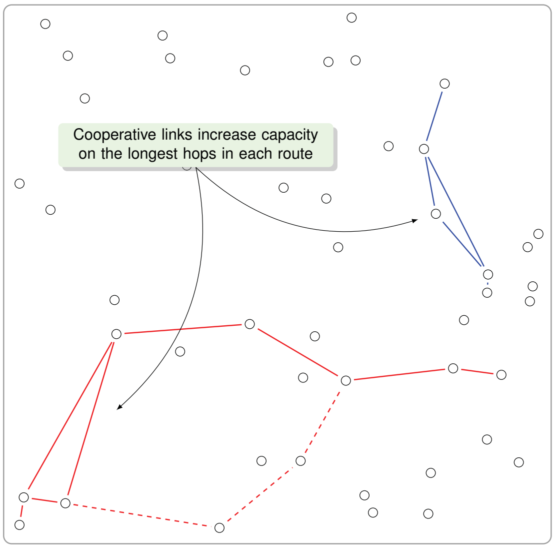

Cooperative links increase capacity
on the longest hops in each route

Fig. 3. Two-flow network in which NFIC resource allocation was used with metanodes available. In the allocation, metanodes corresponding to the cooperative terminals shown above were found optimal by the dynamic programming routines, and in the case of the red flow, use of the cooperative metanode completely changed the routing decision for the flow.

This is clearly illustrated in the example network, where the routing decision for the red flow *changed* as a result of the cooperative metanode being available in the NFIC. Had cooperation not been available, the red flow would have followed the sequence of terminals indicated by the broken line.

B. Simulation Results

1) Cooperative Benefits in Large Networks: To compare the benefits of NFIC-Metanode resource allocation to multihop routing, we simulate networks and allocate resources under the two different paradigms. We assume a pathloss-dominated channel environment. We study the mean throughput for a varying number of flows in the region, where the schedules and routes have been calculated using NFIC techniques. Shown in Figure 4 is a comparison of throughputs for two pathloss environments, free space ($\alpha = 2$) and urban ($\alpha = 4$) as a function of the number of flows demanding resources. The solid lines are mean flow throughputs when three metanodes per terminal are included in the NFIC, allowing the algorithm to exploit cooperative technologies where needed. The broken line indicates throughputs for multi-hop only allocation.

We observe improvements over multihop in both environments and for all levels of congestion, though the improvements are most pronounced with low pathloss and few flows,

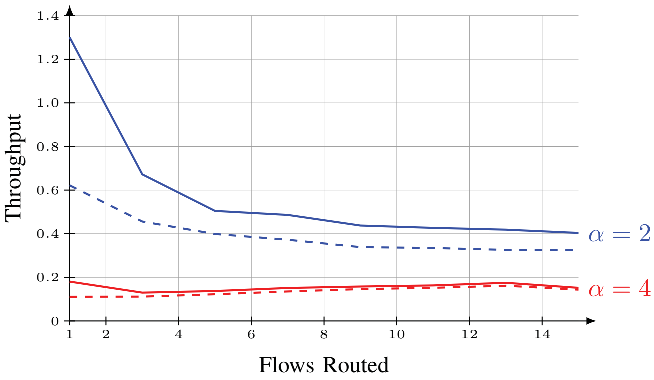


Fig. 4. Mean throughput for networks with cooperative links (solid) and multihop-only routes (broken). Cooperative links, represented by metanodes in the NFIC, considerably increase throughputs for low-pathloss environments, especially when only few flows are competing for resources.

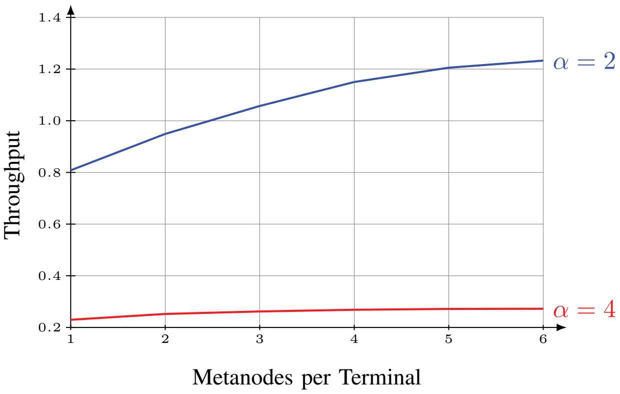


Fig. 5. Mean throughput of two flows as a function of the number of metanodes defined for each terminal in the network. More metanodes are helpful in low-pathloss regimes, though they do increase memory requirements.

where datarates can increase by a factor of 2. Gains diminish for higher pathloss and as congestion increases, since cooperative gains are affected by the overall-higher interference temperature in the network.

2) Number of Metanodes: As discussed above, the number of metanodes chosen for the NFIC affects memory complexity, but it also affects the resource allocation selected. If a relatively small number of metanodes are defined, they may not be useful in the routes required by the flows. Increasing the number of metanodes defined per terminal does increase memory requirements, but also makes cooperative links available to more parts of the network. It is interesting to study network performance as a function of how many metanodes are defined for each terminal in the network.

Figure 5 show this relationship. For each terminal, we define a number of metanodes using local relays as described above, and calculate a resource allocation for that topology with two information flows. We then redefine more metanodes and recalculate the allocation, plotting the mean throughput for the flows as a function of number of metanodes we have defined. This is repeated for a thousand topologies, and average results are reported. Throughput increases considerably if more metanodes are defined in the case of low pathloss, less so if pathloss is high. This is because the throughput on a route is determined by the "bottleneck link" which, once aided with cooperation, may remain the cooperative link. This is the case in high-pathloss environments, where the cooperative advantage is smaller. Cooperative units are much more effective in lower-pathloss environments, where all three channels are stronger.

V. Conclusion & Future Challenges

We have presented a model for wireless networks which captures the salient issues in both physical layer and network layer resource allocation, namely those of rates, schedules, and routes. We have shown how this model can be extended to incorporate cooperative transmissions, and we have used our polynomial-time NFIC allocation technique to show the clear benefits of using cooperation where needed.

Open questions remain, especially in the domain of choosing metanodes. Our geographically localized technique employed here shows benefits, but is heuristic and unoptimized. To maximize the gains possible with NFIC-metanode resource allocation, those metanodes must be carefully selected according to an optimized technique.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, 1948.

[2] G. B. Middleton, B. Aazhang, and J. Lilleberg, "A flexible framework for polynomial-time resource allocation in multiflow wireless networks," *Proceedings of the 47th Allerton Conference on Communication, Control and Computing*, September 2009.

[3] ——, "Efficient resource allocation and interference management for streaming multiflow wireless networks," submitted to *International Conference on Communications*, May 2010.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. part I. System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927 – 1938, November 2003.

[5] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, September 1979.

[6] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Information Theory*, vol. 51, no. 6, pp. 2020–2040, December 2005.

[7] M. A. Khojastepour, "Distributed cooperative communications in wireless networks," Ph.D. dissertation, Rice University, Houston, TX, 2005.

[8] T. H. Corman, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. McGraw-Hill, 1996.

The Causal Cognitive Interference Channel

Shraga I. Bross*, Yossef Steinberg†, and Stephan Tinguely‡
* School of Engineering, Bar-Ilan University, Ramat Gan, 52900, Israel
Email: brosss@macs.biu.ac.il
† Departement of Electrical Engineering, Technion, Haifa 32000, Israel
Email: ysteinbe@ee.technion.ac.il
‡ Email: tinguely@isi.ee.ethz.ch

***Abstract*— The (non-causal) cognitive interference channel, studied recently by Liang *et. al.*, is a model for a classical two-user discrete memoryless interference channel, over which two transmitters send a pair of independent messages. It is assumed that the first message is shared by both encoders, whereas the second message in known only to Encoder 2 – the cognitive transmitter. Receiver 2 needs to decode both messages, and Receiver 1 should decode only the first message while Message 2 should be kept as secret as possible from Receiver 1. The level of secrecy is measured by the equivocation rate. For this model the capacity-equivocation region has been derived by Liang *et. al.*.**

In this work we dispense of the assumption that Message 1 is shared a-priori by both encoders. Instead, we study the case in which Encoder 2 cribs causally from Encoder 1. We derive an achievable rate-equivocation region for this model and establish the capacity-equivocation region for a degraded interference channel.

***Index Terms*—Cognitive interference channel, cribbing encoder.**

I. Introduction

In the classical two-user discrete memoryless interference channel model two encoders transmit a pair of independent messages to a pair of receivers while the signal intended for one receiver causes interference at the other receiver. A cognitive interference channel is an interference channel where it is further assumed that the first message is shared by both encoders whereas the second message is known only to Encoder 2 – the cognitive transmitter. Receiver 2 needs to decode both messages and Receiver 1 should decode only the first message while Message 2 should be kept as secret as possible from Receiver 1. The level of secrecy is measured by the equivocation rate. For this model the capacity-equivocation region has been derived by Liang et al in [1].

In this work we dispense of the assumption that Message 1 is shared a-priori by both encoders. Instead, we study the simplest model in which Encoder 2 acquires Message 1 causally; namely a model in which Encoder 2 "cribs" causally and learns the sequence of channel inputs emitted by Encoder 1 in all past transmissions (in the sense of [2, Situation 2]) before generating its next channel input. The model is depicted in Figure 1.

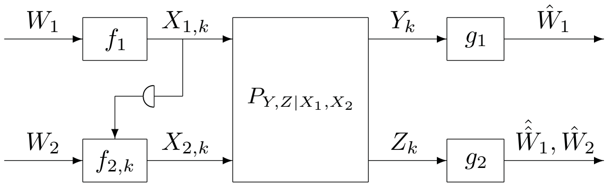


Fig. 1. Interference channel with a cribbing encoder.

First we present an inner bound on the rate-equivocation region for the above model and then we establish the capacity-equivocation region for a degraded interference channel — a model in which conditionally on the first channel input the output at Receiver 1 is degraded w.r.t. the output at Receiver 2.

The paper is organized as follows. In Section II we provide a formal definition for the interference channel with a cribbing encoder and one confidential message. In Section III we present our main results, while Section IV is devoted to the description of our coding scheme establishing the achievability part of our main result.

II. Channel model

A discrete memoryless interference channel is a triple $(\mathcal{X}_1 \times \mathcal{X}_2, p(y,z|x_1,x_2), \mathcal{Y} \times \mathcal{Z})$ where $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite sets corresponding to the input alphabets of Encoder 1 and Encoder 2 respectively, the finite sets $\mathcal{Y}$ and $\mathcal{Z}$ are the output alphabets at Receiver 1 and Receiver 2 respectively, while $p(\cdot,\cdot|x_1,x_2)$ is a collection of probability laws on $\mathcal{Y} \times \mathcal{Z}$ indexed by the input symbols $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. The channel's law extends to n-tuples according to the memoryless law

$$\Pr(y_k, z_k|x_1^k, x_2^k, y^{k-1}, z^{k-1}) = p(y_k, z_k|x_{1,k}, x_{2,k}) \ ,$$

where $x_{1,k}, x_{2,k}, y_k$ and z_k denote the inputs and outputs of the channel at time k, and $x_1^k \triangleq (x_{1,1}, \ldots, x_{1,k})$.

Encoder 1 sends a message W_1 which is drawn uniformly over the set $\{1, \ldots, e^{nR_1}\} \triangleq \mathcal{W}_1$ to both receivers. Encoder 2 sends a message W_2 which is drawn uniformly over the set $\{1, \ldots, e^{nR_2}\} \triangleq \mathcal{W}_2$ to Receiver 2 in such a way that Receiver 1 is unable to decode W_2 reliably. Hence the message W_2 is referred to as the confidential message with respect to Receiver 1. At the same time, given its "partial" knowledge

The work of S. Tinguely was done while he was visiting Bar-Ilan University.
The work of Y. Steinberg was supported by the Israel Science Foundation under Grant 280/07.

about W_1, Encoder 2 assists Encoder 1 in conveying the message W_1 to both receivers.

An (e^{nR_1}, e^{nR_2}, n) code for the interference channel with a cribbing encoder consists of:

1) Encoder 1 defined by a deterministic mapping

$$f_1 \ : \ \mathcal{W}_1 \to \mathcal{X}_1^n \tag{1}$$

which maps a message $w_1 \in \mathcal{W}_1$ to a codeword $x_1^n \in \mathcal{X}_1^n$.

2) Encoder 2 defined by a collection of encoding functions

$$f_{2,k} \ : \ \mathcal{W}_2 \times \mathcal{X}_1^{k-1} \to \mathcal{X}_2 \quad k = 1, 2, \ldots, n \tag{2}$$

which, based on the message $w_2 \in \mathcal{W}_2$ and what was learned from the other encoder by cribbing $x_1^{k-1} \in \mathcal{X}_1^{k-1}$, map into the next channel input $x_{2,k} \in \mathcal{X}_2$.

3) Decoder 1 defined by a mapping

$$g_1 \ : \ \mathcal{Y}^n \to \mathcal{W}_1$$

which maps a received sequence y^n to a message $\hat{w}_1 \in \mathcal{W}_1$.

4) Decoder 2 defined by a mapping

$$g_2 \ : \ \mathcal{Z}^n \to \mathcal{W}_1 \times \mathcal{W}_2$$

which maps a received sequence z^n to a message pair $(\hat{\hat{w}}_1, \hat{w}_2) \in \mathcal{W}_1 \times \mathcal{W}_2$.

For a given code, the block average probability of error is defined as

$$P_e^{(n)} = \frac{1}{e^{n(R_1+R_2)}} \sum_{w_1=1}^{e^{nR_1}} \sum_{w_2=1}^{e^{nR_2}} P_e^{(n)}(w_1, w_2)$$

where

$$P_e^{(n)}(w_1, w_2) = \Pr\left\{(\hat{w}_1, \hat{\hat{w}}_1, \hat{w}_2) \neq (w_1, w_1, w_2) | (w_1, w_2) \text{ sent}\right\}.$$

The secrecy level of W_2 at Receiver 1 is measured by the normalized equivocation $R_e^{(n)} = \frac{1}{n}H(W_2|Y^n)$, and a rate-equivocation tuple (R_1, R_2, R_e) is said to be achievable if there exists a sequence of (e^{nR_1}, e^{nR_2}, n) codes with $\lim_{n\to\infty} P_e^{(n)} = 0$ and $R_e \leq \liminf_{n\to\infty} R_e^{(n)}$.

III. Main results

Our first result for the discrete memoryless interference channel with a cribbing encoder is an achievability result. By combining the coding strategies from [1] and [2] we prove the following.

Proposition 1: Consider the discrete memoryless interference channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, z|x_1, x_2), \mathcal{Y} \times \mathcal{Z})$ with a cribbing encoder. Then, the rate region $\mathcal{R}$ defined by

$$\mathcal{R} = \bigcup_{p_{TUX_1X_2YZ}} \left\{ \begin{array}{c} (R_1, R_2, R_{21}, R_{22}, R_e) : \\ R_1 \geq 0, R_2 \geq 0, R_{21} \geq 0, R_{22} \geq 0, R_e \geq 0 \\ R_2 = R_{21} + R_{22} \\ R_1 \leq H(X_1|T) \\ R_1 + R_{21} \leq I(TUX_1; Y) \\ R_{21} + R_{22} \leq I(UX_2; Z|TX_1) \\ R_1 + R_{21} + R_{22} \leq I(TUX_1X_2; Z) \\ R_{22} \leq I(X_2; Z|TUX_1) \\ R_e \leq R_{22} \\ R_e \leq I(X_2; Z|TUX_1) - I(X_2; Y|TUX_1) \\ R_{21} + R_e \leq I(UX_2; Z|TX_1) - I(X_2; Y|TUX_1) \\ R_1 + R_{21} + R_e \leq I(TUX_1X_2; Z) - I(X_2; Y|TUX_1) \end{array} \right\}, \tag{3}$$

is achievable. The union in (3) is taken over all laws on $T \in \mathcal{T}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}, Z \in \mathcal{Z}$ of the form

$$\begin{aligned} & p_{TUX_1X_2YZ}(t, u, x_1, x_2, y, z) \\ & \quad = \quad p_T(t) p_{U|T}(u|t) p_{X_1|T}(x_1|t) p_{X_2|TU}(x_2|t, u) p(y, z|x_1, x_2). \end{aligned} \tag{4}$$

Applying the *Fourier-Motzkin elimination* to eliminate R_{12} and R_{22}, the achievable rate-equivocation region (3) can be expressed in terms of R_1 and R_2 as follows:

Proposition 2: The achievable rate-equivocation region $\mathcal{R}$ is identical to the rate region $\mathcal{R}'$ defined by

$$\mathcal{R}' = \bigcup_{p_{TUX_1X_2YZ}} \left\{ \begin{array}{c} (R_1, R_2, R_e) : \\ R_1 \geq 0, R_2 \geq 0, R_e \geq 0 \\ R_1 \leq \min\{H(X_1|T), I(TUX_1; Y)\} \\ R_2 \leq I(UX_2; Z|TX_1) \\ R_1 + R_2 \leq \min\{I(TUX_1; Z), I(TUX_1; Y)\} \\ +I(X_2; Z|TUX_1) \\ R_e \leq R_2 \\ R_e \leq I(X_2; Z|TUX_1) - I(X_2; Y|TUX_1) \\ R_1 + R_e \leq I(TUX_1X_2; Z) - I(X_2; Y|TUX_1) \end{array} \right\}, \tag{5}$$

where the union in (5) is taken over all laws of the form (4).

By adding in (5) a bound on R_1, the bound on $R_1 + R_e$ becomes redundant and can thus be removed. This establishes the achievability of the region $\hat{\mathcal{R}} \subseteq \mathcal{R}'$ which is defined as follows.

$$\hat{\mathcal{R}} = \bigcup_{p_{TUX_1X_2YZ}}$$

$$\left\{\begin{array}{c} (R_1, R_2, R_e): \\ R_1 \geq 0, R_2 \geq 0, R_e \geq 0 \\ R_1 \leq \min\{H(X_1|T), I(TUX_1;Y), I(TUX_1;Z)\} \\ R_2 \leq I(UX_2;Z|TX_1) \\ R_1 + R_2 \leq \min\{I(TUX_1;Z), I(TUX_1;Y)\} \\ \quad +I(X_2;Z|TUX_1) \\ R_e \leq R_2 \\ R_e \leq I(X_2;Z|TUX_1) - I(X_2;Y|TUX_1) \end{array}\right\}, \tag{6}$$

where the union in (6) is taken over all laws of the form (4).

Definition 1: A discrete memoryless interference channel is degraded if the channel law decomposes as

$$\begin{aligned} & p_{YZ|X_1X_2}(y_k, z_k|x_{1,k}x_{2,k}) \\ & = p_{Z|X_1X_2}(z_k|x_{1,k}x_{2,k})p_{Y|ZX_1}(y_k|z_kx_{1,k}). \end{aligned} \tag{7}$$

We next consider an interference channel that is *degraded* in which case we have the following conclusive result.

Theorem 1: Consider a discrete memoryless interference channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y,z|x_1,x_2), \mathcal{Y} \times \mathcal{Z})$ with a cribbing encoder that is degraded. The capacity-equivocation rate region for such a channel is given by

$$\mathcal{C} = \bigcup_{p_{TUX_1X_2YZ}}$$

$$\left\{\begin{array}{c} (R_1, R_2, R_e): \\ R_1 \geq 0, R_2 \geq 0, R_e \geq 0 \\ R_1 \leq \min\{H(X_1|T), I(TUX_1;Y), I(TUX_1;Z)\} \\ R_2 \leq I(UX_2;Z|TX_1) \\ R_1 + R_2 \leq \min\{I(TUX_1;Z), I(TUX_1;Y)\} \\ \quad +I(X_2;Z|TUX_1) \\ R_e \leq R_2 \\ R_e \leq I(X_2;Z|TUX_1) - I(X_2;Y|TUX_1) \end{array}\right\}, \tag{8}$$

where the union in (8) is taken over all laws on $T \in \mathcal{T}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}, Z \in \mathcal{Z}$ of the form (4), and given an auxiliary r.v. T, the cardinality of the auxiliary random variable U is bounded by $\|\mathcal{U}\| \leq \|\mathcal{T}\| \cdot \|\mathcal{X}_1\| \cdot \|\mathcal{X}_2\| + 4$.

Proof: Due to lack of space the converse proof of Theorem 1, which establishes that under the technical assumption (7), $\mathcal{C} = \hat{\mathcal{R}} = \mathcal{R}'$, is omitted. For the direct proof of Theorem 1, we limit ourselves to the description of the coding scheme. This is the subject of Section IV.

IV. The coding scheme

We propose a coding scheme that is based on Block-Markov superposition encoding and which combines the coding technique of [1] with the backward decoding idea of [2].

We consider B blocks, each of n symbols. We split the message (W_1, W_2) into a sequence of $B-1$ sub-messages $(W_1^{(b)}, W_2^{(b)})$, for $b = 1, \ldots, B-1$, where $W_2^{(b)}$ consists of the pair $(W_{21}^{(b)}, W_{22}^{(b)})$. Here the sequence $\{W_1^{(b)}\}$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_1}\}$ and independent thereof $\{W_2^{(b)}\}$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_{21}}\} \times \{1, \ldots, e^{nR_{22}}\}$. As $B \to \infty$, for fixed n, the rate pair of the message (W_1, W_2), $(\tilde{R}_1, \tilde{R}_2) = (R_1(B-1)/B, (R_{21} + R_{22})(B-1)/B)$, is arbitrarily close to $(R_1, R_{21} + R_{22})$.

We assume a tuple of random variables $T \in \mathcal{T}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}, Z \in \mathcal{Z}$ of joint law (4).

Random coding and partitioning: In each block $b, b = 1, 2, \ldots, B$, we shall use the following code.

- Generate e^{nR_1} sequences $\boldsymbol{t} = (t_1, \ldots, t_n)$, each with probability $\Pr(\boldsymbol{t}) = \prod_{k=1}^n p_T(t_k)$. Label them $\boldsymbol{t}(\omega_0)$ where $\omega_0 \in \{1, \ldots, e^{nR_1}\}$.
- For each $\boldsymbol{t}(\omega_0)$ generate e^{nR_1} sequences $\boldsymbol{x}_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$, each with probability $\Pr(\boldsymbol{x}_1|\boldsymbol{t}(\omega_0)) = \prod_{k=1}^n p_{X_1|T}(x_{1,k}|t_k(\omega_0))$. Label them $\boldsymbol{x}_1(i, \omega_0), i \in \{1, \ldots, e^{nR_1}\}$.
- For each $\boldsymbol{t}(\omega_0)$ generate $e^{nR_{21}}$ sequences $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, each with probability $\Pr(\boldsymbol{u}|\boldsymbol{t}(\omega_0)) = \prod_{k=1}^n p_{U|T}(u_k|t_k(\omega_0))$. Label them $\boldsymbol{u}(j, \omega_0), j \in \{1, \ldots, e^{nR_{21}}\}$.
- For each $(\boldsymbol{t}(\omega_0), \boldsymbol{u}(j, \omega_0))$ generate $e^{n(R_\alpha + R_\beta)}$ sequences $\boldsymbol{x}_2 = (x_{2,1}, x_{2,2}, \ldots, x_{2,n})$, each with probability $\Pr(\boldsymbol{x}_2|\boldsymbol{t}(\omega_0), \boldsymbol{u}(j, \omega_0)) = \prod_{k=1}^n p_{X_2|TU}(x_{2,k}|t_k(\omega_0), u_k(j, \omega_0))$. Label them $\boldsymbol{x}_2(\alpha, \beta, j, \omega_0), \alpha \in \{1, \ldots, e^{nR_\alpha}\}, \beta \in \{1, \ldots, e^{nR_\beta}\}$ with $R_\beta \triangleq I(X_2;Y|TUX_1)$, $R_\alpha \triangleq R'_{22} - R_\beta \geq 0$ and $R'_{22} = R_{22} + \Delta$ for some $\Delta > 0$. Consequently, let $R_P \triangleq R_{22} - R_\alpha = R_\beta - \Delta$.
- If $R_P > 0$: Randomly partition the set $\{1, \ldots, e^{nR_\beta}\}$ into e^{nR_P} cells. Label the cells $p \in \{1, \ldots, e^{nR_P}\}$ and let $p(s) = c$ if s belongs to cell c. In the sequel we shall refer to this partition as Partition 1.

Encoding : We denote the realizations of the sequences $\{W_1^{(b)}\}$, $\{W_{21}^{(b)}\}$, and $\{W_{22}^{(b)}\}$ by $\{w_1^{(b)}\}$, $\{w_{21}^{(b)}\}$, and $\{w_{22}^{(b)}\}$. The code builds upon a Block-Markov structure in which the message $(w_1^{(b)}, w_{21}^{(b)}, w_{22}^{(b)})$ is encoded over the successive blocks b and $(b+1)$ such that, $\omega_0^{(b)} = w_1^{(b-1)}$, for $b = 1, \ldots, B-1$.

The messages $\{w_1^{(b)}\}$, $\{w_{21}^{(b)}\}$, and $\{w_{22}^{(b)}\}$, $b = 1, 2, \ldots, B-1$ are encoded as follows:

In block 1 the encoders send

$$\begin{aligned} \boldsymbol{x}_1^{(1)} &= \boldsymbol{x}_1(w_1^{(1)}, 1) \\ \boldsymbol{x}_2^{(1)} &= \boldsymbol{x}_2(\alpha(w_{22}^{(1)}), \beta(w_{22}^{(1)}), w_{21}^{(1)}, 1). \end{aligned}$$

Here, the encoding $\alpha(w_{22})$ and $\beta(w_{22})$ is defined as follows:

1) $R_P > 0$: Let $w_{22} = (a, p)$ where $a \in \{1, \ldots, e^{nR_\alpha}\}$ and $p \in \{1, \ldots, e^{nR_P}\}$ then $\alpha(w_{22}) = a$ and $\beta(w_{22}) = s$ where s is chosen randomly within the cell p in Partition 1.
2) $R_P < 0$: Let $\alpha(w_{22}) = w_{22}$ and $\beta(w_{22}) = s$ where s is chosen randomly within the set $\{1, \ldots, e^{nR_\beta}\}$.

Suppose that, as a result of cribbing from Encoder 1, before the beginning of block $b = 2, 3, \ldots, B$, Encoder 2 has an estimate $\hat{\hat{w}}_1^{(b-1)}$ for $w_1^{(b-1)}$. Then, in block $b = 2, 3, \ldots, B-1$, the encoders send

$$\begin{aligned}\boldsymbol{x}_1^{(b)} &= \boldsymbol{x}_1(w_1^{(b)}, w_1^{(b-1)})\\ \boldsymbol{x}_2^{(b)} &= \boldsymbol{x}_2(\alpha(w_{22}^{(b)}), \beta(w_{22}^{(b)}), w_{21}^{(b)}, \hat{\hat{w}}_1^{(b-1)}),\end{aligned}$$

and in block B

$$\begin{aligned}\boldsymbol{x}_1^{(B)} &= \boldsymbol{x}_1(1, w_1^{(B-1)})\\ \boldsymbol{x}_2^{(B)} &= \boldsymbol{x}_2(1, 1, 1, \hat{\hat{w}}_1^{(B-1)}).\end{aligned}$$

Decoding at the receivers: After the reception of block-B both receivers use backward decoding starting from block B downward to block 1 and decode the messages as follows.

In block B Decoder 1 looks for $\hat{w}_1^{(B-1)}$ such that

$$\Big(\boldsymbol{t}(\hat{w}_1^{(B-1)}), \boldsymbol{x}_1(1, \hat{w}_1^{(B-1)}), \boldsymbol{u}(1, \hat{w}_1^{(B-1)}),\\ \boldsymbol{x}_2(1, 1, 1, \hat{w}_1^{(B-1)}), \boldsymbol{y}^{(B)}\Big) \in \mathcal{A}_\epsilon(T, X_1, U, X_2, Y).$$

Next, assume that, decoding backwards up to (and including) block $b+1$, Decoder 1 decoded $\hat{w}_1^{(B-1)}, \hat{w}_1^{(B-2)}, \ldots, \hat{w}_1^{(b)}$. To decode block b, Decoder 1 looks for $\hat{w}_1^{(b-1)}$ such that

$$\Big(\boldsymbol{t}(\hat{w}_1^{(b-1)}), \boldsymbol{x}_1(\hat{w}_1^{(b)}, \hat{w}_1^{(b-1)}), \boldsymbol{u}(\hat{w}_{21}^{(b)}, \hat{w}_1^{(b-1)}), \boldsymbol{y}^{(b)}\Big)\\ \in \mathcal{A}_\epsilon(T, X_1, U, Y),$$

for some $\hat{w}_{21}^{(b)} \in \mathcal{W}_{21}$ — i.e. Decoder 1 looks just for the "cloud center" $\omega_0^{(b)}$ such that $(\boldsymbol{t}(\omega_0^{(b)}), \boldsymbol{x}_1(\hat{w}_1^{(b)}, \omega_0^{(b)}), \boldsymbol{u}(\cdot, \omega_0^{(b)}), \boldsymbol{y}^{(b)})$ is jointly typical.

Similarly, in block B Decoder 2 looks for $\hat{\hat{w}}_1^{(B-1)}$ such that

$$\Big(\boldsymbol{t}(\hat{\hat{w}}_1^{(B-1)}), \boldsymbol{x}_1(1, \hat{\hat{w}}_1^{(B-1)}), \boldsymbol{u}(1, \hat{\hat{w}}_1^{(B-1)}),\\ \boldsymbol{x}_2(1, 1, 1, \hat{\hat{w}}_1^{(B-1)}), \boldsymbol{z}^{(B)}\Big) \in \mathcal{A}_\epsilon(T, X_1, U, X_2, Z).$$

Next, assume that, decoding backwards up to (and including) block $b + 1$, Decoder 2 decoded $\hat{\hat{w}}_1^{(B-1)}, (\hat{w}_{22}^{(B-1)}, \hat{w}_{21}^{(B-1)}, \hat{\hat{w}}_1^{(B-2)}), \ldots, (\hat{w}_{22}^{(b+1)}, \hat{w}_{21}^{(b+1)}, \hat{\hat{w}}_1^{(b)})$. To decode block b, Decoder 2 looks for $(\hat{w}_{22}^{(b)}, \hat{w}_{21}^{(b)}, \hat{\hat{w}}_1^{(b-1)})$ such that

$$\Big(\boldsymbol{t}(\hat{\hat{w}}_1^{(b-1)}), \boldsymbol{x}_1(\hat{\hat{w}}_1^{(b)}, \hat{\hat{w}}_1^{(b-1)}), \boldsymbol{u}(\hat{w}_{21}^{(b)}, \hat{\hat{w}}_1^{(b-1)}),\\ \boldsymbol{x}_2(\alpha(\hat{w}_{22}^{(b)}), \beta(\hat{w}_{22}^{(b)}), \hat{w}_{21}^{(b)}, \hat{\hat{w}}_1^{(b-1)}), \boldsymbol{z}^{(b)}\Big)\\ \in \mathcal{A}_\epsilon(T, X_1, U, X_2, Z).$$

Decoding at Encoder 2: To obtain cooperation, after block $b = 1, 2, \ldots, B-1$, Encoder 2 chooses $\tilde{w}_1^{(b)}$ such that

$$\Big(\boldsymbol{t}(\tilde{\omega}_0^{(b)}), \boldsymbol{x}_1(\tilde{w}_1^{(b)}, \tilde{\omega}_0^{(b)}), \boldsymbol{x}_1^{(b)}\Big) \in \mathcal{A}_\epsilon(T, X_1, X_1),$$

where $\tilde{\omega}_0^{(b)} = \tilde{w}_1^{(b-1)}$ was determined at the end of block $b-1$ and $\tilde{\omega}_0^{(1)} = 1$.

Optional decoding at Decoder 1: When Decoder 1 is given the triple $\omega_0^{(b)}, w_1^{(b)}, w_{21}^{(b)}$ and $\alpha(w_{22}^{(b)})$ it decodes $\beta(w_{22}^{(b)})$ by choosing $\hat{\beta}(w_{22}^{(b)})$ such that

$$\Big(\boldsymbol{t}(\omega_0^{(b)}), \boldsymbol{x}_1(w_1^{(b)}, \omega_0^{(b)}), \boldsymbol{u}(w_{21}^{(b)}, \omega_0^{(b)}),\\ \boldsymbol{x}_2(\alpha(w_{22}^{(b)}), \hat{\beta}(w_{22}^{(b)}), w_{21}^{(b)}, \omega_0^{(b)}), \boldsymbol{y}^{(b)}\Big)\\ \in \mathcal{A}_\epsilon(T, X_1, U, X_2, Y).$$

When a decoding step either fails to recover a unique index (or index triple) which satisfies the decoding rule, or there is more than one index (or index triple), then an index (or an index triple) is chosen at random.

The achievability of the rate region (3) can now be established by upper bounding the probability of the possible error events associated with this coding scheme.

References

[1] Y. Liang, A. Somekh-Baruch, H.V. Poor, S. Shamai (Shitz) and S. Verdú, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inform. Theory,* vol. IT-55, no. 2, pp. 604-619, Feb. 2009.

[2] F.M.J. Willems and E.C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders", *IEEE Trans. Inform. Theory,* vol. IT-31, no. 3, pp. 313-327, May 1985.

State of the cognitive interference channel: a new unified inner bound

Stefano Rini, Daniela Tuninetti and Natasha Devroye
University of Illinois at Chicago
Chicago, IL 60607, USA
Email: srini2, danielat, devroye@uic.edu

Abstract—**The capacity region of the interference channel in which one transmitter non-causally knows the message of the other, termed the cognitive interference channel, has remained open since its inception in 2005. A number of subtly differing achievable rate regions and outer bounds have been derived, some of which are tight under specific conditions. In this work we present a new unified inner bound for the discrete memoryless cognitive interference channel. We show explicitly how it encompasses all known discrete memoryless achievable rate regions as special cases. The presented achievable region was recently used in deriving the capacity region of the linear high-SNR deterministic approximation of the Gaussian cognitive interference channel. The high-SNR deterministic approximation was then used to obtain the capacity of the Gaussian cognitive interference channel to within 1.87 bits.**

I. Introduction

The cognitive interference channel (CIFC)[1] is an interference channel in which one of the transmitters - dubbed the cognitive transmitter - has non-causal knowledge of the message of the other - dubbed the primary - transmitter. The study of this channel is motivated by cognitive radio technology which allows wireless devices to sense and adapt to their RF environment by changing their transmission parameters in software on the fly. One of the driving applications of cognitive radio technology is secondary spectrum sharing: currently licensed spectrum would be shared by primary (legacy) and secondary (usually cognitive) devices in the hope of improving spectral efficiency. The extra abilities of cognitive radios may be modeled information theoretically in a number of ways - see [6], [11] for surveys - one of which is through the assumption of non-causal primary message knowledge at the secondary, or cognitive, transmitter.

The two-dimensional capacity region of the CIFC has remained open in general since its inception in 2005 [7]. However, capacity is known in a number of channels:

- **General deterministic CIFCs.** Fully deterministic CIFCs in the flavor of the deterministic interference channel [1] are being considered in [24, Ch.3], where new inner and outer bounds are shown to meet in certain classes of channels. A special case of the deterministic CIFC is the deterministic linear high-SNR approximation of the Gaussian CIFC, whose capacity region, in the spirit of [2], was obtained in [22].
- **Semi-deterministic CIFCs.** In [4] the capacity region for a class of channels in which the signal at the cognitive receiver is a deterministic function of the channel inputs is derived.
- **Discrete memoryless CIFCs.** First considered in [7], [8], its capacity region was obtained for very strong interference in [13] and for weak interference in [29]. Prior to this work and the recent work of [4], the largest known achievable rate regions were those of [8], [9], [14], [19]. The recent and independently derived region of [4] was shown to contain [14], [19], but was not conclusively shown to encompass [8] or the larger region of [9].
- **Gaussian CIFC.** This capacity region under weak interference was obtained in [15], [29], while that for very strong interference follows from [13]. Capacity for a class of Gaussion MIMO CIFCs is obtained in [28].
- **Z-CIFCs.** Inner and outer bounds when the cognitive-primary link is noiseless are obtained in [3], [18]. The Gaussian causal case is considered in [4], and is related to the general (non Z) causal CIFC explored in [26].
- **CIFCs with secrecy constraints.** Capacity of a CIFC in which the cognitive message is to be kept secret from the primary and the cognitive wishes to decode both messages is obtained in [17]. A cognitive multiple-access wiretap channel is considered in [27].

We focus on the discrete memoryless CIFC (DM-CIFC) and propose a new achievable rate region and show explicitly how it encompasses or reduces to all other known achievable rate regions. The best known outer bounds for the DM-CIFC are those of [19]. The new unified achievable rate region has been shown to be useful as: 1) specific choices of random variables yield the capacity region of the linear high-SNR approximation of the Gaussian CIFC [22], 2) specific choices of random variables yield capacity in certain regimes of the deterministic CIFC [24] and 3) specific choices of Gaussian random variables have resulted in an achievable rate region which lies within 1.87 bits, regardless of channel parameters, of an outer bound [25]. Numerical simulations indicate the actual gap is smaller.

II. Channel Model

The Discrete Memoryless Cognitive InterFerence Channel (DM-CIFC), as shown in Fig. 1, consists of two transmitter-

[1] Other names for this channel include the cognitive radio channel [8], interference channel with degraded message sets [14], [29], the non-causal interference channel with one cognitive transmitter [4], the interference channel with one cooperating transmitter [19] and the interference channel with unidirectional cooperation [13], [20].

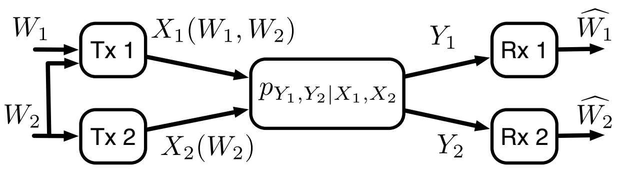


Fig. 1. The Cognitive Interference Channel.

receiver pairs that exchange independent messages over a common channel. Transmitter $i, i \in \{1, 2\}$, has discrete input alphabet $\mathcal{X}_i$ and its receiver has discrete output alphabet $\mathcal{Y}_i$. The channel is assumed to be memoryless with transition probability $p_{Y_1,Y_2|X_1,X_2}$. Encoder i, $i \in \{1, 2\}$, wishes to communicate a message W_i uniformly distributed on $\mathcal{M}_i = [1 : 2^{NR_i}]$ to decoder i in N channel uses at rate R_i. Encoder 1 (i.e., the cognitive user) knows its own message W_1 and that of encoder 2 (the primary user), W_2. A rate pair (R_1, R_2) is achievable if there exist sequences of encoding functions

$$\begin{aligned} X_1^N &= f_1^N(W_1, W_2), & f_1 &: \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}_1^N, \\ X_2^N &= f_2^N(W_2), & f_2 &: \mathcal{M}_2 \to \mathcal{X}_2^N, \end{aligned}$$

with corresponding sequences of decoding functions

$$\begin{aligned} \widehat{W}_1 &= g_1^N(Y_1^N), & g_1 &: \mathcal{Y}_1^N \to \mathcal{M}_1, \\ \widehat{W}_2 &= g_2^N(Y_2^N), & g_2 &: \mathcal{Y}_2^N \to \mathcal{M}_2. \end{aligned}$$

The capacity region is defined as the closure of the region of achievable (R_1, R_2) pairs [5]. Standard strong-typicality is assumed; properties may be found in [16].

III. A NEW UNIFIED ACHIEVABLE RATE REGION

As the DM-CIFC encompasses classical interference, multiple-access and broadcast channels, we expect to see a combination of their achievability proving techniques surface in any unified scheme for the CIFC:

- **Rate-splitting.** As in [12] for the interference-channel and [8], [14], [19] for the CIFC, rate-splitting is not necessary in the weak [29] and strong [13] interference regimes.
- **Superposition-coding.** Useful in multiple-access and broadcast channels [5], the superposition of private messages on top of common ones [14], [19] is proposed and is known to be capacity achieving in very strong interference [13].
- **Binning.** Gel'fand-Pinsker coding [10], often referred to as binning, allows a transmitter to "cancel" (portions of) the interference known to it at its intended receiver. Related binning techniques are used by Marton in deriving the largest known DM-broadcast channel achievable rate region [21].

We now present a new achievable region for the DM-CIFC which generalizes all best known achievable rate regions including [8], [14], [19], [29] as well as [4].

Theorem 1: Region $\mathcal{R}_{RTD}$. A rate pair (R_1, R_2) such that

$$R_1 = R_{1c} + R_{1pb}, \tag{1}$$

$$R_2 = R_{2c} + R_{2pa} + R_{2pb} \tag{2}$$

is achievable for a DM-CIFC if $(R_0', R_1', R_2', R_{1c}, R_{1pb}, R_{2c}, R_{2pa}, R_{2pb}) \in \mathbb{R}_+^8$ satisfies (3a)–(3j) for some input distribution $p_{X_1,X_2,U_{1c},U_{2c},U_{2pa},U_{1pb},U_{2pb}}$.

The encoding scheme used in deriving this achievable rate region is shown in Fig.2. The key aspects of our scheme are the following, where we drop n for convenience:

- We **rate-split** the independent messages W_1 and W_2 uniformly distributed on $\mathcal{M}_1 = [1 : 2^{nR_1}]$ and $\mathcal{M}_2 = [1 : 2^{nR_2}]$ into the messages W_i, $i \in \{1c, 2c, 1pb, 2pb, 2pa\}$, all independent and uniformly distributed on $[1 : 2^{nR_i}]$, each encoded using the random variable U_i, such that

$$\begin{aligned} W_1 &= (W_{1c}, W_{1pb}), & R_1 &= R_{1c} + R_{1pb}, \\ W_2 &= (W_{2c}, W_{2pb}, W_{2pa}), & R_2 &= R_{2c} + R_{2pa} + R_{2pb}. \end{aligned}$$

- **Tx2 (primary Tx):** We **superimpose** U_{2pa}, which encodes the private ("p" for private, "a" for alone) message of Tx2 on top of U_{2c}, which encodes the common ("c" for common) message of Tx2. Tx2 sends X_2 over the channel.
- **Tx1 (cognitive Tx):** The common message of Tx1, encoded by U_{1c}, is **binned** against (U_{2pa}, X_2) conditioned on U_{2c}. The private message of Tx2, encoded by U_{2pb} ("b" for broadcast) and a portion of the private message of Tx1, encoded as U_{1pb}, are **binned** against each other and X_2 as in Marton's region [21] conditioned on U_{1c}, U_{2c}, U_{2pa} and U_{1c}, U_{2c} respectively. Tx1 sends X_1 over the channel. The incorporation of a Marton-like scheme at the cognitive transmitter was initially motivated by the fact that in certain regimes, this strategy was shown to be capacity achieving for the linear high-SNR deterministic CIFC [22]. It is also, independently, a key feature of the region in [4].

The codebook generation, encoding and decoding as well as the error event analysis are provided in [24, Ch.2].

IV. COMPARISON WITH EXISTING ACHIEVABLE REGIONS

We now show that the region of Theorem 1 contains all other known achievable rate regions for the DM-CIFC. We note that showing inclusion of the rate regions [4, Thm.2] and [9] is sufficient to demonstrate the largest known DM-CIFC region, since the region of [4] is shown to contain those of [19, Th.1] and [14]. However we include the independently derived inclusions of the regions of [19, Th.1], [14] and [21, Thm. 2] in our region $\mathcal{R}_{RTD}$ for completeness.

A. Maric et al.'s region [19, Th.1]

Note that, given the encoding and decoding scheme of [19, Th.1], rate splitting of message 2 does not enlarge the region, and hence $X_{2a} = \emptyset$ WLOG. This derivation is included in the Appendix of the long version of this work, found in [23]. To prove inclusion of [19, Th.1] in $\mathcal{R}_{RTCD}$ consider the following

$$\begin{aligned}
R_0' &\geq I(U_{1c};U_{2pa},X_2|U_{2c}) && \text{(3a)}\\
R_0'+R_1'+R_2' &\geq I(U_{1c};U_{2pa},X_2|U_{2c})+I(U_{1pb};U_{2pa},U_{2pb},X_2|U_{2c},U_{1c})\\
&\quad +I(U_{2pb};X_2|U_{2c},U_{2pa},U_{1c}) && \text{(3b)}\\
R_{2c}+R_{1c}+R_{2pa}+R_{2pb}+R_0'+R_2' &\leq I(Y_2;U_{1c},U_{2c},U_{2pa},U_{2pb})+I(U_{1c};U_{2pa}|U_{2c}) && \text{(3c)}\\
R_{1c}+R_{2pa}+R_{2pb}+R_0'+R_2' &\leq I(Y_2;U_{1c},U_{2pa},U_{2pb}|U_{2c})+I(U_{1c};U_{2pa}|U_{2c}) && \text{(3d)}\\
R_{2pa}+R_{2pb}+R_2' &\leq I(Y_2;U_{2pa},U_{2pb}|U_{2c},U_{1c})+I(U_{1c};U_{2pa}|U_{2c}) && \text{(3e)}\\
R_{1c}+R_{2pb}+R_0'+R_2' &\leq I(Y_2;U_{1c},U_{2pb}|U_{2c},U_{2pa})+I(U_{1c};U_{2pa}|U_{2c}) && \text{(3f)}\\
R_{2pb}+R_2' &\leq I(Y_2;U_{2pb}|U_{2c},U_{1c},U_{2pa})+I(U_{2pa};U_{1c}|U_{2c}) && \text{(3g)}\\
R_{2c}+R_{1c}+R_{1pb}+R_0'+R_1' &\leq I(Y_1;U_{2c},U_{1c},U_{1pb}) && \text{(3h)}\\
R_{1c}+R_{1pb}+R_0'+R_1' &\leq I(Y_1;U_{1c},U_{1pb}|U_{2c}) && \text{(3i)}\\
R_{1pb}+R_1' &\leq I(Y_1;U_{1pb}|U_{2c},U_{1c}) && \text{(3j)}
\end{aligned}$$

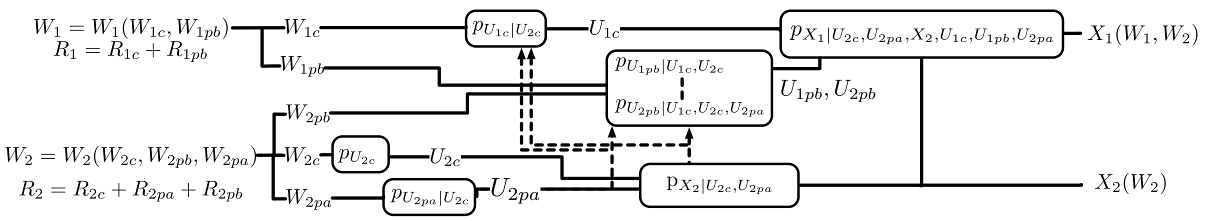

$W_1 = W_1(W_{1c}, W_{1pb})$
$R_1 = R_{1c} + R_{1pb}$
W_{1c}
W_{1pb}
$p_{U_{1c}|U_{2c}}$
U_{1c}
$p_{X_1|U_{2c},U_{2pa},X_2,U_{1c},U_{1pb},U_{2pa}}$
$X_1(W_1, W_2)$
$p_{U_{1pb}|U_{1c},U_{2c}}$
$p_{U_{2pb}|U_{1c},U_{2c},U_{2pa}}$
U_{1pb}, U_{2pb}
W_{2pb}
$W_2 = W_2(W_{2c}, W_{2pb}, W_{2pa})$
$R_2 = R_{2c} + R_{2pa} + R_{2pb}$
W_{2c}
$p_{U_{2c}}$
U_{2c}
W_{2pa}
$p_{U_{2pa}|U_{2c}}$
U_{2pa}
$\mathrm{p}_{X_2|U_{2c},U_{2pa}}$
$X_2(W_2)$

Fig. 2. The achievable encoding scheme of Thm 1. The ordering from left to right and the distributions demonstrate the codebook generation process. The dotted lines indicate binning. We see rate splits are used at both users, private messages $W_{1pb}, W_{2pa}, W_{2pb}$ are superimposed on common messages W_{1c}, W_{2c}. U_{1pb} and U_{2pb} are binned against each other and X_2 conditioned on U_{1c}, U_{2c} and U_{1c}, U_{2c}, U_{2pa} respectively

assignment in (3):

$$\begin{array}{lll}
U_{2c}=Q & R_{2c}=0 & \\
U_{1c}=U_{1c} & R_{1c}=R_{1c} & R_0'=I(U_{1c};U_{2pa}|U_{2c})\\
U_{2pa}=X_{2b} & R_{2pa}=R_{2b} & \\
U_{1pb}=U_{1a} & R_{1pb}=R_{1a} & R_1'=I(U_{1pb};U_{2pa}|U_{1c})\\
U_{2pb}=U_{2pa} & R_{2pb}=0 & R_2'=0
\end{array}$$

Moreover let X_1 and X_2 be deterministic functions, that is $X_2 = f_{X_2}(U_{2c}, U_{2pa})$ and $X_1 = f_{X_1}(U_{2c}, U_{2pa}, U_{1c}, U_{1pb}, U_{2pb})$. With this assignment note that we may drop (3g) and (3f) since incorrect decoding of U_{1c} at decoder 2 is not an error.

Also X_2 can be dropped from the binning rates since $I(X;Y|Z) = I(X;Y,g(Y,Z)|Z)$.

From this we conclude that the region of [19] $\subseteq \mathcal{R}_{RTD}$. The weak interference regions of [15], [29] are special cases of [19, Th.1] by [19, Thm. 3], and are also $\subseteq \mathcal{R}_{RTD}$.

B. Marton's region [21, Thm. 2]

One key ingredient that was missing in all previous regions, as also noted in [4] and first addressed in the context of the CIFC in [3], was the inclusion of a broadcast strategy from the cognitive Tx to both receivers. To remedy this obvious gap, we proposed a Marton-like [21] binning of U_{1pb} and U_{2pb}. Our region may be reduced to Marton's broadcast channel region, using the notation of [21, Thm. 2] by the following assignment of random variables:

$$\begin{array}{ll}
U_{1pb}=U & R_{1pb}=R_x,\\
U_{2pb}=V & R_{2pb}=R_z,\\
 & R_1'+R_2'=I(U_{1pb};U_{2pb}|U_{2c})\\
U_{2c}=U_{1c}=U_{2pa}=W & R_{1c}=R_{2c}=R_{2pa}=0\\
 & R_0'=0\\
X_2=f_{X_2}(U_{2c}). & X_1=f_{X_1}(U_{1pb},U_{2pb},U_{2c})
\end{array}$$

C. Jiang and Xin's region [14]

We compare $\mathcal{R}_{RTD}$ with the region described by (11)-(12), (17)-(19) of [14]. Note that the indices 1 and 2 are switched. Our region may be reduced, with some manipulation, to that of [14] for the following choices of random variables:

$$\begin{array}{lll}
U_{2c}=Q & R_{2c}=0 & \\
U_{1c}=U & R_{1c}=R_{21} & R_0'=I(U_{1c};U_{2pa}|U_{2c})\\
U_{2pa}=W & R_{2pa}=R_1 & \\
U_{1pb}=V & R_{1pb}=R_{22} & R_1'=I(U_{1pb};U_{2pa}|U_{2c},U_{1c})\\
U_{2pb}=(U_{2c},U_{2pa}) & R_{2pb}=0 & R_2'=0
\end{array}$$

Note that we may again drop (3g) and (3f) since incorrect decoding of U_{1c} at decoder 2 is not an error.

D. Devroye et al.'s region [9, Thm. 1]

The comparison of the region of [9, Thm. 1] with that of [4] and [19] has been unsuccessfully attempted in the past. In the Appendix of [23] we show that the region of [9, Thm. 1] $\mathcal{R}_{DMT}$, is contained in our new region $\mathcal{R}_{RTD}$ along the lines:

- We make a correspondence between the random variables and corresponding rates of $\mathcal{R}_{DMT}$ and $\mathcal{R}_{RTD}$.
- We define new regions $\mathcal{R}_{DMT} \subseteq \mathcal{R}_{DMT}^{out}$ and $\mathcal{R}_{RTD}^{in} \subseteq \mathcal{R}_{RTD}$ which are easier to compare: they have identical input distribution decompositions and similar rate equations.
- For any fixed input distribution, an equation-by-equation comparison leads to $\mathcal{R}_{DMT} \subseteq \mathcal{R}_{DMT}^{out} \subseteq \mathcal{R}_{RTD}^{in} \subseteq \mathcal{R}_{RTD}$.

E. Cao and Chen's region [4, Thm. 2]

The independently derived region in [4, Thm. 2] uses a similar encoding structure as that of $\mathcal{R}_{RTD}$ with two exceptions: a) the binning is done sequentially rather than jointly as in $\mathcal{R}_{RTD}$ leading to binning constraints (43)–(45) in [4, Thm. 2] as opposed to (3a)–(3b) in Thm.1. Notable is that both schemes have adopted a Marton-like binning scheme at the cognitive transmitter, as first introduced in the context of the CIFC in [3]. b) While the cognitive messages are rate-split in identical fashions, the primary message is split into 2 parts in [4, Thm. 2] ($R_1 = R_{11} + R_{10}$, note the reversal of indices) while we explicitly split the primary message into three parts $R_2 = R_{2c} + R_{2pa} + R_{2pb}$. In the Appendix of [23] we show that the region of [4, Thm.2], denoted as $\mathcal{R}_{CC} \subseteq \mathcal{R}_{RTD}$:

- We first show that we may WLOG set $U_{11} = \emptyset$ in [4, Thm.2], creating a new region R'_{CC}.
- We next make a correspondence between our random variables and those of [4, Thm.2] and obtain identical regions.

V. Conclusion

A new achievable rate region for the DM-CIFC has been derived and shown to encompass all known achievable rate regions. Of note is the inclusion of a Marton-like broadcasting scheme at the cognitive transmitter. Specific choices of this region have been shown to achieve capacity for the linear high-SNR approximation of the Gaussian CIFC [22], [24], and lead to capacity achieving points in the deterministic CIFC [24]. This region has furthermore been shown to achieve within 1.87 bits of an outer bound, regardless of channel parameters in [25]. Numerical evaluation of the region under Gaussian input distributions for the Gaussian CIFC, and further comparisons with the region of [4] are our short-term goals, while extensions of the CIFC to multiple users will be investigated in the longer term.

References

[1] A. El Gamal and M.H.M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 343–346, Mar. 1982.

[2] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic model for wireless relay networks an its capacity," in *Information Theory for Wireless Networks, 2007 IEEE Information Theory Workshop on*, July 2007, pp. 1–6.

[3] Y. Cao and B. Chen, "Interference channel with one cognitive transmitter," in *Asilomar Conference on Signals, Systems, and Computers*, Oct. 2008.

[4] ——, "Interference Channels with One Cognitive Transmitter," *Arxiv preprint arXiv:09010.0899v1*, 2009.

[5] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[6] N. Devroye, P. Mitran, M. Sharif, S. S. Ghassemzadeh, and V. Tarokh, "Information theoretic analysis of cognitive radio systems," in *Cognitive Wireless Communication Networks*, V. Bhargava and E. Hossain, Eds. Springer, 2007.

[7] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," in *39th Annual Conf. on Information Sciences and Systems (CISS)*, Mar. 2005.

[8] ——, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[9] N. Devroye, "Information theoretic limits of cognition and cooperation in wireless networks," Ph.D. dissertation, Harvard University, 2007.

[10] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of control and information theory*, 1980.

[11] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, 2009.

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

[13] R. D. Y. I. Maric and G. Kramer, "The strong interference channel with unidirectional cooperation," *The Information Theory and Applications (ITA) Inaugural Workshop*, Feb 2006, uCSD La Jolla, CA,.

[14] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 54, no. 10, pp. 4707–4712, Oct. 2008.

[15] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2413–2417, July 2006.

[16] G. Kramer, *Topics in Multi-User Information Theory*, ser. Foundations and Trends in Communications and Information Theory. Vol. 4: No 45, pp 265-444, 2008.

[17] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai, and S. Verdú, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. on Inf. Theory*, vol. 55, no. 2, pp. 604–619, Feb. 2009.

[18] N. Liu, I. Maric, A. Goldsmith, and S. Shamai, "The capacity region of the cognitive z-interference channel with one noiseless component," *http://www.scientificcommons.org/38908274*, 2008. [Online]. Available: http://arxiv.org/abs/0812.0617

[19] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with a cognitive transmitter," *European Transactions on Telecommunications*, vol. 19, pp. 405–420, Apr. 2008.

[20] I. Maric, R. Yates, and G. Kramer, "The capacity region of the strong interference channel with common information," in *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, 2005, pp. 1737–1741.

[21] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *Information Theory, IEEE Transactions on*, vol. 25, no. 3, pp. 306–311, May 1979.

[22] S. Rini, D. Tuninetti, and N. Devroye, "The capacity region of gaussian cognitive radio channels at high snr," *Proc. IEEE ITW Taormina, Italy*, vol. Oct., 2009.

[23] S. Rini, D. Tuninetti, and N. Devroye, "State of the cognitive interference channel: a new unified inner bound," long version of paper in *Proc. IZS*, Mar., 2010, available at `http://www.ece.uic.edu/~devroye`.

[24] S. Rini, "On the role of cognition and cooperation in wireless networks: an information theoretic perspective - a preliminary thesis," http://sites.google.com/site/rinistefano/my-thesis-proposal.

[25] S. Rini, D. Tuninetti, and N. Devroye, "The capacity region of gaussian cognitive radio channels to within 1.87 bits," *Proc. IEEE ITW Cairo, Egypt*, 2010, http://www.ece.uic.edu/~devroye/conferences.html.

[26] S. H. Seyedmehdi, Y. Xin, J. Jiang, and X. Wang, "An improved achievable rate region for the causal cognitive radio," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2009.

[27] O. Simeone and A. Yener, "The cognitive multiple access wire-tap channel," in *Proc. Conf. on Information Sciences and Systems (CISS)*, Mar. 2009.

[28] S. Sridharan and S. Vishwanath, "On the capacity of a class of mimo cognitive radios," in *Information Theory Workshop, 2007. ITW '07. IEEE*, Sept. 2007, pp. 384–389.

[29] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.

Oblivious Relaying for Primitive Interference Relay Channels

O. Simeone
CWCSPR, ECE Dept.,
NJIT, Newark, NJ, USA

E. Erkip
Dept. of ECE, Polytechnic Inst. of NYU
Brooklyn, NY, USA

S. Shamai (Shitz)
Dept. of EE, Technion
Haifa, Israel

***Abstract*—[1]Consider a relay node that needs to operate without knowledge of the codebooks (i.e., modulation, coding) employed by the assisted source-destination pairs. This paper studies the performance of relaying under this condition, termed oblivious relaying, for the primitive relay channel (PRC) and the primitive interference relay channel (PIRC). "Primitive" refers to the fact that the relay-to-destinations links use orthogonal resources with respect to the other links. Assuming discrete memoryless models, the capacity of a PRC with oblivious relaying is derived, along with the capacity region of the PIRC with oblivious relaying and interference-oblivious decoding (i.e., each decoder is unaware of the codebook used by the interfering transmitter). In all cases, capacity is achieved by Compress-and-Forward (CF) with time-sharing. Performance without time-sharing is discussed as well. Finally, it is shown that for the general (non-oblivious) Gaussian PRC, the achievable rate by CF (with Gaussian inputs and test channels and no time-sharing) is suboptimal by at most half bit with respect to the cut-set bound.**

I. Introduction

A standard, and often implicit, assumption in network-information theoretic analyses is that design of encoding/ decoding functions at all nodes is performed jointly in order to optimize the system performance. This implies, in particular, that all nodes must be aware at all times of the operations carried out by any other node. Moreover, in general, addition of a new node, or even only change of operation at one node, calls for a re-design of the entire network. While this may be reasonable in centrally controlled networks such as conventional[2] cellular system, it becomes impractical in decentralized scenarios. In fact, in the latter cases, nodes operate without extensive signalling capabilities, so that full coordination in the choice of encoding/ decoding functions is typically a prohibitive task.

In this work, we investigate design of basic network building blocks, under the assumption that information about the operations carried out at the source encoders (i.e., of the sources' *codebooks*) is not available throughout the network. We emphasize that this may be due to practical constraints, as discussed above, or simply to the need for simple network protocols that do not require continuous reconfiguration (and thus extensive signaling). The analysis is based on the framework of *oblivious processing* first proposed in [1]. We focus on the "primitive" relay channel (PRC, see review in [3]) and on an extension of the PRC to a setting with two source-destination pairs, that we define primitive interference relay channel (PIRC), see Fig. 1. We establish a number of capacity results under the assumption of oblivious processing and the relay and, possibly, at the interfered destinations.

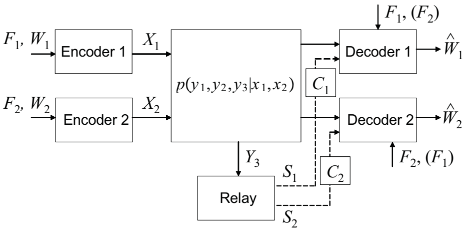


Fig. 1. A Primitive Inteference Relay Channel (PIRC) with oblivious relaying and interference-oblivious or interference-aware decoders.

II. System Model

We study the PRC and PIRC with oblivious processing as depicted in Fig. 2 and 1, respectively. We use the term "primitive" as in [3] to mean that the relay is connected to the destination(s) via finite-capacity orthogonal links. This corresponds to assuming that the relay transmissions occupy a different resource with respect to the other links in the system. As detailed below, oblivious processing, following [1], refers to coding/ decoding operations designed without the knowledge of some of the codebooks in the system.

A discrete memoryless PIRC consists of two source-destination pairs (indexed by subscripts 1 and 2) and is defined by a tuple $(\mathcal{X}_1, \mathcal{X}_2, p(y_1, y_2, y_3|x_1, x_2), \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, C_1, C_2)$ where C_1, C_2 denote the capacities (bits/ channel use) of the links from relay to destination 1 and 2, respectively. Subscript 3 is used for the relay. A special case of the PIRC is the PRC [3], where there is only one source-destination pair, i.e., we set $\mathcal{X}_2 = \mathcal{Y}_2 = \emptyset$. In this case, we drop the subscript 1 for simplicity so that the PRC is defined as $(\mathcal{X}, p(y, y_3|x), \mathcal{Y}, \mathcal{Y}_3, C)$, see Fig. 2. We will also consider a Gaussian model with power constraints, to be introduced below.

[2] I.e., without advanced capabilities such as relaying or network MIMO.

[1] The work of O. Simeone was supported by U.S. NSF grant CCF-0905446. The work of E. Erkip was supported by U.S. NSF grant CCF-0914899. The work of S. Shamai was supported by the European Commision in the framework of the FP7 Network of Excellence in Wireless COMmunications, NEWCOM++ and by the CORNET consortium.

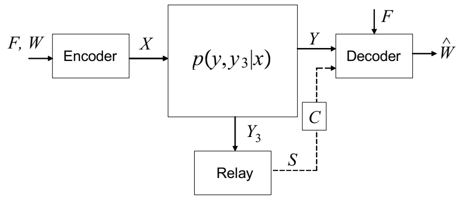


Fig. 2. A Primitive Relay Channel (PRC) with oblivious relaying.

A. Oblivious Processing

In the following, we detail on the considered model of PIRC with oblivious processing. The corresponding model for the PRC with oblivious processing is a special case that follows immediately and will not be detailed explicitly. In the considered model, each source-destination pair agrees on the codebook to be used for communications (i.e., the destination knows the codebook used by the corresponding transmitter), as in regular interference channels. However, we assume that the information about the codebooks may be lacking at the relay (*oblivious relaying*) and possibly at the interfered destination (*oblivious decoding*).

To account for oblivious processing, we follow the model of [1], which we first describe informally in the following. Fix rates R_j [bits/ channel use], $j = 1, 2$, used for transmission between the jth source-destination nodes. According to [1], we assume that the currently employed codebook (say by pair $j = 1, 2$) is identified by an index $F_j \in [1, |\mathcal{X}|^{n2^{nR_j}}]$, which ranges over the set $[1, |\mathcal{X}|^{n2^{nR_j}}]$ of all possible codebooks of rate R_j. Therefore, transmitter j sends a message $W_j \in [1, 2^{nR_j}]$ by transmitting a codeword $x_j^n(F_j, W_j)$ dependent on both message W_j and index F_j. Knowledge of F_j implies awareness of the codebook used by the jth source-destination pair[2]. Moreover, in the absence of knowledge of F_j, it is assumed that the codeword transmitted by the jth source completely lacks any structure, and thus its letters "look" independent identically distributed (i.i.d.) with respect to a given single-letter distribution $p_{X_j}(\cdot)$ over $\mathcal{X}_j$, $j = 1, 2$. Rigorous definitions are given below, highlighting also the role of time-sharing.

B. Formal Setting

Formal definitions are as follows.

Definition 1: A (n, R_1, R_2) code for the PIRC with *oblivious processing* is given by:

a. Message sets $[1, 2^{nR_j}]$ and codebook sets $[1, |\mathcal{X}_j|^{n2^{nR_j}}]$, $j = 1, 2$;

b. Encoding functions: For each user j, the *encoder* is defined by a pair (p_{X_j}, ϕ_j), where p_{X_j} is a single-letter pmf and $\phi^{(j)}$ is a mapping ϕ_j: $[1, |\mathcal{X}_j|^{n2^{nR_j}}] \times [1, 2^{nR_j}] \rightarrow \mathcal{X}_j^n$, that provides the transmitted codeword $x_j^n = \phi_j(F_j, W_j)$ given codebook index F_j and message W_j. The pmf p_{X_j} defines the probability $p_F(f)$ of choosing a certain codebook $F \in [1, |\mathcal{X}_j|^{n2^{nR_j}}]$ as

$$p_{F_j}(f) = \prod_{w\in[1,2^{nR_j}]} p_{X^n}(\phi_j(f, w)), \qquad (1)$$

where $p_{X^n}(x^n) = \prod_{i=1}^n p_X(x_i)$;

c. Relaying function: The *relay*, unaware of the codebooks F_j with $j = 1, 2$ maps the received sequence $y_3^n \in \mathcal{Y}^n$ into two indices $s_j \in [1, 2^{nC_j}]$ to be sent to destinations $j = 1, 2$ as $\phi_3 : \mathcal{Y}_3^n \rightarrow [1, 2^{nC_1}] \times [1, 2^{nC_2}]$, so that $[s_1, s_2] = \phi_3(y_3^n)$.

d. Decoding functions: For interference-aware decoding, decoding is described by a mapping

$$g_j : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times [1, |\mathcal{X}_2|^{n2^{nR_2}}] \times \mathcal{Y}_j^n \rightarrow [1, 2^{nR_j}]$$

from the two codebook indices F_1, F_2 and received signal y_j^n to the decoded message $\hat{W}_j = g_j(f_1, f_2, y_j^n)$; Instead, for interference-oblivious decoding we have

$$g_j : [1, |\mathcal{X}_1|^{n2^{nR_j}}] \times \mathcal{Y}_j^n \rightarrow [1, 2^{nR_j}],$$

so that the decoded message $\hat{W}_j = g_j(f_j, y_j^n)$ depends only on the received signal and index of the codebook of the corresponding transmitter alone (not of the interferer).

We say that we have: (i) *Oblivious relaying*: The relay is not aware of both indices F_1 and F_2; (*ii*) *Interference-oblivious decoding*: Destination j only knows index F_j and not F_i, $i \neq j$.

