EDIC Semester Project: Autumn 2022
Multiple Access Channels Under Maximal Error Probability

Adway Girish
Supervisor: Prof. Emre Telatar
Information Theory Laboratory (LTHI)

Last Updated: February 16, 2023

Abstract

The multiple access channel is the most well-understood problem setup in network information theory, with its capacity region exactly characterized when the average error probability is used as the reliability criterion. Nevertheless, it is known that when the maximal rather than average error probability is used, this leads, in general, to a smaller capacity region, about which very little is known. And yet, allowing for randomization at the encoders surprisingly leads to the same capacity region in both cases. For this result to be practically useful, it is also necessary that the maximal error decays to zero at a comparable rate to the average error (which is exponentially fast in the blocklength). We show that if the encoders have "too little" randomness, the maximal error decay cannot be exponential, and discuss the difficulties in proving a more general result.

1 Preliminaries

For completeness, below is a precise description of the problem that we look to solve, along with the associated definitions and preliminaries.

1.1 Problem Setting

The (2-input) *multiple access channel* (MAC) has two inputs and one output. It is described by its input space, transition probabilities, and output space, $(\mathcal{X}_1 \times \mathcal{X}_2, p_{Y|X_1,X_2}(y|x_1,x_2), \mathcal{Y})$. We assume that the channel is memoryless, i.e., for $(x_1^n, x_2^n, y^n) \in (\mathcal{X}_1^n, \mathcal{X}_2^n, \mathcal{Y}^n)$, we have $\Pr(y^n|x_1^n, x_2^n) = \prod_{i=1}^n p_{Y|X_1,X_2}(y_i|x_{1i}, x_{2i})$, which we represent simply as $p_{Y|X_1,X_2}(y^n|x_1^n, x_2^n)$.

Communicating over this channel are two independent senders at the input, having message sets $\mathcal{M}_i = \{1, \dots, M_i\},\ i = 1, 2$ resp., and a receiver at the output. An (n, R_1, R_2) *deterministic code* over the channel consists of *encoders* $f_i : \mathcal{M}_i \to \mathcal{X}_i^n,\ i = 1, 2$ and a *decoder* $g : \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$ such that $M_i = |\mathcal{M}_i| = \left\lceil 2^{nR_i} \right\rceil,\ i = 1, 2$. In particular, note that $f_1,\ f_2$ and g are deterministic functions.

We also define (n, R_1, R_2) *random codes*, which are simply random variables taking values in a set of (n, R_1, R_2) deterministic codes $\{C_j\}_{j \in \mathcal{J}}$ with some distribution Q. Clearly, such a code is only practically realizable if the encoders and decoder share some common randomness, which is rare, but they will be useful to us in proving the existence of "good" deterministic codes.

Requiring the same randomness at both the encoders and the decoder is too much to ask practically, so by making the decoders deterministic, we have the realizable *codes with stochastic encoders*, where the randomness is only required while encoding. We then have (independent)

encoders $\phi_i : \mathcal{M}_i \to \mathcal{P}(\mathcal{X}_i^n),\ i = 1, 2$ and a decoder $g : \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$ (where $\mathcal{P}(\mathcal{X})$ denotes the set of probability distributions on $\mathcal{X}$). In this case, the encoders define a conditional distribution on the input space given a message, which, by the functional representation lemma [1, Appendix B], is equivalent to the encoder being a deterministic function of the message and an independent random variable. We use this equivalent form, and denote the independent random variables by J_1 and J_2 for encoders ϕ_1 and ϕ_2, taking values in $\mathcal{J}_1$ and $\mathcal{J}_2$ resp. (since the encoders themselves are independent, so must J_1 and J_2). Thus an (n, R_1, R_2) *code with stochastic encoders* consists of encoders $f_i : \mathcal{M}_i \times \mathcal{J}_i \to \mathcal{X}_i^n,\ i = 1, 2$ that have access to random variables J_1 and J_2 with the product distribution Q, and a decoder $g : \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$. Once again, note that f_1, f_2 and g are deterministic functions. For brevity, we will refer to these codes as simply *stochastic codes*, which are not to be confused with random codes.

Note that random codes are the most general class of codes and contain all stochastic codes, which in turn contain all deterministic codes.

