CS3236 - Network Information Theory: Multiple-Access Channel

Christoph Schnabl

1 Introduction

We have studied communication models with one source and one destination that exchange encoded messages over a channel. In a full communication setup, we modeled source and channel coding separately. We argued that this is plausible since any redundancy has likely been compressed by some encoding algorithm. For the channel coding model we considered inputs x and outputs y over finite alphabets $\mathcal{X}, \mathcal{Y}$. We said that y is the output produced for a given input x according to the conditional probability distribution $P_{Y|X}$. We have explored the channel coding problem for a discrete memoryless channels with one source and one receiver. In particular, our focus was to maximize channel capacity while keeping the error probability small.
This report pertains to the general scenario of multiple sources, specifically the discrete memoryless multiple-access channel. One common example of such a multiple-access channel is a set of phones communicating with a base station. Ideally, we would strive for a theory that can be as comprehensive as the one-to-one setting. However, due to the interference caused by multiple sources on each other, developing such a theory for communication problems involving multiple parties is a challenging task. Nonetheless, recent advancements in the field of network information theory have enabled the development of solutions for an increasing number of multiple-party communication problems.

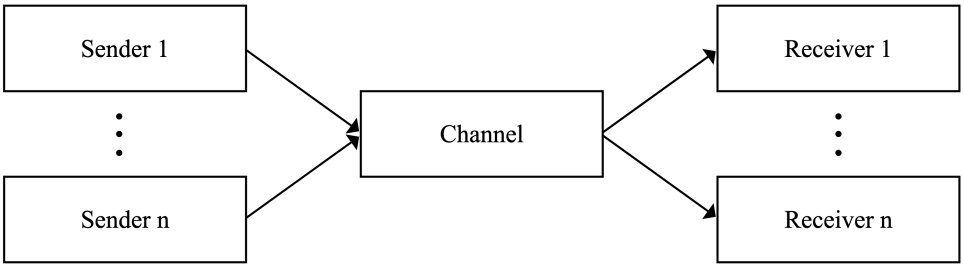

Figure 1: Multiple-user communication setup with more than one sender and receiver

In the following report, we will first examine the setup for multiple-access channels with two sources and relate the terminology to channel coding for a single source. Second, the multiple-access channel capacity theorem is stated describing the trade-off between achievable rates of different sources. Different from the one-to-one scenario in most cases there does not exist one single combination of rates that maximize the capacity for both sources. We will explore the intuition behind these bounds and relate them to the singular

model. In, particular it will be helpful to fix one source and look at the noise that is induced on it by the other source. We will then also give for DM-MAC channels and their respective rate limits.
Additionally, the Appendix includes a detailed proof of the main theorem in this report and uses the same terminology as the channel coding lecture.

2 Setup

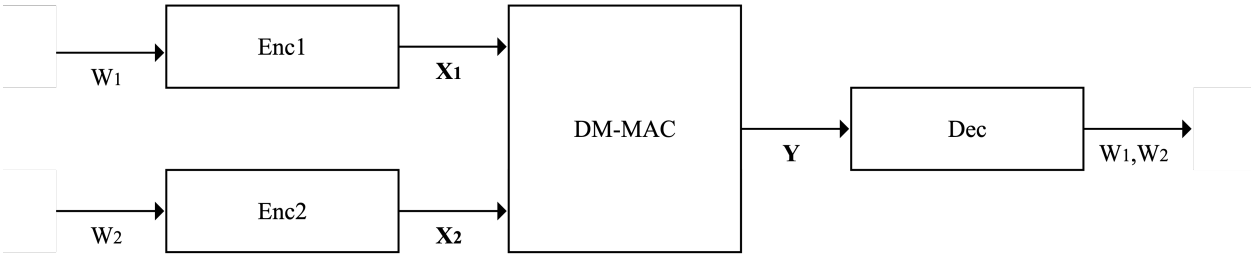

Figure 2: DM-MAC model setup with two senders and one receiver, adapted from [2]