Definition 2: A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of codes such that $\Pr[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)] \rightarrow 0$, where the probability is taken with respect to a uniform distribution of messages W_1 and W_2 and with respect to independent indices F_1 and F_2 whose joint distribution is given by the product of (1) for $j = 1, 2$. The capacity region $\mathcal{C}$ is the closure of the union of all achievable rates.

Remark 1: The definition of oblivious processing obtained from (1), which is the same as in [1], rules out general multiletter input distributions, thus limiting the space of feasible coding schemes, but does not exclude standard "single-letter" coding schemes such as superposition coding and rate-splitting strategies. Moreover, the definition does not allow time-sharing. In fact, in case the transmitters time-share among different codewords, the condition (1) is not satisfied for a given time-sharing schedule. The following alternative definition of oblivious processing instead enables time-sharing.

Definition 3: *Oblivious processing (relaying or decoding) with enabled time-sharing* refers to codes defined as in Definition 1 with the difference that encoders, relay and decoders are all aware of a time-sharing sequence $q^n \in \mathcal{Q}^n$, defined over a finite alphabet $\mathcal{Q}$. Encoding and decoding functions $\{\phi_j, g_j\}$ defined above are modified to depend on q^n. Moreover, codebook generation is constrained so that

$$p_{F_j}(f|q^n) = \prod_{w\in[1,2^{nR_j}]} p_{X^n|Q^n}(\phi_j(f, w)|q^n), \qquad (2)$$

where $p_{X^n|Q^n}(x^n|q^n) = \prod_{i=1}^n p_{X|Q}(x_i|q_i)$ for a conditional pmf $p_{X|Q}(x_i|q_i)$, instead of (1).

[2] This can be seen as a form of randomized encoding.

Remark 3: Depending on the application, it may be feasible or not for the relay to acquire the time-sharing sequence q^n decided by sources and destinations. Notice that acquiring the time-sharing sequence is in any case much less demanding that obtaining the full codebook information. If it is possible to acquire q^n, then the definition (2) is appropriate, otherwise the original definition (1) should be adopted.

As a result of the constraints assumed on the coding function, we have the following facts.

Lemma 1 [1]: Given an oblivious processing code for the PIRC, the distribution of a transmitted codeword of source j is given by $p_{X_j^n}(x^n) = \prod_{i=1}^n p_{X_j}(x_i)$. In other words, in the absence of information regarding the index F_j and the message W_j, a codeword $x_j^n(F_j, W_j)$ taken from a (n, R_1, R_2) codebook is i.i.d. As a consequence, the received signals at destinations and relay are also i.i.d. vectors.

Lemma 2: Given an oblivious codebook code for the PIRC with enabled time-sharing, the distribution of a transmitted codeword of source j, conditioned on the time-sharing sequence is given by $p_{X^n|Q^n}(x^n|q^n) = \prod_{i=1}^n p_{X|Q}(x_i|q_i)$. In other words, in the absence of information regarding the index F_j and the message W_j, a codeword $x_j^n(F_j, W_j)$ taken from a (n, R_1, R_2) codebook has independent, but non-indentically distributed, entries.

Remark 4: While the unconditional pmf $p_{X_j^n}(x^n)$, or $p_{X^n|Q^n}(x^n|q^n)$, factorizes as discussed above, the conditional pmf $p_{X_j^n|F}(x^n|f)$, or $p_{X^n|Q^n,F}(x^n|q^n, f)$, given the key $F_j = f$ does not. In other words, as shown in [2], given a specific "good" code, the empirical distribution with respect to the choice of the message W_j can never be i.i.d. (except for extreme cases such as noiseless channels).

III. Primitive Relay Channel with Oblivious Relaying

We start by analyzing the PRC with oblivious relaying.

Proposition 1: The capacity of a primitive relay channel with oblivious relaying and enabled time-sharing is given by

$$\begin{aligned} \mathcal{C} &= \max I(X; Y\hat{Y}_3|Q) && \text{(3a)} \\ \text{s.t. } C &\geq I(Y_3; \hat{Y}_3|YQ) && \text{(3b)} \end{aligned}$$

where maximization is taken with respect to the distribution $p(q)p(x|q)p(\hat{y}_3|y_3, q)$ and the mutual informations are evaluated with respect to

$$p(q)p(x|q)p(\hat{y}_3|y_3, q)p(y, y_3|x). \quad (4)$$

If time-sharing is not allowed, (3) is still an upper bound on the capacity, and the following rate is achievable (i.e., Q =const)

$$\begin{aligned} \mathcal{C} &= \max I(X; Y\hat{Y}_3) && \text{(5a)} \\ \text{s.t. } C &\geq I(Y_3; \hat{Y}_3|Y) && \text{(5b)} \end{aligned}$$

Proof: See Appendix A.

Remark 5: Capacity is attained by Compress-and-Forward (CF) with time sharing. This may not be surprising, given that the relay is incapable by design of decoding the codeword transmitted by the source. However, notice that in the setting of [1] where multiple relays are present but no direct link between source and destination is in place, optimality of (distributed) CF strategies remains elusive. This is in accordance with the current state of the art on the corresponding *source coding* problems, where the source (rather than being an encoded sequence) is a given i.i.d. process to be reconstructed at the destination. In fact, the source coding counterpart of [1] is the (discrete memoryless) CEO problem, which is still generally unsolved [8], while the source coding counterpart of the PRC is the Wyner-Ziv scenario of source coding with side information, whose solution is well-known (see, e.g., [4]). For a discussion on other scenarios where CF was shown to be optimal, we refer to [6].

Remark 6: In (3), variable Q allows time sharing. The fact that the performance of CF can be generally improved by time-sharing was shown in [5, Theorem 2]. In case, time-sharing is not allowed, rate (5) is achievable, which is generally smaller than (3).

A. Gaussian Model

Here we turn to the memoryless Gaussian PRC, that is defined as

$$Y_{3i} = \sqrt{\alpha}X_i + N_{3i}, \; Y_i = X_i + N_i, \quad \text{(6a)}$$

where N_{3i}, N_i are independent zero-mean unit-power, and the power constraint is given by $1/n \sum_{i=1}^n E[X_i^2] \leq P$. The result of Proposition 1 can be extended using standard arguments to continuous channels and thus to the Gaussian channel (6). However, optimization of the input distribution $p(q)p(x|q)p(\hat{y}_3|y_3, q)$ in (3) remains an open problem. Achievable rates using Gaussian input distribution $p(x|q)$ and quantization test channel $p(\hat{y}_3|y_3, q)$ in (3) can be found in [7] and [5, Theorem 2] without and with time-sharing random variable Q, respectively. However, as discussed in [1], a Gaussian input distribution is generally not optimal and, as seen in [7], non-Gaussian test channels may be advantageous, especially with a non-Gaussian input distribution. Nevertheless, the next proposition shows that the suboptimality of Gaussian channel inputs, Gaussian test channel and no time-sharing, is at most half bit (per (real) channel use), *even if one allows non-oblivious relaying.*

Proposition 2: The rate achievable via CF (and hence oblivious relaying)

$$R_{CF} = \frac{1}{2}\log_2\left(1 + P + \frac{\alpha P}{1 + \frac{1+P+\alpha P}{(2^{2C}-1)(P+1)}}\right) \quad (7)$$

on the Gaussian PRC (6), by employing Gaussian channel inputs, Gaussian test channel and no time-sharing, is at most half bit away from the capacity of the PRC with codebook-aware (and thus also oblivious) relaying.

Proof: The proof is obtained by comparing the achievable rate (7) (that can be found in, e.g., [7]) with the cut-set bound upper bound (which holds even with non-oblivious relaying)

$$R_{UB} = \min\left\{\frac{1}{2}\log_2\left(1+P\right)+C,\ \frac{1}{2}\log_2\left(1+\alpha P+P\right)\right\}. \tag{8}$$

See full derivation in Appendix B.

IV. Primitive Interference Relay Channel with Oblivious Relaying

We not turn to the analysis of the PIRC with oblivious relaying. The following proposition shows that in the presence of interference-oblivious decoding, it is optimal for the relay to employ CF and for the destinations to treat the interfering signal as noise.

Proposition 3: The capacity region of the PIRC with oblivious relaying, interference-oblivious decoding and enabled time-sharing is given by the set of all non-negative pairs (R_1, R_2) that satisfy

$$R_j \leq I(X_j; Y_j\hat{Y}_3^{(j)}|Q),\ \text{for } j=1,2, \tag{9}$$

for some distribution $p(q)\prod_{j=1}^{2}p(x_j|q)p(\hat{y}_3^{(j)}|y_3,q)$ $p(y_1,y_2|x_1,x_2)$ that satisfy

$$C_j \geq I(Y_3;\hat{Y}_3^{(j)}|Y_jQ) \text{ for } j=1,2. \tag{10}$$

If time-sharing is not enabled, the above is an outer bound to the capacity region and setting Q =const leads to an achievable rate region.

Proof: Follows similarly to the proof of Proposition 1.

V. Appendix

A. Appendix-A: Proof of Proposition 1

Achievability follows by CF with Wyner-Ziv coding and time-sharing determined by variable Q (see, e.g., [3] [7] and [5]). For the converse, consider the first the variable S transmitted by the relay to the destination over the finite-capacity link. Denote as $\tilde{Q}$ the vector of time-sharing variables q^n in Definition 2

$$\begin{aligned}
nC &\geq H(S) \geq H(S|\tilde{Q}) \\
&\geq I(S;X^nY_3^n|Y^n\tilde{Q}) \geq \sum_{i=1}^{n} I(S;Y_{3i}|Y^n\tilde{Q}Y_3^{i-1}X^{i-1}) \\
&= \sum_{i=1}^{n} H(Y_{3,i}|Y_i\tilde{Q}) - H(Y_{3,i}|\hat{Y}_{3i}Y_i\tilde{Q}) \\
&= \sum_{i=1}^{n} I(Y_{3i};\hat{Y}_{3i}|Y_i\tilde{Q}),
\end{aligned}$$

where in the third line we used the fact that Y_3^n, Y^n, X^n have conditionally independent entries given $\tilde{Q}$, due to Lemma 2, and we defined $\hat{Y}_{3i}{=}\,[SX^{i-1}Y_3^{i-1}Y^{i-1}Y_{i+1}^n]$. Notice that the following Markov chain $(Y_i, X_i) - (Y_{3i}, \tilde{Q}) - \hat{Y}_{3i}$ holds. Now, introducing a variable Q', independent of all other variables and uniformly distributed in $[1, n]$, defining $Y_3{=}\,Y_{3Q'}$ and similarly for the other variables, and $Q = [\tilde{Q}\ Q']$, we get the constraint (3b). Notice that with these definitions we have the Markov chain $(Y, X) - (Y_3, Q) - \hat{Y}_3$. Turning to the destination, using Fano inequality $H(W|Y^nSF\tilde{Q}) \leq n\epsilon_n$ with $\epsilon_n \rightarrow 0$ for $n \rightarrow \infty$ (for vanishing probability of error), we obtain

$$\begin{aligned}
nR &\leq I(W;Y^nSF|\tilde{Q}) + n\epsilon_n \\
&= H(Y^nS|\tilde{Q}) + H(F|Y^nS\tilde{Q}) \\
&\quad -H(F|W\tilde{Q}) - H(Y^nS|FW\tilde{Q}) + n\epsilon_n \\
&= I(FW;Y^nS|\tilde{Q}) - I(F;Y^nS|\tilde{Q}) + n\epsilon_n \\
&\leq I(X^n;Y^nS|\tilde{Q}) + n\epsilon_n \\
&= \sum_{i=1}^{n} H(X_i|\tilde{Q}) - H(X_i|Y_i\hat{Y}_{3i}\tilde{Q}) \\
&= \sum_{i=1}^{n} I(X_i;Y_i\hat{Y}_{3i}|\tilde{Q}) + n\epsilon_n,
\end{aligned}$$

where in the third equality we have used the fact that F and W are independent and in the last line we have used Lemma 2.

B. Appendix-B: Proof of Proposition 2

We first rewrite (7) as $R_{CF}{=}\frac{1}{2}\log_2\left(\frac{2^{2C}(1+P)(1+\alpha P+P)}{2^{2C}(1+P)+\alpha P}\right)$, which can be proved by standard algebraic manipulations. Now, assume first that $1+P(1+\alpha) \leq 2^{2C}(1+P)$ so that the upper bound (8) reads $R_{UB}{=}\frac{1}{2}\log_2\left(1+\alpha P+P\right)$. Under this condition, the achievable rate R_{CF} satisfies

$$\begin{aligned}
R_{CF} &= R_{UB} - \frac{1}{2}\log_2\left(1+\frac{\alpha P}{2^{2C}(1+P)}\right) \\
&\geq R_{UB} - \frac{1}{2}\log_2\left(1+\frac{(2^{2C}-1)(1+P)}{2^{2C}(1+P)}\right) \\
&\geq R_{UB} - \frac{1}{2},
\end{aligned}$$

where the second inequality follows from the assumed condition. The same inequality is proved in a similar way under the complementary condition $1+P(1+\alpha) \geq 2^{2C}(1+P)$.

References

[1] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, “Communication via decentralized processing,” *IEEE Trans. Inform. Theory,* vol. 54, no. 7, pp. 3008-3023, July 2008.

[2] S. Shamai (Shitz) and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 836-846, May 1997.

[3] Y.-H. Kim, “Coding techniques for primitive relay channels,” in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sept. 2007.

[4] G. Kramer, *Topics in multi-user information theory*, Foundations and Trends in Communications and Information Theory, vol. 4, no. 4-5, pp. 265-444, 2007.

[5] A. El Gamal, M. Mohseni and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1545-1561, Apr. 2006.

[6] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," submitted [arXiv:0808.0948].

[7] R. Dabora and S.D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4409-4431, Oct. 2008.

[8] V. Prabhakaran, D. Tse and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2004), pp. 119, July 2004.

[9] A. Carleial, "Interference channels," *IEEE Trans. Inform. Theory*, vol. 24, no. 1, pp. 60-70, Jan. 1978.

Are Yesterday's Fading Models and Performance Metrics Adequate for Today's Wireless Systems?

Angel Lozano
UPF
08018 Barcelona, Spain
Email: angel.lozano@upf.edu

Nihar Jindal
University of Minnesota
Minneapolis, MN 55455, USA
Email: nihar@umn.edu

Abstract—**This paper examines some of the settings commonly used to represent fading. We raise the question of whether these settings remain meaningful in light of the advances that wireless communication systems have undergone over the last decade. A number of weaknesses are pointed out, and ideas on possible fixes are put forth. Some of the identified weaknesses have to do with models that, over time, have become grossly inadequate; other weaknesses have to do with changes in the operating conditions of modern systems, and others with the coarse and asymptotic nature of some of the most popular performance metrics.**

I. Motivation

Fading is an essential attribute of wireless channels and, as such, the characterization of its impact on fundamental communication limits has been the object of much scrutiny [1]. A few canonical settings have become established over time that offer a compromise between realism and tractability. These settings have served the information theory and communications research communities extremely well for years.

II. Canonical Settings

The marginal modeling is not particularly problematic: application of the central limit theorem to the sum of a large number of multipath components yields a Rayleigh distribution for the fading amplitude, and experimental measurements have repeatedly confirmed the validity of this distribution. It is the modeling of the selectivity over each codeword that presents the most complications, and the two most common canonical settings idealize it in limiting senses:

- Ergodic setting: the fading varies ergodically over the span of each codeword. This setting has a well-defined capacity in the Shannon sense, which entails an expectation over the fading distribution. While analytically convenient, this setting is frequently dismissed as inadequate on the grounds of the latency constraints of many applications.
- Quasi-static setting: the fading is fixed over each codeword, and varies only from codeword to codeword. In Rayleigh fading, the Shannon capacity of this setting is strictly zero. The relevant metric is then the outage probability, i.e., the probability that a given transmission rate is not supported [2]. The quasi-static setting is often regarded as more relevant than its ergodic counterpart to modern systems.

III. A Contemporary Perspective

The canonical settings have been in use for years and are by now deeply ingrained. Wireless systems, however, have evolved greatly since the time when these settings were defined. They have made link adaptation a norm, grown wideband, adopted packet switching and scheduling, and embraced ARQ and H-ARQ (hybrid ARQ), among other advances.

A. Link Adaptation

In contrast with older designs, modern wireless systems exhibit a very high degree of adaptivity. The transmission rate, in particular, is matched to the fading whenever timely CSI (channel-state information) can be had at transmitter, i.e., in slow fading. This fundamentally changes the nature of the communication problem: outages are essentially eliminated.

B. Hybrid ARQ

Another trait that is central to the adaptive nature of modern systems is H-ARQ, whereby the codeword length itself is made adaptive. The combination of link adaptation and H-ARQ allows for a finely tuned match between the rate and the channel in slow fading.

C. Wideband Signaling

While older wireless systems were organized into narrowband channels, modern systems are wideband. Signals occupy many MHz of bandwidth, which has two immediate consequences. First, it renders frequency selectivity a property that cannot be ignored by the models. And second, it allows for long codewords without long latency.

D. Operating Point

Yet another fundamental facet of modern wireless systems is that the physical layer operates at some fixed packet error probability, which depends on the specific parameters of each system but is typically around 1% at H-ARQ termination. The aim of reliable communication is not given up, but the physical layer is no longer alone in the task of ensuring it.

References

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[2] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

Hopfield Neural Networks for Vector Precoding

Vesna Gardašević, Ralf R. Müller and Geir E. Øien
Department of Electronics and Telecommunications
The Norwegian University of Science and Technology
7491 Trondheim, Norway
Email: {vesna.gardasevic, mueller, oien}@iet.ntnu.no

***Abstract*—We investigate the application of Hopfield neural networks (HNN) for vector precoding in wireless multiple-input multiple-output (MIMO) systems. We apply the HNN to vector precoding with N transmit and K receive antennas, and obtain simulation results for the average transmit energy optimization as a function of the system load $\alpha = K/N$. We compare these results with lattice search based precoding performances, and show that the proposed method for nonlinear vector precoding with complexity $O(K^3)$ achieves competitive performances in the range $0 < \alpha \leq 0.9$ in comparison to lattice search based precoders. The proposed method is of a polynomial complexity and therefore, it is an attractive suboptimal approach for vector precoding.**

I. Introduction

In a broadcast MIMO system a transmitter, typically a base station, communicates with a number of receivers. We assume that the receivers cannot cooperate with each other, and that the channel state information is known at the transmitter side. In this scenario, our aim is to delegate most of the signal processing work to the transmitter side. The signal processing at the transmitter includes precoding techniques [1] for predistortion of the transmitted symbols. In this way the transmit energy is reduced and signal detection at the receivers is simplified.

The capacity region of Gaussian broadcast MIMO channels [2] can be achieved by dirty paper coding (DPC) [3]. Since DPC has high complexity demands for implementation, research in the area of precoding techniques has been focused on different linear and nonlinear sub-optimal methods. In linear precoding (e.g. [4]) the transmitted symbols are premultiplied by the pseudo-inverse $\mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^\dagger)^{-1}$ of the channel matrix $\mathbf{H}$ and the receiver applies simple symbol-by-symbol detection. This method is known as zero-forcing precoding (ZF) (e.g. [5]) and the main advantage of this method is its low complexity.

A drawback occurs when the channel matrix has small singular values, such that an inversion operation causes severe transmit power amplification. One method proposed to control power amplification due to ill-conditioned channel matrices is channel regularization [6]. However, this approach does not cancel all the interference at the receiver.

A nonlinear precoding approach employs nonlinear predistortion of the transmitted symbols before the linear operation. In a nonlinear predistortion step the alphabet of transmitted symbols is increased to a larger redundant set, such that the symbols to be sent are subject to optimization. A vector pertubation method [7], for example, modifies the idea of THP. The vector-space search for the closest point in the lattice set that minimizes the energy is performed by an exhaustive search [7]. Often search is performed by a sphere encoder (SE), which is known to reduce the complexity, but still keeping it exponential [8].

Lattice-basis reduction algorithms have been proposed for further reduction of the SE complexity. Lower complexity can be achieved by searching for the approximately closest point in the lattice. There are implementations of vector precoding algorithms based on lattice-basis reduction [9], for example the LLL (Lenstra, Lenstra, and Lovasz) algorithm [10], that provide good performance.

Another approach to nonlinear predistortion is to apply a convex relaxation of the input symbol alphabet. A novel vector precoding method that applies a convex relaxed symbol alphabet instead of a discrete set is presented in [11]. In [12] it is shown that with a convex precoding approach, spectral efficiency can be higher then with lattice precoding at low to moderate signal-to-noise ratios.

The field of artificial neural networks (ANN) has been an active research area with periods of both dynamic and stagnation phases, from the early 1940s. Research has resulted in a great number of publications e.g. [13], and ANNs have been applied to the realization of, for example: content-addressable memory, pattern classifiers, pattern recognition, vector optimization, multiuser detection. ANNs have also been applied for solving optimization problems in different areas.

In this paper we will apply the Hopfield Neural Network (HNN) [13] which belongs to the class of recurrent ANNs as the algorithm for optimization in vector precoding. The structure of extensive parallelism makes the computational capabilities of the HNN very powerful and attractive. We will provide numerical simulation results for loads within $0 < \alpha \leq 1$, for $K = 8$, $K = 16$, $K = 27$ and $K = 64$, and compare the results with the performances of the SE lattice precoding, where the number of redundant representations of each information bit is $L = 2$ and the number of receive antennas are $K = 27$ and $K = 64$. We also compare the HNN precoder performances with the analytical solution for the SE vector precoding [12] and to the performance of convex relaxation (CR) [11] of the input symbol alphabet.

Our simulation results show that for loads $\alpha \leq 0.7$, the performance of the HNN vector precoding is very close to the SE performance. Up to $\alpha \approx 0.9$ there is a controlled increase in the transmit energy of the HNN algorithm compared with the SE vector precoding performances. For loads $\alpha \approx 0.9$ and greater there is a substantial increase in the transmit energy of the HNN solution. Comparing the HNN performances with the analytical result for CR we show that up to $\alpha \approx 0.9$ the HNN precoding outperforms CR, which for loads $0.9 \leq \alpha < 1$ the energies provided by the HNN start to increase considerably.

It is known that the expected complexity of the SE algorithm is exponential [8]. Therefore, due to the fast computational capabilities and robustness of the HNN, our approach is an efficient suboptimal way for vector precoding within a wide load range at low complexity.

II. SYSTEM MODEL

A narrowband multi-user MIMO system is modeled by

$$\mathbf{r} = \mathbf{H}\mathbf{t} + \mathbf{n} \tag{1}$$

where $\mathbf{t}= [t_1, t_2, \cdots, t_N]^T$ and $\mathbf{r}= [r_1, r_2, \cdots, r_K]^T$ denote the transmit and receive signal vectors, respectively, the $K \times N$ channel matrix $\mathbf{H}$ is assumed to have independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance, and $\mathbf{n}$ is the white Gaussian noise. We consider (1) as a MIMO system with a single transmitter with N antennas, and K receivers, each with a single antenna, that cannot cooperate with each other.

Now, assume that the number of transmit antennas is greater or equal to the number of receive antennas ($K \leq N$). The $K \times 1$ data symbol vector is denoted by s, and for binary phase-shift keying (BPSK), the elements of the vector $\mathrm{s} = [s_1, s_2, \cdots, s_K]^T$ belong to the set $\mathbf{S} = \{-1, +1\}$. Let the union of the sets $\mathcal{B}_{-1} = \{-1, +3\}$ and $\mathcal{B}_{+1} = \{+1, -3\}$ be the relaxed alphabet. In a nonlinear predistortion step the data vector s is mapped onto a vector $\mathbf{x} = [x_1, x_2, \cdots, x_K]^T$, with the vector elements chosen to minimize the transmit energy, where $\mathrm{x}_k \in \mathcal{B}_{s_k}$, for $k = 1, 2, \cdots, K$. The following linear predistortion matrix $\mathbf{T}$ is assumed to be

$$\mathbf{T} = \mathbf{H}^{+} = \mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger})^{-\mathbf{1}} \tag{2}$$

The optimization problem can now be formulated as follows:

$$\mathbf{x}^{*} = \min_{\mathbf{x} \in \mathcal{B}_{s_1} \times \cdots \times \mathcal{B}_{s_K}} \|\mathbf{T}\mathbf{x}\|^2 = \min_{\mathbf{x} \in \mathcal{B}_{s_1} \times \cdots \times \mathcal{B}_{s_K}} \mathbf{x}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger})^{-\mathbf{1}}\mathbf{x} \tag{3}$$

This problem is difficult to solve since it is a nonconvex optimization problem in a high dimensional space, and we therefore investigate the application of the HNN for solving (3).

III. QUADRATIC OPTIMIZATION USING HOPFIELD NEURAL NETWORKS

Hopfield [13] proposed the application of ANNs for solving combinatorial optimization problems. A review of the HNN applications for solving mathematical programming problems is given in [14]. The optimization using the HNN is performed by constructing the energy function with the parameters that depend on a practical optimization problem. Hopfield [13] showed that the energy function constructed for an observed problem provides convergence of the system to stable states if the matrix in the objective function is symmetric, with zero diagonal elements. The main drawback of the HNN is that its computational properties do not necessarily provide the best solution by minimizing the appropriate energy function, but the optimization can result in a local minimum. Various modifications of the HNN, for example the combination with stochastic algorithms [15] have been proposed for avoiding said local minima.

The HNN works as follows: the sum of an external threshold value θ_i and the weighted sum of input states are transformed by a nonlinear function called an activation function to become the output of the system. The HNN supports different activation functions, for example: hard limiter (threshold) transfer function, hyperbolic tangent (tanh), sigmoid and other functions. A weight matrix $\mathbf{W}$ can be used to model various effects in an observed system, and depends on the particular problem that is solved by the HNN.

We apply the HNN model with an activation function $f(\cdot)$ described by

$$v_j^{(l+1)} = f\left(\sum_{i=1}^{K} w_{ji} v_i^{(l)} + \theta_j\right) \tag{4}$$

where $l = 0, 1, 2 \cdots$, denotes the number of iterations run by the HNN, $i = 1, 2 \ldots, K$, $j = 1, 2 \ldots, K$, the network states are denoted by v_j, respectively, w_{ji} are the assigned weights between neurons j and i, and θ_j is the external input signal.

The HNN described by (4) minimizes the energy function

$$E = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} w_{ij} v_i v_j + \sum_{i=1}^{K} v_i \theta_i \tag{5}$$

This HNN model can also be considered as an iterative algorithm that performs soft parallel interference cancellation [16], [17]. In our model the solution of the optimization problem defined in (3) corresponds to the minimization of the energy function in (5). The output vector $\mathbf{v}$ of the HNN in (4) corresponds to the vector $\mathbf{x}$ in (3), while the coefficients w_{ij} in (4) correspond to the entries of the channel matrix $\mathbf{H}$.

We assume that the start time of the iterations is $l = 1$, and that a soft decision $v_i^{(l)}$ is calculated in each step. This value is subtracted from the soft decision from the previous iteration and the hypothetical interference is cancelled in each iteration. The iterations can be performed until there is only a minor change between the soft decisions in successive iterations, i.e. until $\max_i |v_i^{(l)} - v_i^{(l-1)}| < \delta$ where δ is a sufficiently small value or the number of iterations exceeds the maximum number of iterations. In our simulations we have chosen the criterion that iterations are performed until the number of iterations exceeds the maximum number of iterations, denoted by $I_{\max}$.

It has been shown that the expected complexity [7] of the SE algorithm depends on the number of dimensions K and the signal-to-noise ratio (SNR). In [8] the expected complexity of the SE algorithm, as well as the asymptotic expression for the complexity, were derived. The expected complexity is defined to represent the expected number of steps performed by the algorithm, and it is a function of the symbols $\mathbf{x}$, and the realization of the channel matrix $\mathbf{H}$. It has been shown that the expected complexity of the SE is exponential.

The optimization algorithm utilizing the HNN is shown in Table I.

TABLE I
HOPFIELD NEURAL NETWORK (HNN) ALGORITHM.

```
Input: Set $\mathbf{H}$, $\mathbf{v} = \mathbf{s}$, $l = 1$, $I_{\max}$
        Set $\mathbf{W} = -\mathrm{diag}[(\mathbf{H}\mathbf{H}^{\dagger})^{-1}] + (\mathbf{H}\mathbf{H}^{\dagger})^{-1}$
Define: $\mathrm{f}_{s_j}(y) = -s_j + 2\tanh\left(2\left(y + s_j\right)\right)$
    while $l \leq I_{\max}$
        for $1 \leq j \leq K$
          Calculate
          $v_j^{(l)} = f_j\left(-\sum_{i=1}^{K} w_{ij} v_i^{(l-1)}\right)$
        end
    end
  for $1 \leq j \leq K$
    $\mathrm{x}_j = -s_j + 2 \cdot \mathrm{sign}(v_j + s_j)$
Output: $\mathbf{x}$
```

For convex precoding, the quadratic programming solver from the MATLAB optimization toolbox [18] is applied. This algorithm has computational complexity $O(K^{3.2})$. Given a fixed number of iterations, the algorithm in Table I contains one loop that is executed K times and involves the summation of K terms. Its complexity is therefore $O(K^2)$.

When we compare computational complexities of the HNN and CR vector precoding methods, we can observe that the dominant complexity is in the linear operation (2). Numerical computation of $(\mathbf{H}\mathbf{H}^{\dagger})^{-1}$ has complexity $O(K^3)$, and the additional complexity due to the application of the HNN is thus negligible in comparison with this pseudo-inverse operation.

IV. NUMERICAL RESULTS

In the HNN vector precoding algorithm the channel matrix $\mathbf{H}$ has been modeled with i.i.d. Gaussian entries. For each realization of the channel matrix $\mathbf{H}$, the number of the iterations was set to be $I_{\max} = 40$.

We simulated the performance for: $K = 8$, $K = 16$, $K = 27$ and $K = 64$. The number of different channel realizations for each system size was 1000. We plot the resulting average transmit energy as a function of the ratio $\alpha = K/N$ of the number of receive antennas K to the number of transmit antennas N, where $0 < \alpha \leq 1$, as shown in Fig. 1. In the same figure we draw for comparison the following plots: the analytical solution for the SE lattice set [12] (the number of the redundant representations of each information bit is $L = 2$), simulation results for the SE lattice precoding ($L = 2$) with the number of receive antennas $K = 27$ and $K = 64$, and the analytical solution for CR [11].

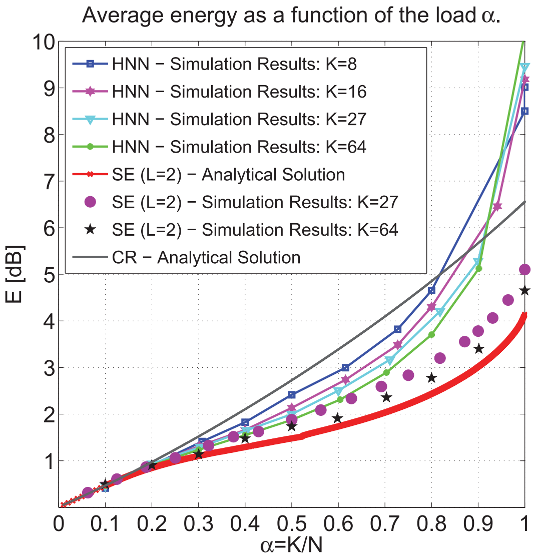


Fig. 1. The average transmit energy as a function of the load $\alpha = K/N$.

The simulation results show that precoding using the HNN provides performances very close to the SE-based precoding performances for the load $0 < \alpha \leq 0.8$. For example, for $K = 27$ and $\alpha = 0.5$ the performance of the SE is 4.88 dB and the HNN performance is 4.99 dB, while for $\alpha = 0.8$ the HNN shows performance penalty of less than 1 dB.

Almost similar results were obtained for $K = 64$; for $\alpha = 0.5$ the SE performance enhancement is 0.11 dB, while for $\alpha = 0.8$ the energy differs by 1 dB. For loads within $0.8 \leq \alpha \leq 0.9$ the average transmit energy by the HNN precoding gradually increases and for $\alpha = 0.9$, the SE outperforms the HNN by 1.5 dB and 1.69 dB for K=27 and $K = 64$, respectively. For loads greater then $\alpha \geq 0.9$, performance of the HNN degrades severely.

In comparison with the analytical results obtained for CR [12] we notice that up to $0 < \alpha \leq 0.9$ the HNN precoding outperforms CR for all simulated values of K, except for $K = 8$, in which case CR is outperformed up to $\alpha \leq 0.8$. The HNN performance enhancement is greatest in the range of $0.5 \leq \alpha \leq 0.8$ and increases when K gets larger. For example, for $\alpha \leq 0.7$, the HNN outperfoms CR by 1.21 dB.

We have thus demonstrated that the HNN-based precoding outperforms the CR theoretical results within $0 < \alpha \leq 0.8$, achieves performances very tight to the SE in the range of $0 < \alpha \leq 0.7$, and has competitive performance in comparison to the SE for $0 < \alpha \leq 0.9$. It is known that lattice precoding is a problem that at loads close to 1 exhibits strong replica symmetry breaking (RSB) [12]. RSB problems are well-known to not being well-approximated by HNNs unlike to those that do not exhibit RSB.

V. COMPUTATIONAL COMPLEXITY

The advantages, limits and computational power of neural networks (for example: [19], [20]) and, in particular, the HNN have been studied over years. There are various realizations of HNN, for example: continuous or discrete time, feedforward or recurrent model, with discrete or analog activation function, finite or infinite network size, asynchronous or synchronous network. The HNN computational complexity has been analyzed depending on the network model and its applications. Some of the results on the computational complexity have been generalized. We will consider a symmetric HNN applied to the energy minimization problems.

The computational complexity can be considered in terms of the memory and time resources required for a particular application. The highest computational cost in terms of the memory resources is due to the allocation of the memory space for storing the weight matrix $\mathbf{W}$. The number of the steps performed by the algorithm before the algorithm converges is convergence time and its trivial upper bound is 2^K. The HNN convergence time may be exponential in the worst case, for both sequential and parallel networks. However, it has been shown that under some mild conditions, the binary HNN converges in only $O(\log \log K)$ parallel steps in the average case. The property allows us to set the maximum number of iterations to a moderate value in practice

VI. CONCLUSIONS

We have presented a method for vector precoding using a HNN as algorithm for combinatorial optimization, and shown that this method can be applied for precoding within a wide load range. We investigated the performance of this scheme by extensive simulations, and compared the results with the simulation results of SE precoding, where the number of redundant representations of each information bit is $L = 2$, with corresponding analytical results for the SE, and with the analytical result for the convex precoding performance. The HNN vector precoding method obtains performances close to the discrete lattice precoding for loads up to $\alpha \leq 0.8$, and for loads $0.8 \leq \alpha \leq 0.9$ there is a gradual increase in the transmit energy within 1.7 dB depending on the number of receiving antennas. When we compare the HNN performances and the CR analytical result we observe that up to $\alpha \approx 0.8$ the HNN precoding outperforms CR for $K = 16$, 27 and 64. For $\alpha \leq 0.7$ the HNN outperforms CR for all simulated values of K. Our simulations showed that this algorithm can be applied for system loads up to $\alpha \leq 0.9$. Therefore, the HNN is an attractive solution for vector precoding of polynomial complexity, with competitive performance within a wide load range in comparison with the SE of exponential complexity. Further modification of the algorithm will be addressed to control the energy penalty for the load up to $\alpha \leq 1$. Furthermore, due to its low complexity, the HNN precoding solution can serve as a starting solution for lattice-based searches with SEs. This allows a greatly improved starting radius for the SE and aid reduction of the SE's complexity.

Finally, we would like to outline that the HNN based lattice precoding, similar to the SE, can be combined with lattice basis reduction. While lattice basis reduction helps to reduce the complexity of the SE, we conjecture that it will not reduce the complexity of the HNN, but improve its performance, particularly at high loads as it reduces the eigenvalue spread of the weight matrix $\mathbf{W}$. This will be investigated in future research.

REFERENCES

[1] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in Multiantenna and Multiuser Communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[3] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[4] M. Joham, W. Utschick, and J. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[5] R. F. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[6] C. Peel, B. Hochwald, and A. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication-part I: Channel Inversion and Regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[7] B. Hassibi and H. Vikalo, "On the Sphere-Decoding Algorithm I. Expected Complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.

[8] J. Jalden and B. Ottersten, "On the Complexity of Sphere Decoding in Digital Communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474–1484, April 2005.

[9] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication Over MIMO Broadcast Channels Using Lattice-Basis Reduction," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4567–4582, 2007.

[10] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring Polynomials with Rational Coefficients," *Math. Ann.*, vol. 261, no. 4, pp. 515–534, 1982.

[11] R. R. Müller, D. Guo, and A. L. Moustakas, "Vector Precoding in High Dimensions: A Replica Analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3., p. 530540, Apr. 2008.

[12] B. M. Zaidel, R. R. Müller, R. de Miguel, and A. L. Moustakas, "On Spectral Efficiency of Vector Precoding for Gaussian MIMO Broadcast Channels," *Proc. IEEE 10th Int. Symp. Spread Spectrum Techniques and Applications (ISSSTA)*, pp. 232–236, Aug. 2008.

[13] A. Cochocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York, NY, USA: John Wiley & Sons, Inc., 1993.

[14] U. P. Wen, K. M. Lan, and H. S. Shih, "A Review of Hopfield Neural Networks for Solving Mathematical Programming Problems," *European Journal of Operational Research*, 2008.

[15] W. H. Schuler, C. J. A. Bastos-Filho, and A. L. I. Oliveira, "A Hybrid Hopfield Network-Simulated Annealing Approach to Optimize Routing Processes in Telecommunications Networks," in *Proc. the Seventh International Conference on Intelligent Systems Design and Applications (ISDA)*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 58–63.

[16] R. R. Müller and J. B. Huber, "Iterated Soft-Decision Interference Cancellation for CDMA," in *Broadband Wireless Communications, Luise and Pupolin, Eds.* Springer, 1998, pp. 110–115.

[17] T. Tanaka and M. Okada, "Approximate Belief Propagation, Density Evolution, and Statistical Neurodynamics for CDMA Multiuser Detection," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 700–706, Feb. 2005.

[18] Matlab Optimization Toolbox, Inc. v3.0.4 (r2006a). [Online]. Available: www.mathworks.com

[19] J. Šíma, "Energy-based Computation with Symmetric Hopfield Nets," in *Limitations and Future Trends in Neural Computation, NATO Science Series: Computer Systems Sciences*, vol. 186. IOS Press, 2003, pp. 45–69.

[20] J. Šíma, P. Orponen, and T. Antti-Poika, "On the Computational Complexity of Binary and Analog Symmetric Hopfield Nets," *Neural Comput.*, vol. 12, no. 12, pp. 2965–2989, 2000.

Multiple-Symbol-Detection-Based Noncoherent Receivers for Impulse-Radio Ultra-Wideband

Andreas Schenk and Robert F.H. Fischer
Lehrstuhl für Informationsübertragung, Universität Erlangen–Nürnberg, Erlangen, Germany
Email: {schenk,fischer}@lnt.de

***Abstract*—Multiple-symbol detection (MSD) is a powerful technique to improve the power efficiency of noncoherent receivers. In this paper, we derive the MSD metric for impulse-radio ultra-wideband for the general case of biorthogonal pulse-position modulation (bPPM) and relate it to its special cases BPSK and PPM. This unified treatment allows us to conduct a comparison of MSD of amplitude- and pulse-position-based impulse-radio signaling schemes in terms of power and spectral efficiency, as well as in terms of complexity.**

I. Introduction

One of the main advantages of the impulse-radio ultra-wideband (IR-UWB) technique in low-complexity transmission systems is its ability to employ noncoherent receivers even in dense multipath propagation scenarios envisioned in typical indoor UWB scenarios [1].

The gap between coherent and noncoherent detection in power efficiency, i.e., in the required signal-to-noise ratio to guarantee a certain bit error rate (BER), can be closed by replacing conventional symbol-by-symbol noncoherent detection with a joint detection of a block of symbols, i.e., performing multiple-symbol detection (MSD). In particular, we consider MSD for differential transmitted reference (DTR) IR-UWB [2], a signaling scheme applying differentially encoded binary phase-shift keying (BPSK). Further considered signaling schemes are orthogonal M-ary pulse-position modulation (M-PPM) [1] and the combination biorthogonal PPM (M-bPPM), i.e., the negatives of the orthogonal PPM signals are included in the signal set, yielding in total $2M$ signal elements. Based on generalized-likelihood ratio testing (GLRT), similar to the approach in [3], we derive the MSD metric of M-bPPM IR-UWB, and relate it to its special cases M-PPM and BPSK.

In [4] these IR-UWB signaling schemes have been compared for transmission over the AWGN channel, while [5] restricts to noncoherent detection of M-PPM in multipath environments. In this paper, we compare the power efficiency of coherent and MSD-based noncoherent receivers for these signaling schemes in a typical UWB multipath propagation scenario. However, only in conjunction with an evaluation of the receiver complexity and the spectral efficiency of the signaling schemes, i.e., the supported number of bits per second per Hertz, we can draw commensurable conclusions from the numerical results. To this end, we evaluate the IR-UWB variants in the power-bandwidth plane [6].

This paper is organized as follows. In Section II, we introduce the system model of M-bPPM IR-UWB used throughout this paper, then derive the MSD metric in Section III, and relate it to the special cases of M-PPM and BPSK. Section IV compares these signaling schemes via numerical results in terms of power and spectral efficiency, and complexity. We conclude with a summary in Section V.

II. System Model

A. Transmit Signal

The transmit signal of biorthogonal M-ary PPM (M-bPPM) IR-UWB is given as

$$s(t) = \sqrt{E_\mathrm{s}/T}\sum_{i=0}^{+\infty} b_i p^\mathrm{TX}(t - a_i\Delta - iT) \qquad (1)$$

where $a_i \in \mathcal{A} = \{0, ..., M-1\}$ are the PPM information symbols and $b_i \in \mathcal{B} = \{\pm 1\}$ are the differentially encoded information symbols $d_i \in \{\pm 1\}$, i.e., $b_i = b_{i-1}d_i$ and $b_0 = 1$. $p^\mathrm{TX}(t)$ is the transmit pulse of unit energy and duration T_{p^TX} in the order of nanoseconds, Δ is the PPM interval, E_s is the energy per bPPM symbol, and $T = M\Delta$ is the symbol duration. Neglecting the reference for differential encoding, b_0, and assuming i.i.d. equal probable data symbols, each symbol conveys $\log_2(M) + 1$ bits, hence the energy per bit is $E_\mathrm{b} = E_\mathrm{s}/(\log_2(M) + 1)$. To preclude inter-pulse and inter-symbol interference even in dense multipath environments and allow for multiple-access capability of a large number of simultaneous users, the PPM interval is chosen sufficiently large, i.e., $\Delta = \beta \cdot T_{p^\mathrm{TX}}$ with $\beta \gg 1$.

B. Spectral Efficiency

Independent of the signaling scheme the transmit signal (1) utilizes a bandwidth approximately proportional to the inverse of the transmit pulse duration, i.e., $\sim \frac{c_p}{T_{p^\mathrm{TX}}}$, with a constant c_p depending on the specific pulse shape ($c_p \approx \pi$ for the Gaussian monocycle considered later). Hence, the spectral efficiency in bits per second per Hertz of M-bPPM is

$$\Gamma_\mathrm{bPPM} = \frac{1 + \log_2(M)}{M}\frac{1}{c_p\beta}\frac{\text{bits/s}}{\text{Hz}} \,. \qquad (2)$$

C. Receive Signal

Having passed a multipath propagation channel with impulse response $h^\mathrm{CH}(t)$ and a receive filter $h^\mathrm{RX}(t)$, the received signal can be written as

$$r(t) = \sum_{i=0}^{+\infty} b_i p(t - a_i\Delta - iT) + n(t) \qquad (3)$$

where $p(t) = \sqrt{E_\mathrm{s}/T}\cdot p^\mathrm{TX}(t) * h^\mathrm{CH}(t) * h^\mathrm{RX}(t)$ is the receive pulse shape, and $n(t) = n_0(t) * h^\mathrm{RX}(t)$ is filtered white Gaussian noise $n_0(t)$ of two-sided power-spectral density $N_0/2$.

This work was supported by Deutsche Forschungsgemeinschaft (DFG) within the framework UKoLoS under grant FI 982/3-1.

D. Coherent Detection

For comparison we recall coherent detection of M-bPPM, assuming ideal knowledge of the receive pulse $p(t)$. For this case, no MSD is necessary and each symbol may be detected by first deciding the transmitted PPM interval based on the magnitude of the crosscorrelation of receive signal and pulse shape, and then the BPSK symbol according to the sign transition in the corresponding interval [6].

III. Multiple-Symbol Detection

At the receiver MSD is performed, i.e., the transmitted sequences $\boldsymbol{a} \in \mathcal{A}^N$ and $\boldsymbol{b} \in \mathcal{B}^N$ are decided blockwise based on the receive signal in the interval $0 \leq t < NT$ (without loss of generality we consider the interval starting at $t = 0$). The channel is assumed to be constant in this interval, which in typical indoor UWB communication scenarios is fulfilled especially for moderate N [7].

A. Decision Metric

Since the additive noise is Gaussian, we base the joint decision of N information symbols on the log-likelihood metric with respect to a receive signal hypothesis $\tilde{s}(t) = \sum_{i=0}^{N-1} \tilde{b}_i \tilde{p}(t - \tilde{a}_i \Delta - iT)$ corresponding to the trial symbols $\tilde{\boldsymbol{a}} = [\tilde{a}_0, ..., \tilde{a}_{N-1}] \in \mathcal{A}^N$, $\tilde{\boldsymbol{b}} = [\tilde{b}_0, ..., \tilde{b}_{N-1}] \in \mathcal{B}^N$ and a hypothesis $\tilde{p}(t)$ for the unknown receive pulse $p(t)$, both assumed to be of duration $T_\mathrm{I} < \Delta$. Due to the differential encoding the reference sign common to all $\tilde{b}_i$, $i \geq 1$, does not influence the decision metric and may be set to $\tilde{b}_0 = 1$.

Following the GLRT approach [8], in contrast to a maximum-likelihood criterion, we perform an explicit optimization over the unknown receive pulse shape $p(t)$ [3], i.e.,

$$[\hat{\boldsymbol{a}}\ \hat{\boldsymbol{b}}] = \underset{\substack{\tilde{\boldsymbol{a}} \in \mathcal{A}^N,\, \tilde{\boldsymbol{b}} \in \mathcal{B}^N \\ \tilde{b}_0 = 1}}{\operatorname{argmax}} \max_{\tilde{p}(t)} \int_0^{NT} \left(2 \cdot r(t)\tilde{s}(t) - \tilde{s}^2(t)\right) \mathrm{d}t$$

and hence do not draw any assumption on the a-priori probability density function of the multipath arrival times or path gains, apart from the assumed pulse duration T_I. However, this GLRT approach leads to the very same decision metric as the ML-approach in [9], derived based on the assumption of a Gaussian distribution of the channel coefficients and a flat power-delay profile of duration T_I. Hence, under these conditions the GLRT estimate is equal to the ML estimate.

Recalling that both $p(t)$ and its hypothesis are assumed of equal duration T_I and $\tilde{b}_i^2 = 1$, with straightforward calculations, we obtain

$$[\hat{\boldsymbol{a}}\ \hat{\boldsymbol{b}}] = \underset{\substack{\tilde{\boldsymbol{a}} \in \mathcal{A}^N,\, \tilde{\boldsymbol{b}} \in \mathcal{B}^N \\ \tilde{b}_0 = 1}}{\operatorname{argmax}} \max_{\tilde{p}(t)} \int_0^{T_\mathrm{I}} \left[\tilde{p}(t) \sum_{i=0}^{N-1} \tilde{b}_i r(t + \tilde{a}_i \Delta + iT) - \frac{N}{2} \cdot \tilde{p}^2(t)\right] \mathrm{d}t\,.$$

Similar to [3], fixing $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$, we solve the maximization over $\tilde{p}(t)$ analytically using variational calculus (omitted for brevity), and obtain a MSD metric for M-bPPM solely based on the receive signal in the observation window $0 \leq t < NT$

$$[\hat{\boldsymbol{a}}\ \hat{\boldsymbol{b}}] = \underset{\substack{\tilde{\boldsymbol{a}} \in \mathcal{A}^N,\, \tilde{\boldsymbol{b}} \in \mathcal{B}^N \\ \tilde{b}_0 = 1}}{\operatorname{argmax}} \int_0^{T_\mathrm{I}} \left[\sum_{i=0}^{N-1} \tilde{b}_i r(t + \tilde{a}_i \Delta + iT)\right]^2 \mathrm{d}t\,. \tag{4}$$

The assumed receive pulse duration T_I, the integration interval, should be set on the one hand large enough to capture sufficient energy of the receive signal, and on the other hand as small as possible not to accumulate too much noise.

Solving (4) requires finding the maximum of $2^{N-1}M^N$ combinations of weighted receive signal intervals, hence, only moderate values of N seem to be applicable. For sufficiently high sampling frequency, (4) can straightforwardly be formulated to work on the sampled and quantized receive signal, analog delay lines can hence be avoided [9]. Implicit restrictions on the sequences $\boldsymbol{a}$ and $\boldsymbol{b}$, as, e.g., in DTR signaling, can be used to reduce the number of candidates, yet, this is not considered here.

Due to the differential encoding, $N = 1$ does not lead to reasonable performance. However, a natural way to overcome this is to perform symbol-by-symbol energy detection (ED) of the M-PPM part ($N = 1$) and differential detection (DD) of the BPSK part ($N = 2$).

B. PPM

The special case of MSD of M-PPM IR-UWB results by setting $b_i = 1$, $\forall i$ (each symbol now conveys $\log_2(M)$ bits, hence $E_\mathrm{b} = E_\mathrm{s}/\log_2(M)$), and the corresponding MSD metric is given as

$$\hat{\boldsymbol{a}} = \underset{\tilde{\boldsymbol{a}} \in \mathcal{A}^N}{\operatorname{argmax}} \int_0^{T_\mathrm{I}} \left[\sum_{i=0}^{N-1} r(t + \tilde{a}_i \Delta + iT)\right]^2 \mathrm{d}t\,. \tag{5}$$

If $N = 1$, (5) corresponds to ED of M-PPM.

C. BPSK

Similarly, MSD of a solely BPSK modulated signal can be viewed as a special case of MSD of M-bPPM. Setting $M = 1$ ($\mathcal{A} = \{0\}$), each symbol represents a single bit, hence $E_\mathrm{b} = E_\mathrm{s}$. Note that b_i still are the differentially encoded information symbols. Using $\tilde{b}_i^2 = 1$, the corresponding MSD metric can be rearranged [3], yielding the triangular structure

$$\hat{\boldsymbol{b}} = \underset{\substack{\tilde{\boldsymbol{b}} \in \mathcal{B}^N \\ \tilde{b}_0 = 1}}{\operatorname{argmax}} \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \tilde{b}_i \tilde{b}_j \int_0^{T_\mathrm{I}} r(t + iT) r(t + jT)\, \mathrm{d}t\,. \tag{6}$$

Hence, for the special case of BPSK signaling the MSD metric in (4) reduces to an autocorrelation of the receive signal with delays being multiples of the symbol duration T, followed by maximization of the decision metric, as shown in [2], [3]. The latter can be formulated as a tree search problem and is efficiently implemented by the sphere decoder (SD), which avoids testing all 2^{N-1} candidate sequences (complexity exponential in N), resulting in effectively polynomial search complexity for a wide range of signal-to-noise ratios. Thus considerably larger MSD window lengths N compared to a full search as required for MSD of PPM become amenable [3], [10].

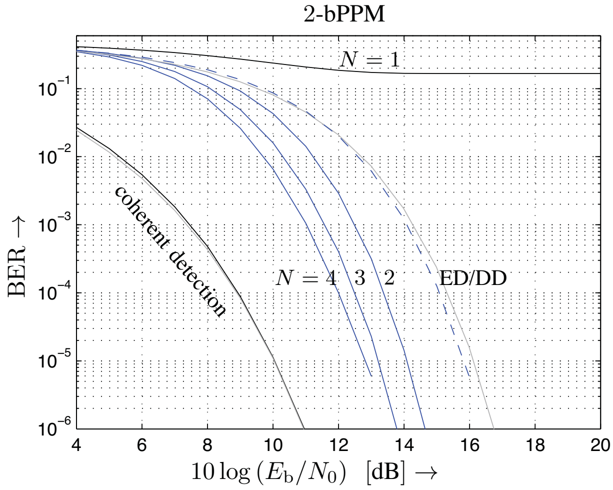


Fig. 1. BER vs. E_b/N_0 in dB for MSD of 2-bPPM IR-UWB with different N in comparison to conventional ED ($N = 1$), coherent detection, and ED/DD detection (dashed). Gray lines: analytical/approximate BER.

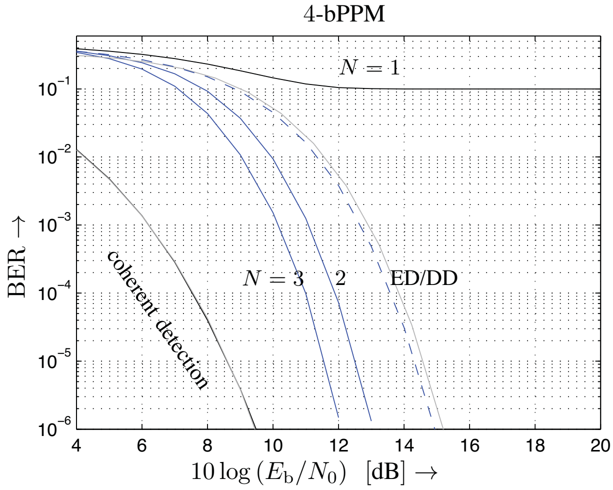


Fig. 2. BER vs. E_b/N_0 in dB for MSD of 4-bPPM IR-UWB with different N in comparison to conventional ED ($N = 1$), coherent detection, and ED/DD detection (dashed). Gray lines: analytical/approximate BER.

In a similar way, MSD of M-bPPM can be realized with M^N parallel autocorrelation receivers, each tuned to one PPM sequence followed by a SD to find the corresponding BPSK part. This puts most of the receiver complexity on detection of the PPM part, while the BPSK part is decided with little additional effort.

IV. Comparison

We compare MSD of the various signaling schemes via numerical results in a typical UWB scenario, where we assume no inter-symbol interference (T chosen sufficiently large), $T_{p^{\mathrm{TX}}} = 1\,\mathrm{ns}$, and $p^{\mathrm{TX}}(t)$ is a Gaussian monocycle with $10\,\mathrm{dB}$ bandwidth of $3.3\,\mathrm{GHz}$ and $2.25\,\mathrm{GHz}$ center frequency. The propagation channel is modeled according to IEEE-CM 2 [7] with each realization normalized to unit energy. The receive filter is modeled as an ideal $3\,\mathrm{GHz}$ bandpass filter around

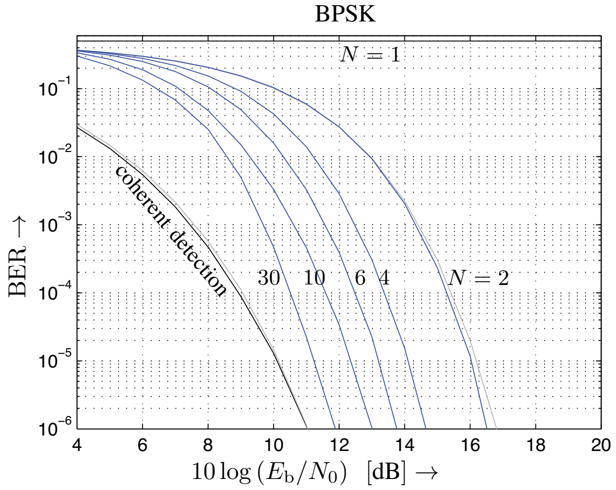


Fig. 3. BER vs. E_b/N_0 in dB for MSD of BPSK IR-UWB with different N in comparison to conventional ED ($N = 1$), DD ($N = 2$) and coherent detection. Gray lines: analytical/approximate BER.

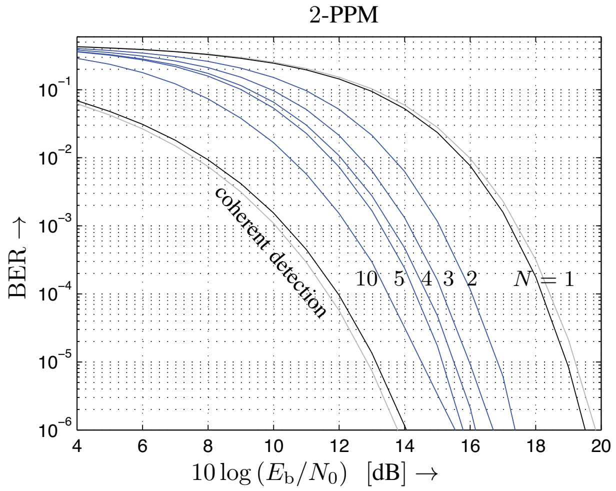


Fig. 4. BER vs. E_b/N_0 in dB for MSD of 2-PPM IR-UWB with different N in comparison to conventional ED ($N = 1$) and coherent detection. Gray lines: analytical/approximate BER.

the pulse center frequency and a good compromise for the integration time is $T_{\mathrm{I}} = 30\,\mathrm{ns}$.

In all figures, gray lines represent approximate BER expressions of ED/DD, which directly result from a Gaussian approximation of the decision metric in the spirit of [2], [11], [12], and the analytical BER for the well known case of coherent detection [6] (omitted here due to lack of space).

Exemplary, Figure 1 and Figure 2 depict the BER of MSD of 2 and 4-bPPM IR-UWB, respectively. The dashed line corresponds to the low-complexity detection (see Section III), i.e., symbol-by-symbol ED of the PPM part followed by DD of the BPSK part. Already MSD with $N = 2$ leads to a gain of $2\,\mathrm{dB}$ in comparison to ED/DD at $\mathrm{BER} = 10^{-5}$. With increasing N performance is improved further and approaches that of coherent detection.

Turning to BPSK, MSD using the SD however enables

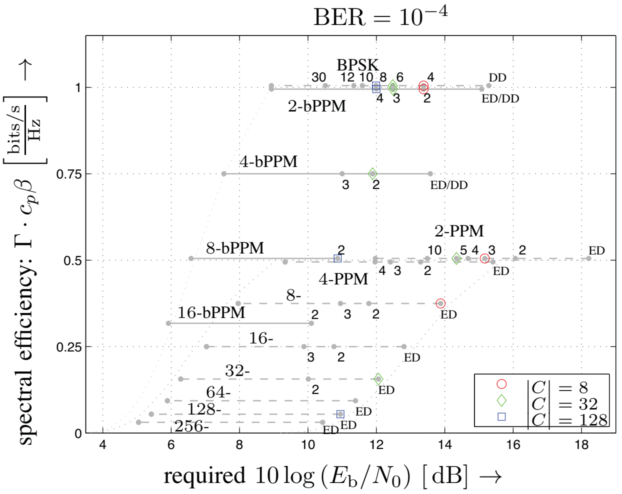


Fig. 5. Trade-off power vs. spectral efficiency at $\mathrm{BER} = 10^{-4}$ of IR-UWB coherent detection and MSD with parameter N as indicated. Solid lines: M-bPPM, dashed: M-PPM, dash-dotted: BPSK. Colored markers: MSD with an equal number of candidates. Dotted lines: analytical/approximate BER.

considerably larger N in the order of 30 and results in performance close to coherent detection (1 dB difference for $N = 30$ at $\mathrm{BER} = 10^{-5}$), as depicted in Figure 3. Here $N = 2$ corresponds to conventional autocorrelation DD.

Considering 2-PPM, from Figure 4 it can be seen, that again already the joint decision of two 2-PPM symbols leads to a gain of more than 2 dB compared to conventional ED. Larger N bridge the gap to coherent detection of 2-PPM with ideal knowledge of the receive pulse shape. In comparison to 2-bPPM, for a fixed N 2-PPM shows a loss of approximately 3 dB, indicating that the PPM part dominates the bPPM performance.

For the same setting as above, Figure 5 visualizes the trade-off between power and spectral efficiency (see (2)) of the studied signaling schemes at $\mathrm{BER} = 10^{-4}$. The lines are parameterized by the MSD window parameter N (indicated next to the dots), where the left most dot corresponds to coherent detection and the right most dot to DD ($N = 2$) for BPSK, to the low-complexity detection as described in Section III (termed ED/DD) for M-bPPM, and to ED ($N = 1$) for M-PPM. Dotted lines represent the performance resulting from analytical/approximate BER expressions. Markers flag an equal number of candidate sequences required for MSD (indicating the receiver complexity–further complexity reduction due to the application of the SD for MSD of BPSK may be possible).

The gain achieved by increasing the MSD window N, naturally accompanied with an increase in receiver complexity, is similar to all the signaling schemes. However, it is important to note that signaling schemes making use of the sign information of the pulse, i.e., BPSK and M-bPPM, lead to a significant increase both in power and in spectral efficiency in comparison to the solely pulse-position-based scheme M-PPM.

Fixing the number of candidate sequences in MSD, i.e., the dimensionality of the search problem (M-bPPM: $2^{N-1}M^N$, BPSK: 2^{N-1}, M-PPM: M^N), from the markers in Figure 5 it can be seen that under this constraint BPSK and 2-bPPM, both signaling schemes using the sign of the pulse, achieve very similar power efficiency, which is substantially higher than that of 2-PPM. Only higher-order PPM overcomes this drawback, however, at the cost of considerably reduced spectral efficiency. This extends the well-known fact for coherent detection [6] and the conclusions of [4] to the case of MSD-based noncoherent detection of IR-UWB in multipath propagation scenarios.

Note that the significant complexity reduction achieved with the application of the SD in MSD of BPSK (and similar in M-bPPM) is not mirrored in Figure 5. The advantages of BPSK and M-bPPM in terms of power and spectral efficiency are accompanied by a reduction in the complexity of noncoherent receivers, which further substantiates to favor sign-based schemes, i.e., BPSK or M-bPPM, over pulse-position-based schemes in IR-UWB systems.

V. Conclusions

In this paper we have compared MSD-based noncoherent receivers for IR-UWB using BPSK, PPM, and biorthogonal PPM with respect to performance, complexity, and spectral efficiency. To this end, we derived the MSD decision metric of biorthogonal PPM IR-UWB and related it to its special cases of PPM and BPSK. While the gain achieved with increasing MSD block length is similar for all IR-UWB signaling schemes, making use of the sign information, i.e., BPSK and biorthogonal PPM, proves preferable to solely PPM not only in terms of power and spectral efficiency, but also in terms of complexity.

References

[1] M. Z. Win and R. A. Scholtz, “Impulse Radio: How It Works,” *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[2] N. Guo and R. C. Qiu, “Improved Autocorrelation Demodulation Receivers Based on Multiple-Symbol Detection for UWB Communications,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2026–2031, Aug. 2006.

[3] V. Lottici and Z. Tian, “Multiple Symbol Differential Detection for UWB Communications,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1656–1666, May 2008.

[4] H. Zhang and T. A. Gulliver, “Biorthogonal Pulse Position Modulation for Time-Hopping Multiple Access UWB Communications,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1154–1162, May 2005.

[5] Y. Souilmi and R. Knopp, “On the Achievable Rates of Ultra-Wideband PPM with Non-Coherent Detection in Multipath Environments,” in *Proc. IEEE International Conference on Communications (ICC ’03)*, vol. 5, pp. 3530–3534, Anchorage, USA, May 11–15, 2003.

[6] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008.

[7] A. F. Molisch, J. R. Foerster, and M. Pendergrass, “Channel Models for Ultrawideband Personal Area Networks,” *IEEE Wireless Commun. Mag.*, vol. 10, no. 6, pp. 14–21, Dec. 2003.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Volume II - Detection Theory*. New Jersey, USA: Prentice-Hall PTR, 1998.

[9] Y. Tian and C. Yang, “Noncoherent Multiple-Symbol Detection in Coded Ultra-Wideband Communications,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2202–2211, Jun. 2008.

[10] A. Schenk, R. F. H. Fischer, and L. Lampe, “A New Stopping Criterion for the Sphere Decoder in UWB Impulse-Radio Multiple-Symbol Differential Detection,” in *Proc. IEEE International Conference on Ultra-Wideband (ICUWB ’09)*, pp. 589–594, Vancouver, Canada, Sep. 9–11, 2009.

[11] T. Q. S. Quek and M. Z. Win, “Analysis of UWB Transmitted-Reference Communication Systems in Dense Multipath Channels,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1863–1874, Sep. 2005.

[12] M. Pausini and G. J. M. Janssen, “Performance Analysis of UWB Autocorrelation Receivers Over Nakagami-Fading Channels,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 443–455, Oct. 2007.

On the Gain of Joint Processing of Pilot and Data Symbols in Stationary Rayleigh Fading Channels

Meik Dörpinghaus, Adrian Ispas, Gerd Ascheid, and Heinrich Meyr
Institute for Integrated Signal Processing Systems, RWTH Aachen University, Germany
Email: {Doerpinghaus, Ispas, Ascheid, Meyr}@iss.rwth-aachen.de

***Abstract*—In many typical mobile communication receivers the channel is estimated based on pilot symbols to allow for a coherent detection and decoding in a separate processing step. Currently much work is spent on receivers which break up this separation, e.g., by enhancing channel estimation based on reliability information on the data symbols. In the present work, we discuss the nature of the possible gain of a joint processing of data and pilot symbols in comparison to the case of a separate processing in the context of stationary Rayleigh flat-fading channels. In addition, we derive a new lower bound on the achievable rate for joint processing of pilot and data symbols.**

I. Introduction

Virtually all practical mobile communication systems face the problem that communication takes place over a time varying fading channel whose realization is unknown to the receiver. However, for coherent detection and decoding an estimate of the channel fading process is required. For the purpose of channel estimation usually pilot symbols, i.e., symbols which are known to the receiver, are introduced into the transmit sequence. In conventional receiver design the channel is estimated based on these pilot symbols. Then, in a separate step, coherent detection and decoding is performed.

In recent years, much effort has been spent on the study of iterative joint channel estimation and decoding schemes, i.e., schemes, in which the channel estimation is iteratively enhanced based on reliability information on the data symbols delivered by the decoder, see, e.g., [1], [2]. In this context, the channel estimation is not solely based on pilot symbols, but also on data symbols. This approach is an instance of a *joint processing* of data and pilot symbols in contrast to the separate processing in conventional receivers. To evaluate the payoff for the increased receiver complexity with joint processing, it is important to study the possible performance gain that can be achieved by a joint processing, e.g., in form of an iterative code-aided channel estimation and decoding based receiver, in comparison to a separate processing.

Therefore, in the present work we will evaluate the performance of a joint processing in comparison to synchronized detection with a solely pilot based channel estimation based on the achievable rate. Regarding the channel statistics we assume a stationary Rayleigh flat-fading channel as it is usually applied to model the fading in a mobile environment without a line of sight component. Furthermore, we assume that the power spectral density (PSD) of the channel fading process is compactly supported, and that the fading process is *non-regular* [3]. Moreover, we assume that the receiver is aware of the law of the channel, while neither the transmitter nor the receiver knows the realization of the fading process.