1.2 Error Probabilities and Capacity Regions

We can now define the error probabilities associated with each type of code and the capacity regions that they describe. Let C_{d} be a deterministic code, and define $p_e(C_{\mathsf{d}}, (m_1, m_2))$ to be the error probability of the code C_{d} for the message pair $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, i.e.,

$$p_e^{C_{\mathsf{d}}}(m_1, m_2) = \sum_{y \in \mathcal{Y}^n : g(y) \neq (m_1, m_2)} p_{Y|X_1,X_2}(y \mid x_1 = f_1(m_1), x_2 = f_2(m_2)).$$

Then the *average error probability* and *maximal error probability* are defined as

$$\begin{aligned} p_{\mathsf{a}}^{C_{\mathsf{d}}} &= \frac{1}{M_1 M_2} \sum_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{d}}}(m_1, m_2), \text{ and} \\ p_{\mathsf{m}}^{C_{\mathsf{d}}} &= \max_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{d}}}(m_1, m_2). \end{aligned}$$

Similarly, for a stochastic code C_{s}, we have

$$\begin{aligned} p_e^{C_{\mathsf{s}}}(m_1, m_2, j_1, j_2) &= \sum_{y \in \mathcal{Y}^n : g(y) \neq (m_1, m_2)} p_{Y|X_1,X_2}(y \mid x_1 = f_1(m_1, j_1), x_2 = f_2(m_2, j_2)), \\ p_e^{C_{\mathsf{s}}}(m_1, m_2) &= \mathbb{E}_Q\, p_e^{C_{\mathsf{s}}}(m_1, m_2, J_1, J_2), \\ p_{\mathsf{a}}^{C_{\mathsf{s}}} &= \frac{1}{M_1 M_2} \sum_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{s}}}(m_1, m_2), \text{ and} \\ p_{\mathsf{m}}^{C_{\mathsf{s}}} &= \max_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{s}}}(m_1, m_2), \end{aligned}$$

and for a random code C_{r} taking values in $\{C_j\}_{j \in \mathcal{J}}$ with distribution Q, where C_j, $j \in \mathcal{J}$ are deterministic codes,

$$\begin{aligned} p_e^{C_{\mathsf{r}}}(m_1, m_2, j) &= p_e^{C_j}(m_1, m_2), \\ p_e^{C_{\mathsf{r}}}(m_1, m_2) &= \mathbb{E}_Q\, p_e^{C_{\mathsf{r}}}(m_1, m_2, J), \\ p_{\mathsf{a}}^{C_{\mathsf{r}}} &= \frac{1}{M_1 M_2} \sum_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{r}}}(m_1, m_2), \text{ and} \\ p_{\mathsf{m}}^{C_{\mathsf{r}}} &= \max_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e^{C_{\mathsf{r}}}(m_1, m_2). \end{aligned}$$

A rate pair (R_1, R_2) is *achievable* with respect to the average (resp. maximal) error criterion if there is a sequence of (n, R_1, R_2) codes C_n such that $\lim_{n\to\infty} p_{\mathsf{m}}^{C_n} = 0$ (resp. $\lim_{n\to\infty} p_{\mathsf{a}}^{C_n} = 0$). The *capacity region* $\mathcal{C}$ is the closure of the set of achievable rate pairs. Depending on the choice of deterministic/stochastic/random codes and average/maximal error probability, we have six different capacity regions that may *a priori* all be unequal. As before, using the subscripts d, s and r to denote the use of deterministic, stochastic and random codes, and a and m to denote the average and maximal error criteria respectively, we have the capacity regions $\mathcal{C}_{\mathsf{d,m}}$, $\mathcal{C}_{\mathsf{d,a}}$, $\mathcal{C}_{\mathsf{s,m}}$, $\mathcal{C}_{\mathsf{s,a}}$, $\mathcal{C}_{\mathsf{r,m}}$, and $\mathcal{C}_{\mathsf{r,a}}$.

