

$x^n$  for which

and a TR-code with a finite rate and

$\mathcal{T}_{[1]}^n \rightarrow C_i$ , is achievable (i.e.  $M_i$  satisfying

achieves his coding rate and the total

$y_{i,2}^n$ , say.

distortion in the

$\bar{y}_1$  was sent or

$y_1^n$  accurately,

on  $\bar{m}_1$  if both

$\frac{1}{m_1}(y_1^n) = y_1^{\sqrt{n}}$ ;

order 2 behaves

and since the

error probabilities

are small. This

can transform a

set of codewords

to  $P$  only. This

leads to a new code, (7)

to appear in

the presence of feedback

in *IEEE Trans.*

in *IEEE Trans.*

in 1978.

## A proof of the coding theorem for the additive white Gaussian noise channel in terms of jointly typical sequences

Frans M.J. Willems\*

*Achievability proofs for additive white Gaussian noise channels are often proved by handwaving. Here we give a rigorous achievability proof for the single-input single-output channel in terms of typical sequences. This approach is generalizable to multi-user situations.*

### I. DEFINITIONS

An additive white Gaussian noise (AWGN) channel is a time-discrete memoryless channel with input  $X \in \mathbb{R}$  and output  $Y \in \mathbb{R}$ . The conditional probability density function of  $y$  given  $x$  is given by

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(y-x)^2/2\sigma^2), \quad (1)$$

hence  $Y-X$  is Gaussian, with mean 0 and variance  $\sigma^2$  (and independent of  $X$ ). The memorylessness of the channel follows from  $p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = \prod_{n=1}^N p_{Y|X}(y_n|x_n)$ , where  $\underline{x} := (x_1, x_2, \dots, x_N)$  is an input sequence and  $\underline{y} := (y_1, y_2, \dots, y_N)$  the corresponding output sequence.

An  $(M, N)$  code for the AWGN channel consists of a message set  $\{1, 2, \dots, M\}$ , the codewords  $\underline{x}(1), \underline{x}(2), \dots, \underline{x}(M)$ , one corresponding to each message, and a decoding function  $D: \mathbb{R}^N \rightarrow \{1, 2, \dots, M\}$ . We assume that the messages are uniformly distributed. The average probability of error  $P_e$  and the maximal signal energy  $S$  of the code are defined as

$$P_e := \frac{1}{M} \sum_{m=1}^M \Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\}, \quad (2a)$$

$$E := \max_{m=1, M} |\underline{x}(m)|^2, \text{ where } |\underline{x}|^2 := \sum_{n=1}^N x_n^2. \quad (2b)$$

\*Eindhoven University of Technology, Electrical Engineering Department, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

A rate  $R$  is said to be  $\mathcal{P}$ -achievable for the AWGN channel if, for every  $\delta > 0$  there exist, for all  $N$  large enough,  $(M, N)$  codes with  $M \geq \exp(N(R-\delta))$ ,  $P_e \leq \delta$ , and  $E \leq N\mathcal{P}(1+\delta)$ . The  $\mathcal{P}$ -capacity of the AWGN channel is the maximum of all  $\mathcal{P}$ -achievable rates. In this contribution we show that the  $\mathcal{P}$ -capacity is at least  $1/2 \cdot \ln(1+\mathcal{P}/\sigma^2)$ . Our proof is based on the concept of Gaussian  $\epsilon$ -typicality which is developed here.

## II. GAUSSIAN $\epsilon$ -TYPICAL SEQUENCES

Let  $X$  and  $Y$  be jointly Gaussian with means 0, variances  $a^2$  and  $b^2$  respectively, and covariance  $\lambda$ , i.e.

$$p_{X,Y}(x,y) := \frac{1}{2\pi\sqrt{a^2b^2-\lambda^2}} \exp\left(-\frac{x^2b^2+y^2a^2-2xy\lambda}{2(a^2b^2-\lambda^2)}\right). \quad (3)$$

