Information Theory
Winter School 2007

La Colle sur Loup
16 March 2007

Zero Error

James L. Massey

Prof.-em. ETH Zürich
Adjunct Prof., Lund Univ., Sweden
Adjunct Prof., Tech. Univ Denmark
Trondhjemsgade 3, 2TH
DK-2100 Copenhagen East

JamesMassey@compuserve.com

The new results in these lectures were obtained in joint research with

Peter C. Massey

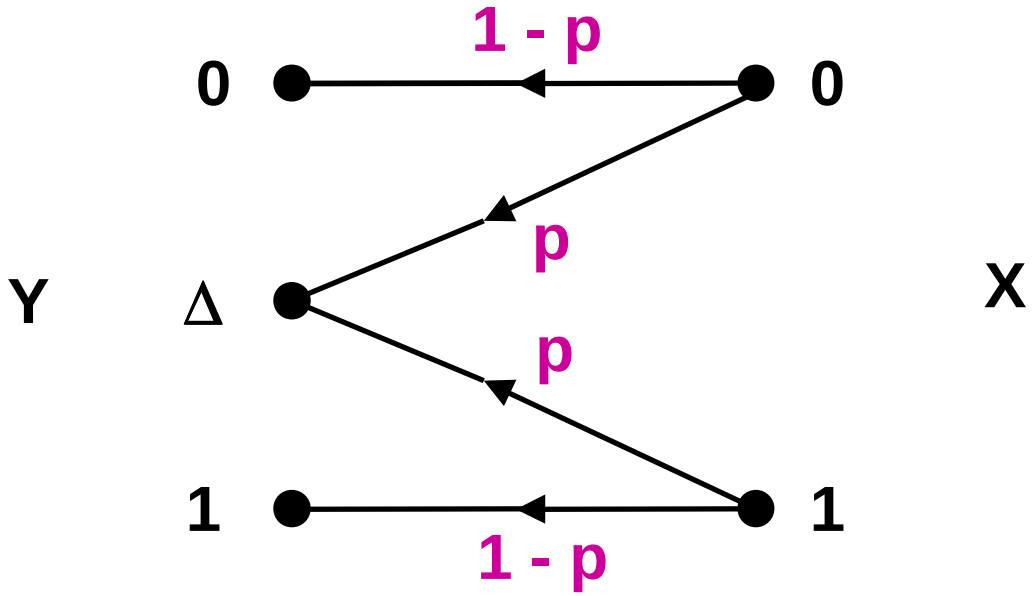
Since its founding by Shannon in 1948, information theory has mostly dealt with the **small-error capacity C** of channels.

C is the largest number such that, for every $\varepsilon > 0$ and every $\delta > 0$, information bits from a Binary Symmetric Source (BSS) can, with the use of proper coding, be sent over the channel at a **rate R** > **C** - δ bits/channel-use and with **bit error probability** $P_b < \varepsilon$.

BSS = Monkey with a fair coin.

Hereafter, we consider only Discrete Memoryless Channels (DMCs).

Example: **The Binary Erasure Channel (BEC)**



As almost everybody knows,

C = 1 - **p** bits/use

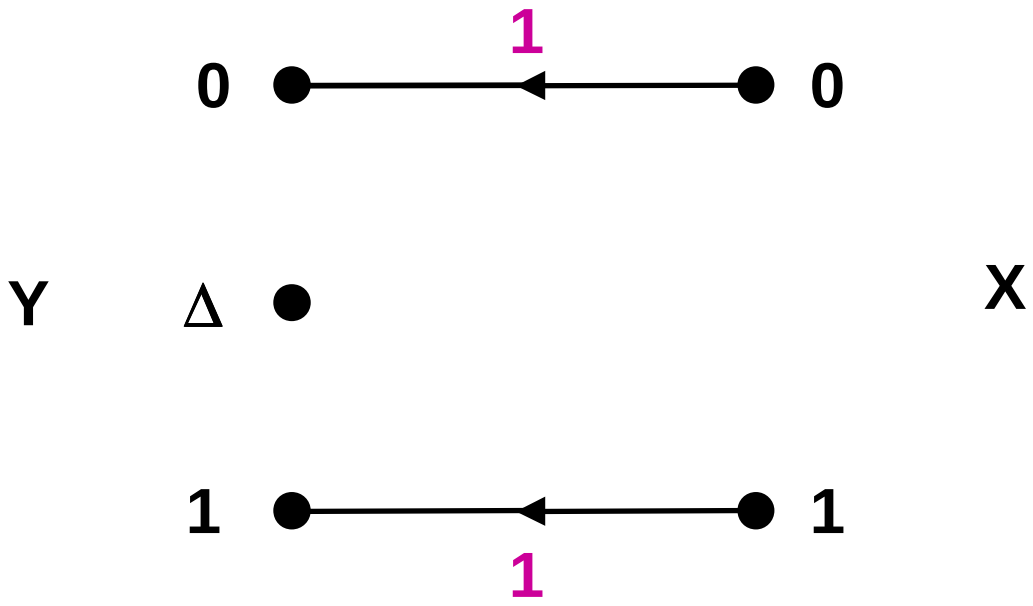
Peter Elias (1923-2001) and friends
Boston 17 August 1998

In his 1956 paper, "The zero error capacity of a noisy channel", Shannon defined the **zero-error capacity C_0** as the largest number such that, for every $\delta > 0$, **K** information bits from a Binary Symmetric Source can, with the use of block code of length **N**, be sent over the channel at a **rate R = K/N** > $C_0 - \delta$ bits/channel-use and with block error probability $P_b = 0$.

