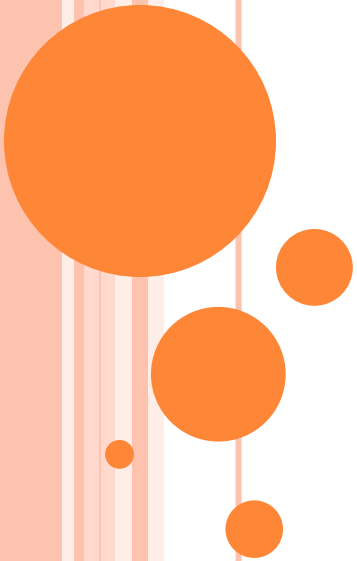


CS3236 Introduction to Information theory

Lecture 10: Error-Correcting Codes

Course given by Pierre Senellart
Material by Stephanie Wehner, with additions by P. Senellart

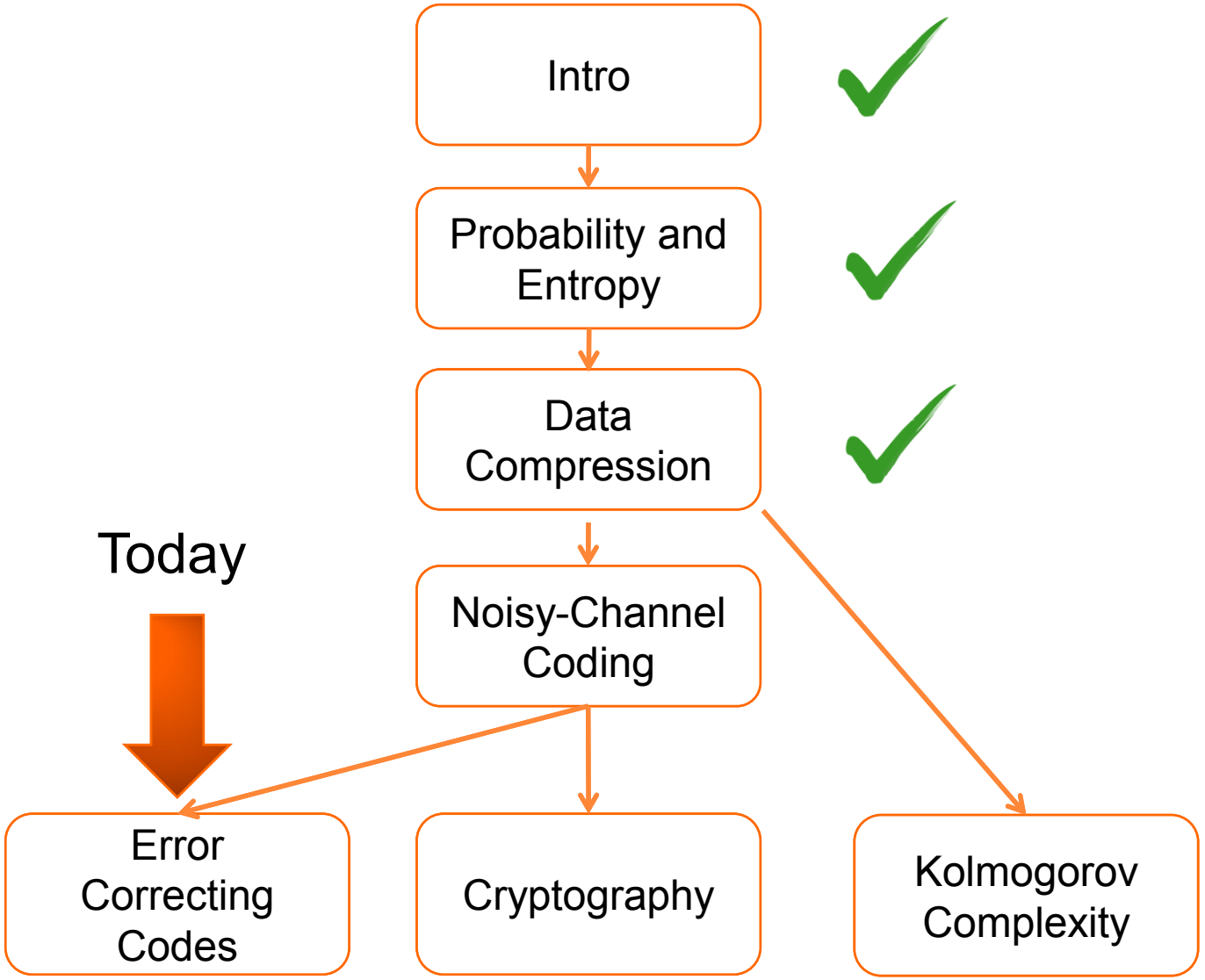
Some admin: mid-term exam coming up!

- Next Monday: Mid-term exam during lecture time!
  - Multiple-choice exam
  - Closed book
  - Counts for 10% of your final mark
  - Example from last year can be found on IVLE
  - This lecture is not in the scope of the mid-term, everything else is (i.e., Chapters 1-2, 4-6, 8-10 of the textbook)
- Caution: the final exam is
  - NOT multiple choice
  - … but open book

What we've done last time

- Shannon-noisy channel coding theorem
- Capacity forms sharp threshold for information transmission
- There exists an arbitrarily good code with rate R as close as the capacity as required, if we are allowed to increase block length arbitrarily

Where do we go from here?