Then for the entropies  $h(X)$ ,  $h(Y)$ , and  $h(X,Y)$  we have that

$$h(X) := -\overline{\ln(p_X(x))} = 1/2 \cdot \ln(2\pi ea^2), \quad (4a)$$

$$h(Y) := -\overline{\ln(p_Y(y))} = 1/2 \cdot \ln(2\pi eb^2), \text{ and} \quad (4b)$$

$$h(X,Y) := -\overline{\ln(p_{X,Y}(x,y))} = 1/2 \cdot \ln((2\pi e)^2 \cdot (a^2b^2-\lambda^2)). \quad (4c)$$

Now let  $p_{\underline{X}}(\underline{x}) := \prod_{n=1}^N p_X(x_n)$ ,  $p_{\underline{Y}}(\underline{y}) := \prod_{n=1}^N p_Y(y_n)$ , and  $p_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) := \prod_{n=1}^N p_{X,Y}(x_n,y_n)$ .

**DEFINITION :** For a fixed  $N$  and  $\epsilon > 0$  the set  $\mathcal{G}_\epsilon(X,Y)$  of Gaussian jointly  $\epsilon$ -typical sequences of length  $N$  is now defined as

$$\begin{aligned} \mathcal{G}_\epsilon(X,Y) := \{(\underline{x},\underline{y}) \in \mathbb{R}^N \times \mathbb{R}^N : & |-\ln(p_{\underline{X}}(\underline{x})) - Nh(X)| \leq N\epsilon \cup \\ & |-\ln(p_{\underline{Y}}(\underline{y})) - Nh(Y)| \leq N\epsilon \cup \\ & |-\ln(p_{\underline{X},\underline{Y}}(\underline{x},\underline{y})) - Nh(X,Y)| \leq N\epsilon\}. \end{aligned} \quad (5)$$

The **MAIN PROPERTY** of  $\mathcal{G}_\epsilon(X,Y)$  is that

$$\Pr\{(\underline{X},\underline{Y}) \notin \mathcal{G}_\epsilon(X,Y)\} \leq \epsilon \text{ for all } N \text{ large enough.} \quad (6)$$

PROOF :

is Gaussian

$(-\ln(p_{\underline{Y}}(\underline{Y}))$

where the

the varian

Using Che

and  $\Pr\{-$

$p_{X,Y}(x,y)$

and that

follows th

the indep

have va

$\Pr\{-\ln(p$

Let

probabili

$p_{Y|X}(\cdot|$

$:= h(X)$

**PROOF :** When  $U$  has density function  $p_U(u) := \exp(-u^2/2v^2)/\sqrt{2\pi v^2}$ ,  $u \in \mathbb{R}$  (i.e.  $U$  is Gaussian with mean 0 and variance  $v^2$ ), we obtain for the variance of  $-\ln(p_U(U))$

$$\begin{aligned} \overline{(-\ln(p_U(U)) - h(U))^2} &= \overline{(U^2/2v^2 + 1/2 \cdot \ln(2\pi v^2) - 1/2 \cdot \ln(2\pi e v^2))^2} \\ &= \overline{(U^2/2v^2 - 1/2)^2} = \overline{U^4}/4v^4 - \overline{U^2}/2v^2 + 1/4 \\ &= 3/4 - 1/2 + 1/4 = 1/2, \end{aligned} \quad (*)$$

where the last step follows from  $\overline{U^4} = 3v^4$  and  $\overline{U^2} = v^2$ . It is important to note that the variance of  $-\ln(p_U(U))$  is independent of  $v^2$ .

Using Chebyshev's inequality we find that  $\Pr\{|-\ln(p_X(X)) - Nh(X)| \leq N\epsilon\} \leq 1/2N\epsilon^2$  and  $\Pr\{|-\ln(p_Y(Y)) - Nh(Y)| \leq N\epsilon\} \leq 1/2N\epsilon^2$ . Next observe that

$$p_{X,Y}(x,y) := \frac{1}{\sqrt{2\pi a^2}} \cdot \exp(-x^2/2a^2) \cdot \frac{1}{\sqrt{2\pi b^2(1-\lambda^2/a^2b^2)}} \cdot \exp\left(-\frac{(y-\lambda x/a^2)^2}{2b^2(1-\lambda^2/a^2b^2)}\right)$$