For the case of synchronized detection with a solely pilot based channel estimation there exist already bounds on the achievable rate [4]. In contrast, for the case of joint processing there is not much knowledge on the achievable rate. Very recently, in [5] the value of joint processing of pilot and data symbols has been studied in the context of a block-fading channel. To the best of our knowledge, there are no results concerning the gain of joint processing of pilot and data symbols for the case of stationary fading channels. Thus, in the present work, we give a lower bound on the achievable rate with joint processing of pilot and data symbols. Besides this lower bound on the achievable rate with a joint processing of pilot and data symbols, we identify the nature of the possible gain of a joint processing in comparison to a separate processing.

II. System Model

We consider a discrete-time zero-mean jointly proper Gaussian flat-fading channel with the input-output relation

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{1}$$

with the diagonal matrix $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. Here the $\mathrm{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal elements are given by the argument vector. The vector $\mathbf{y} = [y_1, \dots, y_N]^T$ contains the channel output symbols in temporal order. Analogously, $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{h}$ contain the channel input symbols, the additive noise samples, and the channel fading weights. All vectors are of length N.

The samples of the additive noise process are i.i.d. zero-mean jointly proper Gaussian with variance σ_n^2.

The channel fading process is zero-mean jointly proper Gaussian with the temporal correlation characterized by $r_h(l) = \mathrm{E}[h_{k+l}h_k^*]$. Its variance is given by $r_h(0) = \sigma_h^2$, and, due to technical reasons, it is assumed to be absolutely summable, i.e., $\sum_{l=-\infty}^{\infty} |r_h(l)| < \infty$. The PSD of the channel fading process is defined as

$$S_h(f) = \sum_{m=-\infty}^{\infty} r_h(m) e^{-j2\pi m f}, \qquad |f| \leq 0.5. \tag{2}$$

We assume that the PSD exists, which for a jointly proper Gaussian fading process implies ergodicity. Furthermore, we assume the PSD to be compactly supported within the interval $[-f_d, f_d]$ with f_d being the maximum Doppler shift and $0 < f_d < 0.5$. This means that $S_h(f) = 0$ for $f \notin [-f_d, f_d]$. The assumption of a PSD with limited support is motivated by the fact that the velocity of the transmitter, the receiver, and of objects in the environment is limited. To ensure ergodicity, we exclude the case $f_d = 0$.

The transmit symbol sequence consists of data symbols with an average power σ_x^2 and periodically inserted pilot symbols with a fixed power σ_x^2. Each L-th symbol is a pilot symbol. The pilot spacing is chosen such that the channel fading process is sampled at least with Nyquist rate, i.e.,

$$L < 1/(2f_d). \tag{3}$$

In the following we use the subvectors $\mathbf{x}_D$ containing all data symbols of $\mathbf{x}$ and $\mathbf{x}_P$ containing all pilot symbols of $\mathbf{x}$. Correspondingly, we define $\mathbf{h}_D$, $\mathbf{h}_P$, $\mathbf{y}_D$, $\mathbf{y}_P$, $\mathbf{n}_D$, and $\mathbf{n}_P$.

The processes $\{x_k\}$, $\{h_k\}$ and $\{n_k\}$ are assumed to be mutually independent. The mean SNR is given by $\rho = \sigma_x^2\sigma_h^2/\sigma_n^2$.

III. The nature of the gain by joint processing of data and pilot symbols

Before we quantitatively discuss the value of a joint processing of data and pilot symbols, we discuss the nature of the possible gain of such a joint processing in comparison to a separate processing of data and pilot symbols. The mutual information between the transmitter and the receiver is given by $\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)$. As the pilot symbols are known to the receiver, the pilot symbol vector $\mathbf{x}_P$ is found at the RHS of the semicolon. We separate $\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)$ as follows

$$\begin{aligned}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) &\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P)+\mathcal{I}(\mathbf{x}_D;\mathbf{y}_P|\mathbf{x}_P)\\ &\quad+\mathcal{I}(\mathbf{x}_D;\mathbf{x}_P) \overset{(b)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P)\end{aligned} \tag{4}$$

where (a) follows from the chain rule for mutual information and (b) holds due to the independency of the data and pilot symbols. The question is, which portion of $\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P)$ can be achieved by synchronized detection with a solely pilot based channel estimation, i.e., with separate processing.

A. Separate Processing

The receiver has to find the most likely data sequence $\mathbf{x}_D$ based on the observation $\mathbf{y}$ while knowing the pilots $\mathbf{x}_P$, i.e.,

$$\hat{\mathbf{x}}_D = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}|\mathbf{x}) = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P) \tag{5}$$

with the set $\mathcal{C}_D$ containing all possible data sequences $\mathbf{x}_D$. The probability density function (PDF) $p(\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P)$ is proper Gaussian and, thus, completely described by the conditional mean and covariance

$$\mathrm{E}\left[\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P\right] = \mathbf{X}_D\mathrm{E}\left[\mathbf{h}_D|\mathbf{y}_P,\mathbf{x}_P\right] = \mathbf{X}_D\hat{\mathbf{h}}_{\text{pil},D} \tag{6}$$

$$\mathrm{cov}[\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P] = \mathbf{X}_D\mathbf{R}_{e_{\text{pil}},D}\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_{N_D} \tag{7}$$

where $\mathbf{X}_D = \mathrm{diag}(\mathbf{x}_D)$ and $\mathbf{I}_{N_D}$ is an identity matrix of size $N_D \times N_D$ with N_D being the length of $\mathbf{n}_D$. The vector $\hat{\mathbf{h}}_{\text{pil},D}$ is an MMSE channel estimate at the data symbol time instances based on the pilot symbols, which is denoted by the index *pil*. Furthermore, the corresponding channel estimation error $\mathbf{e}_{\text{pil},D} = \mathbf{h}_D - \hat{\mathbf{h}}_D$ is zero-mean proper Gaussian and $\mathbf{R}_{e_{\text{pil}},D} = \mathrm{E}\left[\mathbf{e}_{\text{pil},D}\mathbf{e}_{\text{pil},D}^H|\mathbf{x}_P\right]$ is its correlation matrix, which is independent of $\mathbf{y}_P$ due to the principle of orthogonality.

Based on (6) and (7) conditioning of $\mathbf{y}_D$ on $\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P$ is equivalent to conditioning on $\mathbf{x}_D, \hat{\mathbf{h}}_{\text{pil},D}, \mathbf{x}_P$, i.e.,

$$p(\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P) = p(\mathbf{y}_D|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil},D},\mathbf{x}_P) \tag{8}$$

as all information on $\mathbf{h}_D$ delivered by $\mathbf{y}_P$ is contained in $\hat{\mathbf{h}}_{\text{pil},D}$ while conditioning on $\mathbf{x}_P$. Thus, (5) can be written as

$$\hat{\mathbf{x}}_D = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil},D},\mathbf{x}_P) = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P). \tag{9}$$

For ease of notation in the following we will use the metric on the RHS of (9) where $\hat{\mathbf{h}}_{\text{pil}}$ corresponds to $\hat{\mathbf{h}}_{\text{pil},D}$ but also contains channel estimates at the pilot symbol time instances, i.e., $\hat{\mathbf{h}}_{\text{pil}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P\right]$. Based on $\hat{\mathbf{h}}_{\text{pil}}$, (1) can be expressed by

$$\mathbf{y} = \mathbf{X}(\hat{\mathbf{h}}_{\text{pil}} + \mathbf{e}_{\text{pil}}) + \mathbf{n} \tag{10}$$

where $\mathbf{e}_{\text{pil}}$ is the estimation error including the pilot symbol time instances. As the channel estimation is an interpolation, the error process is not white but temporally correlated, i.e.,

$$\mathbf{R}_{e_{\text{pil}}} = \mathrm{E}\left[\mathbf{e}_{\text{pil}}\mathbf{e}_{\text{pil}}^H|\mathbf{x}_P\right] \tag{11}$$

is not diagonal, cf. (21). Thus, the PDF in (9) is given by

$$p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) = \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\text{pil}}, \mathbf{X}\mathbf{R}_{e_{\text{pil}}}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right) \tag{12}$$

where $\mathcal{CN}(\boldsymbol{\mu},\mathbf{C})$ denotes a proper Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ and where $\mathbf{I}_N$ is the $N\times N$ identity matrix.[1]

Note that corresponding to (8), we can also rewrite (4) as

$$\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P) = \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}})$$

and where (a) holds as the pilot symbols are deterministic.

However, typical channel decoders like a Viterbi decoder are not able to exploit the temporal correlation of the channel estimation error. Therefore, the decoder performs mismatch decoding based on the assumption that the estimation error process is white, i.e., $p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P)$ is approximated by

$$p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) \approx \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\text{pil}}, \sigma_{e_{\text{pil}}}^2\mathbf{X}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right). \tag{13}$$

As it is assumed that the channel is at least sampled with Nyquist frequency, see (3), for an infinite block length $N \to \infty$ the channel estimation error variance $\sigma_{e_{\text{pil}}}^2$ is independent of the symbol time instant [4] and is given by

$$\sigma_{e_{\text{pil}}}^2 = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} S_{e_{\text{pil}}}(f)df = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2}+1}df \tag{14}$$

where the PSD of the channel estimation error process $S_{e_{\text{pil}}}(f)$ is given in (21). Hence, the variance of the channel estimation process, i.e., of the entries of $\hat{\mathbf{h}}_{\text{pil}}$, is given by $\sigma_h^2 - \sigma_{e_{\text{pil}}}^2$, which follows from the principle of orthogonality.

As the information contained in the temporal correlation of the channel estimation error is not retrieved by synchronized detection with a solely pilot based channel estimation, the mutual information in this case corresponds to the sum of

[1] Note that for the case of data transmission only (12) becomes $p(\mathbf{y}|\mathbf{x}_D) = \mathcal{CN}(\mathbf{0}, \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2\mathbf{I}_N)$ as in this case $\hat{\mathbf{h}}_{\text{pil}} = \mathbf{0}$ and $\mathbf{R}_{e_{\text{pil}}} = \mathbf{R}_h$.

the mutual information for each individual data symbol time instant. As, obviously, by this separate processing information is discarded, the following inequality holds

$$\lim_{N\to\infty}\frac{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}})}{N}=\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}})\geq\frac{L-1}{L}\mathcal{I}(x_{D_k};y_{D_k}|\hat{\mathbf{h}}_{\text{pil}}) \quad (15)$$

where $\mathcal{I}'$ denotes the mutual information rate and the index D_k denotes an arbitrarily chosen data symbol.

As the LHS of (15) is the mutual information of the channel and the RHS of (15) is the mutual information achievable with synchronized detection with a metric corresponding to (13) and a solely pilot based channel estimation, i.e., a separate processing, the difference of both terms upper-bounds the possible gain due to joint processing of data and pilot symbols. The additional information that can be gained by a joint processing in contrast to separate processing is contained in the temporal correlation of the channel estimation error process.

Regarding synchronized detection in combination with a solely pilot based channel estimation, i.e., separate processing, in [4] bounds on the achievable rate, i.e., on the RHS of (15), are given. In Fig. 1 these bounds are shown for i.i.d. zero-mean proper Gaussian data-symbols. These bounds show that the achievable rate with separate processing is decreased in comparison to perfect channel knowledge in two ways. First, time instances used for pilot symbols are lost for data symbols, and secondly, the average SNR is decreased due to the channel estimation error variance.

IV. Joint processing of data and pilot symbols

Now, we give a new lower bound on the achievable rate for a joint processing of data and pilot symbols. The following approach can be seen as an extension of the work in [5] for the case of a block-fading channel to the stationary Rayleigh flat-fading scenario discussed in the present work. Therefore, analogous to [5] we decompose and lower-bound the mutual information between the transmitter and the receiver as follows

$$\begin{aligned}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)&\overset{(a)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P,\mathbf{h})-\mathcal{I}(\mathbf{x}_D;\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&=\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)+h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&\overset{(b)}{\geq}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P)+h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \quad (16)\end{aligned}$$

where (a) follows from the chain rule for mutual information and (b) is due to the fact that conditioning reduces entropy. The first term on the RHS of (16) is the mutual information in case of perfect channel knowledge.

Now we deviate from [5] and rewrite the RHS of (16) as

$$\begin{aligned}(16)&\overset{(a)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P)+h(\mathbf{h}|\hat{\mathbf{h}}_{\text{joint}},\mathbf{x}_D,\mathbf{x}_P)\\&\overset{(b)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)+h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\&\overset{(c)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\text{logdet}\left(\pi e\mathbf{R}_{e_{\text{pil}}}\right)+\text{logdet}\left(\pi e\mathbf{R}_{e_{\text{joint}}}\right) \quad (17)\end{aligned}$$

where for (a) we have substituted the conditioning on $\mathbf{y}_P$ by $\hat{\mathbf{h}}_{\text{pil}}$, which is possible as the estimate $\hat{\mathbf{h}}_{\text{pil}}$ contains the same information on $\mathbf{h}$ as $\mathbf{y}_P$ while conditioning on $\mathbf{x}_P$. Corresponding to the solely pilot based channel estimate $\hat{\mathbf{h}}_{\text{pil}}$, based on $\mathbf{x}_D$, $\mathbf{x}_P$, $\mathbf{y}_D$, and $\mathbf{y}_P$, we can calculate the estimate $\hat{\mathbf{h}}_{\text{joint}}$, which is based on data and pilot symbols. Like $\hat{\mathbf{h}}_{\text{pil}}$ this estimate is a MAP estimate, which, due to the jointly Gaussian nature of the problem, is an MMSE estimate, i.e.,

$$\hat{\mathbf{h}}_{\text{joint}}=\mathrm{E}\left[\mathbf{h}|\mathbf{y}_D,\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P\right]. \quad (18)$$

Thus, for (a) we have substituted the conditioning on $\mathbf{y}_D$ and $\mathbf{y}_P$ by conditioning on $\hat{\mathbf{h}}_{\text{joint}}$ in the third term, as $\hat{\mathbf{h}}_{\text{joint}}$ contains all information on $\mathbf{h}$ that is contained in $\mathbf{y}_D$ and $\mathbf{y}_P$ while $\mathbf{x}_D$ and $\mathbf{x}_P$ are known. For the second term in equality (b) we have used (10), the fact that the addition of a constant does not change differential entropy and that the estimation error $\mathbf{e}_{\text{pil}}$ is independent of the estimate $\hat{\mathbf{h}}_{\text{pil}}$. Analogously, for the third term we used the separation of $\mathbf{h}$ into the estimate $\hat{\mathbf{h}}_{\text{joint}}$ and the corresponding estimation error $\mathbf{e}_{\text{joint}}$ which depends on $\mathbf{x}_D$ and $\mathbf{x}_P$ and is independent of $\hat{\mathbf{h}}_{\text{joint}}$. Finally, (c) holds as the estimation error processes are zero-mean jointly proper Gaussian. The error correlation matrices are given by (11) and by

$$\mathbf{R}_{e_{\text{joint}}}=\mathrm{E}\left[\mathbf{e}_{\text{joint}}\mathbf{e}_{\text{joint}}^H|\mathbf{x}_D,\mathbf{x}_P\right]. \quad (19)$$

The estimation error $\mathbf{e}_{\text{joint}}$ depends on the distribution of the data symbols $\mathbf{x}_D$. It can be shown that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)=\lim_{N\to\infty}\frac{1}{N}h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ is minimized for a given average transmit power σ_x^2 if the data symbols have constant modulus (CM). Due to lack of space the proof given in [6] is not shown here.

Thus, with (16) and (17) a lower bound for the achievable rate with joint processing of data and pilot symbols is given by

$$\begin{aligned}&\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)=\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&\geq\lim_{N\to\infty}\frac{1}{N}\left\{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\log\det\left(\mathbf{R}_{e_{\text{pil}}}\right)+\log\det\left(\mathbf{R}_{e_{\text{joint,CM}}}\right)\right\}\\&\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)}\right)df \quad (20)\end{aligned}$$

with $\mathbf{R}_{e_{\text{joint,CM}}}$ corresponding to (19), but under the assumption of CM data symbols with transmit power σ_x^2. Note that the CM assumption has only been used to lower bound the third term at the RHS of (17), and not the whole expression at the RHS of (17). For (a) in (20) we have used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [7]. $S_{e_{\text{pil}}}(f)$ and $S_{e_{\text{joint,CM}}}(f)$ are the PSDs of the channel estimation error processes, on the one hand, if the estimation is solely based on pilot symbols, and on the other hand, if the estimation is based on data and pilot symbols, assuming CM data symbols. They are given by [6]

$$S_{e_{\text{pil}}}(f)=\frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2}+1},\quad S_{e_{\text{joint,CM}}}(f)=\frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2}+1}. \quad (21)$$

The first term on the RHS of (20) is the mutual information rate in case of perfect channel state information, which for an average power constraint is maximized with i.i.d. zero-mean proper Gaussian data symbols. Thus, we get the following lower bound on the achievable rate with joint processing

$$\mathcal{R}_{L,\text{joint}}=\frac{L-1}{L}C_{\text{perf}}-\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\frac{\rho}{\sigma_h^2}S_h(f)+1}{\frac{\rho}{L\sigma_h^2}S_h(f)+1}\right)df \quad (22)$$

where C_{perf} corresponds to the coherent capacity with

$$C_{\text{perf}} = \mathrm{E}_h\left[\log\left(1+\rho\frac{|h|^2}{\sigma_h^2}\right)\right] = \int_{z=0}^{\infty}\log\left(1+\rho z\right)e^{-z}dz \quad (23)$$

and the factor $(L-1)/L$ arises as each L-th symbol is a pilot.

A. Lower Bound on the Achievable Rate for a Joint Processing of Data and Pilot Symbols and a Fixed Pilot Spacing

Equation (22) is a lower bound on the achievable rate with joint processing of data and pilot symbols, for a given pilot spacing L and stationary Rayleigh flat-fading.

For the special case of a rectangular PSD[2] $S_h(f)$, i.e.,

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

the lower bound in (22) becomes

$$\mathcal{R}_{L,\text{joint}}\Big|_{\text{rect.}S_h(f)} = \frac{L-1}{L}\int_{z=0}^{\infty}\log(1+\rho z)e^{-z}dz - 2f_d\log\left(\frac{\rho+2f_d}{\frac{\rho}{L}+2f_d}\right). \quad (25)$$

B. Lower Bound on the Achievable Rate for a Joint Processing of Data and Pilot Symbols and an Arbitrary Pilot Spacing

The lower bound in (25) depends on the pilot spacing L and can be enhanced by calculating the supremum of (25) with respect to L. In this regard, it has to be considered that the pilot spacing L is an integer value. Furthermore, we have to take into account that the derivation of the lower bound in (25) is based on the assumption that the pilot spacing is chosen such that the channel fading process is at least sampled with Nyquist rate, see (3). For larger L the estimation error process is no longer stationary, which is required for our derivation.[3]

For these conditions, the lower bound (25) is maximized for

$$L_{\text{opt}} = \lfloor 1/(2f_d) \rfloor \quad (26)$$

which can be observed based on differentiation of (25) w.r.t. L and numerical evaluation. Note that L_{opt} is not necessarily the L which maximizes the achievable rate.

V. Numerical Evaluation

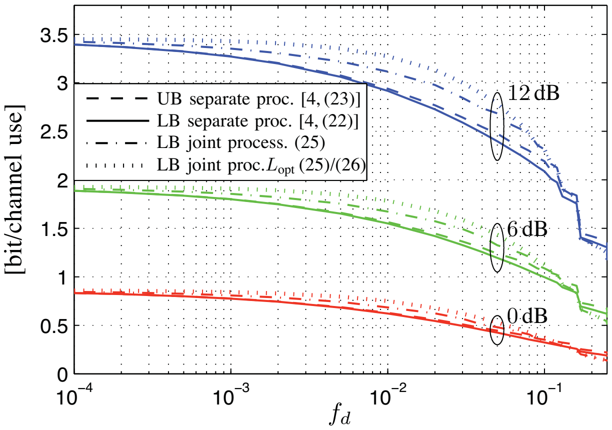


Fig. 1. Comparison of bounds on the achievable rate with separate processing to lower bounds on the achievable rate with joint processing of data and pilot symbols; except of *LB joint proc.* L_{opt}, the pilot spacing L is chosen such that the lower bound for separate processing is maximized; rectangular $S_h(f)$ (24)

Fig. 1 shows a comparison of the bounds on the achievable rate for separate and joint processing of data and pilot symbols. On the one hand, the lower bound on the achievable rate for joint processing in (25) is compared to bounds on the achievable rate with separate processing of data and pilot symbols for a fixed pilot spacing, i.e., [4,(22)] and [4,(23)] for zero-mean proper Gaussian data symbols. As the upper and lower bound on the achievable rate with separate processing are relatively tight, we choose the pilot spacing such that the lower bound on the achievable rate for separate processing in [4,(22)] is maximized. Except of very large f_d the lower bound on the achievable rate for joint processing is larger than the bounds on the achievable rate with separate processing. This indicates the possible gain while using joint processing of data and pilot symbols for a given pilot spacing. The observation that the lower bound for joint processing for very large f_d is smaller than the achievable rate with separate processing indicates that the lower bound is not tight for these parameters.

On the other hand, also the lower bound on the achievable rate with joint processing and a pilot spacing that maximizes this lower bound, i.e., (25) with (26), is shown. Obviously, this lower bound is larger than or equal to the lower bound for joint processing while choosing the pilot spacing as it is optimal for separate processing of data and pilot symbols. This behavior arises from the effect that for separate processing in case of small f_d a pilot rate is chosen that is higher than the Nyquist rate of the channel fading process to enhance the channel estimation quality. In case of a joint processing all symbols are used for channel estimation anyway. Therefore, a pilot rate higher than Nyquist rate always leads to an increased loss in the achievable rate as less symbols can be used for data transmission.

References

[1] M. C. Valenti and B. D. Woerner, “Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697–1705, Sep. 2001.

[2] L. Schmitt, H. Meyr, and D. Zhang, “Systematic design of iterative ML receivers for flat fading channels,” *IEEE Trans. Commun.*, submitted.

[3] J. Doob, *Stochastic Processes*. New York: Wiley, 1990.

[4] J. Baltersee, G. Fock, and H. Meyr, “An information theoretic foundation of synchronized detection,” *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.

[5] N. Jindal, A. Lozano, and T. Marzetta, “What is the value of joint processing of pilots and data in block-fading channels?” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009.

[6] M. Dörpinghaus, A. Ispas, G. Ascheid, and H. Meyr, “On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels,” in preparation.

[7] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA, U.S.A.: Univ. Calif. Press, 1958.

[2] Note that a rectangular PSD $S_h(f)$ corresponds to $r_h(l) = \sigma_h^2\text{sinc}(2f_d l)$ which is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a raised cosine shape, whose corresponding correlation function is absolutely summable.

[3] Periodically inserted pilot symbols do not maximize the achievable rate. However, we restrict to periodical pilot symbols with a spacing fulfilling (3), as this enables detection with manageable complexity.

Outage Analysis of Asynchronous OFDM Non-orthogonal DF Cooperative Networks

Mehdi Torbatian and Mohamed Oussama Damen

Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, Canada
Email: {m2torbat, modamen}@uwaterloo.ca

***Abstract*—Outage behavior of non-orthogonal[1] selection decode-and-forward (NSDF) relaying protocol over an *asynchronous* cooperative network is examined when orthogonal frequency division multiplexing (OFDM) is used to combat synchronization error among the transmitting nodes. It is proved that the asynchronous protocol provides diversity gain greater than or equal to the one of the corresponding synchronous counterpart, synchronous NSDF, in the limit of code word length and throughout the range of multiplexing gain.**

I. Introduction

Cooperative diversity was first proposed as a synchronous technique [1], [2] to provide spacial diversity with the help of surrounding terminals; however, because relays are at different locations (i.e., different propagation delays) and they have their own local oscillators with no common timing reference, it is an *asynchronous* technique in nature.

To combat the synchronization error, two major approaches have been proposed: delay tolerant space-time schemes (see [3], [4] and references therein), and OFDM [5]. While it is usually assumed in the former schemes that asynchronous delays are integer factor of the symbol interval, OFDM allows the delays to be any real number. In [6], the effect of the synchronization error on diversity multiplexing gain tradeoff (DMT) [7] of an orthogonal decode-and-forward (DF) cooperative network with two relays is examined when the maximum possible relative delay between the relays is less than a symbol interval. In [8], authors show that by allowing the source and the relays to transmit over proper portions of a cooperative frame, the better diversity gain can be achieved for each multiplexing gain.

In this paper, we analyze the outage behavior of NSDF protocol over a general two-hop relay network when OFDM is used to offset the synchronization error among transmitting nodes. In contrast to [6], we do not restrict the relative delays to be less than a symbol interval. In addition, we let the source and the relays to transmit over non-symmetric portions of a cooperative frame to maximize the diversity gain at each multiplexing gain. It is proved that the asynchronous protocol outperforms the synchronous counterpart in the limit of code word length and throughout the range of the multiplexing gain.

In the following, the system model and the required background are presented. DMT analysis of the asynchronous OFDM NSDF relaying protocol is detailed afterward. The paper is concluded at the end.

II. Preliminaries

A. Notations, Assumptions, and Definitions

In this work, letters with underline $\underline{x}, \underline{X}$ denote vectors, and boldface uppercase letters $\mathbf{X}$ denote matrices. The superscripts $(\cdot)^T$ and $(\cdot)^\dagger$ denote the transpose and conjugate transpose of the corresponding vector or matrix, respectively. $\mathbf{I}_n$ is the identity matrix of dimension n. diag$\{\cdot\}$ indicates a diagonal or a block diagonal matrix of its arguments. The symbol $\otimes$ indicates the Kronecher product. $\doteq$ is used to show the exponential equality. For example, $f(\rho) \doteq \rho^b$ if $\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = b$. $(x)^+$ is considered as $\max\{0, x\}$.

We assume half-duplex signal transmission. All channels are assumed to be quasi-static. They are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance $\mathbb{C}\mathcal{N}(0,1)$. Each node knows channel state information (CSI) of its incoming links. The destination also knows the asynchronous delays of its incoming links.

Define $\{\mathcal{C}(\rho)\}$ as a family of variable rate codes each of them is used at the corresponding signal to noise ratio ρ. This family of codes is said to achieve the multiplexing gain r and the diversity gain $d(r)$ if

$$\lim_{\rho\to\infty} \frac{R(\rho)}{\log \rho} = r, \qquad \lim_{\rho\to\infty} \frac{\log P_e(\rho)}{\log \rho} = -d(r), \tag{1}$$

where $R(\rho)$ is the rate and $P_e(\rho)$ is the average error probability of the code $\mathcal{C}(\rho)$. The outage diversity is obtained by replacing $P_e(\rho)$ with the outage probability $P_{\mathcal{O}}$ in the above formula. It is proved that the outage diversity is a tight upper bound for the diversity gain of a coding scheme [7].

B. System Model

We consider a network containing one source node, one destination node, and M relay nodes as shown in Fig. 1. h_i and g_i are fading coefficients represent the links from the i-th transmitting node to the destination and from the source to the i-th relay, respectively. Communication between the source and the destination is carried out in two phases. First, the

[1] A relaying protocol is called orthogonal if the source and relays transmit in two non-overlapping intervals; otherwise, it is called non-orthogonal.

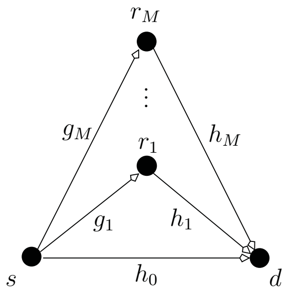


Fig. 1. System structure

source broadcasts its message to relays and the destination in p channel uses. Second, those relays that can fully decode the source message retransmit it to the destination in q channel uses. Assuming ℓ is the length of a cooperative frame, $\ell = p + q$. The cooperation is avoided whenever it is beneficial to do so. In this case, the source transmits to the destination without help of the relays. Each node is supported by an i.i.d. Gaussian random code book which is independent from the other nodes' code books. The source's transmitted signal in the first phase is given by

$$x_0'(t) = \sum_{i=0}^{p-1} x_0'(i)\mathrm{g}_0(t - iT_s), \tag{2}$$

where $\underline{x}_0' = [x_0'(0), x_0'(1), \ldots, x_0'(p-1)]^T$ is the transmitted code word corresponding to the source message, T_s is the symbol interval, and $\mathrm{g}_0(t)$ is a unit energy shaping waveform with non-zero duration T_s over $t \in [0, T_s]$. The received signals at the destination and the i-th relay are modeled by

$$\begin{aligned} y_d(t) &= h_0 x_0'(t) + z_d(t), \quad 0 \leq t \leq pT_s, & (3)\\ y_{r_i}(t) &= g_i x_0'(t) + z_{r_i}(t), \quad 0 \leq t \leq pT_s, & (4)\end{aligned}$$

where $z_d(t)$ and $z_{r_i}(t)$ are additive noises at the destination and the i-th relay modeled by white Gaussian noises with zero mean and variances σ_d^2 and σ_r^2, respectively.

Let $\mathcal{D}$ be a set containing index of the nodes participating in the second phase (not in outage). As the relaying protocol is non-orthogonal, $\mathcal{D}$ contains 0, index of the source. Similarly, the i-th relay, $i \in \mathcal{D}$, uses a unit energy shaping waveform $\mathrm{g}_i(t)$ with nonzero duration T_s to transmit its code words of length q in the second phase. This signal is received at the destination by τ_i second delay with reference to the first received signal. τ_is are finite values less than or equal to τ_{max} which is the maximum amount of asynchronous delay. Without loss of generality, we assume that the source signal is the earliest received signal at the destination, and the delays of the other received signals are measured with reference to this signal, i.e., $\tau_0 = 0$.

Let $x_i(t)$ be the transmitted signal by the i-th node. The received signal at the destination is modeled by

$$y_d(t) = \sum_{i \in \mathcal{D}} h_i x_i(t - \tau_i) + z_d(t), \tag{5}$$

$y_d(t)$ is processed through parallel matched filters corresponding to the transmitting links. The output of the i-th matched filter sampled at $t = (k+1)T_s + \tau_i$, is given by

$$y_{d_i}(k) = \int_{kT_s+\tau_i}^{(k+1)T_s+\tau_i} y_d(t)\mathrm{g}_i^*(t - kT_s - \tau_i)dt. \tag{6}$$

C. Asynchronous OFDM Space-Time Codes

In our work, OFDM is used to combat the synchronization error. Assume that the i-th node participates in the second phase, i.e., $i \in \mathcal{D}$. Its code word of length n, $\underline{x}_i$, is first passed through an inverse discrete Fourier Transform (IDFT) filter, $\mathrm{IDFT}\{\underline{x}_i\} = \underline{X}_i$, and then supported by a cyclic prefix (CP) of length $u = \lceil \frac{\tau_{max}}{T_s} \rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than x, to produce $\underline{X}_i^{cp}$ of length $q = n + u$. The received signal at the destination is given by

$$Y_d(t) = \sum_{i \in \mathcal{D}} h_i \sum_{j=0}^{q-1} X_i^{cp}(j)\mathrm{g}_i(t - jT_s - \tau_i) + Z_d(t), \tag{7}$$

where $X_i^{cp}(j)$ is the j-th entry of $\underline{X}_i^{cp}$. For $i \geq j,\ i, j \in \mathcal{D}$, define the relative delay τ_{ij} as

$$\tau_{ij} \triangleq \tau_i - \tau_j. \tag{8}$$

As $i \geq j$, then $\tau_{i,j} \geq 0$. The fractional delay $\tilde{\tau}_{ij}$ is defined as

$$\tilde{\tau}_{ij} \triangleq \tau_{ij} - a_{ij}T_s, \tag{9}$$

where $a_{ij} = \lfloor \frac{\tau_{ij}}{T_S} \rfloor \geq 0$, with $\lfloor x \rfloor$ denoting the largest integer smaller than or equal to x, and $0 \leq \tilde{\tau}_{ij} < T_s$.

III. Asynchronous OFDM NSDF Protocol

A. Signal Model

Let E_m, the event of any m relays participates in the second phase, occurs. E_0 corresponds to the case that only the source node transmits in the second phase. $\mathcal{D} = \{0, 1, 2, \ldots, m\}$ is the index set pointing out to participating nodes in the second phase. Without loss of generality, we assume that $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots \leq \tau_m$. The sampled signal at the output of the i-th matched filter $(i = 0, 1, \ldots, m)$ is modeled by [9]

$$\begin{aligned} &Y_{d,i}(k) = h_i X_i^{cp}(k) + Z_{d,i}(k) + \\ &\sum_{j=0}^{i-1} h_j[X_j^{cp}(k + a_{ij} + 1)B_{ij}^* + X_j^{cp}(k + a_{ij})C_{ij}^*] + \\ &\sum_{j=i+1}^{m} h_j[X_j^{cp}(k - a_{ji} - 1)B_{ji} + X_j^{cp}(k - a_{ji})C_{ji}], \end{aligned} \tag{10}$$

where $Y_{d,i}(k)$ is the k-th entry of the output of the i-th matched filter, and for $i \geq j$, $B_{ij} = \int_0^{T_s} \mathrm{g}_i(t + T_s - \tilde{\tau}_{ij})\mathrm{g}_j^*(t)dt$, $C_{ij} = \int_0^{T_s} \mathrm{g}_i(t - \tilde{\tau}_{ij})\mathrm{g}_j^*(t)dt$. Define

$$\alpha_{ij}(k) \triangleq [C_{ij} + B_{ij}e^{-j\frac{2\pi}{n}k}]e^{j\frac{2\pi}{n}k\tilde{a}_{ij}}, \tag{11}$$

where $\tilde{a}_{ij} = 0$ when $\tilde{\tau}_{i0} \geq \tilde{\tau}_{j0}$, and $\tilde{a}_{ij} = 1$ when $\tilde{\tau}_{i0} < \tilde{\tau}_{j0}$. It can be checked that, for $j > i, \alpha_{ij}(k) = \alpha_{ji}^*(k)$, $k = 0, 1, \ldots, n-1$ and $i, j = 0, 1, \ldots, m$. Let

$$\mathbf{D}_{ij} = \mathrm{diag}\{\alpha_{ij}(0), \alpha_{ij}(1), \ldots, \alpha_{ij}(n-1)\}, \tag{12}$$

$$\mathbf{E}_i = \mathrm{diag}\{1, e^{-j\frac{2\pi}{n}i}, \ldots, e^{-j\frac{2\pi}{n}(n-1)i}\}. \tag{13}$$

At the output of each matched filter CP is discarded. The result is then passed through a Discrete Fourier Transform (DFT) filter. The outputs can be written in a matrix form as

$$\underline{y} = \mathbf{H}\underline{x} + \underline{z}, \tag{14}$$

where

$$\begin{aligned}
\underline{x} &= \left[\underline{x}_0^T \ \underline{x}_1^T \ \ldots \ \underline{x}_m^T\right]^T \\
\underline{y} &= \left[\underline{y}_{d,0}^T \ \left(\mathbf{E}_1^{\dagger}\underline{y}_{d,1}\right)^T \ \ldots \ \left(\mathbf{E}_1^{\dagger}\underline{y}_{d,m}\right)^T\right]^T, \\
\underline{z} &= \left[\underline{z}_{d,0}^T \ \left(\mathbf{E}_1^{\dagger}\underline{z}_{d,1}\right)^T \ \ldots \ \left(\mathbf{E}_1^{\dagger}\underline{z}_{d,m}\right)^T\right]^T, \\
\mathbf{H} &= \mathbf{\Xi}(\mathbf{I}_n \otimes \hat{\mathbf{H}})\mathbf{U}. \qquad (15)
\end{aligned}$$

$\mathbf{U} = \text{diag}\{\mathbf{I}_n, \mathbf{E}_{a_{10}}, \ldots, \mathbf{E}_{a_{m0}}\}$, $\hat{\mathbf{H}} = \text{diag}\{h_0, h_1, \ldots, h_m\}$, and

$$\mathbf{\Xi} = \begin{bmatrix} \mathbf{I}_n & \mathbf{D}_{10} & \mathbf{D}_{20} & \ldots & \mathbf{D}_{m0} \\ \mathbf{D}_{10}^{\dagger} & \mathbf{I}_n & \mathbf{D}_{21} & \ldots & \mathbf{D}_{m1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{D}_{m0}^{\dagger} & \mathbf{D}_{m1}^{\dagger} & \mathbf{D}_{m2}^{\dagger} & \ldots & \mathbf{I}_n \end{bmatrix}. \tag{16}$$

Equation (14) represents a simple multiple-input multiple-output (MIMO) channel model with correlated noise vector $\underline{z}$ for the underlying system. The covariance matrix of $\underline{z}$ is calculated as [9]

$$\mathbf{\Phi} = n\sigma_d^2 \ \mathbf{\Xi}. \tag{17}$$

Clearly, $\mathbf{\Phi}^{-1}$ exists if and only if $\mathbf{\Xi}^{-1}$ exists.

Proposition 1: $\mathbf{\Xi}$ is semi-positive definite. i.e., $\det \mathbf{\Xi} \geq 0$. The equality holds if and only if $\exists \ \underline{c} \in \mathbb{C}^{1\times m}, \exists \ k \in \{0, \ldots, n-1\}$ such that [9].

$$\left(\sum_{i=0}^{1} \underline{\mathrm{g}}(t + iTs)e^{-j\frac{2\pi}{n}ki}\right)\underline{c}^{\dagger} = 0, \quad \forall t \in [0, T_s], \tag{18}$$

where $\underline{\mathrm{g}}(t) \triangleq [\mathrm{g}_0(t), \mathrm{g}_1(t-\tilde{\tau}_{10}), \mathrm{g}_2(t-\tilde{\tau}_{20}), \ldots, \mathrm{g}_m(t-\tilde{\tau}_{m0})]$, and $\mathbb{C}$ is the field of complex numbers.

B. DMT Analysis

The outage probability $P_{\mathcal{O}}$ is calculated as follows.

$$P_{\mathcal{O}} \ = \ \sum_{m=0}^{M} Pr(I_{E_m} < R \mid \ E_m)Pr(E_m),$$

where I_{E_m} is the mutual information between the source and the destination when E_m occurs.

Lemma 1: $Pr(E_m)$ is given by [9]

$$Pr(E_m) \doteq \begin{cases} \rho^{-(1-\frac{\ell r}{p})(M-m)}, & 0 \leq r \leq \frac{p}{\ell}, \\ 0, & \frac{p}{\ell} < r \leq 1, \ 1 \leq m \leq M \\ 1, & \frac{p}{\ell} < r \leq 1, \ m = 0. \end{cases} \tag{19}$$

When E_m occurs, the mutual information between the source and the destination is given by [9]

$$\begin{aligned} I_{E_m} \ &= \ \frac{p}{\ell}\log(1 + \rho|h_0|^2) + \\ &\quad \frac{1}{\ell}\log\det\left(\mathbf{I}_{(m+1)n} + n\mathcal{E}\mathbf{H}\mathbf{H}^{\dagger}\mathbf{\Phi}^{-1}\right), \qquad (20) \end{aligned}$$

where the first and the second terms on the right hand side are the resulted mutual information between the transmitting nodes and the destination, respectively, in the first and the second phases. Define $\mathcal{A} \triangleq \mathbf{I}_{(m+1)n} + n\mathcal{E}\mathbf{H}\mathbf{H}^{\dagger}\mathbf{\Phi}^{-1}$. By substituting (15) and (17) into (20) and considering the fact that $\mathbf{U}$ is a Hermitian matrix, we have

$$\det \mathcal{A} = \det\left(\mathbf{I}_{(m+1)n} + \rho\mathbf{\Xi}(\mathbf{I}_n \otimes \hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger})\right). \tag{21}$$

$\mathbf{\Xi}$ can be decomposed as $\mathbf{\Xi} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\dagger}$, where $\mathbf{V}$ is a unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix containing eigenvalues of $\mathbf{\Xi}$ on its main diagonal. By assuming proper design of the shaping waveforms, all eigenvalues of $\mathbf{\Xi}$ are finite positive values bounded from zero. Hence, their ρ exponents at high SNR regime is zero. By replacing all the eigenvalues by the smallest one, say ξ, the mutual information between the source and the destination is lower bounded. Since the ρ exponent of ξ is zero, this bound is tight. We have,

$$\begin{aligned} \det \mathcal{A} \ &\doteq \ \det\left(\mathbf{I}_{(m+1)n} + \rho\xi(\mathbf{I}_n \otimes \hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger})\right) \\ &\doteq \ \prod_{i=0}^{m}(1 + \rho|h_i|^2)^n. \end{aligned}$$

Define $\gamma_i \triangleq -\frac{\log|h_i|^2}{\log\rho}$. For large values of ρ, $(1 + \rho|h_i|^2) \simeq \rho^{(1-\gamma_i)^+}$. After some mathematical manipulation, we obtain

$$I_{E_m} = \left[\frac{p+n}{\ell}(1-\gamma_0)^+ + \frac{n}{\ell}\sum_{i=1}^{m}(1-\gamma_i)^+\right]\log\rho. \tag{22}$$

As can be seen, the resulted mutual information among the transmitting nodes and the destination behaves similar to the one of a parallel channel with (m+1) independent links. $P_{\mathcal{O}|E_m}$ is obtained as follows [9].

$$P_{\mathcal{O}|E_m} \ = \ P\left(I_{E_m} < R\right) \ \doteq \ \rho^{-d_{E_m}(r)}$$

where

$$d_{E_m}(r) = \inf_{\frac{p+n}{\ell}(1-\gamma_0)^+ + \frac{n}{\ell}\sum_{i=1}^{m}(1-\gamma_i)^+ < r} \ \sum_{i=0}^{m}\gamma_i. \tag{23}$$

By solving the above optimization problem, we have [9]

Lemma 2:

$$d_{E_m}(r) = \begin{cases} 1 + m - \frac{\ell}{n}r, & 0 \leq r \leq \frac{mn}{\ell}, \\ 1 + \frac{mn}{p+n} - \frac{\ell}{p+n}r, & \frac{mn}{\ell} < r \leq \frac{p+n}{\ell}. \end{cases}$$

Define $\kappa \triangleq \frac{p}{n}$. When $m \geq \kappa + 1$, then $\frac{mn}{\ell} \geq \frac{p+n}{\ell}$. Hence,

$$d_{E_m}(r) = 1 + m - \frac{\ell}{n}r, \quad 0 \leq r \leq \frac{p+n}{\ell}.$$

For the single relay network, theorem 1 concludes the results.

Theorem 1: DMT of the asynchronous OFDM NSDF protocol over the single relay cooperative network for a fix $\kappa \geq 1$ is as follows [9].

If $\quad 1 \leq \kappa \leq \hat{\kappa}$

$$d(r) = \begin{cases} (1 - \frac{\ell}{p}r) + (1 - \frac{\ell}{p+n}r), & 0 \leq r \leq \eta_1 \\ 1 - r, & \eta_1 \leq r \leq 1, \end{cases}$$

else if $\quad \kappa \geq \hat{\kappa}$

$$d(r) = \begin{cases} 2(1-\frac{\ell}{2n}r), & 0 \le r \le \eta_2 \\ 1+\frac{n}{p+n}-\frac{\ell}{p+n}r, & \eta_2 \le r \le \eta_3 \\ (1-\frac{\ell}{p}r)+(1-\frac{\ell}{p+n}r), & \eta_3 \le r \le \eta_1 \\ 1-r, & \eta_1 \le r \le 1, \end{cases}$$

where $\hat{\kappa} = \frac{1+\sqrt{5}}{2}$, $\eta_1 = \frac{(p+n)p}{(2p+n)\ell-(p+n)p}$, $\eta_2 = \frac{n}{\ell}$, and $\eta_3 = \frac{p^2}{(p+n)\ell}$. For the case that κ varies to maximize the diversity gain, for large length code words we have

$$d(r) = \begin{cases} [1-(1+\frac{1}{\hat{\kappa}})r]+(1-r), & 0 \le r \le \frac{1}{\hat{\kappa}+1} \\ (1-\sqrt{r})+(1-r), & \frac{1}{\hat{\kappa}+1} \le r \le 1. \end{cases}$$

The optimum κ corresponding to each r is given by

$$\kappa = \begin{cases} \hat{\kappa}, & 0 \le r \le \frac{1}{\hat{\kappa}+1} \\ \frac{\sqrt{r}}{1-\sqrt{r}}, & \frac{1}{\hat{\kappa}+1} \le r \le 1. \end{cases}$$

Fig. 2 depicts the DMT curves of the asynchronous OFDM NSDF and the corresponding synchronous protocol over a single relay network when κ varies to maximize the diversity gain at each multiplexing gain r. As can be seen, the DMT performance of the asynchronous protocol performs is the same as that of the synchronous one in low multiplexing gains and is better than that in high multiplexing gains.

Calculating DMT in a general network with any number of relays, say M, is straightforward. However, because too many regions for r and κ should be considered, it is cumbersome. Alternatively, this procedure is easier if we assume that DMT of a simpler network containing $(M-1)$ relays is known. Let $d^M(r)$ be the DMT of an M relay cooperative network when the cooperation is not avoided throughout the range of the multiplexing gain. We have,

Theorem 2: DMT of the asynchronous OFDM NSDF relaying protocol over a general two-hop cooperative network with M relays for a fix $\kappa \ge 1$ is as follows [9].

$$\text{If} \quad \kappa \le \frac{M+\sqrt{M^2+4M}}{2}$$

$$d^M(r) = \begin{cases} (1-\frac{\ell}{p}r)+d^{M-1}(r), & 0 \le r \le \frac{p}{\ell} \\ 1-\frac{\ell}{p+n}r, & \frac{p}{\ell} \le r \le \frac{p+n}{\ell}, \end{cases}$$

$$\text{else if} \quad \kappa > \frac{M+\sqrt{M^2+4M}}{2}$$

$$d^M(r) = \begin{cases} (1-\frac{\ell}{p}r)+d^{M-1}(r), & 0 \le r \le \eta_1 \\ 1+M-\frac{\ell}{n}r, & \eta_1 \le r \le \eta_2 \\ 1+\frac{Mn}{p+n}-\frac{\ell}{p+n}r, & \eta_2 \le r \le \eta_3 \\ M(1-\frac{\ell}{p}r)+1-\frac{\ell}{p+n}r, & \eta_3 \le r \le \eta_4 \\ 1-\frac{\ell}{p+n}r, & \eta_4 \le r \le \eta_5, \end{cases}$$

where $\eta_1 = \frac{(M-1)p^2n}{\ell(p^2-np-n^2)}$, $\eta_2 = \frac{Mn}{\ell}$, $\eta_3 = \frac{p^2}{\ell(p+n)}$, $\eta_4 = \frac{p}{\ell}$, and $\eta_5 = \frac{p+n}{\ell}$. The resulted DMT is compared to $(1-r)$ to determine wether or not avoiding the cooperation. When κ is allowed to vary to maximize the diversity gain at each multiplexing gain r, for large length code words we have

$$d(r) = \begin{cases} M[1-(1+\frac{1}{\hat{\kappa}})r]+(1-r), & 0 \le r \le \frac{1}{1+\hat{\kappa}} \\ M(1-\sqrt{r})+(1-r), & \frac{1}{1+\hat{\kappa}} \le r \le 1. \end{cases}$$

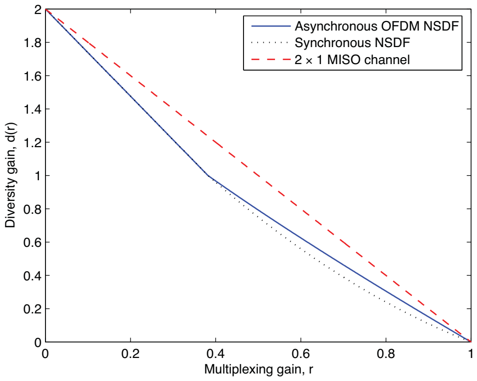


Fig. 2. DMT of the asynchronous OFDM NSDF and the synchronous NSDF protocols over a single relay network with optimum values of κ.

where $\hat{\kappa} = \frac{1+\sqrt{5}}{2}$. The corresponding optimum κ is the same as that of the single relay network.

For $M \ge 2$ the resulted DMT is always better than that of the corresponding synchronous protocol. DMT of the asynchronous orthogonal selection DF (OSDF) relaying protocol is calculated in a similar manner [9].

IV. Conclusion

DMT of the asynchronous OFDM NSDF protocol over a general one-hop cooperative network was examined. It was shown that asynchronous delays among transmitting nodes not only decrease the diversity gain, but also increase it particularly at high multiplexing gains for large length code words.

References

[1] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity-part I: system description,” *IEEE Trans. Commun.*, vol. 51, Issue 11, pp. 1927-1938, Nov. 2003.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity Part II: Implementation aspects and performance analysis,” *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939-1948, Nov. 2003.

[3] M. Torbatian and M. O. Damen, “On the design of delay-tolerant distributed space-time codes with minimum length,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 931-939, Feb. 2009.

[4] Y. Shang and X.G. Xia, “Shift-full-rank matrices and applications in space-time trellis codes for relay networks with asynchronous cooperative diversity,” *IEEE Trans. Info. Theory*, vol. 52, Issue 7, pp. 3-7, July 2006.

[5] Y. Mei, Y. Hua, A. Swami, and B. Daneshrad, “Combating synchronization errors in cooperative relays,” in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing*, Philadelphia, PA, USA, pp. 1-6, Mar. 2005.

[6] Shuangqing Wei, "Diversity multiplexing tradeoff of asynchronous cooperative diversity in wireless networks", *IEEE Trans. Info. Theory*, vol. 53, pp. 0-2, issue. 11, Nov. 2007.

[7] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,"*IEEE Trans. Info. Theory*, vol. 49, May 2003.

[8] P. Elia, K. Vinodh, M. Anand, P. Vijay Kumar, "D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols," submitted to *IEEE Trans. Info. Theory*, July 2006.

[9] M. Torbatian and M. O. Damen, “On the Outage Behavior of Asynchronous OFDM DF and AF Cooperative Networks,” Submitted to *IEEE Trans. Info. Theory* , August 2009.

Multi-Layer Coded Direct Sequence CDMA

Avi Steiner, Shlomo Shamai (Shitz)
Technion—IIT, Haifa 32000, Israel
Department of Electrical Engineering
Email: {savi@tx,sshlomo@ee}.technion.ac.il

Valentin Lupu, Uri Katz
Rafael, Israel
Email: {katzu,valentin}@rafael.co.il

***Abstract*—We consider the problem of multi-user detection for randomly spread direct-sequence (DS) coded-division multiple access (CDMA) over flat fading channels. The analysis focuses on the case of many users, and large spreading sequences such that their ratio, defined as the system load, is kept fixed. Single layer and multi layer coding are analyzed in this setup. The spectral efficiency, for linear multiuser detectors, is derived for different decoding strategies. Iterative decoding of multi-layered transmission with successive interference cancellation (SIC) is optimized, and the optimal layering power distribution is obtained. For small system loads, the achievable spectral efficiency with the broadcast approach and a matched filter detector exhibits significant gains over single layer coding.**

I. Introduction

Consider the case of many users, and large spreading sequences, such that the system load β is kept fixed. That is, $K, N \to \infty$, and $\beta = K/N$, where K denotes the number of users, and N is the spreading sequence length. The spectral efficiency of direct sequence (DS) CDMA with random spreading for this regime over fading multiaccess channels is studied in [1]. In that contribution, ergodic spectral efficiency is studied, assuming that all users are reliably decoded regardless of their received powers. The assumption is that users can adjust their rates according to their experienced fading level, using, for example, an instantaneous feedback from the receiver. Unfortunately, such a feedback and ideal tuning of transmission rates are not always feasible. Thus, these results can be achieved only on fast fading channels, where sufficient fading statistics is observed over a single transmission block. Motivated by practical considerations, decoding of strongest users on block fading channels is studied in [2]. In this work, it is assumed that all users transmit at equal rate and equal power. In this case the receiver can no longer guarantee reliable decoding of all active users. As a result, the receiver ranks all active users by their received power and decodes the transmission of the largest number of users, for which decoding is successful. The maximal expected sum rate is referred to as the outage capacity.

In this work, we first derive the spectral efficiency of successive interference cancellation (SIC) detectors with iterative decoding. The main idea here is to keep on trying to decode users after every SIC stage, as the residual interference is reduced every iteration, which decreases the effective system load during decoding. This concept is adopted for multi-layer multiuser successive decoding, where the optimal power distribution is derived for maximizing the achievable expected spectral efficiency.

II. Channel Model and Preliminaries

We describe here the channel model and the basic assumptions. Consider the following system model,

$$\mathbf{y} = \mathbf{V}\mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} = [x_1, ..., x_K]$ is a vector of length K. An individual term x_k is a sample of a layered coded signal of the k^{th} user, and $\{x_k\}$ are i.i.d. $\{x_k\} \sim \mathcal{CN}(0, P)$, where P sets the power constraint per user. $\mathbf{V}$ is an $[N \times K]$ signature matrix (i.i.d. with elements $v_{i,j} \sim \mathcal{CN}(0, \frac{1}{N})$, and $\mathbf{n}$ is, without loss of generality, a normalized AWGN vector $\mathbf{n} \sim \mathcal{CN}(0, I_N)$. The channel matrix $\mathbf{H}$ is a diagonal matrix $\mathbf{H} = \text{diag}(h_1, h_2, ..., h_K)$ of fading gains, which empirical distribution of $\{s_k\} \triangleq \{|h_k|^2\}$ converges a.s. to a distribution $F_s(s)$ such that $E_{F_s}[s] = 1$. The channel matrix $\mathbf{H}$ remains fixed throughout a transmission block, which corresponds to a slowly fading channel model. Note that, since the additive noise is normalized, $\text{SNR} = P$.

The energy per bit to noise spectral density ratio is used for evaluation of the spectral efficiency and comparison of different strategies. Its definition is

$$\frac{E_b}{N_o} = \frac{\beta}{R_{sum}}\text{SNR} \tag{2}$$

where R_{sum} is the total spectral efficiency, i.e. the sum-rate in bits per second per Hertz.

It is well known that the spectral efficiency of the optimal multiuser detector is achievable with a minimum mean square error (MMSE) detector, with successive decoding and cancellation. It is therefore interesting to study the spectral efficiency gain with successive decoding and practical linear detectors such as matched filter or decorrelator.

For a system load $\beta = \frac{K}{N}$, the ergodic sum-rate is [1],

$$C(\beta, \text{SNR}) = \lim_{K,N\to\infty} \beta E_s \left\{\log(1 + s \cdot \eta(\beta)\text{SNR})\right\} \tag{3}$$

where the expectation is taken w.r.t. the fading gain distribution $F_s(s)$. The ergodic sum-rate is an upper bound, since its achievability requires an instantaneous feedback from receiver to all users. With SIC decoding, after every decoding stage the subtraction of users decreases the effective system load, therefore

$$C_{SIC} = E_s \left\{ \lim_{K,N\to\infty} \sum_{j=0}^{K-1} \frac{1}{N} \log\left(1 + s\eta\left(\frac{K-j}{N}\right)\text{SNR}\right)\right\}$$

which converges a.s. to the following integral expression,

$$C_{SIC,erg}(\beta,\text{SNR}) = E_s \left\{ \int\limits_0^\beta dz \, \log\left(1 + s \cdot \eta\left(z\right)\text{SNR}\right) \right\}. \quad (4)$$

Since the MMSE with SIC achieves optimal receiver performance, we focus on MF and decorellator detectors for spectral efficiency analysis. A matched filter detector efficiency is [1]

$$\eta_{mf}(\beta) = \frac{1}{1+\beta\text{SNR}}. \quad (5)$$

For a decorrelator detector, we use a similar derivation. The detector efficiency is [1]

$$\eta_{dec}(\beta) = \max(0, 1-\beta). \quad (6)$$

III. Strongest Users Detection and SIC

Consider the case that all users transmit the same rate R, using a single layer code,

$$R \triangleq \log(1 + s_{th}\eta\left(\beta\right)\text{SNR}). \quad (7)$$

where s_{th} is a rate allocation parameter which governs the fading gain threshold for reliable decoding. The probability of outage, in parallel decoding, is $F_s(s_{th})$. The achievable rate at the first SIC stage is obtained by decoding in parallel all users that are not in outage (with single user detectors). Hence,

$$R_0(s_{th},\beta) = \beta(1 - F_s(s_{th}))R \quad (8)$$

and after cancelling all the reliably decoded users, there is a fraction $\beta F_s(s_{th})$ of undecoded users. The mutual interference reduces after cancellation, and there may exist more users with fading gains $s < s_{th}$ who can now be decoded. The additional rate, obtainable at the next stage, is given by

$$R_1(s_{th},\beta) = \beta(F_s(s_{th}) - F_s(s_1))R \quad (9)$$

which expresses the expected sum-rate for parallel decoding of all users with fading levels $s_1 \le s < s_{th}$, where s_1 satisfies

$$s_1\eta\left(\beta F_s(s_{th})\right) = s_{th}\eta\left(\beta\right). \quad (10)$$

This procedure continues similarly to the next stage. We can express the total achievable rate as follows

$$R_{out} = \beta\sum_{n=0}^{\infty}(F_s(s_{n-1}) - F_s(s_n)) \ \cdot R = \beta(1 - F_s(s_\infty))R \ (11)$$

where $s_0 \triangleq s_{th}$, and $F_s(s_{-1}) = 1$, and

$$s_n = s_{th}\frac{\eta\left(\beta\right)}{\eta\left(\beta F_s(s_{n-1})\right)}, \qquad n = 1,2,... \quad (12)$$

It can be shown [3] that there exists a limit $0 \le s_\infty \le s_{th}$ for the linear detectors, since $F_s(s)$ is a monotonically non increasing function, and $\eta(\beta)$ is a monotonically decreasing function. Hence s_∞ satisfies the following condition,

$$s_\infty\eta\left(\beta F_s(s_\infty)\right) = s_{th}\eta\left(\beta\right). \quad (13)$$

IV. Two Coded Layers

A higher expected spectral efficiency may be obtained with coded layering at the transmitter for each user. We begin here with analysis of two coded layers for a MF detector at the receiver. Let every user use the following rate allocation

$$\begin{aligned} R_1 &= \log\left\{1 + \frac{s_1^{(1)}\alpha\text{SNR}\eta_{mf}(\beta_1)}{1+s_1^{(1)}\overline{\alpha}\text{SNR}\eta_{mf}(\beta_1)}\right\} \\ R_2 &= \log\left\{1 + s_2^{(1)}\overline{\alpha}\text{SNR}\eta_{mf}\left(\beta_2\right)\right\} \end{aligned} \quad (14)$$

where $s_1^{(1)}, s_2^{(1)}$ are the layering fading gain thresholds, and $\beta_1 \triangleq \overline{\alpha}\beta + \alpha\beta F_s(s_1^{(1)})$, $\beta_2 \triangleq \overline{\alpha}\beta F_s(s_2^{(1)}) + \alpha\beta F_s(s_1^{(1)})$ and αSNR, $\overline{\alpha}$SNR result from the power allocated to the first and second layers, respectively. Note $\overline{\alpha} \triangleq 1-\alpha$, and $\alpha \in [0,1]$. The iterative decoding steps are as follows: 1) Decode, using SIC, the first layer of all decodable users; 2) Repeat the previous step for the next layer; 3) Repeat 1)-2) until there are no more decodable users. The reduced system load after the first iteration is a direct result of (11) and (13) for a single layer. The iterative SIC decoding converges a.s. to the following expected spectral efficiency

$$R_{2L} = \beta\left(1 - F_s\left(s_1^{(\infty)}\right)\right)R_1 + \beta\left(1 - F_s\left(s_2^{(\infty)}\right)\right)R_2 \ (15)$$

where $\{s_i^{(\infty)}\}$ satisfy the following conditions

$$\begin{aligned} s_1^{(\infty)}\eta_{mf}\left(\overline{\alpha}\beta F_s\left(s_2^{(\infty)}\right) + \alpha\beta F_s\left(s_1^{(\infty)}\right)\right) &= s_1^{(1)}\eta_{mf}\left(\beta_1\right) \\ s_2^{(\infty)}\eta_{mf}\left(\overline{\alpha}\beta F_s\left(s_2^{(\infty)}\right) + \alpha\beta F_s\left(s_1^{(\infty)}\right)\right) &= s_2^{(1)}\eta_{mf}\left(\beta_2\right) \end{aligned} \quad (16)$$

A detailed derivation is available at [3]. A similar result is derived for the case a decorrelator detector is used by the receiver. The same expression for the average rate as in (15) can be obtained for a decorrelator, only the rates R_1, R_2 are given by

$$\begin{aligned} R_1 &= \log\left\{1 + \frac{s_1^{(\infty)}\alpha\text{SNR}\eta_{dec}\left(\beta F_s\left(s_2^{(\infty)}\right)\right)}{1+s_1^{(\infty)}\overline{\alpha}\text{SNR}\eta_{dec}\left(\beta F_s\left(s_2^{(\infty)}\right)\right)}\right\} \\ R_2 &= \log\left\{1 + s_2^{(\infty)}\overline{\alpha}\text{SNR}\eta_{dec}\left(\beta F_s\left(s_2^{(\infty)}\right)\right)\right\} \end{aligned} \quad (17)$$

A detailed derivation is available at [3, Proposition 8.1]. The elementary difference between the MF and decorrelator decoding is that with a MF every decoded layer reduces the interference, and thus increases the effective system load. With a decorrelator, the effective system load can be reduced only after all layers of some user are reliably decoded.

V. The Continuous Broadcast Approach

Consider a single-input single-output (SISO) channel,

$$y_i = hx_i + n_i \ , \quad (18)$$

where $\{y_i\}$ are samples of the received symbols, $\{x_i\}$ are the transmitted complex symbols, satisfying the power constraint $E|x|^2 \le P$. $\{n_i\}$ are the additive noise samples, which are complex Gaussian i.i.d with zero mean and unit variance denoted $\mathcal{CN}(0,1)$, and h is the fading coefficient, which remains fixed during a transmission block, and varies over time according to a distribution density function $f_s(h)$. Note that since the additive noise is normalized and $E[|h|^2] = 1$, $\text{SNR} = P$.

In the continuous broadcast approach [4], every layer is associated with a channel state $s = |h|^2$. The incremental differential rate as function of the channel state is

$$dR(s) = \log\left(1 + \frac{s\rho(s)ds}{1 + sI(s)}\right) = \frac{s\rho(s)ds}{1 + sI(s)} \qquad (19)$$

where $\rho(s)$ is the transmit power density function. Thus $\rho(s)ds$ is the transmit power of a layer parameterized by s, associated with fading state s. Information streams intended for receivers indexed by $u > s$ are undetectable and play a role of additional interfering noise, denoted by $I(s)$. The interference for a fading power s is $I(s) = \int_s^\infty \rho(u)du$, which is a monotonically decreasing function of s. The total transmitted power is the overall collected power assigned to all layers $I(0) = P$. The expected rate is achieved with sufficiently many transmission blocks, each experiencing an independent fading realization. Therefore, the expected rate R_{bs} is

$$R_{bs} = \int\limits_0^\infty du\ f_s(u) \int\limits_0^u dR(s)ds = \int\limits_0^\infty du(1 - F_s(u))\frac{u\rho(u)}{1 + uI(u)}$$

where $f_s(u)$ is the pdf of the fading power, and $F_s(u)$ is the corresponding cdf. Optimization of R_{bs} for maximal throughput w.r.t. the power distribution $I(s)$ can be found by solving the associated constrained Eüler equation [4].

A. Matched Filter Detector

For the multiuser channel model defined in (1), the achievable rates strongly depend on the transmission scheme and the decoding strategy. The decoding strategy which is adopted here is the iterative decoding, just like described for two coded layers. The achievable continuous layering rate is given by

$$R_{sum,bs}(I) = \beta \int\limits_0^\infty ds(1 - F_s(s))\frac{s\eta_{mf}(G)\rho(s)}{1 + s\eta_{mf}(G)I(s)}$$

$$\triangleq \int_0^\infty dsJ(s, I, I') \qquad (20)$$

where G corresponds to the remaining layers per user, which induce the mutual interference,

$$G \triangleq \frac{\beta}{\text{SNR}} \int\limits_0^\infty F_s(s)\rho(s)ds \triangleq \int\limits_0^\infty dsZ(s, I, I') \qquad (21)$$

where $\rho(s) = -I'(s)$. The optimization of (20) w.r.t the residual interference constraint in (21) can be solved by fixing the interference parameter G to an arbitrary value such that $0 < G \le \beta$. For such a G the optimization in (20) is a standard variational problem with a residual interference constraint on top of the power constraint $I(0) = P$. The optimization problem is therefore,

$$\begin{array}{ll} \max\limits_{I} & \int\limits_0^\infty dsJ(s, I, I') \\ \text{s.t.} & G \ge \int\limits_0^\infty dsZ(s, I, I') \end{array} \qquad (22)$$

We can write the Lagrangian form

$$L = \int\limits_0^\infty dsJ(s, I, I') \ + \ \lambda\left(G - \int\limits_0^\infty dsZ(s, I, I')\right) \qquad (23)$$

The Eüler-Lagrange condition for extremum can be derived, and the optimal layering power distribution can be expressed in a closed form, as summarized in the next proposition.

Proposition 5.1: The optimal power distribution, which maximizes the expected sum-rate of a continuous broadcast approach (22), with matched-filter multiuser detection and iterative SIC decoding, is given by

$$I(s) = \begin{cases} \mathit{SNR} & s < s_0 \\ \dfrac{-\mathit{SNR} + \sqrt{\mathit{SNR}^2 \ + \ \dfrac{4\lambda\beta(1 - F_s(s))\mathit{SNR}}{\eta_{mf}(G)s^2F'_s(s)}}}{2\lambda\beta} - \dfrac{1}{s\eta_{mf}(G)} & s_0 \le s \le s_1 \\ 0 & s > s_1 \end{cases}$$

with s_1 is the smallest fading gain for which $I(s_1) = 0$, and the left boundary condition on s_0 satisfies $I(s_0) = \mathit{SNR}$. The Lagrangian multiplier λ is obtained by an equality for the residual interference constraint (21), as specified by

$$\int_{s_0}^{s_1} F_s(s)I'(s)ds = -G\frac{\mathit{SNR}}{\beta} \qquad (24)$$

Proof: (sketch) Details are available in [3, Proposition 8.4]. The first step is to explicitly write the extremum condition of the Lagrangian in (23). The extremum condition is given by the Eüler-Lagrange equation which is a necessary condition for a zero variation,

$$J_I - \frac{\partial J_{I'}}{\partial s} - \lambda\left(Z_I - \frac{\partial Z_{I'}}{\partial s}\right) \ = \ 0 \qquad (25)$$

which can be explicitly formulated

$$\lambda\frac{\beta}{\text{SNR}}F'_s(s)T^2 + F'_s(s)s\eta_{mf}T - (1 - F_s(s))\eta_{mf} \ = \ 0 \ (26)$$

where we defined $T \triangleq 1 + s\eta_{mf}I$. Solving I from (26) yields the optimal power allocation. It remains to apply the subsidiary conditions on the optimal solution, such that the power constraint and the residual interference constraint (21) are met with equality.

B. Decorrelator Detector

The decoding algorithm for a decorrelator multiuser detector is similar. In the continuous setting the detector efficiency is updated according to the number of users for which ALL layers are decoded. This is the reason the upper boundary of the power distribution is actually a subject for optimization. The solution is obtained by solving the corresponding variable end point variational optimization problem.