Some relations between these regions are immediate: all deterministic codes are a special type of stochastic codes, which are in turn random codes, hence $\mathcal{C}_{\mathsf{d},\cdot} \subseteq \mathcal{C}_{\mathsf{s},\cdot} \subseteq \mathcal{C}_{\mathsf{r},\cdot}$, where $\cdot$ may be a or m. Further, the maximal error criterion enforces a stronger constraint than the average error criterion, and hence, $\mathcal{C}_{\cdot,\mathsf{m}} \subseteq \mathcal{C}_{\cdot,\mathsf{a}}$, where $\cdot$ may be d, s or r. It is also easy to see that $\mathcal{C}_{\mathsf{d,a}} = \mathcal{C}_{\mathsf{s,a}} = \mathcal{C}_{\mathsf{r,a}}$, since it is always possible to construct a deterministic code with the same rate and at most the same average error probability as a given random code, because

$$p_{\mathsf{a}}^{C_{\mathsf{r}}} = \mathbb{E}_Q \frac{1}{M_1 M_2} \sum_{(m_1,m_2)} p_e^{C_{\mathsf{r}}}(m_1, m_2, J) \geq \frac{1}{M_1 M_2} \sum_{(m_1,m_2)} p_e^{C_j}(m_1, m_2) = p_{\mathsf{a}}^{C_j}$$

for some deterministic code C_j. This gives $\mathcal{C}_{\mathsf{r,a}} \subseteq \mathcal{C}_{\mathsf{d,a}}$ and together with $\mathcal{C}_{\mathsf{d,a}} \subseteq \mathcal{C}_{\mathsf{s,a}} \subseteq \mathcal{C}_{\mathsf{r,a}}$, we are done. (Under the maximal error criterion, the claim that there exists some such j would be invalid, since the expectation occurs before the maximization over messages, and these cannot be interchanged.)

2 Multiple Access Channels: Maximal vs. Average Error

In 1978, Dueck [2] showed by means of an example that the maximal error capacity region can be strictly smaller than the average error capacity region when using only deterministic encoders, i.e., $\mathcal{C}_{\mathsf{d,m}} \subset \mathcal{C}_{\mathsf{d,a}}$. Using Ahlswede's "random code reduction" technique [3], [4], [5, Chapter 12], N. Cai [6] showed that a stochastic code under the maximal error criterion can achieve any rate in the average error capacity region (deterministic or random, as they are the same) of the MAC, i.e., $\mathcal{C}_{\mathsf{s,m}} = \mathcal{C}_{\mathsf{r,m}} = \mathcal{C}_{\mathsf{d,a}} = \mathcal{C}_{\mathsf{r,a}}$. Thus, a practical shortcoming of using the average error probability as the performance criterion, namely that it does not guarantee error-free recovery of every codeword, seems to be a non-issue, as it is possible to achieve all rates in the average error capacity region with arbitrarily low maximal error probabilities. But this is not the case, as we are only guaranteed that the errors eventually decay to zero with sufficiently large blocklengths. To be able to sincerely claim that using stochastic codes, we can achieve the same practical performance for every codeword and not simply on average, we must also consider the rate of decay to zero, and require that the maximum error over all codewords becomes sufficiently small at a comparable blocklength to when the average error does.

It is known [7] that if $(R_1, R_2) \in \mathcal{C}_{\mathsf{d,a}}$, then not only does there exist (as by definition) a sequence of (n, R_1, R_2) codes C_n such that $p_{\mathsf{a}}^{C_n} \to 0$, but this decay is, in fact, exponential in n, i.e., there exist deterministic codes of rate pairs in the average error capacity region, such that the average error probability decays exponentially, or for large enough n,

$$p_{\mathsf{a}}^{C_n} < \exp(-nE + o(n))$$

for some $E > 0$, which is called the *error exponent*, and $\frac{o(n)}{n} \to 0$ as $n \to \infty$. It is also likely (but remains to be checked) that the same holds for maximal error probabilities $p_{\mathsf{m}}^{C_n}$ at rate pairs in the

deterministic maximal error capacity region $\mathcal{C}_{\mathsf{d,m}}$. Allowing for randomization at the encoder, the capacity region under the maximal error criterion increases from $\mathcal{C}_{\mathsf{d,m}}$ to $\mathcal{C}_{\mathsf{d,a}}$. A natural question to ask, then, is whether we can still have exponentially decaying $p_{\mathsf{m}}^{C_n}$ at rate pairs in between the two, i.e., in $\mathcal{C}_{\mathsf{d,a}} \setminus \mathcal{C}_{\mathsf{d,m}}$ (of course, if nonempty).

