EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS
Eur. Trans. Telecomms. 2008; **19**:405–420
Published online 24 April 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/ett.1298

On the capacity of interference channels with one cooperating transmitter

Ivana Marić[1]*, Andrea Goldsmith[1], Gerhard Kramer[2] and Shlomo Shamai (Shitz)[3]

[1]*Department of Electrical Engineering, Stanford University, Stanford, CA, USA*
[2]*Bell Labs, Alcatel-Lucent, Murray Hill, NJ, USA*
[3]*Technion, Haifa, Israel*

SUMMARY

Inner and outer bounds are established on the capacity region of two-sender, two-receiver interference channels (IC) where one transmitter knows both messages. The transmitter with this extra message knowledge is referred to as being cognitive. The inner bound is based on strategies that generalise prior work, and includes rate-splitting, Gel'fand–Pinsker (GP) coding and cooperative transmission. A general outer bound is based on the Nair–El Gamal outer bound for broadcast channels (BCs). A simpler bound is presented for the case in which one of the decoders can decode both messages. The bounds are evaluated and compared for Gaussian channels. Copyright © 2008 John Wiley & Sons, Ltd.

1. INTRODUCTION AND RELATED WORK

Two-sender, two-receiver channel models allow for various forms of transmitter cooperation. When senders are unaware of each other's messages, we have the interference channel (IC) [1, 2]. In wireless networks, however, the broadcast nature of the wireless medium allows nodes to overhear transmissions and possibly decode parts of other users' messages. An encoder that has such knowledge can use it to improve its own rate as well as the other user's rate. The level of cooperation and the resulting performance improvement will depend on the amount of information the encoders share, as demonstrated in Reference [3].

Channel models with cooperating nodes are of interest also for networks with cognitive users. Cognitive radio [4] technology is aimed at developing smart radios that are both aware of and adaptive to the environment. Such radios can efficiently sense the spectrum, decode information from detected signals and use that knowledge to improve the system performance. This technology motivates information–theoretic models that try to capture the cognitive radio characteristics. In that vein, this paper considers a two-sender, two-receiver channel model in which, somewhat idealistically, we assume that cognitive capabilities allow one user to know the full message of the other encoder, as shown in Figure 1. The capacity region for this channel is unknown in general, although it has been determined for special cases. No existing coding scheme is known to be uniformly better than other known techniques for all channel characteristics and topologies.

The IC with one cooperating encoder was dubbed the *cognitive radio channel* and achievable rates were presented in References [3, 5]. The capacity region for the strong interference regime in which both receivers can decode both messages was determined in Reference [6]. The capacity region for the Gaussian case of weak interference was determined in References [7] and [8]. The results of References [7, 8] were extended to the Gaussian MIMO cognitive radio network and shown to achieve the sum-capacity in Reference [9]. The conclusions of Reference [9] apply to the single-antenna cognitive radio channel as well. A general encoding scheme was recently proposed in Reference [10]. Related work can also be found in References [11, 12]. Scaling laws for cognitive networks were analysed in Reference [13].

In this paper, we present a scheme that generalises those in References [6–8]. The scheme is similar to the one in Reference [10]: as in References [10] and [3], an encoder

* Correspondence to: Ivana Marić, Department of Electrical Engineering, Stanford, University, Stanford, CA, USA. E-mail: ivanam@wsl.stanford.edu

Copyright © 2008 John Wiley & Sons, Ltd. *Accepted 29 February 2008*

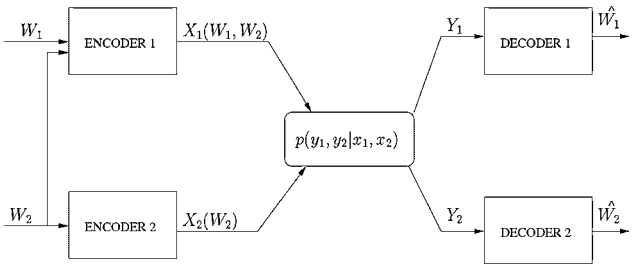


Figure 1. Interference channel with cooperating encoder.

uses *rate-splitting* [2] to enable the other receiver to decode part of the interference; the cognitive transmitter cooperates in sending the other user's message to its intended receivers and uses Gel'fand–Pinsker (GP) binning [14] to reduce interference to its own receiver. The key difference of our contribution to the prior work is in the way the binning is performed.

An overview of the encoding scheme is given in the next section. The channel model is described in Section 3. Details of the encoding scheme are developed in Section 4. In Section 6, the encoding scheme is adapted for Gaussian channels and its performance is compared to the performance of other coding schemes. Our results demonstrate improvements compared to the general scheme of Reference [10]. In Section 5, we present two outer bounds for the IC with one cooperating encoder. The first bound is based on Reference [15] and the only difference is in the input distribution over which the optimisation is performed. We then present an outer bound for the strong interference case that is of the same form as the one in Reference [6, Section V] and compare it to the achievable rate region for Gaussian channels.

2. OVERVIEW OF THE ENCODING STRATEGY

The channel model in Figure 1 has elements of both the IC and the broadcast channel (BC). Encoding techniques developed for either of these channel models are therefore useful for our model. If the message W_2 of encoder 2 was not known at the cognitive encoder, the channel would reduce to the IC. The best achievable rate region for the IC is achieved by rate-splitting [2, 16]: each encoder divides its message into two parts and encodes each of them with a separate codebook. This allows receivers to decode one of the other user's messages and cancel a part of the interference. Rate-splitting was applied to the cognitive radio channel model in References [3, 10]. In the encoding scheme presented in this paper, rate-splitting is performed at the cognitive encoder.

The cognitive encoder can employ a number of techniques in addition to rate-splitting. For example, to improve the rate for the noncognitive communicating pair, the cognitive encoder can *cooperate* by encoding W_2 to help convey it to the other decoder. On the other hand, any signal carrying information about W_2 creates interference to the cognitive encoder's receiver. This interference is known at the cognitive transmitter and therefore techniques for precoding against the interference, for example GP binning [14] or dirty-paper coding (DPC) [17], can be employed. In fact, GP binning is crucial for the cognitive radio channel: together with cooperation, it achieves capacity in certain scenarios [7–9]. It is not surprising that DPC brings gains in the Gaussian cognitive radio channel: if the non-cognitive encoder is silent, we have a BC from the cognitive encoder to the two receivers for which DPC is the optimal strategy [18, 19].

In general, however, there are two differences at the cognitive encoder from the classical GP setting. First, the interference carries useful information for receiver 2. Second, the interference is a *codebook* of some rate and can thus have lower entropy than in the GP setting. As we will see in SubSection 2.2, the latter can be exploited to achieve a higher rate.

We note that due to rate-splitting, there is a common part of W_1 decoded at the both receivers and precoded against interference. Since the signal carrying this common message experiences *different* interference at the two receivers, we use the ideas of References [20] and [21] that respectively extend References [14] and [17] to channels with different states known non-causally to the encoder. For Gaussian channels, DPC is generalised to *carbon-copying* onto dirty paper [21] to adjust to the interference experienced at both receivers.

2.1. Summary of techniques and special cases

Based on the above discussion, a number of techniques may be applied to exploit the additional knowledge of the cognitive encoder:

- Rate-splitting at encoder 1: to improve R_2 through interference cancellation at decoder 2.
- GP binning and binning against a codebook: to improve R_1 by precoding against interference. This approach also allows decoder 1 to decode message W_2 (or part of it) when R_2 is small, as will be shown in SubSection 2.2.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

- Carbon-copying onto dirty paper: to improve the rate of the common message sent by the cognitive encoder.
- Cooperation: Encoder 1 contributes to R_2 by encoding W_2.

A general encoding scheme that combines these techniques is described in Section 4. While this general encoding scheme may not achieve capacity in all scenarios, there are a number of special cases of channels for which a subset of techniques suffice to achieve capacity, as we now describe:

(1) Strong interference: both decoders can decode both messages with no rate penalty, so there is no need for either rate-splitting or binning. Superposition coding achieves capacity [6].
(2) Cognitive decoder must decode both messages: again, there is no need for binning. Rate-splitting and superposition coding achieve capacity [22, 23].
(3) Weak interference at receiver 2: there is no need for a common part of message W_1 and hence for rate-splitting. DPC and cooperation achieve capacity for Gaussian channels [7–9].

2.2. Rate improvement due to binning against a codebook

For the communication between the cognitive transmitter and its receiver, a codebook carrying W_2 creates interference. The situation is depicted in Figure 2, where S plays the role of the codebook of rate R_s interfering with the communication of message W at rate R. While in the GP problem the interference S is generated by a discrete memoryless source, the interference in the cognitive setting is a *codebook* of some rate R_s. The next lemma reflects the fact that when R_s is small, this can be exploited for potential rate gains.

Lemma 1. *For the communication situation of Figure 2, the rate*

$$R \leqslant \max_{P_{U|S},\, f(\cdot)} \min\{I(X;Y|S), \max\{I(U,S;Y) - R_s,\, I(U;Y) - I(U;S)\}\} \quad (1)$$

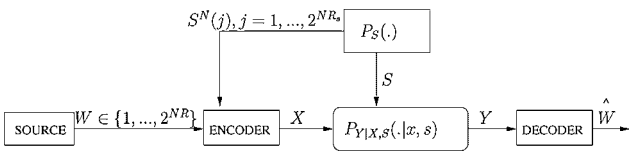


Figure 2. Communication from the cognitive transmitter to its receiver.