What we'll do today

- Error-correction
  - Distance and decoding
  - Comparison to Shannon's noisy channel coding theorem
- Error correcting codes as matrices
  - Generator matrix
  - Syndrome decoding
- Modern error-correction example
  - Polar codes

Explicit error correcting codes

- Remember our examples:
  - Repetition code
  - (7,4)-Hamming Code

- Remember the distance of the code:

$$d(C) = \min_{x_1, x_2 \in C} H(x_1, x_2)$$

Hamming distance

Is the distance of the code all we care about?

- We know that a code with distance d is guaranteed to correct a number of errors equal to

$$t = \left\lfloor \frac{d-1}{2} \right\rfloor$$

Bounded distance decoding

- For a received y=r
  - Find the closest codeword $x^{(s)}$
  - If the distance of $x^{(s)}$ and r is smaller than $t = \lfloor \frac{d-1}{2} \rfloor$ then output s
    otherwise output failure

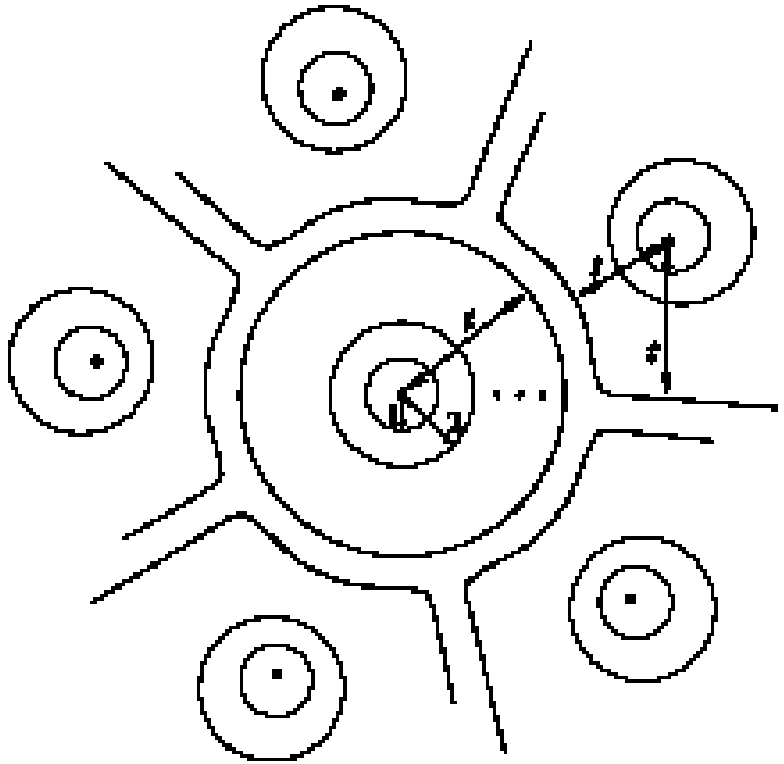
Can such a decoder ever achieve the Shannon limit?

… large distance is certainly a good thing…

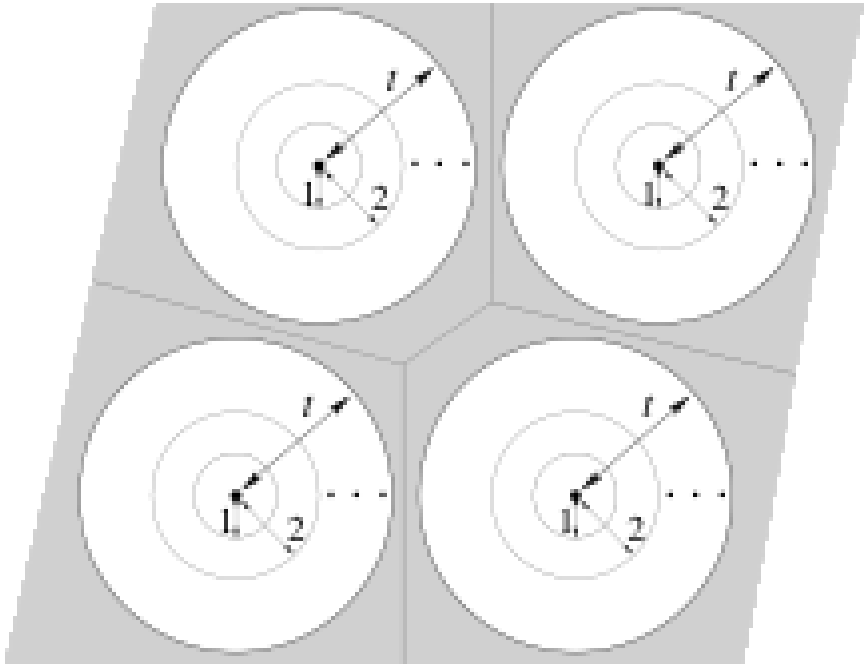
"Perfect" codes

- t-spheres around each code word fill space but don't overlap

Perfect code



Not a perfect code



Examples of perfect codes

- (7,4)-Hamming code (for t=1)
- Generalized Hamming code (for t=1)

- Repetition codes of odd blocklength for t=(N-1)/2
  - Remember they have vanishing rate

- Golay code (see book)

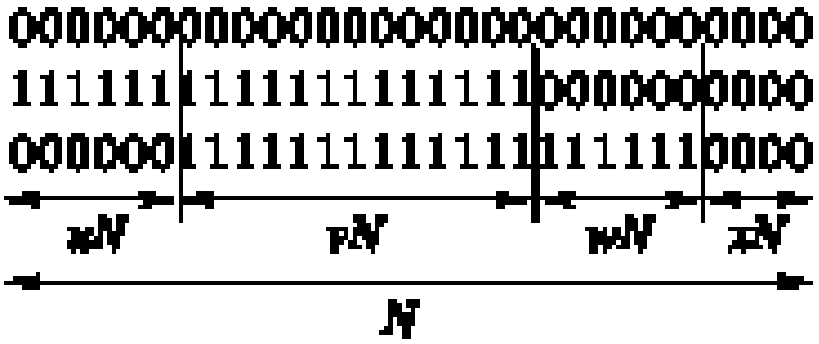
- There are no other perfect binary codes… but does it matter?