Conjecture 1. *Even though it is possible, by employing randomization at the encoder, to obtain codes with rate pairs outside the maximal error capacity region (in $\mathcal{C}_{\mathsf{d,a}} \setminus \mathcal{C}_{\mathsf{d,m}}$) achieving maximal error probabilities that can be made arbitrarily small, these maximal error probabilities cannot decay exponentially in n.*

This follows quite easily if the "amount of randomness" used is not "too large".

Theorem 1. *If the randomization employed at the encoders takes values in a set of cardinality subexponential in n, the maximal error probability cannot decay exponentially in n outside the maximal error capacity region.*

Proof. Let (R_1, R_2) be a rate pair in $\mathcal{C}_{\mathsf{d,a}} \setminus \mathcal{C}_{\mathsf{d,m}}$ (which we assume to be nonempty). Then there is a sequence of (n, R_1, R_2) random codes C_n indexed by random variables $J_n \in \mathcal{J}_n$ with distribution p_J, such that $p_{\mathsf{m}}^{C_n} \to 0$. Now suppose this decay is exponential, i.e., there exists a constant $E > 0$ such that $p_{\mathsf{m}}^{C_n} < \exp(-nE + o(n))$, or equivalently, for every (m_1, m_2) in the message set,

$$\mathbb{E}_{p_J}\, p_e^{C_n}(m_1, m_2, J_n) < \exp(-nE + o(n)).$$

Suppose further that the sets $\mathcal{J}_n$ grow subexponentially in n, i.e., for each $S > 0$, there exists an $n_0(S)$ such that for all $n > n_0(S)$, we have $|\mathcal{J}_n| < \exp(nS)$. Then there must be some instance $j_n^* \in \mathcal{J}_n$ such that

$$p_J(j_n^*) \geq \frac{1}{|\mathcal{J}_n|} > \exp(-nS) \text{ for every } S > 0,\ n > n_0(S).$$

Thus we have for every message pair (m_1, m_2), with any $S > 0$ and $n > n_0(S)$,

$$\begin{aligned} \exp(-nE + o(n)) &> \mathbb{E}_{p_J}\, p_e^{C_n}(m_1, m_2, J_n) \\ &\geq p_J(j_n^*)\, p_e^{C_n}(m_1, m_2, j_n^*) \\ &> \exp(-nS)\, p_e^{C_n}(m_1, m_2, j_n^*) \\ \implies p_e^{C_{j_n^*}} = p_e^{C_n}(m_1, m_2, j_n^*) &< \exp(o(n) - nE + nS). \end{aligned}$$

Choosing $S < E$ and letting $E' = E - S > 0$, we have that for every (m_1, m_2),

$$p_e^{C_{j_n^*}} < \exp(-nE' + o(n))$$

for $n > n_0(S)$, which shows the existence of a sequence of (n, R_1, R_2) deterministic codes $C_{j_n^*}$ with maximal error probability decaying exponentially to zero. This means that $(R_1, R_2) \in \mathcal{C}_{\mathsf{d,m}}$, contradicting our initial assumptions, and we are done. □

This settles Conjecture 1 if only a subexponential amount of randomness is allowed at the encoders. What happens with an exponential amount of randomness? Given that Theorem 1 is true, Conjecture 1 holds if the following conjecture does. This is motivated by Ahlswede's "random code reduction" result [3] – using the Chernoff bound, it is possible to show that from any arbitrary random code achieving exponentially small error probability, it is possible to obtain

a subcode with a subexponential amount of randomness, with the same rate and achieving small (note: not exponentially decaying, that would settle our problem!) error probability, and then it easily follows that there must exist some deterministic code with the same rate and small error since the random variable used in this randomization can also be transmitted with the message at no extra rate (there are only subexponentially many possible values, so $\frac{1}{n}\log|\mathcal{J}_n| \to 0$). If an exponential amount of randomness is used without reduction, then there must be a reduction in rate incurred by transmitting the randomization parameter.