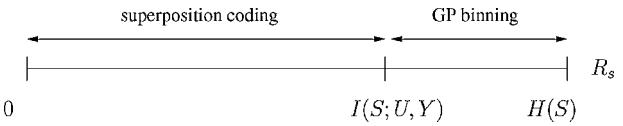


Figure 3. Binning against a codebook.

is achievable. For $I(S;U,Y) \leqslant R_s \leqslant H(S)$*, binning achieves the GP rate given by the second term on the right-hand side in Equation* (1)*. For* $R_s \leqslant I(S;U,Y)$*, superposition coding achieves the rate given by the first term on the right-hand side in Equation* (1)*. The two cases are shown in Figure* 3*.*

Proof. See Appendix B.

Remark 1. Rate (1) can be written as

$$R \leqslant \max_{P_{U|S},\, f(\cdot)} \{I(X,S;Y) - \max\{I(S;Y), \min\{R_s, I(U,Y;S)\}\} \quad (2)$$

From Equations (1) and (2), we observe that $I(S;U,Y) \leqslant R_s \leqslant H(S)$ corresponds to the classical GP setting. Potential rate improvement comes for $R_s \leqslant I(S;U,Y)$. Interestingly, in this case the receiver decodes both indexes (w, j), thus learning both its message and the interference. A related setting in which both data and the channel state information are communicated to the receiver was analysed in References [24, 25].

In the cognitive setting of Figure 1, index W_2 corresponds to index j of Figure 2. The implication is that, when R_s is small, receiver 1 will decode a part (or the whole) of W_2 without having encoder 2 rate-split to send common information in the sense of References [2, 16].

Recall that, due to rate-splitting, encoder 1 uses two codebooks to send a common and a private index. We denote these respective codebooks as (U_{1c}^N, U_{1a}^N). We can distinguish four cases depending on whether the two codebooks are generated through binning or superposition coding with respect to X_2^N:

(1) Binning: both (U_{1c}^N, U_{1a}^N) are binned against the codebook X_2^N of the non-cognitive encoder.
(2) Superposition coding: codebooks are superimposed on X_2^N.
(3) Binning then superposition coding: U_{1c}^N is binned against X_2^N, and U_{1a}^N is superimposed on (X_2^N, U_{1c}^N).
(4) Superposition coding then binning: U_{1c}^N is superimposed on X_2^N; U_{1a}^N is superimposed on U_{1c}^N and binned against X_2^N.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

In the last two cases, decoder 1 can decode W_2 due to superposition coding of U_{1a}^N or U_{1c}^N on X_2^N, as shown in Lemma 1. The setting thus corresponds to the *cognitive radio with degraded message sets*. For this channel model, superposition coding achieves capacity [22, 23]. The last two cases can therefore bring no rate improvement. The achievable rate region is the union of two rate regions achieved by binning or superposition coding. We will derive these regions after formally defining the problem in the next section. We remark that in the above encoding scheme, codebook U_{1a}^N is always superimposed on U_{1c}^N. The other encoding choice would be to use binning for U_{1a}^N against the codebook U_{1c}^N.

Finally, we note that encoder 2 also uses rate-splitting and forms two codebooks (X_{2a}^N, X_{2b}^N) using superposition coding. Encoder 1 bins against both codebooks and does not decode a part of W_2. Following Lemma 1, the respective rates R_{2a} and R_{2b} could be chosen such that (U_{1a}^N, U_{1c}^N) are binned against one of the two codebooks, but superimposed on the other. That would facilitate decoding a part of W_2 at receiver 1.

3. CHANNEL MODEL

Consider a channel with finite input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, finite output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, and a conditional probability distribution $p(y_1, y_2|x_1, x_2)$, where $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ are channel inputs and $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ are channel outputs. Each encoder t, $t = 1, 2$, wishes to send a message $W_t \in \{1, \ldots, M_t\}$ to decoder t in N channel uses. Message W_2 is also known at encoder 1 (see Figure 1). The channel is memoryless and time-invariant in the sense that

$$p(y_{1,n}, y_{2,n}|x_1^n, x_2^n, y_1^{n-1}, y_2^{n-1}, \bar{w}) = p_{Y_1,Y_2|X_1,X_2}(y_{1,n}, y_{2,n}|x_{1,n}, x_{2,n}) \quad (3)$$

for all n, where X_1, X_2 and Y_1, Y_2 are random variables representing the respective inputs and outputs, $\bar{w} = [w_1, w_2]$ denotes the messages to be sent, and $x_t^n = [x_{t,1}, \ldots, x_{t,n}]$. We will follow the convention of dropping subscripts of probability distributions if the arguments of the distributions are lower case versions of the corresponding random variables.

An (M_1, M_2, N, P_e) code has two encoding functions

$$X_1^N = f_1(W_1, W_2) \quad (4)$$

$$X_2^N = f_2(W_2) \quad (5)$$

two decoding functions

$$\hat{W}_t = g_t\left(Y_t^N\right) \quad t = 1, 2 \quad (6)$$

and an error probability

$$P_e = \max\{P_{e,1}, P_{e,2}\} \quad (7)$$

where, for $t = 1, 2$, we have

$$P_{e,t} = \sum_{(w_1,w_2)} \frac{1}{M_1 M_2} P\left[g_t\left(Y_t^N\right) \neq w_t|(w_1, w_2) \text{ sent}\right] \quad (8)$$

A rate pair (R_1, R_2) is achievable if, for any $\epsilon > 0$, there is an (M_1, M_2, N, P_e) code such that

$$M_t \geqslant 2^{NR_t}, \quad t = 1, 2, \text{ and } P_e \leqslant \epsilon$$

The capacity region of the IC with a cooperating encoder is the closure of the set of all achievable rate pairs (R_1, R_2).

4. ACHIEVABLE RATE REGION

To obtain an inner bound, we employ rate-splitting. We let

$$R_1 = R_{1a} + R_c \quad (9)$$

$$R_2 = R_{2a} + R_{2b} \quad (10)$$

for nonnegative $R_{1a}, R_c, R_{2a}, R_{2b}$, which we now specify. In the encoding scheme, encoder 2 uses superposition coding with two codebooks X_{2a}^N, X_{2b}^N. Encoder 1 repeats the steps of encoder 2 and adds binning: it encodes the split message W_1 with two codebooks which are GP precoded against X_{2a}^N, X_{2b}^N. In particular:

(1) Binning against X_{2a}^N, X_{2b}^N is used to create a codebook U_{1c}^N of common rate R_c.
(2) Binning against X_{2a}^N, X_{2b}^N conditioned on U_{1c} is used to create a codebook U_{1a}^N with private rate R_{1a}.

The encoding structure is shown in Figure 4.

We have the following result.

Theorem 1 (joint decoding)**.** *Rates* (9) *and* (10) *are achievable if*

$$R_{1a} \leqslant I(U_{1a}; Y_1|U_{1c}, Q) - I(U_{1a}; X_{2a}, X_{2b}|U_{1c}, Q) \quad (11)$$

$$R_1 \leqslant I(U_{1c}, U_{1a}; Y_1|Q) - I(U_{1c}, U_{1a}; X_{2a}, X_{2b}|Q) \quad (12)$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

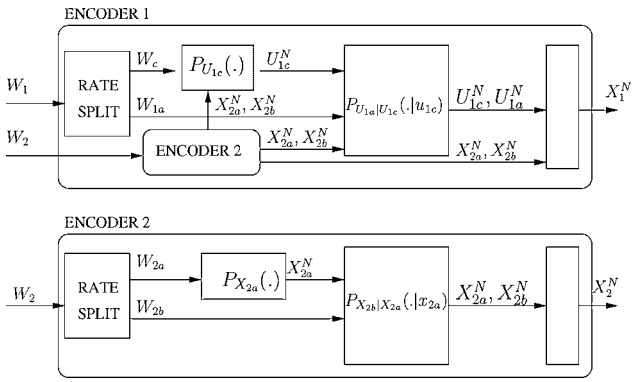


Figure 4. Encoding structure.

$$R_2 \leqslant I(X_2; Y_2, U_{1c}|Q) \tag{13}$$

$$R_2 + R_c \leqslant I(X_2, U_{1c}; Y_2|Q) \tag{14}$$

$$R_{2b} \leqslant I(X_{2b}; Y_2, U_{1c}|X_{2a}, Q) \tag{15}$$

$$R_{2b} + R_c \leqslant I(X_{2b}, U_{1c}; Y_2|X_{2a}, Q) \tag{16}$$

for some joint distribution that factors as

$$p(q)p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2|q)p(y_1, y_2|x_1, x_2) \tag{17}$$

and for which the right-hand sides of Equations (11) *and* (12) *are nonnegative. Q is a time-sharing random variable.*

Proof. See Appendix A.

