27.1 Recap

Last time we defined the multiple access channel as the sequence of random transformations

$$\{P_{Y^n|A^nB^n} : \mathcal{A}^n \times \mathcal{B}^n \to \mathcal{Y}^n, n = 1, 2, \dots\}$$

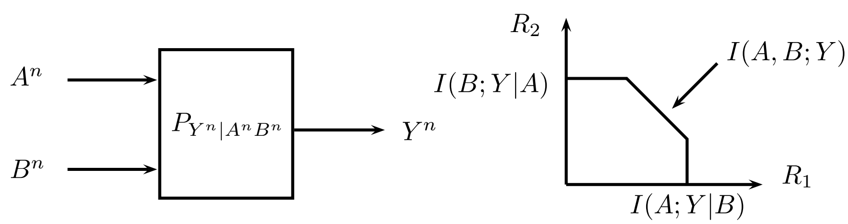
Furthermore, we showed that its capacity region is

$$C = \{(R_1, R_2) : \exists (n, 2^{nR_1}, 2^{nR_2}, \epsilon) - MAC\ code\} = \overline{co} \bigcup_{P_A P_B} \text{Penta}(P_A, P_B)$$

were $\overline{co}$ denotes the convex hull of the sets Penta, and Penta is

$$\text{Penta}(P_A, P_B) = \begin{cases} R_1 \le I(A;Y|B) \\ R_2 \le I(B;Y|A) \\ R_1 + R_2 \le I(A,B;Y) \end{cases}$$

So a general MAC and one Penta region looks like



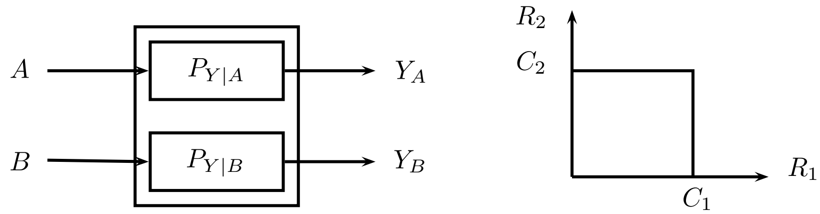
Note that the union of Pentas need not look like a Penta region itself, as we will see in a later example.

27.2 Orthogonal MAC

The trivial MAC is when each input sees its own independent channel: $P_{Y|AB} = P_{Y|A}P_{Y|B}$ where the receiver sees (Y_A, Y_B). In this situation, we expect that each transmitter can achieve it's own capacity, and no more than that. Indeed, our theorem above shows exactly this:

$$\text{Penta}(P_A, P_B) = \begin{cases} R_1 \le I(A;Y|B) = I(A;Y) \\ R_2 \le I(B;Y|A) = I(B;Y) \\ R_1 + R_2 \le I(A,B;Y) \end{cases}$$

Where in this case the last constraint is not applicable; it does not restrict the capacity region.



Hence our capacity region is a rectangle bounded by the individual capacities of each channel.

27.3 BSC MAC

Before introducing this channel, we need a definition and a theorem:

Definition 27.1 (Sum Capacity)**.** $C_{sum} \triangleq \max\{R_1 + R_2 : (R_1, R_2) \in C\}$

Theorem 27.1. $C_{sum} = \max_{A \perp\!\!\!\perp B} I(A, B; Y)$

Proof. Since the max above is achieved by an extreme point on one of the Penta regions, we can drop the convex closure operation to get

$$\max\{R_1 + R_2 : (R_1, R_2) \in \overline{co} \bigcup \text{Penta}(P_A, P_B)\} = \max\{R_1 + R_2 : (R_1, R_2) \in \bigcup \text{Penta}(P_A, P_B)\}$$
$$\max_{P_A, P_B} \{R_1 + R_2 : (R_1, R_2) \in \text{Penta}(P_A, P_B)\} \leq \max_{P_A, P_B} I(A, B; Y)$$

Where the last step follows from the definition of Penta. Now we need to show that the constraint on $R_1 + R_2$ in Penta is active at at least one point, so we need to show that $I(A, B; Y) \leq I(A; Y|B) + I(B; Y|A)$ when $A \perp\!\!\!\perp B$, which follows from applying Kolmogorov identities

$$\begin{aligned} I(A; Y, B) = 0 + I(A; Y|B) = I(A; Y) + I(A; B|Y) &\implies I(A; Y) \leq I(A; Y|B) \\ \implies I(A, B; Y) = I(A; Y) + I(B; Y|A) &\leq I(A; Y|B) + I(B; Y|A) \end{aligned}$$