and that  $Y-\lambda X/a^2$  is Gaussian, with mean 0, and independent of  $X$ . From (\*) it follows that both  $-\ln(p_X(X))$  and  $-\ln(p_{Y-\lambda X/a^2}(Y-\lambda X/a^2))$  have variance 1/2. By the independence  $-\ln(p_{X,Y}(X,Y)) = -\ln(p_X(X)) - \ln(p_{Y-\lambda X/a^2}(Y-\lambda X/a^2))$  must have variance 1. Again applying Chebyshev's inequality yields that  $\Pr\{|-\ln(p_{X,Y}(X,Y)) - Nh(X,Y)| \leq N\epsilon\} \leq 1/N\epsilon^2$ . Using the union bound we finally get

$$\Pr\{(X,Y) \notin \mathcal{G}_\epsilon(X,Y)\} \leq 1/2N\epsilon^2 + 1/2N\epsilon^2 + 1/N\epsilon^2 = 2/N\epsilon^2. \quad \square$$

### III. RANDOM CODING ARGUMENT

Let the  $X$  be a Gaussian random variable with mean 0 and variance  $\mathcal{P}$  (and probability density function  $p_X(\cdot)$ ). Let  $p_{X,Y}(x,y) := p_X(x)p_{Y|X}(y|x)$  with  $p_{Y|X}(\cdot|\cdot)$  as in (1). The mutual information between  $X$  and  $Y$  is defined as  $I(X;Y) := h(X) + h(Y) - h(X,Y)$ . Note that  $Y$  has variance  $\mathcal{P} + \sigma^2$  and that the covariance of

$X$  and  $Y$  is equal to  $\mathcal{P}$ . Therefore  $I(X;Y) = 1/2 \cdot \ln(2\pi e\mathcal{P}) + 1/2 \cdot \ln(2\pi e(\mathcal{P} + \sigma^2)) - 1/2 \cdot \ln((2\pi e)^2 \cdot (\mathcal{P} + \sigma^2) \cdot \mathcal{P}^2) = 1/2 \cdot \ln(1 + \mathcal{P}/\sigma^2)$ . We will now show that for any  $\epsilon > 0$  there exists an  $(M,N)$  code with  $M = \exp(N(I(X;Y) - 4\epsilon))$  and  $P_e \leq 2\epsilon$ .

**RANDOM CODING** (adapted from El Gamal and Cover [1]) : Fix  $N$ . Generate  $M$  i.i.d. sequences  $\underline{x}(1), \underline{x}(2), \dots, \underline{x}(M)$  of length  $N$  each with probability  $p_{\underline{X}}(\underline{x}) := \prod_{n=1}^N p_X(x_n)$ . The encoder sends sequence  $\underline{x}(m)$  when message  $m$  is to be transmitted. The decoder upon receiving  $\underline{y}$  decodes the unique  $\hat{m}$  for which  $(\underline{x}(\hat{m}), \underline{y}) \in \mathcal{G}_\epsilon(X,Y)$ . When there is no such  $\hat{m}$  an error is declared.

We evaluate the average error probability averaged over the ensemble of codes :

$$\begin{aligned} P_e &= \sum_{\text{all codes}} \Pr\{\text{code}\} \cdot \frac{1}{M} \sum_{m=1}^M \Pr\{\underline{Y} \notin D(m) | \underline{X}(m)\} \\ &= \frac{1}{M} \sum_{m=1}^M \sum_{\text{all codes}} \Pr\{\text{code}\} \cdot \Pr\{\underline{Y} \notin D(m) | \underline{X}(m)\} \\ &= \frac{1}{M} \sum_{m=1}^M \Pr\{E_m^c \cup [m' \neq m, E_{m'}]\} \\ &\leq \frac{1}{M} \sum_{m=1}^M \Pr\{E_m^c\} + \frac{1}{M} \sum_{m=1}^M \sum_{m' \neq m} \Pr\{E_{m'}\}, \end{aligned} \quad (7)$$

where the event  $E_m := \{(\underline{X}(m), \underline{Y}) \in \mathcal{G}_\epsilon(X,Y)\}$ , with  $\underline{Y}$  generated by  $\underline{X}(m)$ .