Theorem 2 (sequential decoding)**.** *Rates* (9) *and* (10) *are achievable if*

$$R_{1a} \leqslant I(U_{1a}; Y_1|U_{1c}, Q) - I(U_{1a}; X_2|U_{1c}, Q) \tag{18}$$

$$R_c \leqslant \min\{I(U_{1c}; Y_1|Q), I(U_{1c}; Y_2, X_{2a}|Q)\} - I(U_{1c}; X_2|Q) \tag{19}$$

$$R_{2a} \leqslant I(X_{2a}; Y_2|Q) \tag{20}$$

$$R_{2b} \leqslant I(X_{2b}; Y_2, U_{1c}|X_{2a}, Q) \tag{21}$$

for some joint distribution that factors as

$$p(q)p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2|q)p(y_1, y_2|x_1, x_2)$$

and for which the right-hand sides of Equations (18) *and* (19) *are nonnegative.*

Proof. The proof follows similar steps as the proof of Theorem 1 and is omitted. Details can be found in Reference [26].

Remark 2. The rates of Theorem 1 include the rates of Theorem 2.

Remark 3. Theorem 1 includes the rates of the following schemes:

- The scheme of Reference [7, Proposition 3.1]: set $X_{2a} = \emptyset$, $U_{1c} = \emptyset$, $X_{2b} = (X_2, U)$ and $U_{1a} = V$ so that Equations (11)–(17) become

$$R_2 \leqslant I(X_2, U; Y_2) \tag{22}$$

$$R_1 \leqslant I(V; Y_1) - I(V; X_2, U) \tag{23}$$

  for $p(u, v, x_1, x_2)$.
- The scheme of Reference [8, Lemma 4.2]: set $X_{2a} = \emptyset$, $X_{2b} = X_2$, $U_{1a} = \emptyset$, and $R_1 = R_c$, $R_2 = R_{2b}$ so that Equations (11)–(17) become

$$R_2 \leqslant I(X_2; Y_2|U_{1c})$$

$$R_1 \leqslant \min\{I(U_{1c}; Y_1), I(U_{1c}; Y_2)\} \tag{24}$$

  for $p(x_2)p(u_{1c})$. The strategy in Reference [8] considers the case when

$$I(U_{1c}; Y_1) \leqslant I(U_{1c}; Y_2) \tag{25}$$

- Carbon-copy on dirty paper [21]: set $X_{2a} = \emptyset$, $U_{1a} = \emptyset$.
- For $X_{2a} = \emptyset$, our scheme closely resembles the scheme in Reference [10]. One difference in our scheme is that the two binning steps are not done independently which brings potential improvements. Another difference is in the evaluation of error events.

It is also interesting to compare our scheme to the encoding scheme in Reference [3]. The latter combines rate-splitting at both users, with two-step binning at the cognitive user. Each user sends a private index decoded by its receiver and a common index decoded by both. Again, one difference in our scheme is that the two binning steps are not independent. Another is that in our scheme the cognitive encoder cooperates by encoding W_2.

We next exploit Lemma 1.

4.1. An achievable rate region with superposition coding

Consider a joint distribution (17) and rate R_2 that satisfies

$$R_2 \leqslant I(X_2; U_{1c}, Y_1) \tag{26}$$

$$R_2 \leqslant I(X_2; U_{1a}, Y_1|U_{1c}) \tag{27}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

From Lemma 1, we know that under conditions (26) and (27), superposition of U_{1c}^N and U_{1a}^N with X_2^N should be used instead of binning. The encoding scheme of the cognitive encoder reduces to rate-splitting and superposition coding. The scheme and the obtained rates reduce to that of Reference [10, Theorem 5] derived for the cognitive radio with degraded message sets, in which the cognitive decoder needs to decode both messages. No rate-splitting at encoder 2 is needed. The resulting achievable rates (R_1, R_2) satisfy

$$\begin{aligned} R_{1a} &\leqslant I(X_1; Y_1|X_2, U_{1c}) \\ R_1 &\leqslant I(X_1; Y_1|X_2) \\ R_1 + R_2 &\leqslant I(X_1, X_2; Y_1) \\ R_c + R_2 &\leqslant I(U_{1c}, X_2; Y_2) \end{aligned} \tag{28}$$

for some joint input distribution $p(x_2, u_{1c}, x_1)$. After Fourier–Motzkin elimination [27], we obtain the following region.

Theorem 3 [22]. *Achievable rates* (R_1, R_2) *satisfy*

$$\begin{aligned} R_1 &\leqslant I(X_1; Y_1|X_2) \\ R_2 &\leqslant I(U_{1c}, X_2; Y_2) \\ R_1 + R_2 &\leqslant I(X_1; Y_1|X_2, U_{1c}) + I(U_{1c}, X_2; Y_2) \\ R_1 + R_2 &\leqslant I(X_1, X_2; Y_1) \end{aligned} \tag{29}$$

for some joint input distribution $p(x_2, u_{1c}, x_1)$.

Remark 4. The above region is the capacity region for the cognitive radio with degraded message sets: the converse follows from Reference [23] where a more general case of confidential messages is analysed.

5. OUTER BOUNDS

We next derive two capacity outer bounds.

Theorem 4. *The union of the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leqslant I(V, U_1; Y_1) \tag{30}$$

$$R_2 \leqslant I(V, U_2; Y_2) \tag{31}$$

$$R_1 + R_2 \leqslant \min\{I(V, U_1; Y_1) + I(U_2; Y_2|U_1, V), \tag{32}$$

$$I(U_1; Y_1|U_2, V) + I(V, U_2; Y_2)\} \tag{33}$$

for input distributions $p(v, u_1, u_2, x_1, x_2)$ *that factor as*

$$p(u_1)p(u_2)p(v|u_1, u_2)p(x_2|u_2, v)p(x_1|u_1, u_2, v) \tag{34}$$

is an outer bound to the capacity region.

Proof. See Appendix C.

Remark 5. We observe that Equations (30)–(33) are of the same form as the outer bound for the BC in Reference [15, Section 3]. The difference is the factorisation of the input distribution.

Remark 6. One can restrict attention to distributions (34) where X_2 is a function of (U_2, V) and X_1 is a function of (U_1, U_2, V). The bounds (30)–(33) can then be written as

$$R_1 \leqslant I(V, U_1; Y_1) \tag{35}$$

$$R_2 \leqslant I(V, U_2, X_2; Y_2) \tag{36}$$

$$R_1 + R_2 \leqslant \min\{I(V, U_1; Y_1) + I(X_1, X_2; Y_2|U_1, V), \tag{37}$$

$$I(X_1; Y_1|X_2, U_2, V) + I(V, U_2, X_2; Y_2)\} \tag{38}$$

From Equations (36) and (38), we obtain the outer bound of Reference [7, Theorem 3.2]:

$$R_1 \leqslant I(X_1; Y_1|X_2) \tag{39}$$

$$R_2 \leqslant I(U, X_2; Y_2) \tag{40}$$

$$R_1 + R_2 \leqslant I(X_1; Y_1|X_2, U) + I(U, X_2; Y_2) \tag{41}$$

where we use the notation $U = [U_2, V]$ and also add Equation (39) as it follows by standard methods. The probability distribution factors as

$$p(u, x_1, x_2)p(y_1, y_2|x_1, x_2) \tag{42}$$

Interestingly, Equations (39)–(42) were shown to be tight under weak interference [7] and in particular for Gaussian channels with weak interference [7, 8]. The following theorem gives a simple capacity outer bound in strong interference.

Theorem 5. *For an interference channel with one cooperating encoder satisfying*

$$I(X_1; Y_1|X_2) \leqslant I(X_1; Y_2|X_2) \tag{43}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

for all input distributions $p(x_1, x_2)$*, the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leqslant I(X_1; Y_1|X_2) \tag{44}$$

$$R_1 + R_2 \leqslant I(X_1, X_2; Y_2) \tag{45}$$

for all input distributions $p(x_1, x_2)$ *is an outer bound to the capacity region.*

Proof. See Appendix D.

Remark 7. The bound (45) reflects the fact that, because decoder 2 experiences strong interference, as given by Equation (43), it can decode W_1 with no rate penalty.

This bound is evaluated for Gaussian channels in Equation (55).

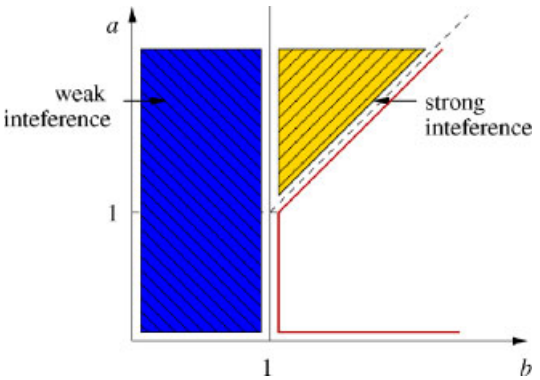


Figure 5. Values of (a, b) for which the capacity region of the Gaussian channel is known. For $b < 1$, the channel has weak interference and the capacity was determined in References [7, 8]. Strong interference conditions for the case $P_1 = P_2$ are $b > 1$ and $a > b$. The capacity was determined in Reference [6].