Conjecture 2. *Using an exponential amount of randomness does not provide any improvement over using a subexponential amount, i.e., any rate pair and maximal error probability that can be achieved using codes with exponentially large randomness at the encoders can also be achieved by a subexponential subcode.*

We have not been able to either prove or disprove Conjecture 2, but consider the following example.

Example 1. Consider the $M \times M$ matrix A, with elements $A(w_1, w_2) = \mathbb{1}\{(w_2 - w_1) \bmod M \leq \sqrt{M}\}$, for $w_1, w_2 \in \{0, \ldots, M-1\}$. Now suppose that the error probabilities with some random code C for the message pair (m_1, m_2), when the random variable $J = j \in \mathcal{J} = \{0, \ldots, M-1\}$ is given by

$$p_e^C(m_1, m_2, j) = A(w_1, (w_2 + j) \bmod M).$$

If J is uniform on $\mathcal{J}$, we have that for all (m_1, m_2), $\mathbb{E}\, p_e^C(m_1, m_2, J) = \frac{1}{\sqrt{M}}$, which decays exponentially with n. However, $\max_{(m_1,m_2)} p_e^C(m_1, m_2, j) = 1$ for every j. Hence we have an exponentially small maximal error probability using an exponential amount of randomness, from which it is impossible to obtain a deterministic code with small (not even necessarily exponentially decaying) maximal error probability. From the contrapositive to Theorem 1, we have that it must also have been impossible to have a random code with a subexponential amount of randomness, which seems to have disproved Conjecture 2, except that

1. this specific example can be expurgated thanks to the presence of an $N \times N$ all-zero submatrix in the top right corner, with $N = \frac{1}{2}(M - \sqrt{M})$, but it does seem likely that there exists a similar example where expurgation is not possible; and

2. this choice of p_e may not be legitimate, i.e., it may not be possible for a code to have such an arrangement of error probabilities, but it is unclear how to characterize such legitimate error probability matrices.

3 Closing Comments

While Theorem 1 is almost trivial, the rest of the problem does not seem to go through quite so easily. Attempting to simply modify the expressions in Cai's proof [6] to incorporate an exponentially decaying error probability does not work. On the other hand, characterizing the maximal error probability itself is a difficult problem, and it is unclear how to make progress without making use of these "good" random codes or expurgation.

A more fundamental approach was first described by Ahlswede in 1973 [8] to show that for multi-user channels, it may not always be possible to extract a "good" subcode w.r.t. the maximal error criterion as is the case with the point-to-point channels, even preceding Dueck's concrete example [2] showing that the capacity regions are strictly unequal. This was further developed into

the "wringing technique" to prove strong converses for the MAC by Ahlswede in 1981 [9] and more recently by Kosut [10]. There may be some scope for development along these lines.

Finally, an important question that remains unanswered (even for the point-to-point channel) is how to optimally decode for the maximal error probability – the maximum *a posteriori* (MAP) decoder minimizes the average error probability – indeed, the optimal decoder for the maximal case is, in general, not even deterministic. It does seem likely that MAP decoding may still be nearly optimal (such as being only a constant factor off and hence optimal w.r.t. the error exponent), as elementary calculations for binary-input binary-output channels show, but generalizing such a result has been difficult because of the inherent difficulty in characterizing the maximal error probability.

References

[1] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2012, ISBN: 1107008735.

[2] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels.," *Prob. Control Inform. Theory, 1978; Vol. 7*, 1978.

[3] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 44, no. 2, 1978.

[4] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998. DOI: 10.1109/18.720535.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511921889.

[6] N. Cai, "The maximum error probability criterion, random encoder, and feedback, in multiple input channels," *Entropy*, vol. 16, no. 3, pp. 1211–1242, 2014.

[7] R. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on information Theory*, vol. 31, no. 2, pp. 124–142, 1985.

[8] R. Ahlswede, "On two-way communication channels and a problem by zarankiewics," *Probl. of Control and Inform. Theory*, 1973.

[9] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Comb. Inform. Syst. Sci*, vol. 7, no. 3, pp. 216–230, 1982.

[10] O. Kosut, "A second-order converse bound for the multiple-access channel via wringing dependence," *IEEE Transactions on Information Theory*, vol. 68, no. 6, pp. 3552–3584, 2022. DOI: 10.1109/TIT.2022.3151711.