The average achievable rate with a decorrelator detector, in its general form, is given by

$$R_{bs,decorr} = \beta \int\limits_{s_a^+}^{s_b^-} ds(1 - F_s(s))\frac{s\rho(s)\eta\left(F_s(s_b)\beta\right)}{1 + sI(s)\eta\left(F_s(s_b)\beta\right)}$$

$$+ (1 - F_s(s_a))R_0(s_a) + (1 - F_s(s_b))R_1(s_b) \qquad (27)$$

where $I(s_a^-) = \text{SNR}$, and $I(s_a^+)$ is the remaining power allocation for the continuous and last layers. The rate of the first and last layers, respectively, is

$$\begin{aligned} R_0(s_a) &= \beta \log\left(1 + \frac{s_a \eta(\beta F_s(s_b))(\text{SNR} - I(s_a^+))}{1 + s_a \eta(\beta F_s(s_b)) I(s_a^+)}\right) \\ R_1(s_b) &= \beta \log\left(1 + s_b \eta\left(\beta F_s(s_b)\right) I(s_b^-)\right) \end{aligned} \tag{28}$$

where $I(s_b^+) = 0$. The optimal power allocation and its derivation are available in [3, Proposition 8.5]. It is shown in [3, Proposition 6.1] that equal rates allocation for all users maximizes the spectral efficiency, for any number of layers.

VI. Numerical Results

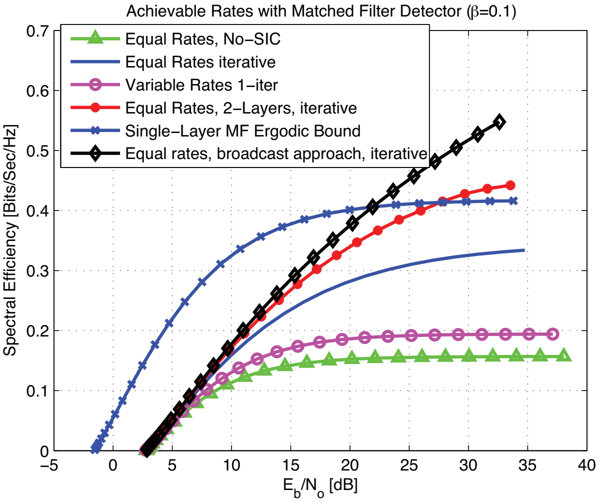


Fig. 1. Expected spectral efficiency for a Rayleigh fading channel, receiver uses a **MF** multiuser detector ($\beta = 0.1$).

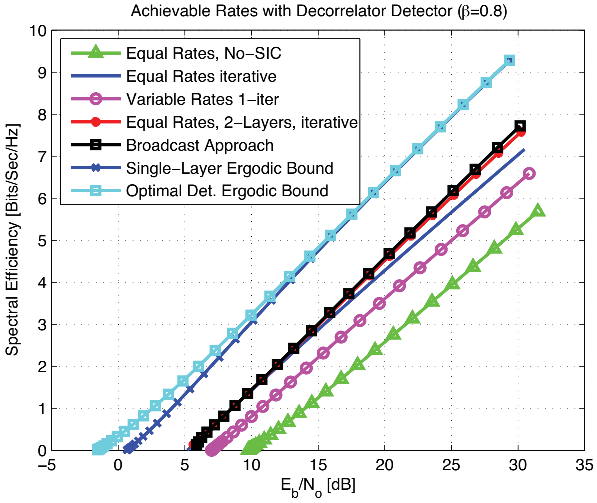


Fig. 2. Expected spectral efficiency for a Rayleigh fading channel, receiver uses a **decorrelator** multiuser detector ($\beta = 0.8$).

Figures 1 and 2 demonstrate the expected spectral efficiency for MF and decorrelator detectors, respectively. Different transmission and decoding strategies are compared. The single layer ergodic bound is given in (4). Equal fixed rate with single iteration refers to the case no SIC is used. The spectral efficiency for single layer iterative decoding is specified in (11), and for two layer coding with iterative decoding in (15). The broadcast approach achievable spectral efficiency, with iterative decoding, is given in (22) for a MF detector.

VII. Conclusion

The spectral efficiency of practical linear multiuser detectors such as MF and decorrelator employing SIC receivers was derived. Single layer and multi-layer coding per user were studied. The multi-layer coding expected sum-rate, under iterative decoding with linear multiuser detectors, is optimized, and the optimal power distribution is obtained. The achievable spectral efficiency for a linear MF detector shows significant gains over the single layer coding approach. The interesting observation here is that the expected spectral efficiency exceeds the single layer ergodic sum-capacity. The ergodic bound assumes that every user transmits at a rate matched to its decoding stage and channel realization. For a single user setting the ergodic bound is always an upper bound for the broadcast approach. However, in our multiuser setting a MF detector is used for the ergodic bound, and the MF detection is information lossy. In the broadcast approach the MF detection is performed over and over for every layer according to the iterative decoding scheme. Therefore the broadcast approach can provide spectral efficiencies exceeding those of a single layer coding with channel side information, when a MF detector is used.

It is worth noting that systems employing decorrelator detection, can significantly gain from using SIC, at system loads close to 1. For such system loads, single user detection is interference limited, and therefore achievable rate can be infinitesimally small. With layering and iterative SIC, the layers decoded first must have low rates. Gradually, the effective system load reduces, and higher expected spectral efficiencies can be achieved.

Acknowledgment

The work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless in Communications NEWCOM++, and the Israel Science Foundation.

References

[1] S. Shamai and S. Verdu, “The impact of frequency-flat fading on the spectral efficiency of cdma,” *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[2] S. Shamai, B. Zaidel, and S. Verdu, “Strongest-users-only detectors for randomly spread cdma,” *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, pp. 20–, 2002.

[3] A. Steiner, S. Shamai, V. Lupu, and U. Katz, “The spectral efficiency of successive cancellation with linear multiuser detection for randomly spread CDMA,” *submitted to IEEE transactions on Information Theory*, September 2009.

[4] S. Shamai (Shitz) and A. Steiner, “A broadcast approach for a single user slowly fading MIMO channel,” *IEEE Trans. on Info, Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

Constructing Optimal Authentication Codes with Perfect Multi-fold Secrecy

Michael Huber
University of Tuebingen
Wilhelm Schickard Institute for Computer Science
Sand 13, D-72076 Tuebingen, Germany
Email: michael.huber@uni-tuebingen.de

***Abstract*—We establish a construction of optimal authentication codes achieving perfect multi-fold secrecy by means of combinatorial designs. This continues the author's work (ISIT 2009, cf. [1]) and answers an open question posed therein. As an application, we present the first infinite class of optimal codes that provide two-fold security against spoofing attacks and at the same time perfect two-fold secrecy.**

I. Introduction

Authentication and secrecy are two crucial concepts in cryptography and information security. Although independent in their nature, various scenarios require that both aspects hold simultaneously. For *information-theoretic* or *unconditional* security (i.e. robustness against an attacker that has unlimited computational resources), authentication and secrecy codes have been investigated for quite some time. The initial construction of authentication codes goes back to Gilbert, MacWilliams & Sloane [2]. A more general and systematic theory of authentication was developed by Simmons (e.g., [3], [4]). Fundamental work on secrecy codes started with Shannon [5].

This paper deals with the construction of optimal authentication codes with perfect multi-fold secrecy. It continues the author's recent work [1], which naturally extended results by Stinson [6] on authentication codes with perfect secrecy. We will answer an important question left open in [1] that addresses the construction of authentication codes with perfect multi-fold secrecy for equiprobable source probability distributions. We establish a construction of optimal authentication codes which are multi-fold secure against spoofing attacks and simultaneously provide perfect multi-fold secrecy. This can be achieved by means of combinatorial designs. As an application, we present the first infinite class of optimal codes that achieve two-fold security against spoofing as well as perfect two-fold secrecy.

The paper is organized as follows: Necessary definitions and concepts from the theory of authentication and secrecy codes as well as from combinatorial design theory will be summarized in Section II. Section III gives relevant combinatorial constructions of optimal authentication codes which bear no secrecy assumptions. In Section IV, we review Stinson's constructions in [6] and recent results from [1]. Section V is devoted to our new constructions.

II. Preliminaries

A. Authentication and Secrecy Codes

We rely on the information-theoretical or unconditional secrecy model developed by Shannon [5], and by Simmons (e.g., [3], [4]) including authentication. Our notion complies, for the most part, with that of [6], [7]. In this model of authentication and secrecy three participants are involved: a *transmitter*, a *receiver*, and an *opponent*. The transmitter wants to communicate information to the receiver via a public communications channel. The receiver in return would like to be confident that any received information actually came from the transmitter and not from some opponent (*integrity* of information). The transmitter and the receiver are assumed to trust each other. Sometimes this is also called an *A-code*.

In what follows, let $\mathcal{S}$ denote a set of k *source states* (or *plaintexts*), $\mathcal{M}$ a set of v *messages* (or *ciphertexts*), and $\mathcal{E}$ a set of b *encoding rules* (or *keys*). Using an encoding rule $e \in \mathcal{E}$, the transmitter encrypts a source state $s \in \mathcal{S}$ to obtain the message $m = e(s)$ to be sent over the channel. The encoding rule is an injective function from $\mathcal{S}$ to $\mathcal{M}$, and is communicated to the receiver via a secure channel prior to any messages being sent. For a given encoding rule $e \in \mathcal{E}$, let $M(e) := \{e(s) : s \in \mathcal{S}\}$ denote the set of *valid* messages. For an encoding rule e and a set $M^* \subseteq M(e)$ of distinct messages, we define $f_e(M^*) := \{s \in \mathcal{S} : e(s) \in M^*\}$, i.e., the set of source states that will be encoded under encoding rule e by a message in M^*. A received message m will be accepted by the receiver as being authentic if and only if $m \in M(e)$. When this is fulfilled, the receiver decrypts the message m by applying the decoding rule e^{-1}, where

$$e^{-1}(m) = s \Leftrightarrow e(s) = m.$$

An authentication code can be represented algebraically by a $(b \times k)$-*encoding matrix* with the rows indexed by the encoding rules, the columns indexed by the source states, and the entries defined by $a_{es} := e(s)$ $(1 \leq e \leq b,\ 1 \leq s \leq k)$.

We address the scenario of a *spoofing attack* of order i (cf. [7]): Suppose that an opponent observes $i \geq 0$ distinct messages, which are sent through the public channel using the same encoding rule. The opponent then inserts a new message m' (being distinct from the i messages already sent), hoping to have it accepted by the receiver as authentic. The cases $i = 0$

and $i = 1$ are called *impersonation game* and *substitution game*, respectively. These cases have been studied in detail in recent years (e.g., [8], [9]), however less is known for the cases $i \geq 2$. In this article, we focus on those cases where $i \geq 2$.

For any i, we assume that there is some probability distribution on the set of i-subsets of source states, so that any set of i source states has a non-zero probability of occurring. For simplification, we ignore the order in which the i source states occur, and assume that no source state occurs more than once. Given this probability distribution p_S on $\mathcal{S}$, the receiver and transmitter choose a probability distribution p_E on $\mathcal{E}$ (called *encoding strategy*) with associated independent random variables S and E, respectively. These distributions are known to all participants and induce a third distribution, p_M, on $\mathcal{M}$ with associated random variable M. The *deception probability* P_{d_i} is the probability that the opponent can deceive the receiver with a spoofing attack of order i. The following theorem (cf. [7]) provides combinatorial lower bounds.

Theorem 1: [Massey] In an authentication code with k source states and v messages, the deception probabilities are bounded below by

$$P_{d_i} \geq \frac{k-i}{v-i}.$$

An authentication code is called *t_A-fold secure against spoofing* if $P_{d_i} = (k-i)/(v-i)$ for all $0 \leq i \leq t_A$.

Moreover, we consider the concept of perfect multi-fold secrecy which has been introduced by Stinson [6] and generalizes Shannon's fundamental idea of perfect (one-fold) secrecy (cf. [5]). We say that an authentication code has *perfect t_S-fold secrecy* if, for every positive integer $t^* \leq t_S$, for every set M^* of t^* messages observed in the channel, and for every set S^* of t^* source states, we have

$$p_S(S^*|M^*) = p_S(S^*).$$

That is, the *a posteriori* probability distribution on the t^* source states, given that a set of t^* messages is observed, is identical to the *a priori* probability distribution on the t^* source states.

When clear from the context, we often only write t instead of t_A resp. t_S.

B. Combinatorial Designs

We recall the definition of a combinatorial t-design. For positive integers $t \leq k \leq v$ and λ, a *t-(v, k, λ) design $\mathcal{D}$* is a pair $(X, \mathcal{B})$, satisfying the following properties:

(i) X is a set of v elements, called *points*,
(ii) $\mathcal{B}$ is a family of k-subsets of X, called *blocks*,
(iii) every t-subset of X is contained in exactly λ blocks.

We denote points by lower-case and blocks by upper-case Latin letters. Via convention, let $b := |\mathcal{B}|$ denote the number of blocks. Throughout this article, 'repeated blocks' are not allowed, that is, the same k-subset of points may not occur twice as a block. If $t < k < v$ holds, then we speak of a *non-trivial* t-design. For historical reasons, a t-(v, k, λ) design with $\lambda = 1$ is called a *Steiner t-design* (sometimes also a *Steiner system*). The special case of a Steiner design with parameters $t = 2$ and $k = 3$ is called a *Steiner triple system STS(v)* of order v. A Steiner design with parameters $t = 3$ and $k = 4$ is called a *Steiner quadruple system SQS(v)* of order v. Specifically, we are interested in Steiner quadruple systems in this paper. As a simple example, the vector space $\mathbb{Z}_2^d$ $(d \geq 3)$ with the set $\mathcal{B}$ of blocks taken to be the set of all subsets of four distinct elements of $\mathbb{Z}_2^d$ whose vector sum is zero, is a non-trivial *boolean* Steiner quadruple system SQS(2^d). More geometrically, these SQS(2^d) consist of the points and planes of the d-dimensional binary affine space $AG(d, 2)$.

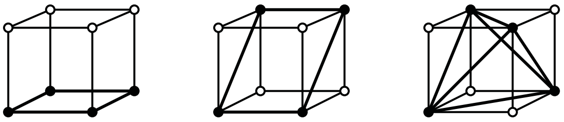


Fig. 1. Illustration of the unique SQS(8), with three types of blocks: faces, opposite edges, and inscribed regular tetrahedra.

For the existence of t-designs, basic necessary conditions can be obtained via elementary counting arguments (see, for instance, [10]):

Lemma 1: Let $\mathcal{D} = (X, \mathcal{B})$ be a t-(v, k, λ) design, and for a positive integer $s \leq t$, let $S \subseteq X$ with $|S| = s$. Then the number of blocks containing each element of S is given by

$$\lambda_s = \lambda \frac{\binom{v-s}{t-s}}{\binom{k-s}{t-s}}.$$

In particular, for $t \geq 2$, a t-(v, k, λ) design is also an s-(v, k, λ_s) design.

It is customary to set $r := \lambda_1$ denoting the number of blocks containing a given point. It follows

Lemma 2: Let $\mathcal{D} = (X, \mathcal{B})$ be a t-(v, k, λ) design. Then the following holds:

(a) $bk = vr$.
(b) $\displaystyle\binom{v}{t}\lambda = b\binom{k}{t}$.
(c) $r(k-1) = \lambda_2(v-1)$ for $t \geq 2$.

For encyclopedic accounts of key results in design theory, we refer to [10], [11]. Various connections of designs with coding and information theory can be found in a recent survey [12] (with many additional references therein).

III. Optimal Authentication Codes

For our further purposes, we summarize the state-of-the-art for authentication codes which bear no secrecy assumptions. The following theorem (cf. [7], [13]) gives a combinatorial lower bound on the number of encoding rules.

Theorem 2: [Massey–Schöbi] If an authentication code is $(t-1)$-fold against spoofing, then the number of encoding rules is bounded below by

$$b \geq \frac{\binom{v}{t}}{\binom{k}{t}}.$$

TABLE I
OPTIMAL AUTHENTICATION CODES WITH PERFECT SECRECY: INFINITE CLASSES

t_A	t_S	k	v	b	Ref.
1	1	$q+1$ q prime power	$\frac{q^{d+1}-1}{q-1}$ $d \geq 2$ even	$\frac{v(v-1)}{k(k-1)}$	[6]
1	1	3	$v \equiv 1 \pmod 6$	$\frac{v(v-1)}{6}$	[1]
1	1	4	$v \equiv 1 \pmod{12}$	$\frac{v(v-1)}{12}$	[1]
1	1	5	$v \equiv 1 \pmod{20}$	$\frac{v(v-1)}{20}$	[1]
2	1	$q+1$ q prime power	q^d+1 $d \geq 2$ even	$\frac{v(v-1)(v-2)}{k(k-1)(k-2)}$	[1]
2	1	4	$v \equiv 2, 10 \pmod{24}$	$\frac{v(v-1)(v-2)}{24}$	[1]

TABLE II
OPTIMAL AUTHENTICATION CODES WITH PERFECT SECRECY: FURTHER EXAMPLES

t_A	t_S	k	v	b	Ref.
2	1	5	26	260	[1]
3	1	5	11	66	[1]
		7	23	253	[1]
		5	23	1.771	[1]
		5	47	35.673	[1]
		5	83	367.524	[1]
		5	71	194.327	[1]
		5	107	1.032.122	[1]
		5	131	2.343.328	[1]
		5	167	6.251.311	[1]
		5	243	28.344.492	[1]
4	1	6	12	132	[1]
		6	84	5.145.336	[1]
		6	244	1.152.676.008	[1]

An authentication code is called *optimal* if the number of encoding rules meets the lower bound with equality. When the source states are known to be independent and equiprobable, optimal authentication codes which are $(t-1)$-fold secure against spoofing can be constructed via t-designs (cf. [6], [13], [14]).

Theorem 3: [DeSoete–Schöbi–Stinson] Suppose there is a t-(v,k,λ) design. Then there is an authentication code for k equiprobable source states, having v messages and $\lambda \cdot \binom{v}{t}/\binom{k}{t}$ encoding rules, that is $(t-1)$-fold secure against spoofing. Conversely, if there is an authentication code for k equiprobable source states, having v messages and $\binom{v}{t}/\binom{k}{t}$ encoding rules, that is $(t-1)$-fold secure against spoofing, then there is a Steiner t-$(v,k,1)$ design.

IV. STINSON'S CONSTRUCTIONS & RECENT RESULTS

Using the notation introduced in Section II-A, we review in Tables I and II previous constructions from [6], [1] for equiprobable source probability distributions. This lists all presently known optimal authentication codes with perfect secrecy.

V. NEW CONSTRUCTIONS

Starting from the condition of perfect t-fold secrecy, we obtain via Bayes' Theorem that

$$\begin{aligned} p_S(S^*|M^*) &= \frac{p_M(M^*|S^*)p_S(S^*)}{p_M(M^*)} \\ &= \frac{\sum_{\{e\in\mathcal{E}:S^*=f_e(M^*)\}} p_E(e)p_S(S^*)}{\sum_{\{e\in\mathcal{E}:M^*\subseteq M(e)\}} p_E(e)p_S(f_e(M^*))} = p_S(S^*). \end{aligned}$$

It follows

Lemma 3: An authentication code has perfect t-fold secrecy if and only if, for every positive integer $t^* \leq t$, for every set M^* of t^* messages observed in the channel and for every set S^* of t^* source states, we have

$$\sum_{\{e\in\mathcal{E}:S^*=f_e(M^*)\}} p_E(e) = \sum_{\{e\in\mathcal{E}:M^*\subseteq M(e)\}} p_E(e)p_S(f_e(M^*)).$$

Hence, if the encoding rules in a code are used with equal probability, then for every $t^* \leq t$, a given set of t^* messages occurs with the same frequency in each t^* columns of the encoding matrix.

We can now establish an extension of the main theorem in [1]. Our construction yields optimal authentication codes which are multi-fold secure against spoofing and provide perfect multi-fold secrecy.

Theorem 4: Suppose there is a Steiner t-$(v,k,1)$ design, where $\binom{v}{t^*}$ divides the number of blocks b for every positive integer $t^* \leq t-1$. Then there is an optimal authentication code for k equiprobable source states, having v messages and $\binom{v}{t}/\binom{k}{t}$ encoding rules, that is $(t-1)$-fold secure against spoofing and simultaneously provides perfect $(t-1)$-fold secrecy.

Proof: Let $\mathcal{D} = (X, \mathcal{B})$ be a Steiner t-$(v,k,1)$ design, where $\binom{v}{t^*}$ divides b for every positive integer $t^* \leq t-1$. By Theorem 3, the authentication code has $(t-1)$-fold security against spoofing attacks. Hence, it remains to prove that the code also achieves perfect $(t-1)$-fold secrecy under the assumption that the encoding rules are used with equal probability. With respect to Lemma 3, we have to show that, for every $t^* \leq t-1$, a given set of t^* messages occurs with the same frequency in each t^* columns of the resulting encoding matrix. This can be accomplished by ordering, for each $t^* \leq t-1$, every block of $\mathcal{D}$ in such a way that every t^*-subset of X occurs in each possible choice in precisely $b/\binom{v}{t^*}$ blocks. Since every t^*-subset of X occurs in exactly $\lambda_{t^*} = \binom{v-t^*}{t-t^*}/\binom{k-t^*}{t-t^*}$ blocks due to Lemma 1, necessarily $\binom{k}{t^*}$ must divide λ_{t^*}. By Lemma 2 (b), this is equivalent to saying that $\binom{v}{t^*}$ divides b. To show that the condition is also sufficient, we consider the bipartite (t^*-subset, block) incidence graph of $\mathcal{D}$ with vertex set $\binom{X}{t^*} \cup \mathcal{B}$, where $(\{x_i\}_{i=1}^{t^*}, B)$ is an edge if and only if $x_i \in B$ $(1 \leq i \leq t^*)$ for $\{x_i\}_{i=1}^{t^*} \in \binom{X}{t^*}$ and $B \in \mathcal{B}$. An ordering on each block of $\mathcal{D}$ can be obtained via an edge-coloring of this graph using $\binom{k}{t^*}$ colors in such a way that each vertex $B \in \mathcal{B}$ is adjacent to one edge of each color, and each vertex $\{x_i\}_{i=1}^{t^*} \in \binom{X}{t^*}$ is adjacent to $b/\binom{k}{t^*}$ edges of each color. Specifically, this can be done by first splitting up each vertex $\{x_i\}_{i=1}^{t^*}$ into $b/\binom{k}{t^*}$ copies, each having degree $\binom{k}{t^*}$, and then by finding an appropriate edge-coloring of the resulting $\binom{k}{t^*}$-regular bipartite graph using $\binom{k}{t^*}$ colors. The claim follows now by taking the ordered blocks as encoding rules, each used with equal probability. ■

Remark 1: It follows from the proof that we may obtain optimal authentication codes that provide $(t-1)$-fold security against spoofing and at the same time perfect $(t'-1)$-fold secrecy for $t' \leq t$, when the assumption of the above theorem holds with $\binom{v}{t^*}$ divides b for every positive integer $t^* \leq t'-1$.

As an application, we give an infinite class of optimal codes which are two-fold secure against spoofing and achieve perfect two-fold secrecy. This appears to be the first infinite class of authentication and secrecy codes with these properties.

Theorem 5: For all positive integers $v \equiv 2 \pmod{24}$, there is an optimal authentication code for $k = 4$ equiprobable source states, having v messages, and $v(v-1)(v-2)/24$ encoding rules, that is two-fold secure against spoofing and provides perfect two-fold secrecy.

Proof: We will make use of Steiner quadruple systems (cf. Section II-A). Hanani [15] showed that a necessary and sufficient condition for the existence of a $\mathrm{SQS}(v)$ is that $v \equiv 2$ or $4 \pmod 6$ $(v \geq 4)$. Hence, the condition $v \mid b$ is fulfilled when $v \equiv 2$ or $10 \pmod{24}$ and the condition $\binom{v}{2} \mid b$ when $v \equiv 2 \pmod{12}$ in view Lemma 2 (b). Therefore, if we assume that $v \equiv 2 \pmod{24}$, then we can apply Theorem 4 to establish the claim. ■

We present the smallest example:

Example 1: An optimal authentication code for $k = 4$ equiprobable source states, having $v = 26$ messages, and $b = 650$ encoding rules, that is two-fold secure against spoofing and provides perfect two-fold secrecy can be constructed from a Steiner quadruple system $\mathrm{SQS}(26)$. Each encoding rule is used with probability $1/650$.

Remark 2: For $v = 26$, the first $\mathrm{SQS}(v)$ was constructed by Fitting [16], admitting a v-cycle as an automorphism (*cyclic* $\mathrm{SQS}(v)$). We generally remark that the number $N(v)$ of non-isomorphic $\mathrm{SQS}(v)$ is only known for $v = 8, 10, 14, 16$ with $N(8) = N(10) = 1$, $N(14) = 4$, and $N(16) =$ 1,054,163 (cf. [17]). Lenz [18] proved that for the admissible values of v, the number $N(v)$ grows exponentially, i.e. $\liminf_{v\to\infty} \frac{\log N(v)}{v^3} > 0$. For comprehensive survey articles on Steiner quadruple systems, we refer the reader to [19], [20]. For classifications of specific classes of highly regular Steiner quadruple systems and Steiner designs, see, e.g., [21], [22].

Acknowledgment

The author thanks Doug Stinson for an interesting conversation on this topic. The author gratefully acknowledges support of his work by the Deutsche Forschungsgemeinschaft (DFG) via a Heisenberg grant (Hu954/4) and a Heinz Maier-Leibnitz Prize grant (Hu954/5).

References

[1] M. Huber, “Authentication and secrecy codes for equiprobable source probability distributions”, in *Proc. IEEE International Symposium on Information Theory (ISIT) 2009*, pp. 1105–1109, 2009.

[2] E. N. Gilbert, F. J. MacWilliams and N. J. A. Sloane, “Codes which detect deception”, *Bell Syst. Tech. J.*, vol. 53, pp. 405–424, 1974.

[3] G. J. Simmons, “Authentication theory/coding theory”, in *Advances in Cryptology – CRYPTO ’84*, ed. by G. R. Blakley and D. Chaum, Lecture Notes in Computer Science, vol. 196, Springer, Berlin, Heidelberg, New York, pp. 411–432, 1985.

[4] G. J. Simmons, “A survey of information authentication”, in *Contemporary Cryptology: The Science of Information Integrity*, ed. by G. J. Simmons, IEEE Press, Piscataway, pp. 379–419, 1992.

[5] C. E. Shannon, “Communication theory of secrecy systems”, *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[6] D. R. Stinson, “The combinatorics of authentication and secrecy codes”, *J. Cryptology*, vol. 2, pp. 23–49, 1990.

[7] J. L. Massey, “Cryptography – a selective survey”, in *Digital Communications*, ed. by E. Biglieri and G. Prati, North-Holland, Amsterdam, New York, Oxford, pp. 3–21, 1986.

[8] D. R. Stinson, “Combinatorial characterizations of authentication codes”, *Designs, Codes and Cryptography*, vol. 2, pp. 175–187, 1992.

[9] D. R. Stinson and R. S. Rees, “Combinatorial characterizations of authentication codes II”, *Designs, Codes and Cryptography*, vol. 7, pp. 239–259, 1996.

[10] Th. Beth, D. Jungnickel and H. Lenz, *Design Theory*, vol. I and II, Encyclopedia of Math. and Its Applications, vol. 69/78, Cambridge Univ. Press, Cambridge, 1999.

[11] C. J. Colbourn and J. H. Dinitz (eds.), *Handbook of Combinatorial Designs*, 2nd ed., CRC Press, Boca Raton, 2006.

[12] M. Huber, “Coding theory and algebraic combinatorics”, in *Selected Topics in Information and Coding Theory*, ed. by I. Woungang *et al.*, World Scientific, Singapore, 38 pages, 2010 (in press). Preprint at arXiv:0811.1254v1.

[13] P. Schöbi, “Perfect authentication systems for data sources with arbitrary statistics” (presented at EUROCRYPT ’86), unpublished.

[14] M. De Soete, “Some constructions for authentication - secrecy codes”, in *Advances in Cryptology – EUROCRYPT ’88*, ed. by Ch. G. Günther, Lecture Notes in Computer Science, vol. 330, Springer, Berlin, Heidelberg, New York, pp. 23–49, 1988.

[15] H. Hanani, “On quadruple systems”, *Canad. J. Math.*, vol. 12, pp. 145–157, 1960.

[16] F. Fitting, “Zyklische Lösungen des Steiner’schen Problems”, *Nieuw. Arch. Wisk.*, vol. 11, pp. 140–148, 1915.

[17] P. Kaski, P. R. J. Östergård and O. Pottonen, “The Steiner quadruple systems of order 16”, *J. Combin. Theory, Series A*, vol. 113, pp. 1764–1770, 2006.

[18] H. Lenz, “On the number of Steiner quadruple systems”, *Mitt. Math. Sem. Giessen*, vol. 169, pp. 55–71, 1985.

[19] A. Hartman and K. T. Phelps, “Steiner quadruple systems”, in: *Contemporary Design Theory*, ed. by J. H. Dinitz and D. R. Stinson, Wiley, New York, pp. 205–240, 1992.

[20] C. C. Lindner and A. Rosa, “Steiner quadruple systems – A survey”, *Discrete Math.*, vol. 22, pp. 147–181, 1978.

[21] M. Huber, “Almost simple groups with socle $L_n(q)$ acting on Steiner quadruple systems”, *J. Combin. Theory, Series A*, 4 pages, 2010 (in press). Preprint at arXiv:0907.1281v1.

[22] M. Huber, *Flag-transitive Steiner Designs*, Birkhäuser, Basel, Berlin, Boston, 2009.

Generalizing the Transfer in Iterative Error Correction: Dissection Decoding

Ulrich Sorger
Computer Science and Communications,
University of Luxembourg,
Luxembourg
Email: ulrich.sorger@uni.lu

Axel Heim
Institute of Telecommunications and
Applied Information Theory
Ulm University, Germany
Email: axel.heim@uni-ulm.de

Abstract—**Iterative decoding with message-passing is considered. The message format is generalized from the classical, single probability value for each code symbol to a probability distribution by introducing an additional logarithmic probability measure. Thereby, the representation of the probability distributions underlying the constituent code constraints by the messages is improved in terms of the Kullback-Leibler divergence. Simulation shows that this improvement can transfer to the error correcting performance.**

I. Introduction

PEARL's belief propagation algorithm (BPA) [1], [2] has attracted major attention in the communication community when it was applied to parallel concatenated convolutional codes (PCCCs) by BERROU et al. [3] in the early 90's. Using the BCJR algorithm [4] to efficiently compute symbol probabilities in the trellises of the constituent codes, the iterative exchange of so-called extrinsic information between the constituent decoders allows for error correcting performance close to the SHANNON limit [5] while maintaining low computational complexity. The field of application was quickly extended to other code constructions like serial concatenations [6] or low-density parity-check codes [7]. The basic principle of the decoding scheme, however, has remained the same ever since.

After recalling the abstract class of intersection codes in Section II, Section III-A emphasizes an observation made in [8]: The symbol probabilities computed in the constituent decoders minimize the KULLBACK-LEIBLER divergence between a) the probability distribution of the code words given the input beliefs and the code constraint, and b) the uncoded distribution given the objective variables. By replacing the latter distribution by a new one with a larger parameter space in Section III-B, this optimization is improved. Simulation in Section IV shows that this improvement can also transfer to the error correcting performance.

II. Intersection Codes

The class of intersection (IS) codes [9] is equivalent to the class of embedding codes [10] or trellis-constrained codes. Every code can be expressed as the intersection of two (or more) super-codes, and hence as an IS code.

Definition 1 (Intersection Code): Let $\mathbb{C}^{(1)}$ and $\mathbb{C}^{(2)}$ be linear block codes of length n. An intersection code $\mathbb{C}^{(\cap)}$ is defined as the intersection

$$\mathbb{C}^{(\cap)} = \mathbb{C}^{(1)} \cap \mathbb{C}^{(2)} \tag{1}$$

of the constituent codes (super codes) $\mathbb{C}^{(1)}$ and $\mathbb{C}^{(2)}$.

The parity check matrix of an intersection code is obtained by stacking the $h^{(l)} \times n$ parity check matrices $\mathbf{H}^{(l)}$, $l = 1, 2$ of its constituent codes $\mathbb{C}^{(l)}$. I.e., for $\boldsymbol{c} = [c_1\, c_2 \ldots c_n]$ being a binary vector, Equation (1) is equivalent to

$$\mathbb{C}^{(\cap)} = \left\{ \boldsymbol{c} : \mathbf{H}^{(\cap)} \cdot \boldsymbol{c}^T = \mathbf{0} \right\} \quad \text{with} \quad \mathbf{H}^{(\cap)} = \begin{bmatrix} \mathbf{H}^{(1)} \\ \mathbf{H}^{(2)} \end{bmatrix},$$

with

$$\mathbb{C}^{(\cap)} \subseteq \mathbb{C}^{(l)} \subseteq \mathbb{S}, \quad l = 1, 2,$$

where $\mathbb{S}$ denotes the n-dimensional binary space.

Example 1 (Turbo Codes): Let

$$\mathbf{G}_{\mathrm{CC}} = \begin{bmatrix} \mathbf{I} & \mathbf{G}^{(p)} \end{bmatrix}$$

denote the generator matrix of the two identical systematic convolutional encoders of a PCCC [3], where $\mathbf{I}$ is the identity matrix and $\mathbf{G}^{(p)}$ generates the parity part of the convolutional code words, including termination bits from both the systematic and the parity output. Let Π denote the Turbo code permutation matrix. The generator matrix of the PCCC then is given by

$$\mathbf{G}^{(\cap)} = \begin{bmatrix} \mathbf{I} & \mathbf{G}^{(p)} & \Pi\mathbf{G}^{(p)} \end{bmatrix}.$$

For interpretation as constituent codes $\mathbb{C}^{(l)}$ of an IS code, the codes defined by $\mathbf{G}_{\mathrm{CC}}$ require *uncoded extension*, i.e.

$$\mathbb{C}^{(1)} = \left\{ \begin{bmatrix} \boldsymbol{u} & \boldsymbol{u}\mathbf{G}^{(p)} & \boldsymbol{v} \end{bmatrix} : \boldsymbol{u} \in \mathbb{F}_2^k, \boldsymbol{v} \in \mathbb{F}_2^{k+2\kappa} \right\}$$
$$\mathbb{C}^{(2)} = \left\{ \begin{bmatrix} \boldsymbol{u}\Pi & \boldsymbol{v} & \boldsymbol{u}\Pi\mathbf{G}^{(p)} \end{bmatrix} : \boldsymbol{u} \in \mathbb{F}_2^k, \boldsymbol{v} \in \mathbb{F}_2^{k+2\kappa} \right\},$$

where k is the dimension of the PCCC, κ is the encoder memory and $\mathbb{F}_2^k$ denotes the binary space of dimension k.

In the following we will implicitly use binary vectors and code words with bipolar values using the mapping

$$\{0, 1\} \mapsto \{+1, -1\}.$$

III. ITERATIVE DECODING

In general, an iterative decoder is a device consisting of two (or more) constituent decoders $\mathcal{D}^{(l)}$, $l = 1, 2$ corresponding to the constituent codes $\mathbb{C}^{(l)}$, which output a set of probabilities. For decoding, the noisy received word $\boldsymbol{r}$ is input to the first constituent decoder $\mathcal{D}^{(1)}$ which computes conditional probabilities given $\boldsymbol{r}$ and the constraint of $\mathbb{C}^{(1)}$. Together with $\boldsymbol{r}$ these probabilities are input to decoder $\mathcal{D}^{(2)}$. $\mathcal{D}^{(2)}$ then computes probabilities under the constraint of $\mathbb{C}^{(2)}$ which are passed back to $\mathcal{D}^{(1)}$ and so forth until some stopping criterion is fulfilled.

We consider transmission over the additive white GAUSSian noise (AWGN) channel. Code words are transmitted with equal probability. Let

$$\boldsymbol{r} = \frac{1}{\sigma^2 \log(2)} \cdot (\boldsymbol{c} + \boldsymbol{\eta}) \tag{2}$$

be the scaled, noisy version of a code word $\boldsymbol{c} \in \mathbb{C}^{(\cap)}$, where $\boldsymbol{\eta}$ is the noise vector, σ^2 is the noise variance,

$$p_{\boldsymbol{R}|\boldsymbol{S}}(\boldsymbol{r}|\boldsymbol{s}) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \cdot \exp\left(-\frac{\left\|(\boldsymbol{c}+\boldsymbol{\eta}) - \boldsymbol{s}\right\|^2}{2\sigma^2}\right) \propto 2^{\boldsymbol{r}\boldsymbol{s}^T}$$

is the probability of $\boldsymbol{r}$ given $\boldsymbol{s} \in \mathbb{S}$, and $\boldsymbol{R}$ and $\boldsymbol{S}$ denote the corresponding random variables, respectively.

Let $\boldsymbol{C}^{(l)}$, $l = 1, 2$ denote the random variable for the words of the codes $\mathbb{C}^{(l)}$, respectively, and S_i the random variable for the i-th bit of a binary vector. Denote by

$$P_{\boldsymbol{C}^{(l)}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}) \propto p_{\boldsymbol{R}|\boldsymbol{S}}(\boldsymbol{r}|\boldsymbol{s}) \cdot \left\langle \boldsymbol{s} \in \mathbb{C}^{(l)} \right\rangle, \ l = 1, 2 \tag{3}$$

the probability of $\boldsymbol{s}$ given $\boldsymbol{r}$ and the constraint of code $\mathbb{C}^{(l)}$, where

$$\sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}^{(l)}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}) = 1 \quad \text{and} \quad \langle b \rangle := \begin{cases} 1 & \text{if } b \text{ is true} \\ 0 & \text{else} \end{cases}$$

denotes the IVERSON bracket. Further, let

$$P_{S_i|\boldsymbol{R}}\left(x|\boldsymbol{r}, \boldsymbol{C}^{(l)}\right) = \sum_{\boldsymbol{s}\in\mathbb{S}: s_i = x} P_{\boldsymbol{C}^{(l)}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}), \quad x \in \{\pm 1\}$$

define the probability for $S_i = x$ given $\boldsymbol{r}$ and the constituent code constraint $\mathbb{C}^{(l)}$, and

$$L_i^{(l)}(\boldsymbol{r}) := \frac{1}{2} \cdot \log_2 \frac{P_{S_i|\boldsymbol{R}}\left(+1|\boldsymbol{r}, \boldsymbol{C}^{(l)}\right)}{P_{S_i|\boldsymbol{R}}\left(-1|\boldsymbol{r}, \boldsymbol{C}^{(l)}\right)} \tag{4}$$

the corresponding logarithmic likelihood ratio (LLR). The probabilities $P_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r})$, $P_{S_i|\boldsymbol{R}}(x|\boldsymbol{r}, \boldsymbol{S})$ and $L_i(\boldsymbol{r})$ without a code constraint, i.e. $\boldsymbol{s} \in \mathbb{S}$, are defined accordingly.

In the following, subscripts may be neglected when clear from the context.

A. Belief Propagation

In belief propagation (BP), the messages passed between the decoders are given by a vector of extrinsic LLRs denoted by $\left(\boldsymbol{d}^{(l)} - \boldsymbol{m}^{(l)}\right)$. This vector is defined by the decoder input $\boldsymbol{m}^{(l)} = \boldsymbol{r} + \left(\boldsymbol{d}^{(h)} - \boldsymbol{m}^{(h)}\right)$ and the decoder output LLRs

$$d_i^{(l)} = \frac{1}{2} \cdot \log_2 \frac{P_{S_i|\boldsymbol{R}}\left(+1|\boldsymbol{m}^{(l)}, \boldsymbol{C}^{(l)}\right)}{P_{S_i|\boldsymbol{R}}\left(-1|\boldsymbol{m}^{(l)}, \boldsymbol{C}^{(l)}\right)}, \ i = 1, \ldots, n \tag{5}$$

Algorithm 1 The Belief Propagation Algorithm

```
1) initialize
     • set $l = 1$, $h = 2$
     • set $\boldsymbol{m}^{(h)} = \boldsymbol{d}^{(h)} = \boldsymbol{0}$
2) iterate
   while (stopping criterion not fulfilled)
     • $\mathcal{D}^{(l)} : \left(\boldsymbol{m}^{(l)} = \boldsymbol{r} + \boldsymbol{d}^{(h)} - \boldsymbol{m}^{(h)}\right) \mapsto \boldsymbol{d}^{(l)}$,   cf. (5)
     • swap $l \leftrightarrow h$
   end
3) output $\hat{\boldsymbol{c}} = \operatorname{sgn}\left(\boldsymbol{d}^{(h)}\right)$
```

given the constraint of code $\mathbb{C}^{(l)}$. This is summarized in Algorithm 1.

The computation (5) can be motivated as follows. For simplicity we consider one constituent decoder and disregard the indices l, h. Let $\boldsymbol{d}$ be a vector of n independent LLRs

$$d_i = \frac{1}{2} \cdot \log_2 \frac{P_{S_i|\boldsymbol{R}}(+1|\boldsymbol{d}, \boldsymbol{S})}{P_{S_i|\boldsymbol{R}}(-1|\boldsymbol{d}, \boldsymbol{S})}, \ i = 1, \ldots, n,$$

where we deliberately choose $\boldsymbol{R}$ as the corresponding random variable. Hence $\boldsymbol{d}$ is considered as being obtained from the same channel as the received word, i.e. $p_{\boldsymbol{R}|\boldsymbol{S}}(\boldsymbol{d}|\boldsymbol{s}) \propto \exp_2(\boldsymbol{d}\boldsymbol{s}^T)$. The following lemma shows that the cross entropy between $P_{\boldsymbol{C}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r})$ and $P_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d})$ is an objective function whose minimization with respect to $\boldsymbol{d}$ yields Equation (5).

Lemma 1 (Cross Entropy [8]): Minimizing the cross entropy

$$H_{\boldsymbol{R}\|\boldsymbol{R}}(\boldsymbol{C}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{d}) := -\sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}) \cdot \log_2 P_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d}) \tag{6}$$

between the distributions $P_{\boldsymbol{C}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r})$ and $P_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d})$ with respect to the vector $\boldsymbol{d}$ of LLRs yields the logarithmic symbol probability ratios

$$d_i = \arg\min_{v_i \in \mathbb{R}} H_{\boldsymbol{R}\|\boldsymbol{R}}(\boldsymbol{C}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{v}) = \frac{1}{2} \log_2 \frac{P_{S_i|\boldsymbol{R}}(+1|\boldsymbol{r}, \boldsymbol{C})}{P_{S_i|\boldsymbol{R}}(-1|\boldsymbol{r}, \boldsymbol{C})}$$

where $\mathbb{R}$ denotes the set of real numbers.[1]

The KULLBACK-LEIBLER divergence (KLD) is an information theoretic measure for the similarity between two distributions over the same probability space. It directly relates to the cross entropy by

$$\begin{aligned} D_{\mathrm{KL}}(\boldsymbol{C}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{d}) := {} & H_{\boldsymbol{R}\|\boldsymbol{R}}(\boldsymbol{C}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{d}) \\ & + \sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}) \cdot \log_2 P_{\boldsymbol{C}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{r}) \end{aligned}$$

and its minimum value is 0 for two identical distributions.

The observation that for belief propagation the computation within the constituent decoders corresponds to the optimization of (6) – or, equivalently, the minimization of the KLD – is essential for the concept of Dissection Decoding below.

[1] When extrinsic information from another decoder is available, $\boldsymbol{r}$ is replaced by the appropriate input $\boldsymbol{m}$.

B. Dissection Decoding

In belief propagation the transfer message can be written as a vector of LLRs. We now increase the transfer complexity by introducing a new dimension to these messages. This new dimension is spanned by the discrete random variable U whose realizations

$$u(\boldsymbol{s}) := H_{\boldsymbol{S}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{d}\right) \tag{7}$$

are given by the *conditional word uncertainties*

$$H_{\boldsymbol{S}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{d}\right) := -\log_2\left(P\left(\boldsymbol{s}|\boldsymbol{d}\right)\right) = -\log_2 \prod_{i=1}^{n} \frac{2^{s_i d_i}}{2^{d_i}+2^{-d_i}}$$

of $\boldsymbol{s}$, given a *dissector* $\boldsymbol{d}$. For now, $\boldsymbol{d}$ is assumed to be constant and is disregarded in the notation for better readability. The finite probability space of U is denoted by $\mathbb{U}$. Its size is determined by $\boldsymbol{d}$. Rather than a transfer vector of length n, we employ a matrix $\boldsymbol{m} = [\boldsymbol{m}_1[u], \ldots, \boldsymbol{m}_n[u]]$ of size $|\mathbb{U}| \times n$. We also introduce a new transfer vector $\boldsymbol{q}$ of length $|\mathbb{U}|$. Let

$$\begin{aligned} P_{\boldsymbol{S}|\boldsymbol{M},\boldsymbol{Q}}(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q}) &\propto q[u(\boldsymbol{s})] \cdot P_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{m}[u(\boldsymbol{s})]) \\ &= q[u(\boldsymbol{s})] \cdot \prod_{i=1}^{n} \frac{2^{s_i \cdot m_i[u(\boldsymbol{s})]}}{2^{m_i[u(\boldsymbol{s})]}+2^{-m_i[u(\boldsymbol{s})]}} \end{aligned} \tag{8}$$

with

$$\sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{S}|\boldsymbol{M},\boldsymbol{Q}}(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q}) = 1$$

denote the symbol-based probability of $\boldsymbol{s}$ given $\boldsymbol{m}$ and $\boldsymbol{q}$. Further define

$$P_{S_i,\boldsymbol{C},U|\boldsymbol{R}}\left(x,\boldsymbol{s},u|\boldsymbol{r}\right) := P_{\boldsymbol{C}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{r}\right) \cdot \langle S_i = x\rangle \cdot \langle H(\boldsymbol{s}|\boldsymbol{d}) = u\rangle$$

from which we obtain probabilities such as

$$P_{S_i,U|\boldsymbol{R}}\left(x,u|\boldsymbol{r},\boldsymbol{C}\right) = \sum_{\boldsymbol{s}\in\mathbb{S}} P_{S_i,\boldsymbol{C},U|\boldsymbol{R}}\left(x,\boldsymbol{s},u|\boldsymbol{r}\right)$$

by marginalization.

Akin to Lemma 1, the following theorem defines the optimum pair $(\boldsymbol{m},\boldsymbol{q})$ for representing the distribution $P_{\boldsymbol{C}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{r}\right)$ in terms of the (uncoded) distribution $P_{\boldsymbol{S}|\boldsymbol{M},\boldsymbol{Q}}(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q})$.

Theorem 1: Minimizing the cross entropy

$$\begin{aligned} H_{\boldsymbol{R}\|\boldsymbol{M},\boldsymbol{Q}}\left(\boldsymbol{C}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{m},\boldsymbol{q}\right) &:= \sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{r}\right) \cdot H_{\boldsymbol{S}|\boldsymbol{M},\boldsymbol{Q}}\left(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q}\right) \\ &= -\sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{r}\right) \cdot \log_2\left(P_{\boldsymbol{S}|\boldsymbol{M},\boldsymbol{Q}}\left(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q}\right)\right) \end{aligned} \tag{9}$$

with respect to $\boldsymbol{m}$ and $\boldsymbol{q}$ yields

$$q[u] \propto \frac{P_{U|\boldsymbol{R}}(u|\boldsymbol{r},\boldsymbol{C})}{P_{U|\boldsymbol{R}}(u|\boldsymbol{m}[u],\boldsymbol{S})}, \tag{10}$$

and $\boldsymbol{m}$ is given by the implicit solution

$$P_{S_i|\boldsymbol{R},U}(x|\boldsymbol{m}[u],u,\boldsymbol{S}) = P_{S_i|\boldsymbol{R},U}(x|\boldsymbol{r},u,\boldsymbol{C}), \quad i = 1,...,n. \tag{11}$$

We observe that for $\boldsymbol{d} = \boldsymbol{0}$, i.e. $|\mathbb{U}| = 1$ it follows from Theorem 1 that

$$m_i[u] = \frac{1}{2} \cdot \log_2 \frac{P_{S_i|\boldsymbol{R},U}(+1|\boldsymbol{r},u,\boldsymbol{C})}{P_{S_i|\boldsymbol{R},U}(-1|\boldsymbol{r},u,\boldsymbol{C})} \tag{12}$$

are the symbol beliefs given $\boldsymbol{r}$ and the code $\mathbb{C}$, and $q[u]$ is a constant, i.e. the computation is as for the BPA. Moreover, due to the larger parameter space of the objective function (9) for $|\mathbb{U}| > 1$ the cross entropy can only decrease. Closer investigation shows that in this case (12) is a near optimum approximation of (11).

We have thus found a (near) optimum pair $(\boldsymbol{m},\boldsymbol{q})$ with respect to the objective function (9) and a given dissector $\boldsymbol{d}$.

From Theorem 1 it does not directly follow how to apply the transfer message $(\boldsymbol{m},\boldsymbol{q})$ in iterative decoding. We now reintroduce superscripts to indicate constituent codes or the decoder where variables originate from. Given a message pair $(\boldsymbol{m}^{(h)},\boldsymbol{q}^{(h)})$ from decoder $\mathcal{D}^{(h)}$ we first need a new dissector $\boldsymbol{d}^{(l)}$ from which then a new message $(\boldsymbol{m}^{(l)},\boldsymbol{q}^{(l)})$ can be computed in $\mathcal{D}^{(l)}$, where $l,h = 1,2$, $l \neq h$. Define by

$$\begin{aligned} &H_{\boldsymbol{M},\boldsymbol{Q}\|\boldsymbol{R}}\left(\boldsymbol{C}|\boldsymbol{m},\boldsymbol{q}\|\boldsymbol{S}|\boldsymbol{d}\right) := \\ &\qquad -\sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}|\boldsymbol{M},\boldsymbol{Q}}\left(\boldsymbol{s}|\boldsymbol{m},\boldsymbol{q}\right) \cdot \log_2\left(P_{\boldsymbol{S}|\boldsymbol{R}}\left(\boldsymbol{s}|\boldsymbol{d}\right)\right) \end{aligned}$$

the cross entropy between the uncoded distribution of $\boldsymbol{s}$ given $\boldsymbol{d}$ and the distribution of $\boldsymbol{s} \in \mathbb{C}$ given the message pair $(\boldsymbol{m},\boldsymbol{q})$. A possible optimization rule for the dissectors $\boldsymbol{d}^{(l)}$ is given in the following.

Proposal 1: Find the dissectors $\boldsymbol{d}^{(1)}$, $\boldsymbol{d}^{(2)}$ minimizing

$$\begin{aligned} (\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)}) = \arg\min_{(\boldsymbol{v}^{(1)},\boldsymbol{v}^{(2)})} & H_{\boldsymbol{M},\boldsymbol{Q}\|\boldsymbol{R}}(\boldsymbol{C}^{(1)}|\boldsymbol{m}^{(2)},\boldsymbol{q}^{(2)}\|\boldsymbol{S}|\boldsymbol{v}^{(1)}) \\ &+ H_{\boldsymbol{M},\boldsymbol{Q}\|\boldsymbol{R}}(\boldsymbol{C}^{(2)}|\boldsymbol{m}^{(1)},\boldsymbol{q}^{(1)}\|\boldsymbol{S}|\boldsymbol{v}^{(2)}) \end{aligned} \tag{13}$$

with $(\boldsymbol{m}^{(l)},\boldsymbol{q}^{(l)})$, $l = 1,2$ chosen to minimize

$$H_{\boldsymbol{R}\|\boldsymbol{M},\boldsymbol{Q}}\left(\boldsymbol{C}^{(l)}|\boldsymbol{r}\|\boldsymbol{S}|\boldsymbol{m}^{(l)},\boldsymbol{q}^{(l)}\right)$$

given $\boldsymbol{d}^{(l)}$ according to Theorem 1.

To derive an algorithm from this proposal, compute the partial derivatives of the entropy terms in (13). We obtain

$$\begin{aligned} &\frac{\partial}{\partial d_i^{(l)}} H_{\boldsymbol{M},\boldsymbol{Q}\|\boldsymbol{R}}(\boldsymbol{C}^{(l)}|\boldsymbol{m}^{(h)},\boldsymbol{q}^{(h)}\|\boldsymbol{S}|\boldsymbol{d}^{(l)}) \\ &= \sum_{\boldsymbol{s}\in\mathbb{S}} P_{\boldsymbol{C}^{(l)}|\boldsymbol{M},\boldsymbol{Q}}(\boldsymbol{s}|\boldsymbol{m}^{(h)},\boldsymbol{q}^{(h)}) \cdot \left(\tanh_2(d_i^{(l)}) - s_i\right), \end{aligned} \tag{14}$$

and the derivative of the second term is approximately zero. Hence we set (14) equal to zero and obtain

$$d_i^{(l)} = \frac{1}{2} \cdot \log_2 \frac{P_{S_i|\boldsymbol{M},\boldsymbol{Q}}(+1|\boldsymbol{m}^{(h)},\boldsymbol{q}^{(h)},\boldsymbol{C}^{(l)})}{P_{S_i|\boldsymbol{M},\boldsymbol{Q}}(-1|\boldsymbol{m}^{(h)},\boldsymbol{q}^{(h)},\boldsymbol{C}^{(l)})} \tag{15}$$

which is a calculation rule. Note that, though not explicitly stated in the formula, the computation (15) requires knowledge of $\boldsymbol{d}^{(h)}$ as $\boldsymbol{m}^{(h)}$ and $\boldsymbol{q}^{(h)}$ are functions of u.

The results in (12) and (15) motivate the Dissection Decoding Algorithm 2 for the decoding of a noisy IS code word. In the beginning, nothing is known about either constituent code and thus $\boldsymbol{d}^{(2)} = \boldsymbol{0}$, $\boldsymbol{m}^{(2)} = \boldsymbol{0}$ and $\boldsymbol{q}^{(2)} = \boldsymbol{1}$ are initialized as all-zero and all-one, respectively, which directly results in $\boldsymbol{d}^{(1)} = \boldsymbol{0}$ when assuming equiprobable code symbols. 'Normal' symbol beliefs $\boldsymbol{m}^{(1)}[u]$ are computed in $\mathcal{D}^{(1)}$ according to (12) and passed to $\mathcal{D}^{(2)}$. There the dissector $\boldsymbol{d}^{(2)}$

Algorithm 2 Dissection Decoding

```
1) initialize
    • set $l = 1$, $h = 2$, $\boldsymbol{d}^{(2)} = \boldsymbol{0}$, $\boldsymbol{m}^{(2)} = \boldsymbol{0}$, $\boldsymbol{q}^{(2)} = \boldsymbol{1}$
2) iterate
    while (stopping criterion not fulfilled)
        • $\mathcal{D}^{(l)}$ : $(\boldsymbol{m}^{(h)}, \boldsymbol{q}^{(h)}, \boldsymbol{d}^{(h)}) \mapsto \boldsymbol{d}^{(l)}$        cf. (15)
                 $\boldsymbol{d}^{(l)} \mapsto (\boldsymbol{m}^{(l)}, \boldsymbol{q}^{(l)})$        cf. (10), (12)
        • swap $l \leftrightarrow h$
    end
3) output $\hat{\boldsymbol{c}} = \operatorname{sgn}\left(\boldsymbol{d}^{(h)}\right)$
```

is computed according to (15). Up to this point the algorithm is identical to the BPA and all computations can be accomplished with the BCJR algorithm. But rather than computing extrinsic symbol beliefs, $\boldsymbol{d}^{(2)}$ is taken to *dissect* (hence the name) the code space $\mathbb{C}^{(2)}$ and to compute the message pair $(\boldsymbol{m}^{(2)}, \boldsymbol{q}^{(2)})$ according to (10) and (12) with which the iterative procedure continues in $\mathcal{D}^{(1)}$.

IV. Implementation and Simulation

For a dissector $\boldsymbol{d}$ with non-zero real-valued elements d_i, the set size or *resolution* $|\mathbb{U}|$ is very large. The result would be a maximum likelihood (ML) decoder with huge matrices $\boldsymbol{m}$ and thus impracticable decoding complexity. Therefore we uniformly quantize the elements of $\boldsymbol{d}$ with a granularity Δ, and limit their magnitude to $|d_i| \leq d_{\max}$. Thus the set of possible word uncertainties is reduced to the values

$$H_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d}) \in \{u_{\min}, u_{\min} + 2 \cdot \Delta, u_{\min} + 4 \cdot \Delta, \dots\}$$

where

$$u_{\min} = \sum_{i=1}^{n} \left(\log_2(2^{d_i} + 2^{-d_i}) - |d_i|\right)$$

is the minimum possible word uncertainty given $\boldsymbol{d}$. We further limit the resolution $|\mathbb{U}|$ by setting

$$u(\boldsymbol{s}) = \begin{cases} H_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d}) & : H_{\boldsymbol{S}|\boldsymbol{R}}(\boldsymbol{s}|\boldsymbol{d}) \leq u_{\max} \\ u_{\max} & : \text{else} \end{cases}$$

with

$$u_{\max} = u_{\min} + (|\mathbb{U}| - 1) \cdot 2 \cdot \Delta.$$

The computations of the distributions in the matrices $\boldsymbol{m}$ are accomplished in the constituent code trellises, cf. [11].

For easy comparison with the BPA we consider a Turbo code according to [3] with dimension $k = 20$ and terminated rate $\mathrm{R} = \frac{1}{2}$ constituent codes with the generator polynomial $G(D) = [1\ \frac{1+D+D^2}{1+D^2}]$. The choice of the rather small code dimension is on purpose as the BPA is known to not perform well for short codes, thus leaving room for improvement, and to keep the requirements for the resolution $|\mathbb{U}|$ small which grow with the code length. On the latter account, the dissector $\boldsymbol{d}$ is allowed to take non-zero values only for the k systematic positions of the code. Figure 1 shows the simulation results for $\Delta = 0.1$, $d_{\max} = 4$, 8 decoding iterations and $|\mathbb{U}| = 50, 75, 100$. We observe that the error correcting performance is superior to Turbo decoding with the BPA, and that it increases with $|\mathbb{U}|$ towards the maximum-likelihood (ML) bound. The gain compared to the BPA is up to 0.3 dB for $|\mathbb{U}| = 100$.

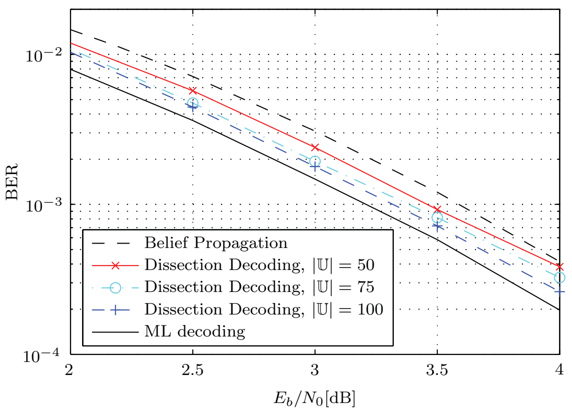


Figure 1. Error Correcting Performance for Turbo Code

V. Discussion & Conclusions

The proposed algorithm shows a distinct error correcting performance gain compared to belief propagation. However, the requirements for the set size $|\mathbb{U}|$ grow approximately proportional to the dissector length – and thus the code length – n. Taking into account the computation of two-dimensional functions over $u \in \mathbb{U}$ in the trellis of length n, the overall decoding complexity is $\mathcal{O}(n^3)$. Ongoing work focuses on Gaussian approximation of the distributions over u [11], leading to $\mathcal{O}(n)$ as for the belief propagation algorithm.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[2] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, “Turbo decoding as an instance of Pearl’s belief propagation algorithm,” *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 140–152, Feb. 1998.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes (1),” in *Proc. IEEE International Conference on Communications*, vol. 2, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[4] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *Information Theory, IEEE Transactions on*, vol. 20, pp. 284 – 287, March 1974.

[5] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[6] S. Benedetto and G. Montorsi, “Serial concatenation of block and convolutional codes,” *Electronics Letters*, vol. 32, pp. 887–888, May 1996.

[7] D. J. MacKay and R. M. Neal, “Good codes based on very sparse matrices,” in *5th IMA Conference on Cryptography and Coding* (C. Boyd, ed.), no. 1025 in Lecture Notes in Computer Science, (Berlin, Germany), pp. 100–111, Springer, 1995.

[8] U. Sorger, “Discriminated Belief Propagation,” Tech. Rep. TR-CSC-07-01, Computer Science and Communications Research Unit, Université du Luxembourg, 2007. http://arxiv.org/abs/0710.5501.

[9] U. Sorger, *Communication Theory*. Books On Demand, 2009.

[10] I. Land, *Reliability Information in Channel Decoding*. PhD thesis, Christian-Albrechts-University of Kiel, Faculty of Engineering, Information and Coding Theory Lab, 2005.

[11] A. Heim, V. Sidorenko, and U. Sorger, “Computation of distributions and their moments in the trellis,” *Advances in Mathematics of Communications (AMC)*, vol. 2, pp. 373–391, November 2008.

Block-Error Performance of Root-LDPC Codes

Iryna Andriyanova
ETIS group
ENSEA/UCP/CNRS-UMR8051
95014 Cergy-Pontoise, France
iryna.andriyanova@ensea.fr

Joseph J. Boutros
Elec. Eng. Department
Texas A&M University at Qatar
23874, Doha, Qatar
boutros@ieee.org

Ezio Biglieri
WISER S.r.l.
Via Fiume 23
57123 Livorno, Italy
e.biglieri@ieee.org

David Declercq
ETIS group
ENSEA/UCP/CNRS-UMR8051
95014 Cergy-Pontoise, France
declercq@ensea.fr

***Abstract*—This paper[1] investigates the error rate of root-LDPC (RLDPC) codes. These codes were introduced in [1], as a class of codes achieving full diversity D over a nonergodic block-fading transmission channel, and hence with an error probability decreasing as SNR^{-D} at high signal-to-noise ratios. As for their structure, root-LDPC codes can be viewed as a special case of multiedge-type LDPC codes [2]. However, RLDPC code optimization for nonergodic channels does not follow the same criteria as those applied for standard ergodic erasure or Gaussian channels. While previous analyses of RLDPC codes were based on their asymptotic *bit* threshold for information variables under iterative decoding, in this work we investigate asymptotic *block* threshold. A stability condition is first derived for a given fading channel realization. Then, in a similar way as for unstructured LDPC codes [3], with the help of Bhattacharyya parameter, we state a sufficient condition for a vanishing block-error probability with the number of decoding iterations.**

I. Introduction and motivation of our work

When a block of encoded data is sent, after being split into F subblocks, through F independent slow-fading channels, the appropriate channel model is nonergodic. This model may correspond to a parallel (MIMO systems) or to a sequential (HARQ protocols) data-transmission scheme.

It turns out that special design criteria are needed for codes to be used with such a model — in particular, full transmit diversity is sought, which guarantees that, at large signal-to-noise ratios (SNR), the error probability of the transmission scheme scales as $1/\mathrm{SNR}^D$, with D the maximum diversity order achievable. It has been shown in [1] that standard sparse-graph code ensembles allow one to obtain error probabilities decreasing only as $1/\mathrm{SNR}$, and hence they are not full-diversity ensembles. Even infinite-length random code ensembles cannot achieve full diversity, as shown via a diversity population evolution technique in [4].

The key idea for codes achieving full diversity is to ensure that each information node is receiving multiple messages affected by independent fading coefficients. This idea has been implemented in RLDPC codes [1] designed for block-fading channels with $F = 2$ by introducing the concept of *root checknodes*. A root checknode protects a message received from the second subchannel when the variable node is received from the first subchannel. RLDPC codes are full-diversity codes (thus, they are also Maximum Distance Separable in the Singleton-bound sense) and can be devised for any diversity order.

In this paper we focus on rate-$1/2$, diversity-2 RLDPC codes, and study their stability under iterative decoding. We also derive a sufficient condition for vanishing block-error probability. As expected, since root checknodes occupy a single edge in each information variable, stability and block-error performance of RLDPC codes depend on the fraction of variables with degrees 2 and 3.

II. Transmission model

Under our assumptions, a block of encoded data (a codeword) is divided into two equal subblocks, each one being transmitted over an independent Rayleigh fading channel with SNR$=\gamma$ and fading coefficients α_1 and α_2. Therefore, the observation y corresponding to the binary transmitted symbol $x = \pm 1$ received from the i-th channel is $y = \alpha_i x + z$, where $\alpha_i \in [0, +\infty)$, and $z \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1/\gamma$.

III. RLDPC Codes: Definition and Density Evolution

A. Definition

Given an initial (λ, ρ) LDPC ensemble, one defines a (λ, ρ) RLDPC ensemble with diversity 2 through the multinomials $\lambda_{\mathrm{root}}(\underline{\mu}, \underline{x})$ and $\rho_{\mathrm{root}}(\underline{\mu}, \underline{x})$, with $\underline{\mu} \triangleq (\mu_1, \mu_2)$ and $\underline{x} \triangleq (x_1, x_2, x_3, x_4, x_5, x_6)$:

$$\lambda_{\mathrm{root}}(\underline{\mu}, \underline{x}) \triangleq \frac{1}{2} \sum_i \left(\frac{\lambda_i}{i} \mu_1 x_1^i + \frac{(i-1)\lambda_i}{i} \mu_1 x_2^i + \lambda_i \mu_1 x_3^i \right.$$
$$\left. + \lambda_i \mu_2 x_4^i + \frac{(i-1)\lambda_i}{i} \mu_2 x_5^i + \frac{\lambda_i}{i} \mu_2 x_6^i \right), \quad (1)$$

$$\rho_{\mathrm{root}}(\underline{\mu}, \underline{x}) \triangleq \frac{1}{2} \sum_i \rho_i \left(x_1 \sum_j \binom{i}{j} f_e^j x_4^j g_e^{i-j} x_5^{i-j} \right.$$
$$\left. + x_6 \sum_k \binom{i}{k} f_e^k x_3^k g_e^{i-k} x_2^{i-k} \right), \quad (2)$$

where the fractions f_e and g_e will be defined in next subsection. In words, the structure of the RLDPC ensemble consists of four types of variable nodes ($1i$, $1p$, $2i$, $2p$), two sets of check nodes ($1c$, $2c$), and 6 different edge classes (see Fig.1a). Permutations of edges within edge classes are chosen uniformly at random. Variable nodes $1i$ and $1p$ correspond to information and redundancy bits, respectively, in a codeword

[1] This work was supported by the European FP7 ICT-STREP DAVINCI project under the contract No. INFSO-ICT-216203.

sent through the first fading subchannel. Similarly, variable nodes $2i$ and $2p$ correspond to bits sent through the second subchannel. Note that the information variable nodes ki ($k = 1, 2$) are connected to check nodes of the same type, kc, through exactly one edge; all other edges are connected to check nodes of the other type. Redundancy variable nodes are always connected to check nodes of different type. In (1)-(2), μ_1 and μ_2 correspond to two fading subchannels, and the variables $x_1, x_2, \ldots, x_6$ to the following edge classes: $1i \to 1c$, $1i \to 2c$, $1p \to 2c$, $2p \to 1c$, $2i \to 1c$, and $2i \to 2c$.