6. GAUSSIAN CHANNEL

To illustrate our results more concretely, consider the Gaussian IC described by

$$Y_1 = X_1 + aX_2 + Z_1 \tag{46}$$

$$Y_2 = bX_1 + X_2 + Z_2 \tag{47}$$

where $Z_t \sim \mathcal{N}[0, 1]$, $E[X_t^2] \leqslant P_t$, $t = 1, 2$ and $\mathcal{N}[0, \sigma^2]$ denotes the normal distribution with zero mean and variance σ^2.

6.1. Previous work: capacity results

The capacity region for Gaussian cognitive radio channels (46)–(47) was determined for the case of weak interference, that is $b \leqslant 1$, in References [7, 8]. The optimum coding strategy at the cognitive encoder consists of encoding message W_1 via DPC while treating X_2^N as interference, and superposition coding to help convey W_2 to receiver 2. Receiver 1 does not suffer interference due to DPC. Because the interference is weak, receiver 2 does not attempt to decode the unwanted message.

For $b \geqslant 1$, the interference at receiver 2 is stronger than in the previous case, and it is plausible to expect that decoding W_1 (or a part of it) may be beneficial. In fact, we observe from Equations (46) and (47) that *after* decoding W_2, decoder 2 has a better observation of X_1 than receiver 1, since $b \geqslant 1$. Therefore, decoder 2 can decode W_1. However, the conditions under which decoder 2 can decode W_1 also depend on the encoding/decoding approach. In particular, the capacity is known for the strong interference regime in which both decoders can decode both messages. The considered channel becomes a *compound* multiaccess (MAC) channel consisting of two MAC channels, one to each receiver. The strong interference conditions under which this is optimal were determined in Reference [6] leading to the capacity region. These conditions depend on P_1 and P_2. For $P_1 = P_2$, they simplify to $b > 1$ and $a > b$. Encoder 1 uses a superposition code

$$X_1 = X_{1c} + \sqrt{\frac{\bar{\alpha} P_1}{P_2}} X_2 \tag{48}$$

where $X_2 \sim \mathcal{N}[0, P_2]$, $X_{1c} \sim \mathcal{N}[0, \alpha P_1]$ and $0 \leqslant \alpha \leqslant 1$. Figure 5 shows the range of channel gains a and b for which the above capacity results apply. In Figure 5, we choose $P_1 = P_2$.

When the channel conditions do not allow decoding of both messages at decoder 1, they may still enable decoder 2 to decode W_1. Reference [8, Lemma 4.2] considers such a case, assuming superposition coding (48). Decoder 2 sequentially decodes: it first decodes W_1, subtracts the part of the received signal that carries it and decodes W_2 in the interference-free channel. The obtained rates are given by Equation (24). The conditions under which this encoding/decoding procedure is optimal are given by Equation (25).

When these conditions are not satisfied, there may be other interesting scenarios for which the presented techniques lead to capacity results. For example, consider the case when $a = 0$ in Equation (46). Then Equations (46)

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

and (47) describe a *Z*-channel where receiver 1 does not suffer interference and the strong interference conditions of Reference [6] are not satisfied. Suppose that encoder 1 is non-cooperating, aimed to achieve the largest possible R_1. It follows that $\alpha = 0$ in Equation (48). Consider the case for which capacity is unknown, that is condition (25) is not satisfied:

$$\frac{1}{2}\log(1+P_1) \geqslant \frac{1}{2}\log\left(1+\frac{b^2P_1}{1+P_2}\right) \tag{49}$$

For this case, the achievable rates evaluate to

$$\begin{aligned} R_1 &\leqslant \frac{1}{2}\log(1+P_1) \\ R_1+R_2 &\leqslant \frac{1}{2}\log(1+b^2P_1+P_2) \end{aligned} \tag{50}$$

For Gaussian channels, the outer bounds (44) and (45) are given by Equation (55). Due to the above assumption of a non-cooperative cognitive encoder, X_1 and X_2 are independent, and $\rho = 0$ in Equation (55). Therefore, the outer bound (55) coincides with the achievable rates (50), yielding capacity.

As the above discussion illustrates, there are still regimes for which the capacity of Gaussian cognitive channels is unknown. This motivates evaluating the general strategy proposed in Theorem 1. Next, we present such an evaluation.

6.2. Numerical results

We evaluate the rates of Theorem 1 for the special case $X_{2a} = \emptyset$ and $Q = \emptyset$. The rates of Theorem 1 reduce to

$$\begin{aligned} R_{1a} &\leqslant I(U_{1a};Y_1|U_{1c}) - I(U_{1a};X_2|U_{1c}) \\ R_1 &\leqslant I(U_{1c},U_{1a};Y_1) - I(U_{1c},U_{1a};X_2) \\ R_2 &\leqslant I(X_2;Y_2,U_{1c}) \\ R_2+R_c &\leqslant I(X_2,U_{1c};Y_2) \end{aligned} \tag{51}$$

To simplify Equation (51), we express the conditional entropies in terms of joint entropies, recall that $R_1 = R_c + R_{1a}$, and apply Fourier–Motzkin elimination to obtain

$$\begin{aligned} R_1 &\leqslant I(U_{1c},U_{1a};Y_1) - I(U_{1c},U_{1a};X_2) \\ R_2 &\leqslant I(X_2;Y_2,U_{1c}) \\ R_2 &\leqslant I(X_2,U_{1c};Y_2) \\ R_1+R_2 &\leqslant I(X_2,U_{1c};Y_2) + I(U_{1a};Y_1,U_{1c}) \\ &\quad - I(U_{1a};X_2,U_{1c}) \end{aligned} \tag{52}$$

It is also interesting to evaluate the rates of Theorem 2 achieved with sequential decoding for $X_{2a} = \emptyset$, $Q = \emptyset$. This evaluation results in

$$\begin{aligned} R_{1a} &\leqslant I(U_{1a};Y_1|U_{1c}) - I(U_{1a};X_2|U_{1c}) \\ R_c &\leqslant \min\{I(U_{1c};Y_1), I(U_{1c};Y_2)\} - I(U_{1c};X_2) \\ R_2 &\leqslant I(X_2;Y_2,U_{1c}) \end{aligned} \tag{53}$$

Remark 8. When $I(U_{1c};Y_1) \leqslant I(U_{1c};Y_2)$, decoder 2 can decode W_1. Thus, there is no need to rate split at encoder 1 and we choose $U_{1a} = \emptyset$. It follows from Equations (52) and (53) that for this case the same rates can be achieved by sequential decoding or by joint decoding.

Remark 9. We observe from Equation (53) that R_c, being a common rate, is bounded by the worst channel, as reflected by the $\min\{I(U_{1c};Y_1), I(U_{1c};Y_2)\}$ term. If $I(U_{1c};Y_1) > I(U_{1c};Y_2)$, transmitting X_{2a} will allow decoder 2 to decode part of W_2 before decoding W_c. It will also serve as an observation when decoding W_c as suggested by the expression $I(U_{1c};Y_2,X_{2a})$ in Equation (19). This will improve the common rate R_c.

We evaluate region given by Equation (52) for

$$X_2 \sim \mathcal{N}[0,P_2],\ X_{1c} \sim \mathcal{N}[0,\alpha\beta P_1],\ X_{1a} \sim \mathcal{N}[0,\alpha\bar{\beta}P_1]$$

$$\begin{aligned} U_{1c} &= X_{1c} + \lambda_1 X_2 \\ U_{1a} &= X_{1a} + \lambda_2 X_2 \\ X_1 &= X_{1c} + X_{1a} + \sqrt{\frac{\bar{\alpha}P_1}{P_2}}X_2 \end{aligned} \tag{54}$$

where $0 \leqslant \alpha,\ \beta \leqslant 1$ and $0 \leqslant \lambda_1,\ \lambda_2$. Parameters α and β determine the amount of power that the cognitive user dedicates for cooperation and for sending the common message, respectively.

We compare the region (52) to the outer bound of Theorem 5 which, for Gaussian channels, is given by the following corollary.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

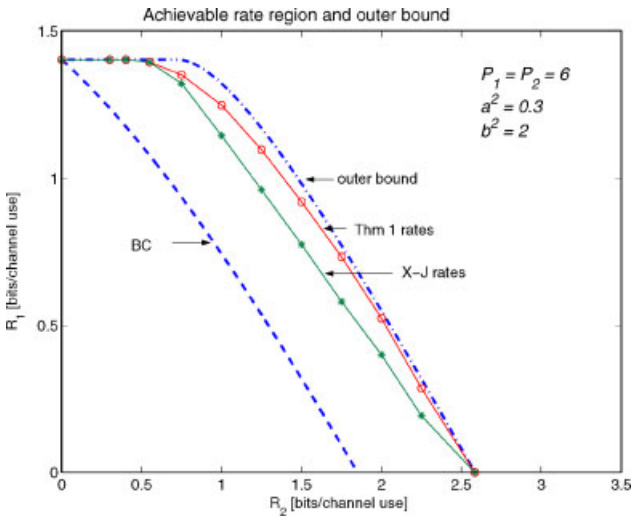


Figure 6. Achievable rates of Theorem 1 and Reference [10, Theorem 5] and the outer bound of Corollary 1. Also shown is the capacity region of a BC from the cooperative encoder, that is the case $P_2 = 0$.

