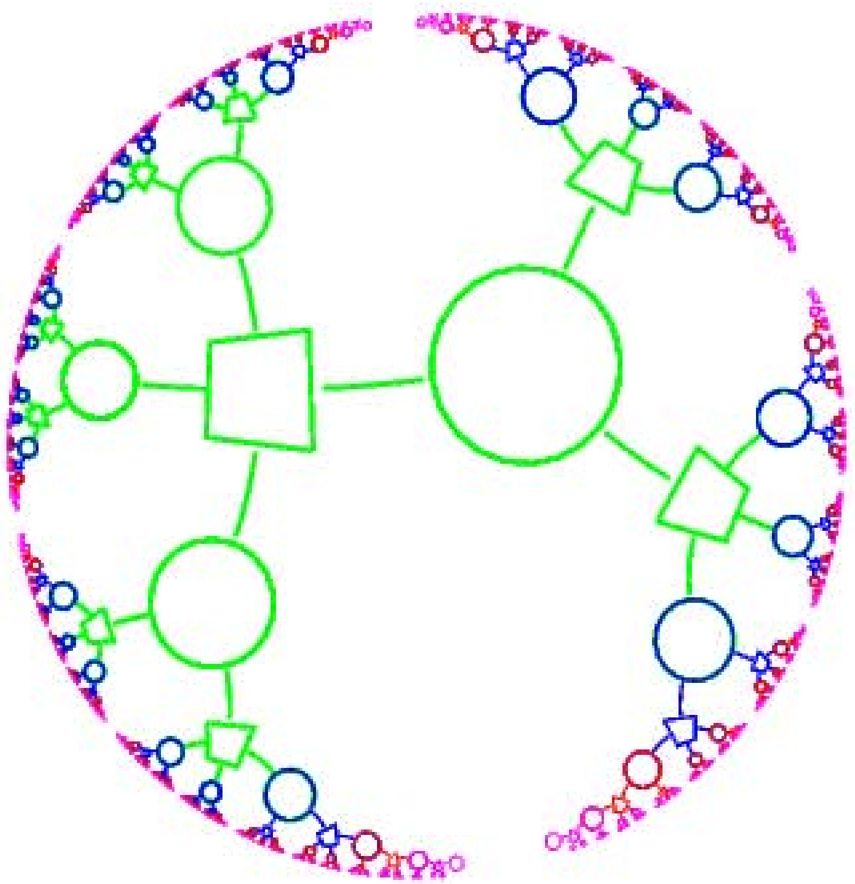
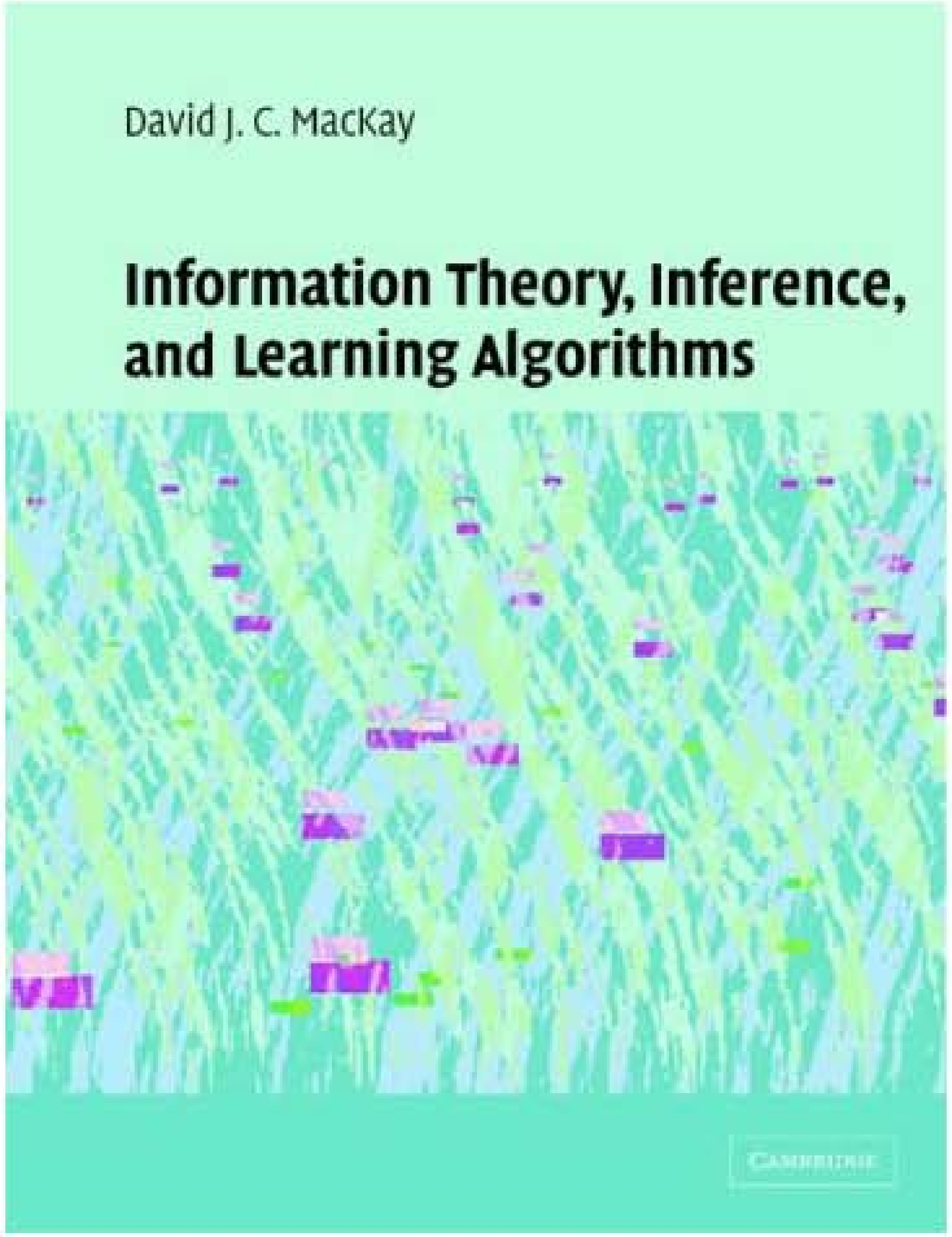


Information theory

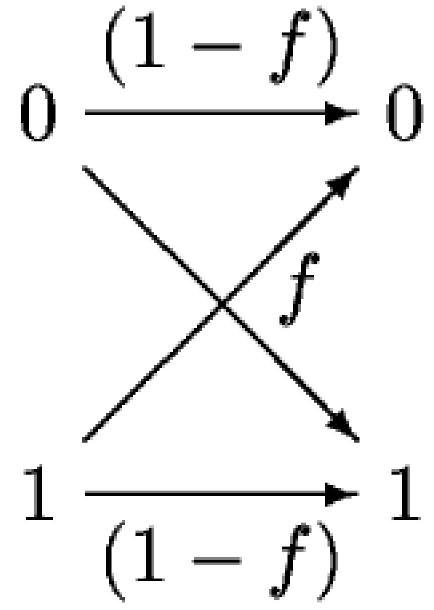
Lecture notes - Chapter 1



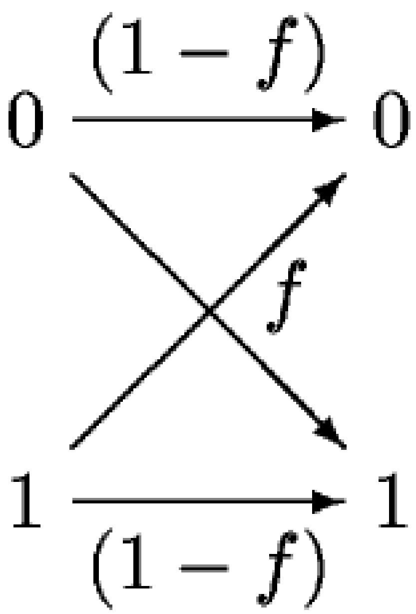
- Cambridge University Press
- 640 pages, 35 pounds
- Also available free online

www.inference.phy.cam.ac.uk/mackay/itila/

This lecture uses the blackboard



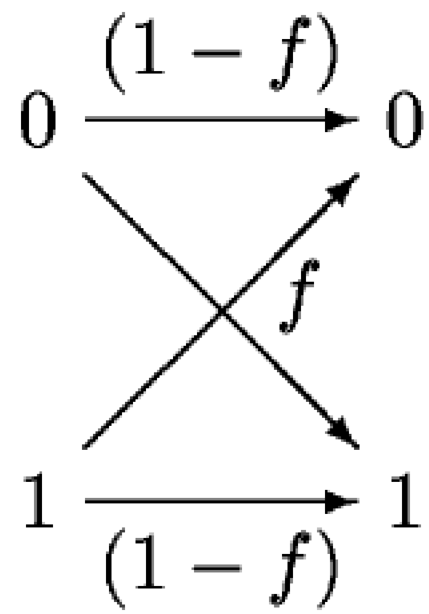
f = 0.1



Q: A file of $N = 10\,000$ bits is stored on this disc drive (with $f = 0.1$), then read.

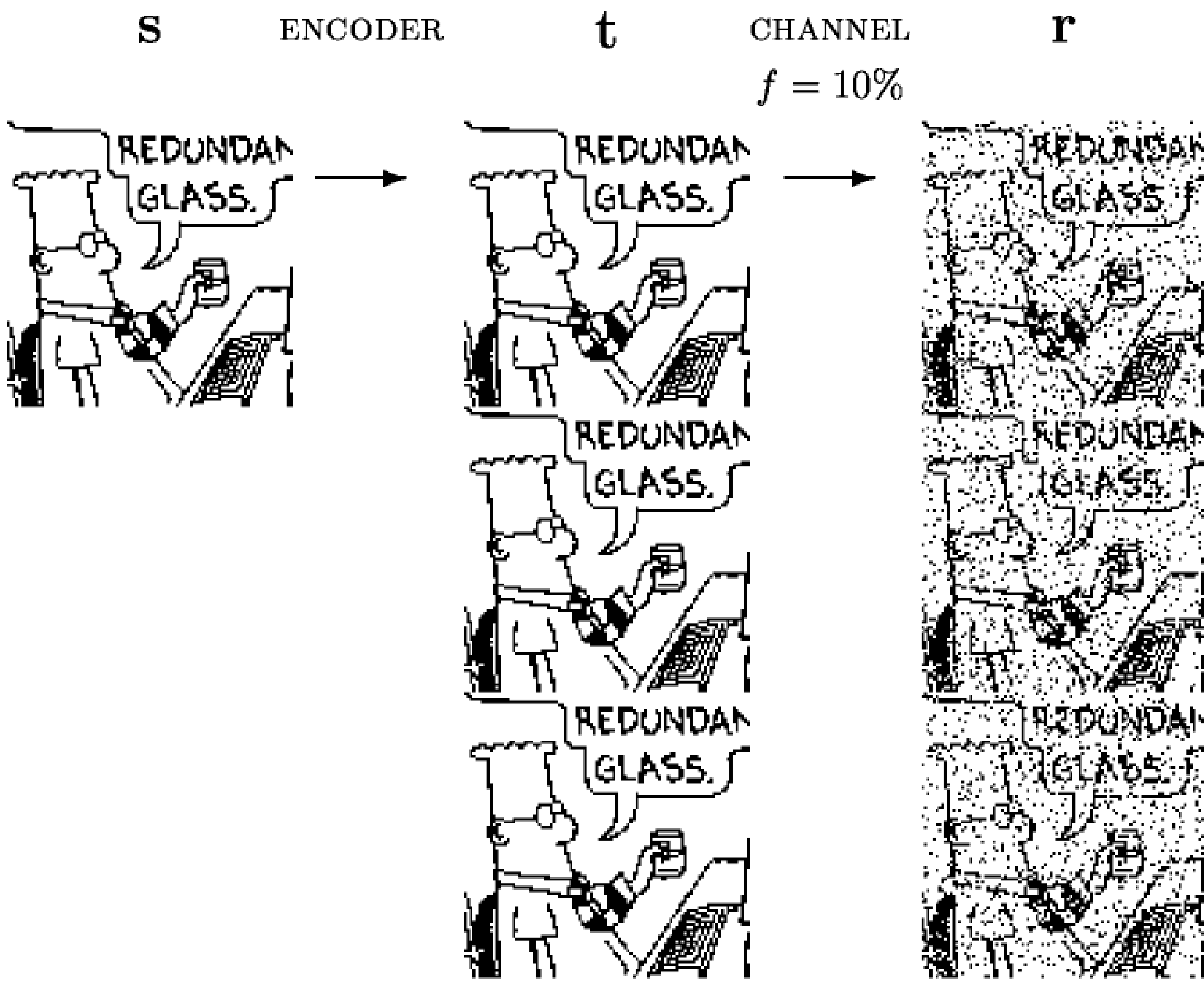
Roughly how many bits are flipped?

$\square \pm \square$

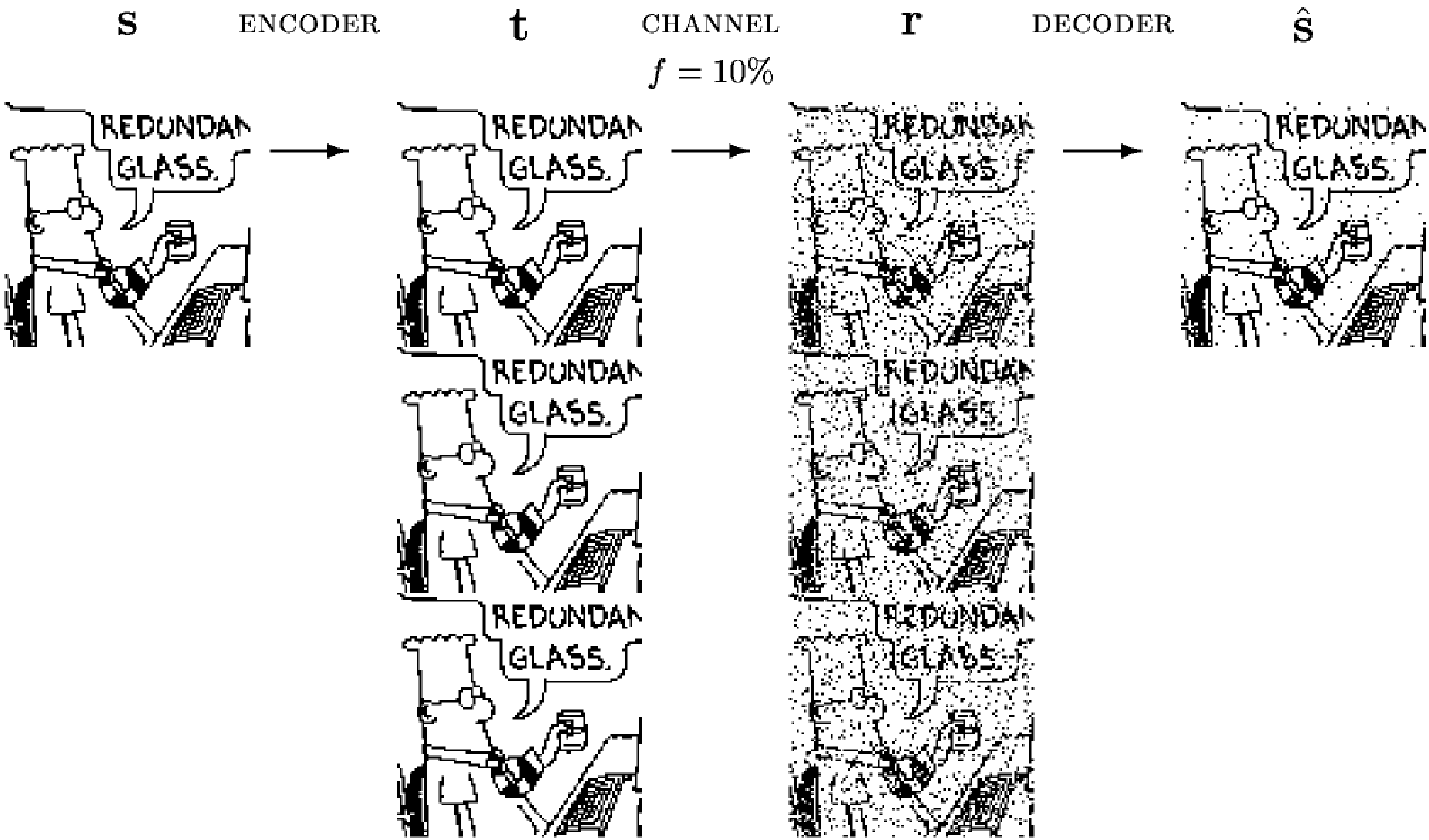


Q: To make a successful business selling 1 Gigabyte disc drives, how small does the flip probability f need to be?

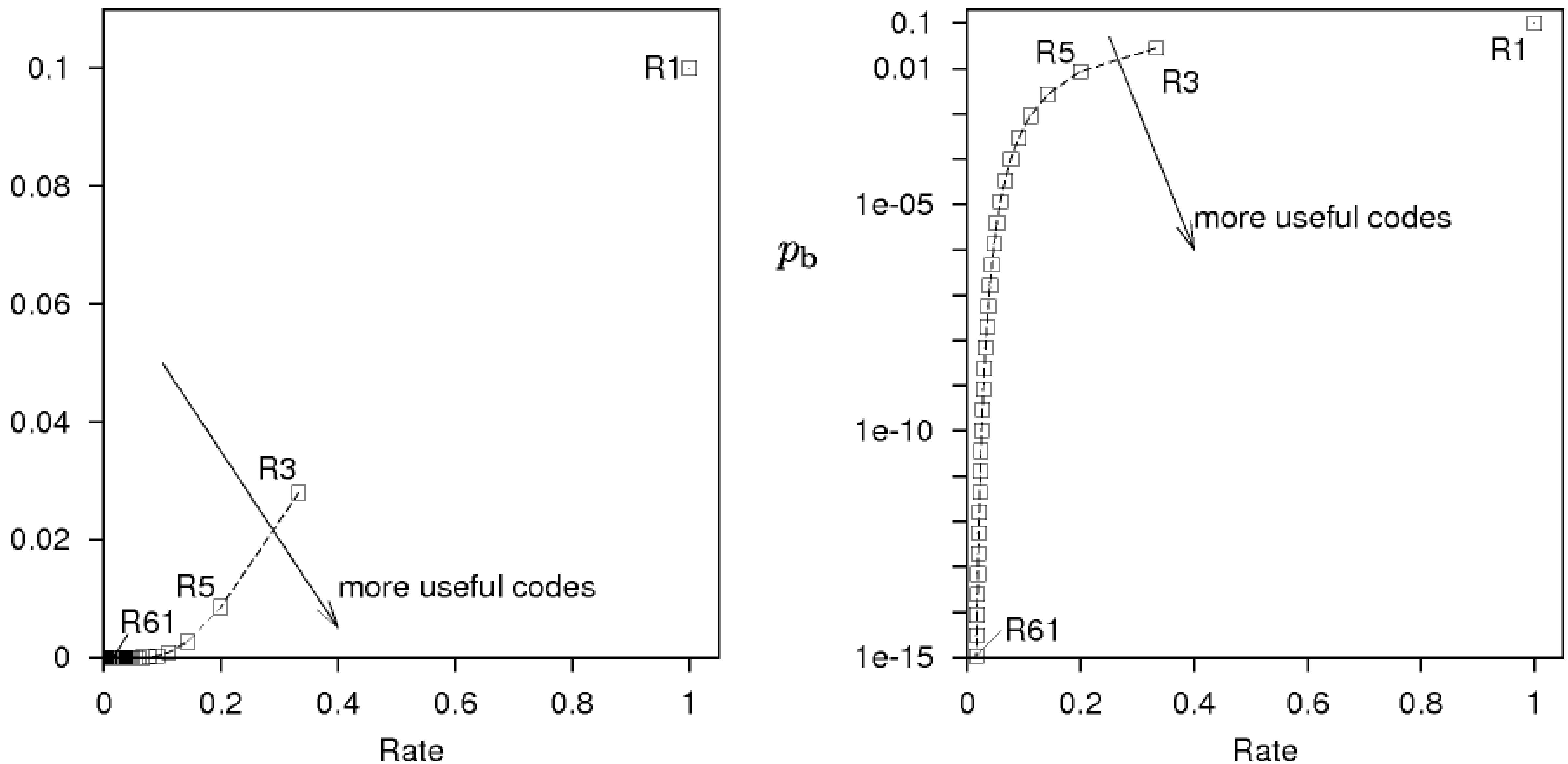
Repetition code 'R3'