Suppose there is more than one source that sends messages over one shared channel. As a first attempt, we could model this in our known channel coding setup where we use a joint alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ to emulate one source. However, it is reasonable to consider that multiple sources using a single channel will influence each other. In such a scenario, we would expect that the more one source transmits, the less transmission capacity would be available to the others. That way, sources have the incentive to coordinate their sending behavior to increase each other's capacity. For that reason, it is interesting to study protocols that alter the parameters of one source to allow for more reliable communication for others.
Throughout this report we will mainly focus on the case of two cases, many of the results generalize to more than two sources. Similar to the one-to-one case we let block length n approaches infinity, to be able to easier reason about capacity bounds.
First, we introduce the *discrete memory multiple-access channel model (DM-MAC)* and the *DM-MAC code*. Their definitions expand on the single-user channel coding model, but with the added complexity of two sources. This entails the use of two input alphabets, two encoding functions, the presence of noise between the two sources, and the need for the decoder to estimate the source of a given message. As in the singular case, we make the assumption that messages are sent at uniformly random and that the two sources send messages independently of each other.

Definition 1 (Discrete Memoryless Multiple-Access Channel (DM-MAC))**.** A discrete memoryless multiple-access channel (DM-MAC) consists of two sources that send messages over one shared medium - the channel, and one receiver:

- Denote x_1, x_2, y as messages with $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$ and output $y \in \mathcal{Y}$ where $\mathcal{X}_1$, $\mathcal{X}_2$ are the *input alphabet*, and $\mathcal{Y}$ the *output alphabet*, respectively

- A given message is produced with probability $P_{Y|X_1,X_2}(y|x_1,x_2)$
- Message pairs (x_1, x_2) are sent uniformly at random with $X_1 \perp X_2$

Definition 2 (DM-MAC Code)**.** A DM-MAC Code consists of two message sets, two encoding functions, one for the messages of each source, and one decoding function that tries to estimate the sent message:

- Two messages sets $\mathcal{W}_1 = \{1, ..., 2^{nR_1}\}$, $\mathcal{W}_2 = \{1, ..., 2^{nR_2}\}$
- Two encoding functions $Enc_1 : \mathcal{W}_1 \to \mathbf{X_1}$, $Enc_2 : \mathcal{W}_2 \to \mathbf{X_2}$
- One decoding function $Dec : \mathbf{Y} \to \mathcal{W}_1 \times \mathcal{W}_2$

We want the destination to be able to correctly estimate the original message m for a given message $\hat{m}$. If $m \neq \hat{m}$ we consider this an error. To analyze the trade-off between redunancy and rate we are interested in the expected error probability. Similar to the one-to-one case we will assume that messages in the joint distribution of $m_1 \times m_2$ are indepentent and equally probable. As stated before, any useless redundancy in the input source has been compressed away by the source coding step.

Definition 3 (Error Probability)**.** An error occurs if the decoder cannot decode a transmitted message. The average error probability for a DM-MAC code with rates R_1, R_2 and block length n is then defined as follows:

$$P_e \overset{(a)}{=} P((\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)) \overset{(b)}{=} \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1,m_2)\in\mathcal{W}_1\times\mathcal{W}_2} P((\hat{m}_1, \hat{m}_2) \neq (m_1, m_2) \mid \textit{chosen } (m_1, m_2))$$

where *(a)* follows from our error definition and *(b)* by total probability. By that $P(\textit{chosen } (m_1, m_2)) = \frac{1}{2^{n(R_1+R_2)}}$ since there exist $2^{n(R_1+R_2)}$ possible uniformly distributed message pairs over n channel uses. For error analysis this formula might be union-bounded.