Hence $\max_{P_A, P_B}\{R_1 + R_2 : (R_1, R_2) \in \text{Penta}(P_A, P_B)\} = \max_{P_A P_B} I(A, B; Y)$ □

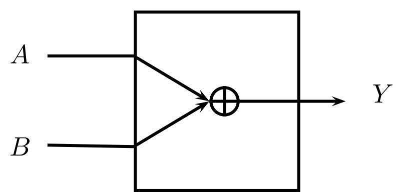
We now look at the BSC MAC, defined by

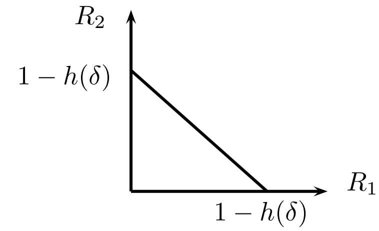
$$\begin{aligned} &Y = A + B + Z \mod 2 \\ &Z \sim \text{Ber}(\delta) \\ &A, B \in \{0, 1\} \end{aligned}$$

Since the output Y can only be 0 or 1, the capacity of this channel can be no larger than 1 bit. If B doesn't transmit at all, then A can achieve capacity $1 - h(\delta)$ (and B can achieve capacity when A doesn't transmit), so that $R_1, R_2 \leq 1 - h(\delta)$. By time sharing we can obtain any point between these two. This gives an inner bound on the capacity region. For an outer bound, we use Theorem 27.1, which gives

$$\begin{aligned} C_{sum} = \max_{P_A P_B} I(A, B; Y) &= \max_{P_A P_B} I(A, B; A + B + Z) \\ &= \max_{P_A P_B} H(A + B + Z) - H(Z) = 1 - h(\delta) \end{aligned}$$

Hence $R_1 + R_2 \leq 1 - h(\delta)$, so by this outer bound, we can do no better than time sharing between the two individual channel capacity points.





Remark: Even though this channel seems so simple, there are still hidden things about it, which we'll see later.

27.4 Adder MAC

Now we analyze the Adder MAC, which is a noiseless channel defined by:

$$
\begin{aligned}
&Y = A + B \quad (\text{over } \mathbb{Z}) \\
&A, B \in \{0, 1\}
\end{aligned}
$$

Intuitively, the game here is that when both A and B send either 0 or 1, we receiver 0 or 2 and can decode perfectly. However, when A sends 0 and B send 1, the situation is ambiguous. To analyze this channel, we start with an interesting fact

Interesting Fact 1: Any deterministic MAC ($Y = f(A, B)$) has $C_{sum} = \max H(Y)$. To see this, just expand $I(A, B; Y)$.

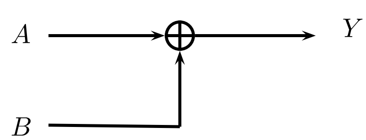
Therefore, the sum capacity of this MAC is

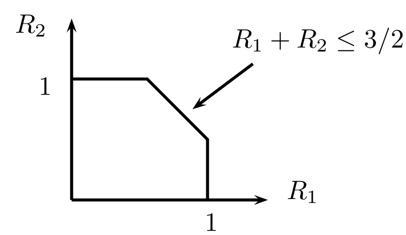
$$
C_{sum} = \max_{A \perp\!\!\!\perp B} H(A + B) = H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) = \frac{3}{2} \text{ bits}
$$

Which is achieved when both A and B are Ber(1/2). With this, our capacity region is

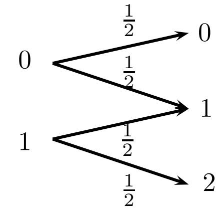
$$
\text{Penta}(\text{Ber}(1/2), \text{Ber}(1/2)) = \begin{cases} R_1 \le I(A; Y|B) = H(A) = 1 \\ R_2 \le I(B; Y|A) = H(B) = 1 \\ R_1 + R_2 \le I(A, B; Y) = 3/2 \end{cases}
$$