Comparison to a capacity achieving code

- We know from Shannon that for the BSC there exist codes with blocklength N and R arbitrarily close to C
- For large N, there will be about f × N errors that the code (almost certainly) corrects
- Can we find a *perfect* code that is fN error correcting?

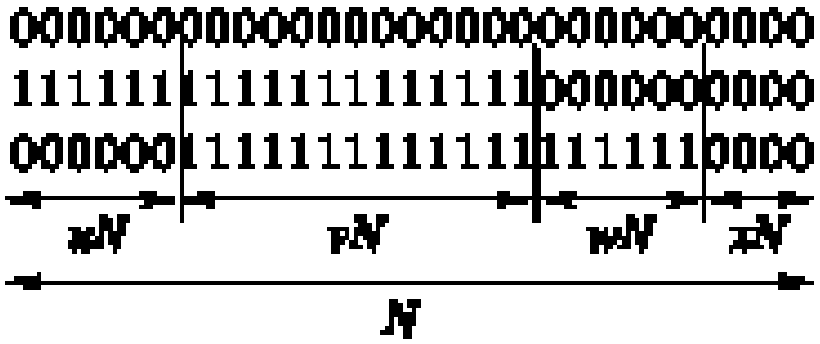
No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords



No such perfect code exists

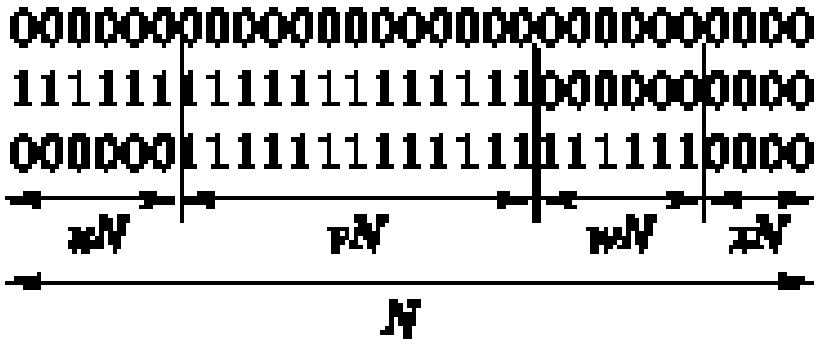
- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords



$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords

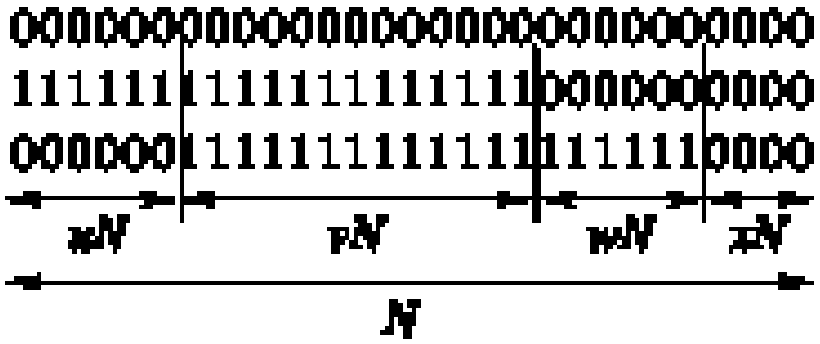


$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

$$u + v + w > 3f$$

No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords

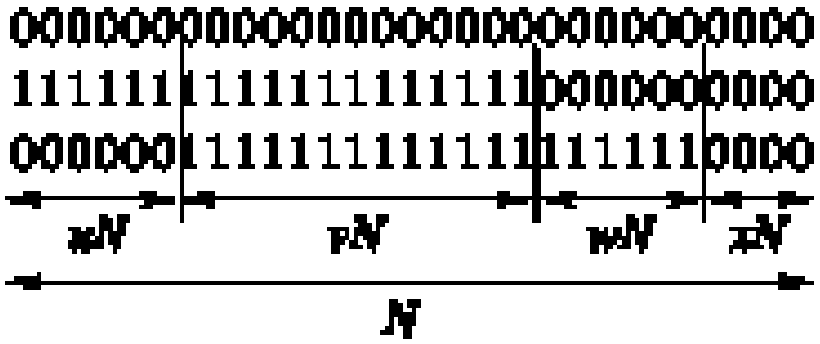


$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

$$u + v + w > 3f \xrightarrow{f > 1/3} u + v + w > 1$$

No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords

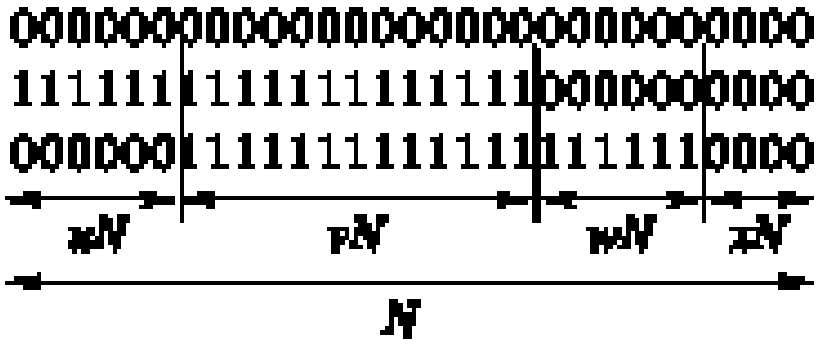


$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

$$u + v + w > 3f \xrightarrow{f > 1/3} u + v + w > 1$$

No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords

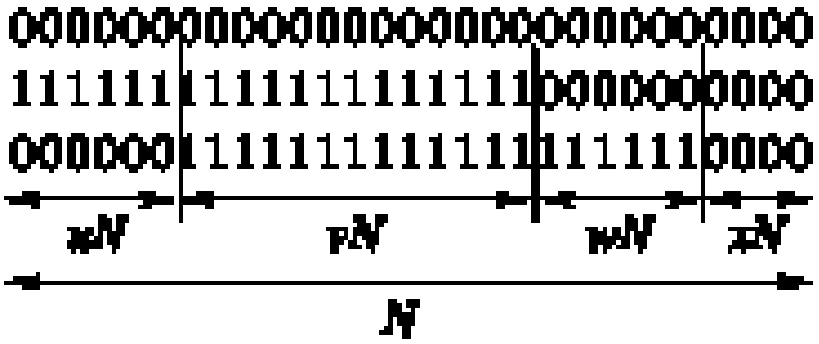


$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

$$u + v + w > 3f \xrightarrow{f > 1/3} u + v + w > 1$$

No such perfect code exists

- Suppose by contradiction that there existed such a code with the same rate
  - Must have distance at least $2fN$
  - Certainly has at least 3 codewords



$$u + v > 2f$$
$$u + w > 2f$$
$$v + w > 2f$$

$$u + v + w > 3f \xrightarrow{f > 1/3} u + v + w > 1$$

Contradiction!

Gilbert Varshamov conjecture

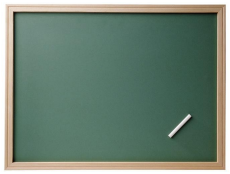
- Gilbert Varshamov distance

$$d_{GV} = NH_2^{-1}(1-R)$$

$$H_2(d_{GV}/N) = 1-R$$

- Gilbert Varshamov conjecture
  - For large N, it is not possible to create a binary code with minimum distance d much larger than d_{GV}

If this conjecture were true…



- Bounded distance decoder can correct a maximum "amount" of flipped bits

$$f \approx \frac{d}{2N}$$

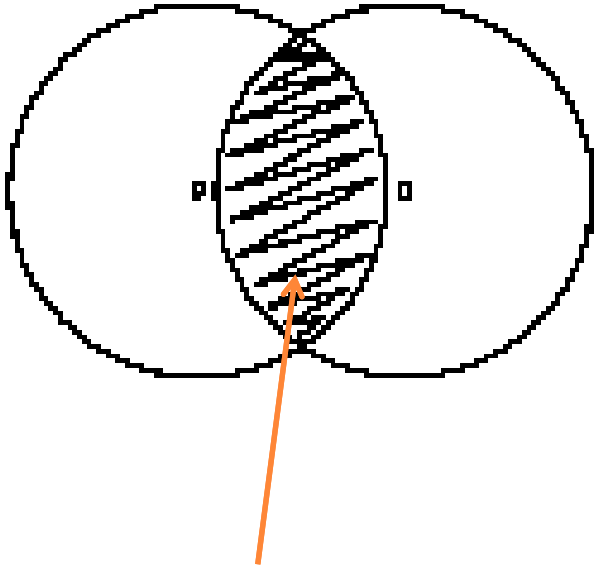
- If the conjecture were true could tolerate $f \approx \frac{d_{GV}}{2N}$

$$H_2(d_{GV}/N) = 1 - R$$

- Thus $R_{GV} = 1 - H_2(2f)$
- But Shannon tells us that we can tolerate twice the noise level with this rate

