Reprinted from Proc. 21st Allerton Conference on Communication, Control and Computing, Monticello, IL. Oct. 5-7, 1983.

# A SIMPLIFIED TREATMENT OF WYNER'S WIRE-TAP CHANNEL

JAMES L. MASSEY
Institute of Telecommunications
Swiss Federal Institute of Technology
CH-8092 Zurich, Switzerland

ABSTRACT

Wyner introduced the wire-tap channel and found its capacity region, which he noted was not convex. This paper introduces a capacity pre-region that is convex, and whose convexity provides a simple proof that points outside Wyner's region cannot be achieved. It is further shown that Wyner's region is a union of simply structured subregions, and that any point in any of these subregions can be achieved by time-sharing the coding systems used to achieve the two critical points of that subregion.

## I. INTRODUCTION

Wyner, in a pioneering paper [1], introduced the wire-tap channel (WTC) and determined its capacity region. The WTC is the one-sender two-receiver channel formed by the cascade of two discrete memoryless channels (DMC's) as shown in Fig.1. The channel structure coincides with that of a degraded broadcast channel [2], but the goal is very different. As shown in Fig.2, the sender seeks to transmit a binary symmetric source (BSS) to the legitimate receiver over the main channel ($DMC_M$) while denying this information as much as possible to the wiretapper, who obtains a noisy version of the received signal available to the legitimate receiver over the channel $DMC_W$. It is assumed that the legitimate sender knows both $DMC_M$ and $DMC_W$, as does also the wiretapper who also knows the code used by the legitimate sender.

Wyner characterized the situation of Fig.2 by the rate, $R = K/N$, at which information is sent to the legitimate receiver, and by the wiretapper's per digit equivocation, $\Delta = H(U^K|Z^N)/K$, about this information. (Our notation here and hereafter is that of Gallager [3].) Wyner defined a rate-equivocation pair, $(r,d)$, to be achievable if $r \geq 0$, $d \geq 0$, and, for every given positive number $\varepsilon$, one can find a blocklength N and a coding system such that $R \geq r - \varepsilon$, $\Delta \geq d - \varepsilon$, and $P_b \leq \varepsilon$, where

$$P_b = \frac{1}{K} \sum_{i=1}^{K} P(\hat{U}_i \neq U_i) \tag{1}$$

is the bit error probability for the legitimate receiver. The capacity region, $\mathcal{C}$, is the set of all achievable rate-equivocation pairs. Wyner proved that



$$\mathcal{C} = \{ (r,d): r \geq 0,\ 0 \leq d \leq 1, \text{ and there exists } p_o(x) \text{ such that } r \leq I_o(X;Y) \text{ and } rd \leq I_o(X;Y|Z) \}, \tag{2}$$

and he noted that, unlike the capacity regions of all previously considered multi-user channels, $\mathcal{C}$ was not a convex region.

Our goal in this paper is to provide a simplified proof that (2) gives the capacity region of the WTC. We do this by introducing a capacity pre-region, $\mathcal{W}$, that uniquely determines $\mathcal{C}$ and vice-versa, but that is convex. The convexity of $\mathcal{W}$ is then exploited in the proof that $\mathcal{C}$ is indeed the capacity region.

We shall have need for Fano's inequality [3, p.79] which asserts that when $U^K$ is estimated from an observation $W^N$ then the bit error probability satisfies

$$h(P_b) \geq \frac{1}{K} H(U^K|W^N), \tag{3}$$

where $h(p) = -p \log p - (1-p) \log (1-p)$ is the binary entropy function. We shall also make use of the result, which was not available to Wyner, that for the WTC, both $I(X;Y)$ and $I(X;Y|Z)$ are concave functions of the input probability distribution $P_X$, as shown by Leung-Yan-Cheong [4] with an elegantly simple argument. Finally, we shall need the fact that, for the WTC,

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|YZ) \\ &= H(X|Z) - H(X|Y) \\ &= I(X;Y) - I(X;Z), \end{aligned} \tag{4}$$

as follows from the fact that X,Y,Z is a Markov Chain.