So the channel can be described by

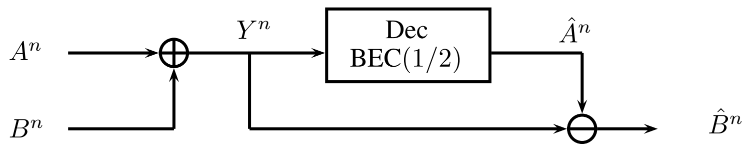




Now we can ask: how do we achieve the corner points of the region, e.g. $R_1 = 1/2$ and $R_2 = 1$? The answer gives insights into how to code for this channel. Take the greedy codebook $B = \{0, 1\}^n$ (the entire space), then the channel $A \to Y$ is a DMC:



Which we recognize as a BEC(1/2) (no preference to either −1 or 1), which has capacity 1/2. How do we decode? The idea is *successive cancellation*, where first we decode A, then remove $\hat{A}$ from Y, then decode B.



Using this strategy, we can use a single user code for the BEC (an object we understand well) to attain capacity.

27.5 Multiplier MAC

The Multiplier MAC is defined as

$$Y = AB$$
$$A \in \{0,1\},\ B \in \{-1,1\}$$

Note that $A = |Y|$ can always be resolved, and B can be resolved whenever $A = 1$. To find the capacity region of this channel, we'll use another interesting fact:

Interesting Fact 2: If $A = g(Y)$, then each $\text{Penta}(P_A, P_B)$ is a rectangle with

$$\text{Penta}(P_A, P_B) = \begin{cases} R_1 \le H(A) \\ R_2 \le I(A,B;Y) - H(A) \end{cases}$$

Proof. Using the assumption that $A = g(Y)$ and expanding the mutual information

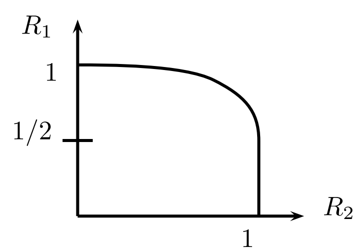
$$I(A;Y|B) + I(B;Y|A) = H(A) - H(Y|A) - H(Y|A,B) = H(A,Y) - H(Y|A,B)$$
$$= H(Y) - H(Y|A,B) = I(A,B;Y)$$

Therefore the $R_1 + R_2$ constraint is not active, so our region is a rectangle. □

By symmetry, we take $P_B = \text{Ber}(1/2)$. When $P_A = \text{Ber}(p)$, the output has $H(Y) = p + h(p)$. Using the above fact, the capacity region for the Multiplier MAC is

$$C = \overline{co} \bigcup \begin{cases} R_1 \le H(A) = h(p) \\ R_2 \le H(Y) - H(A) = p \end{cases}$$

We can view this as the graph of the binary entropy function on its side, parametrized by p:



To achieve the extreme point $(1, 1/2)$ of this region, we can use the same scheme as for the Adder MAC: take the codebook of A to be $\{0,1\}^n$, then B sees a BEC(1/2). Again, successive cancellation decoding can be used.

For future reference we note:

Lemma 27.1. *The full capacity region of multiplier MAC is achieved with zero error.*

Proof. For a given codebook D of user B the number of messages that user A can send equals the total number of erasure patters that codebook D can tolerate with vanishing probability of error. Fix rate $R_2 < 1$ and let D be a row-span of a random linear $nR_2 \times n$ binary matrix. Then randomly erase each column with probability $1 - R_2 - \epsilon$. Since on average there will be $n(R_2 + \epsilon)$ columns left, the resulting matrix is still full-rank and the decoding is possible. In other words,

$$\mathbb{P}[D \text{ is decodable}, \# \text{ of erasures} \approx n(1 - R_2 - \epsilon)] \to 1\,.$$

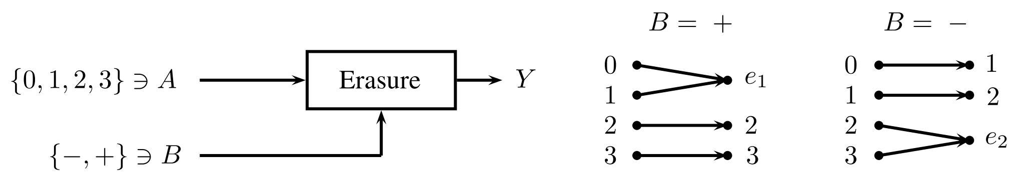
Hence, by counting the total number of erasures, for a random linear code we have

$$\mathbb{E}[\# \text{ of decodable erasure patterns for } D] \approx 2^{nh(1-R_2-\epsilon)+o(n)}\,.$$