3 Multiple-Access Channel Capacity

Similarly to the singular case, we are now interested in codes and their highest rates R_1, R_2 at which we can send messages with arbitrarily small errors, even if that results in adapting very long block lengths. We call such rates achievable, if we can construct a code, for which the error probability approaches 0 with increasing block lengths n or more formally:

Definition 4 (Achievability)**.** If there exists a DM-MAC code with rates R_1, R_2 such that for $\lim_{n\to\infty} P_e = 0$ the same rates are said to be achievable for this DM-MAC code

For the singular case, we defined the capacity as the maximum achievable rate for a capacity-achieving input distribution. We do not assume a total order over all the rate pairs for the multiple-access model. Simply, we do not know to which source we give more "priority" and allow to send more at the expense of the other source's rate. In other words, different achievable rate pairs are mereley trade-offs, e.g. we could increase the rate R_1 for the first source R_1, by decreasing the rate R_2 for the second source using

time-sharing. We will use a similar idea later to bind achievable rate pairs. We might be able to prefer one rate pair over another, e.g. clearly we would prefer (R_1, R_2) over (R_1', R_2') for $R_1 > R_1'$ and $R_2 > R_2'$. However these case are not particularly interesting to study. We proceed to formally define the capacity region as the capacity for the multiple-access scenario. We will later give examples for DM-MAC channels and explore their different capacity regions.

Definition 5 (Capacity Region)**.** The capacity region is the closure of the set of all achievable rate pairs (R_1, R_2)

Recall, that the closure of a set S is the union of $S \cup S'$ where S' denotes all the limit points of S. This notion is useful in the proof achievability of rate pairs which shows the existence of capacity achieving codes. The proof of this, which follows along the lines of the achievability proof from the lecture, can be found in the Cover book [1]. Note that, analogous to the one-to-one case, the proof solely shows the theoretical boundary, and does not inherently construct an encoding and decoding algorithm that can easily be used in practice.

3.1 Simple Capacity Region Bounds

We now want to explore a simple first bound on the capacity region, using the idea from above, time-sharing, and an inequality that draws inspiration from the singular case. We will later see, that these bounds are not tight. For time-sharing we assume that sources can switch instantenously and coordination does not decrease their ability to transmit. First, denote C_1, C_2 as the time-sharing bounds for the two different sources, respectively.

$$R_1 \leq C_1 \ = \max_{x_2, P_{X_1}(x_1)} I(X_1; Y | X_2 = x_2) \tag{1}$$

$$R_2 \leq C_2 \ = \max_{x_1, P_{X_2}(x_2)} I(X_2; Y | X_1 = x_1) \tag{2}$$

We will now give more intuition for this bound on the example for C_1. C_2 follows analogously. The mutual information $I(X_1; Y | X_2 = x_2)$ captures the amount of information the first source is able to transmit, given that the other source has sent x_2. Note that since we are maximizing over x_2 $I(X_1; Y | X_2 = x_2)$ might very well be bigger than $I(X_1; Y | X_2)$ for one specific x_2. This is called time-sharing as users share their access to the channel, similar to the term time-sharing in operating systems. We will now look at a third bound, the upper bound, that we derive from the one-to-one case.

$$R_1 + R_2 \leq C = \max_{P_{X_1}(x_1), P_{X_2}(x_2)} I(X_1, X_2; Y) \tag{3}$$

For the upper bound, we want the total amount of transmission by both sources to not exceed the capacity of the channel. The term $I(X_1, X_2; Y)$ gives us the amount of information that the sources reveal about the output. We then want to maximize the capacity, and thus the mutual information, by choosing capacity-achieving input distributions.
While the time-sharing bounds coincide with the upper bound for the binary multiplier channel, both are not tight in general. We will illustrate that later with the example of the binary erasure MAC.

3.2 Tight Capacity Region Bounds

We now give bounds, that better estimate the achievable rates. For the first inequality this means, that the rate R_1 is less than the information that X_1 given the inference, or noise of X_2. On the right-hand side $I(X_1; Y|X_2)$ quantifies the amount of information X_1 contains about Y given the noise (or uncertainty in information-theoretic terms) introduced by X_2. The second inequality follows analogously. The third bound, $I(X_1; X_2|Y)$, is similar to the capacity in the single user case. In information-theoretic terms it measures the reduction of uncertainty in X_1 after observing X_2, given Y. For channel coding, this can be viewed as the amount of shared information between the two sources. There the conditional mutual information $I(X1; X2|Y)$ represents the amount of shared information between the two signals, which reduces the channel capacity. We want the combined rate of transmission should be strictly less than this capacity $I(X_1; X_2|Y)$.