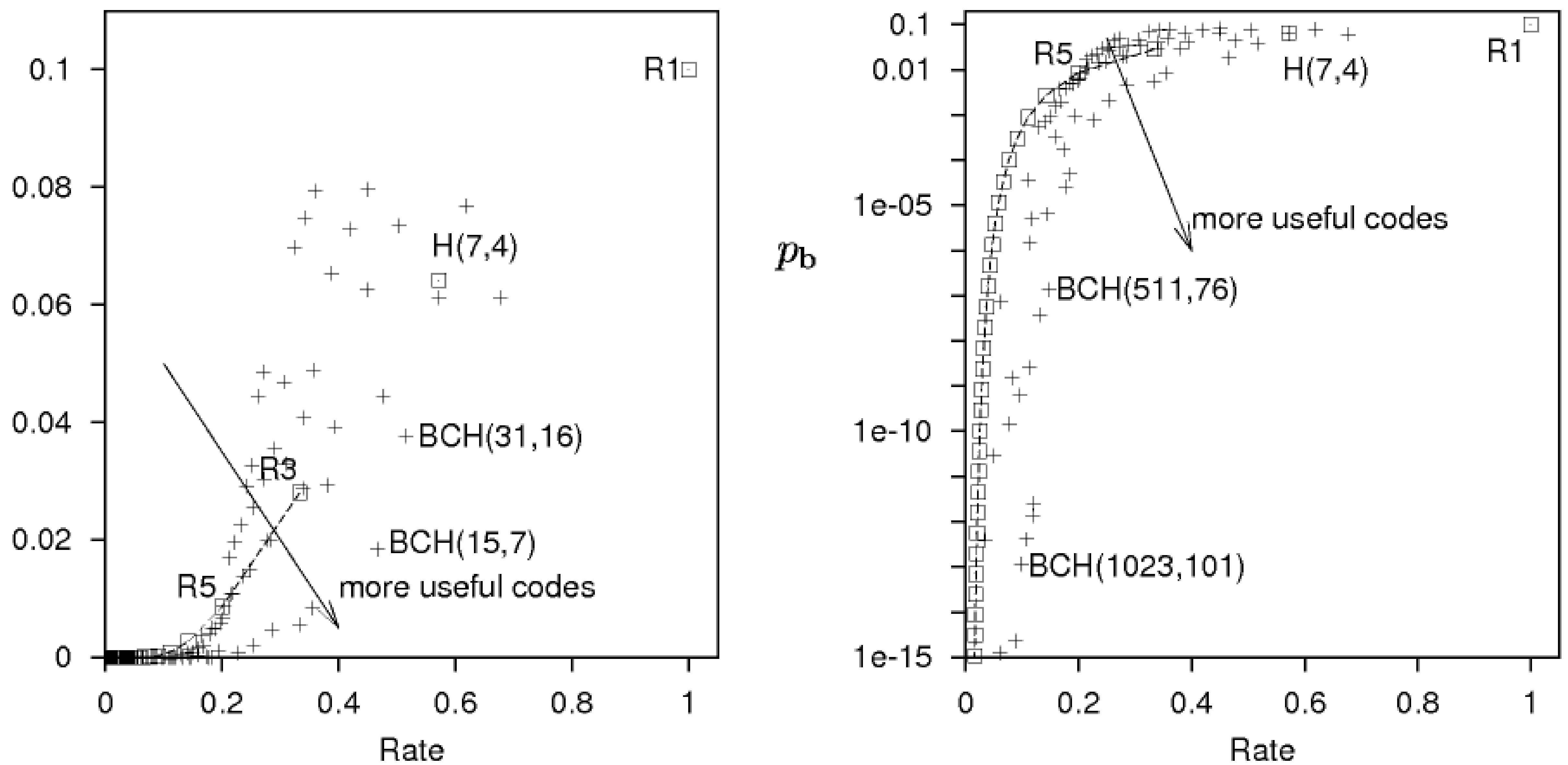


Repetition code 'R3'



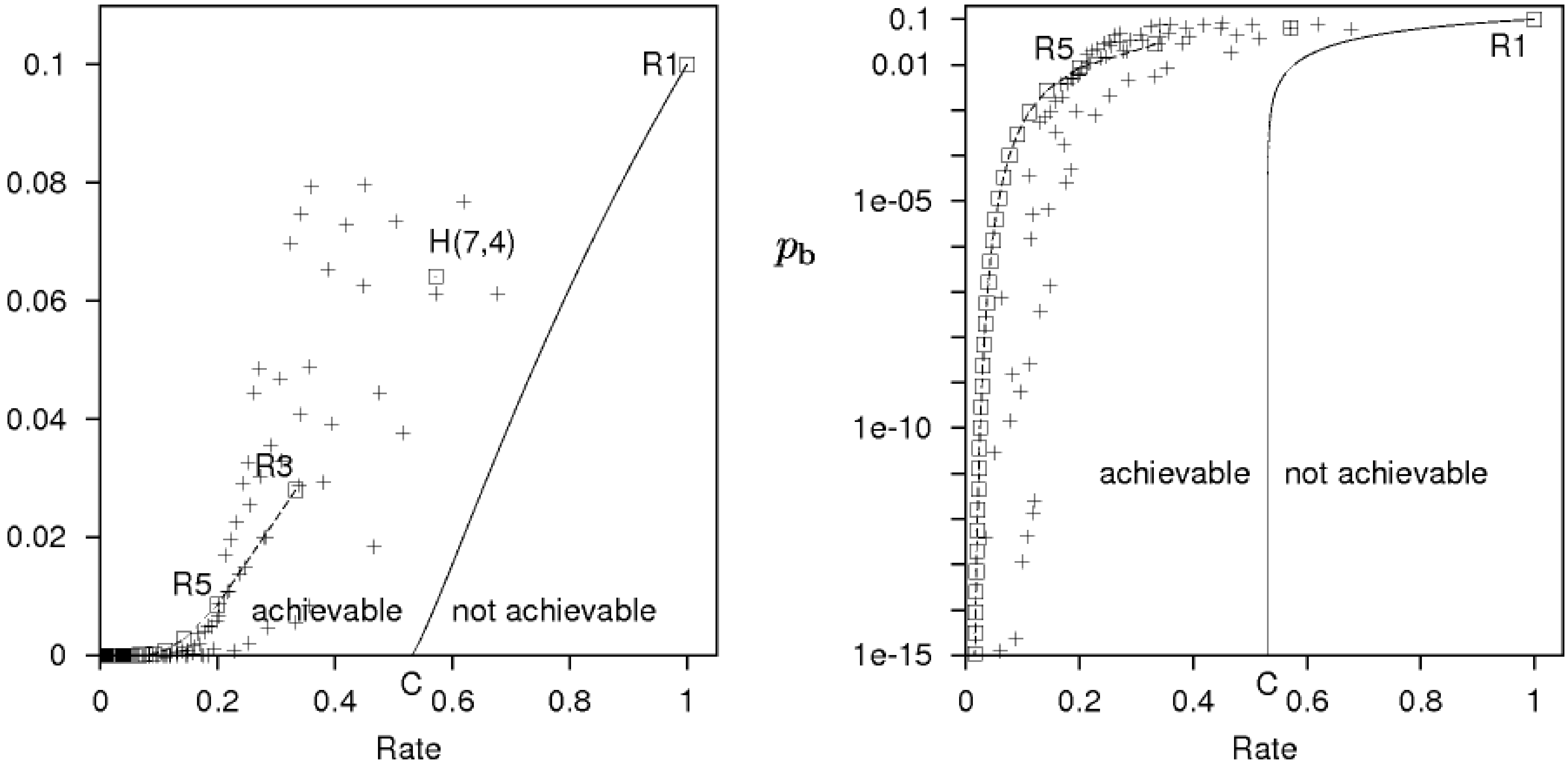
Performance of repetition codes





What's achievable?

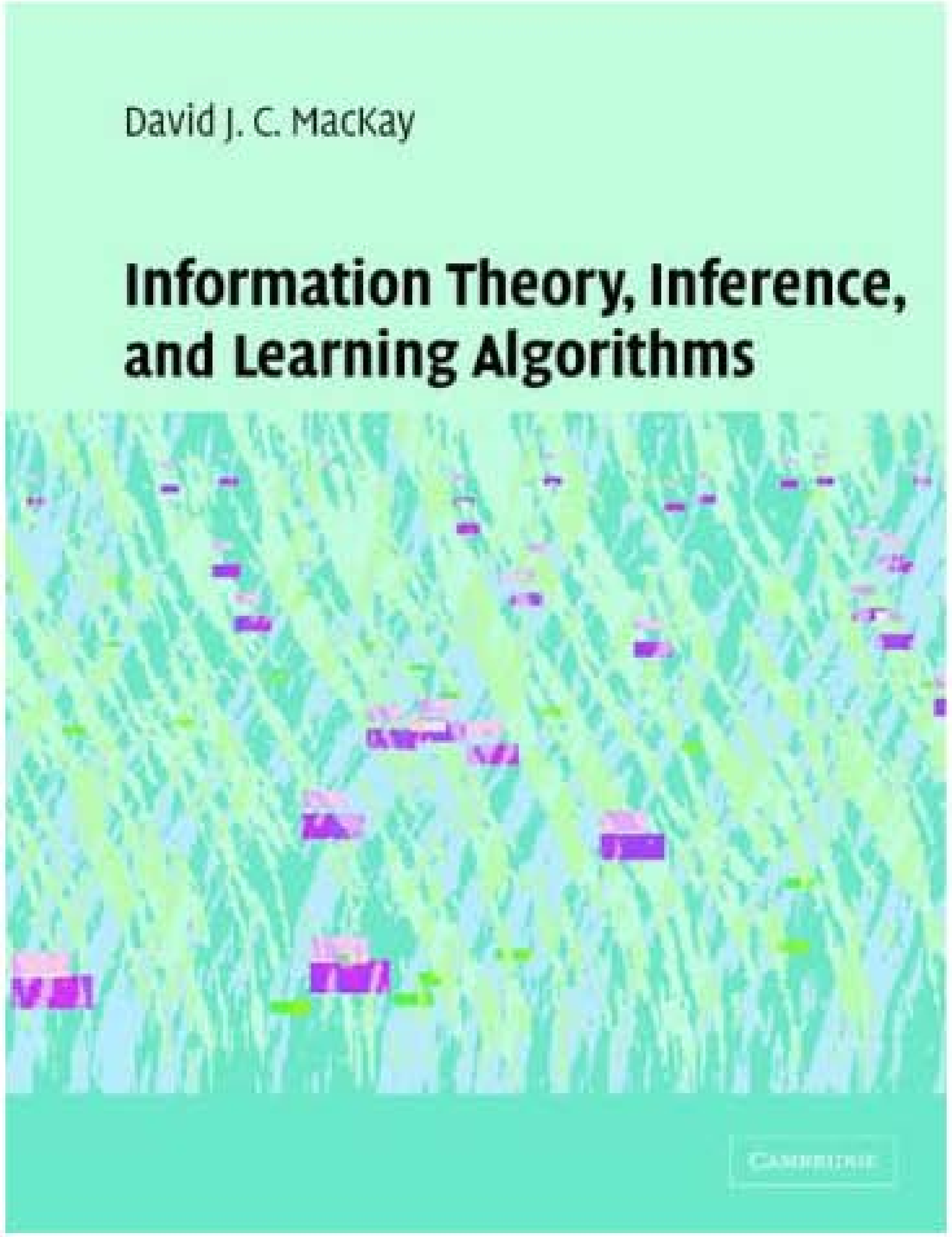
Shannon's noisy-channel coding theorem



$$C_{\mathrm{BSC}}(f) = 1 - H_2(f)$$

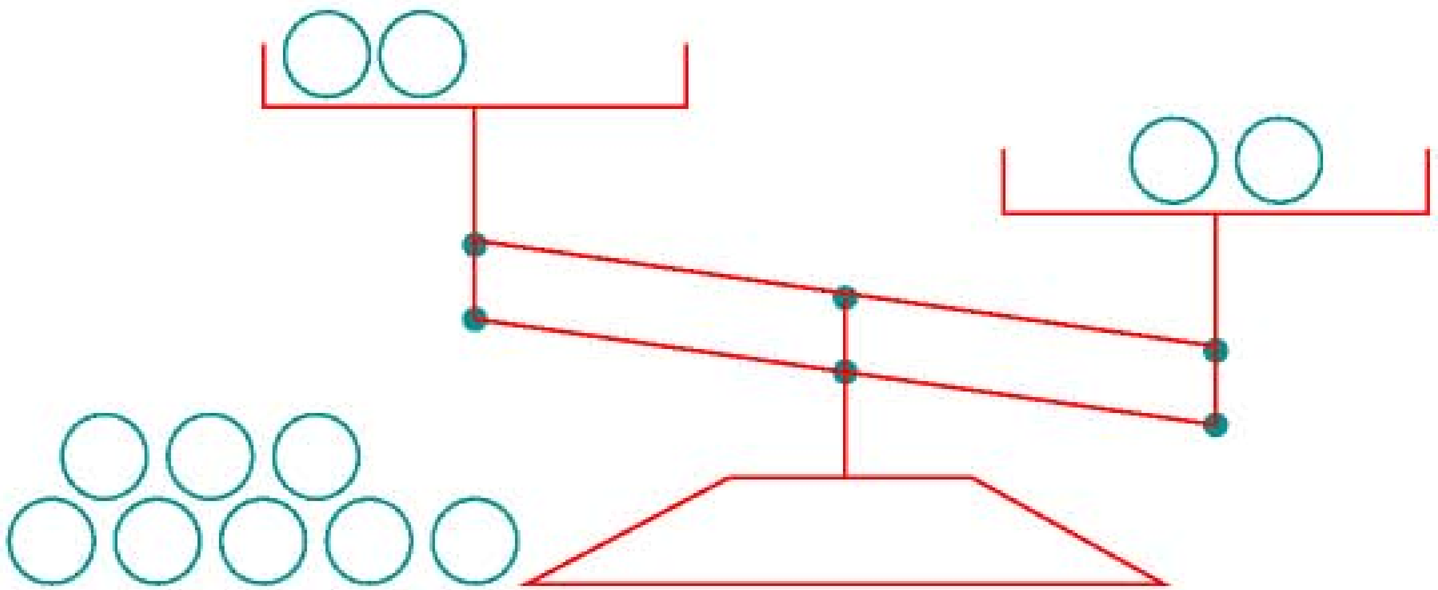
$$H_2(f) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f}$$

Outline of next lecture



- Information Theory
  - Source coding (Data compression)
  - Noisy-channel coding
    - the theorem
    - state-of-the art error-correcting codes

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

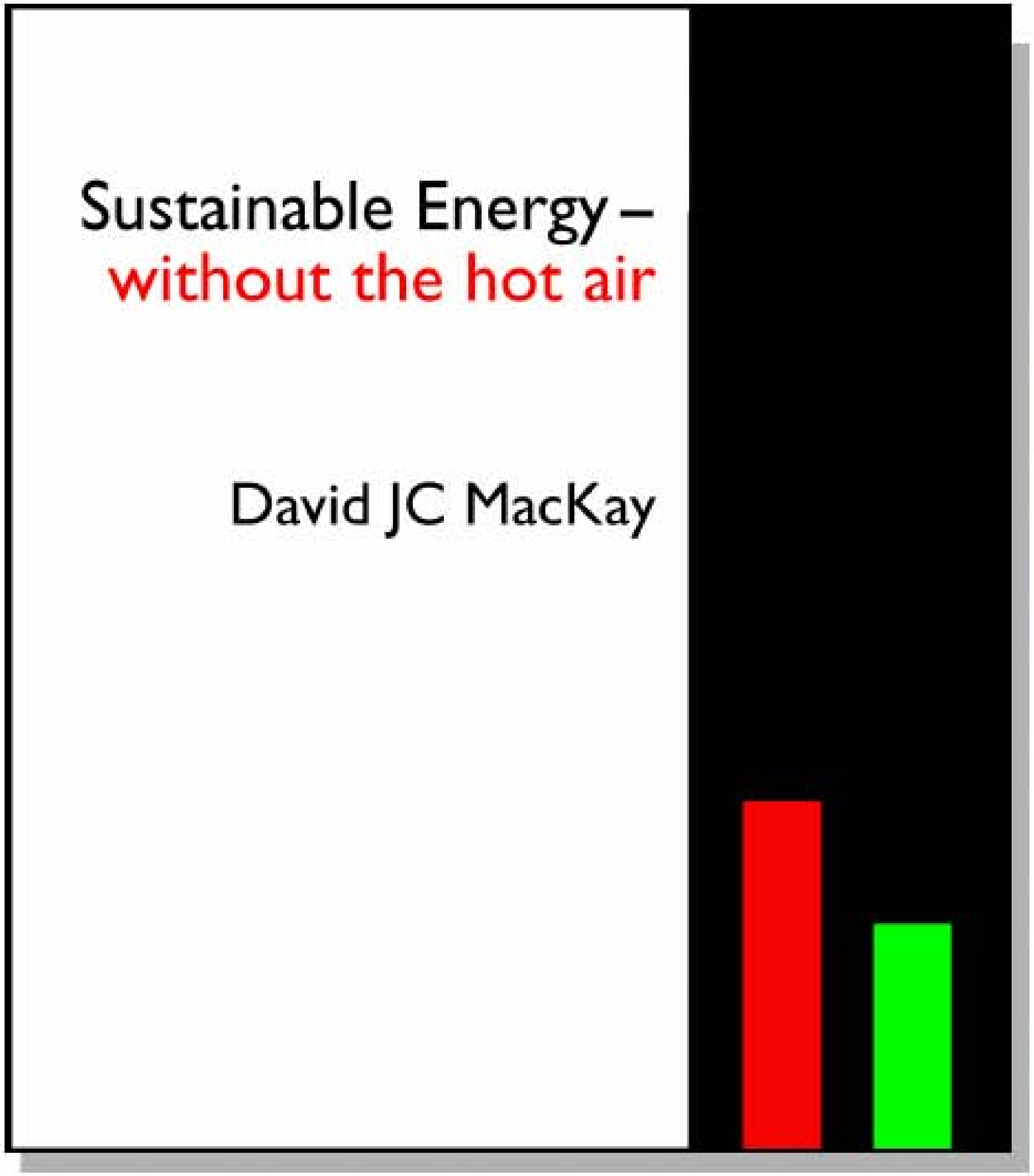
Design a strategy to determine

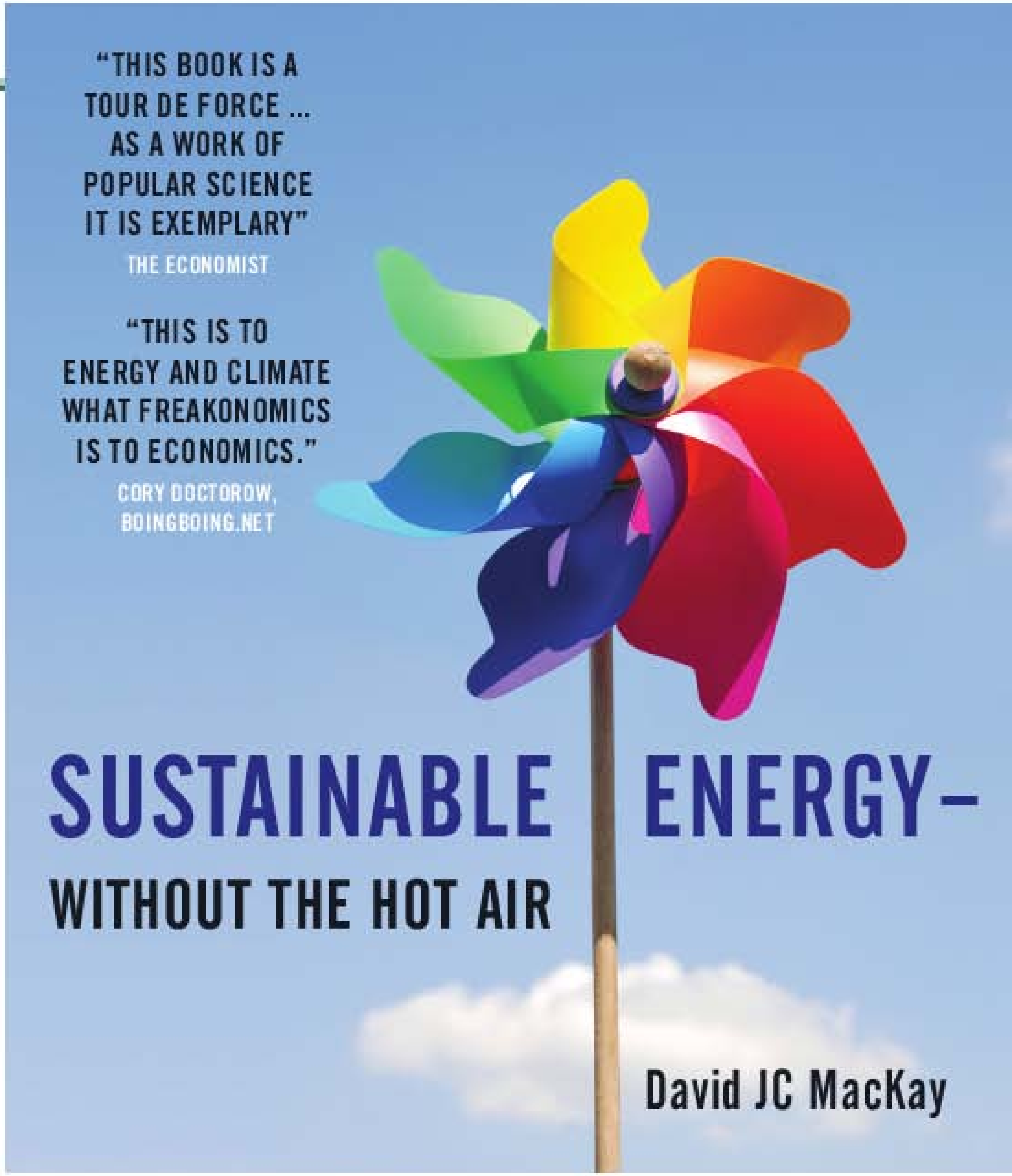
which is the odd ball

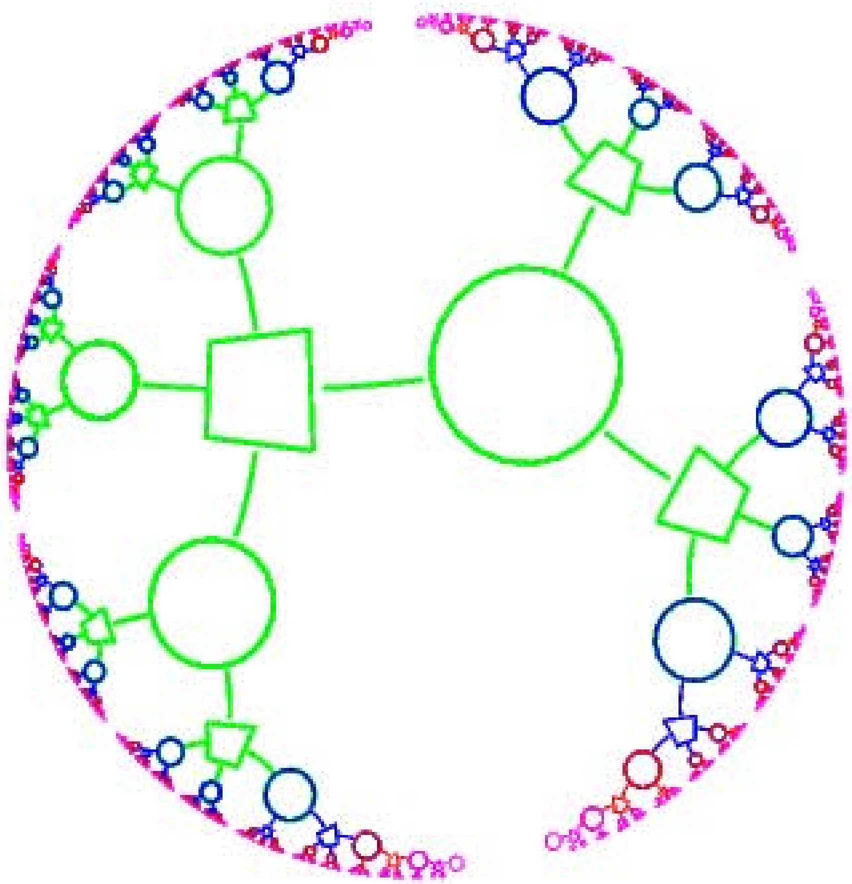
and whether it is heavier or lighter,

in as few uses of the balance as possible.