## II. THE CAPACITY PRE-REGION

We define the capacity pre-region, $\mathcal{W}$, as

$$\mathcal{W} = \{(r,\gamma): r \geq 0,\ \gamma \geq 0 \text{ and there exists } p_o(x) \text{ such that } 0 \leq r \leq I_o(X;Y) \text{ and } 0 \leq \gamma \leq I_c(X;Y|Z) \}. \tag{5}$$

A comparison of (2) and (5) shows that $(r,\gamma)$ is in $\mathcal{W}$ if and only if $(r,r\gamma)$ is in $\mathcal{C}$. Moreover, $\mathcal{W}$ is a convex region as we now prove.

Suppose that $(r_1,\gamma_1)$ and $(r_2,\gamma_2)$ are pairs in $\mathcal{W}$ corresponding to $p_1(x)$ and $p_2(x)$, respectively. We must show that for every $\theta$, $0 \leq \theta \leq 1$, the pair $(r,\gamma) = (\theta r_1 + (1-\theta)r_2,\ \theta\gamma_1 + (1-\theta)\gamma_2)$ is also in $\mathcal{W}$. Clearly, $r \geq 0$

and $\gamma \geq 0$. We now choose

$$p_o(x) = \theta p_1(x) + (1-\theta) p_2(x).$$

By the concavity of $I(X;Y)$ and $I(X;Y|Z)$, it follows that

$$\begin{aligned} I_o(X;Y) &\geq \theta I_1(X;Y) + (1-\theta) I_2(X;Y) \\ &\geq \theta r_1 + (1-\theta) r_2 = r \end{aligned}$$

and

$$\begin{aligned} I_o(X;Y|Z) &\geq \theta I_1(X;Y|Z) + (1-\theta) I_2(X;Y|Z) \\ &\geq \theta \gamma_1 + (1-\theta)\gamma_2 = \gamma, \end{aligned}$$

which proves that $(r,\gamma)$ is indeed in $\mathcal{W}$.

From (5), we see that every input probability distribution, $p_o(x)$, determines the rectangular subregion of $\mathcal{W}$ shown in Fig. 3(a). Moreover, this rectangular region determines, and is determined by, the subregion of $\mathcal{C}$, corresponding to $p_o(x)$, shown in Fig.3(b). $\mathcal{W}$ is the union of all subregions as in Fig.3(a), just as $\mathcal{C}$ is the union of those in Fig.3(b). The interesting segment of the boundary of the capacity subregion of Fig.3(b) occurs over the rate interval

$$I_o(X;Y) - I_o(X;Z) \leq r \leq I_o(X;Y) \tag{6a}$$

for which the equivocation is given by the hyperbola

$$rd = I_o(X;Y) - I_o(X;Z) \tag{6b}$$

as follows from (2) and (4).

## III. THE NECESSITY ARGUMENT

We now wish to show that, for $(R,\Delta)$ to be achievable, a necessary condition is that $(R,\Delta)$ be in $\mathcal{C}$ or, equivalently, that $(R,R\Delta)$ be in $\mathcal{W}$. The following result is the key to this demonstration.

Lemma 1: For the coding scheme of Fig.2, there exists $p_o(x)$ such that

$$I(X^N;Y^N) \leq N I_o(X;Y) \tag{7a}$$

and

$$I(X^N;Y^N|Z^N) \leq N I_o(X;Y|Z). \tag{7b}$$

Remark: Because of the non-negativity of mutual information, Lemma 1 is equivalent to the statement that $(r,\gamma)$ is in $\mathcal{W}$, where $r = I(X^N;Y^N)/N$ and $\gamma = I(X^N;Y^N|Z^N)/N$.