N.B. When one deals with zero error, **bit error probability** and **block error probability** coincide.

What is the zero-error-capacity of the BEC?

First consider the special case of the BEC with **p** = 0.



(**We will not show impossible transitions**.)

The output letter Δ is **unreachable** for this BEC.

Obviously, C_0 = 1 bit/use.

What is the zero-error-capacity of the BEC in the nontrivial case where $p > 0$?

Trivial Lemma:
If you cannot send even one information bit over the channel with zero error no matter how large the block length N, then $C_0 = 0$ bits/use.

To send one information bit, we need a code with only two codewords of length N, e.g., 0 0 … 0 and 1 1 … 1. But no matter which codeword we send, $\Delta\ \Delta \ldots \Delta$ will be received with nonzero probability and the decoder cannot then make a zero-error decision. Therefore, $C_0 = 0$ bits/use.

Let $\mathcal{Y}(x)$ denote the set of all output letters **reachable** (with **positive probability**) from the input letter x.

Theorem (Shannon, 1956):
The zero-error-capacity C_0 of a discrete memoryless channel is zero unless and only unless it has **two** input letters x and x' whose **reachable sets** $\mathcal{Y}(x)$ and $\mathcal{Y}(x')$ are **disjoint**.

N.B. If $C_0 \neq 0$, then $C_0 \geq 1$ bit/use; in fact $R = 1$ bit/use with zero error can be achieved with a block code of length $N = 1$.

The two input letters of the **BEC** with $p > 0$ do not have disjoint reachable sets because the erasure symbol Δ is in both sets. Thus $C_0 = 0$.

Shannon called two input letters x and x' **adjacent** if their reachable sets $\mathcal{Y}(x)$ and $\mathcal{Y}(x')$ are **not** disjoint, i.e., if **there is an output letter reachable from both**. (Shannon was usually a master at choosing appropriate terminology but, in my opinion, not this time.) I am going to use the term **confusable** instead of "adjacent".

We will say that two or more channel input letters are **non-confusable** if no output letter can be reached from more than one of these inputs.

Similarly, we will say that two or more channel input sequences of the same length N (N > 0) are **non-confusable** if there is no output sequence of length N that can result with nonzero probability from transmitting two of these input sequences.

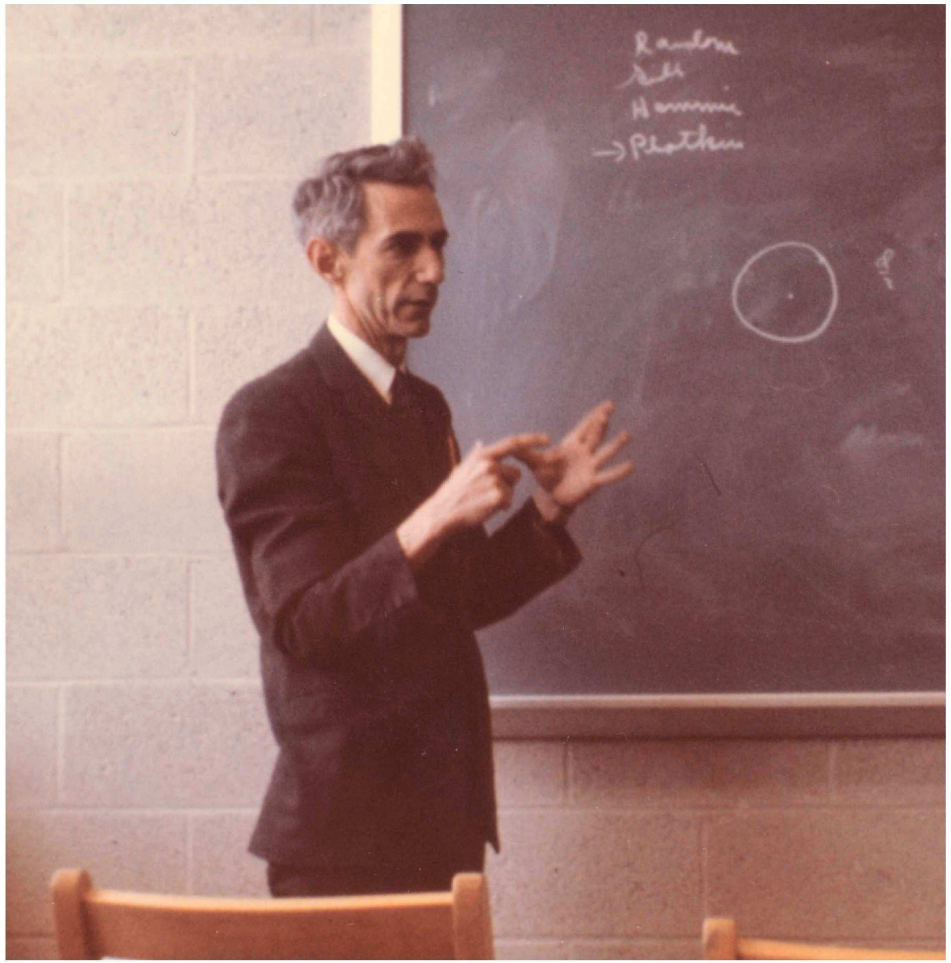
In our language, the previous theorem becomes.

> Theorem (Shannon, 1956):
> The zero-error-capacity C_0 of a discrete memoryless channel is zero unless and only unless it has two **non-confusable** input letters.

Shannon's terminology may have been unfortunate, but **Shannon was putting his finger on the right concept!** When one is considering zero-error capacity, the only thing that counts about a transition probability is whether it is zero or not.