And result follows by selecting a random element of the D-ensemble and then taking the codebook of user A to be the set of decodable erasure patterns for a selected D. □

27.6 Contraction MAC

The Contraction MAC is defined as



Here, B is received perfectly, We can use the fact above to see that the capacity region is

$$C = \begin{cases} R_1 \le \frac{3}{2} \\ R_2 \le 1 \end{cases}$$

For future reference we note the following:

Lemma 27.2. *The zero-error capacity of the contraction MAC satisfies*

$$R_1 \le h(1/3) + (2/3 - p)\log 2\,, \tag{27.1}$$

$$R_2 \le h(p) \tag{27.2}$$

for some $p \in [0, 1/2]$. *In particular, the point* $R_1 = \frac{3}{2}$, $R_2 = 1$ *is not achievable with zero error.*

Proof. Let C and D denote the zero-error codebooks of two users. Then for each string $b^n \in \{+, -\}^n$ denote

$$U_{b^n} = \{a^n : a_j \in \{0, 1\} \text{ if } b_j = +, a_j \in \{2, 3\} \text{ if } b_j = -\}\,.$$

Then clearly for each b^n we have

$$|U_{b^n}| \le 2^{d(b^n, D)}\,,$$

where $d(b^n, D)$ denotes the minimum Hamming distance from string b^n to the set D. Then,

$$|C| \le \sum_{b^n} 2^{d(b^n, D)} \tag{27.3}$$

$$= \sum_{j=0}^{n} 2^j |\{b^n : d(b^n, D) = j\}| \tag{27.4}$$

For a given cardinality $|D|$ the set that maximizes the above sum is the Hamming ball. Hence, $R_2 = h(p) + O(1)$ implies

$$R_2 \le \max_{q \in [p,1]} h(q) + (q - p) \log 2 = h(1/3) + (2/3 - p) \log 2 \,.$$

□

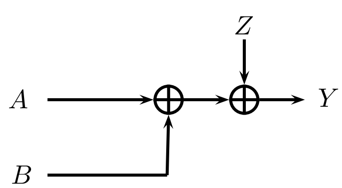
27.7 Gaussian MAC

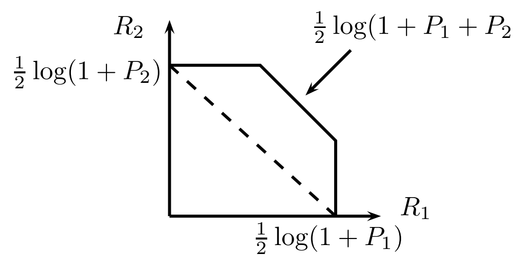
Perhaps the most important MAC is the Gaussian MAC. This is defined as

$$\begin{aligned} &Y = A + B + Z \\ &Z \sim \mathcal{N}(0,1) \\ &\mathbb{E}[A^2] \le P_1,\ \mathbb{E}[B^2] \le P_2 \end{aligned}$$

Evaluating the mutual information, we see that the capacity region is

$$\begin{aligned} &I(A;Y|B) = I(A;A+Z) \le \frac{1}{2}\log(1+P_1) \\ &I(B;Y|A) = I(B;B+Z) \le \frac{1}{2} \le (1+P_2) \\ &I(A,B;Y) = h(Y) - h(Z) \le \frac{1}{2}\log(1+P_1+P_2) \end{aligned}$$





Where the region is Penta$(\mathcal{N}(0,P_1), \mathcal{N}(0,P_2))$. How do we achieve the rates in this region? We'll look at a few schemes.

1. TDMA: A and B switch off between transmitting at full rate and not transmitting at all. This achieves any rate pair in the form

$$R_1 = \lambda \frac{1}{2}\log(1+P_1), \quad R_2 = \bar{\lambda}\frac{1}{2}\log(1+P_2)$$

   Which is the dotted line on the plot above. Clearly, there are much better rates to be gained by smarter schemes.