Proof: By the well-known inequality [3, p.75] for a DMC,

$$\frac{1}{N} I(X^N;Y^N) \leq \sum_{i=1}^{N} \frac{1}{N} I(X_i;Y_i). \tag{8}$$



Thus, defining

$$p_o(x) = \sum_{i=1}^{N} \frac{1}{N} p_i(x) \tag{9}$$

where $p_i(x)$ is the probability distribution for $X_i$, we have by the concavity of $I(X;Y)$ that

$$I_o(X;Y) \geq \sum_{i=1}^{N} \frac{1}{N} I(X_i;Y_i). \tag{10}$$

Combining (8) and (10) now gives (7a).

By the concavity also of $I(X;Y|Z)$ for the WTC, we see that an entirely similar argument will give (7b), provided we can establish the inequality corresponding to (8), namely

$$\frac{1}{N} I(X^N;Y^N|Z^N) \leq \sum_{i=1}^{N} \frac{1}{N} I(X_i;Y_i|Z_i). \tag{11}$$

Inequality (11) was proved by Wyner [1] as follows:

$$\begin{aligned} I(X^N;Y^N|Z^N) &= H(Y^N|Z^N) - H(Y^N|X^N Z^N) \\ &= \sum_{i=1}^{N} \left[H(Y_i|Y^{i-1}Z^N) - H(Y_i|Y^{i-1}X^N Z^N)\right] \\ &\leq \sum_{i=1}^{N} \left[H(Y_i|Z_i) - H(Y_i|Y^{i-1}X^N Z^N)\right] = \\ &= \sum_{i=1}^{N} \left[H(Y_i|Z_i) - H(Y_i|X_i Z_i)\right] \\ &= \sum_{i=1}^{N} I(X_i;Y_i|Z_i), \end{aligned}$$

where the fact that $H(Y_i|Y^{i-1}X^N Z^N) = H(Y_i|X_i Z_i)$ follows from the fact that, given $X_i$ and $Z_i$, $Y_i$ has no further dependence on any of the digits $X$, $Y$, $Z$ as can be seen from Fig.2.

We now prove the following result, of some independent interest, which immediately implies that if $(R,\Delta)$ is achievable, then $(R,R\Delta)$ is in $\mathcal{W}$.

Theorem 1: For the coding scheme of Fig.(2),

$$h(P_b) \geq 1 - I_o(X;Y)/R \tag{12a}$$

and

$$R\Delta \leq I_o(X;Y|Z) + Rh(P_b), \tag{12b}$$

where $p_o(x)$ is defined by (9).

Proof: We begin by noting that

$$\begin{aligned} H(U^K|Y^N) &= H(U^K) - I(X^N;Y^N) \\ &= K - I(X^N;Y^N) \\ &\geq K - N I_o(X;Y) \end{aligned}$$

where we have used the BSS entropy property and inequality (7a). Dividing by K and using (3) now gives (12a).

Next, we note that $U^K, Y^N, Z^N$ form a Markov chain so that

$$I(U^K;Y^N|Z^N) = H(U^K|Z^N) - H(U^K|Y^N Z^N)$$
$$= H(U^K|Z^N) - H(U^K|Y^N).$$

Again invoking Fano's inequality (3), we have thus

$$h(P_b) \geq \frac{1}{K} H(U^K|Y^N) =$$
$$= \frac{1}{K}\left[H(U^K|Z^N) - I(U^K;Y^N|Z^N)\right]$$
$$= \frac{1}{K}\left[K\Delta - I(U^K;Y^N|Z^N)\right]$$
$$\geq \Delta - \frac{1}{K} I(X^N,Y^N|Z^N)$$
$$\geq \Delta - \frac{N}{K} I_o(X;Y|Z)$$

where the last inequality follows from (7b), and where the preceding inequality follows from the fact that $U^K, X^N, Y^N, Z^N$ form a Markov chain. Multiplying by $R = K/N$ and rearranging now gives (12b).