Another book

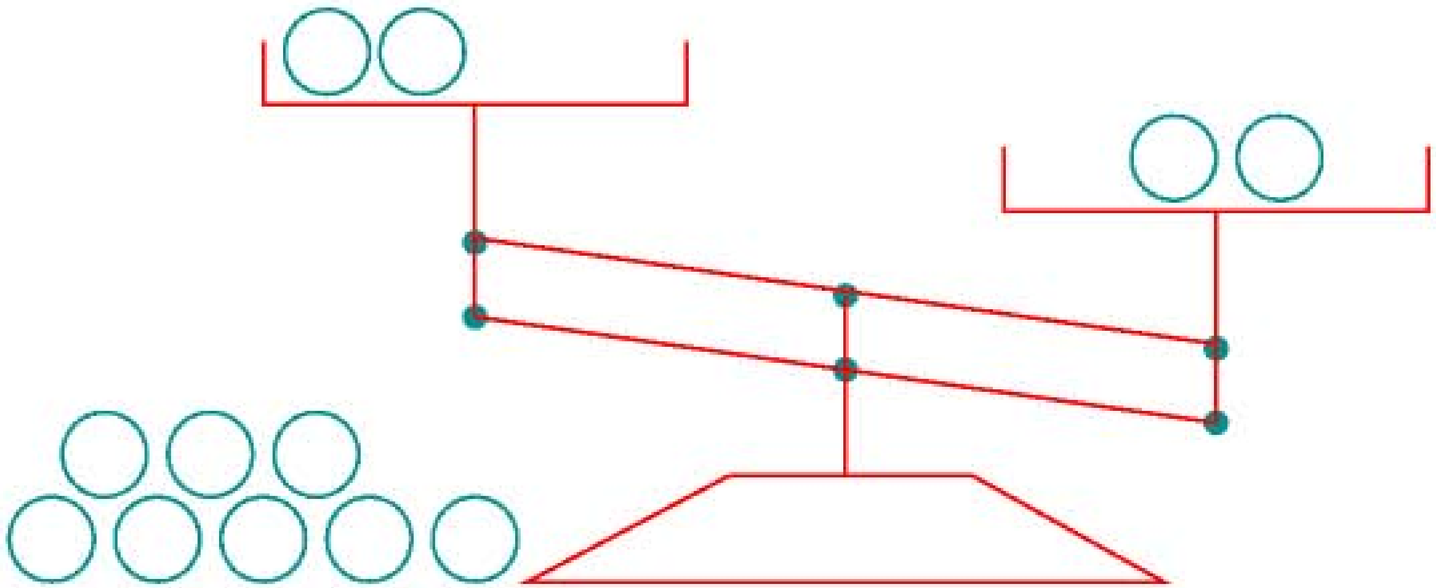






Information theory II

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

Design a strategy to determine

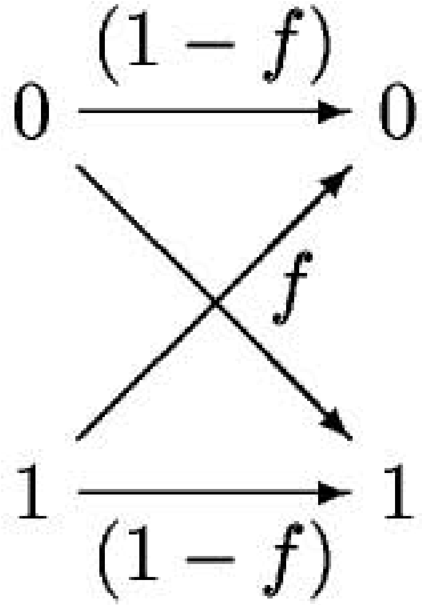
which is the odd ball

and whether it is heavier or lighter,

in as few uses of the balance as possible.

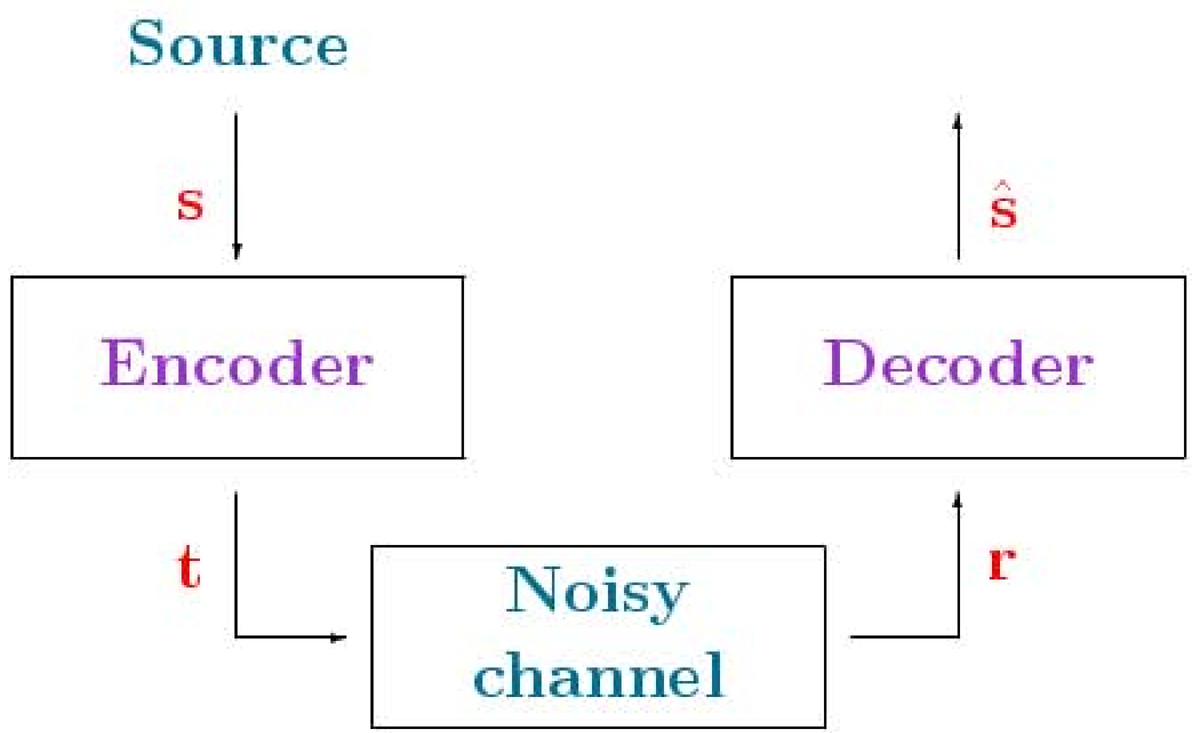
Purpose: reliable communication over unreliable channels

eg, Binary symmetric channel



f = 0.1

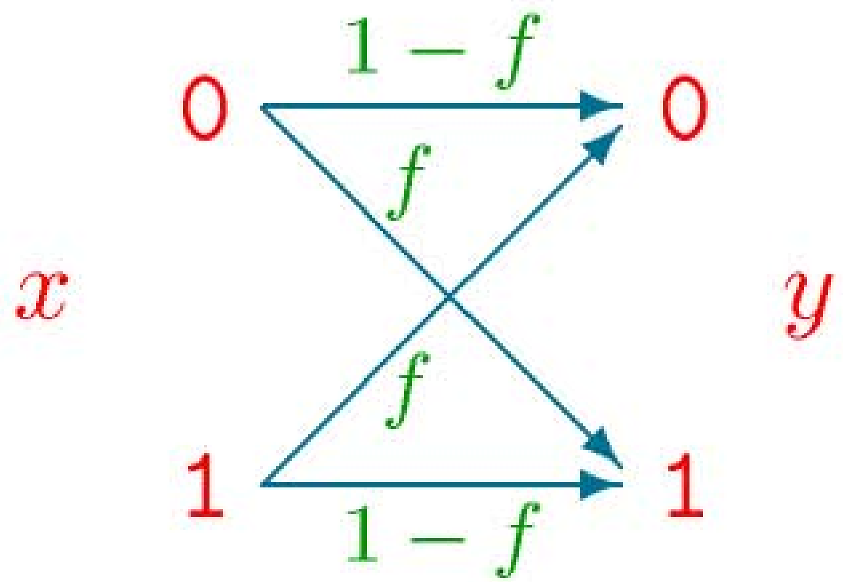
System solution



adds redundancy

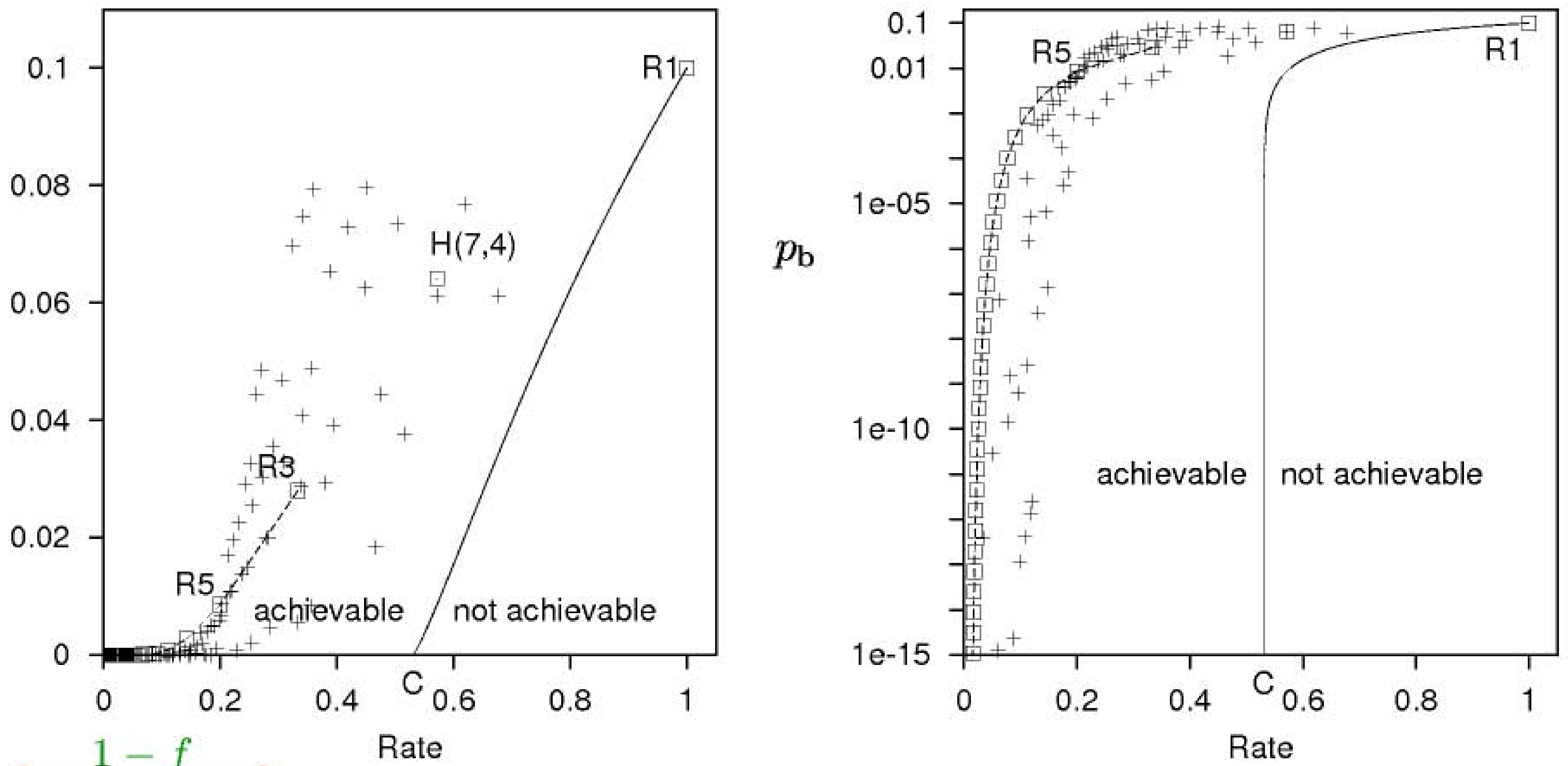
does inference

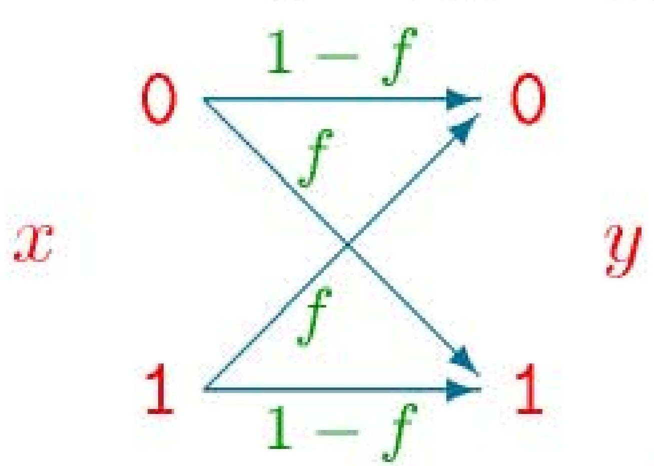
The binary symmetric channel



- Repetition codes
- (7,4) Hamming code

Shannon's noisy-channel coding theorem

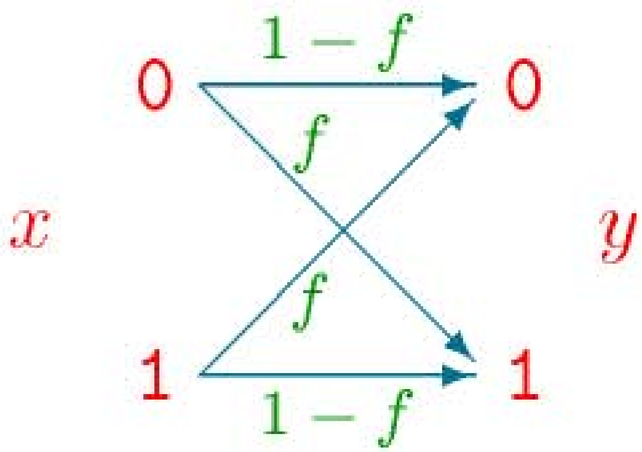


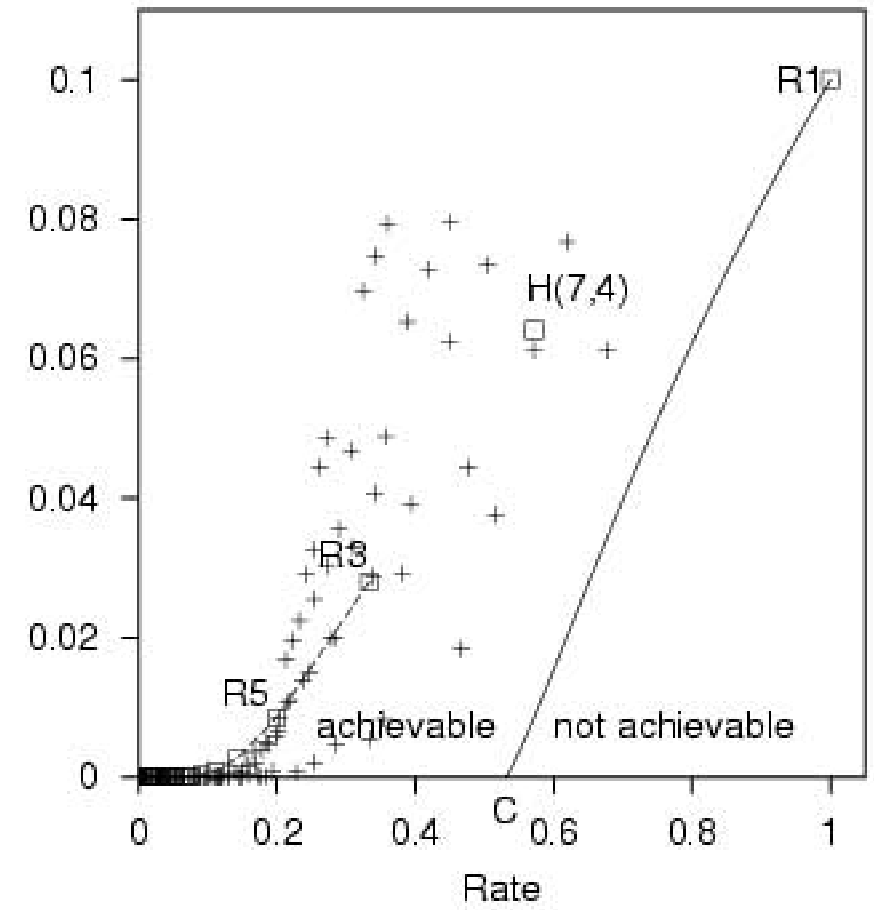


$$C_{\mathrm{BSC}}(f) = 1 - H_2(f)$$

$$H_2(f) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f}$$

Shannon's noisy channel coding theorem



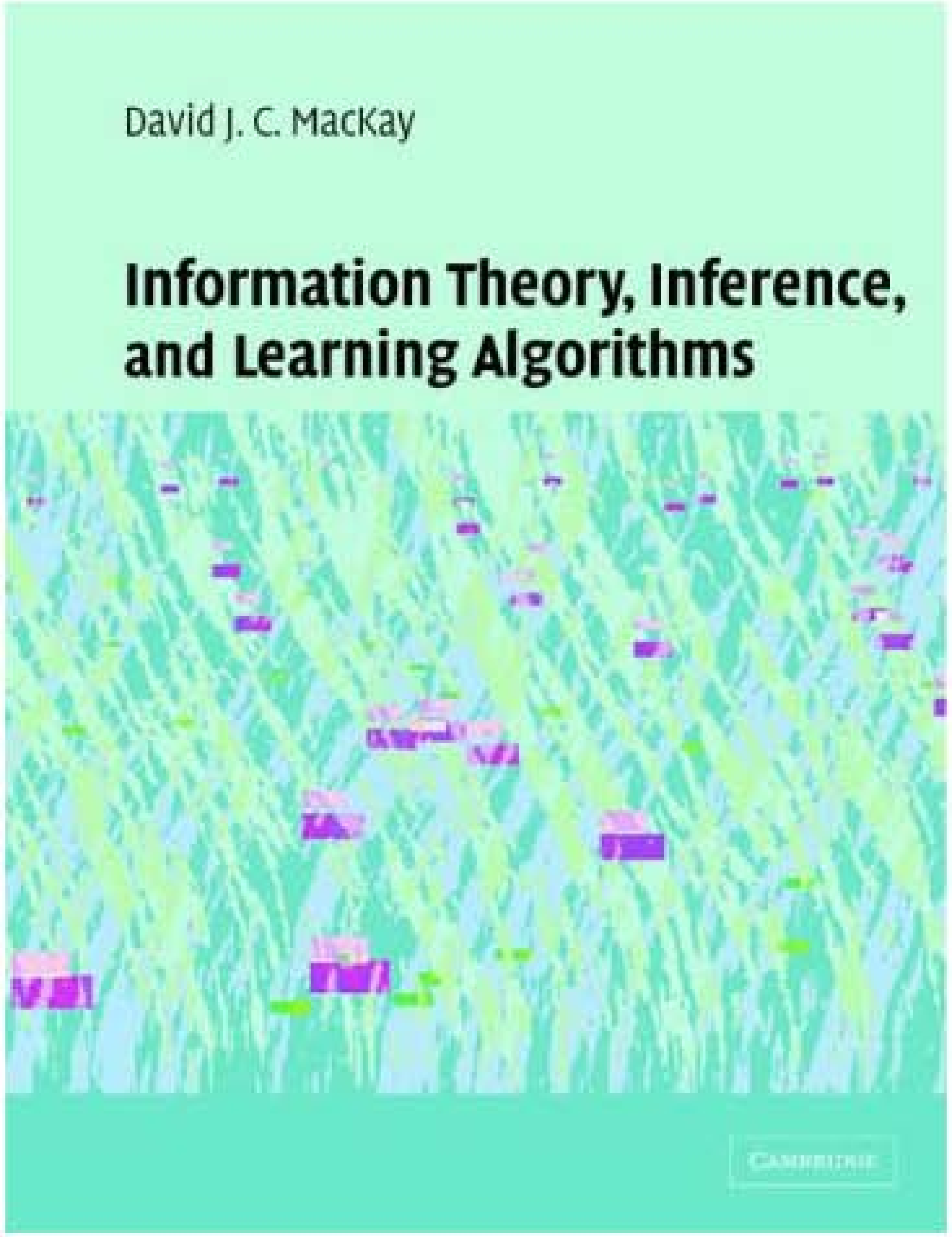


For any channel:

Reliable (virtually error-free) communication is possible

at rates up to C

Outline



- Source coding (Data compression)
  - key ideas
  - optimal symbol codes
  - arithmetic coding
- Noisy-channel coding
  - the theorem
  - state-of-the art error-correcting codes
- Lecture notes - Chapters 4, 5, 6, 47
  - Cambridge University Press
  - 640 pages, 35 pounds
  - Also available free online

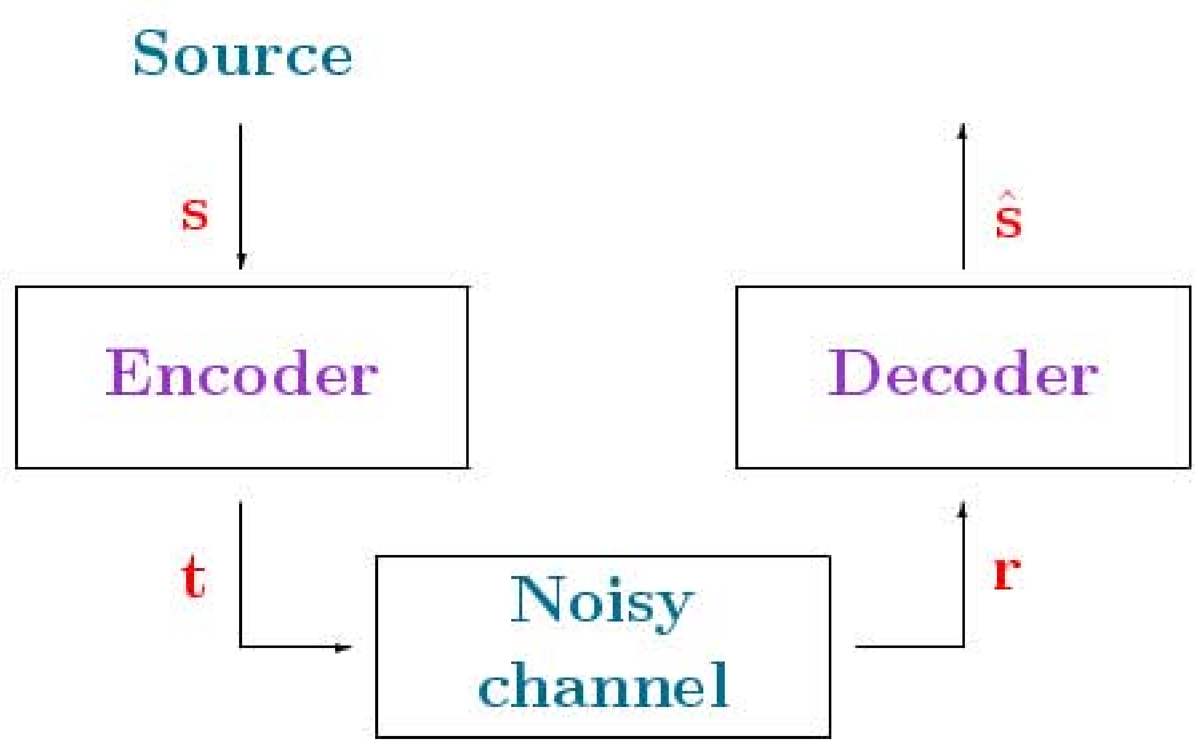
www.inference.phy.cam.ac.uk/mackay/itila/

Emma Woodh*use, hands*me, clever* and rich,*with a
comfortab*e home an* happy di*position,*seemed to*unite som*
of the b*st bless*ngs of e*istence;*and had *ived nea*ly
twenty *ne year* in the*world w*th very*little *o distr*ss
or vex*her. *he was*the yo*ngest *f the *wo dau*hters *f a
most *ffect*onate* indu*gent *ather* and *ad, i* cons*quenc*
of h*r si*ter'* mar*iage* bee* mis*ress*of h*s ho*se f*om a
ver* ea*ly *eri*d. *er *oth*r h*d d*ed *oo *ong*ago*for*her
to*ha*e *or* t*an*an*in*is*in*t *em*mb*an*e *f *er*ca*es*es*
a*d*h*r*p*a*e*h*d*b*e* *u*p*i*d*b* *n*e*c*l*e*t*w*m*n*a*
g**e**e**,**h**h** **l**n**i**l**s**r**o**a**o**e**i*
a***c***n***S***e***y***s***d***s***a***r***e***n***
W****o****s****i****l****a****g****n****t****a****e****v***

Emma Woodhouse, handsome, clever, and rich, with a
comfortable home and happy disposition, seemed to unite some
of the best blessings of existence; and had lived nearly
twenty one years in the world with very little to distress
or vex her. She was the youngest of the two daughters of a
most affectionate, indulgent father; and had, in consequence
of her sister's marriage, been mistress of his house from a
very early period. Her mother had died too long ago for her
to have more than an indistinct remembrance of her caresses;
and her place had been supplied by an excellent woman as
governess, who had fallen little short of a mother in
affection. Sixteen years had Miss Taylor been in Mr
Woodhouse's family, less as a governess than a friend, very

Emma Woodh*use, hands*me, clever* and rich,*with a
comfortab*e home an* happy di*position,*seemed to*unite som*
of the b*st bless*ngs of e*istence;*and had *ived nea*ly
twenty *ne year* in the*world w*th very*little *o distr*ss
or vex*her. *he was*the yo*ngest *f the *wo dau*hters *f a
most *ffect*onate* indu*gent *ather* and *ad, i* cons*quenc*
of h*r si*ter'* mar*iage* bee* mis*ress*of h*s ho*se f*om a
ver* ea*ly *eri*d. *er *oth*r h*d d*ed *oo *ong*ago*for*her
to*ha*e *or* t*an*an*in*is*in*t *em*mb*an*e *f *er*ca*es*es*
a*d*h*r*p*a*e*h*d*b*e* *u*p*i*d*b* *n*e*c*l*e*t*w*m*n*a*
g**e**e**,**h**h** **l**n**i**l**s**r**o**a**o**e**i*
a***c***n***S***e***y***s***d***s***a***r***e***n***
W****o****s****i****l****a****g****n****t****a****e****v***

Channel coding



A simple redundant source - a bent coin

```
000000000000000001001001000000000000000000100001000101000000
000101000000000000000000000100000000010000000000000100000110
000010000000000000001000000000001000010000010000000000000000
000010000001000000000001000000000000000000000000010000000000
000000010001000000100000010010000000000001000000001000000000
000100100001010100000011000010000000000001000000000000000000
000000000000000000000000010100000010000000000000100010100001
100000000010000000000000000000000000000000000000000000000000
010000000000000100000000000000000100001000000000110110000000
101000101000000000000000000000000000100010100000100001000000
000000000000100000100000100001001000000001010001000000000000
000000000000000000000000011000000000101000000000100100010101
000110010101000000000000100000000000000000100000001010000110
100000001000000010110000000010000000010010000010000000000000
100000000000000100000100000000000000100000000000000000000010
000010001000000100000000000000100000000001000001000000000000
110100000100100000000000000000001000000000000000001000000000
100000000000000000000001000000010010001000000001100000011100
000000000001000000000000000000000001000000000000001000000000
000000000010000000000010100000100000000000010000000000001000
101000100100010000000000000000000000000000000000001010000000
010010000000000000000000000001000001100000010000000001000000
000000001000100000000000010000000000000000000000000000001000
000000000001000000010000000000000000000000000000000000000000
000001000000000100100000000100000000100001000000100000000000
```

$p_1 = 0.1$

How to compress a redundant file?

```
000000000000000001001001000000000000000000100001000101000000
000101000000000000000000000100000000010000000000000100000110
000010000000000000001000000000010000100000100000000000000000
000010000000100000000001000000000000000000000000010000000000
000000010001000000100000010010000000000001000000010000000000
000100100001010100000001100001000000000001000000000000000000
000000000000000000000000101000000010000000000000100010100001
100000000010000000000000000000000000000000000000000000000000
010000000000000100000000000000000100001000000000110110000000
101000101000000000000000000000000000000000100010100000100001000000
```

e.g., $N = 1000$ tosses of a bent coin with $p_1 = 0.1$

- How to measure information content?
  - How much compression should we expect is possible?

How to measure information content?

Claims: 1. The *Shannon information content* of an outcome

$$h(x=a_i) \;=\; \log_2 \frac{1}{P(x=a_i)}$$

is a sensible measure of information content.

2. The entropy

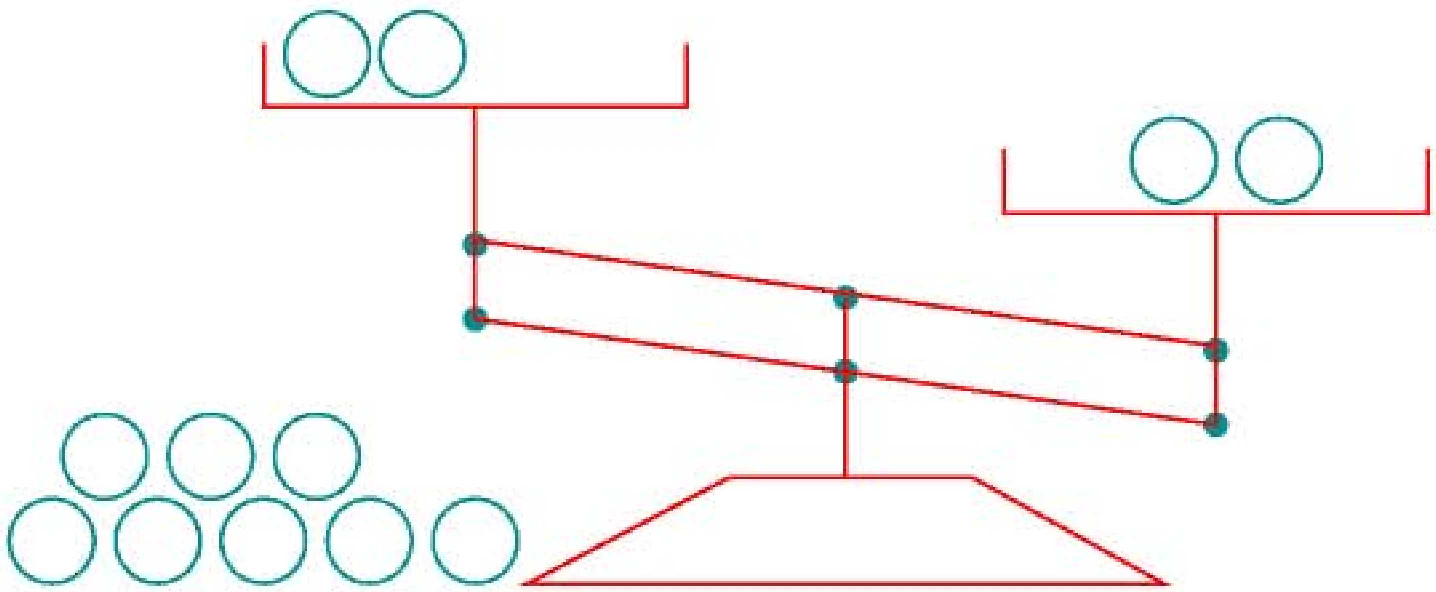
$$H(X) \;=\; \sum_x P(x) \log_2 \frac{1}{P(x)}$$

is a sensible measure of expected information content.

(sketch h(p))

Point out additivity of h.

The weighing problem



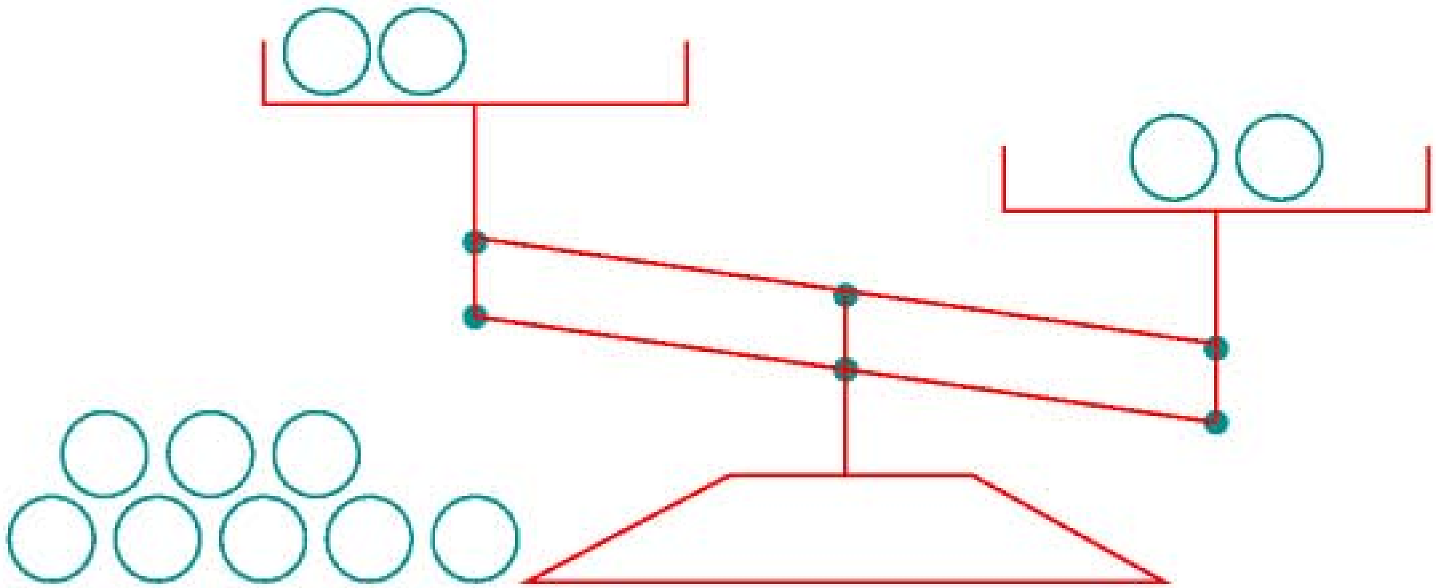
You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

Design a strategy to determine

which is the odd ball

and whether it is heavier or lighter,

in as few uses of the balance as possible.



My strategy will find the odd ball and whether it is heavier or lighter in at most 13 uses of the balance.

12

11

10

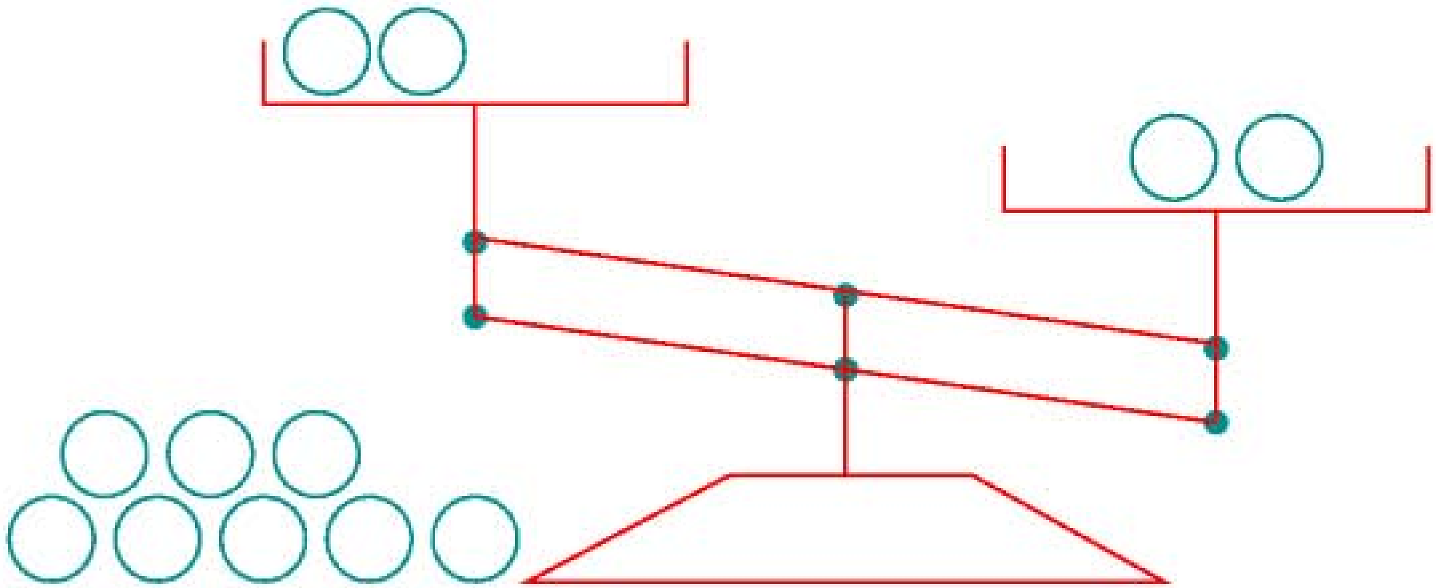
9

8

7

6

5



My strategy's first weighing is

6 v 6

5 v 5

4 v 4

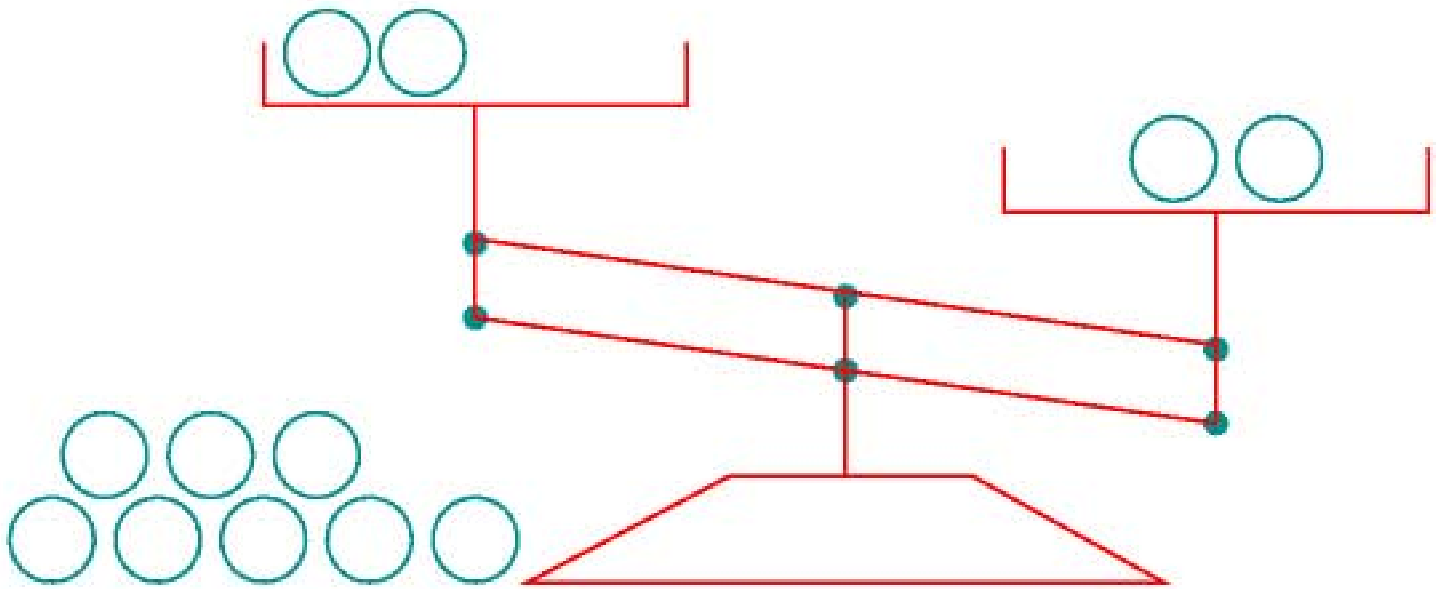
3 v 3

2 v 2

1 v 1

Testing 'Shannon information content' claims

- Weighing problem



  - Shannon says

Most 'informative' experiment is the one with maximum entropy

This lecture uses the blackboard

How to measure information content?

Claims:

1. The *Shannon information content* of an outcome

$$h(x=a_i) = \log_2 \frac{1}{P(x=a_i)}$$

is a sensible measure of information content.