From the main property of Gaussian  $\epsilon$ -typical sequences it follows that for all  $N$  large enough  $\Pr\{E_m^c\} \leq \epsilon$ . Furthermore using the definition of  $\mathcal{G}_\epsilon(X,Y)$  we obtain that

$$\begin{aligned} &\Pr\{E_{m'}\} \\ &= \int_{\underline{x} \in \mathbb{R}^N} \int_{\underline{y} \in \mathbb{R}^N} \int_{\underline{x}' \in \mathbb{R}^N : (\underline{x}', \underline{y}) \in \mathcal{G}_\epsilon(X,Y)} p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y}) \cdot p_{\underline{X}}(\underline{x}') \, d\underline{x} d\underline{y} d\underline{x}' \\ &= \int_{\underline{y} \in \mathbb{R}^N} \int_{\underline{x}' \in \mathbb{R}^N : (\underline{x}', \underline{y}) \in \mathcal{G}_\epsilon(X,Y)} p_{\underline{Y}}(\underline{y}) \cdot p_{\underline{X}}(\underline{x}') \, d\underline{y} d\underline{x}' \\ &= \int_{\underline{x} \in \mathbb{R}^N} \int_{\underline{y} \in \mathbb{R}^N : (\underline{x}, \underline{y}) \in \mathcal{G}_\epsilon(X,Y)} \frac{p_{\underline{X}}(\underline{x}) \cdot p_{\underline{Y}}(\underline{y})}{p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y})} \cdot p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y}) \, d\underline{x} d\underline{y} \\ &= \int_{\underline{x} \in \mathbb{R}^N} \int_{\underline{y} \in \mathbb{R}^N : (\underline{x}, \underline{y}) \in \mathcal{G}_\epsilon(X,Y)} \frac{\exp(-N(h(X) - \epsilon)) \cdot \exp(-N(h(Y) - \epsilon))}{\exp(-N(h(X, Y) + \epsilon))} \cdot p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y}) \, d\underline{x} d\underline{y} \end{aligned}$$

$$\leq \exp(-N(I(X;Y)-3\epsilon)) \cdot \int_{\underline{x} \in \mathbb{R}^N} \int_{\underline{y} \in \mathbb{R}^N} P_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y}) \, d\underline{x} d\underline{y} = \exp(-N(I(X;Y)-3\epsilon)). \quad (8)$$

Substitution in (7) gives us (for all  $N$  large enough)

$$P_e \leq \epsilon + (M-1) \cdot \exp(-N(I(X;Y)-3\epsilon)) \leq \epsilon + \exp(-N\epsilon) \leq 2\epsilon, \quad (9)$$

if  $M = \exp(N(I(X;Y)-4\epsilon))$ . Inequality (9) therefore implies the existence of  $(M, N)$  codes with  $M = \exp(N(I(X;Y)-4\epsilon)) = \exp(N(1/2 \cdot \ln(1+\mathcal{P}/\sigma^2)-4\epsilon))$  and  $P_e \leq 2\epsilon$  for all  $N$  large enough.

#### IV. $\mathcal{P}$ -ACHIEVABILITY

In section III we have demonstrated the existence of a sequence of codes with small  $P_e$  and satisfactory values of  $M$ . However nothing was said about the maximum signal energy  $E$  for such a code. A certain code may contain a number of codewords with signal energy higher than  $N\mathcal{P}(1+2\epsilon)$ . Let  $\mathcal{M}^* := \{m : |\underline{x}(m)|^2 > N\mathcal{P}(1+2\epsilon)\}$  denote the messages corresponding to such codewords. What happens when  $m^* \in \mathcal{M}^*$  is sent? From the definition of  $\mathcal{G}_\epsilon(X, Y)$  it follows that  $-\ln(p_{\underline{X}}(\underline{x}) - Nh(X)) > N\epsilon$ , hence there is no  $\underline{y}$  for which  $(\underline{x}(m), \underline{y}) \in \mathcal{G}_\epsilon(X, Y)$ , in other words  $\Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\} = 1$ . Therefore

$$2\epsilon M \geq \sum_{m=1, M} \Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\} \geq \sum_{m \in \mathcal{M}^*} \Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\} = |\mathcal{M}^*|. \quad (10)$$