## IV. THE SUFFICIENCY ARGUMENT

It remains to show that if (R,d) is in $\mathcal{C}$, then (R,d) can be achieved. But our discussion in Section II showed that if (R,d) is in $\mathcal{C}$, then (R,d) is also in the capacity subregion of Fig.3(b) for some $p_o(x)$. Hence, it suffices to show that if (R,d) is a point in such a subregion, then (R,d) can be achieved. It is obvious that if we can achieve any point on the outer boundary (that part of the boundary not coinciding with the coordinate axes), then we can achieve any point in the subregion. The two critical points on this outer boundary are the endpoints of the hyperbolic segment. It is obvious that if we can achieve these two points, we can also achieve the points on the linear portions of the outer boundary. The following result, however, is more surprising.

Theorem 2: Any point on the hyperbolic segment of the boundary of the capacity subregion in Fig.3(b) can be achieved by time-sharing of the coding schemes that achieve the two endpoints of this segment, $(r_1,d_1)$ and $(r_2,d_2)$.

Proof: Let (r,d) be the pair to be achieved; then $r = (1-\lambda)r_1 + \lambda r_2$ where $0 < \lambda < 1$. Without loss of essential generality, we may suppose that $\lambda$ is a rational number, say $\lambda = m/n$.



Let $\varepsilon > 0$ be specified. Because $(r_1,d_1)$ and $(r_2,d_2)$ are achievable, we can find coding schemes with the same blocklength $N_{12}$ that achieve these points in the manner

$$P_{b1} \leq \varepsilon \text{ and } P_{b2} \leq \varepsilon \tag{13a}$$

$$r_1 \geq R_1 \geq r_1 - \delta \text{ and } r_2 \geq R_2 \geq r_2 - \delta \tag{13b}$$

$$\Delta_1 \geq d_1 - \delta \text{ and } \Delta_2 \geq d_2 - \delta \tag{13c}$$

where $\delta$ satisfies $0 < \delta \leq \varepsilon$ and will be specified later. Now consider the "time-shared" coding system of blocklength $N = n\,N_{12}$ such that the first and second coding schemes are used n - m and m times, respectively, in one block. Note that fractions $1 - \lambda$ and $\lambda$ of the N uses of the WTC are made with the first and second coding system, respectively.

The bit error probability of the time-shared system certainly satisfies

$$P_b \leq \max(P_{b1},P_{b2}) \leq \varepsilon \tag{14a}$$

as required. The rate of the time-shared system is

$$R = (1-\lambda)R_1 + \lambda R_2$$

so that from (13b) we have

$$r \geq R \geq r - \delta \geq r - \varepsilon \tag{14b}$$

as desired. Finally, we note that the fraction of information bits transmitted via the first and second coding system is $(1-\lambda)R_1/R$ and $\lambda R_2/R$, respectively. Thus, the wiretapper's equivocation for the time-shared system is

$$\Delta = (1-\lambda)R_1\Delta_1/R + \lambda R_2\Delta_2/R$$
$$\geq (1-\lambda)(r_1-\delta)(d_1-\delta)/r + \lambda(r_2-\delta)(d_2-\delta)/r$$
$$\geq (1-\lambda)(r_1 d_1 - r_1\delta - d_1\delta)/r + \lambda(r_2 d_2 - r_2\delta - d_2\delta)/r. \tag{15}$$

But now we note that $(r,d), (r_1,d_1)$ and $(r_2,d_2)$ are <u>all</u> on the hyperbolic segment of the capacity subregion boundary so that (6b) implies

$$rd = r_1 d_1 = r_2 d_2. \tag{16}$$

Using (16) in (15) and noting that $d_1 \leq 1$ and $d_2 \leq 1$ now gives

$$\Delta \geq d - (r+2)\delta/r \geq d - \varepsilon \tag{14c}$$

as desired, provided we ensure that the free parameter $\delta$ also satisfies $\delta \leq r\varepsilon/(r+2)$. Equations (14a) - (14c) show that the time-shared system achieves (r,d).

To complete the proof that $\mathcal{C}$ is indeed the capacity region of the WTC, it remains only to show that the rate-equivocation pairs of Theorem 2, namely

$$(r_1,d_1) = (I_o(X;Y),\ 1 - I_o(X;Z)/I_o(X;Y)) \tag{17a}$$

and

$$(r_2,d_2) = (I_o(X;Y) - I_o(X;Z),\ 1) \tag{17b}$$

can be achieved.