2. The entropy

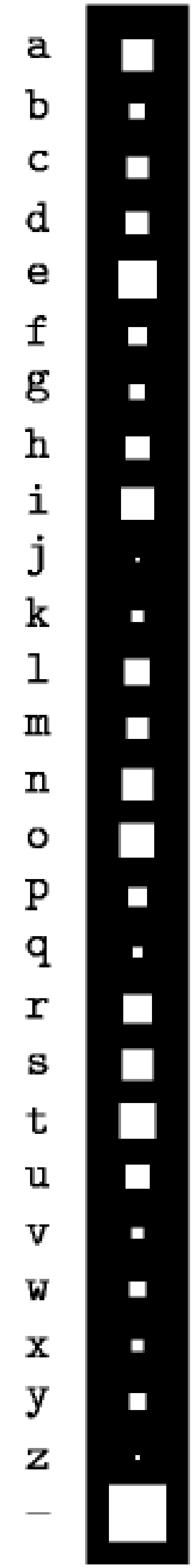
$$H(X) = \sum_x P(x) \log_2 \frac{1}{P(x)}$$

is a sensible measure of expected information content.

3. Source coding theorem –
N outcomes from a source X can be compressed into roughly $NH(X)$ bits.

Symbol codes

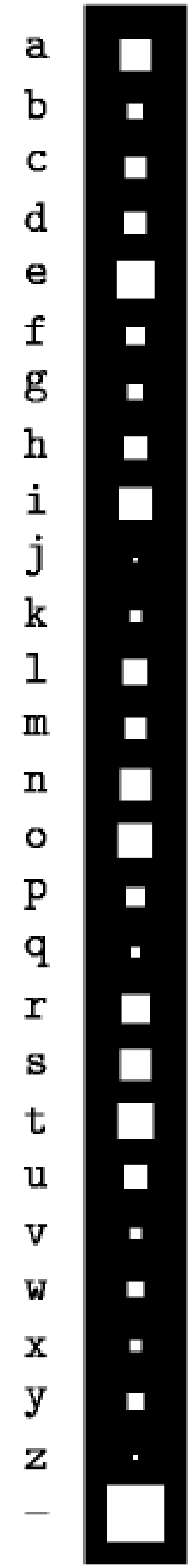
i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928



The symbol-code supermarket

0	00	000	0000
			0001
		001	0010
			0011
	01	010	0100
			0101
		011	0110
			0111
1	10	100	1000
			1001
		101	1010
			1011
	11	110	1100
			1101
		111	1110
			1111

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928



a	0000
b	001000
c	00101
d	10000
e	1100
f	111000
g	001001
h	10001
i	1001
j	1101000000
k	1010000
l	11101
m	110101
n	0001
o	1011
p	111001
q	110100001
r	11011
s	0011
t	1111
u	10101
v	11010001
w	1101001
x	1010001
y	101001
z	1101000001
_	01

Explain the Huffman algorithm on the blackboard

Symbol codes

The ideal codelengths l_i^* are the information contents

$$l_i^* = \log \frac{1}{p_i}$$

The optimal symbol code's expected length L satisfies

$$H(X) \leq L < H(X) + 1$$

The optimal symbol code is generated by the Huffman algorithm

- Does that wrap up compression?
  - What's wrong with optimal symbol codes?
- Arithmetic coding

The Guessing Game

Headline composed of

$\{$ A, B, C, D, E, F, G, H, I, J, K, L, M, . . . , Z, – $\}$

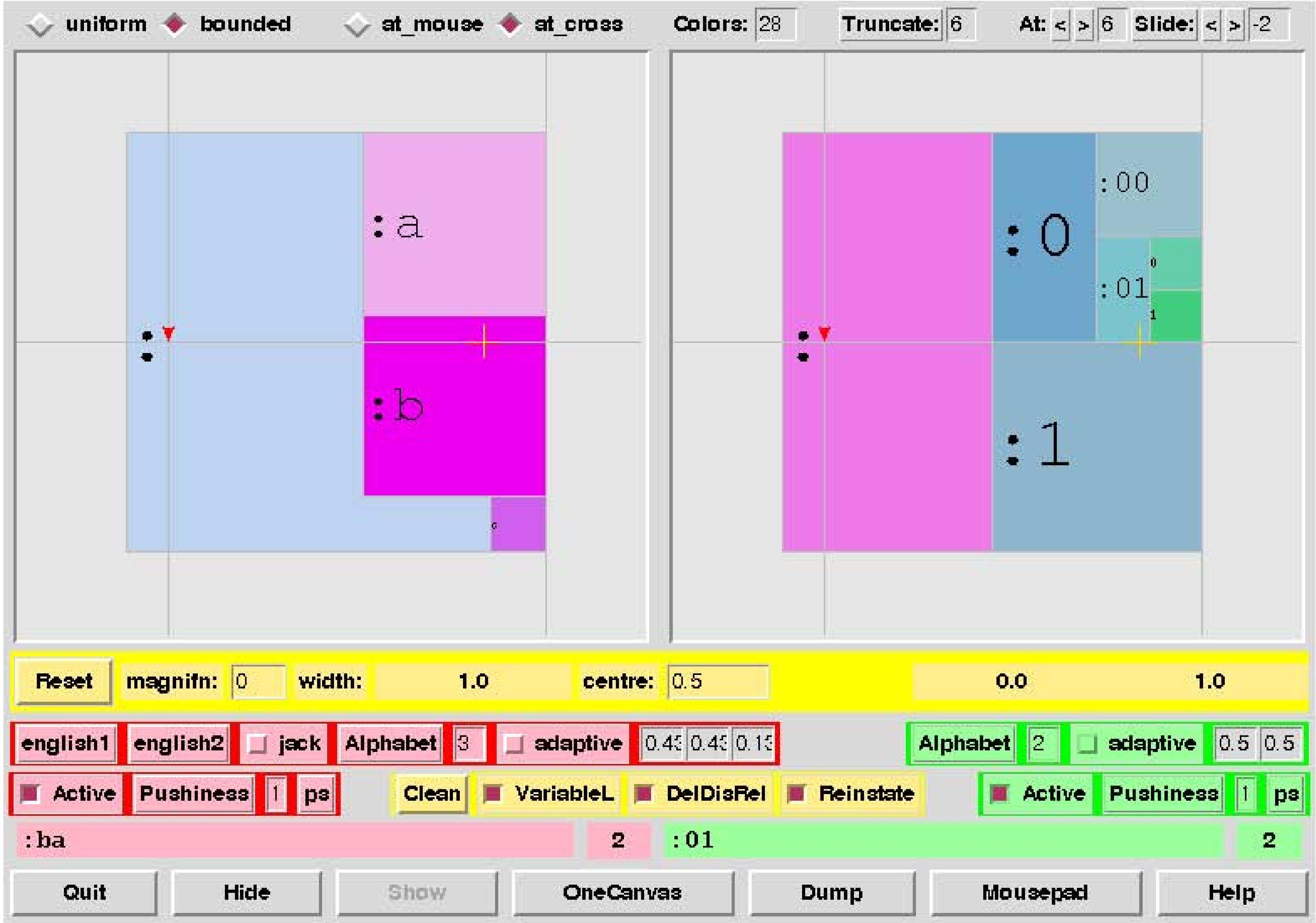
Realistic compression

Optimal symbol code for

a	0.001	00000
b	0.001	00001
c	0.990	1
d	0.001	00010
e	0.001	00011
f	0.001	0100
g	0.001	0101
h	0.001	0110
i	0.001	0111
j	0.001	0010
k	0.001	0011

expected length	1.034
entropy	0.11401
length / entropy	9

Arithmetic coding



Other uses for arithmetic coding

- Efficient writing

Compression:

$$\text{Text} \longrightarrow \text{Bit string}$$

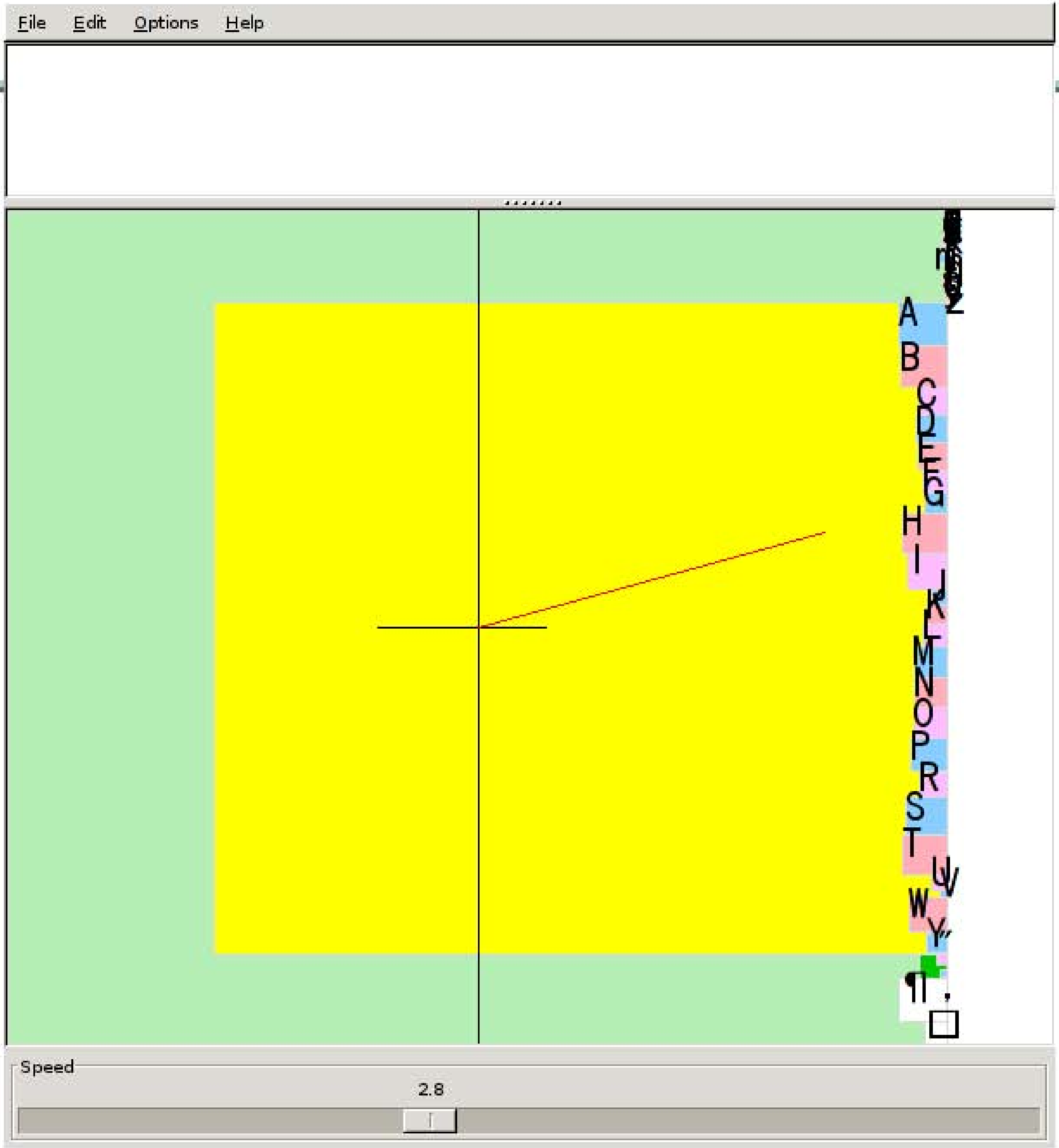
(preferably short)

Writing:

$$\text{Text} \longleftarrow \text{Gesture}$$

(preferably brief)

Dasher

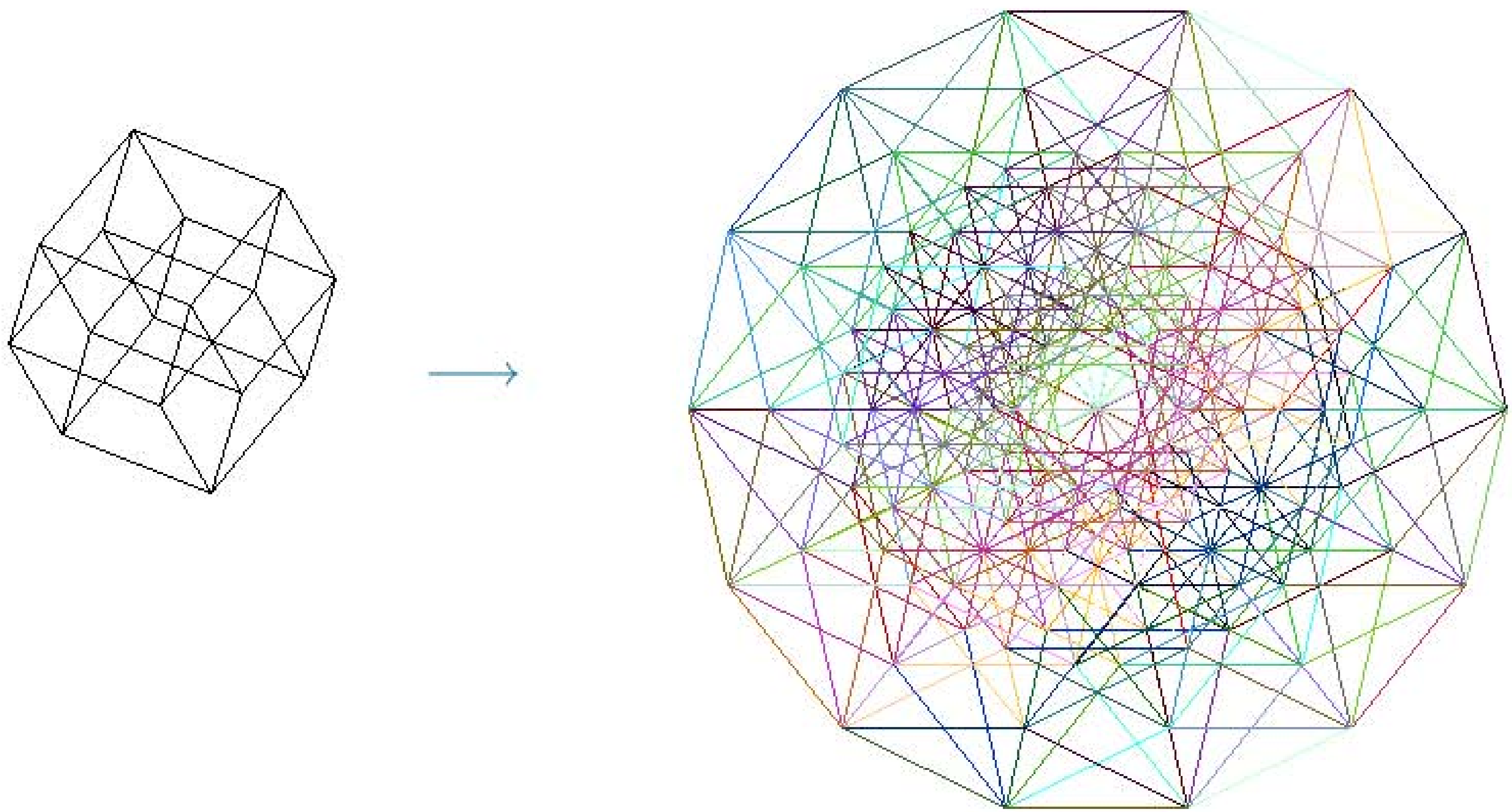


www.inference.phy.cam.ac.uk/dasher/

Block codes

- (7,4) Hamming Code

s	t	s	t	s	t	s	t
0000	0000 000	0100	0100 110	1000	1000 101	1100	1100 011
0001	0001 011	0101	0101 101	1001	1001 110	1101	1101 000
0010	0010 111	0110	0110 001	1010	1010 010	1110	1110 100
0011	0011 100	0111	0111 010	1011	1011 001	1111	1111 111



(7,4) Hamming Code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M = 3 \\ \downarrow \end{matrix}$$

$$\longleftarrow \quad N = 7 \quad \longrightarrow$$

Valid transmissions $\mathbf{t}$ satisfy

$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$

(7,4) Hamming Code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M=3 \\ \downarrow \end{matrix}$$

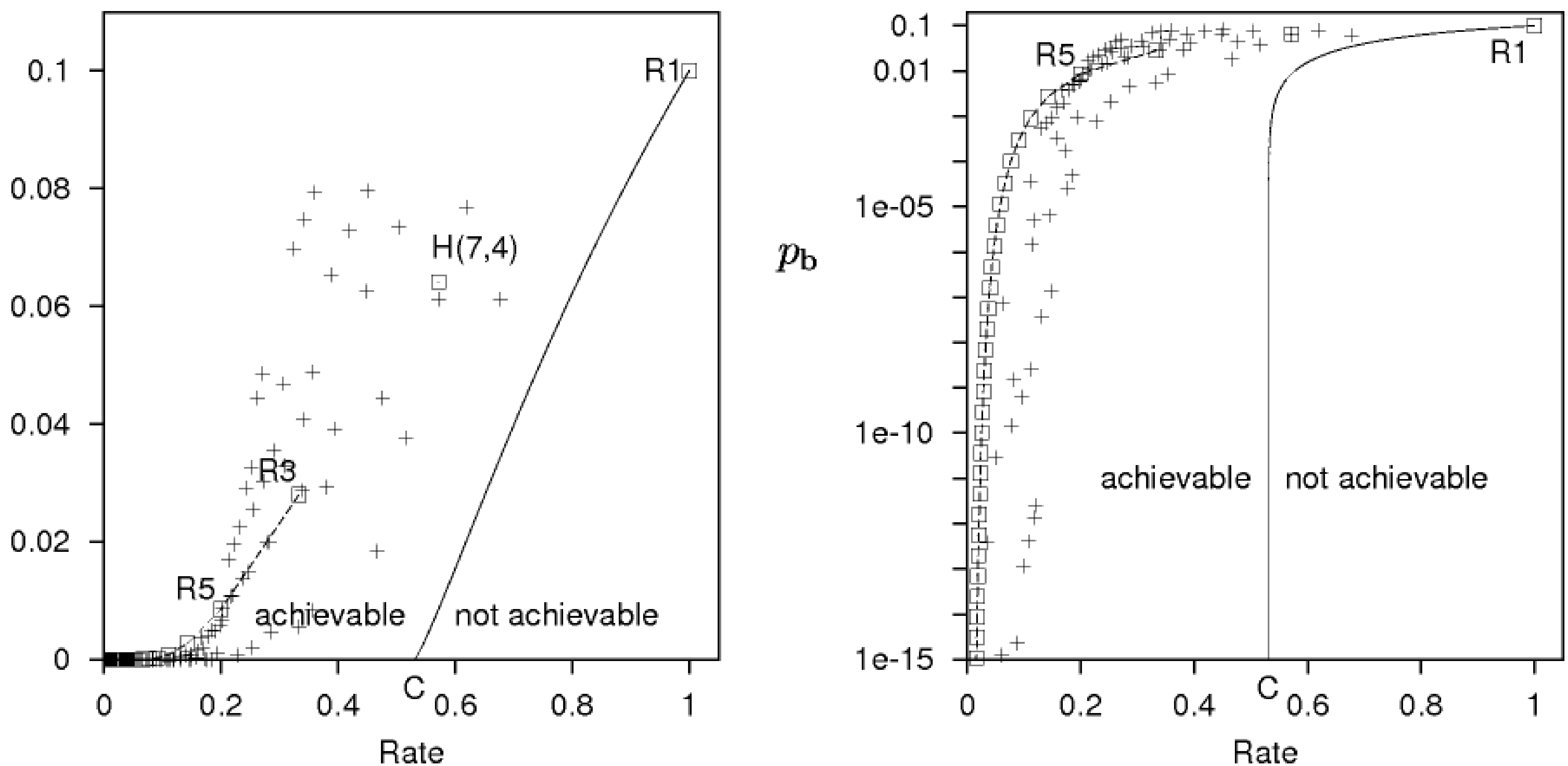
$$\longleftarrow \quad N=7 \quad \longrightarrow$$

Valid transmissions $\mathbf{t}$ satisfy

$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$

Received signal $\mathbf{r} = \mathbf{t} + \mathbf{n}$
Syndrome $\mathbf{z} = \mathbf{Hr} = \mathbf{Hn}$.
Syndrome decoder $\mathbf{z} \longrightarrow \hat{\mathbf{n}}$.