This makes computing zero-error capacity a **combinatorial problem** rather than an analytical probabilistic problem.

Shannon liked ideas more than equations!

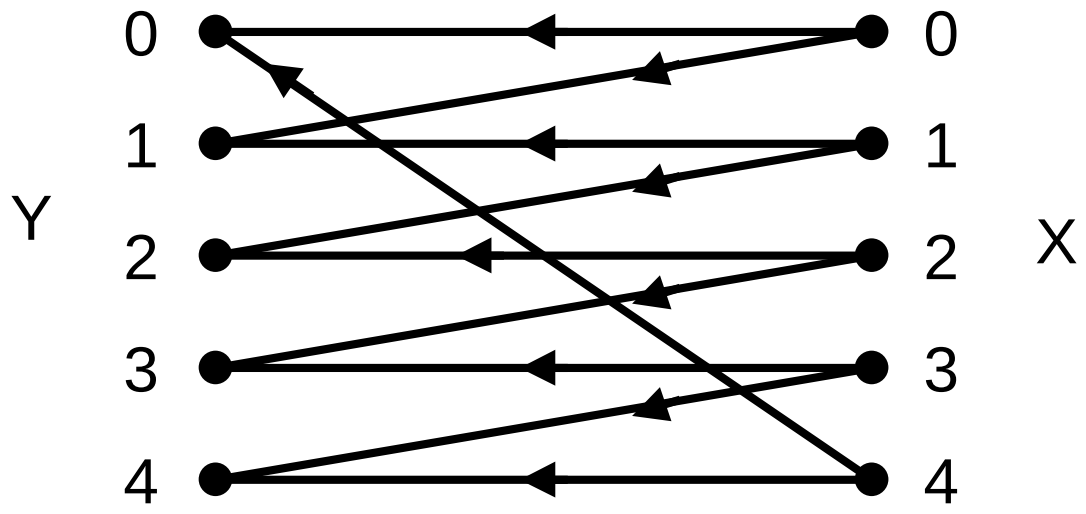


Claude Elwood Shannon (1916–2001)
(photograph 17 April 1961 by Göran Einarsson)

If the maximum number of non-confusable input letters is M, then the zero-error capacity of the channel satisfies

$$C_0 \geq \log_2 M \text{ bits/use.}$$

An example from Shannon:

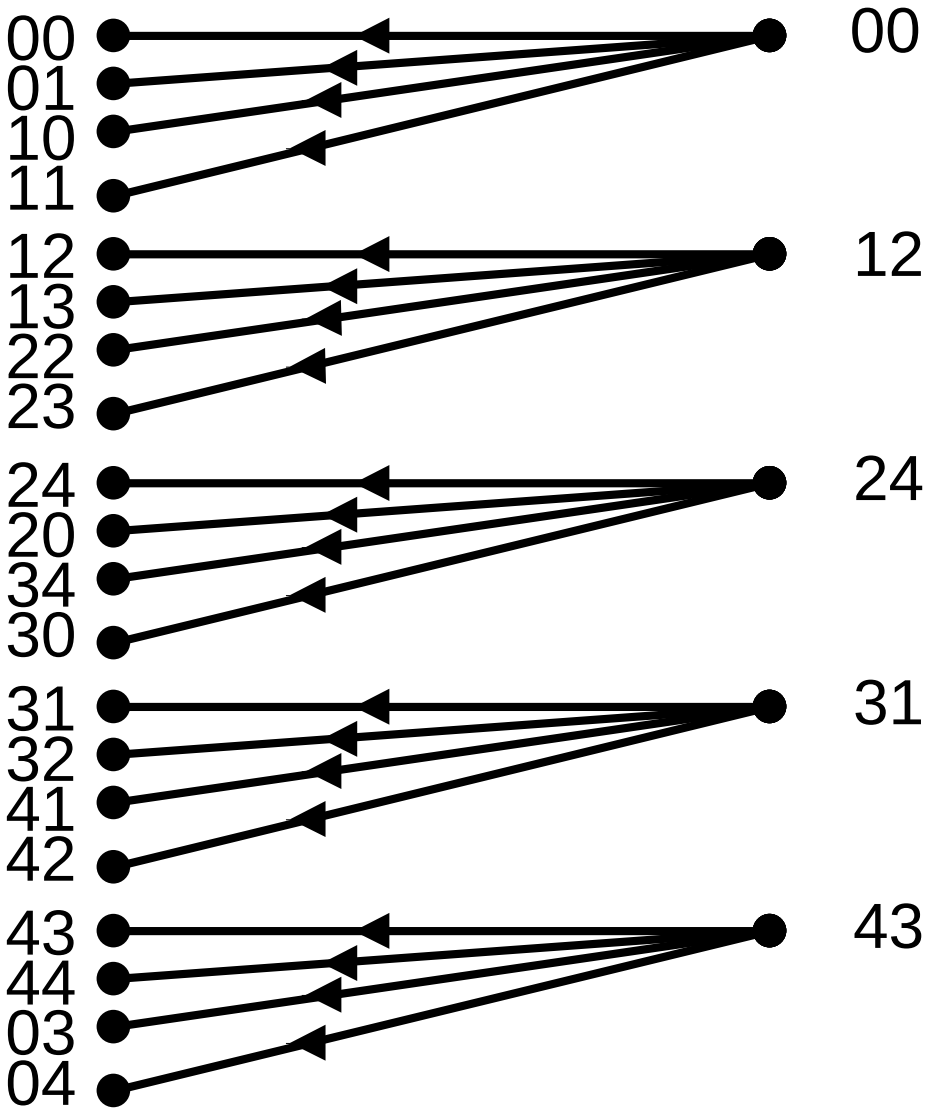


0 and 2 are non-confusable inputs.
0 and 3 are non-confusable inputs.
1 and 3 are non-confusable inputs.
1 and 4 are non-confusable inputs.
2 and 4 are non-confusable inputs.