We have thus obtained a code ensemble of rate $1/2$. As shown in [1], such a construction guarantees transmit diversity 2, which is the maximum we can obtain with two independent transmission subchannels.

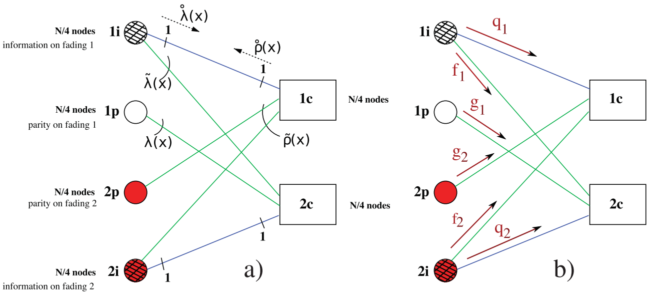


Fig. 1. Structure of a (λ, ρ) RLDPC code ensemble of diversity 2.

B. Density Evolution

RLDPC codes are decoded, as standard LDPC codes, using an iterative algorithm. An asymptotic analysis of iterative decoding is provided in [1], [4] and we shall summarize it here, after giving some notation. We denote the probability density functions (pdfs) of channel LLR outputs from the two transmission subchannels by $\mu_1(x)$ and $\mu_2(x)$, respectively. These are normal pdfs with means $2\alpha_1^2/\gamma$ and $2\alpha_2^2/\gamma$ and variances $4\alpha_1^2/\gamma$ and $4\alpha_2^2/\gamma$, respectively. Further, we denote by $\otimes$ the operation of convolution of two pdfs. We also define the following operation:

Definition 1: The R-convolution of two pdfs $\alpha(x)$ and $\beta(x)$ is

$$\alpha \odot \beta(x) = f(\hat{\alpha}(x) \otimes \hat{\beta}(x)),$$

where

$$\hat{\alpha}(x) \triangleq \frac{2\alpha(2\text{th}^{-1}(x))}{1-x^2}, \quad \hat{\beta}(x) = \frac{2\beta(2\text{th}^{-1}(x))}{1-x^2}$$

and

$$f(x) = \cosh^2\left(\frac{\hat{\alpha} \otimes \hat{\beta}(x)}{2}\right) \text{th}^{-1}(\hat{\alpha} \otimes \hat{\beta}(x)).$$

Note that the R-convolution of pdfs corresponds to the following operation over the corresponding random variables A and B:

$$2\text{th}^{-1}\big(\text{th}(A/2) + \text{th}(B/2)\big),$$

which is exactly the operation performed at the check nodes.

Let us denote the average pdfs for 6 edge sets by $q_1(x)$, $f_1(x)$, $g_1(x)$, $g_2(x), f_2(x)$, and $q_2(x)$ as shown in Fig.1b. Then the evolution of the pdfs at the iteration $m+1$ can be described by the following recursions:

$$\begin{aligned}
q_1^{m+1}(x) &= \mu_1(x) \otimes \mathring{\lambda}(\tilde{\rho}(q_2^m(x), f_e f_1^m(x) + g_e g_1^m(x))) \\
f_1^{m+1}(x) &= \mu_1(x) \otimes \tilde{\lambda}(\tilde{\rho}(q_2^m(x), f_e f_1^m(x) + g_e g_1^m(x))) \\
&\quad \otimes \mathring{\rho}(f_e f_2^m(x) + g_e g_2^m(x)) \\
g_1^{m+1}(x) &= \mu_1(x) \otimes \lambda(\tilde{\rho}(q_2^m(x), f_e f_1^m(x) + g_e g_1^m(x))) \\
g_2^{m+1}(x) &= \mu_2(x) \otimes \lambda(\tilde{\rho}(q_1^m(x), f_e f_2^m(x) + g_e g_2^m(x))) \\
f_2^{m+1}(x) &= \mu_2(x) \otimes \tilde{\lambda}(\tilde{\rho}(q_1^m(x), f_e f_2^m(x) + g_e g_2^m(x))) \\
&\quad \otimes \mathring{\rho}(f_e f_1^m(x) + g_e g_1^m(x)) \\
q_2^{m+1}(x) &= \mu_2(x) \otimes \mathring{\lambda}(\tilde{\rho}(q_1^m(x), f_e f_2^m(x) + g_e g_2^m(x)))
\end{aligned}$$

where we have borrowed from [4] the following notation:

$$\tilde{\lambda}(x) \triangleq \frac{\bar{d}_b}{\bar{d}_b - 1} \sum_i \frac{\lambda_i(i-1)}{i} x^{\otimes(i-2)}; \qquad \bar{d}_b \triangleq 1/\sum_i \lambda_i/i;$$

$$\tilde{\rho}(x) \triangleq \frac{\bar{d}_c}{\bar{d}_c - 1} \sum_i \frac{\rho_i(i-1)}{i} x^{\odot(i-2)}; \qquad \bar{d}_c \triangleq 1/\sum_i \rho_i/i;$$

$$f_e \triangleq \frac{\sum_i (i-1)\frac{\lambda_i}{i}}{\sum_i (i-1)\frac{\lambda_i}{i} + 1} = \frac{\bar{d}_b - 1}{2\bar{d}_b - 1}; \qquad g_e \triangleq 1 - f_e;$$

$$\mathring{\lambda}(x) \triangleq \bar{d}_b \sum_i \frac{\lambda_i}{i} x^{\otimes(i-1)}; \qquad \mathring{\rho}(x) \triangleq \bar{d}_c \sum_i \frac{\rho_i}{i} x^{\odot(i-1)}.$$

Also, we define

$$\tilde{\rho}(q, x) \triangleq \frac{\bar{d}_c}{\bar{d}_c - 1} \sum_i \frac{\rho_i(i-1)}{i} q \odot x^{\odot(i-3)}.$$

IV. Stability Conditions

We are interested in defining stability conditions for RLDPC codes. The main difficulty here lies in the fact that not all messages need be recovered exactly (or, in LDPC jargon, not all pdfs converge to δ_∞). It is not hard to prove that only the pdfs responsible for the convergence of *information* messages, i.e., f_1 and f_2, need to converge for exact recovery of the information bits (this condition is also sufficient). The main concept of the proof is that f_1 and f_2 are strictly "better" than q_1 and q_2.

In this section we derive the stability condition for RLDPC codes based on the recovery of information bits only. Before starting our derivation, let us first apply the traditional stability condition [2] to RLDPC codes, assuming that *all* the code bits should be recovered. In such case the RLDPC codes are simply viewed as a multi-edge code ensemble,

A. RLDPCs as Multi-Edge Codes

The stability condition for multi-edge codes consists in ensuring that the spectral radius of a matrix M is < 1, where $M \triangleq B(\underline{\mu})\Lambda P$, with $B(\underline{\mu})$ the vector of Bhattacharyya parameters for all transmission channels, the Λ matrix corresponding to the variable node side of the graph, and P corresponding to the check node side. Applying the expressions derived in [2], we find that

$$\begin{aligned}
B(\underline{\mu}) &= \begin{pmatrix} B(\mu_1) & B(\mu_1) & B(\mu_1) & B(\mu_2) & B(\mu_2) & B(\mu_2) \end{pmatrix}^T \\
\Lambda &= \begin{pmatrix} \frac{\bar{d}_b\lambda_2}{2} & \frac{\bar{d}_b\lambda_2}{2\bar{d}_b-2} & \lambda_2 & \lambda_2 & \frac{\bar{d}_b\lambda_2}{2\bar{d}_b-2} & \frac{\bar{d}_b\lambda_2}{2} \end{pmatrix} \cdot I \\
P &= \begin{pmatrix} P_2 & P_1 & P_2 & P_3 & P_4 & P_3 \end{pmatrix}^T
\end{aligned}$$

with

$$\begin{aligned}
P_1 &\triangleq \begin{pmatrix} 0 & 0 & 0 & (\bar{d}_c-1)g_e & (\bar{d}_c-1)f_e & 0 \end{pmatrix} \\
P_2 &\triangleq \begin{pmatrix} 0 & \tilde{\rho}'(1)f_e & \tilde{\rho}'(1)g_e & 0 & 0 & \tilde{\rho}(1) \end{pmatrix} \\
P_3 &\triangleq \begin{pmatrix} \tilde{\rho}(1) & 0 & 0 & \tilde{\rho}'(1)f_e & \tilde{\rho}'(1)g_e & 0 \end{pmatrix} \\
P_4 &\triangleq \begin{pmatrix} 0 & (\bar{d}_c-1)f_e & (\bar{d}_c-1)g_e & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

Note that two eigenvalues of M are already 0.

B. RLDPCs as Full-Diversity Codes

By looking at RLDPC as at full-diversity codes, we only ask for the convergence of f_1 and f_2 to δ_∞. To derive a stability condition for this case, assume that, at iteration $m-1$,

$$f_1^{m-1} = \epsilon_1\delta_0 + (1-\epsilon_1)\delta_\infty, \quad f_2^{m-1} = \epsilon_2\delta_0 + (1-\epsilon_2)\delta_\infty.$$

and find an approximation of messages f_1 and f_2 at the next iteration which is linear in ϵ.

To do this, let us first find a linear approximation of $\rho(f_e f(x) + g_e g(x))$:

$$\begin{aligned}
&\rho(f_e f(x)+g_e g(x)) = \rho(f_e\epsilon\delta_0 + f_e(1-\epsilon)\delta_\infty + g_e g(x)) = g(x)^{\odot j-1} \\
&= \sum_j \rho_j \left(g_e^{j-1} g(x)^{\odot j-1} + (j-1)f_e\epsilon \sum_{k=0}^{j-2} \binom{j-2}{k} f_e^{j-2-k} g_e^k \cdot g(x)^{\odot k} \right) \\
&\qquad + c\cdot\delta_\infty = \rho(g_e g(x)) + \epsilon f_e F(g_e g(x)) + c\cdot\delta_\infty,
\end{aligned}$$

where c is a constant, and $\rho(g_e g(x))$ denotes the first term in the sum, while $F(g_e g(x))$ denotes the second one. Over the binary erasure channel, $\rho(g_e g(x))$ and $F(g_e g(x))$ can be computed explicitly, while, in the general case, the two functions should be computed by running the density evolution iterations. Also note that one can bound the pdf of $g(x)$ by the initial pdf corresponding to the channel estimate. If the transmission channel is bad, the bound will be quite tight. Next,

$$\begin{aligned}
&\tilde{\rho}(q_2(x), f_e f(x) + g_e g(x)) = \sum_j \rho_j g_e^{j-2} q_2(x) \odot \left(g(x)^{\odot j-2}\right) \\
&+ \sum_j \rho_j (j-2) f_e\epsilon \sum_{k=0}^{j-3} \binom{j-3}{k} f_e^{j-3-k} g_e^k \cdot q_2(x) \odot \left(g(x)^{\odot k}\right) \\
&\quad + c\cdot\delta_\infty = \tilde{\rho}(q_2(x), g_e g(x)) + \epsilon f_e F(q_2(x), g_e g(x)) + c\cdot\delta_\infty.
\end{aligned}$$

Further calculations yield

$$\begin{aligned}
&\tilde{\lambda}(\tilde{\rho}(q_2(x), f_e f(x) + g_e g(x)) = \\
&\qquad = \tilde{\lambda}(\tilde{\rho}(q_2(x), f_e\epsilon\delta_0 + f_e(1-\epsilon)\delta_\infty + g_e g(x)) \\
&= \tilde{\lambda}_1 + \tilde{\lambda}_2 \left(\tilde{\rho}(q_2(x), g_e g(x)) + \epsilon f_e F(q_2(x), g_e g(x))\right) + c\cdot\delta_\infty.
\end{aligned}$$

Finally, the approximation of f_1 linear in ϵ is obtained as

$$\begin{aligned}
f_1 &= \mu_1(x) \otimes \Big(\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_2(x), g_e g_1(x)) + \tilde{\lambda}_2\epsilon_1 f_e F(q_2(x), g_e g_1(x))\Big) \\
&\quad \otimes (\rho(g_e g_2(x)) + \epsilon_2 f_e F(g_e g_2(x))) + const\cdot\delta_\infty \\
&= \mu_1(x) \otimes \Big([\tilde{\lambda}_1 + \tilde{\lambda}_2\rho(q_2(x), g_e g_1(x))] \otimes \rho(g_e g_2(x)) \\
&\quad + \epsilon_2 f_e[\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_2(x), g_e g_1(x))] \otimes F(g_e g_2(x)) \\
&\quad + \tilde{\lambda}_2\epsilon_1 f_e F(q_2(x), g_e g_1(x)) \otimes (\rho(g_e g_2(x)))\Big) + c\cdot\delta_\infty \\
&= \mu_1(x) \otimes (C_0(x) + \epsilon_1 f_e C_1(x) + \epsilon_2 f_e C_2(x)) + c\cdot\delta_\infty
\end{aligned}$$

where

$$\begin{aligned}
C_0(x) &\triangleq [\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_2(x), g_e g_1(x))] \otimes \rho(g_e g_2(x)) \\
C_1(x) &\triangleq \tilde{\lambda}_2 F(q_2(x), g_e g_1(x)) \otimes \rho(g_e g_2(x)) \\
C_2(x) &\triangleq [\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_2(x), g_e g_1(x))] \otimes F(g_e g_2(x))
\end{aligned}$$

Similarly,

$$f_2 = \mu_2(x) \otimes \left(\tilde{C}_0(x) + \epsilon_1 f_e\tilde{C}_1(x) + \epsilon_2 f_e\tilde{C}_2(x)\right) + c\cdot\delta_\infty$$

with

$$\begin{aligned}
\tilde{C}_0(x) &\triangleq [\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_1(x), g_e g_2(x))] \otimes \rho(g_e g_1(x)) \\
\tilde{C}_1(x) &\triangleq [\tilde{\lambda}_1 + \tilde{\lambda}_2\tilde{\rho}(q_1(x), g_e g_2(x))] \otimes F(g_e g_1(x)) \\
\tilde{C}_2(x) &\triangleq \tilde{\lambda}_2 F(q_1(x), g_e g_2(x)) \otimes \rho(g_e g_1(x))
\end{aligned}$$

Therefore, we have the following relation:

$$\begin{pmatrix} f_1^m \\ f_2^m \end{pmatrix} = \underline{a} + f_e A \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},$$

with

$$\underline{a} \triangleq \begin{pmatrix} \mu_1(x) \otimes C_0(x) \\ \mu_2(x) \otimes \tilde{C}_0(x) \end{pmatrix}$$

and

$$A \triangleq \begin{pmatrix} \mu_1(x) \otimes C_1(x) & \mu_1(x) \otimes C_2(x) \\ \mu_2(x) \otimes \tilde{C}_1(x) & \mu_2(x) \otimes \tilde{C}_2(x) \end{pmatrix}$$

Denote now by $q(f)$ the Bhattacharyya parameter related to the pdf f,

$$B(f) \triangleq \int_R e^{-x/2} f(x)dx.$$

B is closely related to the bit error probability P_b corresponding to $f(x)$, and it has been shown in [5] that $P_b \to 0 \Leftrightarrow B(f) \to 0$. Knowing this, and taking into account the properties of convolution and of R-convolution, we obtain that

$$\begin{pmatrix} B(f_1^m) \\ B(f_2^m) \end{pmatrix} \leq [C + f_e B(A)] \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},$$

where

$$C = \begin{pmatrix} 0 & B(\mu_2)\frac{\lambda_2 g_e}{2}\rho'(g_e) \\ B(\mu_1)\frac{\lambda_2 g_e}{2}\rho'(g_e) & 0 \end{pmatrix}.$$

Note that we simplified the expressions by bounding $1 - B(g_1) \leq 1 - B(q_1)$ and $1 - B(g_2) \leq 1 - B(q_2)$, and further bounding $C_0(x)$ and $\tilde{C}_0(x)$.

Define next $D \triangleq C + f_e B(A)$. Then the following recurrence relation can be obtained:

$$\begin{pmatrix} B(f_1^m) \\ B(f_2^m) \end{pmatrix} \leq D \cdot \begin{pmatrix} B(f_1^{m-1}) \\ B(f_2^{m-1}) \end{pmatrix},$$

and hence, if we perform m iterations of density evolution, we obtain that

$$\left(\begin{array}{c} B(f_1^m) \\ B(f_2^m) \end{array} \right) \leq D^m \cdot \left(\begin{array}{c} B(f_1^0) \\ B(f_2^0) \end{array} \right),$$

where we assume that the messages q and g for any iteration are bounded by q^0 and g^0. We are interested in the case of $B(f^\infty)$ decreasing to 0.

Taking all the above into account, we have the following sufficient stability condition for full-diversity codes:

Theorem 1 (Sufficiency part of the stability condition): The bit error probability P_e for a full-diversity RLDPC ensemble converges to 0 if all the absolute values of the eigenvalues of D are < 1.

Notice that the usual stability condition mentioned in Section IV-A depends on λ_2, while the stability condition derived here depends on both λ_2 and λ_3, "hidden" in $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$.

V. Block-Error Rate of RLDPC Codes

The main result of this paper is the study of the block-error probability P_B of RLDPC codes. Using the sufficient part of the stability condition derived above, we can link P_B to the bit-error probability P_b, and show in which cases $P_b \to 0$ implies $P_B \to 0$.

Using a union bound at some iteration m, we obtain

$$\begin{aligned} P_B^m &\leq \frac{n}{4} P_b^l(1i) + \frac{n}{4} P_b^l(2i) \\ &\leq \frac{1}{4} (\max M_l(1i))^{6+\varepsilon} P_b^l(1i) + \frac{1}{4} (\max M_l(2i))^{6+\varepsilon} P_b^l(2i), \end{aligned}$$

where n is the code length, and $\max M_m(1i)$ ($\max M_m(2i)$) is the maximum number of variable nodes in a computation tree of a variable node from the set $1i$ ($2i$) in the bipartite graph, after m iterations. The second inequality follows from the same reasoning used in [3, Section II], to which we refer the reader desiring a detailed proof. Now, to ensure that, as $m \to \infty$, P_B^m decreases to 0 while $P_b^m \to 0$, one has to ensure that P_b^m decreases with m faster than the maximum number of variable nodes in the computation tree.

A. Case of $\lambda_2 = \lambda_3 = 0$

Let us consider the simple case of both λ_2 and λ_3 being 0. (this is similar to the case of standard LDPC codes with $\lambda_2 = 0$). Repeating the calculations of [3, Section VI.A], we obtain

$$\begin{aligned} &P_B(m+k) \leq \\ &\frac{1}{4} (d_v^{max} d_c^{max})^{6(1+\varepsilon)(m+k)} [B(f_1^m)^{(3/2)^k} + B(f_2^m)^{(3/2)^k}], \end{aligned}$$

which decreases to 0 as $k \to \infty$.

B. General case

Given that

$$B(\text{output}) = \Pi_i B(\text{input}_i)$$

for variable nodes and

$$1 - B(\text{output}) \geq \Pi_i (1 - B(\text{input}_i))$$

for check nodes, and since $B(q^m)$ and $B(g^m)$ for any m, q, and g are no greater than the corresponding $B(\mu)$, one can bound

$$\begin{aligned} B(C_1) &\leq \tilde{\lambda}_2 g_e \rho'(g_e) B(\mu_2) \max\{B(\mu_1), B(\mu_2)\} \\ B(C_2) &\leq (\tilde{\lambda}_1 + \lambda_2 \rho(g_e)) g_e B(\mu_2) \\ B(\tilde{C}_1) &\leq (\tilde{\lambda}_1 + \lambda_2 \rho(g_e)) g_e B(\mu_1) \\ B(\tilde{C}_2) &\leq \tilde{\lambda}_2 g_e \rho'(g_e) B(\mu_1) \max\{B(\mu_1), B(\mu_2)\} \end{aligned}$$

and obtain

$$\begin{aligned} B(f_1^m) &\leq B(\mu_2)(w_1 B(f_1^{m-1}) + w_2 B(f_2^{m-1})) \\ B(f_2^m) &\leq B(\mu_1)(w_2 B(f_1^{m-1}) + w_1 B(f_2^{m-1})) \end{aligned}$$

with $w_1 \triangleq f_e \tilde{\lambda}_2 \rho'(g_e)$ and $w_2 \triangleq f_e(\tilde{\lambda}_1 + \tilde{\lambda}_2 \rho(g_e)) + \frac{\lambda_2 g_e}{2} \rho'(g_e)$. Thus, with a linear approximation,

$$\begin{aligned} B(f_1^{m+2k}) &\leq B(\mu_2)^{2k} w_1^{2k} B(f_1^m) \\ &\quad + B(\mu_2)^k B(\mu_2)^k w_1^k w_2^k B(f_2^m) \\ B(f_2^{m+2k}) &\leq B(\mu_1)^{2k} w_1^{2k} B(f_2^m) \\ &\quad + B(\mu_2)^k B(\mu_2)^k w_1^k w_2^k B(f_1^m). \end{aligned}$$

Consequently, the block error probability

$$\begin{aligned} &P_B(m+k) \leq \\ &\quad \frac{1}{4} (d_v^{max} d_c^{max})^{6(1+\varepsilon)(m+k)} [B(\mu_2)^{2k} w_1^{2k} B(f_1^m) + \\ &B(\mu_1)^{2k} w_1^{2k} B(f_2^m) + B(\mu_2)^k B(\mu_2)^k w_1^k w_2^k \{B(f_1^m) + B(f_2^m)\}], \end{aligned}$$

can be seen to decrease to 0, as $k \to \infty$, if the following conditions are satisfied:

$$\begin{aligned} B(\mu_2) w_1 &\leq (d_v^{max} d_c^{max})^{-3}, \\ B(\mu_1) w_2 &\leq (d_v^{max} d_c^{max})^{-3}. \end{aligned}$$

VI. Conclusion

In this paper we have derived the conditions under which the block-error rate of a RLDPC code ensemble decreases to 0 as the bit-error rate does the same. The interest of our findings lies in the fact that results existing in the literature deal with errors related to all the of code bits, while for RLDPC only errors affecting information bits should be considered.

References

[1] J. Boutros, A. G. i Fabregas, E. Biglieri, and G. Zemor, "Low-density parity-check codes for nonergodic block-fading channels," 2007, submitted to IEEE Trans. Inform. Theory.

[2] T. Richardson and R. Urbanke, "Multi-edge LDPC codes," 2004, submitted to IEEE Trans. Inform. Theory. [Online]. Available: http://ece.iisc.ernet.in/∼ vijay/multiedge.pdf

[3] H. Jin and T. Richardson, "Block error iterative decoding capacity for ldpc codes," in *ISIT'05*, Adelaide, Australia, September 2005.

[4] J. Boutros, "Diversity and coding gain evolution in graph codes," in *ITA'09*, San-Diego, USA, February 2009.

[5] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

On Iterative Decoding of HDPC Codes Using Weight-Bounding Graph Operations

Joakim Grahl Knudsen, Constanza Riera*, Lars Eirik Danielsen, Matthew G. Parker, and Eirik Rosnes
Dept. of Informatics, University of Bergen, Thormøhlensgt. 55, 5008 Bergen, Norway
email: {joakimk, larsed, matthew, eirik}@ii.uib.no
*Bergen University College, Nygårdsgt. 112, 5008 Bergen, Norway, email: csr@hib.no

***Abstract*—In this paper, we extend our work on iterative soft-input soft-output (SISO) decoding of high density parity check (HDPC) codes. Edge-local complementation (ELC) is a graph operation which can be used to give structural diversity during decoding with the sum-product algorithm (SPA). We describe the specific subgraphs required for ELC to not increase the weight of the Tanner graph beyond a specified upper bound. We call this controlled operation weight-bounding ELC (WBELC). A generalized iterative SISO HDPC decoder based on SPA decoding is described, which can be configured to employ our SPA-ELC decoders, or iterative permutation decoding (SPA-PD). The latter is a state-of-the-art decoding algorithm for HDPC codes, using permutations from the automorphism group of the code. We observe performance improvements over SPA-PD when the SISO HDPC decoder is configured to use SPA-ELC in conjunction with WBELC.**

I. Introduction

Iterative soft decision decoding algorithms are known to give results which approach the theoretical limits postulated by Shannon [1]. Specifically, the use of such algorithms for the decoding of random, sparse linear codes yields near-optimum error-rate performance when the blocklength goes to infinity. The best known instance is low density parity check codes, decoded with the sum-product algorithm (SPA). Inspired by these results, the aim of much research has been to develop practical (non-asymptotic) codes and decoders exhibiting comparable performance. Recently, iterative decoding techniques have been adapted to classical linear codes, which have strong structural properties (large minimum distance, and small description complexity in hardware implementation), but are non-sparse. One state-of-the-art decoder for such *high density parity check* (HDPC) codes [2] is the iterative permutation decoder (SPA-PD) [3], which performs very well on Bose-Chaudhuri-Hocquenghem codes, as well as on quadratic residue codes [4, 5], over the additive white Gaussian noise (AWGN) channel. Our paper is an extension of our previous work on iterative, graph-local decoding of HDPC codes using a graph operation known as edge-local complementation (ELC) [5, 6]. The contribution of this work is the description of subgraphs on which ELC will not increase the number of edges in the graph beyond a desired threshold–a trait we call weight-bounding ELC (WBELC). We describe an SPA-WBELC algorithm – an instance of a generalized soft-input soft-output (SISO) HDPC decoder – which gives an improvement over our previous algorithm, SPA-ELC [5]. We also extend our scope towards less structured HDPC codes (i.e., smaller automorphism group), for which we also observe an improvement over SPA-PD. Most significantly, we show a gain when the size of the automorphism group is one–moving towards random codes–in which case SPA-PD ‘reduces’ to SPA.

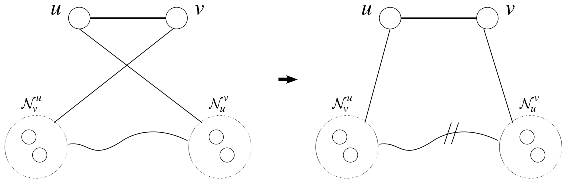
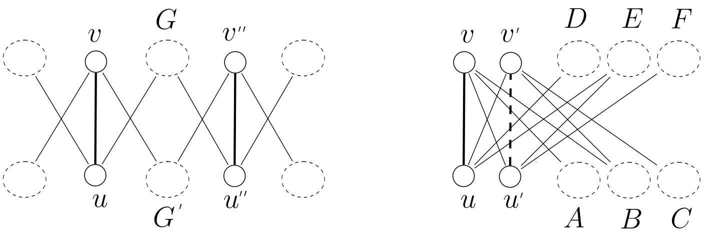


Fig. 1. ELC on edge (u, v) of a bipartite simple graph. Doubly slashed links mean that the edges connecting the two sets have been complemented; edges are replaced by non-edges, and vice versa. This graph may be a subgraph of a larger graph.

A binary linear code $\mathcal{C}$ of length n and dimension k is denoted by $[n, k, d_{\mathbf{min}}]$, and $\mathcal{C}^{\perp}$ is its dual. The automporhism group is denoted by $\mathrm{Aut}(\mathcal{C})$, and if it consists of the identity permutation alone, we say that $\mathrm{Aut}(\mathcal{C})$ is trivial. The $(n-k) \times n$ parity check matrix and the corresponding Tanner graph are denoted by H and $\mathbf{TG}(H)$, respectively. All definitions regarding H have obvious equivalents for $\mathbf{TG}(H)$, and vice versa, so we will use these representations interchangeably. H is said to be systematic if its columns can be reordered into the form $[I\, P]$, where I is the identity matrix of size $n - k$. The transpose of H is written H^T. The weight of H, denoted by $|H|$, is the number of non-zero entries in H, and the minimum weight of H is lower-bounded by $\max(k(d_{\mathbf{min}}(\mathcal{C}) - 1) + n - k,\ (n - k)d_{\mathbf{min}}(\mathcal{C}^{\perp}))$. Accordingly, the number of edges of $\mathbf{TG}(H)$ is $|H|$. The local neighborhood of a node v is the set of nodes adjacent to v, and is denoted by $\mathcal{N}_v$, while $\mathcal{N}_v^u$ is shorthand notation for $\mathcal{N}_v \setminus \{u\}$. $|\mathcal{E}_{A,B}|$ denotes the number of edges in the subgraph induced by the nodes in $A \cup B$. $\mathcal{E}_{u,v}$ is shorthand notation for $\mathcal{E}_{\mathcal{N}_u^v, \mathcal{N}_v^u}$, the local neighborhood of the edge (u, v). ELC requires that H is systematic, so, as a simplification, we may describe the subgraphs on which ELC is WBELC using a simple bipartite graph (undirected, no double edges) $G = \left(\begin{smallmatrix} 0 & P \\ P^T & 0 \end{smallmatrix}\right)$. By taking the P-part as one of the two partitions, G is equivalent to $\mathbf{TG}(H)$, and straight-forward mappings exist to implement ELC operations directly on $\mathbf{TG}(H)$ [5]. The operation of ELC on an edge (u, v) is to complement the edges of $\mathcal{E}_{u,v}$, followed by swapping the nodes u and v–see Fig. 1. In the following,



(a) Theorem 4

(b) Theorem 5

Fig. 2. Depth-2 WBELC. The dashed edge in Theorem 5 is a non-edge. The edges between sets in the (bipartite) subgraphs are not shown.

we will use boldface notation for vectors.

The following section describes WBELC. The remainder of this article details the application of this controlled ELC operation in a SISO HDPC decoding algorithm, in an extension of our previous work on the SPA-ELC decoder. Finally, we present simulations results, and compare the decoding algorithms.

II. WBELC

The effect of repeated random ELC is that the average weight of H tends to $\frac{k(n-k)}{2} + (n-k)$. In this section, we introduce a restriction on the ELC operation, being that an ELC on a certain edge in the graph is only allowed if $|H|$ remains below a given threshold, T. We give a complete description of the conditions that are necessary and sufficient in order to achieve this bound, both for a single ELC and for two consecutive ELCs. Using these conditions, we improve the perormance of the SISO HDPC decoder.

We begin by formalizing the notion of WBELC. If the weight change due to the complementation caused by ELC is bounded, then the weight of the entire graph is bounded, and we say that the ELC is WBELC.

A. Depth 1

There is a simple condition for one ELC to be WBELC.

Theorem 1: ELC on (u, v) does not increase the weight of the graph by more than a threshold T *iff*

$$|\mathcal{N}_u^v||\mathcal{N}_v^u| - 2|\mathcal{E}_{u,v}| \leq T.$$

B. Depth 2

For many graphs, it is simply not possible to bound the weight increase by any reasonable threshold using only a single ELC. The notion of WBELC can be extended to the case of consecutive ELC operations. In this work, we will completely characterize WBELC to within depth 2, where we use the compact notation $\{(u, v), (u', v')\}$ for an ordered pair of edges. Incidentally, the search space can be significantly reduced from checking all pairs of edges in the graph.

Theorem 2: ELC on $\{(u, v), (v, v')\}$, where $v' \in \mathcal{N}_u^v$, gives the same graph as ELC on (u, v'). Consequently, depth-2 WBELC reduces in this case to depth-1 WBELC.

Note that, due to the swap of ELC on (u, v), (v, v') and (u, v') refer to the same edge–see Fig. 1. From this theorem, we see that we need only consider pairs of non-adjacent edges, i.e., at a distance of at least one edge apart. However, it can be shown that the search space can be further reduced by noting that the distance can also not be greater than two edges.

Theorem 3: Let $T \geq -1$. Any depth-2 WBELC where the pair of edges are at a distance greater than two edges apart, will always reduce to either one or two separate instances of depth-1 WBELC.

One implication of Theorem 3 is that depth-2 WBELC, like depth-1 WBELC, only acts locally on a graph. For $T < -1$, there is an additional case (not discussed in this paper), not covered by Theorem 3. Thus, the following three theorems describe all possible depth-2 WBELC cases for $T \geq -1$.

Let us first consider the case where the pair of edges are at a distance of exactly two edges apart, Fig. 2(a). Given an edge (u, v), let $u'', v'' \notin \mathcal{N}_u \cup \mathcal{N}_v$ be such that $G = \mathcal{N}_u^v \cap \mathcal{N}_{u''}^{v''} \neq \emptyset$, and, similarily, $G' = \mathcal{N}_{v''}^{u''} \cap \mathcal{N}_v^u \neq \emptyset$.

Theorem 4: ELC on $\{(u, v), (u'', v'')\}$ does not increase the weight of the graph by more than a threshold T *iff*

$$|\mathcal{N}_u^v||\mathcal{N}_v^u| + |\mathcal{N}_{u''}^{v''}||\mathcal{N}_{v''}^{u''}| - 2|\mathcal{E}_{u,v}| \ + \\ 4|\mathcal{E}_{G,G'}| - 2|\mathcal{E}_{\mathcal{N}_{u''}^{v''},\mathcal{N}_{v''}^{u''}}| - 2|G||G'| \ \leq T.$$

For the next theorem, given an edge (u, v) and two nodes u' and v', we denote by $B = N_v^{u,u'} \cap N_{v'}^{u,u'}$, $A = N_v^{u,u'} \setminus B$, $C = N_{v'}^{v,v'} \setminus B$, $E = N_u^{v,v'} \cap N_{u'}^{v,v'}$, $D = N_u^{v,v'} \setminus E$, and $F = N_{u'}^{v,v'} \setminus E$, see Fig. 2(b).

We now consider the case where both u' and v' are in the neighborhood of (u, v).

Theorem 5: ELC on $\{(u, v), (u', v')\}$ does not increase the weight of the graph by more than a threshold T *iff*

$$|F| - |E| - |B| - 2|\mathcal{E}_{A,E\cup F}| - 2|\mathcal{E}_{B,D\cup E}| - 2|\mathcal{E}_{C,D\cup F}| \ + \\ |C| + |A|(|E| + |F|) + |B|(|D| + |E|) + |C|(|D| + |F|) \leq T.$$

Note that the edge (u', v') is created by the first ELC. Last, we consider the case where either u' or v' belong to $\mathcal{N}_u^v \cup \mathcal{N}_v^u$, but not both. Without loss of generality, let $v' \in \mathcal{N}_u^v$ be connected to $u' \notin \mathcal{N}_v^u$.

Theorem 6: ELC on $\{(u, v), (u', v')\}$ gives the same graph as ELC on $\{(u, v'), (u', v)\}$.

Note that $\{(u, v'), (u', v)\}$ is covered by Theorem 5.

III. Iterative SISO HDPC Decoding

We have previously described the SPA-ELC decoder, which, essentially, consists of SPA iterations interspersed with random ELC operations [5]. Since ELC complements edges, we avoid loss of extrinsic information (on edges) by executing a *flooding* scheduling SPA iteration in the order 'functions, then variables.' At this point, all messages, μ, have been accumulated in variable nodes, making it safe to change the graph. A generalized SISO HDPC decoder is listed in Algorithm 1, which can be configured to perform the decoding algorithms compared in this work–see Section IV.

Both SPA-PD and SPA-ELC suffer a performance loss if the extrinsic contribution of the soft input vector, $\mathbf{L}$, is not scaled down (damped) in between iterations. For each variable node, v, the SPA produces a decision based on two pieces of information; the extrinsic information produced by the decoder, and the input to iteration j, L_j^v. $\mathbf{L}_0$ is the received noisy channel vector and τ is the maximum number of decoder iterations. The damping coefficient, $\alpha_0 \leq \alpha \leq 1$, represents the amount of 'trust' in the extrinsic information versus the input after the current iteration [2], $L_{j+1}^v := L_j^v + \alpha(\Sigma_{u \in \mathcal{N}_v} \mu_j^{v \leftarrow u})$, $\forall\, u \in \mathcal{N}_v$. As the decoder converges, the information produced by the graph is assumed to become more reliable (hopefully converging towards the maximum-likelihood codeword), so our trust in the decoder state may be increased accordingly. This is normally reflected by incrementing α with iteration number j. A *global* damping rule (GD) scales down all variable nodes, and re-initializes all edges, $\mu_{j+1}^{v \rightarrow u} := L_{j+1}^v$, $\forall\, v \in \mathbf{TG}(H)$. We propose an *edge-local* damping rule (LD), which restricts the application of the damping-and-initialization rule to new edges due to ELC on (u, v), $\mu_{j+1}^{v' \rightarrow u'} := L_{j+1}^v$, $\forall\, (u', v') \in \mathcal{E}_{u,v}$. All other edges retain messages computed in iteration j.

SPA-PD applies a random permutation (PD) $\mathbf{L}_j := \sigma(\mathbf{L}_j)$, $\sigma \in \mathrm{Aut}(\mathcal{C})$, before re-initializing $\mathbf{TG}(H)$ with global damping. SPA decoding on a fixed graph suffers a performance loss when global damping is applied, which suggests that the benefit of damping is to moderate the effects of modifications (e.g., permutations, Gaussian elimination, ELC) to $\mathbf{TG}(H)$. Note that damping is disabled by configuring $\alpha_0 := 1$.

A. SPA-WBELC

The SPA-WBELC algorithm uses the theorems in Section II to determine a random WBELC operation on the current $\mathbf{TG}(H)$, and applies the corresponding one or two ELC operations, with edge-local damping. Let H_j denote the matrix after j iterations of the SISO HDPC decoder. It is helpful to reduce the weight of the initial matrix, H_0, in a preprocessing stage, as this has a positive effect on SPA decoding. This can be done using repeated random WBELC with $T = -1$, for non-increasing weight. A simple but effective heuristic, if the preprocessing gets stuck, is to allow one random (i.e., unbounded) ELC. Then, for SPA-WBELC decoding, a threshold $T \geq -1$ must be determined, such that WBELC yields a sufficient number of distinct matrices of weight $|H| \leq |H_0| + T$, to give structural diversity during decoding.

Algorithm 1 SISO-HDPC$(p, I_1, I_2, I_3, \alpha_0, \mathrm{OP}, \mathrm{DR})$

```
$\alpha = \alpha_0$
for $I_3$ times do
    Restart decoder from channel vector
    for $I_2$ times do
        Stop if syndrome check is satisfied
        Apply damping rule, DR, with coefficient $\alpha$
        Apply at random $p$ operations, OP
        for $I_1$ times do
            Apply SPA iteration ('flooding' scheduling)
        end for
    end for
    Increment $\alpha$ towards 1
end for
```

IV. Results

The aim of this paper is to explore the effects of ELC decoding, while maintaining a bound on the weight of $\mathbf{TG}(H)$. We will show that the SPA-WBELC decoder outperforms SPA-PD when $\mathrm{Aut}(\mathcal{C})$ is small. For this work, we chose the best codes we could find at practical blocklengths: two extremal (in terms of minimum distance) self-dual $[36, 18, 8]$ and $[38, 19, 8]$ codes from [7], and an extremal double circulant self-dual $[68, 34, 12]$ code from [8]. We use the notation $\mathcal{C}^n$ to refer to these codes, and we have that $|\mathrm{Aut}(\mathcal{C}^n)| \approx n$, except $\mathcal{C}^{38}$ which has a trivial $\mathrm{Aut}(\mathcal{C})$.

The matrices used were optimized on weight, both in non-systematic form (for SPA and SPA-PD), as well as systematic form (for SPA-ELC and SPA-WBELC). For $\mathcal{C}^{36}$ and $\mathcal{C}^{38}$, we were able to compute the entire ELC orbit of the codes, to find optimal-weight matrices in systematic form to be $|H_0^{36}| = 156$ and $|H_0^{38}| = 166$. For $\mathcal{C}^{68}$, the orbit is infeasibly large, yet, using WBELC preprocessing, we were able to find a systematic matrix of weight $|H_0^{68}| = 488$. For non-systematic form, minimum-weight codewords of $\mathcal{C}^{\perp}$ were combined to assemble matrices of weight 152, 154, and 492, respectively, which is very close to the lower bound based on $d_{\mathbf{min}}(\mathcal{C}^{\perp})$.

The simulation results compare the proposed SPA-WBELC$(p, I_1, I_2, I_3, \alpha_0, T)$ = SISO-HDPC$(p, I_1, I_2, I_3, \alpha_0,$WBELC$(T), \mathrm{LD})$ decoder against standard SPA(τ) = SISO-HDPC$(0, 1, \tau, 1, 1, -, -)$, where we ensure that $\tau = I_1 I_2 I_3$; SPA-PD$(I_1, I_2, I_3, \alpha_0)$ = SISO-HDPC$(1, I_1, I_2, I_3, \alpha_0, \mathrm{PD}, \mathrm{GD})$; and our previous ELC decoder, SPA-ELC$(p, I_1, I_2, I_3, \alpha_0)$ = SISO-HDPC$(p, I_1, I_2, I_3, \alpha_0, \mathrm{ELC}, \mathrm{LD})$. We compare frame-error rate (FER) when signalling over the AWGN channel, and measure complexity in SPA messages, $\frac{1}{F}\Sigma_F \Sigma_{j=0}^{J \leq \tau} |H_j|$, where J is the number of iterations used for a frame, and F the total number of frames simulated. For comparisons between SPA-ELC and SPA-WBELC, we use a comparative number, p, of ELC operations (one WBELC is one or two ELC operations). The most significant result is that SPA-WBELC outperforms SPA-PD in FER on $\mathcal{C}^{38}$ and $\mathcal{C}^{36}$, even when $\mathrm{Aut}(\mathcal{C})$ is non-trivial. For $\mathcal{C}^{68}$, we approach the performance of SPA-PD quite closely. In addition, we see that SPA-WBELC will generally result in an improvement over SPA-ELC. This gain is consistent for all codes attempted, and is most significant at low signal-to-noise ratio (SNR). At high SNR, the performance of SPA-WBELC will, in general, approach that of SPA-ELC. This is assumed to be linked to the average number of iterations per frame approaching zero, such that the number of operations (ELC or WBELC) also goes down, diminishing the difference between the respective decoders. The point at which the performance of SPA-WBELC 'breaks off' towards SPA-ELC is influenced by the choice of T. By increasing T, the break occurs at higher SNR. Yet, this is obviously at the expense of increased average weight, such that, for some T sufficiently high, SPA-WBELC equals SPA-ELC also at low SNR.

For $\mathcal{C}^{38}$, $\mathrm{Aut}(\mathcal{C})$ is trivial, such that SPA-PD 'reduces' to SPA. In this extreme setting, ELC-based decoding has its most interesting gain. The SISO HDPC decoder is sensitive to choice of parameters, so various configurations (of T, I_1, I_2, I_3, and p) were systematically attempted in order to arrive at the presented data.

For complexity, we observe the desired effect of bounding the weight increase due to ELC. For SPA-ELC, the average weight of H quickly settles around $k(k+2)/2$ (the codes are self-dual), whereas for SPA-WBELC, the average weight is $|H_0| + T$. The SPA-WBELC decoder has a uniform improvement in complexity over both SPA and SPA-ELC, and can also be pushed down quite close to SPA-PD. We have also simulated SPA and SPA-PD on systematic matrices (not shown), to verify that FER performance is not significantly sensitive to this.

V. Conclusion

We have developed a new algorithm for the decoding of linear codes on graphs, which is particularly suited for HDPC codes. The main idea of this work is to use a graph operation, ELC, in a controlled manner. We described the necessary and sufficient conditions for this operation to be weight-bounding, and discuss its application in SPA decoding. The results show a significant improvement over standard (flooding) SPA, our previous algorithm SPA-ELC, as well as over SPA-PD in codes with limited structure.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] J. Jiang and K. R. Narayanan, "Iterative soft decision decoding of Reed-Solomon codes," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 244–246, Apr. 2004.

[3] T. R. Halford and K. M. Chugg, "Random redundant iterative soft-in soft-out decoding," *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 513–517, Apr. 2008.

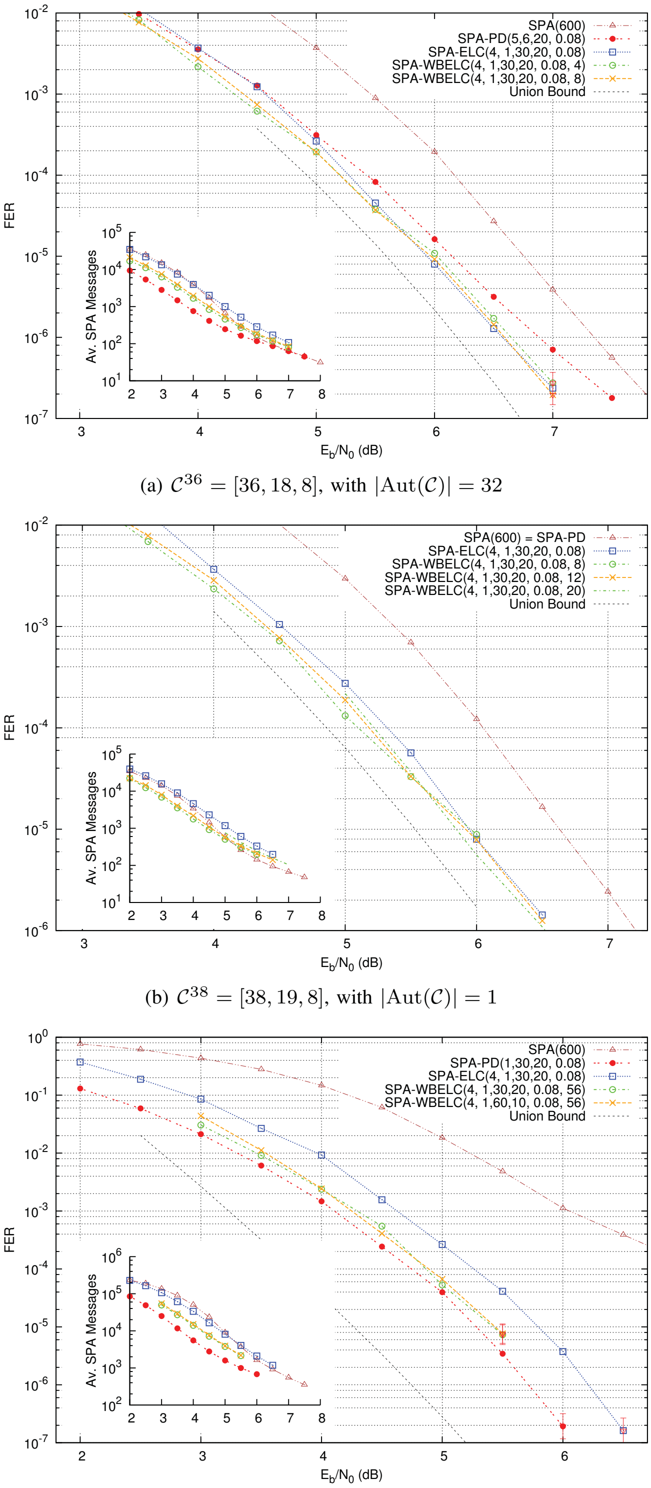
[4] I. Dimnik and Y. Be'ery, "Improved random redundant iterative HDPC decoding," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1982–1985, Jul. 2009.

[5] J. G. Knudsen, C. Riera, L. E. Danielsen, M. G. Parker, and E. Rosnes, "Random edge local complementation and iterative soft-decision decoding," in *Proc. Int. Symp. Inform. Theory*, Seoul, Korea, Jul. 2009, pp. 899–903.

[6] A. Bouchet, "Isotropic systems," *European J. Comb.*, vol. 8, pp. 231–244, Jul. 1987.

[7] M. Harada, "New extremal self-dual codes of lengths 36 and 38," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2541–2543, Nov. 1999.

[8] T. A. Gulliver and M. Harada, "Classification of extremal double circulant self-dual codes of lengths 64 to 72," *Des. Codes Cryptogr.*, vol. 13, pp. 257–269, 1998.



(a) $\mathcal{C}^{36} = [36, 18, 8]$, with $|\mathrm{Aut}(\mathcal{C})| = 32$

(b) $\mathcal{C}^{38} = [38, 19, 8]$, with $|\mathrm{Aut}(\mathcal{C})| = 1$

(c) $\mathcal{C}^{68} = [68, 34, 12]$, with $|\mathrm{Aut}(\mathcal{C})| = 68$

Fig. 3. Simulations results. Each SNR point is simulated until at least 100 frame-error events were observed (otherwise, error bars indicate significance). The union bound is calculated based on the full weight enumerator of the code.

Generalized Minimum Distance Decoding for Correcting Array Errors

Vladimir R. Sidorenko, Martin Bossert
Inst. of Telecommunications and Applied Information Theory
Ulm University, Ulm, Germany,
{vladimir.sidorenko | martin.bossert}@uni-ulm.de

Ernst M. Gabidulin
Moscow Institute (State University) of Physics and Technology
Dolgoprudny, Russia
ernst.gabidulin@gmail.com

***Abstract*—We consider an array error model for data in matrix form, where the corrupted symbols are confined to a number of lines (rows and columns) of the matrix. Codes in array metric (maximum term rank metric) are well suited for error correction in this case. We generalize the array metric to the case when the reliability of every line of the matrix is available. We propose a minimum distance decoder for the generalized metric and estimate the guaranteed error correcting radius for this decoder.**

I. Introduction

Consider transmission of matrices C with elements from the field $\mathbb{F}_q$ over a channel with array (or crisscross) errors. This channel corrupt a number of lines (rows and columns) of the matrix C, i.e., the channel may erase some lines and replace components of some other lines by arbitrary elements of $\mathbb{F}_q$. Array errors can be found in various data storage applications and in OFDM systems. The array metric, which is also known as the maximal-term-rank metric, suits well for the channels with array errors. Array-error-correcting codes, i.e., codes having a distance d in the array metric, were proposed in [1], [2], [3], and in other publications. These codes have algebraic decoders, which are able to correct up to $(d-1)/2$ erroneous lines in the received matrix. More precisely, these decoders correct ε erroneous lines and θ erased lines as soon as

$$\lambda\varepsilon + \theta \leq d-1, \tag{1}$$

where $\lambda = 2$ is the tradeoff rate between errors and erasures for these decoders.

Assume that the decoder has side information about reliabilities of lines in the received matrix. Can we correct more than $(d-1)/\lambda$ erroneous lines in this case?

For the case of correction of independent errors (using codes in Hamming metric) the answer "yes" was done by Forney [4]. He introduced generalized Hamming distance, which is the weighted Hamming distance, where weights are the reliabilities of the received symbols. Forney also suggested a decoding algorithm, which uses an algebraic decoder (with $\lambda = 2$) in multi-trial manner to decode the received vector in generalized metric. Later, Kovalev [5] suggested an adaptive form of the Forney algorithm to decrease twice the number of decoding trials. The Forney-Kovalev decoding algorithm was refined by Weber and Abdel-Ghaffar in [6] and extended for $\lambda \leq 2$ in [7].

In this paper we introduce generalized array distance, which is array-distance weighted by reliabilities. We show that this generalized array distance suits well to the channel with array errors and with side reliabilities information. Then we show, that decoding of codes in the new generalized metric can be done by a modification [7] of the Forney-Kovalev algorithm. This allows us to estimate the error correcting radius of the decoding algorithm for all λs.

II. Array-error model and array metric

A. Channel

We consider transmission of $m \times n$ matrix C over $\mathbb{F}_q$. Let us enumerate lines (rows and columns) of C by numbers $1, \dots, m+n$. The received matrix Y is $Y = C + E$, where error-matrix E is constructed by the channel as follows. The channel selects s different lines of E with probability $P(s)$ and fills these lines randomly by elements of $\mathbb{F}_q$, independently and equiprobable. All the rest components of E are zeros. We assume that $P(s)$ decreases with s.

B. Array metric

The array (or maximal term rank) metric is defined as follows. Assume that all nonzero elements of the matrix $A \in \mathbb{F}_q^{m\times n}$ are contained in t lines with indexes $\{i_1, \dots, i_t\}$, then we call this set *a covering of* A and denote it by $\mathcal{I}(A) = \{i_1, \dots, i_t\}$. The *array-weight* (or array-norm) $w^{(\mathrm{a})}(A)$ of a matrix A is defined as follows

$$w^{(\mathrm{a})}(A) \triangleq \min_{\mathcal{I}(A)} |\mathcal{I}(A)|. \tag{2}$$

In other words, the array-weight of a matrix A is the minimum number of lines that contain all nonzero elements of A. The maximum possible array weight of a matrix A is $\min\{m, n\}$.

The *array-distance* $d^{(\mathrm{a})}(A, B)$ between matrices A and B is defined as

$$d^{(\mathrm{a})}(A, B) \triangleq w^{(\mathrm{a})}(A - B). \tag{3}$$

The array-norm (2) satisfies the axioms of a norm, and hence the array distance (3) satisfies the axioms of a distance.

This work of V.R. Sidorenko has been supported by DFG (German Research Council) under grant BO 867/21. V.R. Sidorenko is on leave from IITP, Russian Academy of Sciences, Moscow, Russia.

The work of E.M. Gabidulin has been supported by DFG under grant 436 RUS 113/941/0-1

III. Algebraic codes correcting array-errors

A linear (nm, k, d) code $\mathcal{C}$ of rate $R = \frac{k}{mn}$ is a linear subspace of $\mathbb{F}_q^{m\times n}$ of dimension k, where the array code distance $d^{(\mathrm{a})}(\mathcal{C}) = d$ is the minimum array distance between two different codewords of $\mathcal{C}$.

From now on let us assume without loss of generality that

$$m \geq n. \tag{4}$$

The code-dimension k satisfies the following Singleton-type bound [1]

$$k \leq m(n - d + 1). \tag{5}$$

In [1] the following construction of array-error-correcting codes was proposed. Assume we have an $(n, k, d^{(H)})$ block linear code $\mathcal{C}^{(H)}$ over $\mathbb{F}_q$ with distance $d^{(H)}$ in *Hamming* metric. A code matrix $C = ||c_{i,j}||, i = 1, \ldots, m, j = 1, \ldots, n$ of an array-error-correcting code $\mathcal{C}$ we design as follows. We say that the set $\{c_{(i+j) \bmod m+1, j+1} : \quad j = 0, \ldots, n-1\}$ forms the $(i+1)$st diagonal of the matrix C, $i = 0, \ldots, m-1$. By writing m arbitrary words of the code $\mathcal{C}^{(H)}$ into m diagonals of the matrix C we obtain a codeword of the code $\mathcal{C}$. Notice, that every corrupted line (erased line or line with errors) in C affects at most one symbol of every diagonal of C. As a result we obtain an $(nm, km, d^{(H)})$ code $\mathcal{C}$ with array distance $d^{(H)}$.

Assume we have a decoder of the code $\mathcal{C}^{(H)}$ correcting up to t errors in Hamming metric. Then, by correcting errors in every diagonal of a received matrix Y, we will correct every error matrix E of array-weight up to t. Standard algebraic decoders allow to correct up to $(d^{(H)} - 1)/2$ errors. If the order q of the field is large enough, $q > (n+1)^l$ then we can use l-punctured, $l = 1, 2, ...$, Reed–Solomon codes [8], which allows to correct up to $\frac{l}{l+1} d^{(H)}$ errors [9]. More precisely the decoder corrects ε errors and θ erasures if (1) holds and fails otherwise. Here, the real number $\lambda = 1 + 1/l$, $1 < \lambda \leq 2$ is the tradeoff rate between errors and erasures for this decoder. This is an example of array-error-correcting (mn, k, d) code $\mathcal{C}$ with array-distance d, which reaches the Singleton-type upper bound (5). If $q > (n+1)^l$ then there is a decoder Φ for this code, which corrects ε errors and θ erasures as soon as (1) is satisfied with $\lambda = 1 + 1/l$.

Another class of array-error-correcting codes is based on codes in rank metric. Rank distance between $m \times n$ matrices A and B is defined as $d^{(\mathrm{r})}(A, B) = \mathrm{rank}\,(A - B)$. Since $\mathrm{rank}\,(A - B) \leq d^{(\mathrm{a})}(A, B)$, every code having distance d in the rank metric has distance at least d in the array metric. There is a class of (mn, k, d) Gabidulin codes [2], [3], which have distance d in rank metric satisfying the Singleton-type upper bound (5) with equality. Hence, every (mn, k, d) Gabidulin code is simultaneously (mn, k, d) code with array-distance d. There are known algebraic decoders of Gabidulin codes, which correct up to $(d-1)/2$ errors in rank metric, and hence in the array metric as well. This is another example of array-error-correcting codes, having the decoder Φ, which corrects ε errors and θ erasures as soon as (1) is satisfied with $\lambda = 2$.

IV. Generalized distance and GMD decoding

A. Generalized weight and distance

Given a vector $h = (h_1, \ldots, h_{n+m})$ of line-reliabilities, where $0 \leq h_i \leq 1$, we define generalized distance as follows. First we define h-weight of a matrix $A \in \mathbb{F}_q^{m\times n}$ as

$$|A|_h = \min_{\mathcal{I}(A)} \sum_{i \in \mathcal{I}(A)} h_i. \tag{6}$$

Theorem 1 *The defined h-weight satisfies the axioms of a seminorm, i.e., for every $A, B \in \mathbb{F}_q^{m\times n}$ holds*

1) $|A|_h \geq 0$,
2) $|A|_h = |-A|_h$,
3) $|A - B|_h \leq |A|_h + |B|_h$.

Proof: The first two properties follow immediately from definition (6). Let us prove the third one. Indeed, $\mathcal{I}(A) \cup \mathcal{I}(B)$ covers $A - B$. Hence

$$|A - B|_h = \min_{\mathcal{I}(A-B)} \sum_{i \in \mathcal{I}(A-B)} h_i \leq \min_{\mathcal{I}(A), \mathcal{I}(B)} \sum_{i \in \mathcal{I}(A) \cup \mathcal{I}(B)} h_i$$

and since $h_i \geq 0$

$$\leq \min_{\mathcal{I}(A)} \sum_{i \in \mathcal{I}(A)} h_i + \min_{\mathcal{I}(B)} \sum_{i \in \mathcal{I}(B)} h_i = |A|_h + |B|_h.$$

■

Notice that the h-weight does not satisfy the axiom of positive definiteness, i.e. the axiom $|A|_h = 0$ iff $A = 0$ does not hold. For example, if $h = 0$ then $|A|_h = 0$ for every matrix A.

Let us modify the h-wight (6) by multiplying it by 2 and adding a fixed (for a fixed h) positive number $m + n - \sum h_i$. The new h-norm remains to be a seminorm. As a result, we obtain the following new definition, which corresponds to a traditional definition of generalized distance.

Definition 1 *For a given vector h of reliabilities and for matrices $A, B \in \mathbb{F}_q^{m\times n}$ a seminorm (or h-norm) $|A|_h$ is defined as follows.*

$$|A|_h = \min_{\mathcal{I}(A)} \left(\sum_{i \in \mathcal{I}(A)} (1 + h_i) + \sum_{i \notin \mathcal{I}(A)} (1 - h_i) \right). \tag{7}$$

A generalized array semidistance (or h-distance) between matrices A and B is defined as

$$d_h(A, B) = |A - B|_h. \tag{8}$$

Notice, for $h = (1, \ldots, 1)$ the h-distance coincides with doubled array distance. For a given h the h-distance $d_h(\mathcal{C})$ of a code $\mathcal{C}$ is defined as the minimum h-distance between two different codewords. For a linear code, h-distance of the code is the minimum h-norm of a nonzero codeword.

Theorem 2 *If array distance of the code $\mathcal{C}$ is $d^{(a)}(\mathcal{C}) = d$ then the minimum h–distance of $\mathcal{C}$ over all h is*

$$\min_h d_h(\mathcal{C}) = d.$$

Proof:

$$\min_h d_h(\mathcal{C}) = \min_h \min_{C:w^{(\mathrm{a})}(C)\geq d} |C|_h = \min_{C:w^{(\mathrm{a})}(C)\geq d} \min_h |C|_h$$

$$= \min_{C:w^{(\mathrm{a})}(C)\geq d} w^{(\mathrm{a})}(C) = d.$$

■

Theorem 2 explains why we modified the definition of the h-norm.

B. Generalized minimum distance decoder

Given a received matrix Y and reliability vector h, the *goal* of the Generalized Minimum Distance (GMD) decoder is to find the list $\mathcal{L}$ of codewords C which are at the minimum h-distance $d_h(Y, C)$ from the received vector Y, i.e., to decode the code $\mathcal{C}$ in the generalized metric.

The guaranteed error correcting radius ρ of a particular GMD decoder is the infimum of real numbers $\tilde{\rho}$, for which there exist two matrices $C \in \mathcal{C}$, $Y \in F_q^{m\times n}$ and a vector $h \in [0,1]^n$, such that $d_h(Y, C) = \tilde{\rho}$, and the GMD decoder fails to decode Y, h, i.e., it outputs a list, which does not contain C. In other words, we guarantee correction of every error of generalized weight less than ρ, where the generalized error-weight is defined to be $|Y - C|_h = d_h(Y, C)$. It follows from Theorem 2 that the error-correcting radius ρ of GMD decoder can not be greater than d.

C. Generalized distance matches the array-error channel

Let p_i, $i = 1, \ldots, m+n$, be the a posteriori probability that the ith line was selected by the channel to be erroneous. Then joint probability that the ith line of length l_i, $l_i \in \{m, n\}$, was selected by the channel and filled by particular l_i symbols from $\mathbb{F}_q$ is $\tilde{p}_i = p_i q^{-l_i}$. Given the received matrix Y and the vector of probabilities $p = (p_1, \ldots, p_{n+m})$, for every codematrix C probability of the error matrix $E = Y - C$ can be estimated neglecting the fact of line-intersection as follows

$$\begin{aligned} P(E) &\approx \sum_{\mathcal{I}(E)} \prod_{i\in\mathcal{I}(E)} \tilde{p}_i \prod_{i\notin\mathcal{I}(E)} (1-p_i) \\ &\approx \max_{\mathcal{I}(E)} \prod_{i\in\mathcal{I}(E)} \frac{\tilde{p}_i}{1-p_i} \prod_{i=1}^{m+n} (1-p_i). \end{aligned} \tag{9}$$

Denote the second product in (9) by $a(p)$ and

$$h_i = -\ln \frac{\tilde{p}_i}{1-p_i}. \tag{10}$$

Using definition (6) we obtain

$$\begin{aligned} P(E) &\approx a(p) \max_{\mathcal{I}(E)} \exp\left(-\sum_{i\in\mathcal{I}(E)} h_i\right) \\ &= a(p) \exp\left(-\min_{\mathcal{I}(E)} \sum_{i\in\mathcal{I}(E)} h_i\right) \\ &= a(p) \exp\left(-|E|_h\right). \end{aligned} \tag{11}$$

The maximum likelihood decoding rule becomes

$$\arg\max_{C\in\mathcal{C}} P(Y|C) = \arg\max_{C\in\mathcal{C}} P(Y-C) = \arg\min_{C\in\mathcal{C}} |Y-C|_h. \tag{12}$$

Let us make a realistic assumption that $p_i \leq (1+q^{-l_i})^{-1} \triangleq p_i^{(\max)} \approx 1$. If the assumption is not satisfied then we can replace $p_i > p_i^{(\max)}$ by $p_i^{(\max)}$. Then from (10) it follows that $h_i \geq 0$. Notice that the result of decoding rule (12) will not change if we mutiply every h_i by a positive number. Denote $h_{\max} = \max\{h_i\}$ and divide every h_i by $h_{\max}$ then we have $0 \leq h_i \leq 1$. As a result, up to approximation in (9), maximum likelihood decoder coincides with generalized minimum distance decoder according to definition (6) and hence according to Definition 1 as well, since the result of decoding rule (12) will not change if we replace definition (6) by Definition 1.

V. Forney-Kovalev (FK) decoding

To implement GMD decoding we use the FK algorithm. Given an array-error-and-erasure decoder Φ of the code $\mathcal{C}$, the *FK list decoding* is as follows. For $j = 1, \ldots, s$ we make a trial to decode the received matrix Y in which the τ_j least reliable lines are erased. Performing s decoding trials using decoder Φ we obtain a list $\mathcal{L}$ of codewords. If this list is empty, we declare a decoding failure, otherwise we leave in the list only codewords C having the minimum $d_h(C, Y)$ and output the new list. FK decoders may differ by using different decoders Φ (having different λ) or by different number s of decoding trials or by different selection of the erasure vector $\tau = (\tau_1, \ldots, \tau_s)$. If the erasure vector is fixed we get the Forney algorithm. If the erasure vector is selected adaptive depending on the received vector h of reliabilities, we obtain the Kovalev algorithm, having better performance. Later we consider the adaptive approach only.

Let us estimate the guaranteed error correcting radius ρ of the adaptive FK algorithm. Recall that we consider a FK decoder based on an array-error-correcting algebraic decoder Φ which satisfies (1) with tradeoff rate λ. At the input of the FK decoder we have a received word Y and a vector of reliabilities h. From now on, assume w.l.o.g. that the lines of matrices Y and C are ordered according to their reliabilities as follows

$$0 \leq h_1 \leq h_2 \leq \cdots \leq h_{m+n} \leq 1. \tag{13}$$

So, we denote by $h = (h_1, ..., h_{m+n})$ the vector of *ordered* reliabilities, and by $\mathcal{H}$ the set of all possible real-valued vectors h satisfying (13).

Definition 2 *Given the vector h of reliabilities, by $\delta_\tau(h)$ we denote the minimum h-weight of the error in the channel that causes a failure of the FK decoder with erasing strategy defined by the vector τ. In other words, $\delta_\tau(h)$ is error-correcting radius for fixed h and τ.*

Lemma 3 *Error-correcting radius $\delta_\tau(h)$ is as follows*

$$\delta_\tau(h) = \sum_{j=1}^{m+n}(1-h_j) + 2\sum_{i=1}^{s}\sum_{j=\tau_i+1}^{\tau_i+\varepsilon(\tau_i)-\varepsilon(\tau_{i+1})} h_j, \qquad (14)$$

where we denote the function

$$\varepsilon(\theta) = \left\lfloor \frac{d-\theta-1}{\lambda} \right\rfloor + 1,$$

and τ_{s+1} is formally defined such that $\varepsilon(\tau_{s+1}) = 0$.

Let $\mathcal{T}$ be the set of all integer valued vectors $\tau = (\tau_1, ..., \tau_s)$ such that $0 \leq \tau_1 \leq \cdots \leq \tau_s \leq d-1$. To specify a particular FK decoder we are free to select a vector τ. For a given h we will select τ to maximize the error-correcting radius $\delta_\tau(h)$:

$$\tau(h) = \arg\max_{\tau\in\mathcal{T}} \delta_\tau(h). \qquad (15)$$

The algorithm with this $\tau(h)$ we will call *adaptive algorithm* and denote by A. The error correcting radius $\rho_A(\lambda)$ of algorithm A is

$$\rho_A(\lambda) = \inf_{h\in\mathcal{H}} \max_{\tau\in\mathcal{T}} \delta_\tau(h). \qquad (16)$$

To find vector $\tau(h)$ from (15) one should consider $|\mathcal{T}|$ vectors τ, thus the complexity of this step is $\mathcal{O}(d^s)$. Remark, that the decoder should compute $\tau(h)$ for every received h, thus the computation is only feasible for one or two decoding trials, i.e., for $s = 1, 2$. This is a big disadvantage of the adaptive approach using the erasing vector (15).

VI. Decoding algorithm

Fortunately Kovalev suggested a simplification of the adaptive decoding algorithm where vector of erasures $\tau(h)$ should be selected from a set of two vectors only! In [7] this simplified algorithm was extended for all the range of λ and the final decoder is given by Algorithm 1. To compute $\tau(h)$ Algorithm 1 requires $\mathcal{O}(d)$ operations only. Error-correcting radius $\rho_A(\lambda)$ of the initial algorithm A based on $\tau(h)$ given by (15) and radius of the simplified Algorithm 1 coincide!