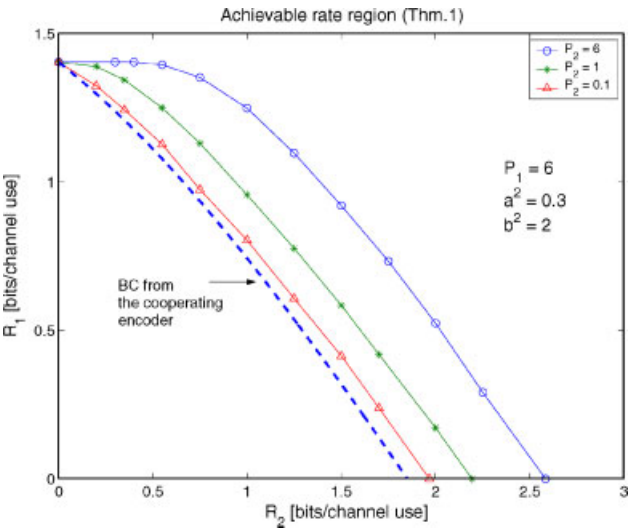


Figure 7. Achievable rates for different values of P_2.

Corollary 1. *When $b \geqslant 1$, any achievable rate pair (R_1, R_2) satisfies*

$$R_1 \leqslant \frac{1}{2}\log\left(1 + (1-\rho^2)P_1\right)$$
$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + b^2 P_1 + P_2 + 2\rho\sqrt{b^2 P_1 P_2}\right) \quad (55)$$

for some ρ, $0 \leqslant \rho \leqslant 1$.

Figure 6 shows the achievable rate region (52) and the outer bound (55) for channel gain values $a^2 = 0.3$, $b^2 = 2$ and equal powers $P_1 = P_2 = 6$. Theorem 1 gives larger rates than those of Reference [10, Theorem 5].

When encoder 2 does not transmit (i.e. $P_2 = 0$), the channel reduces to the BC in which there is only the cooperating encoder communicating to two receivers. Unlike the BC channel rate region, the region for the IC with one cooperating encoder is flat for small values of R_2, reflecting that a cognitive transmitter does not need to cooperate in this regime. It can instead use its full power to precode and transmit W_1 at the single-user rate as if the second user was not present. On the other hand, at $R_1 = 0$ the cooperating encoder fully helps encoder 2, that is $\alpha = 0$ and user 2 benefits from the coherent combining gain as indicated by the rate expression

$$R_{2,\max} = \frac{1}{2}\log\left(1 + \left(1 + b\sqrt{\frac{P_1}{P_2}}\right)^2 P_2\right)$$

The achievable rates are very close to the outer bound, especially for large values of R_2, in the regime where the cognitive encoder dedicates more of its power to cooperate.

Figure 7 shows achievable rates for different values of P_2 and fixed P_1. As P_2 decreases, the performance approaches the rate achieved in the BC with only the cooperating encoder transmitting to the two receivers. Since in the BC encoder 2 is not present, the rate region does not depend on P_2 and is given by the dashed line. Figure 8 shows the effect of reducing power at the cognitive encoder, keeping P_2 constant. This has a strong impact, drastically reducing R_1.

For the Gaussian channel, the rates achieved with sequential encoding (53) can be evaluated for the choice of random variables X_2, X_1, U_{1a}, X_{1a} as in Equation (54). U_{1c} carries a common message and is precoded against interference. Since the two channels from encoder 1 to the two receivers experience different interference, the carbon-copy method of Reference [21] can be used. More details on this approach are presented in Reference [26]. Figure 9 shows the performance of the two decoding schemes which can differ significantly.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

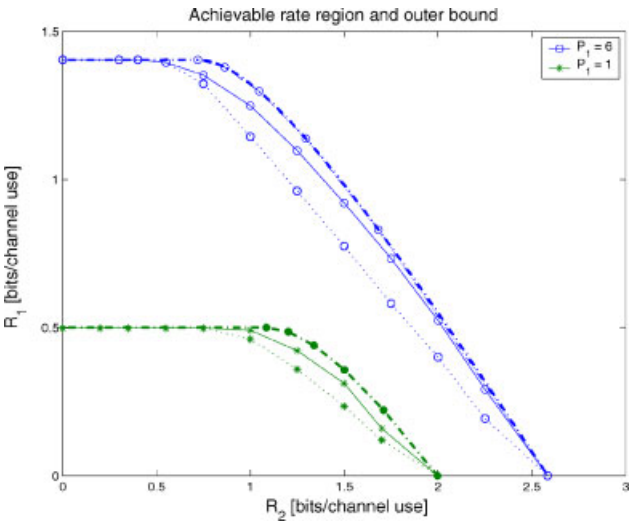


Figure 8. Impact of reduced power of the cognitive transmitter. Rates achieved with Theorem 1 are shown in solid lines and rates of Reference [10] are shown with dotted lines. Dash-dotted lines show the outer bound (55).

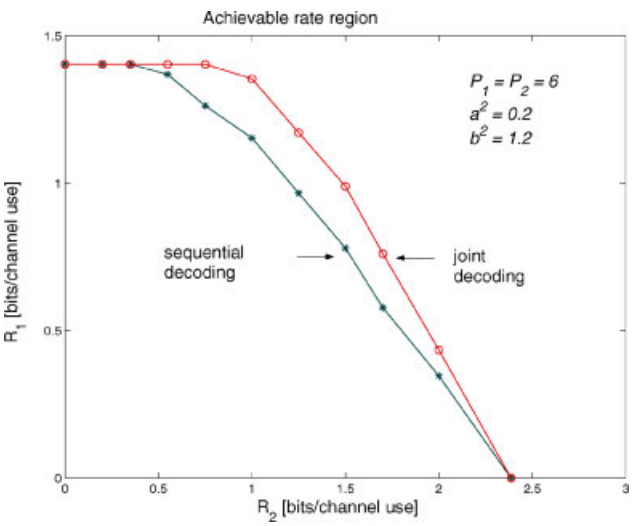


Figure 9. Comparison of achievable rates with joint and sequential decoding.

7. CONCLUSIONS AND FUTURE WORK

We have developed an encoding strategy for the IC with one cooperating encoder that generalises previously proposed strategies. We evaluated its performance and compared it to the performance of other schemes, focusing on the Gaussian channel. A comparison with Reference [3] would be an interesting next step. As explained in Section 4, the latter combines rate-splitting at both users, with two-step binning at the cognitive user. One of the differences is that in our scheme the cognitive encoder cooperates by encoding W_2. However, it is unclear whether our strategy generalises the scheme in Reference [3], or whether a combination of the two techniques would achieve higher rates. We also compared the proposed scheme to the outer bound that we developed for the strong interference regime. We further developed a new outer bound that extends the Nair–El Gamal broadcast outer bound. Evaluating this bound for specific channels such as Gaussian channels may give capacity results for special cases.

The cognitive radio channel shares some characteristics of both ICs and BCs. Combining encoding strategies developed for either of the two channel models therefore seems a natural approach. However, the optimality of a particular encoding scheme seems to be in part dictated by the channel conditions: for the Gaussian channel in which decoder 2 experiences weak interference, dirty-paper coding achieves capacity. On the other hand, strong interference conditions may allow the cognitive receiver to decode the message not intended for him and therefore DPC against that message is not needed; superposition coding and rate-splitting achieve capacity. An even simpler scheme suffices when both receivers experience strong interference and can both decode the two messages. Neither DPC nor rate-splitting is needed; superposition coding achieves capacity. The encoding scheme presented in this paper is a combination of rate-splitting, GP binning and superposition coding. We believe that this general encoding scheme may achieve capacity for certain special cases related to the channel or specific encoding/decoding constraints. Finding such special cases is a topic of ongoing investigation.

APPENDIX A: PROOF OF THEOREM 1

Proof (Theorem 1). **Code construction:** Ignore Q. Choose a distribution $p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2)$.

- Split the rates as in Equations (9) and (10).
- Generate $2^{NR_{2a}}$ codewords $x_{2a}^N(w_{2a})$, $w_{2a} = 1, \ldots, 2^{NR_{2a}}$, by choosing $x_{2a,n}(w_{2a})$ independently according to $P_{X_{2a}}(\cdot)$.
- For each w_{2a}, generate $2^{NR_{2b}}$ codewords $x_{2b}^N(w_{2a}, w_{2b})$ using $\prod_{n=1}^N P_{X_{2b}|X_{2a}}(\cdot|x_{2a,n}(w_{2a}))$, $w_{2b} = 1, \ldots, 2^{NR_{2b}}$.
- For each pair (w_{2a}, w_{2b}), generate $x_2^N(w_{2a}, w_{2b})$ where x_2 is a deterministic function of (x_{2a}, x_{2b}).
- Generate $2^{N(R_{1c}+R_{1c'})}$ codewords $u_{1c}^N(w_c, b_c)$, $w_c = 1, \ldots, 2^{NR_{1c}}$, $b_c = 1, \ldots, 2^{NR_{1c'}}$ using $P_{U_{1c}}(\cdot)$.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

- For each $u_{1c}^N(w_c, b_c)$, generate $2^{N(R_{1a}+R'_{1a})}$ codewords $u_{1a}^N(w_c, b_c, w_{1a}, b_{1a})$, $w_{1a} = 1, \ldots, 2^{NR_{1a}}$, $b_{1a} = 1, \ldots, 2^{NR'_{1a}}$ using $\prod_{n=1}^N P_{U_{1a}|U_{1c}}(\cdot|u_{1c,n}(w_c, b_c))$.
- For (w_1, w_2), generate $x_1^N(w_{2a}, w_{2b}, w_c, b_c, w_{1a}, b_{1a})$ where x_1 is a deterministic function of $(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_2)$.