No three input letters are non-confusable.

Is C_0 = 1 bit/use ?

Shannon pointed out that if we use only the following pairs of inputs: 00, 12, 24, 31, 43. Then the channel equivalently becomes



These **five inputs** for the compound channel are **non-confusable** so we know that

$C_o \geq (\log_2 5)/2 \approx 1.16$ bits/use. Is this number C_o ?

In a celebrated 1979 paper* that won the Information Theory Society's annual paper award, Lovász proved that $C_o = (\log_2 5)/2 \approx 1.16$ bits/use for Shannon's channel.

This was 23 years after Shannon's paper was published!

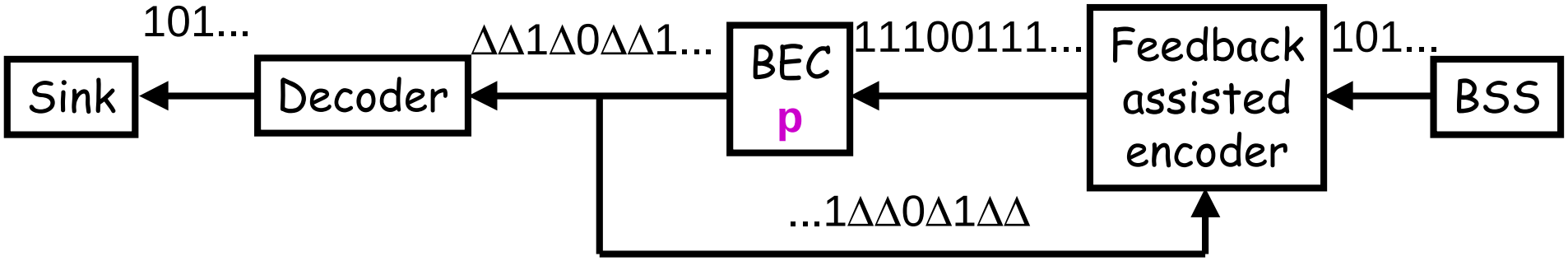
*László Lovász; On the Shannon capacity of a graph, IEEE Trans. Info. Theory, vol. IT-25, pp. 1 - 7, January 1979.

In spite of what I've said until now, I'm not really interested in zero-error capacity except when there is a feedback channel available.

First we consider what Shannon called **complete feedback** by which he meant that "there exists a return channel sending back from the receiving point to the transmitting point, **without error**, the letters actually received. It is assumed that this information is **received** at the transmitting point **before the next letter is transmitted**, and can be used, therefore, if desired, in choosing the next transmitted letter."

Shannon wrote C_{0F} to denote the **zero-error capacity with complete feedback**.

Example of the BEC:



(There is assumed to be a slight delay in the complete feedback line.)

The encoder is using the rule: transmit each information bit repeatedly until it is received unerased.

This coding scheme clearly gives **zero error** ! Moreover, the rate of transmission is **R** = 1 - **p** bits/use (the probability of success on each transmission) so that the BEC is used 1/(1 - **p**) times on average for each info. bit.

Can we conclude that $C_{oF} = 1 - p = C$???

Shannon says NO!.

Shannon proved in 1956 that if $C_o = 0$ then $C_{oF} = 0$.

How can we reconcile this statement with the example of the BEC ???

Shannon's allowed only "block" codes!

In his 1956 paper, "The zero error capacity of a noisy channel", Shannon in fact defined the **zero-error capacity C_{0F} with complete feedback** as the largest number such that, for every $\delta > 0$, **K** information bits from a Binary Symmetric Source can be sent with **N** uses of the DMC with complete feedback at a **rate R = K/N** > $C_{0F} - \delta$ bits/channel-use and with zero error.

But if $C_0 = 0$, then one cannot send one information bit with zero error over a BEC, even with an adaptive code of length N and no matter how large N may be because the received sequence can be $\Delta\ \Delta \ldots \Delta$ for both values of the information bit. The same argument applies to any channel with $C_0 = 0$.

In 1948, Shannon used the following definition:

The **small-error capacity C** is the largest number such that, for every $\varepsilon > 0$ and every $\delta > 0$, information bits from a Binary Symmetric Source can, with the use of a block code of sufficiently large length N, be sent over the channel at a **rate R** > **C** - δ bits/channel-use and with block error probability $P_B < \varepsilon$.

When one deals with "small error", there is no loss of generality in restricting oneself to block coding, but for "zero error" one loses generality. Shannon seems to have overlooked this point.

In his 1956 paper, Shannon proved that the **small-error capacity** of a discrete memoryless channel (the only kind of channel that we have been, and will be, talking about) is **not increased by the availability of complete feedback**. It may have seemed natural to him then that the zero-error capacity also should not be increased by the availability of complete feedback. Thus, he may not have reflected much over the generality of his restriction to "block" codes in his definition of C_{oF}.

There are very few places in his enormous record of contributions to information theory that Shannon adopted an unnecessarily restrictive approach—this appears to be one of them.