Shannon's noisy-channel coding theorem



$$C_{\text{BSC}}(f) = 1 - H_2(f)$$

$$H_2(f) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f}$$

How to prove good codes exist

- Constructive proof

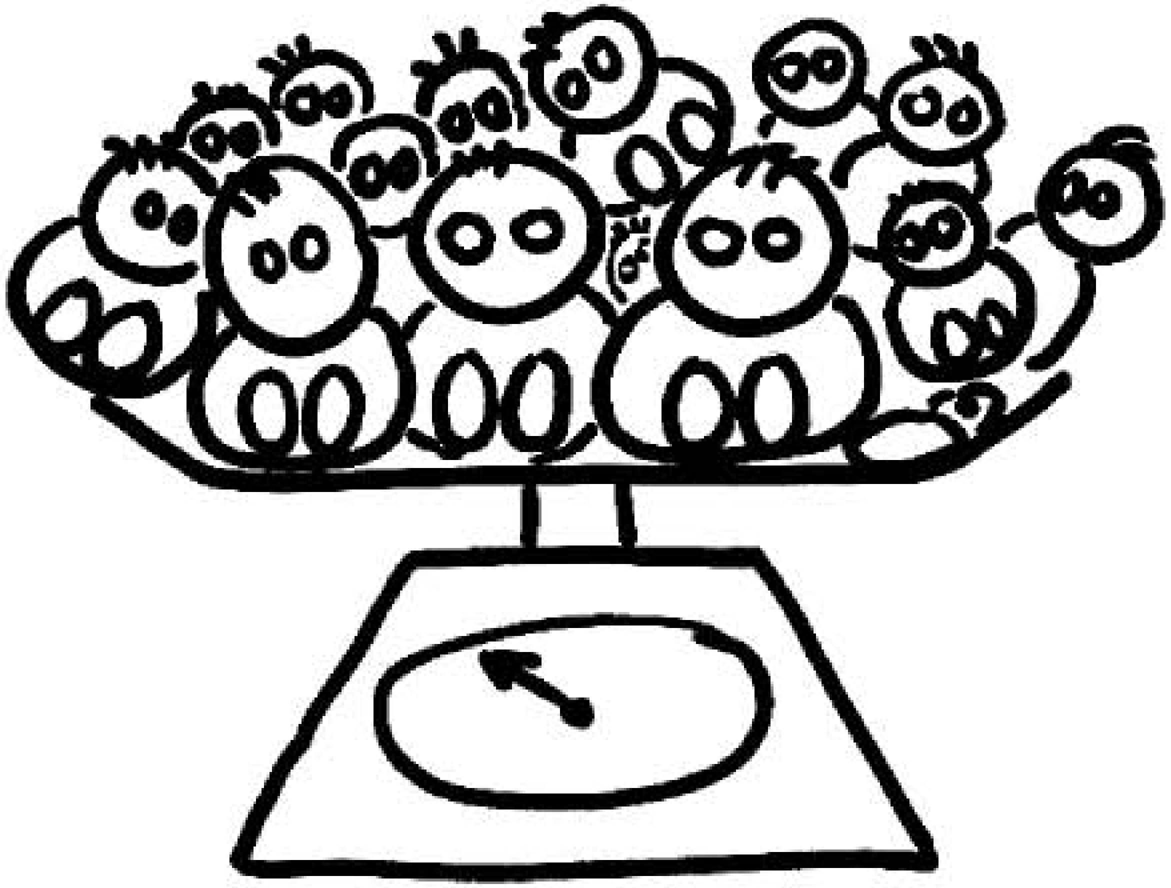
  Given required $R < C$, and $\epsilon > 0$,

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & & & \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & & & \\ \vdots & & & \vdots & & & \vdots & \ddots & & \\ \vdots & & & \vdots & & & \vdots & & \ddots & \\ \vdots & & & \vdots & & & \vdots & & & \ddots \end{bmatrix}$$

- Non-constructive proof

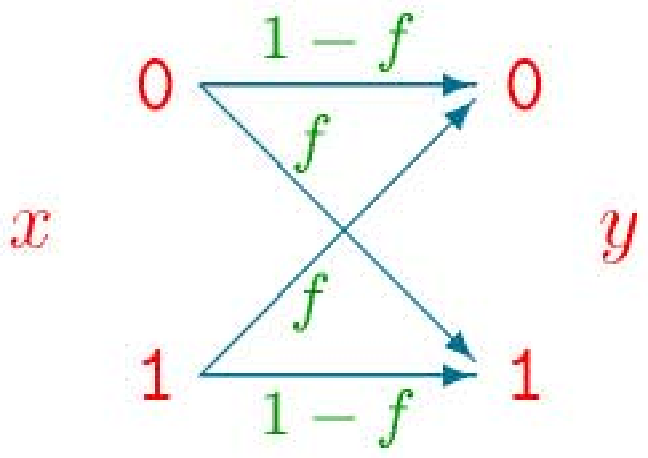
Iraq baby malnutrition analogy

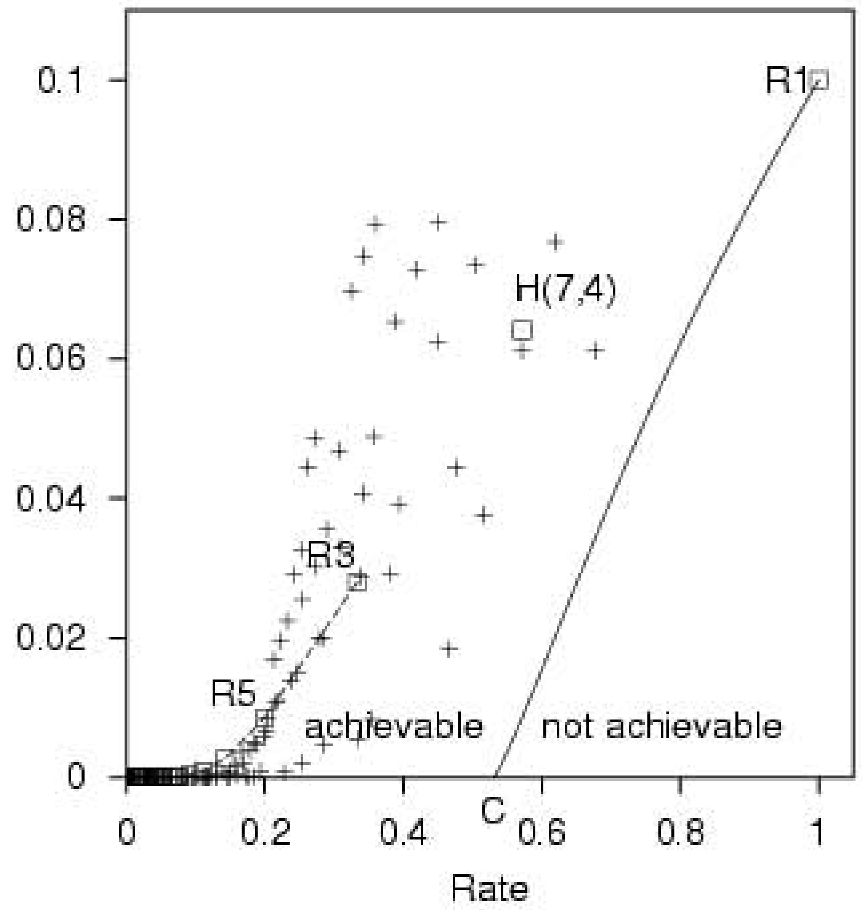
Shannon's way of proving malnutrition



If average weight
of all babies is $< \epsilon$,
there must be (at least!)
one baby with weight $< \epsilon$.

Shannon's noisy channel coding theorem





For any channel:

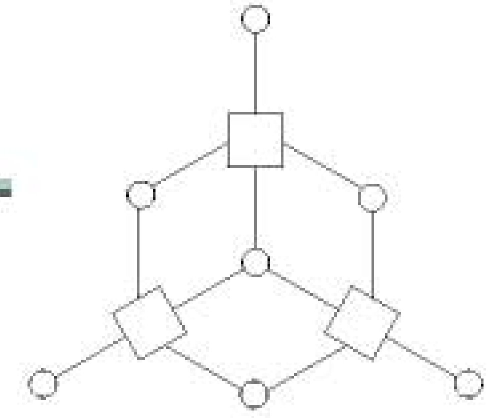
Reliable (virtually error-free) communication is possible at rates up to C

Information theory Shannon, 1948

Coding theory Hamming, 1948; Reed-Solomon; Forney (Convolutional & concatenated codes)

Skip to page 44

Idea



- Decoding problems, such as

$$P_1(\mathbf{x}) = \frac{1}{Z_1} e^{\beta\left[x_1x_2x_3x_5 + x_2x_3x_4x_6 + x_1x_3x_4x_7\right] + \sum_{n=1}^{N} b_n x_n}$$

look a bit like Boltzmann machines + Hopfield networks.... so

- Solve the decoding problem

$$\max_{\mathbf{x}} P_1(\mathbf{x})$$

using variational methods?

Electronics Letters, March 1995

Free energy minimisation algorithm for decoding and cryptanalysis

D.J.C. MacKay

Indexing terms: Decoding, Cryptography

An algorithm is derived for inferring a binary vector **s** given nois observations of **As** modulo 2, where **A** is a binary matrix. Th binary vector is replaced by a vector of probabilities, optimised b free energy minimisation. Experiments on the inference of th state of a linear feedback shift register indicate that this algorithr supersedes the Meier and Staffelbach polynomial algorithm.

Decoding error-correcting codes

- For which codes are approximate message-passing methods **effective?**

Electronics Letters, August 1996

Near Shannon limit performance of low density parity check codes

D.J.C. MacKay and R.M. Neal

[Low Density Parity Check Codes: Gallager 1962]

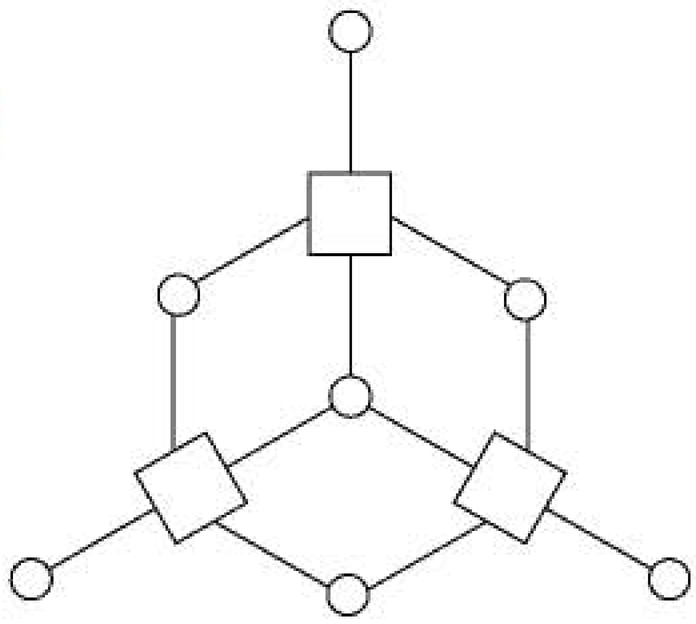
Indexing terms: Probabilistic decoding, Error correction codes

The authors report the empirical performance of Gallager's low density parity check codes on Gaussian channels. They show that performance substantially better than that of standard convolutional and concatenated codes can be achieved; indeed the performance is almost as close to the Shannon limit as that of turbo codes.

(7,4) Hamming Code - recap

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M \\ \downarrow \end{matrix}$$

$$\longleftarrow \quad N = 7 \quad \longrightarrow$$



factor graph

Valid transmissions $\mathbf{t}$ satisfy

$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$

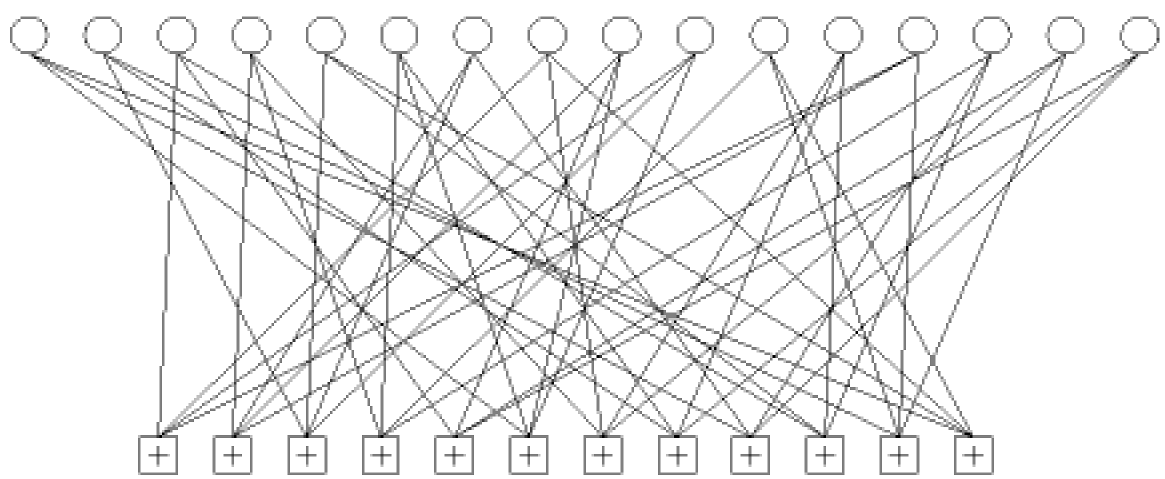
Received signal $\mathbf{r} = \mathbf{t} + \mathbf{n}$

Syndrome $\mathbf{z} = \mathbf{H}\mathbf{r} = \mathbf{H}\mathbf{n}$.

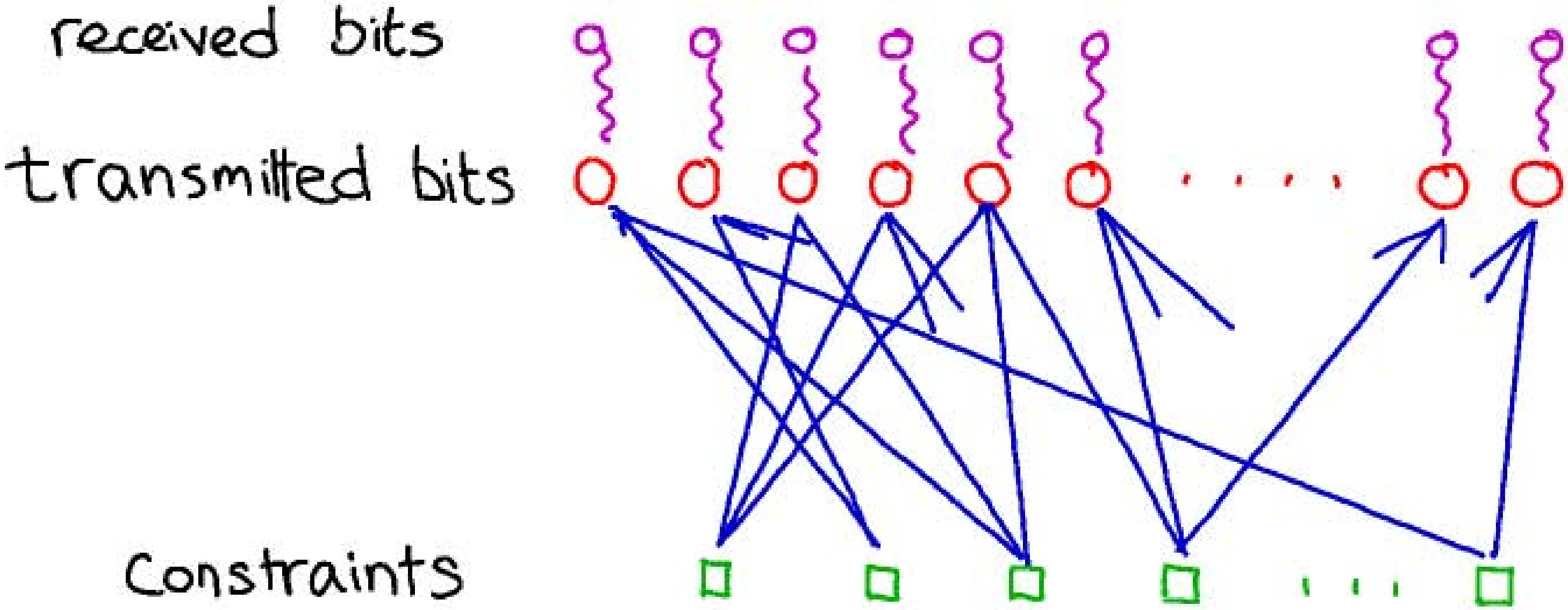
Syndrome decoder $\mathbf{z} \longrightarrow \hat{\mathbf{n}}$.

Low-density parity-check code

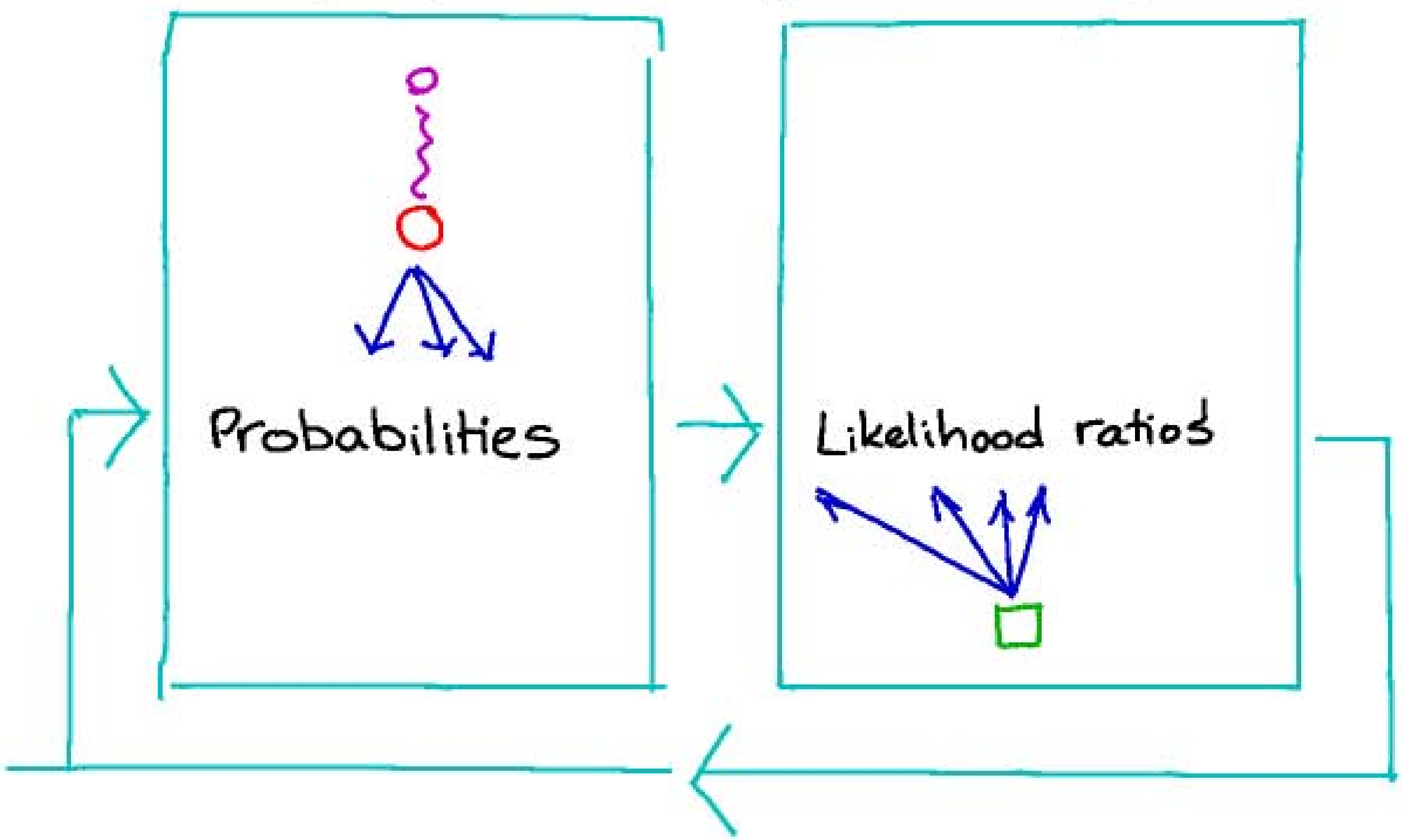
$$\mathbf{H} = \begin{array}{|cccccccccccccccc|}
 & & 1 & & & & & 1 & & 1 & & & 1 & & & \\
 & & & 1 & & & 1 & & 1 & & & & 1 & & & \\
 & 1 & & & 1 & & 1 & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & 1 & & & 1 & & \\
 & & 1 & & & & & & 1 & & & & & & 1 & 1 \\
1 & & & & & 1 & & & 1 & 1 & & & & & & \\
 & & & 1 & & & & 1 & & & & 1 & & & 1 & \\
 & 1 & & & & 1 & & & & & & 1 & & & & 1 \\
1 & & & & & & 1 & & & & & & & 1 & & 1 \\
 & & 1 & & 1 & & & & & & & 1 & & 1 & & \\
 & 1 & & & & & & & & & 1 & & 1 & & 1 & \\
1 & & & & 1 & & & 1 & & & 1 & & & & &
\end{array}$$