$$C(BSC) = 1 - H_2(f)$$

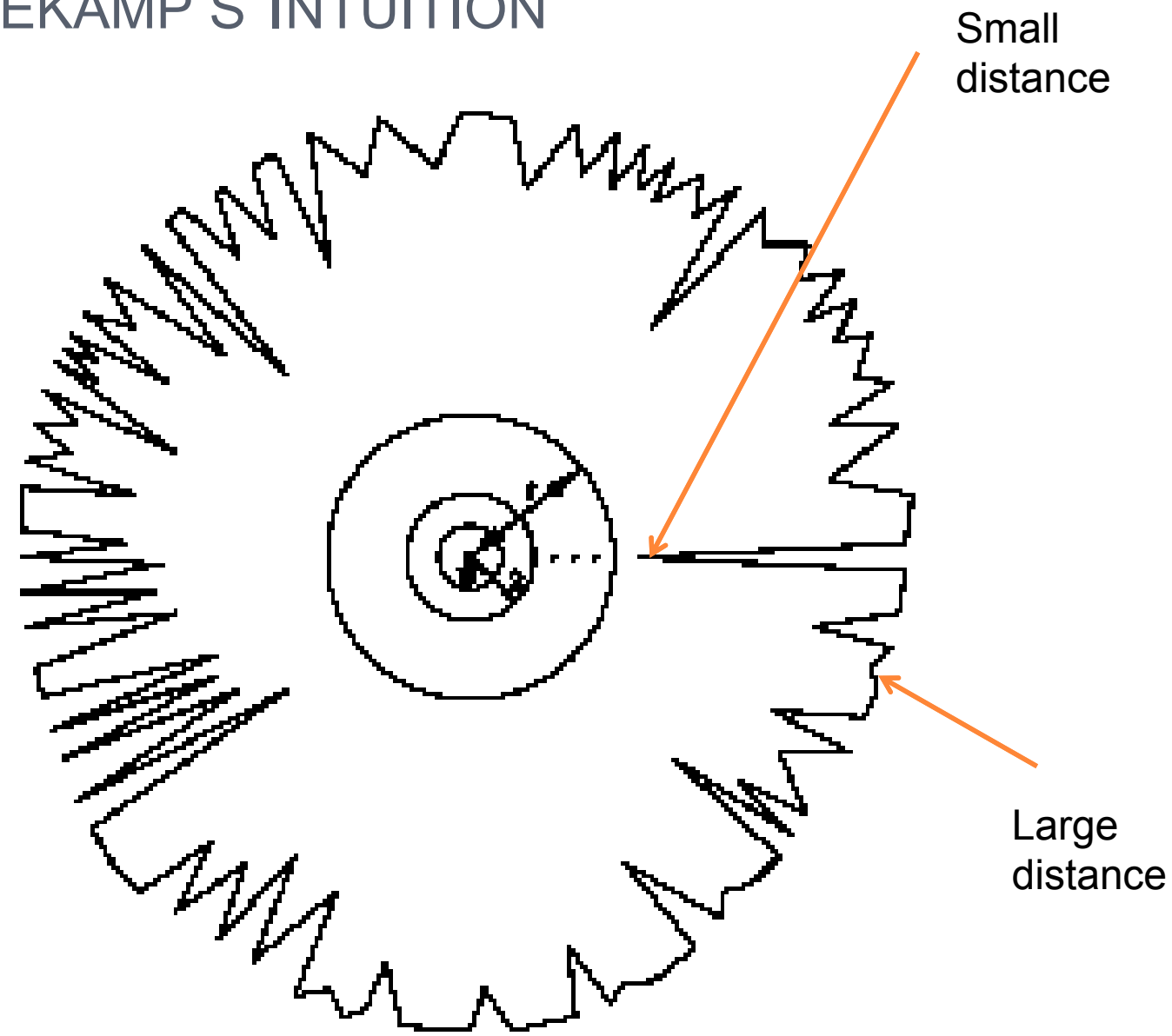
Why sheer distance does not always matter



Overlap between t-spheres around code words

Intuition: in high dimension the overlap is small, so the probability that we end up in it is small.