Definition 6 (Convex Hull)**.** The convex hull for a set $S \in \mathbb{R}^n$ is the smallest convex set such that it contains all points from S.

Theorem 1. *The capacity for a DM-MAC is the closure of the convex hull of all* (R_1, R_2) *that satisfy*

$$R_1 < I(X_1; Y|X_2) \tag{4}$$

$$R_2 < I(X_2; Y|X_1) \tag{5}$$

$$R_1 + R_2 < I(X_1; X_2|Y) \tag{6}$$

for some input probability distribution $P_{X_1}(x_1)P_{X_2}(x_2)$

The proof for the convere of this Theorem can be found in the Appendix 6.

4 Examples

In the following section, we will study three examples of DM-MAC and state their capacity region, as well as the intuition behind them.

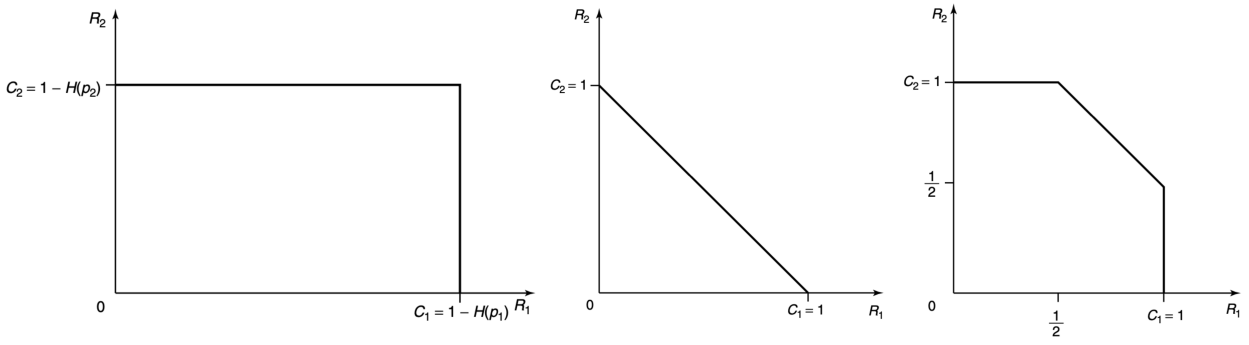

Figure 3: Capacity regions for the IBSC, BMC, BEMAC (from left to right) [1]

4.1 Indepenent Binary Symmetric Channels (IBSC)

The independent binary symmetric channel is an interesting case to study, as it provides a special case where our initial idea to model multiple sources without interference would work. That is, both sources send over a binary symmetric channel (BSC) without disturbing each other. Then, each source sees an independent channel such that $P_{Y|X_1,X_2} = P_{Y|X_1}P_{Y|X_2}$. By that we know $I(X_1;Y|X_2) = I(X_1;Y)$ and likewise $I(X_2;Y|X_1) = I(X_2;Y)$. From that follows, both sources can send at their individual capacity which we know to equal $C_1 = 1-H_2(p_1)$ and $C_2 = 1-H_2(p_2)$ for some capacity-achieving input distributions with parameters p_1 and p_2. The capacity region for the independent BSC is shown in Figure 3. The x-axis shows the rate R_1, and the y-axis the rate R_2. The horizontal and vertical lines display respective capacity regions given p_1 and p_2. Observe that in the diagram, as we analyzed earlier, the capacity of one source remains constant regardless of the capacity of the other source.

4.2 Binary Symmetric Channel (BSC)

Similarly to the singular case, suppose a channel that randomly flips bits in the output. In contrast, to the noiseless IBSC, this channel has some added noise Z. That is, the channel produces Y according to the following random variable: $Y = X_1 + X_2 + Z \pmod 2$ for $Z \sim Ber(p)$. We can then describe Y using the following probability distributions for $X \in \{X_1, X_2\}$: $P_{Y|X}(1|0) = P_{Y|X}(0|1) = p$ and $P_{Y|X}(1,1) = P_{Y|X}(0,0) = 1-p$. Clearly, setting $X_1 = 0$ or $X_2 = 0$ achieves the same rates as the BSC for the singular case. That is, $R_1 = 1-H_2(p)$ or $R_2 = 1-H_2(p)$ respectively. By that, with time-sharing all rate pairs with $R_1 + R_2 = 1 - H_2(p)$ are achievable. We can show, that this is the best we could do using Theorem 1.