Gallager 1962; MacKay & Neal 1995



Decoding by the sum-product algorithm

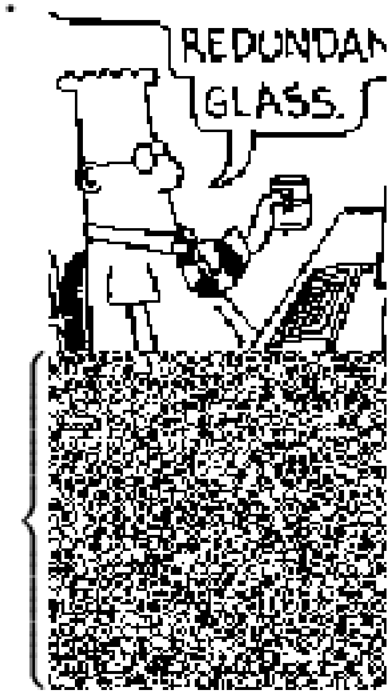


loopy belief propagation Gallager 1962

Low Density Parity Check Code

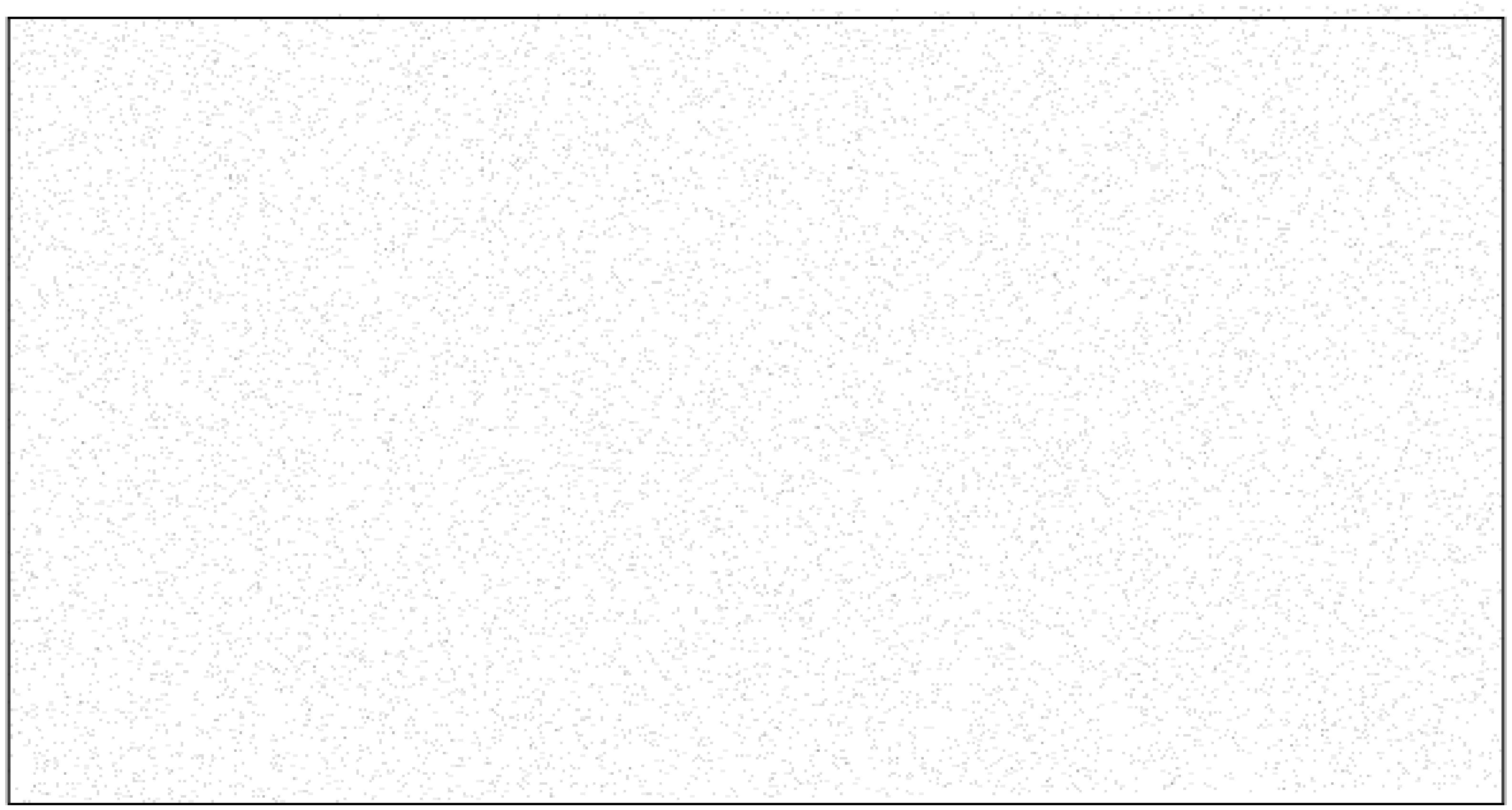
We demonstrate a large code that encodes $K = 10000$ source bits into $N = 20000$ transmitted bits.
Each parity bit depends on about 5000 source bits.
The encoder is derived from a very sparse 10000×20000 matrix $\mathbf{H}$ with three 1s per column.

TRANSMITTED:



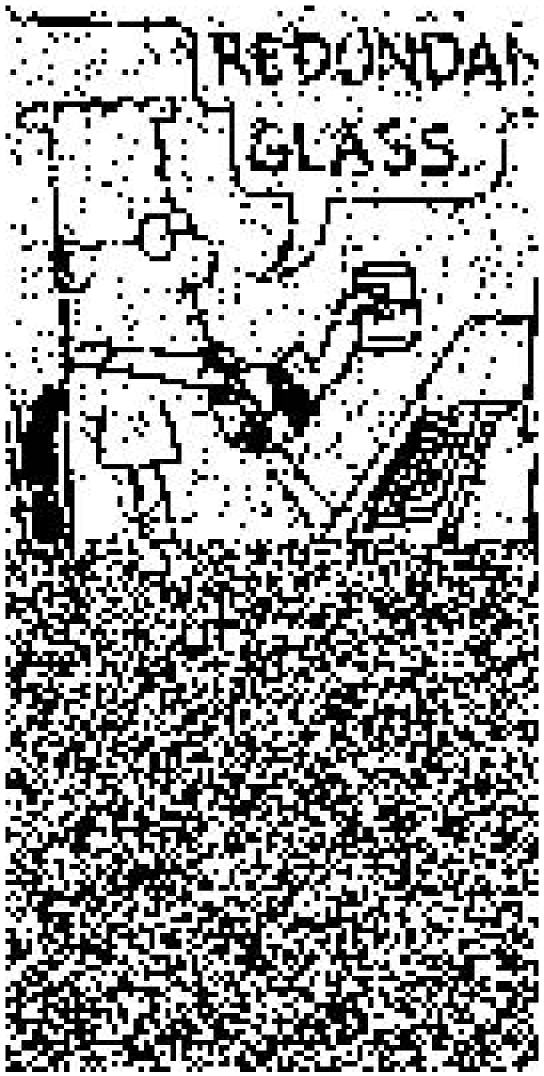
parity bits

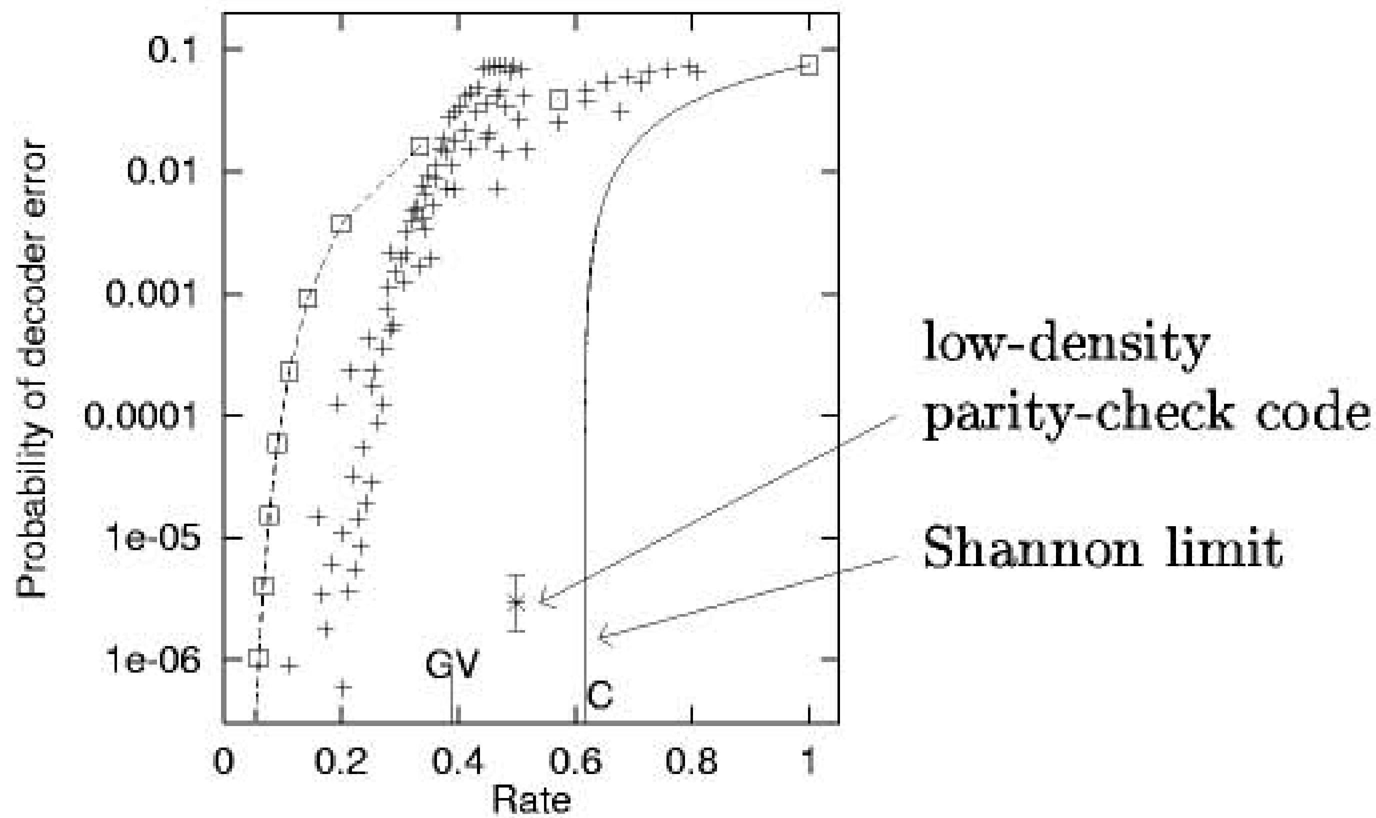
$\mathbf{H} =$



Low Density Parity Check Code (f = 7.5%)

Iterative probabilistic decoding



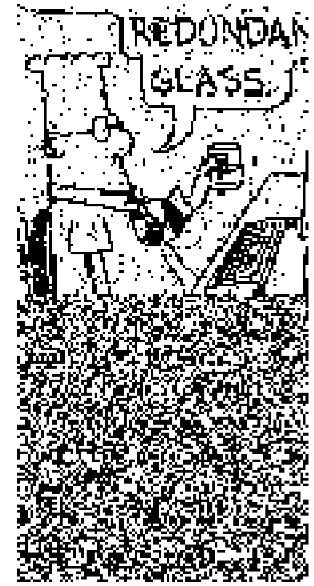


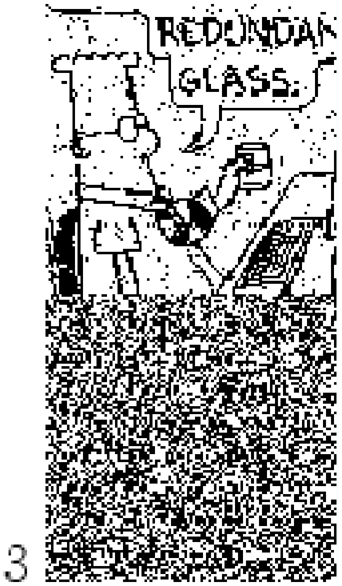
After the transmission is sent over a channel with noise level $f = 7.5\%$:

RECEIVED:

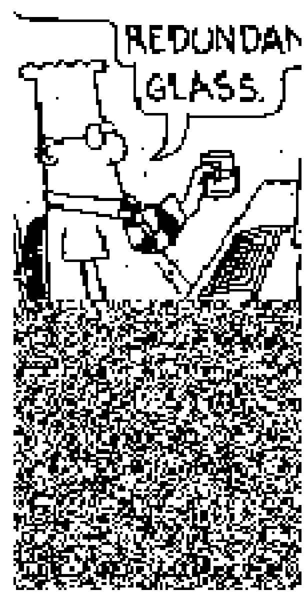
0 

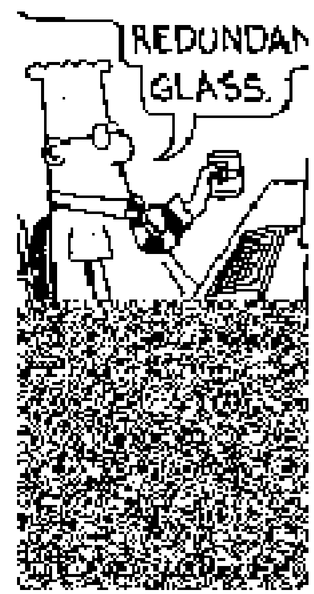
1 

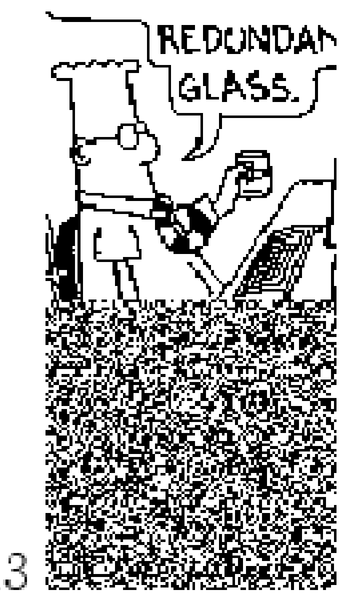
2 

3 

10 

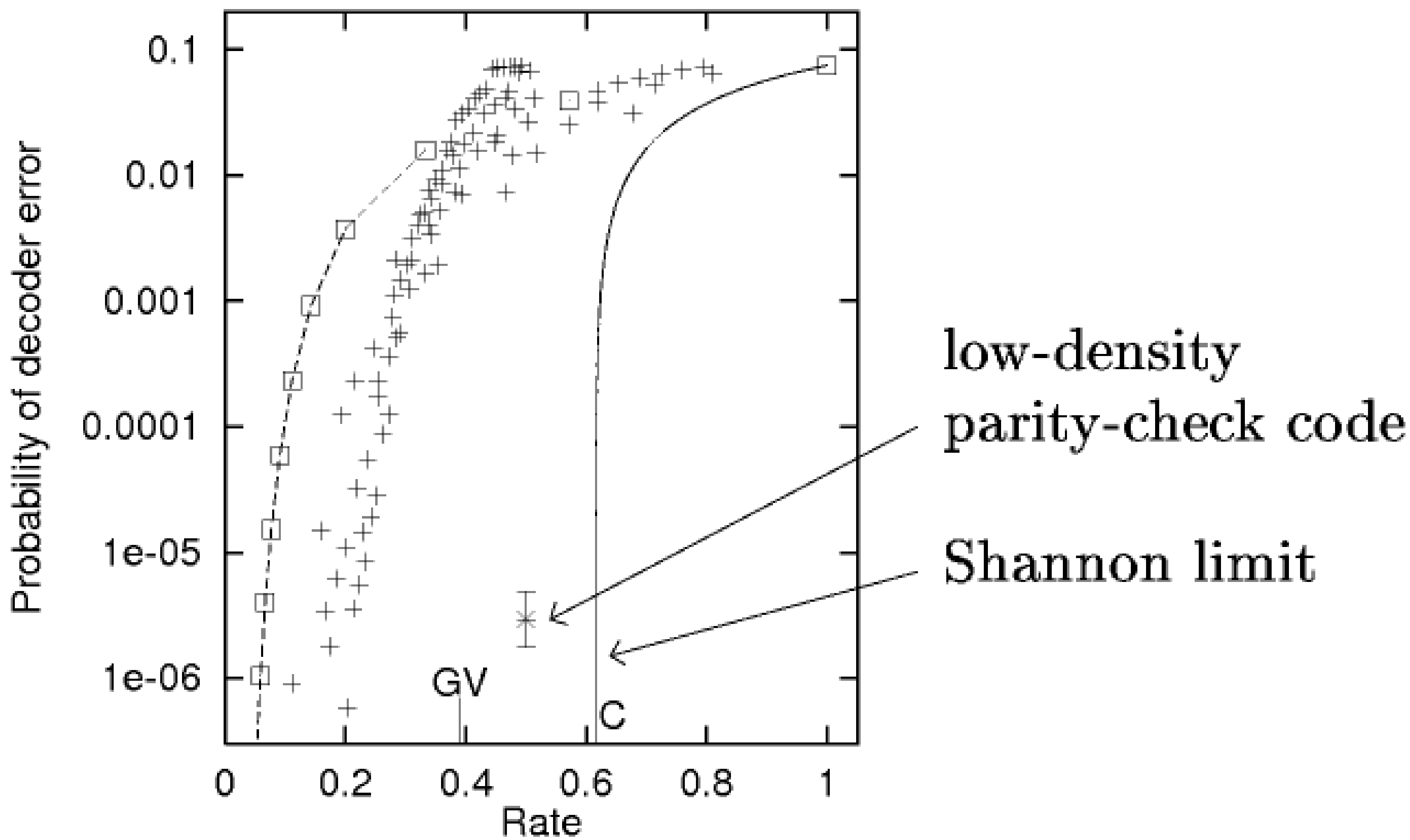
11 

12 

13 

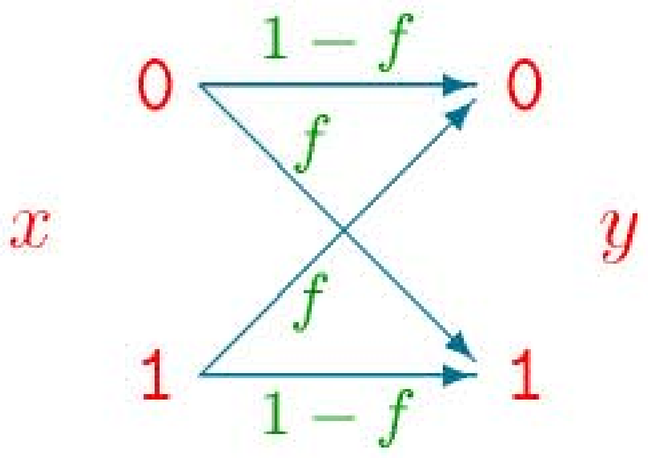
$\rightarrow$ **DECODED:**

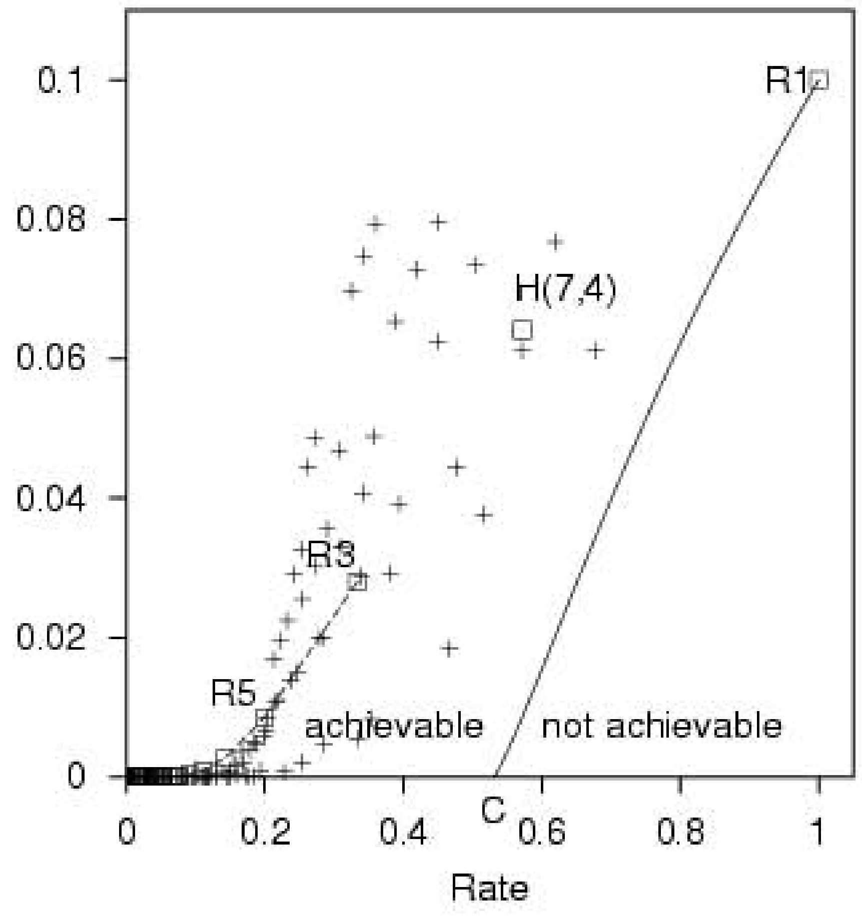
BSC: f=7.5%



0.1
0.01
0.001
0.0001
1e-05
1e-06
Probability of decoder error
GV
C
0
0.2
0.4
0.6
0.8
1
Rate
low-density
parity-check code
Shannon limit