Theorem 4 ([7]) *The error correcting radius of Algorithm 1 is lower bounded by $\underline{\rho}_A(\lambda)$*

$$\rho_A(\lambda) \geq \underline{\rho}_A(\lambda) = \varepsilon(0) + \varepsilon(\tau_1), \qquad (17)$$

where τ_1 is a solution of recurrent inequalities

$$\tau_i \geq \tau_{i-1} + \varepsilon(\tau_{i-1}) - \varepsilon(\tau_{i+1}), \quad i = 1, \ldots, 2s-1, \qquad (18)$$

with boundary conditions

$$\tau_0 = 0, \quad \tau_{2s} = \lfloor d-1+\lambda \rfloor. \qquad (19)$$

The lower bound (17) is nearly tight [7] and can be approximated as follows.

Corollary 5 *For $1 < \lambda < 2$ s-trial decoding radius is*

$$\underline{\rho}_A(\lambda) \approx d\left(1 - \frac{(2-\lambda)(\lambda-1)^{2s}}{\lambda(1-(\lambda-1)^{2s})}\right) \approx d\left(1-(\lambda-1)^{2s}\right). \qquad (20)$$

To reach $\rho_A(2) = d$ it is sufficient to have $s = \frac{1}{2}\left(\log_{\frac{1}{\lambda-1}} d + 1\right)$ decoding trials.

```
Algorithm 1: Simplified $s$-trial adaptive decoding
Precomputations: Solve (18), get vectors
$\tau^{(a)} = (\tau_0, \tau_2, ..., \tau_{2(s-1)})$ and $\tau^{(b)} = (\tau_1, \tau_3, ..., \tau_{2s-1})$;
Input: received matrix $Y$ and (ordered) vector $h$;
Select vector $\tau' = \arg\max_{\tau\in\{\tau_a,\tau_b\}} \delta_\tau(h)$;
for each $j$ from $1$ to $s$ do
    decode $Y$ with erased first $\tau'_j$ positions by the decoder $\Phi$ of the code $\mathcal{C}$, add obtained codeword (if any) to the list $\mathcal{L}$;
Output:
if the list $\mathcal{L}$ is empty then
    declare a decoding failure;
else
    leave in $\mathcal{L}$ only codewords nearest to $Y$ in $h$-metric, output $\mathcal{L}$
```

Corollary 6 *For $\lambda = 2$ s-trial decoding radius is*

$$\rho_A(2) \geq d + 1 - \left\lceil \frac{d+1}{4s} \right\rceil, \qquad (21)$$

which coincides with Kovalev's result. To reach $\rho_A(2) = d$ it is sufficient to have $s = \lceil \frac{d+1}{4} \rceil$ decoding trials.

Notice, to reach $\rho_A(2) = d$, the number s of decoding trials grows linearly with d for the classical case $\lambda = 2$ and only logarithmically for $\lambda < 2$. As a result, for $\lambda < 2$ the error-correcting radius of Algorithm 1 quickly approaches d with increasing number of trials, and 2 or 3 trials are sufficient to reach $\rho_A(2) = d$ in many practical cases.

References

[1] E.M. Gabidulin, B.I. Korjik, "Lattice-error-correcting codes," Izv. Vyssh. Uchebn. Zaved., Radioelektron.,15, no. 4,492-498, 1972.

[2] E.M. Gabidulin, "Theory of codes with maximum rank distance,", Probl. Inform. Transm. 21(1), pp. 3-16, 1985.

[3] R.M. Roth, "Maximum-rank array codes and their application to criss-cross error correction," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 328-336, Mar. 1991.

[4] G. D. Forney Jr., "Generalized minimum distance decoding," *IEEE Trans. Inf. Theory*, vol. 12, pp. 125-131, Apr. 1966.

[5] S. I. Kovalev, "Two classes of minimum generalized distance decoding algorithms," *Probl. Pered. Inform.*, vol. 22, no. 3, pp. 35-42, 1986.

[6] J. H. Weber, K. A. S. Abdel-Ghaffar, "Reduced GMD decoding," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1013-1027, April 2003.

[7] V. Sidorenko, A. Chaaban, Ch. Senger, M. Bossert, On Extended Forney-Kovalev GMD decoding, IEEE International Symposium on Information Theory, June-July, 2009, Seoul, Korea.

[8] V. R. Sidorenko, G. Schmidt, M. Bossert, "Decoding punctured Reed-Solomon codes up to the Singleton bound," in *Proc. of Int. ITG Conference on Source and Channel Coding*, Ulm, January 2008.

[9] G. Schmidt, V. R. Sidorenko, and M. Bossert, "Collaborative decoding of interleaved Reed–Solomon codes and concatenated code designs,"*IEEE Trans. Inf. Theory*, vol. 55, n. 7, pp. 2991-3012, July 2009.

Optimizing BICM with Convolutional Codes for Transmission over the AWGN Channel

Clemens Stierstorfer, Robert F.H. Fischer, and Johannes B. Huber
Lehrstuhl für Informationsübertragung, Friedrich–Alexander–Universität Erlangen–Nürnberg
Cauerstraße 7/LIT, 91058 Erlangen, Germany, Email: {clemens,fischer,jbhuber}@LNT.de

Abstract—**The usual comparison of trellis coded modulation and bit-interleaved coded modulation, both using convolutional codes and Viterbi decoding, leads to the well-known result that for the AWGN channel TCM clearly outperforms BICM. In fading scenarios BICM shows superior results. Based on recent results on optimized bit mappings and bit-interleaver designs for BICM, we demonstrate that the BER of BICM on AWGN channels can be significantly lowered at no additional cost. Depending on the signal constellation size and the constraint length of the convolutional code gains up to 7 dB can be achieved over BICM with random bit interleavers.**

I. Introduction

The comparison of trellis coded modulation (TCM) [8] and bit-interleaved coded modulation [11] as for example performed in [2], leads to the well-known result that on the additive white Gaussian noise (AWGN) channel TCM outperforms BICM in terms of capacity and bit error ratio (BER). In fading scenarios, on the contrary, BICM shows superior results. The provided numerical results are mostly based on non-iteratively decoded implementations using convolutional codes and Viterbi decoders. This classical variant of BICM recently was identified as optimum solution for low-delay applications [4]. Here, we show that in part the performance of conventional BICM on AWGN channels has been considerably underestimated so far.

For the analysis of BICM an equivalent channel model initially introduced in the context of multilevel codes [5], [9] has proven to be helpful. Based on this model we recently studied bit mappings for BICM transmission [6] and optimized the design of the bit interleaver [7]. In particular the latter offers a large potential for optimizations at no additional costs.

In this contribution, our recent insights are combined with the knowledge that in non-fading scenarios bit interleaving is not necessarily beneficial, but may be even counter-productive [9]. We propose some slight modifications for BICM transmission over AWGN channels which at higher spectral efficiency significantly lower the resulting BER.

II. General System Model

We investigate block-based coded transmission over an AWGN channel (see Fig. 1). A rate-$R_\mathrm{c} = k/n$ convolutional encoder (ENC) is used to encode a sequence $\boldsymbol{q}$ of K_bs binary source bits q_κ, $\kappa = 1, \ldots, K_\mathrm{bs}$, originating from a discrete i.i.d. memoryless source, into a binary sequence $\boldsymbol{c}$ of $N_\mathrm{bs} = K_\mathrm{bs}/R_\mathrm{c}$ encoded symbols c_ν, $\nu = 1, \ldots, N_\mathrm{bs}$. The symbol rate of the source bits is denoted by $1/T_\mathrm{b}$, that of the encoded bits by $1/T_\mathrm{c}$. The sequence of encoded bits is passed on to a block bit interleaver Π which permutes the encoded binary symbols and generates an output stream of R_M-tuples $\boldsymbol{x} = [x_\delta^{(1)}, x_\delta^{(2)}, \ldots, x_\delta^{(R_M)}]$ with index $\delta = 1, \ldots, \Delta$; $\Delta = N_\mathrm{bs}/R_M$. These binary R_M-tuples $\boldsymbol{x}$ are then bijectively mapped onto channel symbols a

$$\mathcal{M} : \boldsymbol{x}_\delta \in \mathbb{F}_2^{R_M} \mapsto a_\delta \in \mathcal{A} \subset \mathbb{C} \,. \tag{1}$$

Here, $\mathcal{A}$ denotes an M-ary signal constellation and $M = 2^{R_M} = |\mathcal{A}|$ holds. For simplicity, we restrict our considerations to M-ary amplitude-shift keying (ASK) constellations ($\mathcal{A} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$), i.e., we focus on one of the quadrature components of an M^2-QAM constellation. If $\mathcal{M}$ is a Gray mapping the two quadrature components of a QAM constellation are independent an can be processed successively.

The received and AWGN-corrupted signal reads

$$y_\delta = a_\delta + n_\delta \,. \tag{2}$$

The variance of the channel symbols[1] is given as $\sigma_\mathrm{A}^2 = E_\mathrm{s}/T_\mathrm{s}$ with the average energy per symbol E_s and the channel symbol rate $1/T_\mathrm{s}$. The variance of the noise samples per quadrature component reads $\sigma_\mathrm{N}^2 = N_0/(2T_\mathrm{s})$ with N_0 denoting the one-sided noise power spectral density. We define a signal-to-noise ratio (SNR) as $E_\mathrm{s}/N_0 = \sigma_\mathrm{A}^2/(2\sigma_\mathrm{N}^2)$.

At the receiver R_M-tuples $\boldsymbol{\Lambda}_\delta$ of pairs of bit metrics $\boldsymbol{\lambda}_\delta^{(\mu)}$, $\mu = 1, \ldots, R_M$, are determined by the bit-metric calculator $\mathcal{L}$ from the y_δ. After deinterleaving the sequence of N_bs pairs of deinterleaved bit metrics $\boldsymbol{\lambda}_\nu$ is processed by the Viterbi decoder (DEC) which finally returns a sequence $\hat{\boldsymbol{q}}$ of K_bs estimates $\hat{q}_\kappa$ on the initial source symbols q_κ.

A. Equivalent Channel Model

Due to the bijection between the R_M-tuples $\boldsymbol{x}$ and the channel symbols[2] a, the combination of $\mathcal{M}$ and AWGN channel can be equivalently represented by a set of R_M parallel subchannels (aka. bit levels) with binary inputs and continuous output, cf. [9] and Fig. 1. The binary labels $\boldsymbol{x}$ are the discrete channel input; the received signal y is the continuous output. The μ-th label bit $x^{(\mu)}$ is transmitted via the μ-th bit level.

B. Bit-Level Capacity and Parallel-Decoding Capacity

The bit levels are characterized by the respective bit-level capacity ${}^\mathsf{bl}C$. For the μ-th bit level ${}^\mathsf{bl}C^{(\mu)}$ is defined as the mutual information between y and $x^{(\mu)}$ [9], i.e.,

$${}^\mathsf{bl}C^{(\mu)} \triangleq I(Y; X^{(\mu)}) \,, \tag{3}$$

which for AWGN and fading channels can only be evaluated numerically. Exemplary results for 16-ASK ($R_M = 4$) and 64-ASK ($R_M = 6$) are depicted in Fig. 3. Obviously, there is a

[1] Upper case letters denote the respective random variables to scalars. Vectors and matrices are set in lower resp. upper case boldface letters.

[2] The discrete time index δ is dropped for convenience.

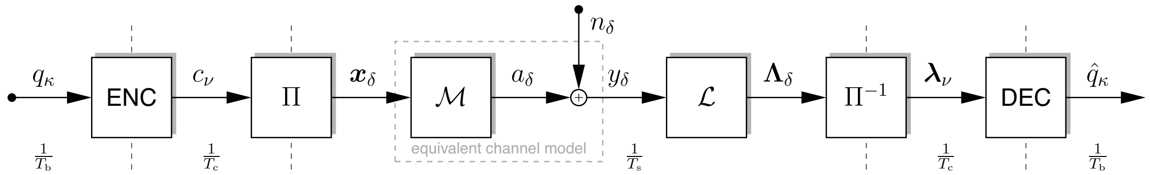
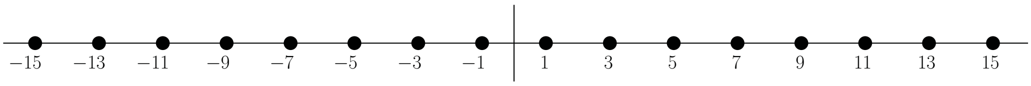


Fig. 1. System model of BICM transmission over AWGN channel.



$$\boldsymbol{X} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Fig. 2. Illustration of exemplary bit mapping for 16-ASK constellation. $\boldsymbol{X}$ realizes a binary-reflected Gray mapping.

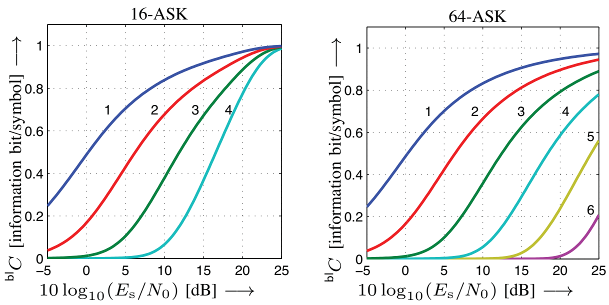


Fig. 3. Exemplary bit-level capacities over $10\log_{10}(E_\mathrm{s}/N_0)$ for AWGN channel. Left: $R_M = 4$ curves for 16-ASK. Right: $R_M = 6$ curves for 64-ASK. Respective level indices μ for binary-reflected Gray mapping given in plots.

strong relation between the ${}^{\mathsf{bl}}C$'s and μ resp. M. Furthermore, there is also a significant dependency of the average reliability of the bit metrics $\boldsymbol{\lambda}^{(\mu)}$ on the related ${}^{\mathsf{bl}}C^{(\mu)}$, cf. [7].

The sum over the R_M individual bit-level capacities yields the parallel-decoding capacity (PDC) [9]

$$ {}^{\mathsf{pd}}C = \sum_{\mu=1}^{R_M} {}^{\mathsf{bl}}C^{(\mu)} \,. \tag{4}$$

In contrast to successive or joint decoding, where knowledge of preceeding or even all bit levels is exploited at the decoder, parallel decoding neglects any knowledge originating from other subchannels. The decoding of BICM realizes a parallel-decoding approach, cf. [2], [9], and thus the PDC usually is referred to as 'BICM capacity' in the literature, e.g., [3]. Due to the omitted knowledge, the PDC is lower than the constellation-constraint capacity. The gap between these two variables depends on the bit mapping $\mathcal{M}$. In [6] binary-reflected Gray mappings (BRGM) were shown to minimize the loss in terms of the PDC for medium to high SNRs and thus are used in the remainder. In [1] BRGM were proven to be optimal for uncoded transmission.

C. Bit Mapping

The binary labeling rules $\mathcal{M}$ can be described by using an $(R_M \times M)$-matrix $\boldsymbol{X}$ and a function $\mathrm{ColNr}_{\boldsymbol{X}}\{\boldsymbol{x}\}$. The columns of $\boldsymbol{X}$ contain the M potential binary R_M-tuples, i.e., $\boldsymbol{X} = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_M]$. $\mathrm{ColNr}_{\boldsymbol{X}}\{\boldsymbol{x}\}$ returns the column index (from 1 to M) of $\boldsymbol{x}$ in $\boldsymbol{X}$. We can then write the mapping as

$$\mathcal{M} : \boldsymbol{x} \mapsto a = 2 \cdot \underset{\boldsymbol{X}}{\mathrm{ColNr}}\{\boldsymbol{x}\} - M - 1 \,. \tag{5}$$

Fig. 2 shows an exemplary matrix for 16-ASK and a BRGM.

Obviously, the ${}^{\mathsf{bl}}C^{(\mu)}$ is entirely determined by the structure of the μ-th row of $\boldsymbol{X}$ denoted by $\chi^{(\mu)}$. Respective ${}^{\mathsf{bl}}C$'s for the exemplary $\mathcal{M}$ given in Fig. 2 are depicted in Fig. 3. The first row of $\boldsymbol{X}$ leads to the left-most curve and so forth.

The matrix $\boldsymbol{X}$ given in Fig. 2 describes only one of many potential $\mathcal{M}$'s applicable to 16-ASK. Some of these mappings perform entirely identical in terms of, e.g., the BER of uncoded transmission or the PDC. In [1] *trivial operations* on $\boldsymbol{X}$ were defined which do not affect the BER of uncoded tranmission or the PDC. These operations comprise complementations of rows of $\boldsymbol{X}$, interchanging rows in $\boldsymbol{X}$, and reflection of rows wrt. to the symmetry axes of the signal constellation. Note, in the BICM scheme the interchanging of rows of $\boldsymbol{X}$ could be performed implicitly by a respectively designed bit interleaver Π.

III. The Viterbi Decoder and Bit Interleaving

The optimization of BICM for AWGN channels is based on an analysis of the Viterbi decoder and its interaction with the bit interleaver Π. Formally, the latter implements a mapping described as

$$\Pi : \boldsymbol{c} \mapsto [\boldsymbol{x}_1 \ldots \boldsymbol{x}_\Delta] \,. \tag{6}$$

A sequence of N_bs encoded binary symbols c is mapped onto a sequence of Δ binary R_M-tuples $\boldsymbol{x}$.

The inverse operation Π^{-1} at the receiver reads

$$\Pi^{-1} : [\boldsymbol{\Lambda}_1 \ldots \boldsymbol{\Lambda}_\Delta] \mapsto [\boldsymbol{\lambda}_1 \ldots \boldsymbol{\lambda}_{N_\mathrm{bs}}] \,, \tag{7}$$

i.e., a sequence of Δ R_M-tuples $\boldsymbol{\Lambda}_\delta$ of pairs of bit metrics is converted into a sequence of N_bs pairs of bit metrics $\boldsymbol{\lambda}_\nu$.

The problem finally to be solved by the decoder (DEC) is

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c} \in \mathcal{C}}{\mathrm{argmin}} \left\{ \sum_{\nu=1}^{N_\mathrm{bs}} \lambda_{\nu, c_\nu} \right\} \,. \tag{8}$$

The estimated sequence $\hat{\boldsymbol{c}}$ minimizes the overall path metric. Here, $\lambda_{\nu,b}$ denotes the bit metric (Euclidean distance) for the ν-th encoded bit to be $b \in \{0,1\}$ and $\boldsymbol{\lambda}_\nu = [\lambda_{\nu,0}, \lambda_{\nu,1}]^\mathsf{T}$.

The importance of bit interleaving is determined by the 'sliding window' characteristic of the Viterbi decoder. Decoding results strongly rely on localized arrangements of bit metrics within the sequence. In turn, aggregations of unreliable bit metrics most likely lead to errors; the processing window covers only a very small fraction of the trellis. Bit interleavers are mostly designed to compensate for the effects of fading scenarios and to (randomly) spread initially neighbored bit metrics—usually affected by similar fading states and thus equally reliable—over the entire codeword.

In [7] the individual bit levels were identified as an inherent source of 'fading'. The varying bit-level capacities induce varying average bit-metric reliabilities. In contrast to the 'real' fading process, the 'bit-level' fading is known prior to transmission. An approach taking advantage of this knowledge for BICM is, e.g., *adaptive bit interleaving* [7] which leads to significant gains compared to conventional designs.

IV. Optimization of BICM for AWGN Channels

A. Bit Interleaver

For the optimization of BICM for transmission over AWGN channels the usually block-based bit interleaver Π is reduced to simply provide binary R_M-tuples to the mapper $\mathcal{M}$; no 'real' bit interleaving is performed. This approach was already vaguely discussed in some early publications on coded modulation, e.g., [9]. The optimization is entirely shifted to the bit mapping $\mathcal{M}$.

B. Bit Mapping

The optimization of the bit mapping is motivated by the bit interleaver designs presented in [7]. For simplicity, we introduce the idea for rate-$1/2$ convolutional codes; an extension to other code rates is straightforward.

Consider the decoding of a rate-$1/2$ convolutional code as illustrated by the trellis diagram in Fig. 4. In each trellis segment two bit metrics $\lambda_{\nu,b}$ are combined into a segmental path metric. If $R_M = 2$ (4-ASK) is assumed, bit metrics originating from the only two existing bit levels are combined in each segment. The virtual sliding processing window of the decoder comprises several trellis segments and we can easily see that a shift of the window does not affect the average reliability of the bit metrics within its span. Regarding in contrast $R_M = 4$, i.e., 16-ASK transmission and the mapping $\mathcal{M}$ specified by $\boldsymbol{X}$ given in Fig. 2, distinct variations of the average reliability of the segmental path metrics, can be observed, cf. Fig. 4. Segments combining bit metrics of the stronger bit levels (1st/2nd) are succeeded by segments where the bit metrics of the weaker bit levels (3rd/4th) are combined. Shifting the window may affect the average bit-metrics reliability within its span. For larger signal constellations, e.g., 64-ASK, this effect is even more pronounced.

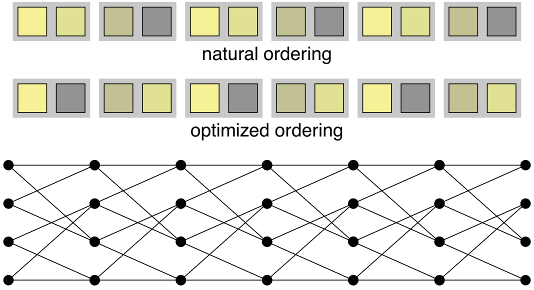


Fig. 4. Illustration of ordering of bit metrics in trellis for rate-$1/2$ code ($\nu_\mathrm{c} = 2$) and 16-ASK transmission ($R_M = 4$). Squares represent bit metrics $\lambda_{\nu,b}$. Brightness grows with average bit metric reliability.

To avoid such disadvantageous arrangements of the bit metrics, we slightly modify $\mathcal{M}$. Consider again 16-ASK transmission using the matrix $\boldsymbol{X}$ given in Fig. 2. By interchanging $\boldsymbol{\chi}_2$ and $\boldsymbol{\chi}_4$ of $\boldsymbol{X}$ we obtain a new matrix $\tilde{\boldsymbol{X}}$, cf. (9), and the average reliability of the segmental path metrics in the trellis is equalized, cf. Fig. 4. In terms of the BER of uncoded transmission and the PDC this modification is a trivial operation, i.e., irrelevant. With regard to the BER of coded transmission, however, $\tilde{\boldsymbol{X}}$ represents a new mapping. For larger signal constellation sizes M the procedure follows the same line: weaker levels are combined with stronger ones.

V. Numerical Results

A. Simulation Settings

Numerical results for the BER of coded transmission are provided for three scenarios: $R_M = 2$, 4, and 6. We used non-recursive, non-systematic convolutional rate-$1/2$ encoders (best known wrt. free distance [10]) with $\nu_\mathrm{c} = 2$, 10, and 13. $\Delta = 1000$ channel symbols were transmitted per block and the initial binary data sequences were zero-padded to ensure terminated trellises. We assessed three different configurations:

(C1) random bit interleaving,
(C2) no interleaving, standard BRGM,
(C3) no interleaving, optimized BRGM acc. to Sec. IV.

The analytically derived BERs of uncoded ASK transmission with equal spectral efficiencies are given for comparison.

B. Results

On the left of Fig. 5 the results of coded 4-ASK transmission and the BER of uncoded 2-ASK transmission are shown. Here, (C2) and (C3) coincide; an optimization is not feasible.

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0 \\ 0&0&1&1&1&1&0&0&0&0&1&1&1&1&0&0 \\ 0&0&0&0&1&1&1&1&1&1&1&1&0&0&0&0 \end{bmatrix} \tag{9}$$

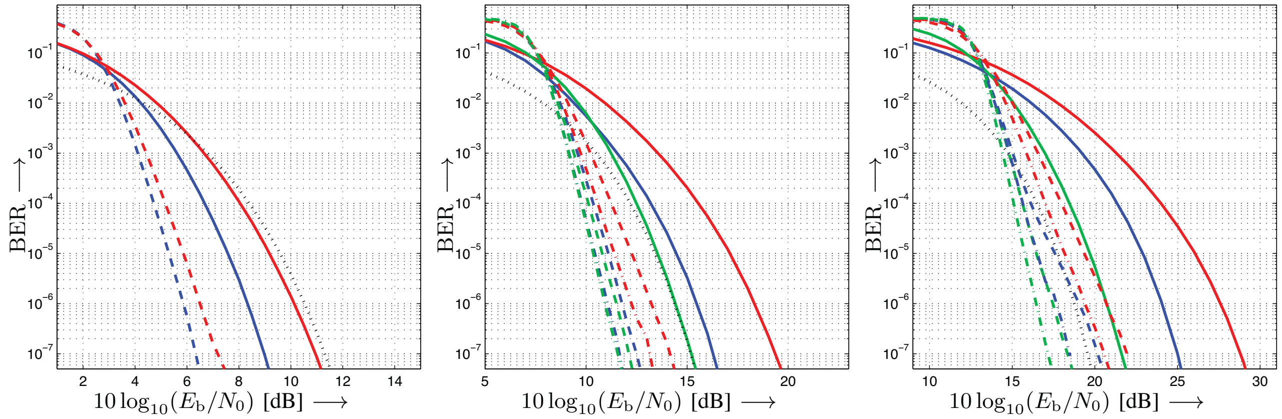


Fig. 5. Bit error ratio of rate-1/2 coded 4- (left), 16- (center), and 64-ASK (right) transmission over $10\log_{10}(E_\mathrm{b}/N_0)$ [dB]. $\nu_\mathrm{c} = 2$ (solid lines), $\nu_\mathrm{c} = 10$ (dashed lines), and $\nu_\mathrm{c} = 13$ (dash-dotted lines). Random bit interleaving (C1) (red), no bit interleaving (C2) (blue), optimized bit mapping (C3) (green). BER of uncoded 2- (left), 4- (center), and 8-ASK (right) transmission given for comparison (dotted).

Nevertheless, the curves illustrate the advantage of not interleaving on AWGN channels. Especially for $\nu_\mathrm{c} = 2$ a careful arrangement of the bit metrics is beneficial. At $\mathrm{BER} = 10^{-6}$ a gain of about 1.8 dB is achieved at no additional cost compared to the standard scheme. Actually, we can even reduce latency as the block interleaver is removed and such show to advantage a convenient feature convolutional codes exhibit over block-based coding like, e.g., LDPC codes [4].

In the center of Fig. 5 the BERs of coded 16-ASK transmission ($R_M = 4$) and uncoded 4-ASK are given. The plot reveals a significant advantage of (C2) over (C1). Simply dropping the bit interleaver leads to a gain of 3.2 dB at $\mathrm{BER} = 10^{-6}$ for $\nu_\mathrm{c} = 2$. Optimizing the bit mapping (C3)adds another decibel resulting in 4.2 dB at $\mathrm{BER} = 10^{-6}$. For larger ν_c the gains decrease a little, but are still remarkable. For $\nu_\mathrm{c} = 10$ we obtain 1.4 dB gain of (C2) over (C1) at $\mathrm{BER} = 10^{-6}$. The optimized $\mathcal{M}$ leads to an additional 0.3 dB. The improvements for $\nu_\mathrm{c} = 13$ are only slightly smaller.

The results for $R_M = 6$ (64-ASK) are depicted on the right of Fig. 5. Here, the gaps between (C1), (C2), and (C3) are tremendous. For $\nu_\mathrm{c} = 2$ the difference between random bit interleaving and no bit interleaving amounts to more than 3.5 dB at $\mathrm{BER} = 10^{-6}$. Optimizing the bit mapping yields almost another 3.5 dB adding up to a total of over 7 dB at $\mathrm{BER} = 10^{-6}$. The weak code with four encoder states achieves a performance similar to the code with $\nu_\mathrm{c} = 10$ and random bit interleaving. For $\nu_\mathrm{c} = 10$ the gap between no bit interleaving and the optimized bit mapping is 1.5 dB at $\mathrm{BER} = 10^{-6}$; we gain 3 dB over random bit interleaving. For $\nu_\mathrm{c} = 13$ the optimization yields comparable gains.

C. Discussion

The presented results emphasize the need for a sensible choice of bit mapping and bit interleaver for BICM transmission over AWGN channels. The gains due to the suggested modifications over conventional BICM with convolutional codes are tremendous. This holds in particular for shorter constraint lengths. Stronger codes better mitigate the ‘fading’ effect of the individual bit levels and the gains due to optimized bit mappings decrease. Nevertheless, the introduced improvements induce no additional complexity and latency is even reduced.

VI. Conclusion

We proposed modifications for BICM transmission over AWGN channels using convolutional codes and Viterbi decoders. Based on a brief analysis of the Viterbi decoder and some recent results on the design of bit interleavers tailored to fading channels, we suggested to abandon bit interleaving on AWGN channels and rearrange the usually employed bit mapping. Such, we achieve a more equalized average bit metric reliability within the relevant processing range of the decoder. The BER of BICM can be significantly lowered at no additional cost and latency is reduced. Consequently, the gap of BICM to TCM in non-fading scenarios is not as large as usually stated in the literature.

References

[1] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson. On the optimality of the binary reflected Gray code. *IEEE Transactions on Information Theory*, 50(12):3170–3182, 2004.

[2] G. Caire, G. Taricco, and E. Biglieri. Bit-Interleaved Coded Modulation. *IEEE Transactions on Information Theory*, 44(3):927–946, 1998.

[3] A. Guillén i Fàbregas, A. Martinez, and G. Caire. Bit-Interleaved Coded Modulation. *Foundations and Trends in Communications and Information Theory*, 5(1/2):1–153, 2009.

[4] T. Hehn and J. B. Huber. LDPC codes and convolutional codes with equal structural delay: a comparison. *IEEE Transactions on Communications*, 57(6):1683–1692, June 2009.

[5] H. Imai and S. Hirakawa. A new multilevel coding method using error-correcting codes. *IEEE Transactions on Information Theory*, 23(3):371–377, May 1977.

[6] C. Stierstorfer and R. F. H. Fischer. (Gray) mappings for bit-interleaved coded modulation. In *Proceedings IEEE Vehicular Technology Conference Spring (VTC Spring)*, Dublin, Ireland, Apr. 2007.

[7] C. Stierstorfer and R. F. H. Fischer. Adaptive interleaving for bit-interleaved coded modulation. In *Proceedings 7th International ITG Conference on Source and Channel Coding (SCC)*, Ulm, Germany, Jan. 2008.

[8] G. Ungerböck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1):55–67, 1982.

[9] U. Wachsmann, R. F. H. Fischer, and J. B. Huber. Multilevel Codes: Theoretical Concepts and Practical Design Rules. *IEEE Transactions on Information Theory*, 45(5):1361–1391, July 1999.

[10] S. B. Wicker. *Error Control Systems for Digital Communications and Storage*. Prentice-Hall, Upper Saddle River, NJ, USA, 1 edition, 1995.

[11] E. Zehavi. 8-PSK Trellis Codes for a Rayleigh Channel. *IEEE Transactions on Communications*, 40(5):873–884, 1992.

X-Codes

(Invited Paper)

Saif Khan Mohammed*, Emanuele Viterbo,†, Yi Hong†, and Ananthanarayanan Chockalingam*
* Indian Institute of Science, India, saifind2007@yahoo.com, achockal@ece.iisc.ernet.in
† DEIS, Università della Calabria, Italy, yi.winnie.hong@gmail.com, viterbo@deis.unical.it

***Abstract*—We propose X-Codes for a time division duplex system with $n_t \times n_r$ multiple-input multiple-output (MIMO), using singular value decomposition (SVD) precoding at the transmitter. It is known that SVD precoding transforms the MIMO channel into parallel subchannels, resulting in a diversity order of only one. To improve the diversity order, X-Codes can be used prior to SVD precoding to pair the subchannels, i.e., each pair of information symbols is encoded by a fixed 2×2 real rotation matrix. X-Codes can be decoded using n_r low complexity two-dimensional real sphere decoders. Error probability analysis for X-Codes enables us to choose the optimal pairing and the optimal rotation angle for each pair. Finally, we show that our new scheme outperforms other existing precoding schemes.**

I. Introduction

In time division duplex (TDD) MIMO systems, where channel state information (CSI) is fully available at the transmitter, precoding techniques can provide large performance improvements and therefore have been extensively studied [1], [2], [4], [5], [11], [12].

In this paper, we consider singular value decomposition (SVD) of the channel, i.e., the MIMO channel can be seen as parallel subchannels [1], [2]. Note that it results in no diversity gain. To improve it, we propose X-Codes, whose name is due to the structure of their encoding matrix. Specifically, the X-Code pairs subchannels with low diversity orders with those having high diversity orders. The pairing is achieved by jointly coding the two subchannels with a two-dimensional real orthogonal matrix (which is effectively parametrized by a single angle). These angles are chosen *a priori* and do not change with each realization of the channel, and therefore we use the term "Code" instead of "Precoder". At the receiver, low complexity sphere decoders (SDs) can be used for maximum likelihood (ML) decoding.

Another precoding scheme that pairs subchannels to improve diversity has been recently proposed in [10], called E-dmin, which is only optimized for 4-QAM symbols. Hence for higher spectral efficiencies, X-Codes have better error performance. Moreover, X-Codes can be decoded with n_r 2-dimensional real SDs, whereas E-dmin requires $\frac{n_r}{2}$ 4-dimensional real SDs.

II. System model

We consider a TDD system with $n_t \times n_r$ MIMO ($n_r \leq n_t$), where the channel state information (CSI) is known perfectly at both the transmitter and receiver. Let $\mathbf{x} = (x_1, \ldots, x_{n_t})^T$ be the vector of symbols transmitted by the n_t transmit antennas, where $(\cdot)^T$ denotes transposition, and let $\mathbf{H} = (h_{ij})$, $i = 1, \ldots, n_r$, $j = 1, \ldots, n_t$, be the $n_r \times n_t$ channel coefficient matrix, with h_{ij} as the complex channel gain between the j-th transmit antenna and the i-th receive antenna. The standard Rayleigh flat fading model is assumed with $h_{ij} \sim \mathcal{N}_c(0,1)$, i.e., i.i.d. complex Gaussian random variables with zero mean and unit variance. The received vector with n_r symbols is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{1}$$

where $\mathbf{z}$ is a spatially uncorrelated Gaussian noise vector such that $\mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] = N_0 \mathbf{I}_{n_r}$, where † denotes the Hermitian transpose and $\mathbb{E}[.]$ is the expectation operator. Such a system has a maximum multiplexing gain of n_r. Let the number of information symbols transmitted be n_s ($n_s \leq n_r$). Let $\mathbf{T}$ be the $n_t \times n_s$ precoding matrix which is applied to the information vector $\mathbf{u} = (u_1, \ldots, u_{n_s})^T$ to yield the transmitted vector $\mathbf{x} = \mathbf{T}\mathbf{u}$. In general $\mathbf{T}$ is derived from the perfect knowledge of $\mathbf{H}$ at the transmitter. The transmission power constraint is given by $\mathbb{E}[\|\mathbf{x}\|^2] = P_T$ where $\| \cdot \|$ denotes the Euclidean norm. Finally, we define the signal-to-noise ratio as $\gamma \triangleq \frac{P_T}{N_0}$.

III. SVD precoding and X-Codes

SVD precoding is based on the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}$ ($\mathbf{U} \in \mathbb{C}^{n_r \times n_r}$, $\mathbf{\Lambda} \in \mathbb{C}^{n_r \times n_r}$ and $\mathbf{V} \in \mathbb{C}^{n_r \times n_t}$), where $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}_{n_r}$, $\mathbf{V}\mathbf{V}^\dagger = \mathbf{I}_{n_r}$ and $\mathbf{I}_{n_r}$ denotes the $n_r \times n_r$ identity matrix. The diagonal matrix $\mathbf{\Lambda}$ contains the singular values λ_i ($i = 1, \ldots n_r$) of $\mathbf{H}$ in decreasing order ($\lambda_1 \geq \lambda_2 \cdots \geq \lambda_{n_r} \geq 0$). Let $\tilde{\mathbf{V}} \in \mathbb{C}^{n_s \times n_t}$ be the submatrix with the first n_s rows of $\mathbf{V}$. The precoder uses $\mathbf{T} = \tilde{\mathbf{V}}^\dagger$ and the received vector is $\mathbf{y} = \mathbf{H}\mathbf{T}\mathbf{u} + \mathbf{z}$. Let $\tilde{\mathbf{U}} \in \mathbb{C}^{n_r \times n_s}$ be the submatrix with the first n_s columns of $\mathbf{U}$. The receiver then computes

$$\mathbf{r} = \tilde{\mathbf{U}}^\dagger \mathbf{y} = \tilde{\mathbf{\Lambda}}\mathbf{u} + \mathbf{w} \tag{2}$$

where $\mathbf{w} \in \mathbb{C}^{n_s}$ is still a uncorrelated Gaussian noise vector ($\mathbb{E}[\mathbf{w}\mathbf{w}^\dagger] = N_0 \mathbf{I}_{n_s}$). $\tilde{\mathbf{\Lambda}} \triangleq diag(\lambda_1, \lambda_2, \cdots \lambda_{n_s})$, and $\mathbf{r} = (r_1, \ldots, r_{n_s})^T$. The overall error performance is dominated by the minimum singular value λ_{n_s}. In the special case of full-rate transmission ($n_s = n_r$), the resulting diversity order is only one. This problem is alleviated by the proposed X-Codes, where pairs of subchannels are jointly coded.

We consider only the full-rate SVD precoding scheme with even n_r and $n_s = n_r$ (In general it is possible to have X-Codes with $n_s < n_r$ and odd n_s). Prior to SVD precoding,

*Saif K. Mohammed is visiting DEIS, Università della Calabria, Italy.

we now add a linear encoder $\mathbf{X} \in \mathbb{C}^{n_r \times n_r}$, which allows us to pair different subchannels in order to improve the diversity order of the system. The precoding matrix $\mathbf{T} \in \mathbb{C}^{n_t \times n_r}$ and the transmitted vector $\mathbf{x}$ are then given by

$$\mathbf{T} = \mathbf{V}^{\dagger}\mathbf{X}, \quad \mathbf{x} = \mathbf{V}^{\dagger}\mathbf{X}\mathbf{u} \tag{3}$$

The code matrix $\mathbf{X}$ is determined by the list of pairings of the subchannels and the linear code generating matrix for each pair. Let the list of pairings be $\{(i_k, j_k),\ k = 1, 2 \cdots \frac{n_r}{2}\}$, where all i_k and j_k are distinct positive integers between 1 and n_r and $i_k < j_k$. On the k-th pair of subchannels i_k and j_k, the symbols u_{i_k} and u_{j_k} are jointly coded using a 2×2 matrix $\mathbf{A}_k$. In order to reduce the ML decoding complexity, we restrict the entries of $\mathbf{A}_k$ to be real valued. In order to avoid transmitter power enhancement, we impose an orthogonality constraint on each $\mathbf{A}_k$ and parametrize it with a single angle θ_k.

$$\mathbf{A}_k = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \quad k = 1, \ldots n_r/2 \tag{4}$$

Each $\mathbf{A}_k$ is a 2×2 submatrix of the code matrix $\mathbf{X}$ as shown below.

$$\begin{aligned} X_{i_k, i_k} &= \cos(\theta_k)\,, X_{i_k, j_k} = \sin(\theta_k) \\ X_{j_k, i_k} &= -\sin(\theta_k)\,, X_{j_k, j_k} = \cos(\theta_k) \end{aligned} \tag{5}$$

where $X_{i,j}$ is the entry of $\mathbf{X}$ in the ith row and jth column. The orthogonality constraint on each $\mathbf{A}_k$ therefore implies that $\mathbf{X}$ is also orthogonal. We shall see later, that an optimal pairing in terms of achieving the best diversity order is one in which the k-th subchannel is paired with the $(n_r - k + 1)$th subchannel. The code matrix $\mathbf{X}$ for this pairing has a cross-form structure and thus the name "X-Codes". Each symbol in $\mathbf{u}$ takes values from a regular M^2-QAM constellation which consists of the M-PAM constellation $\mathcal{S} \stackrel{\Delta}{=} \{\beta(2i - (M - 1))\ |i = 0, 1, \cdots (M-1)\}$ used in quadrature on the real and the imaginary components of the channel. $\beta \stackrel{\Delta}{=} \sqrt{\frac{3E_s}{2(M^2-1)}}$ and $E_s = \frac{P_T}{n_r}$ is the average symbol energy for each information symbol in the vector $\mathbf{u}$. Gray mapping is used to map the bits separately to the real and imaginary component of the symbols in $\mathbf{u}$.

IV. Decoding of X-Codes

Given the received vector $\mathbf{y}$, the receiver computes $\mathbf{r} = \mathbf{U}^{\dagger}\mathbf{y}$. Using (1) and (3), we have $\mathbf{r} = \mathbf{\Lambda}\mathbf{X}\mathbf{u} + \mathbf{w} = \mathbf{M}\mathbf{u} + \mathbf{w}$, where $\mathbf{M} \stackrel{\Delta}{=} \mathbf{\Lambda}\mathbf{X}$ is the equivalent channel gain matrix and $\mathbf{w} \stackrel{\Delta}{=} \mathbf{U}^{\dagger}\mathbf{z}$ is a noise vector with the same statistics as $\mathbf{z}$.

Further let $\mathbf{r}_k \stackrel{\Delta}{=} [r_{i_k}, r_{j_k}]^T$, $\mathbf{u}_k \stackrel{\Delta}{=} [u_{i_k}, u_{j_k}]^T$, $\mathbf{w}_k \stackrel{\Delta}{=} [w_{i_k}, w_{j_k}]^T$, for $k = 1, 2, \cdots n_r/2$. For each $k \in \{1, 2, \cdots \frac{n_r}{2}\}$, let $\mathbf{M}_k \in \mathbb{R}^{2 \times 2}$ denote the 2×2 submatrix of $\mathbf{M}$ consisting of entries in the i_k and j_k rows and columns. Using (5) and the definition of $\mathbf{M}$ we have

$$\mathbf{M}_k = \begin{bmatrix} \lambda_{i_k}\cos(\theta_k) & \lambda_{i_k}\sin(\theta_k) \\ -\lambda_{j_k}\sin(\theta_k) & \lambda_{j_k}\cos(\theta_k) \end{bmatrix} \tag{6}$$

With these new definitions, $\mathbf{r}$ can be equivalently written as

$$\mathbf{r}_k = \mathbf{M}_k\mathbf{u}_k + \mathbf{w}_k,\ k\ =\ 1, 2, \cdots \frac{n_r}{2}. \tag{7}$$

Since $\mathbf{M}$ has real entries ML decoding for the k-th pair can be separated into independent ML decoding of the real and imaginary components of $\mathbf{u}_k$.

V. Performance evaluation and design of X-Codes

In this section, we analyze the word (block) error probability of X-Codes. Towards this end, we shall find the following Lemma useful ([13]).

Lemma 1: Given a real scalar channel modeled by $y = \sqrt{\alpha}x + n$, where $x = \pm\sqrt{E_s}$, $n \sim \mathcal{N}(0, \sigma^2)$, and the square fading coefficient α has $\mathbb{E}[\alpha] = 1$ and a cdf (Cumulative Density Function) $F(\alpha) = C\alpha^k + o(\alpha^k)$, for $\alpha \rightarrow 0^+$, where C is a constant and k is a positive integer, then the asymptotic error probability for $\gamma = E_s/\sigma^2 \rightarrow \infty$ is given by

$$P = \frac{C((2k-1).(2k-3)\ \cdots\ 5.3.1)}{2}\gamma^{-k} + o(\gamma^{-k})$$

■

Let P_k denote the ML word error probability for the k-th pair of subchannels. The overall word error probability for the transmitted information symbol vector is given by

$$P = 1 - \Pi_{k=1}^{\frac{n_r}{2}}(1 - P_k). \tag{8}$$

It is also clear that the word error probability for the real and the imaginary components of the k-th pair are the same. Therefore without loss of generality we can compute the word error probability only for the real component (denoted by $P_k^{'}$) and then $P_k = 1 - (1 - P_k^{'})^2$. Let us further denote by $P_k^{'}(\Re(\mathbf{u}_k))$ the probability of the real part of the ML decoder decoding not in favor of $\Re(\mathbf{u}_k)$ when $\mathbf{u}_k$ is transmitted on the k-th pair.

Getting an exact analytic expression is difficult, and therefore we try to get tight upper bounds. Towards this end let $\{\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k)\}$ denote the pairwise error event, whose probability is denoted by $P_k^{'}(\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k))$ (PEP) ($\Re(\cdot)$ denotes the real parts of a complex argument). Using the union bounding technique, $P_k^{'}(\Re(\mathbf{u}_k))$ is then upper bounded by the sum of all the possible PEPs. It is clear that this upper bound on $P_k^{'}(\Re(\mathbf{u}_k))$ induces an upper bound on $P_k^{'}$. The difference vector $\mathbf{z}_k = \Re(\mathbf{u}_k) - \Re(\mathbf{v}_k)$ can be written as $\sqrt{\frac{6E_s}{(M^2-1)}}(p\ ,\ q)^T$, where $(p, q) \in \mathbb{S}_M$ and $\mathbb{S}_M \stackrel{\Delta}{=} \{(p, q) | 0 \leq p \leq (M-1), 0 \leq q \leq (M-1), (p, q) \neq (0, 0)\}$. Then, the PEP $P_k^{'}(\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k))$ is given by

$$P_k^{'}(\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k)) = \mathbb{E}_{(\lambda_{i_k}, \lambda_{j_k})}\left[Q\left(\sqrt{\frac{3\gamma d_k^2(p, q, \theta_k)}{n_r(M^2-1)}}\right)\right] \tag{9}$$

where

$$\begin{aligned} d_k^2(p, q, \theta_k) \stackrel{\Delta}{=}\ & \lambda_{i_k}^2(p\cos(\theta_k) + q\sin(\theta_k))^2 \\ & + \lambda_{j_k}^2(q\cos(\theta_k) - p\sin(\theta_k))^2 \end{aligned}$$

and $Q(x)$ is the Gaussian tail function. Since $\lambda_{i_k} \geq \lambda_{j_k} \geq 0$, we have the inequality

$$\lambda_{i_k}^2 (p\cos(\theta_k) + q\sin(\theta_k))^2 < d_k^2(p,q,\theta_k) < \lambda_{i_k}^2 (p^2+q^2). \tag{10}$$

Since $Q(x)$ is a monotonically decreasing function with increasing argument, the PEP in (9) can be bounded as

$$P_k^{'}(\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k)) < \mathbb{E}_{\lambda_{i_k}}\left[Q\left(\sqrt{\frac{3\gamma\, \tilde{d}_k(p,q,\theta_k)\, \lambda_{i_k}^2}{n_r(M^2-1)}} \right) \right] \tag{11}$$

where $\tilde{d}_k(p,q,\theta_k) \triangleq (p^2+q^2)\cos^2(\theta_k - \tan^{-1}(\frac{q}{p}))$. Using Lemma 1 and the marginal pdf of the s-th eigenvalue λ_s^2 (for $\lambda_s^2 \rightarrow 0$) as given in [9], the bound in (11) can be further written as

$$P_k^{'}(\Re(\mathbf{u}_k) \rightarrow \Re(\mathbf{v}_k)) < b_k\Big(\frac{3\gamma\tilde{d}_k(p,q,\theta_k)}{n_r(M^2-1)}\Big)^{-\delta_k} + o(\gamma^{-\delta_k}) \tag{12}$$

where $\delta_k \triangleq (n_t - i_k + 1)(n_r - i_k + 1)$ and $b_k \triangleq \frac{C(i_k)((2\,\delta_k - 1)\,.\,\cdots\,5\,.\,3\,.\,1)}{2\,\delta_k}$, where C is defined in [9]. Using the upper bound in (12), the union bound is given by

$$P_k^{'} \leq \frac{b_k}{M^2}\Big[\sum_{(p,q)\in\mathbb{S}_M}\Big(\frac{3\gamma\tilde{d}_k(p,q,\theta_k)}{n_r(M^2-1)}\Big)^{-\delta_k}\Big] + o(\gamma^{-\delta_k}) \tag{13}$$

We further define $g(\theta_k, M)$ as follows,

$$g(\theta_k, M) = \min_{(p,q)\in\mathbb{S}_M} \tilde{d}_k(p,q,\theta_k) \tag{14}$$

Using (14) in (13), we can further upper bound $P_k^{'}$ as follows.

$$P_k^{'} \leq \frac{4(M-1)b_k}{M}\Big(\frac{3\gamma g(\theta_k,M)}{n_r(M^2-1)}\Big)^{-\delta_k} + o(\gamma^{-\delta_k}) \tag{15}$$

From (15) it is clear that the diversity order achievable by the k-th pair is at least δ_k. The diversity order achievable for the overall system (combined effect of all the pairs) is determined by the pair with the lowest diversity order. Let δ_{ord} denote the overall diversity order. Based on the above discussion δ_{ord} can be lower bounded as follows.

$$\delta_{ord} \geq \min_k \delta_k. \tag{16}$$

For a given MIMO configuration (n_t, n_r), the design of optimal X-Codes depends upon the optimal pairing of subchannels and the optimal angle for each pair. From the lower bound on δ_{ord} (16) it is clear that the following pairing of subchannels achieves the best lower bound

$$i_k = k \quad j_k = (n_r - k + 1),\ k\ = 1, 2\cdots\frac{n_r}{2}. \tag{17}$$

Note that this corresponds to a cross-form generator matrix $\mathbf{X}$. The lower bound on the overall diversity order is then given by $\delta_{ord} \geq (\frac{n_r}{2}+1)(n_t - \frac{n_r}{2}+1)$. Finding the optimal angle for the k-th pair is a difficult problem, hence we choose the angle which maximizes $g(\theta_k, M)$. Maximization of $g(\theta_k, M)$ can be computed offline as the angles for X-Codes are fixed *a priori*.

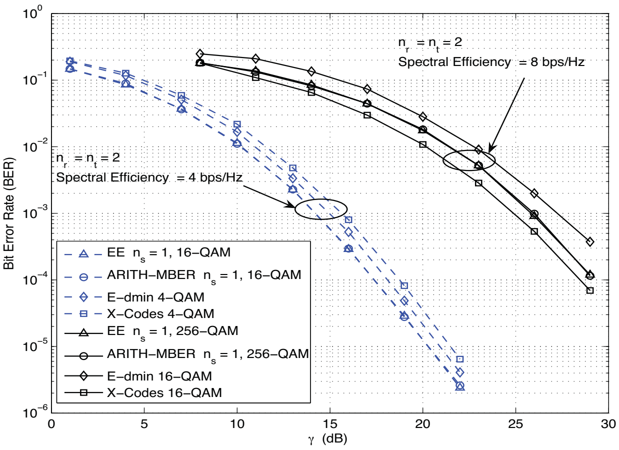


Fig. 1. Comparison between various precoders for $n_r = n_t = 2$ and target spectral efficiency = 4,8 bps/Hz.

VI. Simulation results

For all the simulations we assume $n_r = n_t$. The subchannel pairing for the X-Code is given by (17). The angle used for the subchannels is derived as discussed in section V (by optimizing upper bounds on the error probability expression).

Comparisons are made with *i*) the E-dmin (equal dmin precoder proposed in [10]), *ii*) the Arithmetic mean BER precoder (ARITH-MBER) proposed in [11], *iii*) the Equal Energy linear precoder (EE) based upon optimizing the minimum eigenvalue for a given transmit power constraint [12]), *iv*) the THP precoder based upon the idea of Tomlinson-Harashima precoding applied in the MIMO context [6]) and *v*) the channel inversion (CI) known as Zero Forcing precoder [3].

Among all the considered precoding schemes (except CI), E-dmin and X-Codes have the best diversity order. Though CI achieves infinite diversity, it suffers from power enhancement at the transmitter. We also observed that THP exhibit poor performance, when compared to the other precoders.

In Fig. 1, we plot the bit error rate (BER) for $n_r = n_t = 2$, and a target spectral efficiency of 4,8 bps/Hz. It is observed that for a target spectral efficiency of 4 bps/Hz, the best performance is achieved by ARITH-MBER and EE using only n_s=1 subchannel with 16-QAM modulation. X-Codes with 4-QAM modulation performs the worst. X-codes perform about 1.2 dB worse (at BER = 10^{-3}) compared to ARITH-MBER and EE. For a target spectral efficiency of 8 bps/Hz the results are totally different. X-Codes with 16-QAM modulation performs the best, and E-dmin performs the worst. Also the performance of X-codes is better than that of ARITH-MBER/EE by about 0.8 dB (at BER = 10^{-3}).

In Fig. 2, we plot the BER for $n_r = n_t = 4$, and a target spectral efficiency of 8,16 bps/Hz. It is observed that for a target spectral efficiency of 8 bps/Hz, the best performance is achieved by E-dmin with 4-QAM modulation. ARITH-MBER with N=3 subchannels (16-QAM modulation on one channel and 4-QAM on the other two) has the worst performance. X-codes perform worse than the E-dmin precoder by about 1 dB

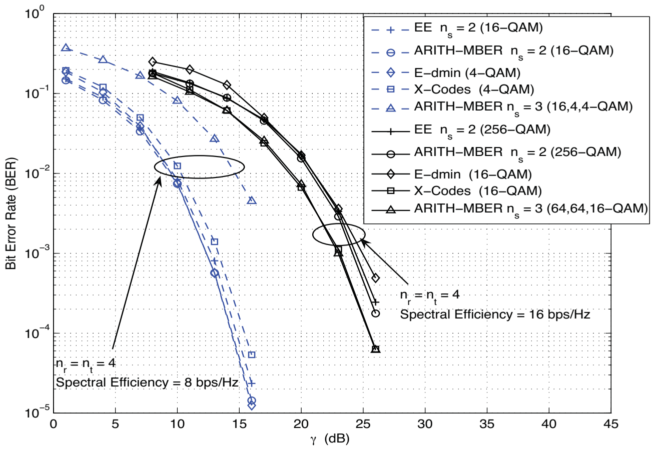


Fig. 2. Comparison between various precoders for $n_r = n_t = 4$ and target spectral efficiency = 8,16 bps/Hz.

(at BER = 10^{-3}).

For a target spectral efficiency of 16 bps/Hz X-codes with 16-QAM modulation performs the best. E-dmin performs the worst and is 2 dB away from X-Codes (at BER = 10^{-3}). E-dmin has poor performance since the precoder proposed in [10] has been optimized only for 4-QAM modulation, and therefore it does not perform that well for higher spectral efficiencies. E-dmin optimization for higher order QAM modulation is prohibitively too complex. It can be observed from Figs. 1 and 2 that for higher spectral efficiencies X-Codes perform the best when compared to other precoders.

VII. Complexity

All the considered precoders need to compute either SVD, QR or the pseudo-inverse of $\mathbf{H}$, whose complexity is $O(n_r^3)$. Generally, TDD is employed in a slowly fading channel, and therefore these computations can be performed at a very low rate compared to the rate of transmission. We, therefore, do not account for the complexity of these decompositions in the discussion below.

The encoding complexity of all the schemes have the same order. The complexity of the transmit pre-processing filter is $O(n_r n_t)$. If the number of operations were to be computed, CI and X-Codes would have the lowest complexity, since the linear and the THP precoders need extra pre-processing. E-dmin and X-Codes need to only compute SVD, which automatically gives the pre-processing matrices. X-Codes have lower encoding complexity compared to E-dmin, since the coding matrices $\mathbf{A}_k$ are fixed *a priori*. CI has an even lower complexity since there is no spatial coding.

The decoding complexity of all the schemes have a square dependence on n_r. This is due to the post-processing matrix filter at the receiver. The linear precoders, CI and THP employ post processing at the receiver, which enables independent ML decoding for each subchannel. E-dmin and X-Codes on the other hand use sphere decoding to jointly decode pairs of subchannels. ML decoding for X-Codes is accomplished by using n_r two-dimensional real sphere decoders.

However E-dmin requires $\frac{n_r}{2}$ 4-dimensional real sphere decoders. The average complexity of sphere decoding is cubic in the number of dimensions (and is invariant w.r.t modulation alphabet size M) [7], and therefore X-Codes have a much lower decoding complexity when compared to E-dmin.

VIII. Conclusion and Future Work

The proposed X-Codes are able to achieve full-rate and high diversity at a low complexity by pairing the subchannels before SVD precoding. Future work will focus on a generalization of X-Codes, which jointly codes more than two subchannels. Additional work will also address the reduction in decoder complexity and the generation of soft outputs.

Acknowledgment

The first author would like to thank the Italian Ministry of University and Research (MIUR) for the collaborative research program “Bando per borse a favore di giovani ricercatori indiani A.F. 2008”.

References

[1] G. Raleigh and J. Cioffi, “Spatio-Temporal Coding for Wireless Communication”, *IEEE Trans. Commun.*, pp. 357–366, March 1998.

[2] R. Knopp and G. Caire, “Power Control Schemes for TDD Systems with Multiple Transmit and Receive Antennas”, *Proc. of IEEE Global Telecommunications Conference (Globecom)*, pp. 2326–2330, Rio de Janeiro, Nov. 1999.

[3] P.W. Baier, M. Meurer, T. Weber, and H. Troeger, “Joint Transmission (JT), an alternative rationale for the downlink of Time Division CDMA using multi-element transmit antennas”, *Proc. of IEEE Int. Symp. on Spread Spectrum Techniques and Applications (ISSSTA)*, pp. 1–5, Parsippany, NJ, Sep. 2000.

[4] H. Harashima, and H. Miyakawa, “Matched transmission technique for channels with inter-symbol interference”, *IEEE Trans. on Communications*, vol. 20, pp. 774–780, 1972.

[5] M. Tomlinson, “New automatic equaliser employing modulo arithmetic”, *Electronics Letters*, 7, pp. 138–139, 1971.

[6] R.F.H. Fischer, C. Windpassinger, A. Lampe, and J.B. Huber, “Space-Time Transmission using Tomlinson-Harashima Precoding”, *Proc. Int. Zurich Seminar on Broadband Communications (IZS’02)*, Zurich, Switzerland, Feb. 2002.

[7] B. Hassibi and H. Vikalo, “On the Sphere-Decoding Algorithm I. Expected Complexity”, *IEEE Trans. on Information Theory*, vol. 53, no. 8, Aug. 2005.

[8] E. Biglieri, Y. Hong and E. Viterbo, “On fast-decodable space-time block codes”, *IEEE Trans. on Inform. Theory*, pp. 524–530, vol. 55, no. 2, Feb. 2009.

[9] L.G. Ordonez, D.P. Palomar, A.P. Zamora, and J.R. Fonollosa, “High-SNR analytical performance of Spatial Multiplexing MIMO Systems With CSI”, *IEEE Trans. on Signal Processing*, pp. 5447–5463, vol. 55, no. 11, Nov. 2007.

[10] B. Vrigneau, J. Letessier, P. Rostaing, L. Collin, and G. Burel, “Extension of the MIMO Precoder Based on the Minimum Euclidean Distance: A Cross-Form Matrix”, *IEEE J. of Selected Topics in Signal Processing*, pp. 135–146, vol. 2, no. 2, April. 2008.

[11] D. Perez Palomar, J.M. Cioffi, and M.A. Lagunas,“Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A unified Framework for Convex Optimization”, *IEEE. Trans. on Signal Processing*, pp. 2381–2401, vol. 51, no. 9, Sept. 2003.

[12] A. Scaglione, P. Stoica, S. Barbarossa, and G.B. Giannakis, Hemanth Sampath,“Optimal Designs for Space-Time Linear Precoders and Decoders”, *IEEE Trans. on Signal Processing*, pp. 1051–1064, vol. 50, no. 5, May. 2002.

[13] Z. Wang and G.B. Giannakis, “A Simple and General Parametrization Quantifying Performance in Fading Channels”, *IEEE Trans. on Communications*, pp. 1389–1398, vol. 51, no. 8, Aug. 2003.

Channel Coding LP Decoding and Compressed Sensing LP Decoding: Further Connections*

Alexandros G. Dimakis
Dept. of EE-Systems
Univ. of Southern California
Los Angeles, CA 90089, USA
dimakis@usc.edu

Roxana Smarandache[†]
Dept. of Math. and Stat.
San Diego State University
San Diego, CA 92182, USA
rsmarand@sciences.sdsu.edu

Pascal O. Vontobel
Hewlett–Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304, USA
pascal.vontobel@ieee.org

***Abstract*—Channel coding linear programming decoding (CC-LPD) and compressed sensing linear programming decoding (CS-LPD) are two setups that are *formally* tightly related. Recently, a connection between CC-LPD and CS-LPD was exhibited that goes beyond this formal relationship. The main ingredient was a lemma that allowed one to map vectors in the nullspace of some zero-one measurement matrix into vectors of the fundamental cone defined by that matrix.**

The aim of the present paper is to extend this connection along several directions. In particular, the above-mentioned lemma is extended from real measurement matrices where every entry is equal to either zero or one to complex measurement matrices where the absolute value of every entry is a non-negative integer. Moreover, this lemma and its generalizations are used to translate performance guarantees from CC-LPD to CS-LPD.

In addition, the present paper extends the formal relationship between CC-LPD and CS-LPD with the help of graph covers. First, this graph-cover viewpoint is used to obtain new connections between, on the one hand, CC-LPD for binary parity-check matrices, and, on the other hand, CS-LPD for complex measurement matrices. Secondly, this graph-cover viewpoint is used to see CS-LPD not only as a well-known relaxation of some zero-norm minimization problem but (at least in the case of real measurement matrices with only zeros, ones, and minus ones) also as a relaxation of a problem we call the zero-infinity operator minimization problem.

I. Introduction

This paper is a direct extension of a line of work that was started in [1] and that connects channel coding linear programming decoding [2], [3] and compressed sensing linear programming decoding [4]. Because the motivation and the aim for the results presented here are very much the same as they were in [1], we refer to that paper for an introduction. We remind the reader that **CC-MLD**, **CC-LPD**, **CS-OPT**, and **CS-LPD** stand for "channel coding maximum likelihood decoding," "channel coding linear programming decoding," "compressed sensing (sparsity) optimal decoding," and "compressed sensing linear programming decoding," respectively. Moreover, all vectors are column vectors.

The present paper is structured as follows. Section II presents three generalizations of [1, Lemma 11], which was the key result in [1]. First, this lemma is generalized from real measurement matrices where every entry is equal to either zero or one to complex measurement matrices where the absolute value of every entry is equal to either zero or one. In that process we also generalize the mapping that is applied to the vectors in the nullspace of the measurement matrix. Secondly, this lemma is generalized to hold also for complex measurement matrices where the absolute value of every entry is a non-negative integer. Finally, the third generalization of this lemma extends the types of mappings that can be applied to the vectors in the nullspace of the measurement matrix. With this, Section III translates performance guarantees from **CC-LPD** to **CS-LPD**. Afterwards, Section IV tightens the already close formal relationship between **CC-LPD** and **CS-LPD** with the help of graph covers, a line of results that is continued in Section V, which presents **CS-LPD** for certain measurement matrices not only as the well-known relaxation of some zero-norm minimization problem but also as the relaxation of some other minimization problem. Finally, some conclusions are presented in Section VI.

Besides the notation defined in [1], we will also use the following conventions and extensions of notions previously introduced. For any $M \in \mathbb{Z}_{>0}$, we let $[M] \triangleq \{1, \ldots, M\}$. We remind the reader that in [1] we extended the use of the absolute value operator $|\cdot|$ from scalars to vectors. Namely, if $\boldsymbol{a} = (a_i)_i$ is a complex vector then we define $|\boldsymbol{a}|$ to be the complex vector $\boldsymbol{a}' = (a'_i)_i$ of the same length as $\boldsymbol{a}$ with entries $a'_i = |a_i|$ for all i. Similarly, in this paper we extend the use of the absolute value operator $|\cdot|$ from scalars to matrices.

We let $|\cdot|_*$ be an arbitrary norm for the complex numbers. As such, $|\cdot|_*$ satisfies for any $a, b, c \in \mathbb{C}$ the triangle inequality $|a+b|_* \leqslant |a|_* + |b|_*$ and the equality $|c \cdot a|_* = |c| \cdot |a|_*$. In the same way the absolute value operator $|\cdot|$ was extended from scalars to vectors and matrices, we extend the norm operator $|\cdot|_*$ from scalars to vectors and matrices.

We let $\|\cdot\|_*$ be an arbitrary vector norm for complex vectors that reduces to $|\cdot|_*$ for vectors of length one. As such, $\|\cdot\|_*$ satisfies for any $c \in \mathbb{C}$ and any complex vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of equal length the triangle inequality $\|\boldsymbol{a}+\boldsymbol{b}\|_* \leqslant \|\boldsymbol{a}\|_* + \|\boldsymbol{b}\|_*$ and the equality $\|c \cdot \boldsymbol{a}\|_* = |c| \cdot \|\boldsymbol{a}\|_*$.

For any complex vector $\boldsymbol{a}$ we define the zero-infinity operator to be $\|\boldsymbol{a}\|_{0,\infty} \triangleq \|\boldsymbol{a}\|_0 \cdot \|\boldsymbol{a}\|_\infty$, i.e., the product of the zero norm $\|\boldsymbol{a}\|_0 = \#\operatorname{supp}(\boldsymbol{a})$ of $\boldsymbol{a}$ and of the infinity

[*] Due to space limitations, some of the proofs and some of the sections have been omitted. A complete version will be posted on arxiv.

[†] Supported by NSF Grants DMS-0708033 and TF-0830608.

norm $\|\boldsymbol{a}\|_\infty = \max_i |a_i|$ of $\boldsymbol{a}$. Note that for any $c \in \mathbb{C}$ and any complex vector $\boldsymbol{a}$ it holds that $\|c \cdot \boldsymbol{a}\|_{0,\infty} = |c| \cdot \|\boldsymbol{a}\|_{0,\infty}$.

Finally, for any $n, M \in \mathbb{Z}_{>0}$ and any length-n vector $\boldsymbol{a}$ we define the M-fold lifting of $\boldsymbol{a}$ to be the vector $\boldsymbol{a}^{\uparrow M} = (a^{\uparrow M}_{(i,m)})_{(i,m)} \in \mathbb{C}^{Mn}$ with components given by

$$a^{\uparrow M}_{(i,m)} \triangleq a_i, \quad (i,m) \in [n] \times [M].$$

Moreover, for any vector $\tilde{\boldsymbol{a}} = (\tilde{a}_{(i,m)})_{(i,m)}$ of length $M \cdot n$ over $\mathbb{C}$ or $\mathbb{F}_2$ we define the projection of $\tilde{\boldsymbol{a}}$ to the space $\mathbb{C}^n$ to be the vector $\boldsymbol{a} \triangleq \boldsymbol{\varphi}_M(\tilde{\boldsymbol{a}})$ with components given by

$$a_i \triangleq \frac{1}{M} \sum_{m \in [M]} \tilde{a}_{(i,m)}, \quad i \in [n].$$

(In the case where $\tilde{\boldsymbol{a}}$ is over $\mathbb{F}_2$, the summation is over $\mathbb{C}$ and we use the the standard embedding of $\{0, 1\}$ into $\mathbb{C}$.)

II. Beyond Measurement Matrices with Zeros and Ones

The aim of this section is to extend [1, Lemma 11], which is a reformulation of [5, Lemma 6], to matrices beyond zero-one measurement matrices. In that vein we will present three generalizations in Lemmas 2, 5, and 6. For ease of reference, let us restate [1, Lemma 11].