4.3 Binary Multiplier Channel (BMC)

The binary multiplier channel (BMC) is a channel where the two sources influence each other in the following way: $Y = X_1X_2$. This implies that the more X_1 transmits, the less X_2 can transmit, and vice versa. We use the time-sharing capacity bounds to achieve any combination of rates as follows: $R_1 \leq 1$ (with $X_2 = 1$), $R_2 \leq 1$ (with $X_1 = 1$). As mentioned earlier, we will now show that these bounds coincide with the outer bound $R_1+R_2 \leq 1$, and also show the same for two of our tight bounds. We will prove that $I(X_1;Y|X_2) \leq 1$, similarly $I(X_2;Y|X_1) \leq 1$, and $I(X_1,X_2;Y) \leq 1$.

$$I(X_1;Y|X_2) \overset{(a)}{=} H(Y|X_2) - H(Y|X_1,X_2) \overset{(b)}{=} H(Y|X_2) \overset{(c)}{\leq} H(Y) \leq 1 \tag{7}$$

In the above equation, *(a)* follows from the definition of mutual information, *(b)* from the fact that X_1 and X_2 determine Y and by that $H(Y|X_1,X_2) = 0$. For *(c)* we use that conditioning reduces entropy since Y is a binary random variable. The cases for X_2, and Y can be proved similarly.
With above results, we can depict the capacity region for the binary multiplier channel in 3 which is just the line connecting the two extreme cases $(R_1, R_2) = (0,1)$ and $(1,0)$.

4.4 Binary Erasure Multiple-Access Channel

The binary erasure multiple-access channel is modeled by adding both inputs as follows: $Y = X_1 + X_2$ with $Y \in \{0,1,2\}$. Clearly, setting either $X_1 = 0$ or $X_2 = 0$ $Y = X_1$ we can achieve a rate of $R_1 = 1$ or $R_2 = 1$

for a noiseless channel. In fact, we can even do better than time-sharing in this case, and we will now show that the upper bound does not coincide with the time-sharing bounds for this channel.
For this, consider that a decoder is able to correctly decode a message $\hat{m}$ as long as $X_1 = X_2$. In other words, if both sources send the same symbol, their output is either 0 or 2 and can correctly be decoded. Ambiguity only arises, if one source sends 1 while the other one sends 0.
To now see, why this channel is called erasure channel, let's analyze the possible capacity region. Using the intuition from above, we can (without loss of generality) assume that X_1 is sent noiselessly with $R_1 = 1$.
We, now treat X_1 as noise induced on X_2 and try to bind the rate for X_2. Observe, that for X_2 the channel looks like a binary erasure channel with erasure probability $\frac{1}{2}$. This is also where binary-erasure comes from in the name of this channel, even if our initial definition of $Y = X_1 + X_2$ did not look at all like the singular case. Note, that different from the singular case, there is no extra symbol e. By the results from the lecture, we know that the capacity of the binary erasure channel equals $\frac{1}{2}$. Hence, our strategy of sending reliably with X_1 and over an emulated binary erasure channel with X_2 yields a combined rate of 1.5 bits per channel use. This channel can serve as an example, that the time-sharing bounds do not coincide with the outer bound in general. For time-sharing, the bound would again be a line between the two extreme points. We can do better using the strategy described above achieving rate pairs of $(1, 0.5)$ or $(0.5, 1)$. By time-sharing between these two extreme points, e.g. $(0.75, 0.75)$ is another achievable rate pair. Gardener and Wolf showed that the capacity region for the BEMAC is indeed $C = \{(R_1, R_2) \,|\, R_1 \leq 1, R_2 \leq 1 \wedge R_1 + R_2 \leq 1.5\}$. Their proof roughly works by setting $I(X_1; Y|X_2) = H(\alpha)$, $I(X_2; Y|X_1) = H(\beta)$ and calculating the resulting mutual information $I(X_1; Y|X_2)$. They then show that all three are maximized iff. $\alpha = \beta = \frac{1}{2}$, which correspond to the extreme points in our drawn capacity region.
The capacity region for the binary erasure multiple-access channel is shown in Figure 3. Observe, that by setting either of the rates $R = 1$ we can achieve $(1, 0.5)$ or $(1, 0.5)$ and any time-share between those two extremes.