Following the proof in Reference [28, Appendix D], it can be shown that it is sufficient to choose respective x_2 and x_1 to be deterministic functions of (x_{2a}, x_{2b}) and $(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_2)$.

Encoders: Encoder 1:

(1) Split the NR_1 bits w_1 into NR_{1a} bits w_{1a} and NR_c bits w_c. Similarly, split the NR_2 bits w_2 into NR_{2a} bits w_{2a} and NR_{2b} bits w_{2b}. We write this as

$$w_1 = (w_{1a}, w_c), \quad w_2 = (w_{2a}, w_{2b})$$

(2) Try to find a bin index b_c so that $(u_{1c}^N(w_c, b_c), x_{2a}^N(w_{2a}), x_{2b}^N(w_{2a}, w_{2b})) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})$ where $T_\epsilon(P_{XY})$ denotes the jointly ϵ-typical set with respect to P_{XY}, see Reference [29, Section 8.6]. If no such b_c is found, choose $b_c = 1$.
(3) For each (w_c, b_c), try to find a bin index b_{1a} such that $(u_{1a}^N(w_c, b_c, w_{1a}, b_{1a}), u_{1c}^N(w_c, b_c)), x_{2a}^N(w_{2a}), x_{2b}^N(w_{2a}, w_{2b}) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})$. If one cannot, choose $b_{1a} = 1$.
(4) Transmit $x_1^N(w_{2a}, w_{2b}, w_c, b_c, w_{1a}, b_{1a})$.

Encoder 2: Transmit $x_2^N(w_{2a}, w_{2b})$.

Decoders: Decoder 1: Given y_1^N, choose $(\hat{w}_c, \hat{b}_c, \hat{w}_{1a}, \hat{b}_{1a})$ if $(u_{1c}^N(\hat{w}_c, \hat{b}_c), u_{1a}^N(\hat{w}_c, \hat{b}_c, \hat{w}_{1a}, \hat{b}_{1a}), y_1^N) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})$. If there is more than one such a quadruple, choose one. If there is no such quadruple, choose (1, 1, 1, 1).

Decoder 2: Given y_2^N, choose $(\hat{w}'_{2a}, \hat{w}'_c, \hat{b}'_c, \hat{w}'_{2b})$ if $(x_{2a}^N(\hat{w}'_{2a}), u_{1c}^N(\hat{w}'_c, \hat{b}'_c), x_{2b}^N(\hat{w}'_{2a}, \hat{w}'_{2b}), y_2^N) \in T_\epsilon(P_{X_{2a}U_{1c}X_{2b}Y_2})$. If there is more than one such a quadruple, choose one. If there is no such quadruple, choose (1, 1, 1, 1).

Analysis: Suppose $(w_{1a}, w_c, w_{2a}, w_{2b}) = (1, 1, 1, 1)$ was sent. An encoder error occurs if

(1) Encoder 1 cannot find a bin index b_c such that $(u_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})$ which happens with probability

$$\begin{aligned} P_{e,enc1}^{(1)} &= P\left[\bigcap_{b_c=1}^{2^{NR'_c}} (U_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \right. \\ &\qquad \left. \notin T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})\right] \\ &= \left(1 - P\left[(U_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \right.\right. \\ &\qquad \left.\left. \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})\right]\right)^{2^{NR'_c}} \\ &\leqslant \left(1 - (1-\epsilon)2^{-N[I(U_{1c};X_{2a},X_{2b})+\delta]}\right)^{2^{NR'_c}} \\ &\leqslant \exp\left(-(1-\epsilon)2^{N[R'_c - I(U_{1c};X_{2a},X_{2b})-\delta]}\right) \end{aligned} \quad (56)$$

where the first inequality follows from Reference [29, Theorem 8.6.1] and the second from $(1-x)^m \leqslant e^{-mx}$. From Equation (56), the error probability $P_{e,enc1}^{(1)}$ can be made arbitrarily small if

$$R'_c > I(U_{1c}; X_{2a}, X_{2b}) + \delta \quad (57)$$

(2) After it has determined bin index b_c, say $b_c = 1$, encoder 1 cannot find a bin index b_{1a} such that $(u_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})$ which happens with probability

$$\begin{aligned} P_{e,enc1}^{(2)} &= P\left[\bigcap_{b_{1a}=1}^{2^{NR'_{1a}}} (U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), \right. \\ &\qquad \left. x_{2b}^N(1, 1)) \notin T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})\right] \\ &= \left(1 - P\left[(U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), \right.\right. \\ &\qquad \left.\left. x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})\right]\right)^{2^{NR'_{1a}}} \end{aligned} \quad (58)$$

We have

$$\begin{aligned} &P[(U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), \\ &x_{2b}^N(1, 1) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})] \\ &= \sum_{u_{1a}^N \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}}|u_{1c}^N x_{2a}^N x_{2b}^N)} P\left[u_{1a}^N | u_{1c}^N\right] \\ &\geqslant (1-\epsilon)2^{-N(H(U_{1a}|U_{1c}X_{2a}X_{2b})-H(U_{1a}|U_{1c})+\delta)} \\ &= (1-\epsilon)2^{-N(I(U_{1a};X_{2a}X_{2b}|U_{1c})+\delta)} \end{aligned} \quad (59)$$

where, as in Reference [30, Handout 2], we denote

$$T_\epsilon(P_{XY}|x^N) = \{y^N : (x^N, y^N) \in T_\epsilon(P_{XY})\}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

Table 1. Error events in joint decoding and corresponding rate bounds.

	Error event	Arbitrarily small positive error probability if
E_1	$(\hat{w}_c \neq 1, \hat{w}_{1a} = 1)$	$R_c + R'_c \leqslant I(U_{1c}, U_{1a}; Y_1)$
E_2	$(\hat{w}_c = 1, \hat{w}_{1a} \neq 1)$	$R_{1a} + R'_{1a} \leqslant I(U_{1a}; Y_1 \mid U_{1c})$
E_3	$(\hat{w}_c \neq 1, \hat{w}_{1a} \neq 1)$	$R_c + R'_c + R_{1a} + R'_{1a} \leqslant I(U_{1c}, U_{1a}; Y_1)$
E'_1	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c = 1)$	$R_{2a} \leqslant I(X_{2a}, X_{2b}; Y_2, U_{1c})$
E'_2	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1)$	$R_{2a} + R_{2b} \leqslant I(X_{2a}, X_{2b}; Y_2, U_{1c})$
E'_3	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c \neq 1)$	$R_{2a} + R_c + R'_c \leqslant I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b})$
E'_4	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c \neq 1)$	$R_{2a} + R_{2b} + R_c + R'_c \leqslant I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b})$
E'_5	$(\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1)$	$R_{2b} \leqslant I(X_{2b}; Y_2, U_{1c} \mid X_{2a})$
E'_6	$(\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c \neq 1)$	$R_{2b} + R_c + R'_c \leqslant I(X_{2b}, U_{1c}; Y_2 \mid X_{2a}) + I(U_{1c}; X_{2a}, X_{2b})$

The first inequality follows from the fact that U_{1a}^N was generated according to $P_{U_{1a}|U_{1c}}$ (also directly from Reference [30, Handout 3]). Employing Equation (59), we can bound Equation (58) as

$$P_{e,enc1}^{(2)} \leqslant \left(1 - (1-\epsilon)2^{-N[I(U_{1a};X_{2a},X_{2b}|U_{1c})+\delta]}\right)^{2^{NR'_{1a}}} \leqslant \exp\left(-(1-\epsilon)2^{N[R'_{1a}-I(U_{1a};X_{2a},X_{2b}|U_{1c})-\delta]}\right) \quad (60)$$