We will use the definition:

The **zero-error capacity C_{oFa} with complete feedback** is the largest number such that, for every $\delta > 0$, information bits from a Binary Symmetric Source can, with the use of some **adaptive coding scheme**, be sent with zero error over the channel at **rate R** > $C_{oFa} - \delta$ bits/channel-use and with **average coding delay** at most D_a, $D_a < \infty$, for every information bit.

(The "**a**" in C_{oFa} is intended as a reminder that we are considering "**a**verage coding delay".)

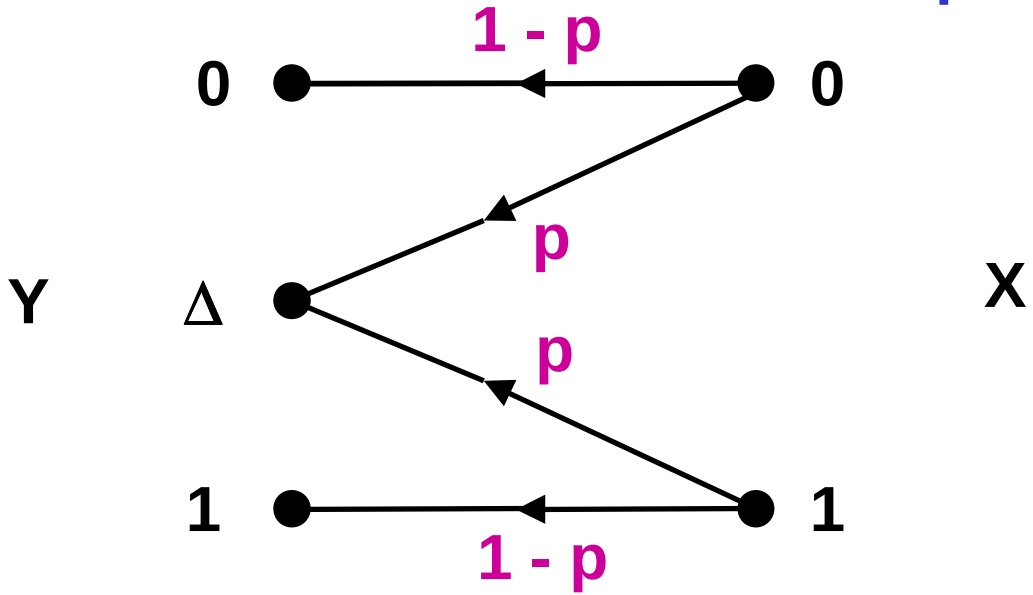
By the **coding delay** for an information bit, we mean the number of channel uses beginning when the information bit enters the encoder and ending when the information bit is assigned its value by the decoder.

Let D_i be the decoding delay for the i^{th} information bit. For a block code of length N, $D_i \leq N$ for all i.

Our previous example of adaptive coding for the BEC gives $D_i \leq 1/(1-p)$ for all i and hence

$$C_{oFa} = 1 - p = C.$$

We will say that an **output letter** y of a discrete memoryless channel is a **disprover for the input letter** x if y **cannot be reached from** x but **can be reached from at least one other input letter**.



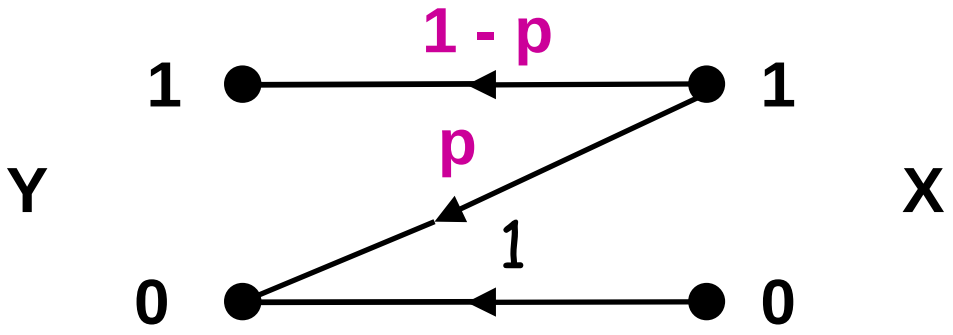
For the BEC with $0 \leq p < 1$, the output letter 0 is a disprover for the input letter 1, and the output letter 1 is a disprover for the input letter 0.

The idea behind the definition of a "disprover": if y is a disprover for x, then the appearance of y in the received sequence **proves** that the corresponding transmitted letter was **not** x.

What is the **simplest** noisy discrete memoryless channel with $C_0 = 0$ whose output letters include a disprover ?

Is it the Binary Erasure Channel ?

No! It is the **Z-channel** !