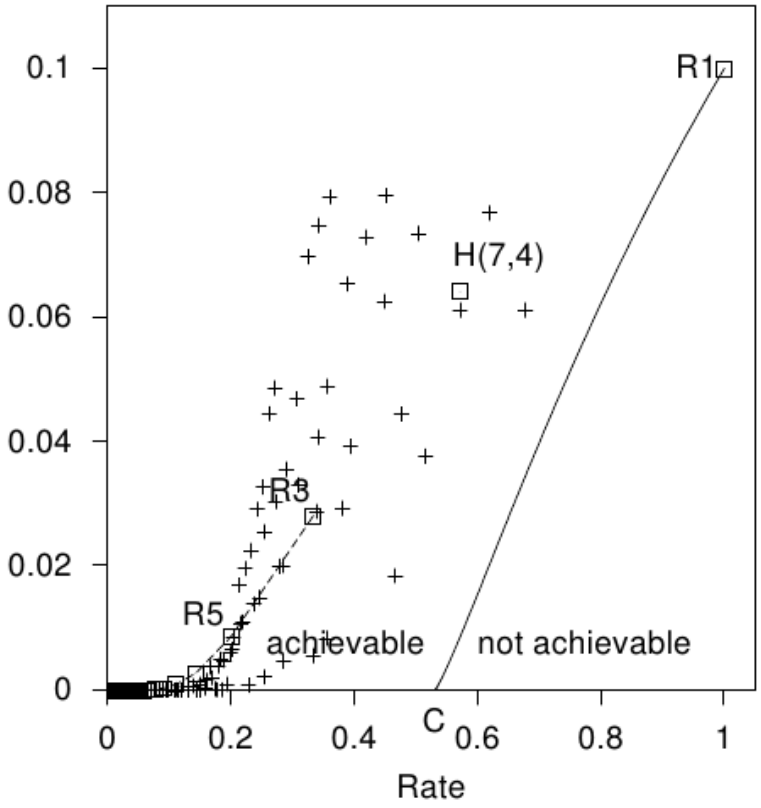
Berlekamp's Intuition

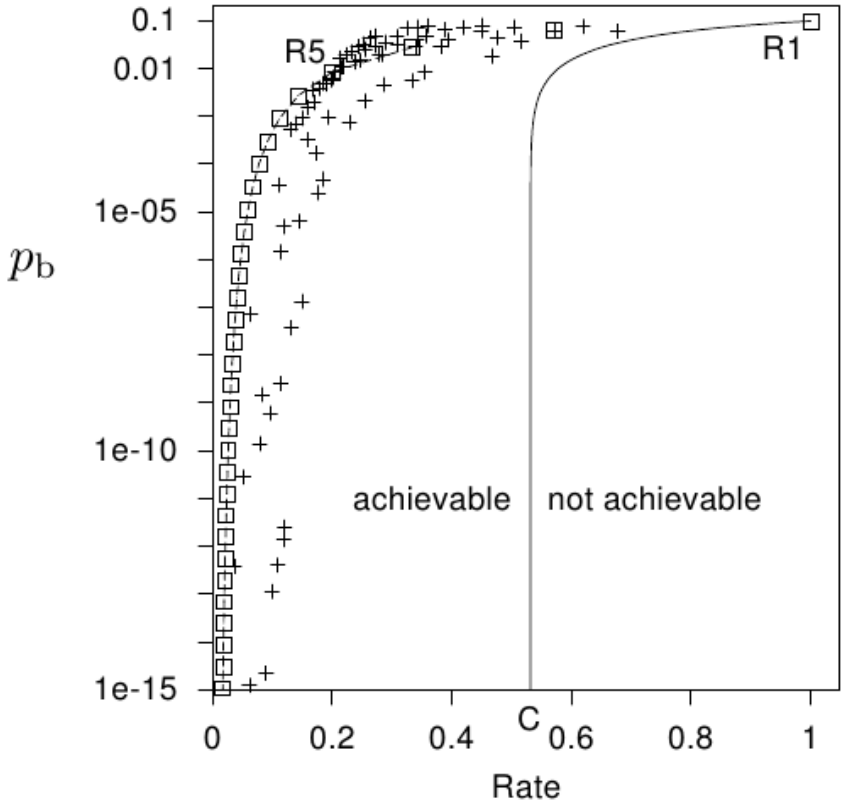


Modern Error-correcting codes

- Many modern error-correcting codes have relatively small minimum distance
- Yet, they perform very well (close to the Shannon limit)
- Examples: LDPC, Polar codes,…

ARE THERE AT ALL BETTER CODES?





- Maybe to have the error arbitrarily close to 0 we must accept an arbitrarily small rate? (after all… no pain, no gain…)

Shannon's noisy channel coding theorem (Part 1: Achievability)

- Associated with every discrete memoryless channel, there is a non-negative number *C* (the capacity) such that
  - For any error $\epsilon > 0$ and $R \leq C$ for large enough *N*, there exists a block code of length *N* and rate *R*, an a decoding algorithm, such that the maximum probability of block error is $< \epsilon$

There “exists” an excellent code…

There "exists" an excellent code…

But how on earth can we find one that's practical ?

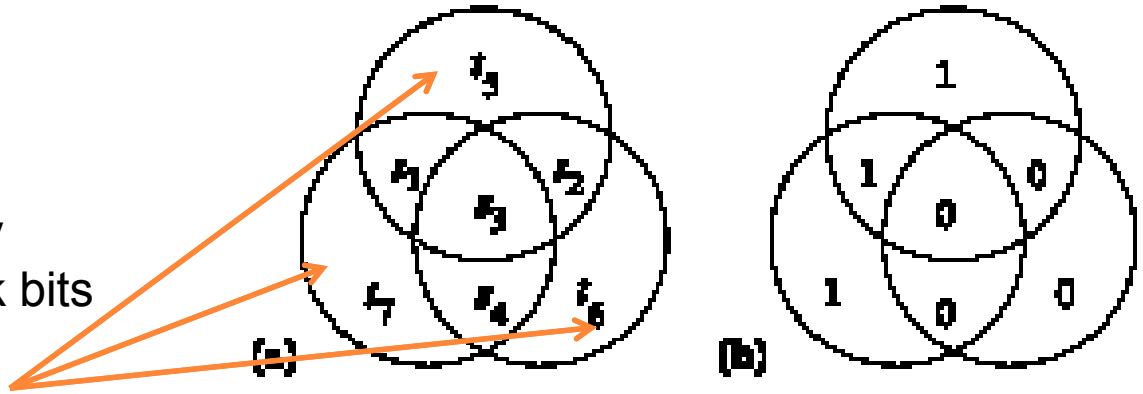
☺

(7,4)-Hamming code in the matrix

- Remember error correcting codes
- Example from the first lecture

(7,4)-Hamming code

Parity check bits



$s_1 s_2 s_3 s_4 t_5 t_6 t_7$

s	t	s	t	s	t	s	t
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

Hamming code in terms of matrices

$$1 + 1 \mod 2 = 0$$

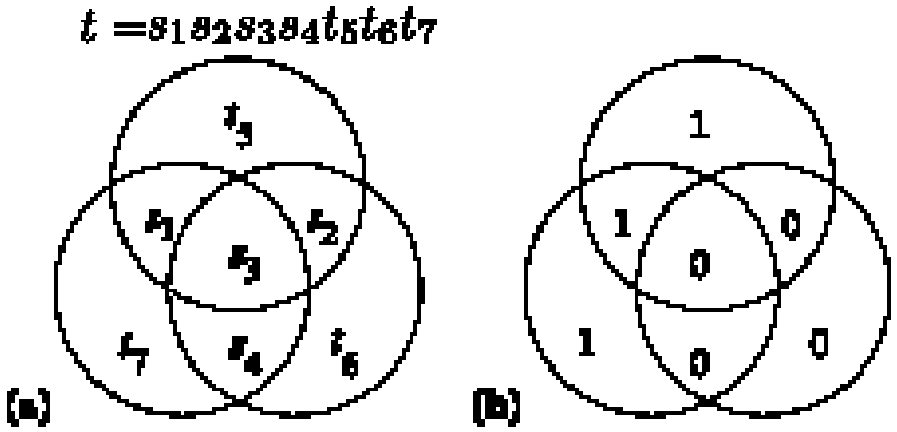
- All bits are a linear function of the source bits!