Lemma 1 ([1, Lemma 11]) *Let $\boldsymbol{H}_{\mathrm{CS}}$ be a measurement matrix that contains only zeros and ones. Then*

$$\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \quad \Rightarrow \quad |\boldsymbol{\nu}| \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CS}}).$$

■

Because in the proofs of the upcoming lemmas we will have to show that certain vectors lie in the fundamental cone $\mathcal{K} \triangleq \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}})$ [2], [3], [6], [7] of the parity-check matrix $\boldsymbol{H}_{\mathrm{CC}}$ of some binary linear code, for convenience let us list here a set of inequalities that characterize $\mathcal{K}$. Namely, $\mathcal{K}$ is the set of all vectors $\boldsymbol{\omega} \in \mathbb{R}^n$ that satisfy

$$\omega_i \geqslant 0 \qquad \text{(for all } i \in \mathcal{I}\text{)}\,, \tag{1}$$

$$\omega_i \leqslant \sum_{i' \in \mathcal{I}_j \setminus i} \omega_{i'} \qquad \text{(for all } j \in \mathcal{J},\ \text{ for all } i \in \mathcal{I}_j\text{)}\,. \tag{2}$$

With this, we are ready to discuss our first generalization of [1, Lemma 11], which generalizes [1, Lemma 11] from real measurement matrices where every entry is equal to either zero or one to complex measurement matrices where the absolute value of every entry is equal to either zero or one. Note that the upcoming lemma also generalizes the mapping that is applied to the vectors in the nullspace of the measurement matrix.

Lemma 2 *Let $\boldsymbol{H}_{\mathrm{CS}} = (h_{j,i})_{j \in \mathcal{J}, i \in \mathcal{I}}$ be some measurement matrix over $\mathbb{C}$ such that $|h_{j,i}| \in \{0, 1\}$ for all $(j,i) \in \mathcal{J} \times \mathcal{I}$, and let $|\cdot|_*$ be an arbitrary norm for complex numbers. Then*

$$\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{C}}(\boldsymbol{H}_{\mathrm{CS}}) \quad \Rightarrow \quad |\boldsymbol{\nu}|_* \in \mathcal{K}\big(|\boldsymbol{H}_{\mathrm{CS}}|\big).$$

Remark: Note that $\mathrm{supp}(\boldsymbol{\nu}) = \mathrm{supp}(|\boldsymbol{\nu}|_*)$.

Proof: Omitted. ■

The second generalization of [1, Lemma 11] generalizes that lemma to hold also for complex measurement matrices where the absolute value of every entry is an integer. In order to present this lemma, we need the following definition, which will be illustrated by Example 4.

Definition 3 *Let $\boldsymbol{H}_{\mathrm{CS}} = (h_{j,i})_{j \in \mathcal{J}, i \in \mathcal{I}}$ be some measurement matrix over $\mathbb{C}$ such that $|h_{j,i}| \in \mathbb{Z}_{\geqslant 0}$ for all $(j,i) \in \mathcal{J} \times \mathcal{I}$, and let $M \in \mathbb{Z}_{>0}$ be such that $M \geqslant \max_{(j,i) \in \mathcal{J} \times \mathcal{I}} |h_{j,i}|$. We define an M-fold cover of $\tilde{\boldsymbol{H}}_{\mathrm{CS}}$ of $\boldsymbol{H}_{\mathrm{CS}}$ as follows: for $(j,i) \in \mathcal{J} \times \mathcal{I}$, $h_{j,i}$ is replaced by $h_{j,i}/|h_{j,i}|$ times the sum of $|h_{j,i}|$ arbitrary $M \times M$ permutation matrices with non-overlapping support.* □

Note that the entries of the matrix $\tilde{\boldsymbol{H}}_{\mathrm{CS}}$ in Definition 3 all have absolute value equal to either zero or one.

Example 4 Let

$$\boldsymbol{H}_{\mathrm{CS}} \triangleq \begin{pmatrix} 1 & 0 & \sqrt{2}(1+i) \\ -2 & i & 3 \end{pmatrix}.$$

Clearly

$$|\boldsymbol{H}_{\mathrm{CS}}| \triangleq \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 3 \end{pmatrix},$$

and so, choosing $M = 3$,

$$\tilde{\boldsymbol{H}}_{\mathrm{CS}} \triangleq \left(\begin{array}{ccc|ccc|ccc} 0 & 1 & 0 & 0 & 0 & 0 & \frac{1+i}{\sqrt{2}} & \frac{1+i}{\sqrt{2}} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & \frac{1+i}{\sqrt{2}} & 0 & \frac{1+i}{\sqrt{2}} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \frac{1+i}{\sqrt{2}} & \frac{1+i}{\sqrt{2}} \\ \hline 0 & -1 & -1 & i & 0 & 0 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & i & 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & 0 & 0 & i & 1 & 1 & 1 \end{array}\right).$$

is a possible matrix obtained by the procedure defined in Definition 3. □

Lemma 5 *Let $\boldsymbol{H}_{\mathrm{CS}} = (h_{j,i})_{j \in \mathcal{J}, i \in \mathcal{I}}$ be some measurement matrix over $\mathbb{C}$ such that $|h_{j,i}| \in \mathbb{Z}_{\geqslant 0}$ for all $(j,i) \in \mathcal{J} \times \mathcal{I}$. With this, let $M \in \mathbb{Z}_{>0}$ be such that $M \geqslant \max_{(j,i) \in \mathcal{J} \times \mathcal{I}} |h_{j,i}|$, and let $\tilde{\boldsymbol{H}}_{\mathrm{CS}}$ be a matrix obtained by the procedure in Definition 3. Moreover, let $|\cdot|_*$ be an arbitrary norm for complex numbers. Then*

$$\begin{aligned} \boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{C}}(\boldsymbol{H}_{\mathrm{CS}}) \quad &\Rightarrow \quad \boldsymbol{\nu}^{\uparrow M} \in \mathrm{nullspace}_{\mathbb{C}}(\tilde{\boldsymbol{H}}_{\mathrm{CS}}) \\ &\Rightarrow \quad \big|\boldsymbol{\nu}^{\uparrow M}\big|_* \in \mathcal{K}\big(|\tilde{\boldsymbol{H}}_{\mathrm{CS}}|\big). \end{aligned}$$

Additionally, with respect to the first implication sign we have the following converse: for any $\tilde{\boldsymbol{\nu}} \in \mathbb{C}^{Mn}$ we have

$$\boldsymbol{\varphi}_M(\tilde{\boldsymbol{\nu}}) \in \mathrm{nullspace}_{\mathbb{C}}(\boldsymbol{H}_{\mathrm{CS}}) \quad \Leftarrow \quad \tilde{\boldsymbol{\nu}} \in \mathrm{nullspace}_{\mathbb{C}}(\tilde{\boldsymbol{H}}_{\mathrm{CS}}).$$

Proof: Omitted. ■

Finally, we present our third generalization of [1, Lemma 11], which generalizes the mapping that is applied to the vectors in the nullspace of the measurement matrix.

Lemma 6 *Let $\boldsymbol{H}_{\mathrm{CS}} = (h_{j,i})_{j \in \mathcal{J}, i \in \mathcal{I}}$ be some measurement matrix over $\mathbb{C}$ such that $|h_{j,i}| \in \{0,1\}$ for all $(j,i) \in \mathcal{J} \times \mathcal{I}$. Moreover, let $L \in \mathbb{Z}_{>0}$, and let $\|\cdot\|_*$ be an arbitrary vector norm for complex numbers. Then*

$$\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(L)} \in \mathrm{nullspace}_{\mathbb{C}}(\boldsymbol{H}_{\mathrm{CS}}) \quad \Rightarrow \quad \boldsymbol{\omega} \in \mathcal{K}\big(|\boldsymbol{H}_{\mathrm{CS}}|\big),$$

where $\boldsymbol{\omega} \in \mathbb{R}^n$ is defined such that for all $i \in \mathcal{I}$,

$$\omega_i = \left\| \left(\nu_i^{(1)}, \ldots, \nu_i^{(L)} \right) \right\|_* .$$

Proof: Omitted. ■

We conclude this section with two remarks. First, it is clear that Lemma 6 can be extended in the same way as Lemma 5 extends Lemma 2. Secondly, similarly to the approach in [1] where [1, Lemma 11] was used to translate "positive results" about **CC-LPD** to "positive results" about **CS-LPD**, the new Lemmas 2, 5, and 6 can be the basis for translating results from **CC-LPD** to **CS-LPD**.

III. Translating Performance Guarantees

In this section we use [1, Lemma 11] to transfer "positive performance results" for **CC-LPD** of low-density parity-check (LDPC) codes to "positive performance results" for **CS-LPD** of zero-one measurement matrices. In particular, three positive threshold results for **CC-LPD** are used to obtain three results that are, to the best of our knowledge, novel for compressed sensing. At the end of the section we will also use Lemma 2 with $|\cdot|_* = |\cdot|$ to study dense measurement matrices with entries in $\{-1, 0, +1\}$.

We will need the notion of an *expander graph*.

Definition 7 *Let G be a bipartite graph where the nodes in the two node classes are called left-nodes and right-nodes, respectively. If $\mathcal{S}$ is some subset of left-nodes, we let $\mathcal{N}(\mathcal{S})$ be the subset of right-nodes that are adjacent to $\mathcal{S}$. Then, given parameters $d_{\mathrm{v}} \in \mathbb{Z}_{>0}$, $\gamma \in (0,1)$, $\delta \in (0,1)$, we say that G is a $(d_{\mathrm{v}}, \gamma, \delta)$-expander if all left-nodes of G have degree d_{v} and if for all left-node subsets $\mathcal{S}$ with $\#\mathcal{S} \leqslant \gamma n$ it holds that $\#\mathcal{N}(\mathcal{S}) \geqslant \delta d_{\mathrm{v}} \cdot \#\mathcal{S}$.* □

Expander graphs have been studied extensively in past work on channel coding and compressed sensing (see, e.g., [8], [9]). It is well-known that randomly constructed left-regular bipartite graphs are expanders with high probability (see, e.g., [10]).

In the following, similar to the way a Tanner graph is associated with a parity-check matrix [11], we will associate a Tanner graph with a measurement matrix. Note that the variable and constraint nodes of a Tanner graph will be called left-nodes and right-nodes, respectively.

Corollary 8 *Let $d_{\mathrm{v}} \in \mathbb{Z}_{>0}$, let $\gamma \in (0,1)$, and let $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{n' \times n}$ be a measurement matrix. Moreover, assume that the Tanner graph of $\boldsymbol{H}_{\mathrm{CS}}$ is a $(d_{\mathrm{v}}, \gamma, \delta)$-expander with sufficient expansion, more precisely, with*

$$\delta > \frac{2}{3} + \frac{1}{3d_{\mathrm{v}}}$$

*(along with the technical condition $\delta d_{\mathrm{v}} \in \mathbb{Z}_{>0}$). Then **CS-LPD** based on the measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ can recover all k-sparse vectors, i.e., all vectors whose support size is at most k, for*

$$k < \frac{3\delta - 2}{2\delta - 1} \cdot (\gamma n - 1).$$

Proof: This result is obtained by combining the results in [1] with [10, Theorem 1]. ■

Interestingly, for $\delta = 3/4$ the recoverable sparsity k matches exactly the performance of the fast compressed sensing algorithm in [9] and the performance of the simple bit-flipping channel decoder of Sipser an Spielman [8], but our result holds for the basis pursuit LP relaxation **CS-LPD**. Expansion has been shown to suffice for **CS-LPD** in [12] but with a different proof and yielding different constants. For $n'/n = 1/2$ and $d_{\mathrm{v}} = 32$, the result of [10] establishes that sparse expander-based zero-one measurement matrices will recover all $k = \alpha n$ sparse vectors for $\alpha \leqslant 0.000175$.

Whereas the above result gave a deterministic guarantee, the following result is based on a so-called weak bound for **CC-LPD** and gives a probabilistic guarantee.

Corollary 9 *Let $d_{\mathrm{v}} \in \mathbb{Z}_{>0}$. Consider a random measurement matrix $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{n' \times n}$ that is formed by placing d_{v} random ones in each column, and zeros elsewhere. This measurement matrix succeeds to recover a randomly supported $k = \alpha n$ sparse vector with probability $1 - o(1)$ if α is below some threshold function $\alpha_{n'}(d_{\mathrm{v}}, n'/n)$.*

Proof: The result is obtained by combining the results in [1] with [13, Theorem 1]. The latter paper also contains a way to compute achievable threshold values $\alpha_{n'}(d_{\mathrm{v}}, n'/n)$. ■

For $n'/n = 1/2$ and $d_{\mathrm{v}} = 8$, a random measurement matrix will recover a $k = \alpha n$ sparse vector with random support with high probability if $\alpha \leqslant 0.002$. This is, of course, a much higher threshold compared to the one presented above but it only holds with high probability over the vector support (therefore it is a so-called weak bound). To the best of our knowledge, this is the first weak bound obtained for random sparse measurement matrices.

The best thresholds known for LP decoding were recently obtained by Arora, Daskalakis, and Steurer [14] but require matrices that are both left and right regular and also have logarithmically growing girth. A random bipartite matrix will not have this latter property but there are explicit deterministic constructions that achieve this (for example the construction presented in Gallager's thesis [15, Appendix C]). By translating the results from [14] to the compressed sensing setup we obtain the following result.

Corollary 10 *Let $d_{\mathrm{v}}, d_{\mathrm{c}} \in \mathbb{Z}_{>0}$. Consider a measurement matrix $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{n' \times n}$ whose Tanner graph is a $(d_{\mathrm{v}}, d_{\mathrm{c}})$-regular bipartite graph with $\Omega(\log n)$ girth. This measurement matrix succeeds to recover a randomly supported $k = \alpha n$ sparse vector with probability $1 - o(1)$ if α is below some threshold function $\alpha'_{n'}(d_{\mathrm{v}}, d_{\mathrm{c}}, n'/n)$.*

Proof: The result is obtained by combining the results in [1] with [14, Theorem 1]. The latter paper also contains a way to compute achievable threshold values $\alpha'_{n'}(d_{\mathrm{v}}, d_{\mathrm{c}}, n'/n)$. ■

For $n'/n = 1/2$, an application of the above result to a $(3,6)$-regular Tanner graph with logarithmic girth (obtained from the Gallager construction) tells us that sparse vectors with sparsity $k = \alpha n$ are recoverable with high probability for $\alpha \leqslant 0.05$. Therefore, measurement matrices based on Gallager's deterministic construction (of low-density parity-check matrices) form the best known class of sparse measurement matrices for the compressed sensing setup considered here.

We conclude this section with some considerations about dense measurement matrices, highlighting our current understanding that the translation of positive performance guarantees from **CC-LPD** to **CS-LPD** displays the following behavior: the denser a measurement matrix is the weaker are the translated performance guarantees.

Remark 11 Consider a randomly generated $n' \times n$ measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ where every entry is generated i.i.d. according to the distribution

$$\begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{with probability } 2/3 \\ -1 & \text{with probability } 1/6 \end{cases} .$$

This matrix, after multiplying it by the scalar $\sqrt{3/n}$, has the restricted isometry property (RIP). (See [16], which proves this property based on results in [17], which in turn proves that this family of matrices has a non-zero threshold.) On the other hand, one can show that the family of parity-check matrices where every entry is generated i.i.d. according to the distribution

$$\begin{cases} 1 & \text{with probability } 1/3 \\ 0 & \text{with probability } 2/3 \end{cases}$$

does *not* have a non-zero threshold under **CC-LPD** for the BSC [18]. □

Therefore, we conclude that the connection between **CS-LPD** and **CC-LPD** given by Lemma 2 is not tight for dense matrices in the sense that the performance of **CS-LPD** based on dense measurement matrices can be much better than predicted by the performance of **CC-LPD** based on their parity-check matrix counterpart.

IV. Reformulations based on Graph Covers

(This section has been omitted.)

V. Minimizing the Zero-Infinity Operator

(This section has been omitted.)

VI. Conclusions and Outlook

In this paper we have extended the results of [1] along various directions. In particular, we have translated performance guarantees from **CC-LPD** to performance guarantees for the recovery of *exactly sparse* vectors under **CS-LPD**. As part of future work we plan to investigate the translation of performance guarantees from **CC-LPD** to performance guarantees for the recovery of *approximately sparse* vectors under **CS-LPD**.

References

[1] A. G. Dimakis and P. O. Vontobel, "LP decoding meets LP decoding: a connection between channel coding and compressed sensing," in *Proc. 47th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, Illinois, USA, Sep. 30–Oct. 2 2009.

[2] J. Feldman, "Decoding error-correcting codes via linear programming," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[3] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 954–972, Mar. 2005.

[4] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] R. Smarandache and P. O. Vontobel, "Absdet-pseudo-codewords and perm-pseudo-codewords: definitions and properties," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, June 28–July 3 2009.

[6] R. Koetter and P. O. Vontobel, "Graph covers and iterative decoding of finite-length codes," in *Proc. 3rd Intern. Symp. on Turbo Codes and Related Topics*, Brest, France, Sept. 1–5 2003, pp. 75–82.

[7] P. O. Vontobel and R. Koetter, "Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes," *accepted for IEEE Trans. Inform. Theory, available online under* `http://www.arxiv.org/abs/cs.IT/0512078`, 2007.

[8] M. Sipser and D. Spielman, "Expander codes," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1710–1722, Nov. 1996.

[9] W. Xu and B. Hassibi, "Efficient compressive sensing with determinstic guarantees using expander graphs," in *Proc. IEEE Information Theory Workshop*, Tahoe City, CA, USA, Sept. 2–6 2007, pp. 414–419.

[10] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 82–89, Jan. 2007.

[11] R. M. Tanner, "A recursive approach to low-complexity codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[12] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery," in *Proc. 46th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, Illinois, USA, Sept. 23–26 2008.

[13] C. Daskalakis, A. G. Dimakis, R. M. Karp, and M. J. Wainwright, "Probabilistic analysis of linear programming decoding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3565–3578, Aug. 2008.

[14] S. Arora, C. Daskalakis, and D. Steurer, "Message-passing algorithms and improved LP decoding," in *Proc. 41st Annual ACM Symp. Theory of Computing*, Bethesda, MD, USA, May 31–June 2 2009.

[15] R. G. Gallager, *Low-Density Parity-Check Codes.* M.I.T. Press, Cambridge, MA, 1963.

[16] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[17] D. Achlioptas, "Database-friendly random projections," in *Proc. 20th ACM Symp. on Principles of Database Systems*, Santa Barbara, CA, USA, 2001, pp. 274–287.

[18] P. O. Vontobel and R. Koetter, "Bounds on the threshold of linear programming decoding," in *Proc. IEEE Information Theory Workshop*, Punta Del Este, Uruguay, Mar. 13–16 2006, pp. 175–179.

Achievable Rates for Multicell Systems with Femtocells and Network MIMO

O. Simeone
CWCSPR, ECE Dept.,
NJIT, Newark, USA

E. Erkip
Dept. of ECE, Polytechnic Inst. of NYU
Brooklyn, NY, USA

S. Shamai (Shitz)
Dept. of EE, Technion
Haifa, Israel

***Abstract*—[1]The uplink of a cellular system where macrocells are overlaid with femtocells is studied. Each femtocell is served by a home base station (HBS) that is connected to the macrocell base station (BS) via a last-mile access link, such as DSL or cable followed by the Internet. Decoding at the BSs takes place via either standard single-cell processing or multicell processing (i.e., network MIMO). Closed and open-access femtocells are considered. Achievable per-cell sum-rates are derived in this setting for a linear cellular network. Overall, the analysis lends evidence to the performance advantages of open-access femtocells and sheds light on the performance trade-offs between single/multi-cell processing and different relaying strategies at the femtocells.**

I. Introduction

With the recent advances in coding and multiantenna technology, interference is becoming the performance-limiting factor in terms of area and spectral efficiency of cellular systems. To cope with interference, two diametrically opposite strategies are currently being investigated. On one end, *femtocells* reduce the size of a cell to contain only the customer's premises, thus allowing transmission with smaller powers and the possibility to reuse the spectrum more aggressively [1]. On the other end, *network MIMO* or *multicell processing* (MCP) [2][3] creates clusters of macrocells for joint coding/ decoding in order to better manage *inter-cell interference*.

A femtocell consists of a short-range low-cost *home base station (HBS)*, installed within the customer's premises, that serves either only indoor users, in case of *closed-access* femtocells, or possibly also outdoor users that are within the HBS coverage range, in case of *open-access* femtocells. Femtocells in open-access mode provide an asset that the network designer can exploit to manage the interference created by outdoor users towards the femtocell and other macrocells. In this work, we provide an information-theoretic look at the performance trade-offs between open and closed-access femtocells, on the one hand, and the deployment of femtocells and MCP, on the other. Analysis is performed by resorting to a simple cellular model that extends [2] and by deriving achievable rates that are then compared via numerical results.

Notation: We define $\mathcal{C}(\mathbf{A}) = 1/2\log_2\det(\mathbf{I}+\mathbf{A})$ for a positive definite $\mathbf{A}$; Notation $[1,N]$ represents the set of numbers $\{1,...,N\}$.

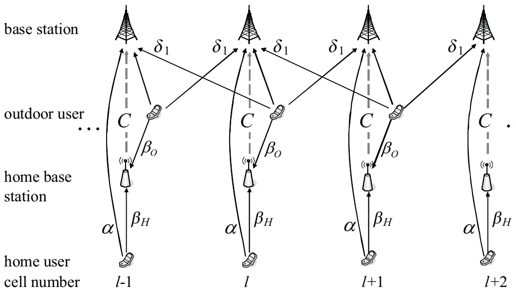


Fig. 1. A linear multicell system where each macrocell is overlaid with a femtocell. Each HBS is connected to the local BS via a last-mile link of capacity C ($L = 1$ in the figure).

II. System Model

Consider a linear cellular system similar to [2], where M cells are arranged on a line, as for a corridor or a highway, as shown in Fig. 1. Each cell, served by a base station (BS), contains a single femtocell, served by a HBS, and presents the same number of outdoor (i.e., outside the femtocell) and home (i.e., within the femtocell) users. Assuming that the channel gains are the same for different home/outdoor users in the same cell, and focusing the analysis on achievable sum-rates, we can concentrate without loss of generality on a single outdoor and home user per cell, as shown in Fig. 1 [5].

Signals generated within each femtocell are received with relevant power only by the local BS with power gain α and the local HBS with power gain β_H, while outdoor users are received not only by the local BS and HBS (with power gains $\delta_0 = 1$ and β_O, respectively), but also by L adjacent BSs on either side with symmetric power gains δ_l, $l \in [1,L]$. Given the above, the received signals at a given time instant for the BS and home BS in the lth cell can be expressed as, respectively

$$Y_l = \sum_{i=-L}^{L} \sqrt{\delta_i} X_{O,[l+i]} + \sqrt{\alpha} X_{H,l} + N_{Y,l} \qquad (1a)$$

$$\text{and } Z_l = \sqrt{\beta_O} X_{O,l} + \sqrt{\beta_H} X_{H,l} + N_{Z,l}, \qquad (1b)$$

[1] The work of O. Simeone was supported by U.S. NSF grant CCF-0905446. The work of E. Erkip by U.S. NSF grant CCF-0914899. The work of S. Shamai was supported by the European Commision in the framework of the FP7 Network of Excellence in Wireless COMmunications, NEWCOM++.

where $X_{O,l}$ and $X_{H,l}$ are the signals transmitted by the outdoor ("O") and home ("H") user in the lth cell, and $(N_{Y,l}, N_{Z,l})$ are independent Gaussian noise processes with unit-power. Power constraints for outdoor and home users are defined as $P_O, P_{H.}$, respectively. Moreover, to avoid edge effects, in (1), we have assumed that inter-cell interference affects cells in a circulant fashion, so that every cell is impaired by the same number of interferers (we have defined $[l+i]$ as the modulo-M operation and assumed $M \geq 2L+1$).

Finally, the HBS is assumed to be connected to the corresponding BS via a last-mile connection (such as DSL or cable) followed by the Internet, whose overall capacity is C bits/ dim[2]. This link is wired and orthogonal to the wireless channels [1]. The scenario at hand can be seen as an extension of the model in [2] to include femtocells and is related to the models in [6] and references therein for mesh networks.

We consider two alternatives for decoding at the BSs: (*i*) *Single-cell Processing* (SCP): The BS in each cell decodes independently; (*ii*) *Multicell Processing* (MCP): All BSs in the system are connected to a central processor (CP) for joint decoding. The CP collects the signals of all BSs and jointly decodes all the M outdoor and M home messages jointly. Furthermore, for both SCP and MCP, we will study the performance of closed-access (CA) and open-access (OA) femtocells. CA femtocells treat the signal of the outdoor user as interference, whereas OA femtocells may serve as relays towards the BS for the outdoor users. The aim is to identify pairs of home user and outdoor users rates R_H and R_O, respectively, that are achievable in each cell according to the usual definitions.

III. Single-Cell processing (SCP)

In this section, we study achievable rate pairs (R_H,R_O) with SCP and OA or CA femtocells.

A. Closed-Access Femtocells

We start with CA femtocells.

Proposition 1 (CA,SCP): Rates satisfying the following conditions

$$\begin{aligned} R_H &< \min\left\{\mathcal{C}\left(\frac{\beta_H P_H}{1+\beta_O P_O}\right), \mathcal{C}\left(\frac{\alpha P_H}{1+\Delta P_O}\right)+C\right\} \\ R_O &< \mathcal{C}\left(\frac{P_O}{1+\Delta P_O}\right) \\ R_O + R_H &< \mathcal{C}\left(\frac{P_O+\alpha P_H}{1+\Delta P_O}\right)+C, \end{aligned}$$

are achievable with SCP and CA femtocells, where with $\Delta = 2\sum_{l=1}^{L}\delta_l$.

Proof (sketch): The HBS decodes the home user's message by treating the outdoor user as noise (of power $\beta_O P_O$). Having decoded, the HBS provides C bits/dim of the decoded message to the BS. The BS performs joint decoding of home and outdoor users' messages by treating inter-cell signals as noise (of power ΔP_O). In this process, thanks to the C bits received from the HBS, the equivalent rate of the home user to be decoded by the BS is reduced to $R_H - C$ (see, e.g., [4]). The proof is completed using standard arguments.

B. Open-Access Femtocells

Turning to OA femtocells, we consider two classes of strategies. In the first, the HBS decodes both home and outdoor users' messages and then shares the last-mile link capacity C for transmission of bits from both messages (Decode-and-Forward, DF). In the second, the HBS simply compresses and forwards (CF) the received signal. It is noted that the latter scheme does not require codebook information at the HBS and thus reduces the signaling overhead.

1) Decode-and-Forward: Proposition 2 (OA-DF,SCP): The convex hull of the union of the rates that satisfy

$$\begin{aligned} R_H &< \min\left\{\mathcal{C}\left(\beta_H P_H\right), \mathcal{C}\left(\frac{\alpha P_H}{1+\Delta P_O}\right)+\gamma C\right\} \\ R_O &< \min\{\mathcal{C}\left(\beta_O P_O\right), \mathcal{C}\left(\frac{P_O}{1+\Delta P_O}\right)+(1-\gamma)C\} \\ R_O + R_H &< \min\{\mathcal{C}\left(\beta_H P_H + \beta_O P_O\right), \\ &\qquad \mathcal{C}\left(\frac{\alpha P_H + P_O}{1+\Delta P_O}\right)+C\} \end{aligned}$$

for some $0 \leq \gamma \leq 1$ is achievable with SCP and OA femtocells employing DF relaying.

Proof (sketch): The HBS decodes both home and outdoor users' messages and then sends γC bits/dim of the decoded home message and $(1-\gamma)C$ bits/dim of the decoded outdoor message to the BS. The BS performs joint decoding as discussed for Proposition 1, but on codebooks of equivalent rates $R_H - \gamma C$ and $R_O - (1-\gamma)C$.

2) Compress-and-Forward: Proposition 3 (OA-CF,SCP): Rates satisfying the following conditions

$$\begin{aligned} R_H &< \mathcal{C}\left(\frac{\alpha P_H}{1+\Delta P_O}+\frac{\beta_H P_H}{1+\sigma^2}\right) \\ R_O &< \mathcal{C}\left(\frac{P_O}{1+\Delta P_O}+\frac{\beta_O P_O}{1+\sigma^2}\right) \\ R_O + R_H &< \mathcal{C}\left(\mathbf{A}\right) \end{aligned}$$

with

$$\mathbf{A} = \begin{bmatrix} \frac{P_O+\alpha P_H}{1+\Delta P_O} & \frac{\sqrt{\beta_O}P_O+\sqrt{\alpha\beta_H}P_H}{\sqrt{(1+\Delta P_O)(1+\sigma^2)}} \\ \frac{\sqrt{\beta_O}P_O+\sqrt{\alpha\beta_H}P_H}{\sqrt{(1+\Delta P_O)(1+\sigma^2)}} & \frac{\beta_H P_H+\beta_O P_O}{1+\sigma^2} \end{bmatrix}$$

are achievable with SCP and OA femtocells employing CF relaying, where

$$\sigma^2 = \frac{\left[1+\beta_O P_O+\beta_H P_H - \frac{(\sqrt{\beta_O}P_O+\sqrt{\alpha\beta_H}P_H)^2}{P_O+\alpha P_H+\Delta P_O}\right]}{2^{2C}-1}. \quad (2)$$

Proof (sketch): The HBS compresses the received signal to a description $\hat{Z}_l$ of C bits/dim using Wyner-Ziv quantization, exploiting the fact that the BS has side information Y_l. The compression noise (2) is found by imposing $I(Z_l;\hat{Z}_l|Y_l) = C$ following standard arguments (see, e.g., [7]). The lth BS performs joint decoding based on the signals $(Y_l, \hat{Z}_l)$.

[2] We measures the rates in bits per (real) dimension (dim).

IV. Multicell Processing (MCP)

In this section, we address achievable rates in the presence of MCP. We recall that, with MCP, decoding is performed at a CP connected via ideal links to all BSs. For notational convenience, we define the channel matrix $\mathbf{H}$ between outdoor users and the M BSs as the $M \times M$ circulant matrix whose first column is given by

$$[\sqrt{\delta_0}\sqrt{\delta_1}\cdots\sqrt{\delta_{L_C}}\mathbf{0}_{L-(2L_C+1)}\sqrt{\delta_{L_C}}\sqrt{\delta_{L_C-1}}\cdots\sqrt{\delta_1}].$$

We also denote the eigenvalues of $\mathbf{HH}^T$ as $\lambda_l = \left(1+2\sum_{l=1}^{L_C}\sqrt{\delta_l}\cos\left(\frac{2\pi}{L}l\right)\right)^2$, $l \in [0, M-1]$.

A. Closed Access

Proposition 4 (CA,MCP): Rates satisfying the following conditions

$$\begin{aligned}
R_H &< \min\left\{\mathcal{C}\left(\frac{\beta_H P_H}{1+\beta_O P_O}\right),\ \mathcal{C}(\alpha P_H)+C\right\} \\
R_O &< \frac{1}{L}\sum_{l=0}^{L=1}\mathcal{C}(\lambda_l P_O) \\
R_O+R_H &< \frac{1}{L}\sum_{l=0}^{L=1}\mathcal{C}(\lambda_l P_O+\alpha P_H)+C
\end{aligned}$$

are achievable with MCP and CA femtocells.

Proof (sketch): The HBS operates as for Proposition 1. The CP decodes jointly all the home and outdoor users' messages based on the signals Y_l, $l \in [1, M]$ and the MC bits/dim received from the HBSs. The equivalent rate of the home users to be decoded is $R_H - C$ due to the bits received from the HBS, as, e.g., for Proposition 1.

B. Open Access

Turning to OA femtocells, as for SCP, we study both DF and CF strategies.

1) Decode-and-Forward: Proposition 5 (OA-DF,MCP): The convex hull of the union of the rates that satisfy

$$\begin{aligned}
R_H &< \min\{\mathcal{C}(\beta_H P_H),\ \mathcal{C}(\alpha P_H)+\gamma C\} \\
R_O &< \min\left\{\begin{array}{l}\mathcal{C}(\beta_O P_O), \\ \frac{1}{L}\sum_{l=0}^{L=1}\mathcal{C}(\lambda_l P_O)+(1-\gamma)C\end{array}\right\} \\
R_O+R_H &< \min\left\{\begin{array}{l}\mathcal{C}(\beta_H P_H+\beta_O P_O), \\ \frac{1}{L}\sum_{l=0}^{L=1}\mathcal{C}(\lambda_l P_O+\alpha P_H)+C\end{array}\right\}
\end{aligned}$$

for some $0 \leq \gamma \leq 1$ is achievable with MCP and OA femtocells employing DF relaying.

Proof (sketch): The HBS operates as for Proposition 2 and the CP performs joint decoding as for Proposition 4.

2) Compress-and-Forward: Proposition 6 (OA-CF,MCP): Rates satisfying the following conditions

$$\begin{aligned}
R_H &< \mathcal{C}\left(\alpha P_H+\frac{\beta_H P_H}{1+\sigma^2}\right) \\
R_O &< \frac{1}{L}\sum_{l=0}^{L=1}\mathcal{C}\left(\lambda_l P_O+\frac{\beta_O P_O}{1+\sigma^2}\right) \\
R_O+R_H &< \frac{1}{L}\mathcal{C}(\mathbf{B})
\end{aligned}$$

with (2) and

$$\mathbf{B} = \left[\begin{array}{cc} P_O\mathbf{HH}^T+\alpha P_H\mathbf{I} & \frac{\sqrt{\beta_O}P_O\mathbf{H}+\sqrt{\alpha\beta_H}P_H\mathbf{I}}{\sqrt{1+\sigma^2}} \\ \frac{\sqrt{\beta_O}P_O\mathbf{H}^T+\sqrt{\alpha\beta_H}P_H\mathbf{I}}{\sqrt{1+\sigma^2}} & \left(\frac{\beta_O P_O}{1+\sigma^2}+\frac{\beta_H P_H}{1+\sigma^2}\right)\mathbf{I} \end{array}\right]$$

are achievable with MCP and OA femtocells employing CF relaying.

Proof (sketch): The HBS operates as for Proposition 3 and the CP decodes jointly all messages based on the signals $(Y_l, \hat{Z}_l)$, $l \in [1, M]$. It is noted that using σ^2 in (2) implies that decompression of $\hat{Z}_l$ is performed at the lth BS. However, with MCP, one could potentially improve the performance by moving decompression from the BSs to the CP, which has better side information (namely, all Y_l with $l \in [1, M]$). We do not pursue this further here.

V. Numerical Results

In this section, we provide some insight into the performance comparison of different scenarios and strategies through numerical results. Throughout, we set parameters $P_O = P_H = 4$, $\beta_H = 20dB$ and $\alpha = -10dB$, which implies that the indoor channel gain between home user and HBS is 30dB better than the channel home user-BS [1], $M = 30$, $L = 1$. We focus on maximum achievable equal rates $R_H = R_O$ for the different considered techniques.

We start by concentrating on the performance comparison between CA and OA femtocells, by varying the outdoor-HBS power gain β_O with fixed $\delta_1 = 0.4$ and $C = 1.5$. Fig. 2 shows that CA femtocells, due to the macro-to-femto interference, are largely outperformed by OA techniques for increasing β_O. More specifically, OA-DF becomes advantageous over CA for sufficiently large β_O, while OA-CF, for the range of β_O shown in the figure, performs always at least as well as CA. As for the comparison between OA-CF and OA-DF, on the one hand, OA-CF has the advantage of enabling joint decoding at the receiver (BS for SCP or CP for MCP), while having the drawback of adding extra noise via compression. On the other hand, OA-DF has the advantage of providing "clean" information bits to the receiver, at the cost of causing a potential performance bottleneck at the home BS for decoding. This trade-off is clear from Fig. 2: Whenever decoding at the HBS does not set the performance bottleneck (i.e., for β_O large enough), OA-DF outperforms OA-CF, while the opposite is true when β_O is small so that decoding of the outdoor users at the home BS limits the performance of OA-DF[3].

We further discuss the comparison between the performance of MCP and SCP in Fig. 3 for $\beta_O = 10$, and varying intercell interference power gain δ_1. It can be seen that as the inter-cell interference δ_1 increases, the advantages of MCP become more pronounced for all techniques. It is also noted, similar to the example above, that CF appears to be performing better when deployed with MCP than with SCP. This is further discussed below.

[3] For β_O larger than β_H (not shown in the figure given the minor relevance of this regime), the performance of CF keeps degrading as β_O increases due to the larger quantization noise, down to the performance attainable with $C = 0$.

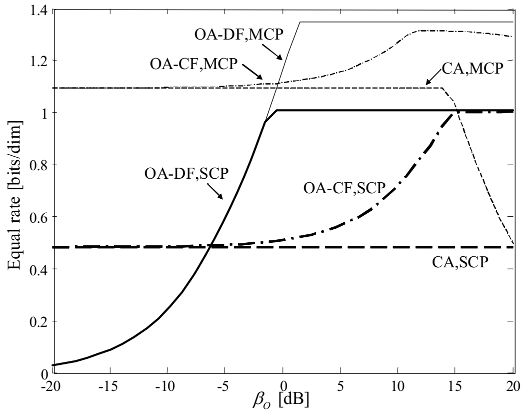


Fig. 2. Equal achievable rate $R_H = R_O$ versus the outdoor-HBS power gain β_O ($\delta_1 = 0.4$, $C = 1.5$, $P_O = P_H = 4$, $\beta_H = 20dB$, $\alpha = -10dB$, $M = 30$, $L = 1$).

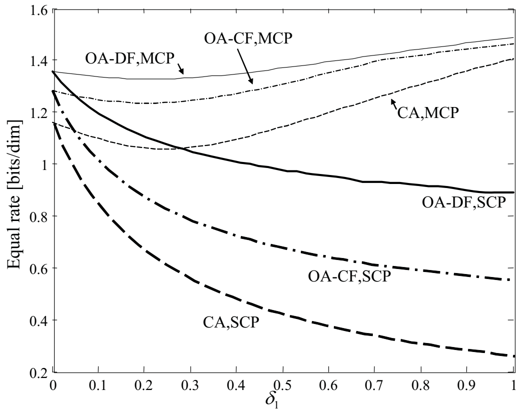


Fig. 3. Equal achievable rate $R_H = R_O$ versus the inter-cell interference power gain δ_1 ($\beta_O = 10$, $C = 1.5$, $P_O = P_H = 4$, $\beta_H = 20dB$, $\alpha = -10dB$, $M = 30$, $L = 1$).

Fig. 4 shows the maximum equal rate of different techniques versus the last-mile link capacity C for $\delta_1 = 0.5$ and $\beta_O = -3dB$. It is seen, following the discussion above, that, if C is small, OA-DF is appropriate since the performance is limited by decoding at the BS. However, as C increases, the equal rate achievable by OA-DF saturates to the maximum equal rate decodable at the HBS (which is the same for both SCP and MCP), while OA-CF does not suffer from such saturation and keeps exploiting larger values of C to improve the quality of the compressed signal provided to the receiver. It is also noted that with MCP the crossing point between the performance of OA-DF and OA-CF occurs for smaller values of C than SCP, due to the greater decoding power at the CP with respect to the single BS.

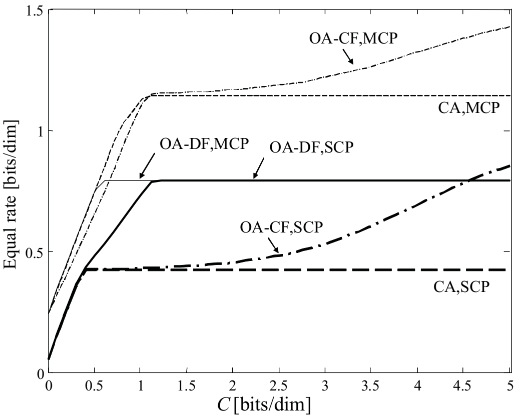


Fig. 4. Equal achievable rate $R_H = R_O$ versus the last-mile HBS-BS link capacity C ($\delta_1 = 0.5$, $P_O = P_H = 4$, $\beta_O = -3dB$, $\beta_H = 20dB$, $\alpha = -10dB$, $M = 30$, $L = 1$).

VI. Concluding Remarks

While network MIMO and femtocells are being mostly developed and studied in separation, this paper has argued for a joint analysis, given the interplay between the two technologies. An important observation is that femtocells, when allowed to work in an open-access mode, have a potentially relevant role for interference management, since they can exploit their dedicated (wired) connection to the BS to reduce radio interference by serving also outdoor users. However, the relaying strategy must be carefully designed according to whether decoding at the BSs implements network MIMO or not, in order not to create performance bottlenecks. This increased interference margin may be dually turned into a corresponding reduction in power emissions, thus moving towards "greener" wireless communications.

References

[1] V. Chandrasekhar, J. G. Andrews and A. Gatherer, “Femtocell networks: a survey,” *IEEE Comm. Magazine*, vol. 46, no. 9, pp. 59-67, Sept. 2008.

[2] A. D. Wyner, “Shannon-theoretic approach to a Gaussian cellular multiple-access channel,” *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1713-1727, Nov. 1994.

[3] M. K. Karakayali, G.J. Foschini and R.A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56-61, Aug. 2006.

[4] Y.-H. Kim, “Coding techniques for primitive relay channels,” in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[5] O. Simeone, E. Erkip and S. Shamai (Shitz), "Robust transmission and interference management for femtocells with unreliable network access," submitted.

[6] O. Somekh, O. Simeone, H. V. Poor, and S. Shamai (Shitz), “Cellular Systems with Full-Duplex Compress-and-Forward Relaying and Cooperative Base-Stations,” in *Proc. IEEE International Symposium Inform. Theory (ISIT 2008)*, Toronto, Canada, July 6-11, 2008.

[7] G. Kramer, *Topics in Multi-User Information Theory*, Foundations and Trends in Communications and Information Theory, vol. 4, no. 4-5, pp. 265-444, 2007.

On the value of data sharing in constrained-backhaul network MIMO

Randa Zakhour and David Gesbert
Mobile Communications Department
Eurecom
06560 Sophia Antipolis, France
{zakhour, gesbert}@eurecom.fr

***Abstract*—This paper addresses the problem of cooperation in a multicell environment where base stations wish to jointly serve multiple users, under a constrained-capacity backhaul. Such a constraint limits, among other things, data sharing and network MIMO concepts need to be revised accordingly. More precisely, we focus on the downlink, and propose to use the backhaul to transmit several messages to each user: some are common to several transmitters and joint precoding is possible, others are private and only local precoding may be done. For the two-cell setup we derive achievable rate regions, optimizing the corresponding beamforming design. Numerical results show how this added flexibility improves performance.**

I. Introduction

A major issue in several types of wireless networks is that of interference. This problem is especially acute in cellular networks with full reuse of the spectrum across all base stations. In traditional designs, each base station obtains from the backbone the data the signals intended for users in its coverage area, i.e., if one ignores cases of soft handover, data for users is not available at multiple base stations. Recent research rooted in MIMO theory has suggested the benefits of relaxing this constraint, thereby allowing for data to be shared at multiple transmitters so that a giant broadcast MIMO channel results. In such a scenario, multicell processing in the form of joint precoding is realized: this scheme is referred to as network MIMO (a.k.a. multicell MIMO).

Full data sharing subsumes very high capacity backhaul links, which may not be feasible, or even simply desirable, in certain applications. Some previous authors have tackled the problem of joint transmission when the backhaul links between a central unit and the transmitters (the base stations), or amongst the latter, are finite, in which case the resulting multicell channel no longer corresponds to a MIMO broadcast channel, nor does it correspond to the so-called interference channel. Among others, in [3] and [4], joint encoding for the downlink of a cellular system is studied under the assumption that the base stations are connected to a central unit via finite-capacity links. The authors investigate different transmission schemes and ways of using the backhaul capacity in the context of a modified version of Wyner's channel model. One of their main conclusions is that ”central encoding with oblivious cells”, whereby quantized versions of the signals to be transmitted from each base station, computed at the central unit, are sent over the backhaul links, is shown to be a very attractive option for both ease of implementation and performance, unless high data rate are required. If this is the case, the base stations need to be involved in the encoding, i.e. at least part of the backhaul link should be used for sending the messages themselves not the corresponding codewords.

In [5], an optimization framework, for an adopted backhaul usage scheme, is proposed for the downlink of a large cellular system. A so-called joint transmission configuration matrix is defined: this specifies which antennas in the system serve each user, along with the number of quantization bits, for each antenna, associated with that user. Thus the transmit signal of all users are transmitted centrally and different quantized versions of each user's signal are transmitted to the appropriate base stations: this is similar to the central encoding with oblivious cells scheme in [4], except that a more realistic system model is assumed, and the number of quantization bits per user and per antenna are optimized.

In [6], a more information-theoretic approach is taken and a two-cell setup is considered in which, in addition to links between the network and each base station, a finite-capacity link connects the two multi-antenna base stations: the authors view the thus formed channel as a superposition of an interference channel and a broadcast channel. The backhaul is used to share the data to be jointly transmitted: this could be in the form of the full messages, or of quantized versions of the signals to transmit, depending on whether the data is coming from the network directly or shared over the link between the base stations.

In this work, we also consider a two-cell setup, but limit the backhaul to be between the network and each of the base stations. Some of the questions we try to answer are:

- Given the backhaul constraints, what kind of rates can we expect to achieve?
- How much of the data needs to be shared to achieve these rates? I.e. how useful is network MIMO when backhaul constraints are present?

We thus specify a transmission scheme whereby superposition coding is used to transmit signals to each user: this allows us to formulate a continuum between full message sharing between base stations and the conventional network with single serving base stations; the data rate is in fact split between two distinct forms of data to be received by the users, a private form to be sent by the 'serving' base alone and a common form to be transmitted via multiple bases. We express the corresponding rate region in terms of the backhaul constraints and the beamforming vectors used to carry the different signals, and reduce finding the boundary of said region to solving a set of convex optimization problems. This is in contrast to the schemes in [6] where the nonconvexity of the problem makes it difficult to characterize the optimum beamforming vectors to use, and the suboptimal scheme of maximum ratio transmission is resorted to. We also formulate the problem in a way that both the rates that correspond to conventional transmission (each base station receives the signals for one user only) are accounted for in the backhaul.

II. System Model

The system considered is shown in Figure 1. In this preliminary study, we focus on a two transmitter two receiver setup. Receivers have a single antenna each, whereas transmitters have $N_t \geq 1$ antennas: $\mathbf{h}_{ij}$ is the channel between transmitter j and user i; $\mathbf{h}_i$ is user i's whole channel. We assume a backhaul link of capacity C_i between the central processor (or backbone network) and transmitter i, for $i = 1, 2$: it will be used to transmit the messages for each user. We distinguish between different types of messages:

- private messages, which are known at, and consequently only sent from, one of the transmitters, and
- shared or common messages, which are known to both transmitters and consequently jointly transmitted. Note that this notion of a common message is different from that commonly used in the context of interference channels for example, as they do not correspond to messages to be decoded by both receivers, but rather to messages to be sent by both transmitters.

Assumptions We assume each receiver does single user detection, in the sense that the other user's signal is treated as noise. Moreover, we do not rely on dirty paper coding (DPC) to avoid inter-user interference. Furthermore, full channel state information (CSI) is available at both transmitters, since we want to focus on the cost of sharing data.

Notation In what follows, $\bar{i} = \mod (i, 2) + 1, i = 1, 2$ and is used to denote the other transmitter/receiver depending on the context.

A. Backhaul usage

Let $r_{i,p}$ denote the rate of the private message transmitted from transmitter i to receiver i, and $r_{i,c}$ denote the rate of the shared message intended for receiver i. Thus, his total rate is

$$r_i = r_{i,p} + r_{i,c} \tag{1}$$

The backhaul link to transmitter $i, i = 1, 2$ will be used to transmit the messages (so no quantization is done here) that transmitter i is meant to know, i.e. the private message for receiver i along with both shared messages.

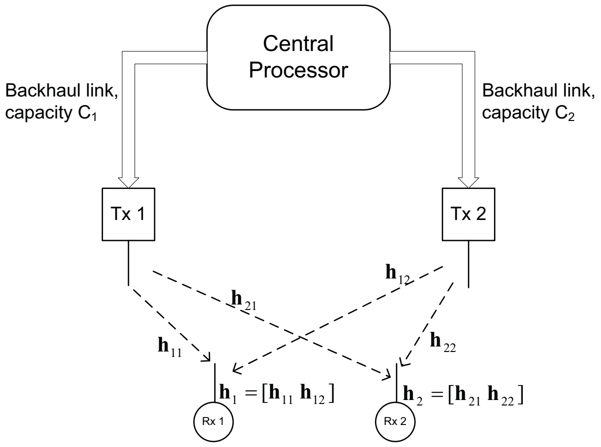

Central Processor
Backhaul link, capacity C_1
Backhaul link, capacity C_2
Tx 1
Tx 2
$\mathbf{h}_{12}$
$\mathbf{h}_{21}$
$\mathbf{h}_{11}$
$\mathbf{h}_{22}$
$\mathbf{h}_1 = [\mathbf{h}_{11}\ \mathbf{h}_{12}]$
$\mathbf{h}_2 = [\mathbf{h}_{21}\ \mathbf{h}_{22}]$
Rx 1
Rx 2

Fig. 1. Constrained backhaul setup.

B. Background: MAC with Common Message

Given our system assumptions, if transmission to user $\bar{i}$ has already been specified, we are left with a MAC with a common message between the two transmitters and user i [1]. Denoting by σ_i^2 the power of the interference, and restricting the transmitted signals to have *rank-1 covariance matrices*, the following rate region is achievable

$$r_{i,p} \leq \log_2\left(1 + \frac{|\mathbf{h}_{ii}\mathbf{w}_{i,p}|^2}{\sigma_i^2}\right),$$

$$r_i = r_{i,p} + r_{i,c} \leq \log_2\left(1 + \frac{|\mathbf{h}_{ii}\mathbf{w}_{i,p}|^2 + |\mathbf{h}_i\mathbf{w}_{i,c}|^2}{\sigma_i^2}\right), \tag{2}$$

where the covariance matrix of user i's private message is $\mathcal{C}_{i,p} = \mathbf{w}_{i,p}\mathbf{w}_{i,p}^H$, and that of the common message is $\mathcal{C}_{i,c} = \mathbf{w}_{i,c}\mathbf{w}_{i,c}^H$, and where $\mathbf{w}_{i,p}$ and $\mathbf{w}_{i,c}$ are such that the power constraints at the transmitters are met. Note that Gaussian signaling is optimal for a two-transmitter MAC with a common message (see [2], where this is shown in the context of a MAC with cooperating encoders.).

C. Over the air transmission

In light of the previous subsection, the transmitted signal may be modeled as follows:

$$\mathbf{x} = \begin{bmatrix} \mathbf{w}_{1,c} & \mathbf{w}_{2,c} \end{bmatrix} \begin{bmatrix} s_{1,c} \\ s_{2,c} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{1,p} \\ \mathbf{0} \end{bmatrix} s_{1,p} + \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_{2,p} \end{bmatrix} s_{2,p}, \tag{3}$$

where $\mathbf{x} \in \mathbb{C}^{2Nt}$ is the transmitted signal, such that the first N_t elements are transmitter 1's transmit signal, the remaining N_t elements are transmitter 2's signal. Though not necessarily optimal, Gaussian signaling is assumed, so that $s_{1,p}, s_{1,c}, s_{2,p}, s_{2,c}$ are all $\mathcal{CN}(0, 1)$. Per base station power constraints $P_i, i = 1, 2$ imply that:

$$\|\mathbf{D}_i\mathbf{w}_{1,c}\|^2 + \|\mathbf{D}_i\mathbf{w}_{2,c}\|^2 + \|\mathbf{w}_{i,p}\|^2 \leq P_i, \tag{4}$$

where $\mathbf{D}_i$ is a matrix whose only non-zero elements are elements $(N_t-1)i+1 : iN_t$ along the diagonal and are equal to 1.

D. Achievable rates

The signal received at receiver i will be given by (see (3)):

$$\begin{aligned} y_i &= \mathbf{h}_i\mathbf{x} + n_i = \begin{bmatrix} \mathbf{h}_{i1} & \mathbf{h}_{i2} \end{bmatrix} \mathbf{x} + n_i \\ &= \mathbf{h}_i\mathbf{w}_{1,c}s_{1,c} + \mathbf{h}_i\mathbf{w}_{2,c}s_{2,c} + \mathbf{h}_{i1}\mathbf{w}_{1,p}s_{1,p} \\ &+ \mathbf{h}_{i2}\mathbf{w}_{2,p}s_{2,p} + n_i \end{aligned} \tag{5}$$

Given our single-user detection (SUD) assumption, user i's rates will satisfy (2) with σ_i^2 given by:

$$\sigma_i^2 = \sigma^2 + \left|\mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i},p}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{\bar{i},c}\right|^2. \tag{6}$$

III. Rate Region

An achievable rate region $\mathcal{R}$ is the set of $(r_1, r_{1,p}, r_2, r_{2,p})$, as specified above, that satisfy the given backhaul and power constraints.

One way to obtain the rate region boundary is to solve the following problem for $\alpha \in [0,1]$, which maximizes the sum rate, subject to a given split between the two users.

$$\begin{aligned} \text{max. } & r \\ \text{s.t. } & r_1 \geq \alpha r \\ & r_2 \geq (1-\alpha)r \\ & r_1 + r_2 - r_{2,p} \leq C_1, \quad r_1 + r_2 - r_{1,p} \leq C_2 \\ & r_i \leq \log_2\left(1 + \frac{\left|\mathbf{h}_{ii}\mathbf{w}_{i,p}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{i,c}\right|^2}{\sigma^2 + \left|\mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i},p}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{\bar{i},c}\right|^2}\right), i = 1,2, \\ & r_{i,p} \leq \log_2\left(1 + \frac{\left|\mathbf{h}_{ii}\mathbf{w}_{i,p}\right|^2}{\sigma^2 + \left|\mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i},p}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{\bar{i},c}\right|^2}\right), i = 1,2, \\ & \|\mathbf{w}_{i,p}\|^2 + \|\mathbf{D}_i\mathbf{w}_{i,c}\|^2 + \|\mathbf{D}_i\mathbf{w}_{\bar{i},c}\|^2 \leq P_i, i = 1,2 \end{aligned} \tag{7}$$

We solve this problem using a bisection method.

1) $r_{min} = 0, r_{max} = C_1 + C_2$
2) Repeat until $r_{max} - r_{min} < \epsilon$
   a) Set $r = (r_{min} + r_{max})/2$
   b) Determine feasibility of r: this is detailed in subsection III-A below.
   c) If feasible, $r_{min} = r$, else $r_{max} = r$.

A. Establishing feasibility of a given rate

Assume sum rate r and α to be fixed. Thus, $r_1 = \alpha r$, $r_2 = (1-\alpha)r$. Establishing feasibility of a given rate pair hinges on the following two remarks:

- For r_i to be supported, it cannot possibly exceed C_i, and
- Sharing information whenever possible outperforms not doing so. Thus if a rate pair is not achievable for the minimum possible private message rates, it is not achievable at all. Given the backhaul constraints, the minimum possible private message rate $r_{i,p}$, $i = 1,2$ is given by:

$$\left(r_{i,p}\right)_{min} = \min(r_i, \max(0, r_1 + r_2 - C_{\bar{i}})). \tag{8}$$

How to establish whether a given rate tuple $(r_1, r_{1,p}, r_2, r_{2,p})$ and determine a beamforming scheme to achieve it is specified in section III-B below. If this procedure yields a valid solution for rate tuple $\left(r_1, \left(r_{1,p}\right)_{min}, r_2, \left(r_{2,p}\right)_{min}\right)$, then r is feasible [1].

B. Feasibility of $(r_1, r_{1,p}, r_2, r_{2,p})$

Assume r_1, r_2, $r_{1,p}$ and $r_{2,p}$ are fixed. Solve

$$\begin{aligned} \text{min. } & \sum_{i=1}^{2} \left[\|\mathbf{w}_{i,c}\|^2 + \|\mathbf{w}_{i,p}\|^2\right] \\ \text{s.t. } & 2^{r_i} - 1 \leq \frac{\left|\mathbf{h}_{ii}\mathbf{w}_{i,p}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{i,c}\right|^2}{\sigma^2 + \left|\mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i}}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{\bar{i},c}\right|^2}, i = 1,2, \\ & 2^{r_{i,p}} - 1 \leq \frac{\left|\mathbf{h}_{ii}\mathbf{w}_{i,p}\right|^2}{\sigma^2 + \left|\mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i}}\right|^2 + \left|\mathbf{h}_i\mathbf{w}_{\bar{i},c}\right|^2}, i = 1,2, \\ & \|\mathbf{D}_i\mathbf{w}_{i,c}\|^2 + \|\mathbf{D}_i\mathbf{w}_{\bar{i},c}\|^2 + \|\mathbf{w}_{i,p}\|^2 \leq P_i, i = 1,2. \end{aligned}$$

We can transform the above problem into an equivalent convex optimization problem.

- If $r_{i,p} \equiv 0$ or $r_i \equiv r_{i,p}$, we can reduce the problem as follows:
  - If $r_{i,p} \equiv 0$, the corresponding constraint becomes redundant, and $\mathbf{w}_{i,p} = 0$.
  - If $r_i \equiv r_{i,p}$, then $\mathbf{w}_{i,c} = \mathbf{0}$ at the optimum and we can remove the constraint corresponding to r_i.

  In both cases, the remaining constraint can be transformed into a second-order cone program [7], [8], [9].
- Otherwise, the problem is reformulated as follows. Consider the inequalities related to user i's rates. Imposing the decoding order to be common message, then private message , both inequalities must be met with equality at the optimum. Combining these two equations, we get:

  $$\frac{2^{r_{i,p}} - 1}{2^{r_i} - 2^{r_{i,p}}} \left|\mathbf{h}_i\mathbf{w}_{i,c}\right|^2 = \left|\mathbf{h}_{ii}\mathbf{w}_{i,p}\right|^2. \tag{9}$$

  Further noting that $\mathbf{h}_i\mathbf{w}_{i,c}$ and $\mathbf{h}_{ii}\mathbf{w}_{i,p}$ being real does not restrict the solution, we can transform our original problem into a convex optimization problem [7], [8], [9]:

$$\begin{aligned} \text{min. } & \sum_{i=1}^{2} \left[\|\mathbf{w}_{i,c}\|^2 + \|\mathbf{w}_{i,p}\|^2\right] \\ \text{s.t. } & \sqrt{2^{r_{i,p}} - 1}\left\|\begin{bmatrix} \sigma & \mathbf{h}_{i\bar{i}}\mathbf{w}_{\bar{i},p} & \mathbf{h}_i\mathbf{w}_{\bar{i},c} \end{bmatrix}\right\| \leq \mathbf{h}_{ii}\mathbf{w}_{i,p}, i = 1,2 \\ & \sqrt{\frac{2^{r_i} - 2^{r_{i,p}}}{2^{r_{i,p}} - 1}}\mathbf{h}_{ii}\mathbf{w}_{i,p} = \mathbf{h}_i\mathbf{w}_{i,c}, i = 1,2 \\ & \|\mathbf{D}_i\mathbf{w}_{i,c}\|^2 + \|\mathbf{D}_i\mathbf{w}_{\bar{i},c}\|^2 + \|\mathbf{w}_{i,p}\|^2 \leq P_i, i = 1,2 \end{aligned}$$

IV. Numerical Results

Figures 2 and 3 show the rate region for different values of the backhaul, for two different instances of the channels with $N_t = 1$. We let $C_1 = C_2 = C$, i.e. similar size backhaul links between the central processor/network and each of the

[1] Note that in our simulations, since not sharing messages yields a simpler beamforming scheme, we first check for the feasibility of rate tuple (r_1, r_1, r_2, r_2).

transmitters. For $N_t = 1$, the sum rate of the interference channel (IC) with SUD is known to be maximized by having the transmitters being either off or transmitting at full power. For the first channel instance shown, the maximum sum rate is achieved by transmitter 1 transmitting at full power and transmitter 2 being off, whereas in the second instance both transmitters transmit at full power. Moreover, in this second case, the IC rate region corresponds to a larger portion of the network MIMO region. When C is low, the same rate region is achieved in both cases. As it increases, the difference between the two setups becomes quite significant.

Finally, Figure 4 compares the maximum average sum rates achieved for $\alpha = .5$ ($r_1 = r_2$) and different channel statistics. Let $\mathbf{h}_{ij} \sim \mathcal{CN}(0, \sigma_{ij}^2 \mathbf{I}_{N_t})$, then the curves marked with x have $\sigma_{ij}^2 = 1$, for $i, j = 1, 2$, whereas those marked with $\triangle$ have $\sigma_{ii}^2 = 1$, and $\sigma_{i\bar{i}}^2 = .5, i = 1, 2$. Note that for lower $\sigma_{i\bar{i}}^2$, higher IC rates are achieved but lower network MIMO rates when the backhaul constraints are ignored. The situation is not as clear-cut when it is. The figure also shows how much of the rates achieved correspond to private messages alone.

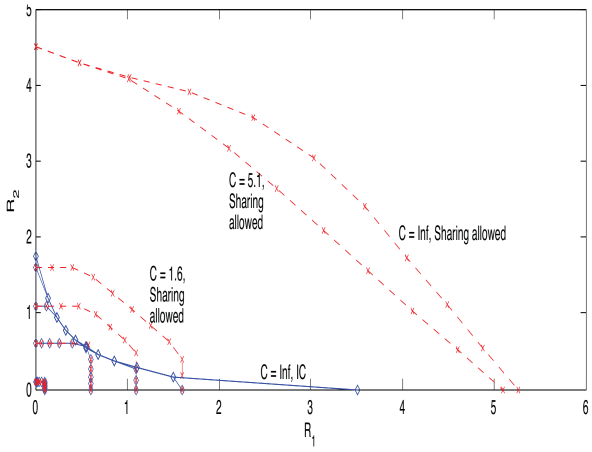


Fig. 2. Sample Rate Region, for $N_t = 1$, SNR = 10dB, and different backhaul rates $C_1 = C_2 = C$. 'x' denotes the scheme proposed, '$\Diamond$' the IC.

V. Conclusion

In this paper, we proposed to use the backhaul capacity to convey different types of messages: private messages transmitted from the serving base station, and common messages jointly transmitted from several base stations. A corresponding achievable rate region was characterized and simulations have shown that unless both interference and backhaul capacity are relatively low, the benefit of data sharing is quite significant.

Acknowledgment

This work was funded in part by french project ORMAC.

References

[1] D. Slepian and J. K. Wolf, *A coding theorem for multiple access channels with correlated sources*, Bell Syst. Tech. J., vol. 52, pp. 1037-1076, Sept. 1973.

[2] S.I. Bross, A. Lapidoth and M.A. Wigger, *The Gaussian MAC with conferencing encoders*, ISIT 2008.

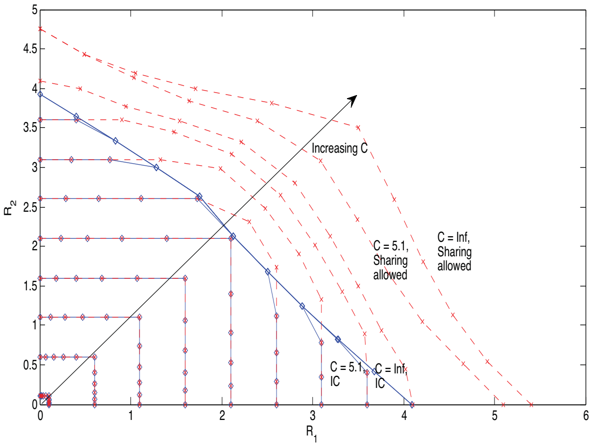


Fig. 3. Sample Rate Region, for $N_t = 1$, SNR = 10dB, and different backhaul rates $C_1 = C_2 = C$. 'x' denotes the scheme proposed, '$\Diamond$' the IC.

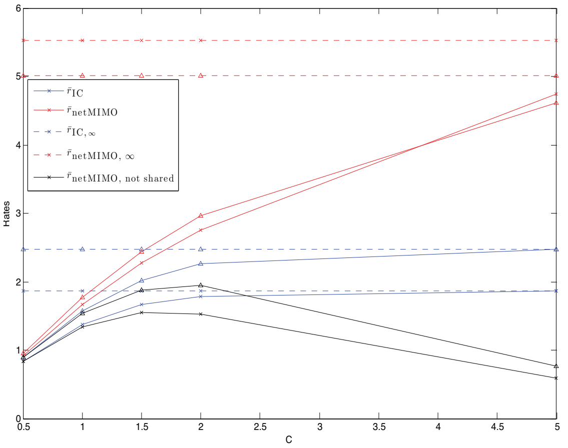


Fig. 4. Average Max Min Rates vs. Backhaul, for $N_t = 1$, SNR = 10dB,

[3] S. Shamai (Shitz), O. Simeone, O. Somekh and H.V. Poor, *Joint Multi-Cell Processing for Downlink Channels with Limited-Capacity Backhaul*, ITA 2008.

[4] O. Simeone, O. Somekh, H.V. Poor and S. Shamai (Shitz), *Downlink Multicell Processing with Limited-Backhaul Capacity*, EURASIP Journal on Advances in Signal Processing, Volume 2009, May 2009.

[5] P. Marsch and G. Fettweis, *A Framework for Optimizing the Downlink Performance of Distributed Antenna Systems under a Constrained Backhaul*, in Proceedings of the 13th European Wireless Conference (EW'07), Paris, France, April 2007.

[6] P. Marsch and G. Fettweis, *On Base Station Cooperation Schemes for Downlink Network MIMO under a Constrained Backhaul*, GLOBECOM 2008, Nov.-Dec. 2008.

[7] S.G. Vorobyov, A.B. Gershman and Z.-Q. Luo, *Robust Adaptive Beamforming Using Worst-Case Performance Optimization: A Solution to the Signal Mismatch Problem*, IEEE Transactions on Signal Processing, Vol. 51, No. 2, February 2003.

[8] A. Wiesel, Y.C. Eldar and S. Shamai, *Linear Precoding via Conic Optimization for Fixed MIMO Receivers*, IEEE Transactions on Signal Processing, Vol. 54, No. 1, January 2006.

[9] H. Dahrouj and W. Yu, *Coordinated beamforming for the multi-cell multi-antenna wireless system*, in Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS '08), pp. 429-434, Princeton, NJ, USA, March 2008.

Orthogonalization to Reduce Overhead in MIMO Interference Channels

Steven W. Peters and Robert W. Heath, Jr.
Wireless Networking and Communications Group
The University of Texas at Austin
1 University Station C0806, Austin, TX, 78712-0240
Email: {speters,rheath}@mail.utexas.edu

Abstract—**Interference channels are useful analytical models for distributed wireless communication networks with multiple simultaneously transmitting users. The degrees-of-freedom optimal transmit strategy for interference channels is interference alignment, which requires substantial channel state knowledge throughout the network. As the network grows, the sum capacity, theoretically, increases linearly. This result, however, neglects overhead from training and feedback. This paper accounts for overhead in the MIMO constant-coefficient interference channel with linear precoding and proposes an orthogonalization approach to maximizing sum throughput when overhead is considered. The optimization's solution, assuming each group uses interference alignment, is found to require full channel state information and a brute-force search, so a greedy partitioning method with reduced CSI requirements is proposed.**

I. Introduction

Interference channels are useful models for networks where non-causal sharing of data across multiple transmitters, such as for base station coordination, is infeasible. Such cases include ad hoc networks and cellular networks with low-bandwidth backhauls between base stations. Interference channels model the case of simultaneous point-to-point transmission by two or more transmitters such that the respective receivers observe the superposition of all transmissions in the network. The transmissions observed from transmitters not intentionally communicating with a given receiver are termed interference.