The pair $(r_1,d_1)$ is easy to obtain. A standard random coding scheme (cf. [3,pp.131-141]) in which each digit of each codeword is chosen independently according to $p_o(x)$ will, for codes of rate $R = r_1 - \delta = I_o(X;Y)-\delta$, yield

$$E[P_b] \leqslant \frac{\varepsilon^2}{2(1-d_1-\varepsilon)} \tag{18a}$$

for all sufficiently large N, where this and all subsequent explicit expectations are over the codes in the randomly selected ensemble, where we restrict $\varepsilon$ henceforth to satisfy $\varepsilon < 1 - d_1$, and where $\delta$ is a parameter satisfying $0 < \delta \leqslant \varepsilon$ that we specify below. Moreover, over this ensemble of codes, we see that

$$\begin{aligned} E[\Delta] &= E[H(U^K|Z^N)/K] \\ &= E[H(U^K)/K - I(U^K;Z^N)/K] \\ &\geqslant 1 - E[I(X^N;Z^N)/K] = \\ &\geqslant 1 - I_o(X;Z)/R \\ &> 1 - I_o(X;Z)/I_o(X;Y) - \varepsilon/2 = d_1 - \varepsilon/2, \end{aligned} \tag{18b}$$

where the last inequality requires us to specify $\delta$ appropriately small. The fact that $P_b \geqslant 0$, together with (18a), implies

$$P[P_b \geqslant \varepsilon] < \frac{\varepsilon/2}{1-d_1-\varepsilon}\ . \tag{19a}$$

Similarly, the fact that $\Delta \leqslant 1$, together with (18b), implies

$$P[\Delta \leqslant d - \varepsilon] \leqslant \frac{1-\varepsilon/2}{1-d_1-\varepsilon}\ . \tag{19b}$$

But (19a) and (19b) now imply that for at least one code in the ensemble <u>both</u> $P_b < \varepsilon$ <u>and</u> $\Delta > d_1-\varepsilon$. Since the code rate is $R = r_1 - \delta \geqslant r_1-\varepsilon$, it follows that $(r_1,d_1)$ is achievable.

It is only to achieve the pair $(r_2,d_2)$ of (17b) that we must resort to the non-obvious (and extremely clever!) coding scheme introduced by Wyner[1]. In this scheme, the "encoder" of Fig.2 now contains both an internal BSS as well as an internal encoder. The real BSS output $U^K$ together with the internal BSS output $V^L$ form the input to the internal encoder. We choose $R = K/N = r_2 - \delta$ (where again $\delta$ satisfies $0 < \delta \leqslant \varepsilon$ and will be specified later), but we also choose $R^* = L/N = I_o(X;Z) - \delta$. We then consider the same random coding scheme as described above except that the code rate is now



$R + R^* = I_o(X;Y) - 2\delta$. Thus, for sufficiently large N, the legitimate receiver will decode both sources, and in particular the real BSS, with

$$E[P_b] \leqslant \varepsilon/2. \tag{20}$$

The key part of the argument is now to note that if $U^K$ were available to the wire-tapper, then only $2^L$ of the $2^{K+L}$ codewords would remain as possibilities so that the coding scheme would then appear to him to be exactly a random coding scheme of rate $R^* = I_o(X;Z)$, and hence he could (whether he wishes to or not!) decode $V^L$ with average bit error probability at most $\delta$ for N sufficiently large. But Fano's inequality thus requires

$$E[H(V^L|U^KZ^N)]/L \leqslant h(\delta) \tag{21}$$

which as we shall soon see, dooms the wiretapper to complete ignorance about $U^K$. To show this, we note that

$$\begin{aligned} E[I(U^KV^L;Z^N)] &= E[H(U^KV^L) - H(U^KV^L|Z^N)] \\ &= K + L - E[H(U^K|Z^N)] - E[H(V^L|U^KZ^N)] \end{aligned} \tag{22}$$