2. FDMA (OFDM): Dividing users into different frequency bands rather than time windows gives an enormous advantage. Using frequency division, we can attain rates

$$R_1 = \lambda \frac{1}{2}\log\left(1+\frac{P_1}{\lambda}\right), \quad R_2 = \bar{\lambda}\frac{1}{2}\log\left(1+\frac{P_2}{\bar{\lambda}}\right)$$

In fact, these rates touch the boundary of the capacity region at its intersection with the $R_1 = R_2$ line. The optimal rate occurs when the power at each transmitter makes the noise look white:

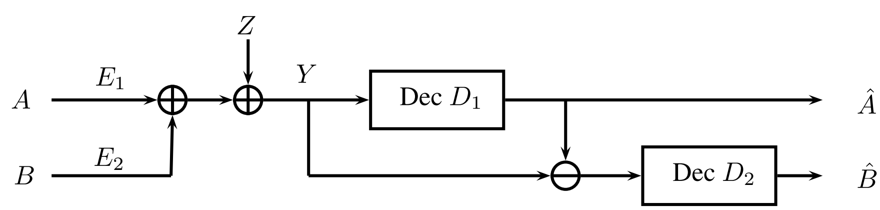
$$\frac{P_1}{\lambda} = \frac{P_2}{\bar{\lambda}} \implies \lambda^* = \frac{P_1}{P_1 + P_2}$$

While this touches the capacity region at one point, it doesn't quite reach the corner points. Note, however, that practical systems (e.g. cellular networks) typically employ *power control* that ensures received powers P_i of all users are roughly equal. In this case (i.e. when $P_1 = P_2$) the point where FDMA touches the capacity boundary is at a very desirable location of *symmetric rate* $R_1 = R_2$. This is one of the reasons why modern standards (e.g. LTE 4G) do not employ any specialized MAC-codes and use OFDM together with good single-user codes.

3. Rate Splitting/Successive Cancellation: To reach the corner points, we can use successive cancellation, similar to the decoding schemes in the Adder and Multiplier MACs. We can use rates:

$$R_2 = \frac{1}{2}\log(1 + P_2)$$
$$R_1 = \frac{1}{2}(\log(1 + P_1 + P_2) - \log(1 + P_2)) = \frac{1}{2}\log\left(1 + \frac{P_1}{1 + P_2}\right)$$

The second expression suggests that A transmits at a rate for an AWGN channel that has power constraint P_1 and noise $1 + P_2$, i.e. the power used by B looks like noise to A.



Theorem 27.2. *There exists a successive-cancellation code (i.e.* (E_1, E_2, D_1, D_2)*) that achieves the corner points of the Gaussian MAC capacity region.*

Proof. Random coding: $B^n \sim \mathcal{N}(0, P_2)^n$. Since A^n now sees noise $1 + P_2$, there exists a code for A with rate $R_1 = \frac{1}{2}\log(1 + P_1/(1 + P_2))$. □

This scheme (unlike the above two) can tolerate frame un-synchronization between the two transmitters. This is because any chunk of length n has distribution $\mathcal{N}(0, P_2)^n$. It has generalizations to non-corner points and to arbitrary number of users. See [RU96] for details.

27.8 MAC Peculiarities

Now that we've seen some nice properties and examples of MACs, we'll look at cases where MACs differ from the point to point channels we've seen so far.

1. Max probability of error ≠ average probability of error.

Theorem 27.3. $C^{(max)} \neq C$

Proof. The key observation for deterministic MAC is that $C^{(max)} = C_0$ (zero error capacity) when $\epsilon \leq 1/2$. This is because when any two strings can be confused, the maximum probability of error

$$\max_{m,m'} \mathbb{P}[\hat{W}_1 \neq m \cup \hat{W}_2 \neq m' | W_1 = m, W_2 = m']$$

Must be larger than 1/2. □

For some of the channels we've seen

- Contraction MAC: $C_0 \neq C$
- Multiplier MAC: $C_0 = C$
- Adder MAC: $C_0 \neq C$. For this channel, no one yet can show that $C_{0,sum} < 3/2$. The idea is combinatorial in nature: produce two sets (Sidon sets) such that all pairwise sums between the two do not overlap.