4.5 Gaussian Multiple-Access Channel

Throughout this report, we focussed on discrete channels only. However, we are now giving one example of a channel that's not discrete, the Gaussian Multiple-Access Channel. Let $Y_i = X_{1,i} + X_{2,i} + Z_i$ by the output Y at time i where all Z_i are i.i.d and $Z_i \sim \mathcal{N}(0, N)$. In addition, we assume that each sender j is power restricted with P_j, that is for each sender and all their messages, the following must hold: $\frac{1}{n} x_{ji}^2 \leq P_j$. We are also able to extend our capacity region definition to this case and we can show that $I(X_1; Y|X_2) = h(X_1 + Z) - h(Z) \leq \frac{1}{2} log(1 + \frac{P_1}{N})$. We can use this to achieve the following rate bounds: $R_1 \leq C(\frac{P_1}{N})$, $R_2 \leq C(\frac{P_2}{N})$ and $R_1 + R_2 \leq C(\frac{P_1+P_2}{N})$. This capacity region can be achieved using a technique called code-division multiple-access, where each sender uses a different code. More details on this may be found in the Cover [1] or the El Gamal [2] book.

5 m-User Multiple Access Channels

The following section, gives a quick outlook on channel coding for more than two sources. The setup and the code definition follow along the lines of the $m = 2$ case. The capacity region is defined as follows:

Theorem 2. *The capacity region for the m-User MAC is the closure of the convex hull of all rate vectors* $R(S) = \sum_{i \in S} R_i$ *such that* $X(S) = \{X_i : i \in S\}$ *that satisfy:*

$$R(S) \leq I(X(S); Y | X(S^c)) \text{ for all } S \subseteq [m]$$

for some distribution $P_{X_1}(x_1)...P_{X_m}(x_m)$

6 Conclusion

To conclude we have first introduced a model for discrete memoryless multiple-access channels and explained the differences to the one-to-one case. We then explored bounds on the achievability of rates and their intuition, in particular the capacity region. Lastly, we explained examples of DM-MAC channels, their respective meaning, and achievable rates. In addition, the Appendix explains the converse part of proof for Theorem 1.

References

[1] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2012.

[2] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge university press, 2011.

[3] N Gaarder and J Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.). *IEEE Transactions on Information Theory*, 21(1):100–102, 1975.

[4] Tadao Kasami and Shu Lin. Coding for a multiple-access channel. *IEEE Transactions on Information Theory*, 22:129–137, 1976.

[5] Yury Polyanskiy. Mit open coursware 6.441s information theory, 2016.

[6] Claude E Shannon. Two-way communication channels. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pages 611–645. University of California Press, 1961.

[7] David Slepian and Jack Keil Wolf. A coding theorem for multiple access channels with correlated sources. *Bell System Technical Journal*, 52(7):1037–1076, 1973.

Appendix

Proof of the Capacity for the Gaussian Multiple Access Channel

We first use the definition of mutual information and insert the definition of $Y = X_1 + X_2 + Z$):

$$I(X_1; Y|X_2) = h(Y|X_2) - h(Y|X_1, X_2) = h(X_1 + X_2 + Z|X_2) - h(X_1 + X_2 + Z|X_1, X_2) \tag{8}$$

We then use the fact that $X_1 + X_2 + Z$ gives no more information about X_2 than $X_1 + Z$ and then that $Z \perp X_1$ and $Z \perp X_2$.