By expurgating the codewords corresponding to messages in  $\mathcal{M}^*$  we get a code with  $M' = M - |\mathcal{M}^*| \geq M(1-2\epsilon)$  codewords that satisfy the signal energy constraint  $|\underline{x}(m)|^2 \leq N\mathcal{P}(1+2\epsilon)$ . For the error probability of this code we find that

$$\begin{aligned} P_e' &= \frac{1}{M - |\mathcal{M}^*|} \cdot \sum_{m \notin \mathcal{M}^*} \Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\} \\ &= \frac{1}{M - |\mathcal{M}^*|} \cdot \left[ \sum_{m=1, M} \Pr\{D(\underline{Y}) \neq m | \underline{x}(m)\} - |\mathcal{M}^*| \right] \end{aligned}$$

$$= (MP_e^{-|\mathcal{M}^*|}) / (M^{-|\mathcal{M}^*|}) \leq (MP_e^{-|\mathcal{M}^*|} P_e) / (M^{-|\mathcal{M}^*|}) = P_e \leq 2\epsilon. \quad (11)$$

We finally conclude that for any  $\epsilon > 0$  there exist for all  $N$  large enough  $(M', N)$  codes with  $M' \geq (1-2\epsilon) \cdot \exp(N(1/2 \cdot \ln(1+\mathcal{P}/\sigma^2) - 4\epsilon)) \geq \exp(N(1/2 \cdot \ln(1+\mathcal{P}/\sigma^2) - 5\epsilon))$ ,  $P_e' \leq 2\epsilon$  and  $E \leq N\mathcal{P}(1+2\epsilon)$ . This implies that  $1/2 \cdot \ln(1+\mathcal{P}/\sigma^2)$  is  $\mathcal{P}$ -achievable for the AWGN channel and that the  $\mathcal{P}$ -capacity of this channel is at least  $1/2 \cdot \ln(1+\mathcal{P}/\sigma^2)$ .

#### IV. CONCLUSION

Typical sequences are used in many achievability proofs for multi-user channels. Inspection of these proofs however (see e.g. El Gamal and Cover [1]), tells us that this technique does not work for the Gaussian case, although some author(s) claim the opposite. The reason for this is that the cardinality of the typical set in the Gaussian (continuous) case cannot be bounded. However as shown here, this difficulty can be overcome. Since the proof in section III does carry over to the discrete case, the cardinality bounds for the typical set must be secondary properties. Handling the signal energy constraints is left to the reader in most achievability proofs for Gaussian multi-user channels. Our proof shows that with typical sequences there is a natural way to incorporate these constraints. It is obvious that the proof technique in this report can be generalized from the single-input single-output channel to multi-user channels as e.g. the broadcast channel and the multiple access channel.

The first achievability proof for the AWGN channel was given by Shannon [2]. His approach was geometrical. Gallager [3] showed that it is possible to approximate the Gaussian channel by a channel with discrete input and output alphabets with capacity as close to  $1/2 \cdot \ln(1+\mathcal{P}/\sigma^2)$  as desired. Both the geometrical and the approximation approach have disadvantages that make them inefficient when proving coding theorems for multi-user channels.

#### REFERENCES

- [1] A. El Gamal and T.M. Cover, "Multiple user information theory," Proc. of the IEEE, vol. 68, No. 12, Dec. 1980, pp. 1466-1483.
- [2] C.E. Shannon, "Communication in the presence of noise," Proc. IRE, vol. 37, No. 1, Jan. 1949, pp. 10-21.
- [3] R.G. Gallager, **Information Theory and Reliable Communication**. New York : Wiley, 1968.

In  
der un  
propo  
egy an

In  
the m  
Gi  
elem  
which  
time  
the b  
(resp  
to hi  
howe  
a) T  
objec  
deno  
b) T  
c) T  
d) T  
time  
e) O  
to v  
deno  
,  
mar  
gies  
1  
Ger  
2  
lesti  
Leu