2. Separation does not hold: In the point to point channel, through joint source channel coding we saw that an optimal architecture is to do source coding then channel coding separately. This doesn't hold for the MAC. Take as a simple example the Adder MAC with a correlated source and bandwidth expansion factor $\rho = 1$. Let the source (S,T) have joint distribution

$$P_{ST} = \begin{bmatrix} 1/3 & 1/3 \\ 0 & 1/3 \end{bmatrix}$$

We encode S^n to channel input A^n and T^n to channel input B^n. The simplest possible scheme is to not encoder at all; simply take $S_j = A_j$ and $T_j = B_j$. Take the decoder

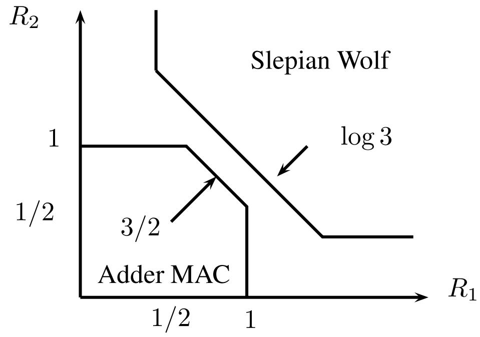
$$\begin{array}{rcc} & \hat{S} & \hat{T} \\ Y_j = 0 \implies & 0 & 0 \\ Y_j = 1 \implies & 0 & 1 \\ Y_j = 2 \implies & 1 & 1 \end{array}$$

Which gives $\mathbb{P}[\hat{S}^n = S^n, \hat{T}^n = T^n] = 1$, since we are able to take advantage of the zero entry in joint distribution of our correlated source.

Can we achieve this with a separated source? Amazingly, even though the above scheme is so simple, we can't! The compressors in the separated architecture operate in the Slepian Wolf region

$$\begin{cases} R_1 \geq H(S|T) \\ R_2 \geq H(T|S) \\ R_1 + R_2 \geq H(S,T) = \log 3 \end{cases}$$

Hence the sum rate for compression must be $\geq \log 3$, while the sum rate for the Adder MAC must be $\leq 3/2$, so these two regions do not overlap, hence we can not operate at a bandwidth expansion factor of 1 for this source and channel.



3. Linear codes beat generic ones: Consider a BSC-MAC and suppose that two users A and B have independet k-bit messages $W_1, W_2 \in \mathbb{F}_2^k$. Suppose the receiver is only interested in estimating $W_1 + W_2$. What is the largest ratio k/n? Clearly, separation can achieve

$$k/n \approx \frac{1}{2}(\log 2 - h(\delta))$$

by simply creating a scheme in which both W_1 and W_2 are estimated and then their sum is computed.

A more clever solution is however to encode

$$\begin{aligned} A^n &= G \cdot W_1 \,, \\ B^n &= G \cdot W_2 \,, \\ Y^n &= A^n + B^n + Z^n = G(W_1 + W_2) + Z^n \,. \end{aligned}$$

where G is a generating matrix of a good k-to-n linear code. Then, provided that

$$k < n(\log 2 - h(\delta)) + o(n)$$

the sum $W_1 + W_2$ is decodable (see Theorem 16.2). Hence even for a simple BSC-MAC there exist clever ways to exceed MAC capacity for certain scenarios. Note that this "distributed computation" can also be viewed as lossy source coding with a distortion metric that is only sensitive to discrepancy between $W_1 + W_2$ and $\hat{W}_1 + \hat{W}_2$.

4. Dispersion is unknown: We have seen that for the point-to-point channel, not only we know the capacity, but the next-order terms (see Theorem 20.2). For the MAC-channel only the capacity is known. In fact, let us define

$$R^*_{sum}(n, \epsilon) \triangleq \sup\{R_1 + R_2 : (R_1, R_2) \in \mathcal{R}^*(n, \epsilon)\} \,.$$

Now, take Adder-MAC as an example. A simple exercise in random-coding with $P_A = P_B = \text{Ber}(1/2)$ shows

$$R^*_{sum}(n, \epsilon) \geq \frac{3}{2}\log 2 - \sqrt{\frac{1}{4n}} Q^{-1}(\epsilon) \log 2 + O(\frac{\log n}{n}) \,.$$

In the converse direction the situation is rather sad. In fact the best bound we have is only slightly better than the Fano's inequality [?]. Namely for each $\epsilon > 0$ there is a constant $K_\epsilon > 0$ such that

$$R^*_{sum}(n, \epsilon) \leq \frac{3}{2}\log 2 + K_\epsilon \frac{\log n}{\sqrt{n}} \,.$$

So it is not even known if sum-rate approaches sum-capacity from above or from below as $n \to \infty$! What is even more surprising, is that the dependence of the residual term on ϵ is not clear at all. In fact, despite the decades of attempts, even for $\epsilon = 0$ the best known bound to date is just the Fano's inequality(!)

$$R^*_{sum}(n, 0) \leq \frac{3}{2}\,.$$

6.441 Information Theory
Spring 2016