We need

$$R'_{1a} > I(U_{1a}; X_{2a}, X_{2b}|U_{1c}) + \delta \quad (61)$$

Decoder errors: The possible error events at the decoders are shown in the first column of Table 1. We next derive the corresponding rate bounds given in the second column of the same table, which guarantee that the error probability of each event can be made small as N gets large. Bounds for E_1, E'_1, E'_3 are redundant. The rest of the rate expressions in Table 1 yield Equations (11)–(16). Consider the probability of event E_1 :

$$P[\hat{W}_c \neq 1, \hat{W}_{1a} = 1] = \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} P\Big[\big(U_{1c}^N(w_c, b_c), U_{1a}^N(w_c, b_c, 1, 1), Y_1^N\big) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})\Big] \leqslant 2^{-N[I(U_{1c},U_{1a};Y_1)-(R_c+R'_c)-\delta]} \quad (62)$$

by References [29, Theorem 8.6.1] and [30, Handout 1, Theorem 2]. From Equation (62), achieving arbitrarily small error probability of E_1 requires

$$R_c + R'_c < I(U_{1c}, U_{1a}; Y_1) \quad (63)$$

Similarly, the probability of E_2 is

$$P[\hat{W}_c = 1, \hat{W}_{1a} \neq 1] = \sum_{w_{1a}=2}^{2^{NR_{1a}}} \sum_{b_{1a}=1}^{2^{NR'_{1a}}} P\Big[\big(U_{1c}^N(1,1), U_{1a}^N(1,1,w_{1a},b_{1a}), Y_1^N\big) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})\Big] \leqslant 2^{-N[I(U_{1a};Y_1|U_{1c})-(R_{1a}+R'_{1a})-\delta]} \quad (64)$$

where the inequality follows by Reference [29, Theorem 14.2.3, Handout 3]. We need

$$R_{1a} + R'_{1a} < I(U_{1a}; Y_1|U_{1c}) \quad (65)$$

The probability of E_3 is, similarly as in Equation (62),

$$P[\hat{W}_c \neq 1, \hat{W}_{1a} \neq 1] = \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} \sum_{w_{1a}=2}^{2^{NR_{1a}}} \sum_{b_{1a}=1}^{2^{NR'_{1a}}} P\Big[\big(U_{1c}^N(w_c,b_c), U_{1a}^N(w_c,b_c,w_{1a},b_{1a}), Y_1^N\big) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})\Big] \leqslant 2^{-N[I(U_{1c},U_{1a};Y_1)-(R_c+R'_c+R_{1a}+R'_{1a})-\delta]} \quad (66)$$

and thus requires

$$R_c + R'_c + R_{1a} + R'_{1a} < I(U_{1c}, U_{1a}; Y_1) \quad (67)$$

We next consider the error events at decoder 2. For E'_1, we have

$$P[\hat{W}'_{2a} \neq 1, \hat{W}'_{2b} = 1, \hat{W}'_c = 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} P\Big[\big(U_{1c}^N(1,1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a},1), Y_2^N\big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] \quad (68)$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

We further have

$$P\Big[\Big(U_{1c}^N(1,1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a},1), Y_2^N\Big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N)\in T_\epsilon} P\Big[x_{2a}^N, x_{2b}^N\Big] P\Big[y_2^N, u_{1c}^N\Big] \leqslant 2^{-N[I(X_{2a},X_{2b};Y_2,U_{1c})-\delta]} \quad (69)$$

From Equations (68) and (69), we require

$$R_{2a} < I(X_{2a}, X_{2b}; Y_2, U_c) \quad (70)$$

The probability of event E_2' is

$$P[\hat{W}_{2a}' \neq 1, \hat{W}_{2b}' \neq 1, \hat{W}_c' = 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} \sum_{w_{2b}=2}^{2^{NR_{2b}}} P\Big[\Big(U_{1c}^N(1,1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, w_{2b}), Y_2^N\Big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] \quad (71)$$

Following the same steps as in Equation (69) and using Equation (71), it can be shown that arbitrarily small $P[E_2']$ requires

$$R_{2a} + R_{2b} < I(X_{2a}, X_{2b}; Y_2, U_{1c}) \quad (72)$$

We next consider E_3':

$$P[\hat{W}_{2a}' \neq 1, \hat{W}_{2b}' = 1, \hat{W}_c' \neq 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR_c'}} P\Big[\Big(U_{1c}^N(w_c, b_c), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a},1), Y_2^N\Big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] \quad (73)(74)$$

We have

$$P\Big[(U_{1c}^N(w_c, b_c), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a},1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N)\in T_\epsilon} P\Big[x_{2a}^N, x_{2b}^N\Big] P\Big[u_{1c}^N\Big] P\Big[y_2^N\Big] \leqslant 2^{-N[I(X_{2a},X_{2b},U_{1c};Y_2)+I(U_{1c};X_{2a},X_{2b})-\delta]} \quad (75)$$

From Equations (74) and (75), it follows that

$$R_{2a} + R_c + R_c' < I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b}) \quad (76)$$

For E_4', we use the same approach as in Equation (74) and reuse Equation (75) to obtain

$$R_{2a} + R_{2b} + R_c + R_c' < I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b}) \quad (77)$$

We continue by considering error event E_5':

$$P[\hat{W}_{2a}' = 1, \hat{W}_{2b}' \neq 1, \hat{W}_c' = 1] = \sum_{w_{2b}=2}^{2^{NR_{2b}}} P\Big[\Big(U_{1c}^N(1,1), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N\Big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] \quad (78)$$

We have

$$P\Big[\Big(U_{1c}^N(1,1), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N\Big) \in T_\epsilon(P_{U_{1c}X_{2a},X_{2b},Y_2})\Big] = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N)\in T_\epsilon} P\Big[x_{2a}^N, x_{2b}^N\Big] P\Big[u_{1c}^N, y_2^N | x_{2a}^N\Big] \leqslant 2^{-N[I(X_{2b},U_{1c};Y_2|X_{2a})-\delta]} \quad (79)$$

From Equations (78) and (79), it follows that

$$R_{2b} < I(X_{2b}, U_{1c}; Y_2|X_{2a}) \quad (80)$$

For the error event E_6', we have

$$P[\hat{W}_{2a}' = 1, \hat{W}_{2b}' \neq 1, \hat{W}_c' \neq 1] = \sum_{w_{2b}=2}^{2^{NR_{2b}}} \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR_c'}} P\Big[\big(U_{1c}^N(w_c, b_c), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N\big) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})\Big] \quad (81)$$

Again

$$P\Big[\Big(U_{1c}^N(w_c, b_c), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N\Big) \in T_\epsilon\Big(P_{U_{1c}X_{2a}X_{2b}Y_2}\Big)\Big]$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

$$= \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N) \in T_\epsilon} P\left[x_{2a}^N, x_{2b}^N\right] P\left[u_{1c}^N\right] P\left[y_2^N | x_{2a}^N\right]$$

$$\leqslant 2^{-N[I(X_{2b}, U_{1c}; Y_2|X_{2a}) + I(U_{1c}; X_{2a}, X_{2b}) - \delta]} \quad (82)$$

From Equations (81) and (82), it follows that

$$R_{2b} + R_c + R_c' < I(X_{2b}, U_{1c}; Y_2|X_{2a}) + I(U_{1c}; X_{2a}, X_{2b}) \quad (83)$$

APPENDIX B: PROOF OF LEMMA 1

Proof (Lemma 1). For $I(S; Y, U) \leqslant R_s \leqslant H(S)$, use GP coding [14]. The achieved rate is

$$R \leqslant I(U; Y) - I(U; S) \quad (84)$$

Note that $I(U; Y) - I(U; S) \leqslant I(X; Y|S)$.

For $R_s < I(S; Y, U)$, proceed as follows.

Code construction: For every codeword $s^N(j)$, $j = 1, \ldots, 2^{NR_s}$ generate 2^{NR} codewords $u^N(w, j)$, $w = 1, \ldots, 2^{NR}$ using $\prod_{n=1}^N P_{U|S}(\cdot|s_n(j))$.

Encoder: Given w and $s^N(j)$, choose $u^N(w, j)$ and transmit $x^N = f^N(u^N(w, j), s^N(j))$.

Decoder: Given y^N, try to find (w, j) such that $(u^N(w, j), s^N(j), y^N) \in T_\epsilon(P_{SUY})$.

Analysis: Suppose $w = 1$, $j = 1$ was sent. Error $\{\hat{w} \neq 1\}$ occurs if $\{\hat{w} \neq 1, \hat{j} \neq 1\}$ or $\{\hat{w} \neq 1, \hat{j} = 1\}$. The probability of error is

$$P_e = \sum_{j=2}^{2^{NR_s}} \sum_{w=1}^{2^{NR}} P\left[\left(U^N(w, j), S^N(j), Y^N\right) \in T_\epsilon(P_{SUY})\right] + \sum_{w=2}^{2^{NR}} P\left[\left(U^N(w, 1), s^N(1), Y^N\right) \in T_\epsilon(P_{SUY})\right]$$

$$\leqslant 2^{N(R + R_s - I(US;Y) + \delta_1)} + 2^{N(R - I(U;Y|S) + \delta_2)} \quad (85)$$

where $\delta_1, \delta_2 \to 0$ as $N \to \infty$. From Equation (85) and $I(U; Y|S) = I(X; Y|S)$, it follows that

$$R \leqslant \min\{I(U, S; Y) - R_s, I(X; Y|S)\} \quad (86)$$

Note that we could have chosen $u^N = x^N$ in the superposition coding above, so that Equation (86) is

$$R \leqslant \min\{I(X, S; Y) - R_s, I(X; Y|S)\} \quad (87)$$

APPENDIX C: PROOF OF THEOREM 4

Proof (Theorem 4). Consider a code (M_1, M_2, N, P_e) for the IC with one cooperating encoder. We first consider the bound (33). Fano's inequality implies that for reliable communication we require