Column vectors

$$t = G^T s$$

$$G^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = G^T \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$



Generator matrix

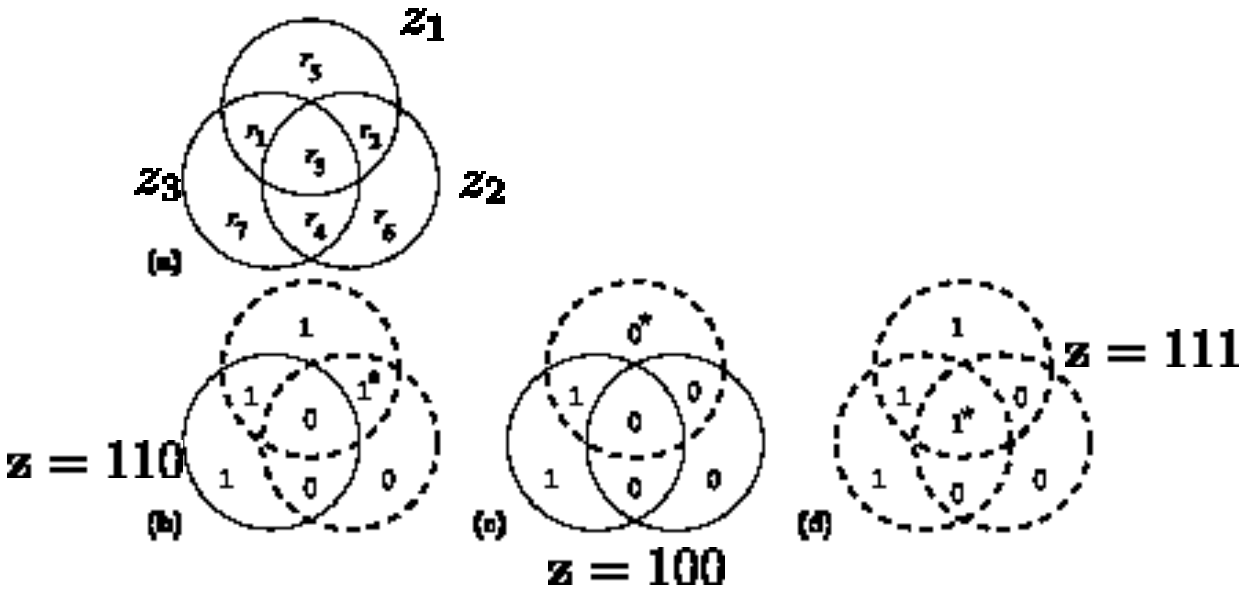
- In our example, G is called the generator matrix

$$G^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \qquad G = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & 0 & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & 1 & 1 & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & 0 & 1 & \mathbf{1} \end{bmatrix}$$

- Rows of the generator matrix define a basis for the codeword space
- Any codeword is a linear combination of these rows
- In standard form when G starts with the identity matrix

Remember.. Syndrome decoding

- The syndrome $\mathbf{z} = z_1 z_2 z_3$ tells us which circle is violated



If $\mathbf{z} = 000$ it means that no error is detected!

Now we will see how this code can easily be implemented as a matrix and decoded without the visual picture.

Syndrome decoding

- We check what the parities should be for the first 4 bits, and whether it matches what we got
- The difference (mod 2) is called the syndrome

- Can be computed using matrices!

$$G^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} I_4 \\ P \end{bmatrix}$$

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

Parity check matrix

$$G^T = \begin{bmatrix} I_4 \\ P \end{bmatrix} \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

3x4

- The parity check matrix is given by

$$H = [-P \; I_3]$$

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- Can find the syndrome by computing

$$z = Hr$$

$$z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

Syndrome

Received vector

$$r = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \end{pmatrix}$$

Codewords and parity check matrix

- If no error occurred then the syndrome is 0
  - For any codeword $t = G^T s$ we have

$$Ht = 0$$

- Syndrome decoding
  - What we receive is a codeword plus noise

$$r = G^T s + n$$

Noise vector flipping bits (all additions mod 2!)

  - Syndrome decoding

$$Hr = H(G^T s) + Hn = 0 + Hn = Hn = z$$

  - Find the most probable noise vector such that

$$Hn = z$$

Matrices for (n,k)-linear codes

- All linear codes can be specified with the help of the generator matrix (or equivalently the parity check matrix)
- Both can be brought into systematic (standard) form

$$G = [I_K \; P^T]$$
$$H = [-P \; I_{N-K}]$$

- K x N generator matrix G: K vectors from which all codewords can be build by taking linear combinations
- (N-K) x N parity check matrix: vectors to which all codewords are orthogonal

Another example

$$G = [I_K \; P^T]$$
$$H = [P \; I_{N-K}]$$

- 3 bit repetition code. K = 1, N = 3

$$G = [1 \; 1 \; 1]$$

$$P^T = [1 \; 1]$$

$$H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Dual code $C^{\perp}$

- Set of all vectors of length N that are orthogonal to all codewords in a code C

$$h^T t = (h_1, \ldots, h_N) \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix} = 0$$