$$h(X_1 + X_2 + Z|X_2) - h(X_1 + X_2 + Z|X_1, X_2) = h(X_1 + Z|X_2) - h(Z|X_1, X_2) = h(X_1 + Z|X_2) - h(Z) \tag{9}$$

From $X_1 \perp X_2$ and the definition of $h(Z)$ know:

$$h(X_1 + Z|X_2) - h(Z) = h(X_1 + Z) - h(Z) = h(X_1 + Z) - \frac{1}{2}log(2\pi e)N \tag{10}$$

Similar to discrete entropy, where uniformity maximizes entropy, we know that normal l maximizes differential entropy:

$$h(X_1 + Z) - \frac{1}{2}log(2\pi e)N \leq \frac{1}{2}log(2\pi e)(P_1 + N) - \frac{1}{2}log(2\pi e)N = \frac{1}{2}log(1 + \frac{P_1}{N}) \tag{11}$$

Proof of Theorem 1 (Converse for the DM-MAC)

Proof. The proof presented by [7] and [1] involves many different steps, but each of them has an interesting information-theoretic intuition. The proof is presented in three parts. First, we want to bind the rates R_1, R_2 individually by taking the average mutual information over the codebook. In the second step, we bind the sum of both rates. The last step is to show that these averages converge to the true mutual information by taking the limit for $n \to \infty$. Intuitively for the rate R_1 this means that the noise induced on the channel by every symbol of m_2 converges to the overall noise induced by m_2 for large enough block sizes n. Similar to the lecture we denote m_1, m_2, and $\hat{m}$ as the random variables for the unencoded messages.

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}) + \epsilon_n \tag{12}$$

$$nR_1 \overset{(a)}{=} H(m_1) \overset{(b)}{=} I(m_1; \mathbf{Y}) + H(m_1|\mathbf{Y}) \overset{(c)}{\leq} I(m_1; \mathbf{Y}) + H(m_1, m_2|\mathbf{Y}) \overset{(d)}{\leq} I(m_1; \mathbf{Y}) + n\epsilon_n \tag{13}$$

We first bind the rate of R_1 as follows, *(a)* by the uniformity of m_1, *(b)* by using the definition of mutual information. For *(c)* expand using that $H(m_1, m_2|\mathbf{Y}) \geq H(m_1, m_2, \mathbf{Y})$ since $H(\mathbf{Y}) > 0$. Then by applying the chain rule again follows that $H(m_1, m_2, \mathbf{Y}) = H(m_2|m_1, \mathbf{Y}) + H(m_1|\mathbf{Y}) + H(\mathbf{Y}) \geq H(m_1|\mathbf{Y})$. In *(d)* we apply Fano's Inequality, to obtain $H(m_1, m_2|\mathbf{Y}) \leq n(R_1 + R_2)P_e + H(P_e) = n\epsilon_n$ and denote the last term as ϵ_n.

$$I(m_1;\mathbf{Y})+n\epsilon_n \overset{(a)}{\leq} I(\mathbf{X_1};\mathbf{Y})+n\epsilon_n \overset{(b)}{=} H(\mathbf{X_1})-H(\mathbf{X_1}|\mathbf{Y})+n\epsilon_n \overset{(c)}{\leq} H(\mathbf{X_1}|\mathbf{X_2})-H(\mathbf{X_1}|\mathbf{Y},\mathbf{X_2})+n\epsilon_n \quad (14)$$

We know *(a)* by the data-processing inequality since $m_1 \to \mathbf{X_1} \to \mathbf{Y} \to \hat{m}$ form a Markov chain, *(b)* by the definition of mutual information, an *(c)* since $\mathbf{X_1} \perp \mathbf{X_2}$ and conditioning reduces entropy. Then by repeated application of mutual information, we can compute the following:

$$H(\mathbf{X_1}|\mathbf{X_2})-H(\mathbf{X_1}|\mathbf{Y},\mathbf{X_2})+n\epsilon_n = I(\mathbf{X_1};\mathbf{Y}|\mathbf{X_2})+n\epsilon_n = H(\mathbf{Y}|\mathbf{X_2})-H(\mathbf{Y}|\mathbf{X_1},\mathbf{X_2})+n\epsilon_n \quad (15)$$