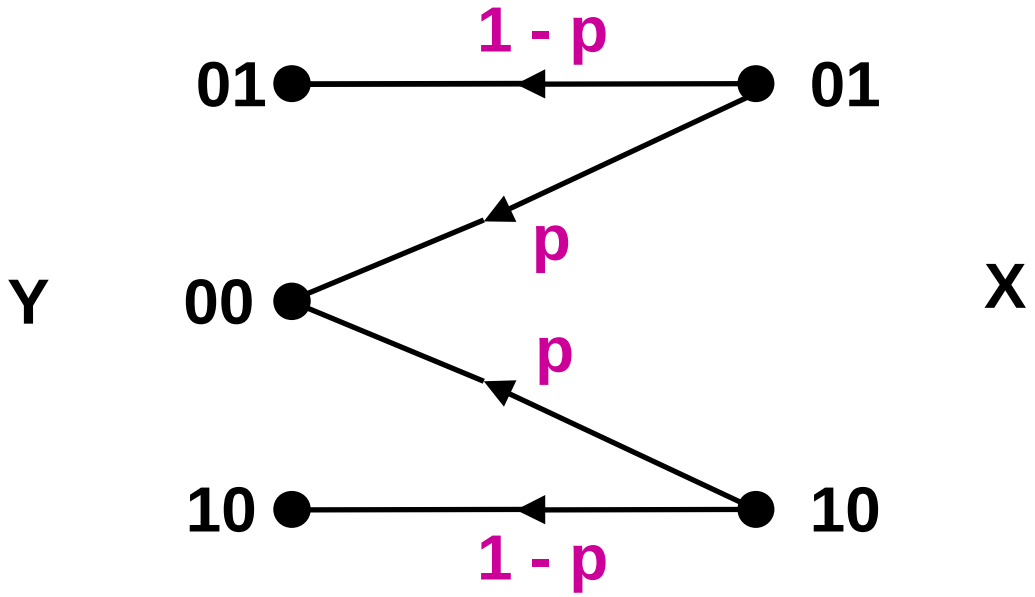
(0 < **p** < 1)

The output letter 1 is a disprover for the input letter 0.

Does this channel have $C_{oFa} > 0$?

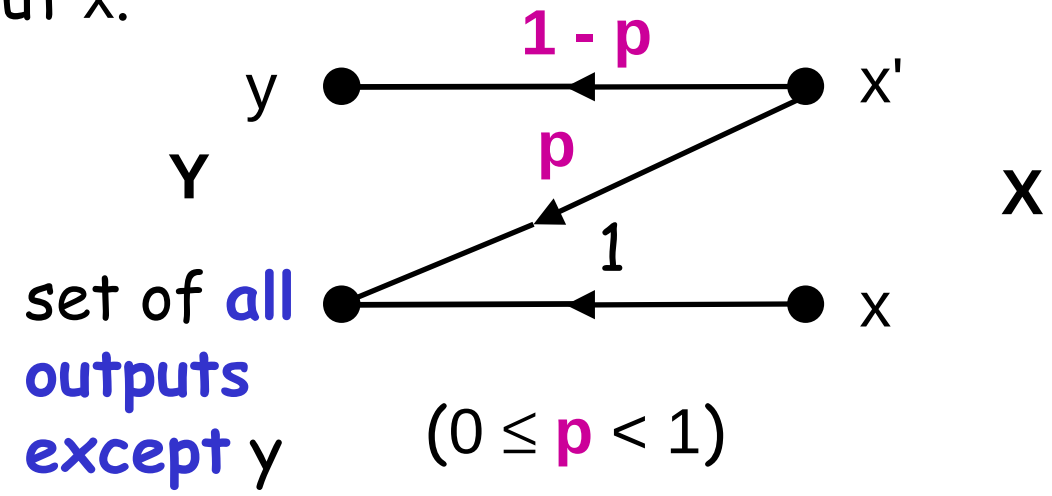
Yes!

Suppose we create a new channel by using the Z-channel for the input pairs 01 and 10 only.



Thus, by our previous result for the BEC, we know that for the Z-channel $C_{oFa} \geq (1 - p)/2$.

We can do this same "trick" for any discrete memoryless channel whose output alphabet contains at least one disprover y for some input x.



(x' is any input letter that can reach y.)

It follows that any such channel has

$C_{oFa} \geq (1 - p)/2 > 0.$

We have proved:

Theorem 1:
For a discrete memoryless channel, $C_{oFa} > 0$ if and only if the output alphabet contains at least one disprover y for some input letter x.

But what is the actual value of C_{oFa} ?
Is it **C** or is it in general smaller?

Theorem 2:
If a discrete memoryless channel has $C_{oFa} > 0$, i.e., if the output alphabet contains at least one disprover y for some input letter x, then $C_{oFa} = C$ (the small-error capacity).

Proof:
We can use the disprover y for the input letter x to create a Z-channel with probability **p** ($0 \leq p < 1$) to the receiver. We can use two uses of this Z-channel to create a Binary Erasure Channel (BEC) with erasure probability **p** as shown on slide 26. We will use this BEC to send a one-bit **ACK** or **NAK** to the receiver in the following way each time coded block is sent on the forward channel:

Suppose we have a block code of length N for sending K information bits in N uses of the forward channel with rate $R_B = K/N > C - \delta_B$ and block error probability $P_B < \varepsilon_B$.

After we send a block, we know from the complete feedback whether the decoding was correct or not. If **yes**, we use our created BEC to send a **1** (**ACK**) to the receiver without error with an average of $D = 2/(1 - p)$ channel uses. If **no**, we use our created BEC to send a **0** (**NAK**) to the receiver without error, again with an average of $D = 2/(1 - p)$ channel uses. The **ACK** informs the receiver that a **new block** will now be **transmitted**; the **NAK** informs the receiver that the **previous block** will be **retransmitted**.

The probability of success (an **ACK**) on each block transmission is $1 - P_B$ so the average transmission rate is

$$R = \frac{K}{N + D}(1 - P_B) = R_B \frac{1}{1 + D/N}(1 - P_B)$$

$$\geq (C - \delta_B) \frac{1}{1 + D/N}(1 - \varepsilon_B)$$

Thus, for any given $\delta > 0$, we can choose δ_B and ε_B sufficiently small and N sufficiently large so that

$$R \geq C - \delta,$$

which proves Theorem 2.

Consider now **noiseless**, but **not necessarily complete**, feedback.

How much noiseless (but not complete) feedback is needed to approach the zero-error capacity with feedback, C_{oFa} ?