and also that

$$E[I(U^KV^L;Z^N)] \leqslant E[I(X^N;Z^N)] \leqslant N\ I_o(X;Z). \tag{23}$$

Combining (22) and (23) and noting that $\Delta = H(U^K|Z^N)/K$ gives

$$E[\Delta] \geqslant 1 + L/K - I_o(X;Z)/R - E[H(V^L|U^KZ^N)]/K,$$

which with the aid of (21) and the fact that $L/K = R^*/R = [I_o(X;Z)-\delta]/R$ now gives

$$E[\Delta] \geqslant 1 - \delta/R - h(\delta)R^*/R.$$

Thus, for sufficiently small $\delta$, we have

$$E[\Delta] > 1 - \varepsilon/2. \tag{24}$$

The fact that $P_b \geqslant 0$, together with (20), implies that $P[P_b \geqslant \varepsilon] \leqslant 1/2$. The fact that $\Delta \leqslant 1$, together with (24), implies that $P(\Delta \geqslant 1 - \varepsilon) < 1/2$. Thus, at least one code in the ensemble must have both $P_b < \varepsilon$ and $\Delta > 1 - \varepsilon = d_2 - \varepsilon$. This code has rate $R = r_2 - \delta \geqslant r_2 - \varepsilon$ so it follows that $(r_2,d_2)$ is achievable.

## REFERENCES


[1] A.D.Wyner, "The Wire-Tap Channel", <u>Bell Sys.Tech.J.</u>, vol. 54, pp.1355-1387, Oct. 1975.

[2] T.M.Cover, "Broadcast Channels", <u>IEEE Trans.Info.Th.</u>, vol. IT-18, pp.2-14, Jan. 1972.

[3] R.G.Gallager, <u>Information Theory and Reliable Communication.</u> New York: Wiley, 1968.

[4] S.K.Leung-Yan-Cheong, "On a Special Class of Wiretap Channels", <u>IEEE Trans.Info.Th.</u>, vol. IT-23, pp.625-627, Sept. 1977.

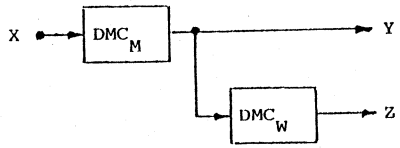


Fig.1: Wyner's Wire-Tap Channel (WTC).

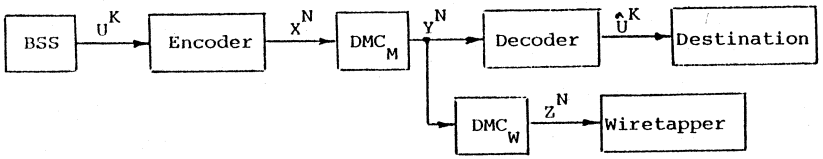


Fig.2: Coding Situation for the WTC.

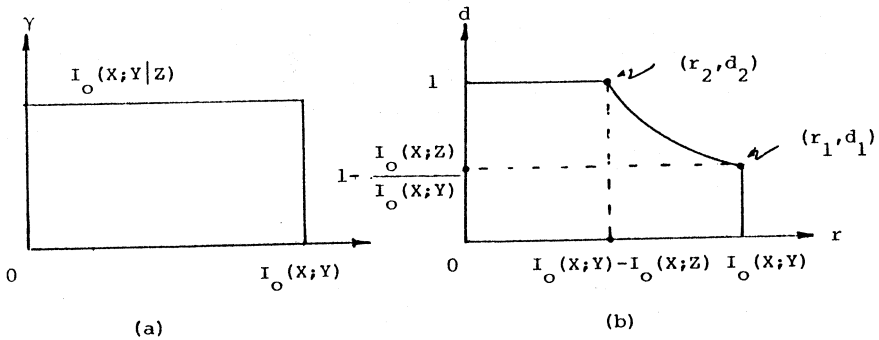


Fig.3: The Subregion (a) of $\mathcal{W}$ and the Subregion (b) of $\mathcal{C}$ Determined by $p_o(x)$.