Recent work on interference channels has shown that, theoretically, the capacity of such networks increases linearly with the number of transmit/receive pairs in the network [1], [2]. In particular, by intelligently precoding the transmitted symbols, all the interference can be forced into a subspace of the received space at all receivers simultaneously. This precoding operation is termed interference alignment (IA). With two users, previous work has shown a loss in degrees of freedom when channel coefficients are not known at the transmitters [3], [4]. There is no prior work analyzing the interference channel without training for channel estimation at the receivers. All current methods for maximizing degrees of freedom for the interference channel require channel training and estimation at some node even if no feedback mechanism is employed. The number of total links grows with the square of the number of users in the network, meaning the overhead associated with training these links will outpace the capacity growth with many users. Similarly, the requirement of CSI, even if only at the receivers, is known to effectively reduce the degrees of freedom of a point-to-point block fading link [5]. Extending this model to the interference channel, overhead associated with training is expected to dominate an interference channel with many users, diminishing the promised capacity increase.

Prior work has considered the impact of imperfect CSI on the achievable sum rate of interference alignment [6], and the number of bits of limited feedback desired for single-antenna interference alignment [7]. Overhead due to training was neglected in both cases. Others have considered clustering a cellular network based on spatial proximity [8], but this clustering is done a priori and does not explicitly consider overhead. To our knowledge there is no prior work explicitly considering training overhead in MIMO interference channels.

This paper presents a model for analyzing overhead in MIMO interference channels and finds that the achieved sum rate with overhead of interference alignment will go to zero with a large number of users. We consider a fully connected interference channel, where spatial clustering is ineffective because of the proximity of all users. Thus, we propose to partition the users into orthogonally transmitting groups. The groups take turns transmitting, with interference alignment used as for transmission inside each group. Although such partitioning still results in an asymptotically zero sum rate, we show that for moderate number of users, partitioning can result in multiplicative gains in sum rate over applying IA to the entire network.

For the model outlined in this paper, the sum-rate-optimal partitioning is shown to be a highly complex optimization that requires global CSI and an exhaustive search over all possible partitioning combinations, calculating the IA precoders for each combination. We therefore propose a greedy algorithm that requires only channel quality information (CQI) on the link for each transmit/receive pair. Based on an approximation to the achievable sum rate for interference alignment using linear precoding, the proposed algorithm is shown to efficiently partition the network into IA groups.

The log refers to $\log_2$. Bold uppercase letters, such as $\mathbf{A}$, denote matrices, bold lowercase letters, such as $\mathbf{a}$, denote column vectors, and normal letters a denote scalars. The letter $\mathbb{E}$ denotes expectation, $\mathbb{C}$ is the complex field, $\max\{a, b\}$ denotes the maximum of a and b, $\|\mathbf{A}\|_F$ is the Frobenius

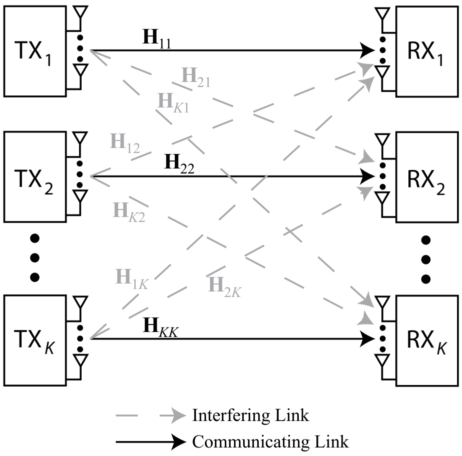


Fig. 1. The MIMO interference channel. Each transmitter is paired with a single receiver. In the model considered in this paper, the channels $\mathbf{H}_{k,\ell}$ are block fading with coherence time $T_{k,\ell}$.

norm of matrix $\mathbf{A}$, and $|\mathbf{A}|$ is the determinant of square matrix $\mathbf{A}$. The identity matrix of appropriate dimension is $\mathbf{I}$ and $[a]^+ = \max\{a, 0\}$.

II. System Model

We consider a distributed synchronized network with $2K$ nodes, each with M antennas. K of the nodes have data to transmit to the other K nodes, with no multiuser or cooperative transmission. In particular, transmitter $k \in \{1, \ldots, K\}$ has data destined only for receiver k. We assume a narrowband block fading model where the channel $\mathbf{H}_{k,\ell}$ between transmitter $\ell \in \{1, \ldots, K\}$ and receiver k is independently generated every T transmission periods $\forall k, \ell$. We assume transmissions are frame and frequency synchronous. Thus, at any fixed moment in time, we have a K-user MIMO interference channel with M antennas at each node, as illustrated in Figure 1. The assumption that all nodes have identical coherence times models the case where the nodes are fixed in relation to each other and a moving environment is causing time selectivity, for example, with fixed infrastructure near highways. Analysis for different coherence times for each link is left for future work.

Communication is divided into frames of period T symbols. At the beginning of each frame, the transmitters send mutually orthogonal training sequences to allow the receivers to estimate the channels. This training is necessary not only for coherent detection but also for CSI feedback required to exploit the full degrees of freedom in the network [3], [4]. Training plus feedback time is $\mathcal{L}(K, M) < T$ symbol periods such that in general we do not make assumptions about how many links must be estimated or how many symbols are required for estimation.

The data transmission portion of the frame begins after the first $\mathcal{L}(K, M)$ symbols and ends when the channel changes. Information theoretic results, which neglect overhead, suggest that all transmitters should send simultaneously to achieve the maximum degrees of freedom in the channel and thus approach its sum capacity with high transmit power [1], [2]. In particular, transmitter ℓ sends S_ℓ spatial streams to receiver ℓ. At symbol period n, the signal observed by receiver $k \in \{1, \ldots, K\}$ is

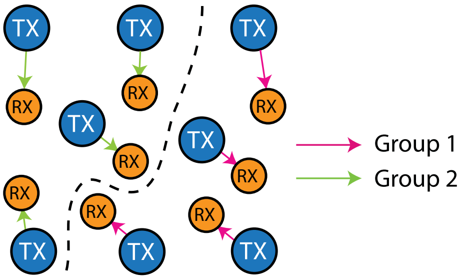


Fig. 2. Illustration of a partition of the K-user interference channel into two $K/2$-user interference channels transmitting orthogonally to each other.

$$\mathbf{y}_k[n] = \mathbf{H}_{k,k}\mathbf{F}_k\mathbf{s}_k[n] + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{H}_{k,\ell}\mathbf{F}_\ell\mathbf{s}_\ell[n] + \mathbf{v}_k[n], \tag{1}$$

where $\mathbf{F}_\ell$ is the $M \times S_\ell$ linear precoder used at transmitter ℓ, $\mathbf{s}_\ell$ is the $S_\ell \times 1$ vector of symbols sent by transmitter ℓ, and $\mathbf{v}_k$ is zero-mean white circularly symmetric zero-mean complex Gaussian noise with covariance matrix $\mathbb{E}\mathbf{v}_k\mathbf{v}_k^* = \mathbf{R}_k$. The sum rate of the network in bits per transmission for a frame is then

$$\begin{aligned} R_{\text{sum}} \;=\; & \left[\frac{T - \mathcal{L}(K, M)}{T}\right]^+ \sum_{k=1}^{K} \log \Bigg| \mathbf{I} + \Bigg(\mathbf{R}_k + \\ & \sum_{\ell \neq k}^{K} \mathbf{H}_{k,\ell}\mathbf{F}_\ell\mathbf{F}_\ell^*\mathbf{H}_{k,\ell}^*\Bigg)^{-1} \mathbf{H}_{k,k}\mathbf{F}_k\mathbf{F}_k^*\mathbf{H}_{k,k}\Bigg|. \end{aligned} \tag{2}$$

From (2) we observe that the overhead term $\mathcal{L}(K, M)$ effetively decreases the degrees of freedom in this network. Previous work has shown that at least M symbols are required for estimation of an $M \times M$ MIMO channel [5]. Although there are K^2 MIMO links, the K receivers can use the training from a given transmitter without any extra use of resources. Therefore, $\mathcal{L}(K, M) \geq KM$. Even assuming feedback requires no overhead, $R_{\text{sum}} = 0$ for $K \geq T/M$. In short, simultaneous transmissions requiring coherent CSIR, such as interference alignment, break down with large K.

To regain degrees of freedom for a given number of users, we propose to partition the users into groups that share the frame orthogonally in time or frequency. This concept is illustrated in Figure 2. Note that since the original K users were modeled as a connected interference channel, where all receivers observe a signal from all transmitters above the noise floor, any subset of transmit/receive pairs, in isolation, may also be modeled as a connected interference channel. If the users are partitioned into P index sets $\{\mathcal{K}_p\}$, with $|\mathcal{K}_p| = K_p$ users in the pth group, then the sum rate of the network becomes

$$\hat{R}_{\text{sum}} = \sum_{p=1}^{P} \sum_{k \in \mathcal{K}_p} \frac{T/P - \mathcal{L}(K_p, M)}{T/P} \log\left|\mathbf{I} + \left(\mathbf{R}_k + \sum_{\substack{\ell \in \mathcal{K}_p \\ \ell \neq k}} \mathbf{H}_{k,\ell}\mathbf{F}_\ell\mathbf{F}_\ell^*\mathbf{H}_{k,\ell}^*\right)^{-1} \mathbf{H}_{k,k}\mathbf{F}_k\mathbf{F}_k^*\mathbf{H}_{k,k}^*\right| \quad (3)$$

We then aim to solve the following optimization:

$$\begin{array}{rc} \text{maximize} & \hat{R}_{\text{sum}} \\ \text{with respect to} & P \in \mathbb{N}_1, K_p \in \mathbb{N}_1 \forall p, \mathbf{F}_\ell \in \mathbb{C}^{M \times S_\ell} \forall \ell \\ \text{subject to} & \sum_{p=1}^{P} K_p = K \\ & \|\mathbf{F}_\ell\| \leq 1. \end{array} \quad (4)$$

The solution to this optimization is computationally complex and involves not only a brute force search over every possible grouping, but also the calculation of the desired precoders for each grouping. Further, such an optimization requires full CSI at a central controller. In the next section we present a suboptimal greedy method for performing this grouping with only channel quality information (CQI).

III. Greedy Partitioning

To develop a greedy algorithm for partitioning the network, we must first define a *selection function* that assigns a value of placing a user in a group. This function would ideally be the sum rate increase of placing a user in a group. This is difficult in multiuser networks since the actual sum rate increase will depend on which future users are assigned to the group—knowledge that is unavailable in a greedy algorithm. Instead we resort to an approximation of this sum rate increase.

After partitioning the K-user interference channel into P orthogonal groups, group p will be a K_p-user interference channel that is restricted to utilizing only $1/P$ of the spectrum or coherence interval. Thus, interference alignment is a reasonable choice for precoder design in each group. Although interference alignment requires extensive CSI and calculation of precoders to find the exact sum rate, we note that the precoder solutions are independent from the direct links $\{\mathbf{H}_{k,k}\}, \forall k$. Thus, with interference alignment, the expected throughput will be approximately the rate obtained from randomly generating orthogonal precoders $\mathbf{Q}$ and combiners $\mathbf{\Phi}$ of correct rank drawn uniformly from the Grassmann manifold in the absence of interferers because of the lack of bias in direction through the channel realization in the algorithm. We then approximate the expected rate for user k in group p to be

$$\overline{R}_{k,p} \approx \mathbb{E}_{\mathbf{\Phi},\mathbf{Q}} \log\left|\mathbf{I} + \mathbf{\Phi}^*\mathbf{H}_{k,k}\mathbf{Q}\mathbf{Q}^*\mathbf{H}_{k,k}^*\mathbf{\Phi}\right| \quad (5)$$

$$\approx \frac{\tilde{d}(K_p, M)}{K_p} \log\left(1 + \frac{\|\mathbf{H}_{k,k}\|_F^2}{M^2}\right). \quad (6)$$

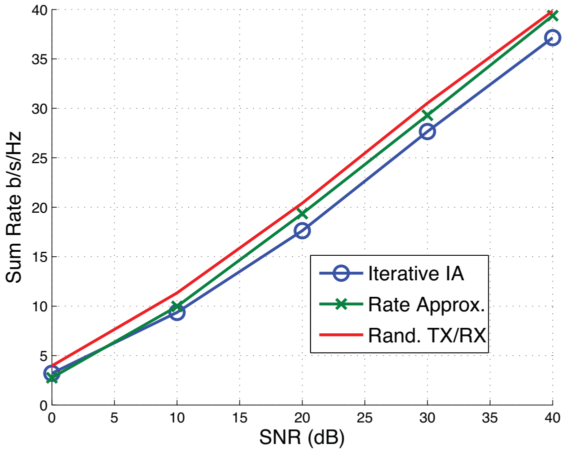


Fig. 3. Sum rate versus SNR of the approximation in (6).

```
1. Find $K_O$ according to (8)
2. $P = \text{round}(\frac{K}{K_O})$
3. Order users such that $\max_p \overline{R}_{1,p} > \max_p \overline{R}_{2,p} > \ldots$
4. Set $u = 1$
5. Let $p = \arg\max_p \overline{R}_{1,p}$
6. Add $u$ to the set $\mathcal{K}_p$
7. If $u < K$, increment $u$ and return to 4; else done
```

TABLE I
Greedy algorithm based on IA rate and group size approximations.

This approximation is justified via the plot in Figure 3. The difficulty with (6) is that the group size K_p is, in general, unknown at the time user k is being assigned. To remedy this, we define K_O, where

$$\tilde{d}(K_O, M, T) > \tilde{d}(K_O - 1, M, T) \quad (7)$$

$$\tilde{d}(K_O, M, T) > \tilde{d}(K_O + 1, M, T). \quad (8)$$

Here, $\tilde{d}(K, M, T)$ is the degrees of freedom with overhead and is defined as

$$\tilde{d}(K, M, T) = \frac{T - \mathcal{L}(\mathcal{K}, \mathcal{M})}{T} d(K, M). \quad (9)$$

Then we set $P = \text{round}(\frac{K}{K_O})$. Once P is found, we can assign users to each group by their approximate rate function $R_{k,p}$. The algorithm is summarized in Table I. Note that, this algorithm is based on a model with linear precoding, which does not likely result in a linear relationship between K and $d(K, M)$ [11]. This algorithm can work for non-linear precoding [2], which may increase the degrees of freedom in a constant-coefficient interference channel, with an appropriate approximation of $\overline{R}_{k,p}$. This problem is beyond the scope of this paper.

IV. Simulations

This section presents numerical results comparing the greedy partitioning method of Section III. The simulations are done using iterative interference alignment with linear precoding [9], [10] with 100 iterations, although the IA

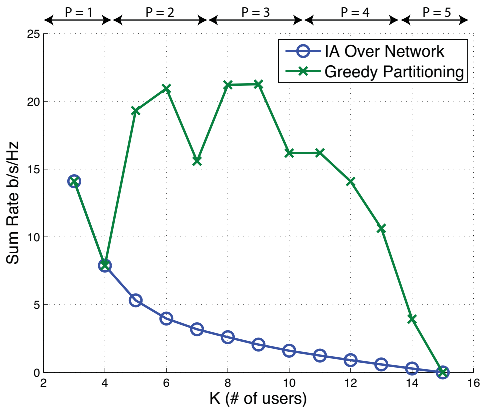


Fig. 4. Sum rate versus number of users K for IA applied to the entire network and the greedy partitioning approach proposed in this paper. In these simulations, $\mathcal{L}(K, M) = KM$ symbols are required for training for the K-user interference channel, which is the minimum for training K links. The coherence interval is $T = 30$ and $M = 2$ antennas per node.

precoders can be found with any IA solution. The degrees of freedom using this method has been conjectured to be $\tilde{d}(K, M) = 2MK/(K + 1)$ [11].

The first simulation gives the sum rate versus the number of users K for greedy partitioning and IA applied to the entire network with $\mathcal{L}(K, M) = KM$, which is the minimum amount of overhead required for training [5]. The coherence interval $T = 30$, and each node is equipped with $M = 2$ antennas. The plot is shown in Figure 4. The greedy partitioning sum rate does not have a monotonic relationship with K since each group cannot have exactly K_O users unless K/K_O is an integer. Nevertheless, the suboptimal partitioning's sum rate drastically outperforms applying IA to the entire network.

The second simulation, whose plot is illustrated in Figure 5, shows the sum rate performance of the greedy partitioning method and the exhaustive partitioning method for $K = 3$ users for various T. As with the previous simulation, $M = 2$ antennas are at each node. With a small coherence interval, the two perform very similarly as the greedy method partitions the network into 3 groups with one user that transmits interference-free. With larger T, the exhaustive search outperforms the greedy method because it still partitions the network into three interference-free groups. Partitioning into one group would result in non-zero interference with only 100 iterations of the iterative IA design, reducing the achievable sum rate. With perfect IA precoders and large T, the sum rate of the partitioning methods approaches the sum rate of IA without overhead.

V. Conclusions

We have demonstrated the importance of considering overhead associated with training and feedback in practical design for the interference channel. In particular, as the network grows, the sum rate with overhead of IA goes to zero. To increase sum rate with a finite number of users, we propose partitioning the network into orthogonally transmitting groups.

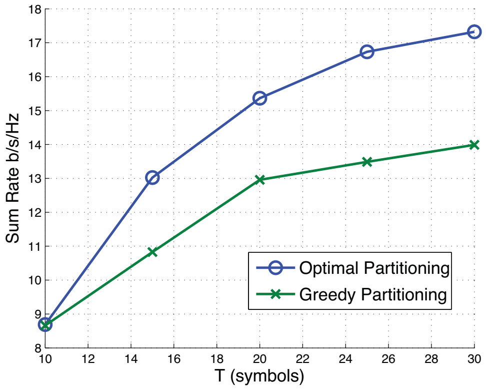


Fig. 5. Sum rate versus coherence interval T for exhaustive search and greedy partitioning. For this simulation, the users are kept at $K = 3$ and there are $M = 2$ antennas at each node, thus one stream is sent by each transmitter. As the coherence interval increases, the overhead percentage decreases and more time is allotted to transmitting data, thus sum rate increases to the rate of IA without overhead.

Although the optimum partition requires a complex brute-force search with global CSI, we propose a greedy algorithm that requires only direct-link CQI, and much of the gains of an exhausitve search can be made.

Acknowledgment

This work was supported by a gift from Huawei Technologies, Inc.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[2] A. Ghasemi, A. S. Motahari, and A. K. Khandani. (2009, September) Interference alignment for the K-user MIMO interference channel. [Online]. Available: http://arxiv.org/abs/0909.4604

[3] Y. Zhu and D. Guo. (2009) Isotropic MIMO interference channels without CSIT: The loss of degrees of freedom. [Online]. Available: http://arxiv.org/abs/0910.2961

[4] C. Huang, S. A. Jafar, S. Shamai (Shitz), and S. Vishwanath. (2009) On degrees of freedom region of MIMO networks without CSIT. [Online]. Available: http://arxiv.org/abs/0909.4017

[5] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[6] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *Proc. IEEE International Conference on Communications (ICC)*, Dresden, Germany, June 2009.

[7] J. Thukral and H. Bölcskei, "Interference alignment with limited feedback," in *IEEE Int. Symposium on Information Theory (ISIT)*, June 2009.

[8] R. Tresch and M. Guillaud, "Clustered interference alignment in large cellular networks," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communicatins (PIMRC)*, 2009.

[9] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 30–Dec. 4 2008, pp. 1–6.

[10] S. W. Peters and R. W. Heath, Jr., "Interference alignment via alternating minimization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009, pp. 2445–2448.

[11] C. M. Yetis, S. A. Jafar, and A. H. Kayran, "Feasibility conditions for interference alignment," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2009.

Beamforming in Interference Networks: Multicast, MISO IFC and Secrecy Capacity

Eduard A. Jorswieck[1]
Communications Theory, Communications Laboratory
Faculty of Electrical Engineering and Information Technology
D-01062 Dresden, Germany, Email: eduard.jorswieck@tu-berlin.de

***Abstract*—Motivated by the two-user beamforming in multi-antenna interference channels, we characterize the upper boundary of the achievable single-user gain-region. The eigenvector corresponding to the maximum eigenvalue of the weighted sum of Hermitian forms of channel vectors is shown to achieve all points on the boundary in some given direction. Thereby, we solve three different beamforming problems, namely the multicast beamforming problem, the beamforming optimization in MISO interference channels, and beamforming in MISO systems with secrecy constraints for arbitrary number of users. We are confident that the framework can be applied to beamforming problems in other interference networks as well. Numerical simulations illustrate the achievable gain-region.**

I. Introduction

Interference channels are one of the basic elements of complex networks. Future wireless communication systems will suffer from interference since the number of subscribers as well as the required data rate increases. Therefore, it is important to exploit carefully the spatial dimension by using multiple transmit or receive antennas. In the current work, we focus on a generic K-user multiple-input single-output (MISO) interference channel [1]. Information-theoretic studies of the IFC have a long history [2], [3], [4], [5]. These references have provided various achievable rate regions, which are generally larger in the more recent papers than in the earlier ones. However, the capacity region of the general IFC remains an open problem. For certain limiting cases, for example when the interference is weak or very strong, respectively, the sum capacity is known [6]. If the interference is weak, it can simply be treated as additional noise. For very strong interference, successive interference cancellation (SIC) can be applied at one or more of the receivers. Multiple antenna interference channels are studied in [1]. Multiple-input multiple-output (MIMO) interference channels have also recently been studied in [7], from the perspective of spatial multiplexing gains. In [8], the rate region of the single-input single-output (SISO) IFC was characterized in terms of convexity and concavity.

The linear combination of the egoistic and altruistic beamformers is proved to be Pareto optimal in the 2-user MISO interference channel [9]. In [10], this idea is extended to the MIMO interference channel. Their proposed egoism and altruism balancing beamforming algorithm has connections with some important works such as rate optimization [11], [12] and interference alignment [13], [14]. In [15], the term coordinated beamforming is coined, and the optimal transmit beamforming and receive combining vectors under a zero inter-user interference constraint are derived for a two-user interference system in the context of two-cell coordination. Using ideas from game theory, the multi-antenna interference channel is studied in [16], [17], [18].

The contribution and outline of the paper is as follows:

1) In Section II, we define the MISO single-user gain-region and show that it is convex.
2) Then, the main result is a characterization of the boundary of the single-user gain-region in a given direction $\boldsymbol{e}$ which follows from the convexity of the gain-region.
3) In Section III, the result is applied in order to completely characterize the Pareto boundary of
   a) the achievable rate region of multi-cast transmission (e.g., broadcast phase of two-way relaying),
   b) the achievable rate region of the MISO interference channel with K users,
   c) and the achievable secrecy and eavesdropping rates in MISO wiretap channels.

The characterization of the Pareto boundary for the MISO interference channel with K users improves a former result in [9]. The characterization of the Pareto boundary of the achievable secrecy and equivocation rates contains the optimum beamforming derived in [19]. The theoretical results are illustrated by a numerical simulation and the paper is concluded in Section IV.

II. Boundary of the single-user gain-region

A. Preliminaries

Consider a multiple-antenna user k in a K-user interference network and denote the flat-fading vector channels from user k to single-antenna receiver $\ell, 1 \leq \ell \leq K$ as $\boldsymbol{h}_{k,\ell}$. Define the channel gain as a function of the beamforming vector $\boldsymbol{w}$ as $x_\ell(\boldsymbol{w}) = \left\| \boldsymbol{h}_{k,\ell}^H \boldsymbol{w} \right\|^2$ for $1 \leq \ell \leq K$. Define the achievable gain-region for user k as

$$\Omega_k = \bigcup_{\|\boldsymbol{w}\|=1} (x_1(\boldsymbol{w}), ..., x_K(\boldsymbol{w})) . \tag{1}$$

[1] Part of this work is funded by the European Union within the SAPHYRE (Sharing Physical Resources – Mechanisms and Implementations for Wireless Networks) project (FP7-ICT-248001): http://saphyre.eu. The author would like to acknowledge the contributions of his colleagues.

The operational meaning of the gain-region Ω_k will be discussed in Section III. Before we illustrate the region and its boundary, we note the following important property.

Lemma 1: The gain region Ω_k is always convex, i.e., for $\boldsymbol{x}, \boldsymbol{y} \in \Omega_k$ it follows that $\boldsymbol{x}(t) = t\boldsymbol{x} + (1-t)\boldsymbol{y} \in \Omega_k$ for $t, 0 \leq t \leq 1$. □

The proof of this lemma follows from direct computation of $\boldsymbol{w}(t) = t\boldsymbol{w}_x + (1-t)\boldsymbol{w}_y$ and $\boldsymbol{x}(\boldsymbol{w})$ where $\boldsymbol{w}_x$ achieves $\boldsymbol{x}$ and $\boldsymbol{w}_y$ achieves $\boldsymbol{y}$. The complete proof can be found in [20].

In Figure 1, an example gain-region is shown for two users $K = 2$. The operational meaning of the gain-region and the directions $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ in Figure 1 will be discussed in Section III. The arrows in Figure 1 correspond to interesting directions $\boldsymbol{e}_1 = [1, 1], \boldsymbol{e}_2 = [1, -2], \boldsymbol{e}_3 = [-1, 1]$.

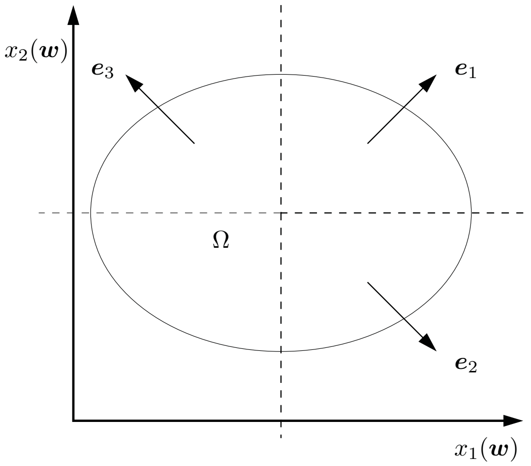


Fig. 1. Example of two-dimensional gain-region and its upper boundaries in directions $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ are illustrated.

B. Main result

Following the discussion on the boundary of Ω_k, we formalize the upper boundary following the definitions in [21]. There, this definition was used to derive the solution of a monotonic optimization problem [22].

Definition 1: A point $\boldsymbol{y} \in \mathbb{R}_+^n$ is called *upper boundary point* of a convex set $\mathcal{C}$ if $\boldsymbol{y} \in \mathcal{C}$ while the set

$$\mathcal{K}_{\boldsymbol{y}} = \boldsymbol{y} + \mathbb{R}_{++}^n = \{\boldsymbol{y}' \in \mathbb{R}_+^n | \boldsymbol{y}' > \boldsymbol{y}\} \subset \mathbb{R}_+^n \setminus \mathcal{C}. \tag{2}$$

The set of upper boundary points of $\mathcal{C}$ is called the upper boundary of $\mathcal{C}$ and it is denoted by $\partial^+\mathcal{C}$. □

The straightforward extension to include also the right boundary of a convex set $\mathcal{C}$ is to define the upper boundary of $\mathcal{C}$ in direction $\boldsymbol{e}$.

Definition 2: A point $\boldsymbol{y} \in \mathbb{R}_+^n$ is called upper boundary point of a convex set $\mathcal{C}$ in direction $\boldsymbol{e}$ if $\boldsymbol{y} \in \mathcal{C}$ while the set

$$\mathcal{K}_{\boldsymbol{y}}(\boldsymbol{e}) = \{\boldsymbol{y}' \in \mathbb{R}_+^n | y'_\ell e_\ell \geq y_\ell e_\ell \; \forall \; 1 \leq \ell \leq n\} \subset \mathbb{R}_+^n \setminus \mathcal{C} \tag{3}$$

where the inequality has at least one strict inequality and directional vector $\boldsymbol{e} \in \{-1, +1\}^n$. We denote the set of upper boundary points in direction $\boldsymbol{e}$ as $\partial^{\boldsymbol{e}}\mathcal{C}$. □

For the choice $\boldsymbol{e} = \boldsymbol{1}$ the upper boundary in direction $\boldsymbol{e}$ is the usual upper boundary, i.e., $\partial^+\mathcal{C} = \partial^{\boldsymbol{1}}\mathcal{C}$.

In the following, we omit the index k when considering only one user for convenience. For efficient operation the boundary points of Ω in all directions (except $\boldsymbol{e} = -\boldsymbol{1}$) are of interest. Define the set $\mathcal{E} = \{-1, 1\}^n \setminus \{-1\}^n$. The following result is the main theorem of the paper. Interestingly, it follows easily from the convexity of the gain-region.

Theorem 1: All upper boundary points of the convex set Ω in direction $\boldsymbol{e} \in \mathcal{E}$ can be achieved by

$$\boldsymbol{w}(\boldsymbol{\lambda}) = \boldsymbol{v}_{max}\left(\sum_{\ell=1}^{K} \lambda_\ell e_\ell \boldsymbol{h}_{k,\ell}\boldsymbol{h}_{k,\ell}^H\right) \tag{4}$$

with $\boldsymbol{v}_{\max}(\boldsymbol{Z})$ denoting the eigenvector which belongs to the maximum eigenvalue of the Hermitian matrix $\boldsymbol{Z}$, $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_{K-1}, \lambda_1, ..., \lambda_{K-1}$ with $0 \leq \lambda_\ell \leq 1$, $1 \leq \ell \leq K-1$ and $\lambda_K = 1 - \sum_{\ell=1}^{K-1} \lambda_\ell$. □

Proof: We provide the sketch of the proof. The complete proof can be found in [20]. The boundary points in direction $\boldsymbol{e}$ of the convex set Ω can be achieved by maximization of the weighted sum gain, i.e.,

$$\max_{\boldsymbol{w}:||\boldsymbol{w}||^2=1} \sum_{\ell=1}^{K} \lambda_\ell e_\ell |\boldsymbol{w}^H \boldsymbol{h}_{k,\ell}|^2. \tag{5}$$

The objective function in (5) can be rewritten as

$$\begin{aligned} y(\boldsymbol{w}) &= \sum_{\ell=1}^{K} \lambda_\ell e_\ell |\boldsymbol{w}^H \boldsymbol{h}_{k,\ell}|^2 \\ &= \boldsymbol{w}^H \left(\underbrace{\sum_{\ell=1}^{K} \lambda_\ell e_\ell \boldsymbol{h}_{k,\ell}\boldsymbol{h}_{k,\ell}^H}_{\boldsymbol{Z}}\right) \boldsymbol{w}. \end{aligned} \tag{6}$$

Note that the matrix $\boldsymbol{Z}$ in (6) is not necessarily positive semidefinite because the directional vector $\boldsymbol{e}$ can contain negative components. However, it is Hermitian and therefore, the solution to (5) is the eigenvector which corresponds to the maximum eigenvalue of $\boldsymbol{Z}$. ■

The interesting observation from Theorem 1 is that all upper boundary points of the K-dimensional gain-region can be achieved by a parameterization using $K-1$ real parameters between zero and one, i.e.,

$$\boldsymbol{\lambda} \in \boldsymbol{\Lambda} = \{\boldsymbol{\lambda} \in [0, 1]^K : \sum_{\ell=1}^{K} \lambda_\ell = 1\}. \tag{7}$$

Depending on the application context different directions or even certain operating points are to be optimized. The gain-region and its boundary are illustrated in Figure 2.

The four colours in Figure 2 correspond to the four directions $\boldsymbol{e}_1 = [1, 1, 1]$, $\boldsymbol{e}_2 = [1, -1, 1]$, $\boldsymbol{e}_3 = [1, 1, -1]$, and $\boldsymbol{e}_4 = [1, -1, -1]$. Inside the three nets (constructed by varying the parameter vector $\boldsymbol{\lambda}$ on a grid with 100×100 points) there is the convex hull of 100.000 gain points achieved by random generated beamforming vectors. The channels $\boldsymbol{h}_{11}, \boldsymbol{h}_{12}, \boldsymbol{h}_{13}$ are randomly generated with three transmit antennas.

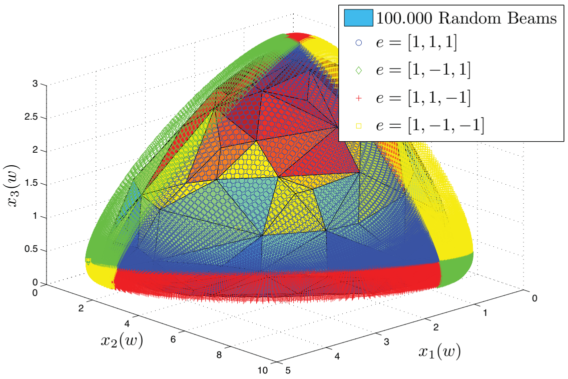


Fig. 2. Example of three-dimensional gain-region and its upper boundary in four directions are illustrated. The upper boundary is computed by the parameterization in Theorem 1. The four colors correspond to the four directions.

III. Applications

The result from Theorem 1 can be applied to an interference network in which (virtual) users are equipped with multiple antennas. We present three representatives for different applications. Needless to say that it can be applied to other scenarios as well.

A. *Multicast beamforming*

We start with a trivial example and consider the simple multicast beamforming scenario in which one transmitter sends common information to K receivers. The motivation for this illustrative scenario could be the second phase of a two-way relaying system with decode-and-forward relaying [23]. In the broadcast phase, one transmitter with multiple antennas transmit the data to the terminals which then subtract the self-interference (analog network coding). Denote the channels from the relay to the terminals by $\boldsymbol{h}_1, ..., \boldsymbol{h}_K$. In this simple scenario, the achievable rate of user terminal k is given by

$$R_k(\boldsymbol{w}) = \log\left(1 + \frac{|\boldsymbol{w}^H \boldsymbol{h}_k|^2}{\sigma^2}\right). \tag{8}$$

The multicast beamforming rate region $\mathcal{R}$ is defined as

$$\mathcal{R} = \bigcup_{\|\boldsymbol{w}\|=1} (R_1(\boldsymbol{w}), ..., R_K(\boldsymbol{w})). \tag{9}$$

The next corollary follows from Theorem 1 since the upper boundary of $\mathcal{R}$ corresponds exactly to the upper boundary of Ω in direction $\boldsymbol{e} = \boldsymbol{1}$.

Corollary 1: Any point on the upper boundary of the rate region $\mathcal{R}$ in (9) can be achieved by

$$\boldsymbol{w}(\boldsymbol{\lambda}) = \boldsymbol{v}_{max}\left(\sum_{\ell=1}^{K} \lambda_\ell \boldsymbol{h}_{k,\ell} \boldsymbol{h}_{k,\ell}^H\right) \tag{10}$$

with $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ in (7). □

B. *MISO interference channel with $K \geq 2$ users*

The MISO interference channel with K users is studied in [9]. All base stations BS$_k$ have N transmit antennas each, that can be used with full phase coherency. The mobiles MS$_k$, however, have a single receive antenna each. We shall assume that transmission consists of scalar coding followed by beamforming, and that all propagation channels are frequency-flat. This leads to the following basic model for the matched-filtered, symbol-sampled complex baseband data received at MS$_k$:

$$y_k = \boldsymbol{h}_{kk}^T \boldsymbol{w}_k s_k + \sum_{l=1, l\neq k}^{K} \boldsymbol{h}_{lk}^T \boldsymbol{w}_l s_l + e_k, \tag{11}$$

where s_l, $1 \leq l \leq K$ is the symbol transmitted by BS$_l$, $\boldsymbol{h}_{ij}$ is the (complex-valued) $N \times 1$ channel-vector between BS$_i$ and MS$_j$, and $\boldsymbol{w}_l$ is the beamforming vector used by BS$_l$. The variables e_k are noise terms which we model as i.i.d. complex Gaussian with zero mean and variance σ^2.

We assume that each base station can use the transmit power P, but that power cannot be traded between the base stations. Without loss of generality, we shall take $P = 1$. This gives the power constraints

$$||\boldsymbol{w}_k||^2 \leq 1, \quad 1 \leq k \leq K \tag{12}$$

Throughout, we define the SNR as $1/\sigma^2$. The precoding scheme that we will discuss requires that the transmitters (BS$_k$) have access to channel state information (CSI) for some of the links. However, at no point we will require phase coherency between the base stations. In [9], a characterization of the beamforming vectors that reach the Pareto boundary of the achievable rate region with interference treated as additive Gaussian noise is provided by a complex linear combination.

In what follows we will assume that all receivers treat co-channel interference as noise, i.e. they make no attempt to decode and subtract the interference. For a given set of beamforming vectors $\{\boldsymbol{w}_1, ..., \boldsymbol{w}_K\}$, the following rate is then achievable for the link BS$_k$ →MS$_k$, by using codebooks approaching Gaussian ones:

$$R_k(\boldsymbol{w}_1, ..., \boldsymbol{w}_K) = \log_2\left(1 + \frac{|\boldsymbol{w}_k^T \boldsymbol{h}_{kk}|^2}{\sum_{l\neq k} |\boldsymbol{w}_l^T \boldsymbol{h}_{lk}|^2 + \sigma^2}\right). \tag{13}$$

We define the *achievable rate region* to be the set of all rates that can be achieved using beamforming vectors that satisfy the power constraint:

$$\begin{aligned}\mathcal{R} \triangleq \bigcup_{\{\boldsymbol{w}_k : ||\boldsymbol{w}_k||^2 \leq 1, 1 \leq k \leq K\}} & \{R_1(\boldsymbol{w}_1, ..., \boldsymbol{w}_K), ..., \\ & ..., R_K(\boldsymbol{w}_1, ..., \boldsymbol{w}_K)\} \subset \mathbb{R}_+^K. \end{aligned} \tag{14}$$

The outer boundary of this region is called the *Pareto boundary*, because it consists of operating points $(R_1, ..., R_K)$ for which it is impossible to improve one of the rates, without simultaneously decreasing at least one of the other rates. More precisely we define the *Pareto optimality* of an operating point as follows.

Definition 3: A rate tuple $(R_1, ..., R_K)$ is Pareto optimal if there is no other tuple $(Q_1, ..., Q_K)$ with $(Q_1, ..., Q_K) \geq (R_1, ..., R_K)$ and $(Q_1, ..., Q_K) \neq (R_1, ..., R_K)$ (the inequality is component-wise). □

Theorem 2: All points of the Pareto boundary of the achievable rate region of the MISO interference channel can be reached by beamforming vectors

$$\boldsymbol{w}_k(\boldsymbol{\lambda}_k) = \boldsymbol{v}_{max}\left(\sum_{\ell=1}^{K} \lambda_{k,\ell} e_\ell \boldsymbol{h}_{k,\ell}\boldsymbol{h}_{k,\ell}^H\right) \tag{15}$$

with $\boldsymbol{\lambda}_k \in \boldsymbol{\Lambda}$ defined in (7) and

$$e_\ell = \begin{cases} +1 & \ell = k \\ -1 & \text{otherwise} \end{cases}.$$

□

Note that for two users $K = 2$, the characterization in [9, Corollary 1] follows as a special case.

The proof of Theorem 2 follows from the observation that the Pareto boundary of the achievable rate region $\mathcal{R}$ in (14) corresponds for user k with the upper boundary of Ω_k in direction of $\boldsymbol{e}_k = [-1, ..., -1, 1, -1, ..., -1]$ with a 1 at the k-th position. The complete proof is provided in [20].

C. Secrecy capacity in MISO systems

As a brief third example, consider the scenario where the transmitter called Alice has multiple antennas n_T to send a confidential message to the legitimate receiver called Bob with a single antenna while the eavesdropper Eve with single antenna overhears the message. This is the MISO wiretap channel for which the secrecy capacity for perfect information at Alice is computed in [19]. Denote the channel from Alice to Bob by $\boldsymbol{h}_A$ and the channel from Alice to Eve by $\boldsymbol{h}_E$. The secrecy rate achievable with beamforming vector $\boldsymbol{w}$ is given by

$$R_s(\boldsymbol{w}) = \log\left(1 + \frac{|\boldsymbol{w}^H\boldsymbol{h}_A|^2}{\sigma^2}\right) - \log\left(1 + \frac{|\boldsymbol{w}^H\boldsymbol{h}_E|^2}{\sigma^2}\right) \tag{16}$$

By application of Theorem 1, the secrecy rate maximization is simply obtained as the solution in [19].

IV. Conclusions

The characterization of the Pareto boundary of the achievable rate regions in interference channels is a necessary prerequisite in order to develop efficient resource allocation strategies. Motivated by the simple characterization of the rate region of the two-user MISO interference channel, this paper develops a general theory for beamforming in interference networks. The idea is to study the problem for one terminal separately based on its gain-region. Since only operating points on the boundary are of interest, we characterize the beamforming vectors which achieve boundary points in a given direction in Theorem 1. Thus it is possible to obtain operating points which maximize the gain in one direction and minimize it in another direction. Finally, we apply the characterization to three representative scenarios: the multicast beamforming, the MISO interference channel rate region, and the secrecy capacity in MISO system. Currently, we study the extension to MIMO interference channels.

References

[1] S. Vishwanath and S. A. Jafar, "On the capacity of vector Gaussian interference channels," *IEEE ITW*, 2004.

[2] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, 1974.

[3] A. B. Carleial, "Interference channels," *IEEE Trans. on Inf. Theory*, vol. 24, pp. 60–70, 1978.

[4] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. on Information Theory*, vol. 27, pp. 49–60, 1981.

[5] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. on Inf. Theory*, vol. 31, pp. 607–615, 1985.

[6] X. Shang, B. Chen, and M. J. Gans, "On the achievable sum rate for mimo interference channels," *IEEE Trans. on Inf. Theory*, vol. 52, pp. 4313–4320, 2006.

[7] S. A. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. on Inf.ormation Theory*, vol. 53, pp. 2637–2642, 2007.

[8] M. Charafeddine, A. Sezgin, and A. Paulraj, "Rate region frontiers for n-user interference channel with interference as noise," *Proc. Allerton*, 2007.

[9] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[10] K. M. Ho, M. Kaynia, and D. Gesbert, "Distributed power control and beamforming on MIMO interference channels," in *Proc. European Wireless, invited paper*, 2010.

[11] S. Ye and R. S. Blum, "Optimized signaling for mimo interference systems with feedback," *IEEE Trans. on Signal Processing*, vol. 51, pp. 2839–2848, November 2003.

[12] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Distributed resource allocation schemes," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 53–63, Sept. 2009.

[13] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3457 – 3470, Aug. 2008.

[14] K. S. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," *preprint*, 2008. [Online]. Available: http://arxiv.org/abs/0803.3816

[15] C.-B. Chae, D. Mazzarese, N. Jindal, and R. Heath, "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1505–1515, Oct. 2008.

[16] E. Larsson and E. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE Journal on selected areas in Communications*, vol. 26, pp. 1059–1069, September 2008.

[17] M. Nokleby and A. L. Swindlehurst, "Bargaining and the MISO interference channel," *EURASIP Journal on Advances in Signal Processing*, vol. ID 368547, p. 13 pages, 2009.

[18] G. Scutari, D. Palomar, and S. Barbarossa, "Competitive design of multiuser mimo systems based on game theory: A unified view," vol. 26, no. 7, pp. 1089–1103, Sep. 2008.

[19] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. IEEE ISIT*, 2007.

[20] E. A. Jorswieck, "Optimal beamforming in interference networks," *IEEE Trans. on Signal Processing, in preparation*, 2010.

[21] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. on Communications, submitted*, 2009.

[22] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM Journal on Optimization*, vol. 11, pp. 464–494, 2000.

[23] T. J. Oechtering, E. A. Jorswieck, R. F. Wyrembelski, and H. Boche, "On the optimal transmit strategy for the MIMO bidirectional broadcast channel," *IEEE Trans. on Wireless Communication*, vol. 57, no. 12, pp. 3817–3826, Dec. 2009.

Optimal Interference Management In Multi-Antenna, Multi-Cell Systems

Johannes Brehmer, Wolfgang Utschick
Associate Institute for Signal Processing
Technische Universität München
{brehmer, utschick}@tum.de

***Abstract*— Intercell interference is a major limiting factor in wireless multi-cell networks. Recently, it has been shown that significant performance gains can be achieved by cooperation between base stations. Different degrees of cooperation are possible: From full cooperation, where multiple base stations form a virtual antenna array, to weak cooperation, where base stations take into consideration the interference caused to users in neighboring cells. In this work, weak cooperation in the form of interference management is investigated. The base stations are equipped with multiple antennas, while the mobile terminals only have a single antenna. Due to the spatial degrees of freedom, a base station can serve multiple users in the same slot. Each base station performs beamforming, user group selection, and scheduling, while the terminals treat interference as noise. The corresponding resource allocation problem is cast as a utility maximization problem, which includes common performance objectives such as sum-throughput, max-min fairness, and proportional fairness. Due to interference, the resulting utility maximization problem is a nonconvex optimization problem. Still, after a suitable reformulation, the problem can be solved to global optimality using the framework of monotonic optimization. In other words, we provide a framework for computing the jointly optimal beamforming, user selection, and scheduling strategy for each base station, under an arbitrary utility objective.**

I. Introduction

The interference management problem in the downlink of a cellular system is considered. The cellular system consists of multiple base stations (BS) and a set of mobile stations (MS). Moreover, the system is partitioned into cells, where each cell consists of a base station and a subset of the mobile stations. The base stations are assumed to have multiple antennas. In the downlink, the base stations transmit independent information to the mobile stations. In a conventional cellular system, each base station transmits to the mobile stations within its cell, without taking into consideration the interference caused in neighboring cells. As a result, the performance of the cellular downlink is limited by intercell interference.

Cooperation between base stations can help mitigate intercell interference and thereby improve system performance. Different degrees of cooperation are possible. Maximal performance is achieved by coordinated transmission [1]. In coordinated transmission, the base stations are connected by a high-speed backbone link, enabling them to act as a single transmitter, meaning that the antennas of all base stations form a single antenna array, and the signals of all users are jointly encoded across all base stations [1]. The coordinated transmission scheme requires that the data signals and channel state information for all users are available at each base station. Moreover, in order to enable coherent reception, each mobile station needs to be synchronized with all base stations.

In this work, a weaker form of cooperation between base stations is considered. As in a conventional system, base stations act as separate transmitters, meaning that the data signals of one user are only available at one of the base stations. Moreover, each mobile station is only synchronized with one base station. Interference from signals intended for other users is treated as noise. Due to the availability of multiple transmit antennas, base stations can choose transmit covariance matrices for transmission to their associated mobile stations. In the following, a choice of transmit covariance matrices for all base stations is denoted as a transmit strategy. Evidently, system performance can be improved if the choice of a transmit strategy is coordinated among base stations, taking into account intercell interference.

Each choice of a transmit strategy yields a certain system performance. Different models for the map from transmit strategy to system performance are possible. In this work, a generic utility model is used. Utility-based models have seen wide application in resource allocation for wireless networks, see, e.g., [2]. By allowing the base stations to switch between transmit strategies during one transmit interval, a further improvement of system performance is possible. Such a switching between strategies can be interpreted as *scheduling*.

Finding the optimal transmit strategies (with or without scheduling) in a coordinated manner represents a utility maximization problem. The presence of interference generally results in a nonconvex optimization problem. There exist resource allocation problems in the multi-cell downlink that can be reformulated as convex problems, such as the minimization of total transmit power under target rate constraints [3]. For the utility maximization problem considered in this work, however, it is generally not possible to find a convex reformulation. As a result, standard tools from convex optimization cannot be applied to find the optimal transmit strategies. Based on a framework proposed in [4], this work uses methods from deterministic global optimization to compute the optimal transmit strategies.

In the case that each base station serves only one mobile station, our system setup corresponds to a multiple-input, single-output interference channel (MISO IFC) with single-user decoding. Recently, a number of works have explored the properties of the MISO IFC under single-user decoding [5], [6], [7]. For the two-user MISO IFC without scheduling, a method to find the optimal transmit strategies for a given utility model is proposed in [8].

Notation: Lowercase bold letters and uppercase bold letters denote vectors and matrices, respectively. The trace of a square matrix $\boldsymbol{Q}$ is $\operatorname{tr}(\boldsymbol{Q})$. We write $\boldsymbol{Q} \succeq \boldsymbol{0}$ to say that a Hermitian matrix $\boldsymbol{Q}$ is positive semidefinite. The symbol $\mathbb{R}_+$ denotes the set of nonnegative real numbers. Order relations $\geq$ and $\leq$ are defined component-wise. A subset $\mathcal{R}$ of $\mathbb{R}_+^K$ is comprehensive if $\boldsymbol{s} \in \mathcal{R}$ and $\boldsymbol{0} \leq \boldsymbol{s}' \leq \boldsymbol{s}$ implies $\boldsymbol{s}' \in \mathcal{R}$. A function u is increasing if $\boldsymbol{s}' \leq \boldsymbol{s}$ implies $u(\boldsymbol{s}') \leq u(\boldsymbol{s})$, provided both $\boldsymbol{s}$ and $\boldsymbol{s}'$ are in the domain of u.

II. System Model

Downlink transmission in a cellular network is considered. The network consists of B multi-antenna base stations and K single-antenna mobile stations. Base station b is equipped with M transmit antennas and sends independent information to each of its associated MS, where the set of associated MS is denoted by $\mathcal{K}_b \subset \{1, \ldots, K\}$. Each MS is associated with one BS, i.e., $\mathcal{K}_b \cap \mathcal{K}_c = \emptyset$ if $b \neq c$ and

$$\bigcup_{b=1}^{B} \mathcal{K}_b = \{1, \ldots, K\}.$$

Let $\boldsymbol{x}_k$ denote the signal transmitted to MS k by the associated BS. The signal transmitted by base station b is the superposition of the signals transmitted to each of its associated MS. Accordingly, the received signal at MS k is given by

$$y_k = \sum_{q=1}^{K} \boldsymbol{h}_{q,k}^{\mathrm{H}} \boldsymbol{x}_q + \eta_k,$$

where $\boldsymbol{h}_{q,k}^{\mathrm{H}} \in \mathbb{C}^{1\times N}$ is the channel from the base station associated with MS q to MS k, and η_k is circularly symmetric AWGN with zero mean and variance σ^2.

Each BS encodes information separately for each of its associated MS using Gaussian codebooks. Each MS receives independent information. Accordingly, the signal $\boldsymbol{x}_k$ sent to MS k is independent of the signals to all other MS. Furthermore, it is assumed that each transmit signal $\boldsymbol{x}_k$ is a circularly symmetric Gaussian random variable with zero mean and covariance matrix $\boldsymbol{Q}_k \in \mathbb{C}^{M\times M}$. Finally, all MS treat interference as noise. A transmit strategy $\boldsymbol{Q}$ is a K-tuple of transmit covariance matrices, one for each MS:

$$\boldsymbol{Q} = (\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K).$$

For each transmit strategy $\boldsymbol{Q}$, an achievable rate of MS k is given by

$$r_k(\boldsymbol{Q}) = \log_2\left(1 + \frac{\boldsymbol{h}_{k,k}^{\mathrm{H}} \boldsymbol{Q}_k \boldsymbol{h}_{k,k}}{\sigma^2 + \sum_{q\neq k} \boldsymbol{h}_{q,k}^{\mathrm{H}} \boldsymbol{Q}_q \boldsymbol{h}_{q,k}}\right)$$

The transmitted signal from each BS is subject to a transmit power constraint,

$$\sum_{k\in\mathcal{K}_b} \operatorname{tr}(\boldsymbol{Q}_k) \leq P_b,\ b = 1, \ldots, B.$$

Accordingly, the set of feasible transmit strategies is given by

$$\mathcal{Q} = \left\{ \boldsymbol{Q} : \boldsymbol{Q}_k \succeq \boldsymbol{0}, \forall k, \sum_{k\in\mathcal{K}_b} \operatorname{tr}(\boldsymbol{Q}_k) \leq P_b, \forall b \right\}.$$

A rate region $\mathcal{R}$ is defined as the set of rate tuples achievable by a feasible choice of $\boldsymbol{Q}$,

$$\mathcal{R} = \{\boldsymbol{r}(\boldsymbol{Q}) : \boldsymbol{Q} \in \mathcal{Q}\}.$$

The rate region $\mathcal{R}$ is compact and comprehensive. In general, however, the rate region $\mathcal{R}$ is not convex. A convex rate region $\mathcal{C}$ is obtained by taking the convex hull of $\mathcal{R}$. Due to the fact that $\mathcal{R}$ is a comprehensive set, each point in $\mathcal{C}$ can be written as the convex combination of at most K points in $\mathcal{R}$: For each $\boldsymbol{s} \in \mathcal{C}$, there exist K transmit strategies $\boldsymbol{Q}^1, \ldots, \boldsymbol{Q}^K$ and coefficients $b_1, \ldots, b_K$ such that $\boldsymbol{Q}^k \in \mathcal{Q}$, $b_k \geq 0$, $\sum b_k = 1$, and

$$\boldsymbol{s} = \sum_{k=1}^{K} b_k \boldsymbol{r}(\boldsymbol{Q}^k).$$

Accordingly, the convex hull operation can be interpreted as *scheduling* between K transmit strategies, with scheduling coefficients $b_1, \ldots, b_K$. Moreover, the convex hull of a comprehensive set is comprehensive, hence $\mathcal{C}$ is comprehensive. In the following, let $\boldsymbol{Q}'$ denote a vector of transmit strategies, $\boldsymbol{Q}' = (\boldsymbol{Q}^1, \ldots, \boldsymbol{Q}^K)$, and let $\boldsymbol{b} = (b_1, \ldots, b_K)$.[1]

III. Interference Management

In general, transmission to MS k causes interference at all MS q with $q \neq k$. On the other hand, reducing the interference caused at MS q reduces the achievable rate for MS k. The goal of interference management is to adapt the system parameters in such a way that overall system performance is maximized. In this work, it is assumed that overall system performance is measured by a utility function u that maps a rate vector $\boldsymbol{s} \in \mathbb{R}_+^K$ into a scalar utility value $u(\boldsymbol{s})$. The utility function u is assumed to be continuous and increasing. Commonly used utility models are

$$\begin{aligned} u_{\text{WSR}}(\boldsymbol{s}) &= \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{s} && \text{(weighted sum-rate)}, \\ u_{\text{MM}}(\boldsymbol{s}) &= \min_k s_k && \text{(max-min fairness)}, \\ u_{\text{PF}}(\boldsymbol{s}) &= \sum_k \ln(s_k) && \text{(proportional fairness)}. \end{aligned}$$

Without scheduling, interference management corresponds to determining a feasible transmit strategy $\boldsymbol{Q}$ such that $u(\boldsymbol{r}(\boldsymbol{Q}))$ is maximized:

$$\max_{\boldsymbol{Q}} u(\boldsymbol{r}(\boldsymbol{Q})) \quad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{Q}. \tag{1}$$

Due to the nonconcavity of the rate map $\boldsymbol{r}$, problem (1) is generally a nonconvex optimization problem, regardless of the properties of u. Moreover, problem (1) offers no

[1] By adapting the results from [6] and [7], it can be shown that beamforming is optimal, i.e., it is sufficient to consider covariance matrices of rank 1. Based on this result, the problem can also be formulated using beamforming vectors instead of covariance matrices, cf. [7].

further structure with respect to the parameters $\boldsymbol{Q}$. Including scheduling makes the interference management problem even harder: Instead of finding a single transmit strategy $\boldsymbol{Q}$, it is now necessary to find a vector $\boldsymbol{Q}'$ of K feasible transmit strategies and a scheduling vector $\boldsymbol{b}$ such that the resulting rate vector maximizes utility.

Interference management is optimal if the globally optimal solution is found. However, finding a globally optimal solution of problem (1) directly by operating in the space of transmit strategies is practically impossible, due to the fact that problem (1) is nonconvex and the dimension of the search space is prohibitively high for global methods.[2] With scheduling, the dimension of the search space is further increased.

The key to finding globally optimal solutions is a rate space approach [4], which basically corresponds to a change of the optimization domain. Without scheduling, the rate space problem is given by

$$\max_{\boldsymbol{s}} u(\boldsymbol{s}) \quad \text{s.t.} \quad \boldsymbol{s} \in \mathcal{R}. \tag{2}$$

Clearly, if $\boldsymbol{s}^*$ is a global maximizer of (2), then there exists $\boldsymbol{Q}^*$ such that $\boldsymbol{s}^* = \boldsymbol{r}(\boldsymbol{Q}^*)$ and $\boldsymbol{Q}^*$ is a global maximizer of (1). The rate space approach provides two major advantages: First, the rate region $\mathcal{R}$ is comprehensive, while the utility function u is increasing. Hence, the rate space problem is a monotonic optimization problem [9], and can be solved by using a generic algorithm for monotonic optimization. Second, the dimension of the search space is reduced to K, the number of MS, and is independent of M.

The rate space problem for the case that scheduling is included is obtained by replacing $\mathcal{R}$ by $\mathcal{C}$ in (2):

$$\max_{\boldsymbol{s}} u(\boldsymbol{s}) \quad \text{s.t.} \quad \boldsymbol{s} \in \mathcal{C}. \tag{3}$$

Due to the fact that $\mathcal{C}$ is also comprehensive, the resulting rate space problem is again a monotonic optimization problem. If the utility function u is concave, the rate space problem with rate region $\mathcal{C}$ is a convex problem.

IV. Solving The Rate Space Problem

A general framework for solving rate space problems in the form of (2) and (3) is provided in [4]. The framework is based on the polyblock algorithm [9], a deterministic global optimization algorithm for solving monotonic optimization problems. As a global method that uses a black-box model of objective function and feasible set, the worst case computational complexity of the polyblock algorithm increases at least exponentially in K [10]. In practice, it can be observed that computing the globally optimal solutions is practically feasible for a small to moderate number of users only ($K \leq 10$). Moreover, the computational complexity of the polyblock algorithm limits the applicability of the framework to off-line computation. Nevertheless, by using global methods it is possible to compute the ultimate performance bounds for a given system configuration and a corresponding interference management strategy which is guaranteed to be globally optimal. The only prerequisite for applying the framework from [4] is the availability of a membership test for the rate region $\mathcal{R}$. In [4], the single-cell case is considered. For the multi-cell case, a membership test can be formulated as follows: A rate vector $\boldsymbol{s}$ is element of $\mathcal{R}$ if and only if there exists $\boldsymbol{Q}$ in $\mathcal{Q}$ such that

$$\boldsymbol{s}_k = r_k(\boldsymbol{Q}), \forall k. \tag{4}$$

Re-arranging (4) yields the condition

$$\boldsymbol{h}_{k,k}^{\mathrm{H}} \boldsymbol{Q}_k \boldsymbol{h}_{k,k} - \beta_k \sum_{q \neq k} \boldsymbol{h}_{q,k}^{\mathrm{H}} \boldsymbol{Q}_q \boldsymbol{h}_{q,k} = \beta_k \sigma^2, \forall k,$$

with $\beta_k = 2^{s_k} - 1$. The following feasibility test is obtained:

$$\begin{aligned} \text{find} \quad & (\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K) \\ \text{s.t.} \quad & \boldsymbol{Q}_k \succeq \boldsymbol{0}, \forall k, \\ & \sum_{k \in \mathcal{K}_b} \operatorname{tr}(\boldsymbol{Q}_k) \leq P, \forall b, \\ & \boldsymbol{h}_{k,k}^{\mathrm{H}} \boldsymbol{Q}_k \boldsymbol{h}_{k,k} - \beta_k \sum_{q \neq k} \boldsymbol{h}_{q,k}^{\mathrm{H}} \boldsymbol{Q}_q \boldsymbol{h}_{q,k} = \beta_k \sigma^2, \forall k. \end{aligned} \tag{5}$$

Problem (5) is a semidefinite program (SDP), i.e., a convex problem and efficiently solvable.[3]

V. Numerical Results

In order to illustrate the impact of optimal interference management, the optimal transmit strategies are computed for an exemplary channel realization. A system with $B = 2$ base station and $K = 4$ mobile stations is considered, with $\mathcal{K}_1 = \{1, 2\}$ and $\mathcal{K}_2 = \{3, 4\}$. Each base station has $M = 2$ transmit antennas and a transmit power budget of $P = 10^{1.5}$. The noise variance at each receiver is $\sigma^2 = 1$. As a reference strategy, we consider the case where $\boldsymbol{Q}_k$ is chosen such that it perfectly matches its channel and transmit power is divided equally among all associated MS:

$$\boldsymbol{Q}_k = 0.5P\, \boldsymbol{h}_{k,k} \boldsymbol{h}_{k,k}^{\mathrm{H}} / \operatorname{tr}\big(\boldsymbol{h}_{k,k} \boldsymbol{h}_{k,k}^{\mathrm{H}}\big), \forall k.$$

This case is denoted as *no coordination*, as it considers neither intra- nor inter-cell interference. Figure 1 shows the path gains $\boldsymbol{h}_q^{\mathrm{H}} \boldsymbol{Q}_k \boldsymbol{h}_q$ in case of no coordination. The diagonal entries in Figure 1 correspond to the signal paths to the four MS. It can be observed that the channels to MS 2 and 4 are best, while MS 3 has the weakest channel. The off-diagonal entries in Figure 1 correspond to interference. As an example, the signal to MS 4 causes significant interference at MS 1.

Figure 2 shows the path gains resulting from a choice of covariance matrices that maximizes the sum of rates. MS 1 and MS 3 are allocated zero transmit power – it is optimal to switch them off. Moreover, it can be observed that the signals to the active MS only cause interference at the inactive MS.

In Figure 3, the transmit strategy is chosen such that the resulting rate vector is max-min fair in $\mathcal{R}$ (i.e., no scheduling). For max-min fairness, no MS can be switched-off. The result is

[2] Clearly, there exist special cases that result in a sufficiently low problem dimension, such as $M = 1$ and K small.

[3] Based on the optimality of beamforming, the feasibility test can also be formulated as a second order cone program (SOCP), cf. [7].

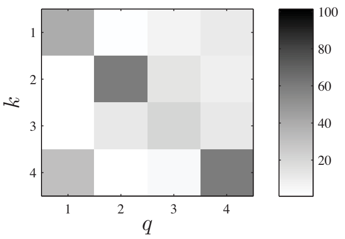


Fig. 1. Path gains $\boldsymbol{h}_q^{\mathrm{H}}\boldsymbol{Q}_k\boldsymbol{h}_q$, no coordination.

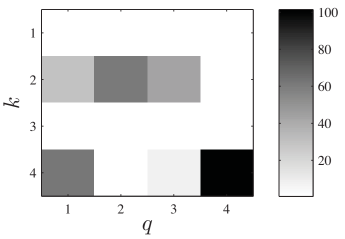


Fig. 2. Path gains $\boldsymbol{h}_q^{\mathrm{H}}\boldsymbol{Q}_k\boldsymbol{h}_q$, sum-rate maximization.

a significant amount of interference, due to insufficient degrees of freedom. Moreover, comparing Figure 3 to Figure 1 shows that it is optimal for BS 1 to transmit with a total power less than P.

Table I shows the optimal rate vectors and corresponding utility values for different performance objectives. The first row corresponds to the rate vector resulting from no cooperation. In the second row, the transmit strategy maximizes the sum of the users' rates. For sum-rate maximization, scheduling is not needed. Rows 3 and 4 correspond to a transmit strategy that is optimal under the max-min and proportional fairness objective, respectively. For the results in rows 3 and 4, optimization is over $\mathcal{R}$ (no scheduling). Whereas sum-rate maximization can achieve a significant gain over the no cooperation case, the benefit of cooperation is significantly lower in case of max-min and proportional fairness. This result is due to the fact that max-min and proportional fairness enforce nonzero rate for all users. Without scheduling, this implies that all users have to be active at the same time. However, there are only two spatial degrees of freedom available, hence it is not possible to properly separate users. Rows 5 and 6 show the optimal rates for the case that jointly optimal scheduling and beamforming is performed. The gains of optimal scheduling are significant – in case of max-min, the minimal rate more than doubles by including scheduling.

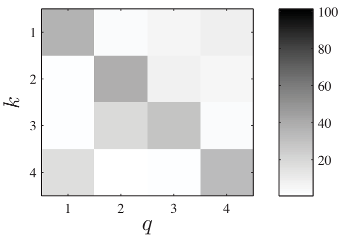


Fig. 3. Path gains $\boldsymbol{h}_q^{\mathrm{H}}\boldsymbol{Q}_k\boldsymbol{h}_q$, max-min fairness.

	r_1	r_2	r_3	r_4	u_{SR}	u_{MM}	u_{PF}
NoCo	1.14	2.21	0.85	1.47	5.68	0.85	1.15
SR	0.00	5.96	0.00	6.81	**12.77**	0.00	-Inf
MM	1.40	1.40	1.40	1.40	5.61	**1.40**	1.35
PF	1.63	1.81	0.96	1.43	5.83	0.96	**1.39**
MM-S	2.96	2.96	2.96	2.96	11.84	**2.96**	4.34
PF-S	2.97	2.98	2.72	3.41	12.08	2.72	**4.41**

TABLE I

RATES AND UTILITY VALUES FOR DIFFERENT STRATEGIES

k	r_1	r_2	r_3	r_4	b_k
1	0.00	5.96	0.00	6.81	0.43
2	5.95	0.00	5.44	0.00	0.50
3	0.00	5.42	3.75	0.00	0.05
4	0.00	5.58	3.65	0.00	0.02

TABLE II

OPTIMAL SCHEDULE FOR MAX-MIN FAIRNESS

Table II shows the optimal rate vectors and scheduling coefficients for max-min fairness. It can be observed that it is optimal to have only two users active at a given time. While this result can be expected (as $M = 2$), it is not a priori clear which two users are grouped together.

REFERENCES

[1] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, August 2006.

[2] S. Stanczak, M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks - Theory and Algorithms*, 2nd ed., ser. Foundations in Signal Processing, Communications and Networking, W. Utschick, H. Boche, and R. Mathar, Eds. Springer-Verlag, 2009.

[3] H. Dahrouj and W. Yu, "Coordinated beamforming for the multi-cell multi-antenna wireless system," in *Proceedings of the Conference on Information Sciences and Systems (CISS)*, Princeton, USA, March 2008.

[4] J. Brehmer and W. Utschick, "Utility maximization in the multi-user MISO downlink with linear precoding," in *Proceedings of the IEEE International Conference on Communications (ICC)*, June 2009.

[5] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, 2008.

[6] X. Shang, B. Chen, and H. V. Poor, "Multi-user MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," July 2009, submitted to IEEE Transactions on Information Theory. [Online]. Available: arXiv:0907.0505v1

[7] R. Zhang and S. Cui, "Cooperative interference management in multi-cell downlink beamforming," October 2009, submitted to IEEE Transactions on Signal Processing. [Online]. Available: arXiv:0910.2771v1

[8] E. Jorswieck and E. Larsson, "Monotonic optimization framework for the MISO IFC," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2009, pp. 3633–3636.

[9] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM Journal on Optimization*, vol. 11, no. 2, pp. 464–494, 2000.

[10] S. A. Vavasis, "Complexity issues in global optimization: a survey," in *Handbook of Global Optimization*. Kluwer, 1995, pp. 27–41.

Author Index