A first answer:

There are coding schemes using noiseless feedback for sending with zero error at any rate less than $C_{oFa} = C$ over a DMC with $C_{oFa} > 0$, that approach arbitrarily closely to **one binary digit of noiseless feedback** for **each bit of information** transmitted on the forward channel.

In our **ACK**/**NAK** block coding scheme for approaching C_{oFa}, **$K = NR_B$ binary digits of noiseless feedback suffice for each block of length N transmitted on the forward channel**. The receiver can simply send back the **K decoded information bits** over the noiseless channel. (This would mean that the receiver has to wait until decoding was complete to send the feedback and hence the sender would have to interleave the transmission of two blocks to keep the forward channel busy.) The repetitions of "**NAK**"ed blocks will cause the average number of binary digits sent on the noiseless feedback channel to slightly exceed K for each block of K information bits, but one can make this average approach K arbitrarily closely by choosing N sufficiently large.

Are there DMCs with $C_{oFa} > 0$ and $C_o = 0$ for which one can send with zero error using much **less than one binary digit of noiseless feedback for each bit of information** sent over the forward channel?

Is it possible to make the number of binary digits of noiseless feedback for each bit of information sent over the forward channel **approach zero**?

YES, to both questions.

Proposition 1:
There are coding schemes using noiseless feedback for sending with zero error at any rate less than $C_{oFa} = C = 1 - p$ over a BEC with $0 < p < 1$, in which the number of **binary digits of noiseless feedback** for **each bit of information** can be made to approach **0** arbitrarily closely.

The main idea:
For block coding on the BEC, if there is **only one codeword** that **agrees** with the received word in **all unerased positions**, then the decoder can output this codeword with no possibility of error.

We will call such certain-to-be-correct decoding **unambiguous decoding** and denote its probability by P_{UA} and we write $P_A = 1 - P_{UA}$ for the probability of **ambiguous decoding**.

Let P_{ML} be the probability of error for **maximum-likelihood decoding**.

Proposition 2:
For block coding on a BEC with $0 < p < 1$,
$P_A \leq 2 P_{ML}$.

Suppose **y** is the received block and **x** is a codeword. $P(\mathbf{y}|\mathbf{x}) = 0$ unless **x** agrees with **y** in all unerased positions in which case $P(\mathbf{y}|\mathbf{x}) = (1-p)^{N-e} p^e$ where e is the number of erasures.

$\Rightarrow$ every "all-agreeing" codeword is a valid choice for the maximum-likelihood decoding decision.

The correct codeword must agree with **y** in all unerased positions. Thus the conditional probability of error for the ML decoder must be 0 when the decoding is unambiguous and is at least 1/2 when decoding is ambiguous. It follows that $P_{ML} \geq (1/2)P_A$.

Suppose we have a block code of length **N** for sending **K** information bits in **N** uses of the BEC at rate $R_B = K/N > C_{oFa} - \delta_B = 1 - p - \delta_B$ with maximum-likelihood block error probability $P_{ML} < \varepsilon_B$.

After receiving a block, the receiver sends a **1** (**ACK**) to the sender on the noiseless feedback channel if the decoding was **unambiguous** (which tells the sender to **transmit a new block**). If the decoding was **ambiguous**, the receiver sends a **0** (**NAK**) on the noiseless feedback channel (which tells the sender to **retransmit the current block**).

The probability that a **NAK** will be sent for any block is $P_A \leq 2P_{ML} < 2\varepsilon_B$. Thus the average number of binary digits of noiseless feedback for $K = NR_B$ information bits is less than $1/(1 - 2\varepsilon_B)$.

Do coding schemes for sending with zero error at rates approaching $C_{oFa} = C = 1 - p$ over a BEC with $0 < p < 1$ actually **require noiseless feedback**?

Would a Z-channel with $0 < p_z < 1$ be good enough for the feedback channel?

Would any DMC with $C_{oFa} > 0$ be good enough for the feeback channel?

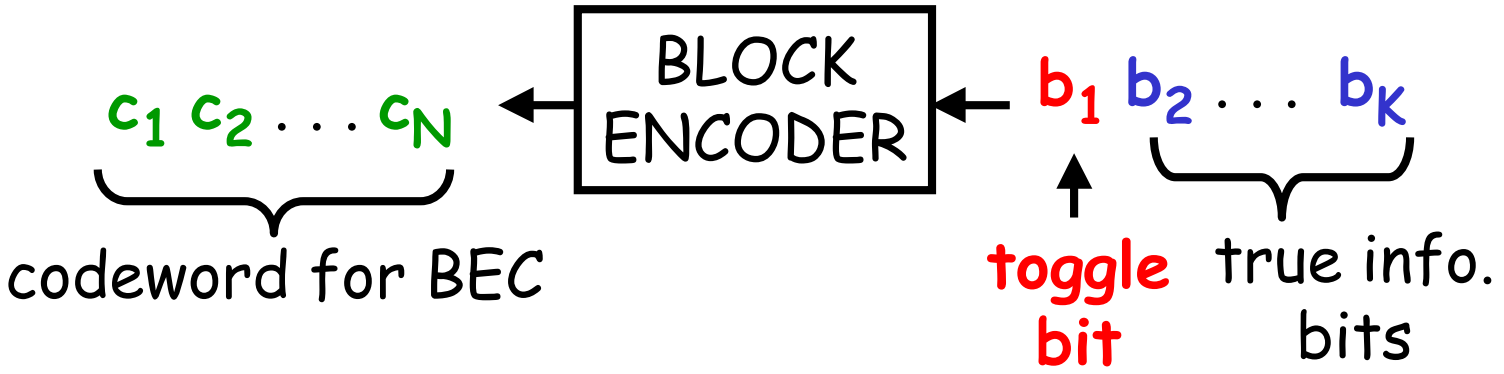
Are there DMCs with $C_{oFa} = 0$ that are good enough for the feedback channel?