- But we know a basis for that set of vectors!

$$Ht = 0$$

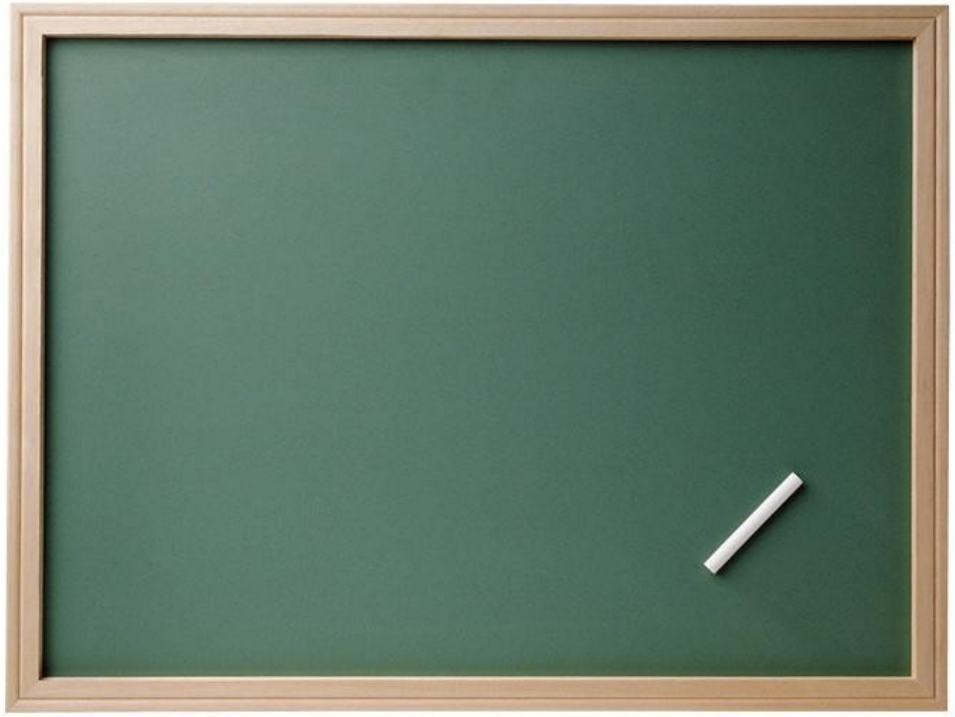
- Dual code given by the parity check matrix (seen as generator matrix!)
- All codewords of the dual code are linear combinations of the rows of the parity check matrix

Caution: What is orthogonal here?

- Set of all vectors of length N that are orthogonal to all codewords in a code c

$$h^T t = (h_1, \ldots, h_N) \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix} = 0$$

Example: repetition code

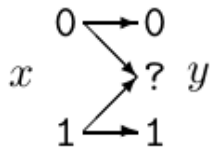


Capacity-Achievable codes exist!

- Polar codes: Invented in 2009 by Erdal Arikan
- Linear binary codes for any block of length $N = 2^n$
- Achieve capacity for binary symmetric channels!
  - … and in general "symmetric" capacity, where P(x) is fixed to be uniform
- Efficient to encode
  - O(*N* log *N*)
- Efficient to decode
  - O(*N* log *N*)

EXAMPLE: BINARY ERASURE CHANNEL (BEC)

Binary erasure channel. $\mathcal{A}_X = \{0, 1\}$. $\mathcal{A}_Y = \{0, ?, 1\}$.



$$
\begin{array}{lcllcl}
P(y=0\,|\,x=0) & = & 1-f; & P(y=0\,|\,x=1) & = & 0; \\
P(y=?\,|\,x=0) & = & f; & P(y=?\,|\,x=1) & = & f; \\
P(y=1\,|\,x=0) & = & 0; & P(y=1\,|\,x=1) & = & 1-f.
\end{array}
$$

$$C = \max_{\mathrm{Pr}_X} I(X;Y)$$

How can such a code be constructed?

- Need to find
  - Generator matrix
  - … Show encoding is more efficient than simply multiplying
  - … Show there is an efficient decoder

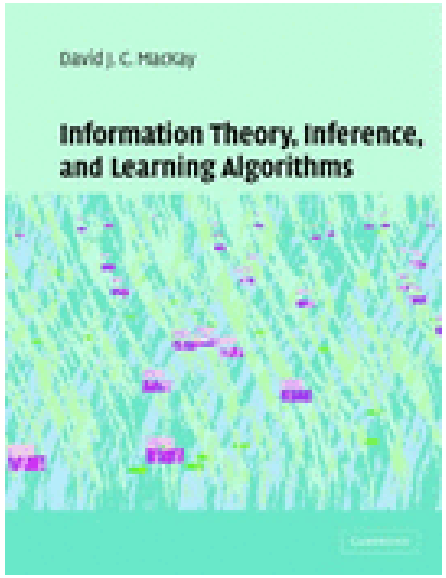
Could make the code depend on the channel… as in Shannon's theorem!

What we did today

- Error-correction
  - Distance and decoding
  - Comparison to Shannon's noisy channel coding theorem
  - Error correction as matrices
- Example of modern error-correction: Polar codes!

Reading for this lecture

- Chapter 13 in the textbook



Information Theory, Inference and Learning Algorithms
by David J. C. MacKay
Cambridge University Press, 2003

- Homework due by Monday 2pm two weeks from now (mid-term next week, no tutorial next week because of Deepavali)