By applying the chain rule one more time we arrive at:

$$H(\mathbf{Y}|\mathbf{X_2})-H(\mathbf{Y}|\mathbf{X_1},\mathbf{X_2})+n\epsilon_n = H(\mathbf{Y}|\mathbf{X_2})-\sum_{i=1}^{n} H(Y_i|Y_{i-1},...,Y_1,\mathbf{X_1},\mathbf{X_2})+n\epsilon_n \quad (16)$$

Since the channel is memoryless, Y_i only depends on $X_{1,i}$ and $X_{2,i}$ and not on previous $Y_{i-1},...,Y_1$:

$$H(\mathbf{Y}|\mathbf{X_2})-\sum_{i=1}^{n} H(Y_i|Y_{i-1},...,Y_1,\mathbf{X_1},\mathbf{X_2})+n\epsilon_n = H(\mathbf{Y}|\mathbf{X_2})-\sum_{i=1}^{n} H(Y_i|X_{1,i}X_{2,i})+n\epsilon_n \quad (17)$$

By applying the chain rule, removing the conditioning and the definition of mutual information we get:

$$H(\mathbf{Y}|\mathbf{X_2})-\sum_{i=1}^{n} H(Y_i|X_{1,i}X_{2,i})+n\epsilon_n \leq \sum_{i=1}^{n} H(Y_i|\mathbf{X_2})-\sum_{i=1}^{n} H(Y_i|X_{1,i}X_{2,i})+n\epsilon_n = \sum_{i=1}^{n} I(X_{1,i};Y_i|X_{2,i})+n\epsilon_n \quad (18)$$

Analogous for R_2 :

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2,i};Y_i|X_{1,i})+\epsilon_n \quad (19)$$

We bind the sum of the two rates R_1+R_2 in a similar way to R_1 and R_2. Many of the steps will seem familiar.

$$n(R_1+R_2) \overset{(a)}{=} H(m_1,m_2) \overset{(b)}{=} I(m_1,m_2;\mathbf{Y})+H(m_1,m_2|\mathbf{Y}) \overset{(c)}{\leq} I(m_1,m_2;\mathbf{Y})+n\epsilon_n \quad (20)$$

We know *(a)* from the uniformity, *(b)* using the definition of mutual information, and *(c)* from Fano's inequality. All of these steps are similar to the R_1 case.

$$I(m_1,m_2;Y)+n\epsilon_n \overset{(a)}{\leq} I(\mathbf{X_1},\mathbf{X_2};\mathbf{Y})+n\epsilon_n \overset{(b)}{=} H(\mathbf{Y})-H(\mathbf{Y}|\mathbf{X_1},\mathbf{X_2})+n\epsilon_n \quad (21)$$

The transformation *(a)* follows from the data-processing inequality, and *(b)* by the definition of mutual information.

$$H(\mathbf{Y})-H(\mathbf{Y}|\mathbf{X_1},\mathbf{X_2})+n\epsilon_n \overset{(a)}{=} H(\mathbf{Y})-\sum_{i=1}^{n} H(Y_i|Y_{i-1},...,Y_1,\mathbf{X_1},\mathbf{X_2})+n\epsilon_n \quad (22)$$

The equality *(a)* is a result of first applying the chain rule and then by the memoryless property Y_i only depends on $X_{1,i}$ and $X_{2,i}$. By applying the chain rule, removing the conditioning and the definition of mutual information:

$$H(\mathbf{Y}) - \sum_{i=1}^{n} H(Y_i|Y_{i-1}, ..., Y_1, \mathbf{X_1}, \mathbf{X_2}) + n\epsilon_n \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n \tag{23}$$

We then take the limit for $n \to \infty$ and $P_e \to 0$ resulting in:

$$R_1 \leq \sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}) + n\epsilon_n = I(\mathbf{X_1}; \mathbf{Y}|\mathbf{X_2}) \tag{24}$$

The same can be done analogous for the other two cases R_2, and $R_1 + R_2$.

□