The answers are **NO**, **YES**, **YES** and **NO**.

Suppose when the decoding is **unambiguous** the receiver feeds back a string of **n 1**'s (**ACK**s) to the sender on the Z-channel and when the decoding is **ambiguous** feeds back a string of **n 0**'s (**NAK**s) on the Z-channel. The probability that at least one **ACK** will get through is $1 - p_z^n$. If the sender gets no **ACK** for a block, the sender repeats the block.

But **this doesn't quite work** as is because the receiver would not know in general whether the decoded block is a repetition or is a new block.

The sender solves this problem by making the first information bit in each block a **toggle bit**, which is initially set to 0 and which is complemented each time the sender transmits a new block.



The receiver maintains a **test bit**, which is initially 0, and **recognizes a new unambiguously decoded block** by the fact that it yields a **toggle bit equal to the test bit**, following which recognition the receiver complements the test bit.

The analysis is the same as for the case of noiseless feedback except that

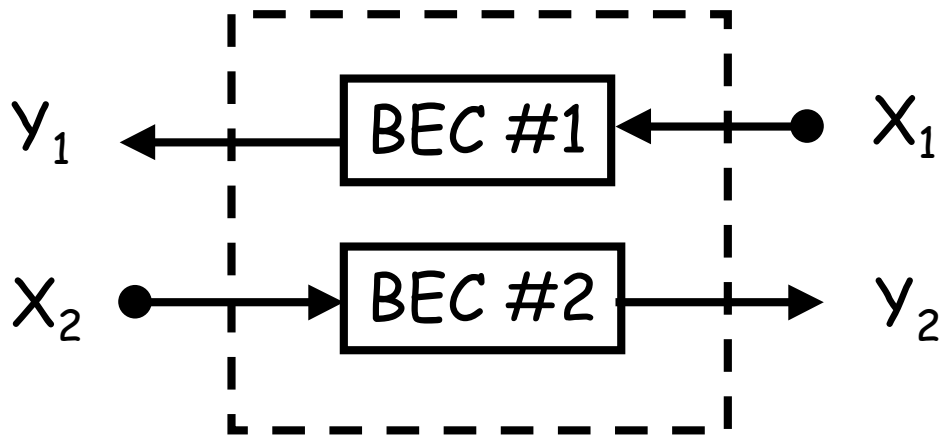
- The probability of block retransmission, P_A, is replaced by $P_A + (1- P_A)p_z^n$ when computing the average number of uses of the feedback channel per block of information bits, and
- there are now $K - 1 = NR_B-1$ information bits, rather than $K = NR_B$, in each block.

It should be obvious that, for proper choice of n, these changes have a negligible effect on the average rate of information transmission and that the average number of binary digits of noiseless feedback per information bit can still be made arbitrarily small.

Theorem 1 implies that, as illustrated on slide 27, **every DMC with $C_{oFa} > 0$** contains an **embedded Z-channel** and hence can be used as the feedback channel in the previously described zero-error C_{oFa}-approaching coding scheme for the BEC.

If $C_{oFa} = 0$, then it follows from Theorem 1 that the DMC output alphabet contains **no disprovers** and hence from a received sequence on this channel **one can never be certain that one of two possible codewords was not sent**. Thus, such a DMC cannot be used as the feedback channel in any zero-error coding scheme for the BEC.

The foregoing results imply that for a **two-way channel** consisting of two BECs, i.e.,



one can transmit with zero error at rates R_1 and R_2 **simultaneously approaching the capacities** C_1 and C_2 of the individual BECs (because the ACK/NAK feedback transmissions can be made with an arbitrarily small fraction of channel usage).

Do coding schemes for sending with zero error at rates approaching $C_{oFa} > 0$ with erasure feedback* over a DMC with $C_o = 0$ **require that this forward channel be an erasure channel**?

N.B. One can certainly send with zero error at some positive rate with erasure feedback over any DMC with $C_o = 0$ and $C_{oFa} > 0$ because such a channel can be converted to a BEC with positive capacity.

*or, equivalently, with any DMC having $C_o = 0$ and $C_{oFa} > 0$ as the feedback channel, because every such channel can be converted to a BEC channel as shown on slides 27 and 28.

Our **guess** is that this **forward channel must be an erasure channel**, but we are not sure of this.

The next example suggests that we can come fairly close to capacity when the forward channel is a Z-channel.

Suppose the forward channel is a Z-channel with $0 < p < 1$. If one uses a **constant-weight block code**, then $P(\mathbf{y}|\mathbf{x}) = 0$ **unless the codeword** $\mathbf{x}$ **agrees with the received block** $\mathbf{y}$ **in all positions where** $\mathbf{y}$ **contains a 1**, in which case

$$P(\mathbf{y}|\mathbf{x}) = p^{w-w'}(1-p)^{w'}$$

where w is the Hamming weight of a codeword and w' is the Hamming weight of $\mathbf{y}$. Thus decoding is **unambiguous** if there is only one such codeword and otherwise all (and at least two) such agreeing codewords are valid choices for a maximum-likelihood decoder. Again this implies $P_A \leq 2P_{ML}$. But it does not seem possible to reach the capacity of the Z-channel with such constant-weight coding.