$$\begin{aligned}
N(R_1 + R_2) &\leqslant I\left(W_1; Y_1^N\right) + I\left(W_2; Y_2^N\right) \\
&\overset{(a)}{\leqslant} I\left(W_1; Y_1^N | W_2\right) + I\left(W_2; Y_2^N\right) \\
&= \sum_{i=1}^N I\left(W_1; Y_1^i | W_2, Y_{2,i+1}^N\right) - I\left(W_1; Y_1^{i-1} | W_2, Y_{2,i}^N\right) \\
&\quad + I\left(W_2; Y_{2,i} | Y_{2,i+1}^N\right) \\
&= \sum_{i=1}^N I\left(W_1; Y_1^i | W_2, Y_{2,i+1}^N\right) \\
&\quad - \left[I\left(W_1, Y_{2,i}; Y_1^{i-1} | W_2, Y_{2,i+1}^N\right)\right. \\
&\quad \left. - I\left(Y_{2,i}; Y_1^{i-1} | W_2, Y_{2,i+1}^N\right)\right] + I\left(W_2; Y_{2,i} | Y_{2,i+1}^N\right) \\
&\overset{(b)}{=} \sum_{i=1}^N I(W_1; Y_{1,i} | W_2, V_i) - I\left(Y_{2,i}; Y_1^{i-1} | W_1, W_2, Y_{2,i+1}^N\right) \\
&\quad + I\left(W_2, Y_1^{i-1}; Y_{2,i} | Y_{2,i+1}^N\right) \\
&\leqslant \sum_{i=1}^N I(W_1; Y_{1,i} | W_2, V_i) + I(W_2, V_i; Y_{2,i}) \quad (88)
\end{aligned}$$

where (a) follows from the independence of W_1, W_2; in (b), we let $Y_{t,i}^j = (Y_{t,i}, \ldots, Y_{t,j})$ and $V_i = \left[Y_1^{i-1}, Y_{2,i+1}^N\right]$.

We next consider the bound (31). Fano's inequality implies that reliable communication requires

$$NR_2 \leqslant I(W_2; Y_2^N) = \sum_{i=1}^N I(W_2; Y_{2,i} | Y_{2,i+1}^N)$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett

$$\leqslant \sum_{i=1}^{N} I(W_2, Y_1^{i-1}, Y_{2,i+1}^{N}; Y_{2,i})$$

$$= \sum_{i=1}^{N} I(W_2, V_i; Y_{2,i}) \tag{89}$$

Note that for Equations (88) and (89), we have used only the independence of W_1 and W_2, and the non-negativity of mutual information. The bounds (30) and (32) thus follow by symmetry.

We introduce random variables $U_{1,i} = W_1$ and $U_{2,i} = W_2$ for all i, and include a time-sharing random variable with V_i, to get the bounds in the form (30)–(33). Observe that $U_{1,i}$ and $U_{2,i}$ are independent. Furthermore, due to unidirectional cooperation, the joint probability distribution factors as in Equation (34).

APPENDIX D: PROOF OF THEOREM 5

Proof (Theorem 5). The bound (44) follows by standard methods. To prove Equation (45), consider Equation (33) and

$$\begin{aligned} I(U_1; Y_1|U_2, V) &\leqslant I(U_1; Y_1, X_2|U_2, V) \\ &= I(U_1; Y_1|U_2, V, X_2) \\ &\leqslant I(U_1, X_1; Y_1|U_2, V, X_2) \\ &= I(X_1; Y_1|U_2, V, X_2) \\ &\leqslant I(X_1; Y_2|U_2, V, X_2) \end{aligned} \tag{90}$$

where the second step follows by the Markov chain (Equation (34)), and the last step follows by Equation (43). We similarly have

$$I(V, U_2; Y_2) \leqslant I(U_2, V, X_2; Y_2) \tag{91}$$

Combining inequalities (33), (90) and (91) gives Equation (45).

ACKNOWLEDGEMENTS

The work by I. Marić and A. Goldsmith was supported in part from the DARPA ITMANET program under grant 1105741-1-TFIND, Stanford's Clean Slate Design for the Internet Program and the ARO under MURI award W911NF-05-1-0246. The work of G. Kramer was partially supported by the Board of Trustees of the University of Illinois Subaward No. 04-217 under NSF Grant No. CCR-0325673 and by ARO Grant W911NF-06-1-0182. The work of S. Shamai was supported by the EU 7th framework program via NEWCOM++, and by the ISRC Consortium.

REFERENCES

1. Sato H. Two user communication channels. *IEEE Transactions on Information Theory* 1977; **23**(3):295–304.
2. Carleial AB. Interference channels. *IEEE Transactions on Information Theory* 1978; **24**(1):60–70.
3. Devroye N, Mitran P, Tarokh V. Achievable rates in cognitive radio channels. *IEEE Transactions on Information Theory* 2006; **52**(5):1813–1827.
4. Mitola J. *Cognitive Radio Architecture*. John Wiley Sons, Inc.: New York, 1991.
5. Devroye N, Mitran P, Tarokh V. Limits on communications in a cognitive radio channel. In *IEEE Communications Magazine* 2006; **44**(6):44–49.
6. Marić I, Yates RD, Kramer G. Capacity of interference channels with partial transmitter cooperation. *IEEE Transactions on Information Theory* 2007; **53**(10):3536–3548.
7. Wu W, Vishwanath S, Arapostathis A. Capacity of a class of cognitive radio channels: interference channels with degraded message sets. *IEEE Transactions on Information Theory* 2007; **53**(11): 4391–4399.
8. Jovičić A, Viswanath P. Cognitive radio: an information-theoretic perspective. *IEEE Transactions on Information Theory* 2006 (submitted for publication). http://www.arxiv.org/pdf/cs.IT/0604107.pdf
9. Sridharan S, Vishwanath S. On the capacity of a class of MIMO cognitive radios. In *IEEE Information Theory Workshop (ITW 2007)*, Lake Tahoe, CA, September 2007.
10. Jiang J, Xin Y. On the achievable rate regions for interference channels with degraded message sets. *IEEE Transactions on Information Theory* 2007 (submitted for publication 2007).
11. Tuninetti D. The interference channels with generalized feedback. In *Proceedings of IEEE International Symposium on Information Theory*, Nice, France, June 2007.
12. Cao Y, Chen B, Zhang J. A new achievable rate region for interference channels with common information. In *Proceedings of IEEE Wireless Communications and Networking Conference*, Hong Kong, March 2007.
13. Devroye N, Vu M, Tarokh V. Achievable rates and scaling laws for cognitive radio channels. *EURASIP Journal on Wireless Communications and Networking* 2008. Vol. 2008, February 2008.
14. Gel'fand SI, Pinsker MS. Coding for channel with random parameters. *Problemy Peredachi Informatsii* 1980; **9**(1):19–31.
15. Nair C, Gamal AE. An outer bound to the capacity region of the broadcast channel. *IEEE Transactions on Information Theory* 2007; **53**(1):350–355.
16. Han T, Kobayashi K. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory* 1981; **27**(1): 49–60.
17. Costa MHM. Writing on dirty paper. *IEEE Transactions on Information Theory* 1983; **29**(3):439–441.
18. Weingarten H, Steinberg Y, Shamai (Shitz) S. The capacity region of the Gaussian multiple input multiple output broadcast channel. *IEEE Transactions on Information Theory* 2006; **52**(9):3936–3964.
19. Vishwanath S, Jindal N, Goldsmith A. Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channels. *IEEE Transactions on Information Theory* 2003; **49**(10):2658–2668.
20. Mitran P, Devroye N, Tarokh V. On compound channels with side-information at the transmitter. *IEEE Transactions on Information Theory* 2006; **52**(4):1745–1755.

Copyright © 2008 John Wiley & Sons, Ltd.

21. Khisti A, Erez U, Lapidoth A, Wornell GW. Carbon copying onto dirty paper. *IEEE Transactions on Information Theory* 2007; **53**(5):1814–1827.
22. Jiang J, Xin Y, Garg H. Interference channels with common information. *IEEE Transactions on Information Theory* 2008; **54**(1):171–187.
23. Liang Y, Somekh-Baruch A, Poor V, Shamai (Shitz) S, Verdú S. Cognitive interference channels with confidential messages. In *45th Annual Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, September 2007.
24. Cover T, Kim Y, Sutivong A. Channel capacity and state uncertainty reduction for state-dependent channels. In *Proceedings of IEEE International Symposium on Information Theory*, Nice, France, June 2007.
25. Cover T, Kim Y, Sutivong A. Simultaneous communication of data and state. http://arxiv.org/PScache/cs/pdf/0703/0703005v1.pdf, 2007.
26. Marić I, Goldsmith A, Kramer G, Shamai (Shitz) S. On the capacity of interference channels with a cognitive transmitter. In *Information Theory and Applications (ITA)*, UCSD, La Jolla, CA, January 2007. http://ita.ucsd.edu/workshop/07/files/paper/paper431.pdf
27. Lall S. *Advanced Topics in Computation for Control*. Lecture notes, Stanford University, 2004.
28. Willems FMJ. The discrete memoryless multiple access channel with cribbing encoders. *IEEE Transactions on Information Theory* 1985; **31**(3):313–327.
29. Cover T, Thomas J. *Elements of Information Theory*. John Wiley Sons, Inc.: New York, 1991.
30. Kramer G. *Multiple User Information Theory*. Lecture Notes, unpublished, 2006.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/ett