Rate, error probability, complexity

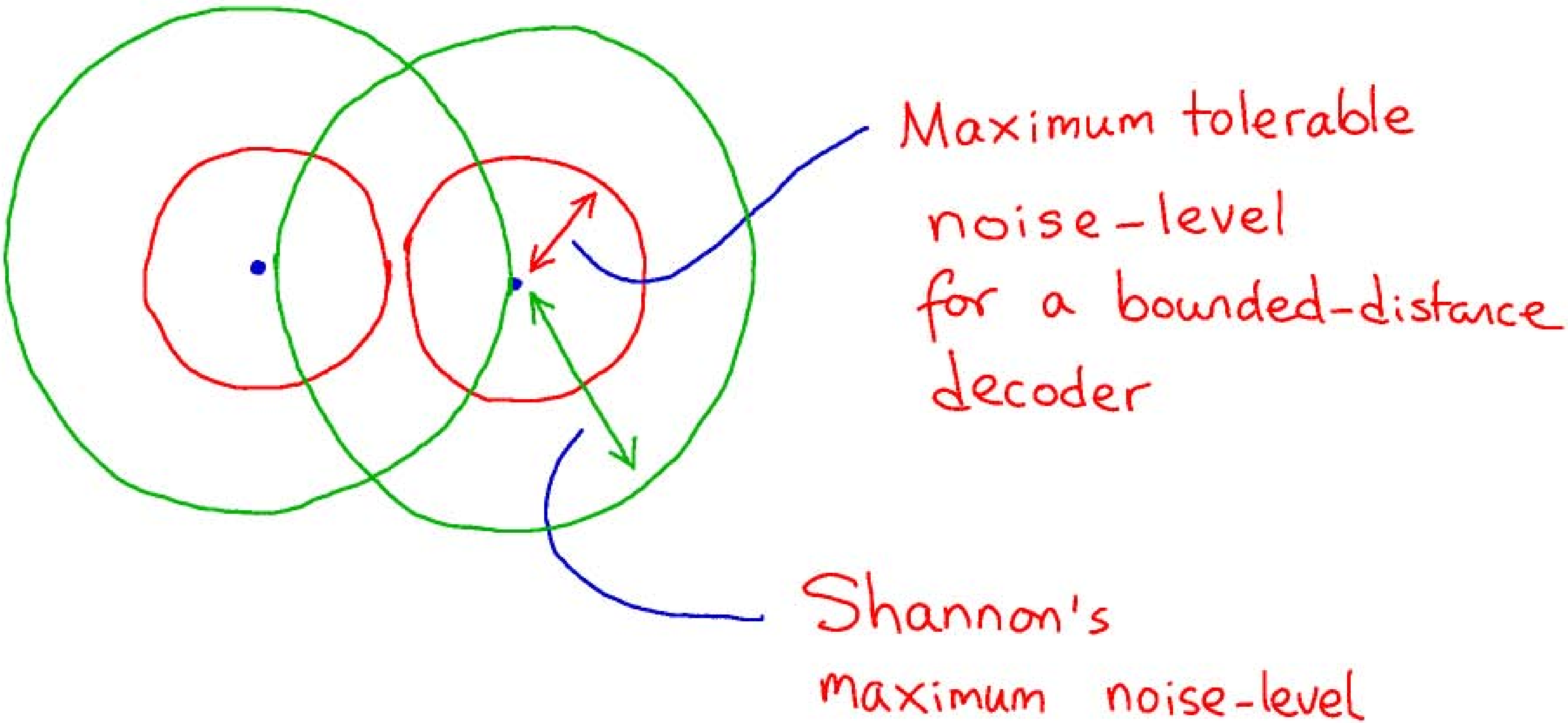




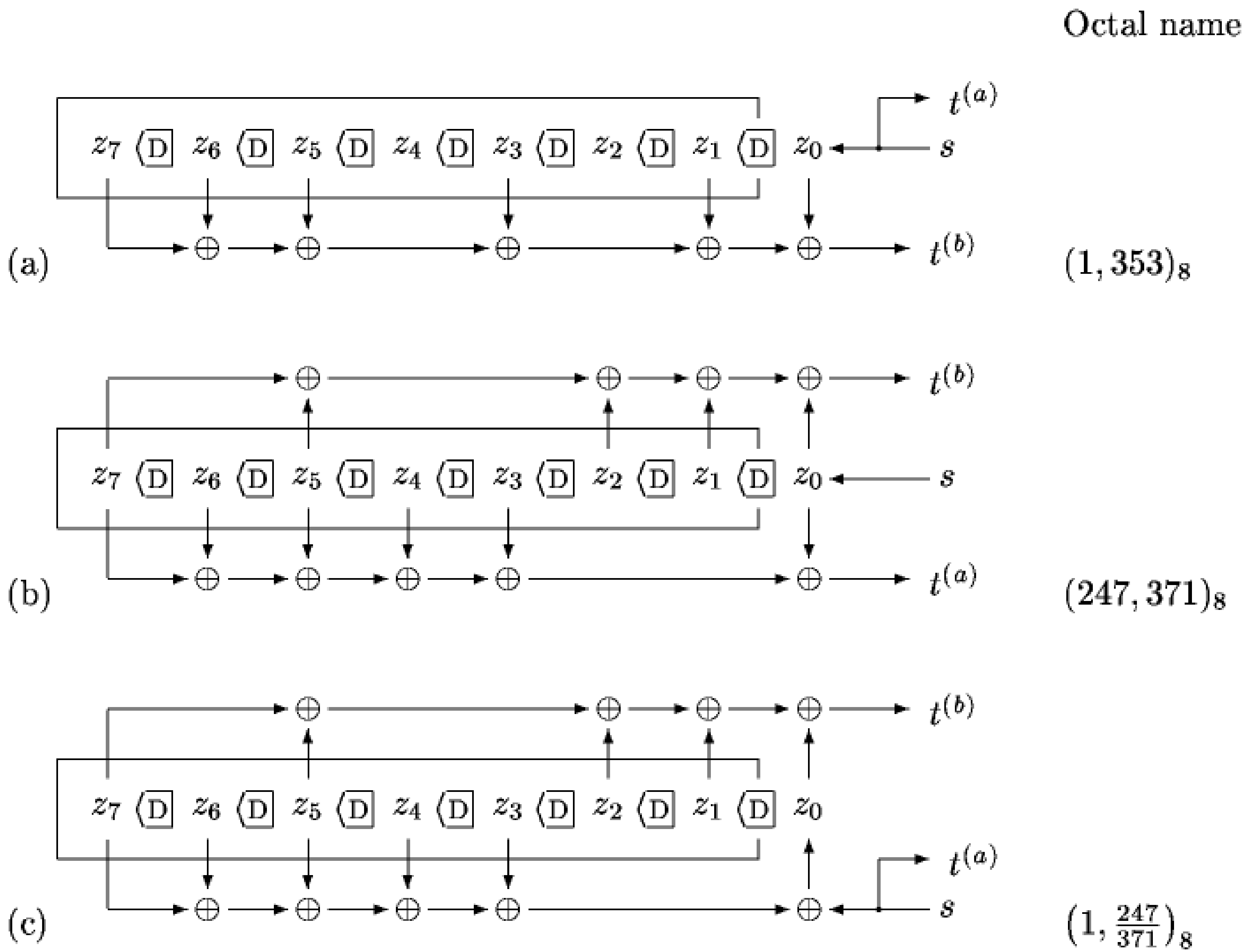
For any channel:

Reliable (virtually error-free) communication is possible at rates up to C

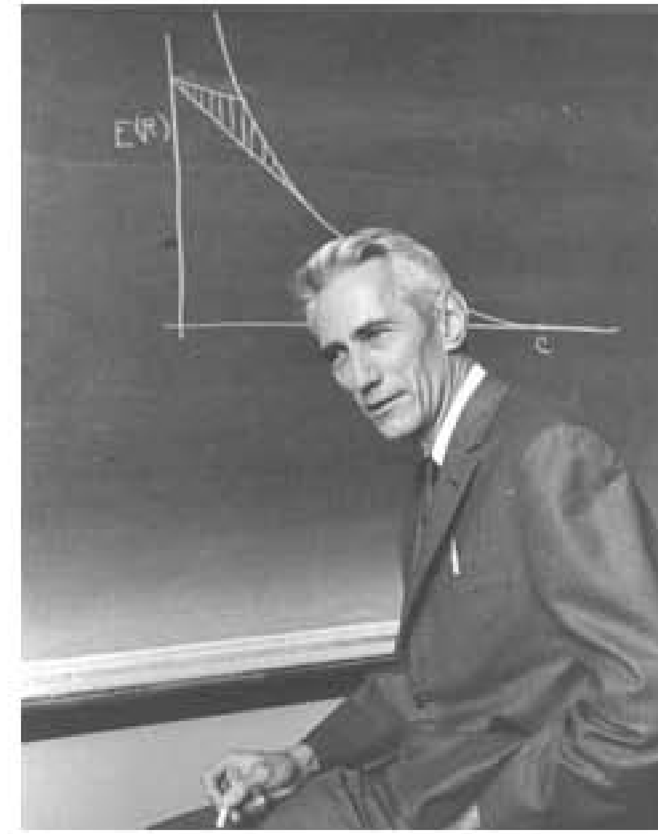
Two codewords



Convolutional codes



Feedback



`Feedback? Pah! Who needs feedback?
Just use a random code!'

Sphere packing view

Count inputs and outputs → get a bound on what's achievable.
Given a transmission of length N,
the output will probably have Nf bits flipped,
so it will be in a typical set of size

$$\binom{N}{Nf} \simeq 2^{NH_2(f)}$$

So if we have 2^K alternative inputs, and almost all these typical outputs are distinct, we must have

$$\overbrace{2^K \times 2^{NH_2(f)}}^{\text{TOTAL NUMBER OF TYPICAL OUTPUTS}} \leq \overbrace{2^N}^{\text{TOTAL SIZE OF OUTPUT SPACE}}$$

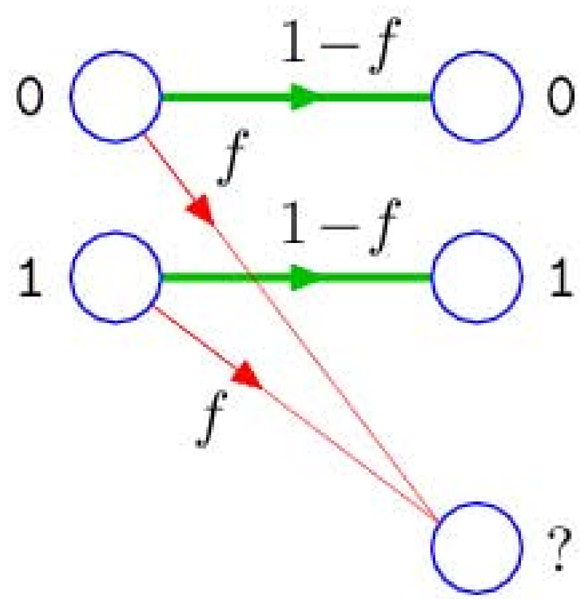
i.e.,

$$K + NH_2(f) \leq N$$

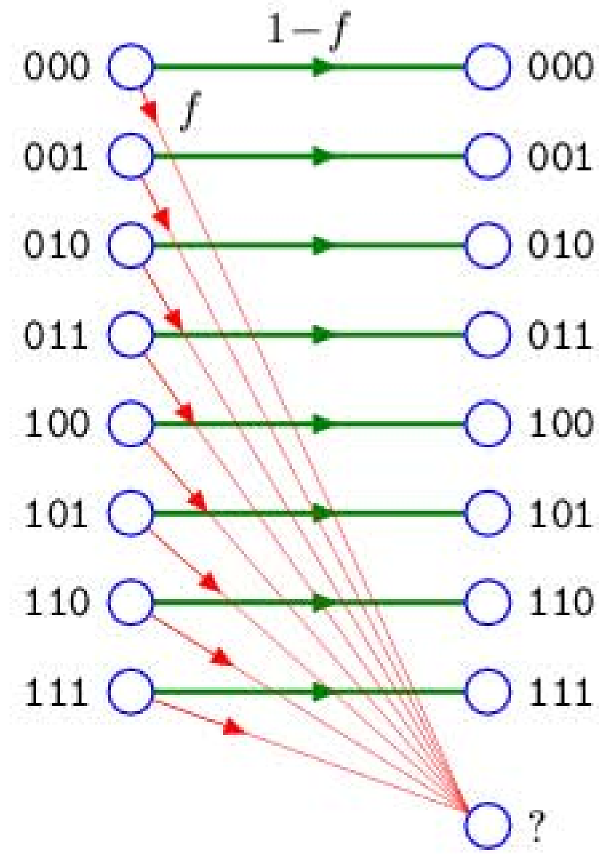
i.e.,

$$\frac{K}{N} \leq 1 - H_2(f)$$

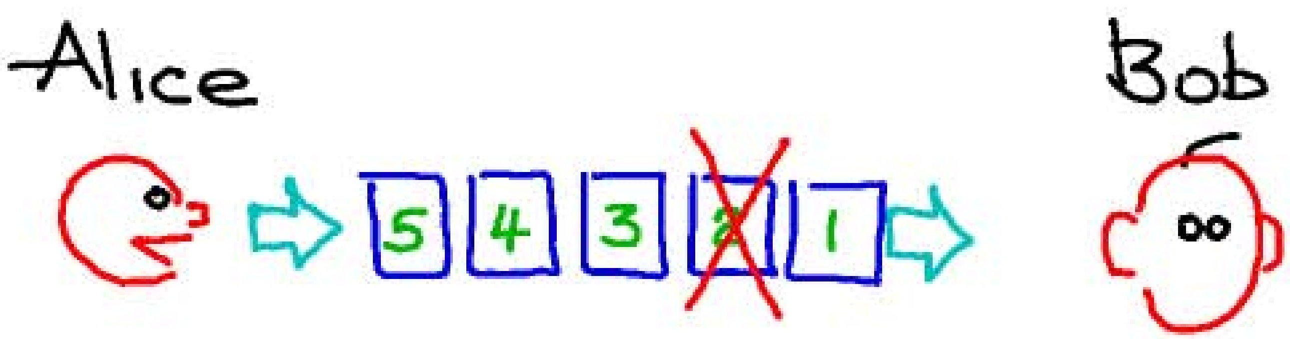
Channels with erasures



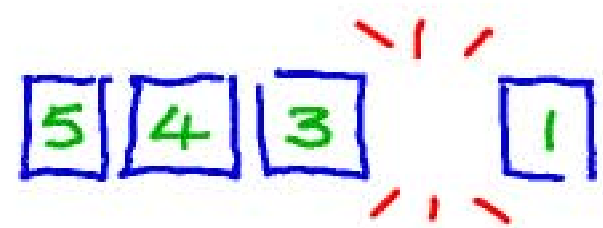
Binary erasure channel



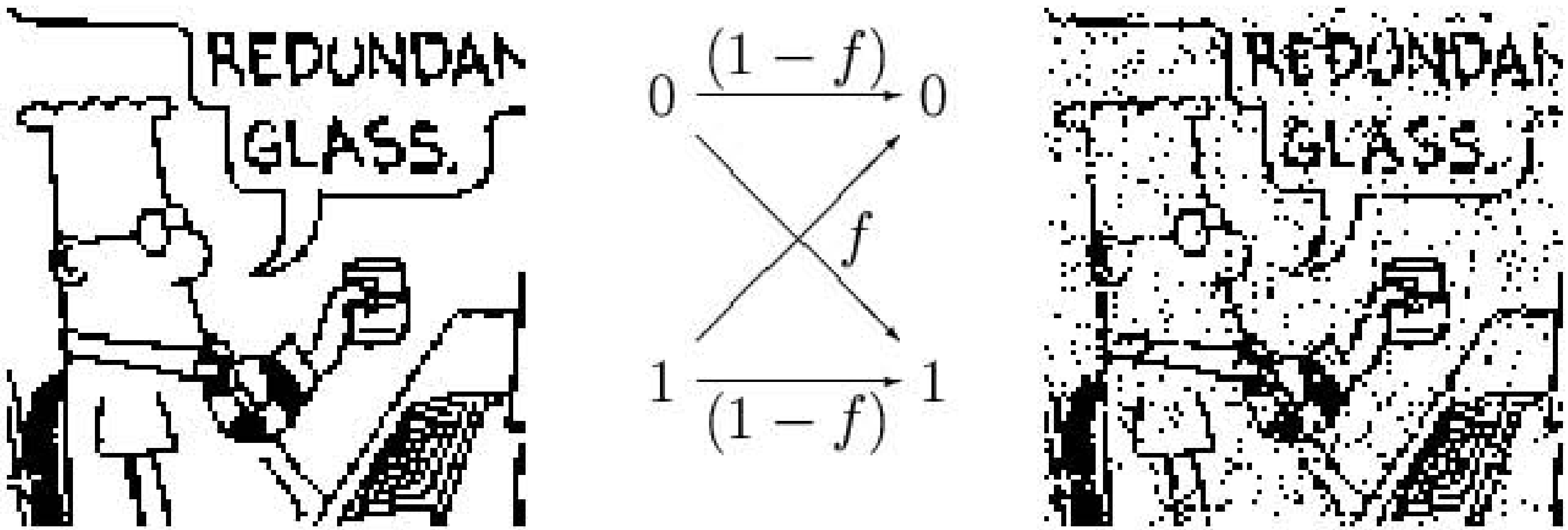
8-ary erasure channel

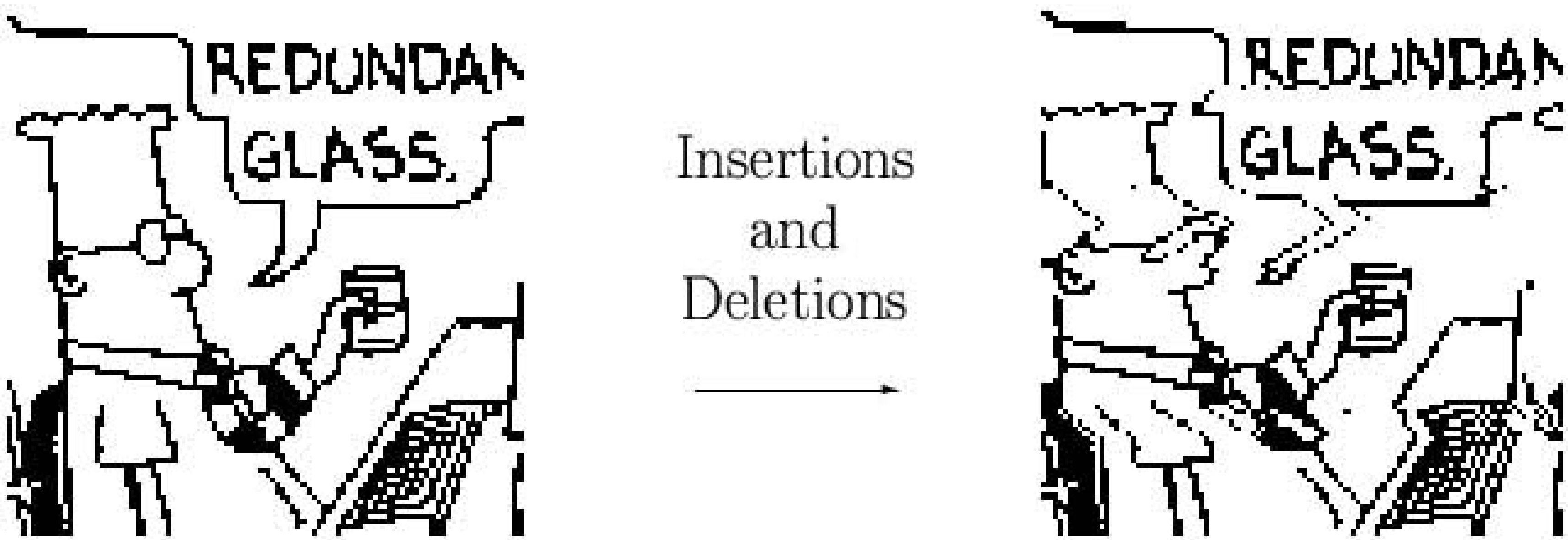


Packet-erasing channel



There are other noisy channels...





synchronization errors

Write free software and free books, and who knows where you will end up

TIMESONLINE

NEWS | COMMENT | BUSINESS | MONEY | SPORT | LIFE & STYLE | TRAVEL | DRIVING

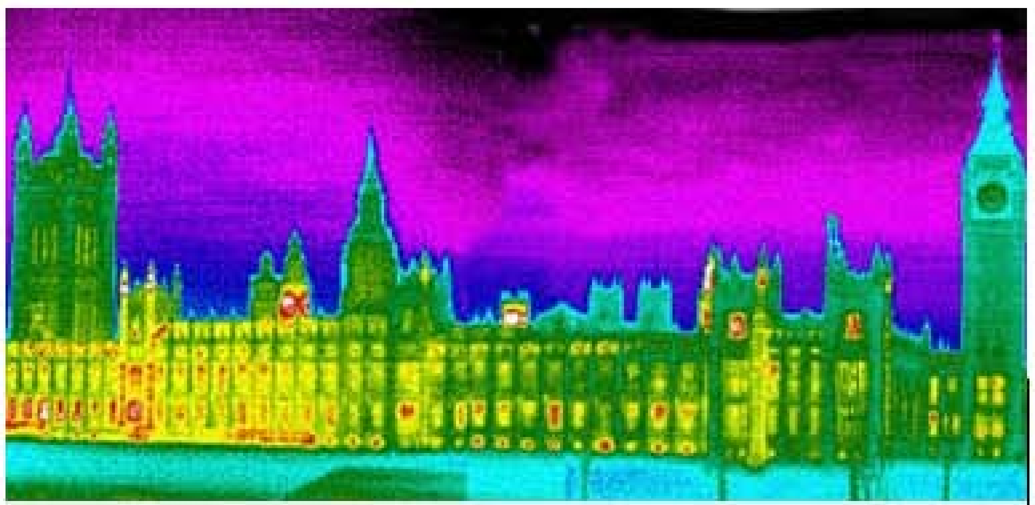
UK NEWS | WORLD NEWS | POLITICS | SCIENCE | ENVIRONMENT | WEATHER | TECH & WEB | VIDEO

Where am I? > Home > News > Environment

From The Sunday Times

August 30, 2009

Bestselling guru David MacKay to lead climate fight



A thermal imaging camera reveal large amounts of hot air escaping from the Houses of Parliament

Dominic O'Connell

✔ RECOMMEND?

A CAMBRIDGE academic who has suggested importing solar energy from the Sahara and using Scottish lakes as giant batteries is to be named the government's scientific adviser on climate change.

REBECCA FRAYN:

THE INDEPENDENT ON SUNDAY

BUSINESS

News | Opinion | Environment | Sport | Life & Style | Arts & Entertainment | Tr

UK | World | Business | People | Science | Media | Education | Obituaries | Corrections

Home > News > Business > Business News

Nuclear energy firms welcome policy adviser

Government to appoint Professor David MacKay to help with its ambitious climate change strategy

By Mark Leftly

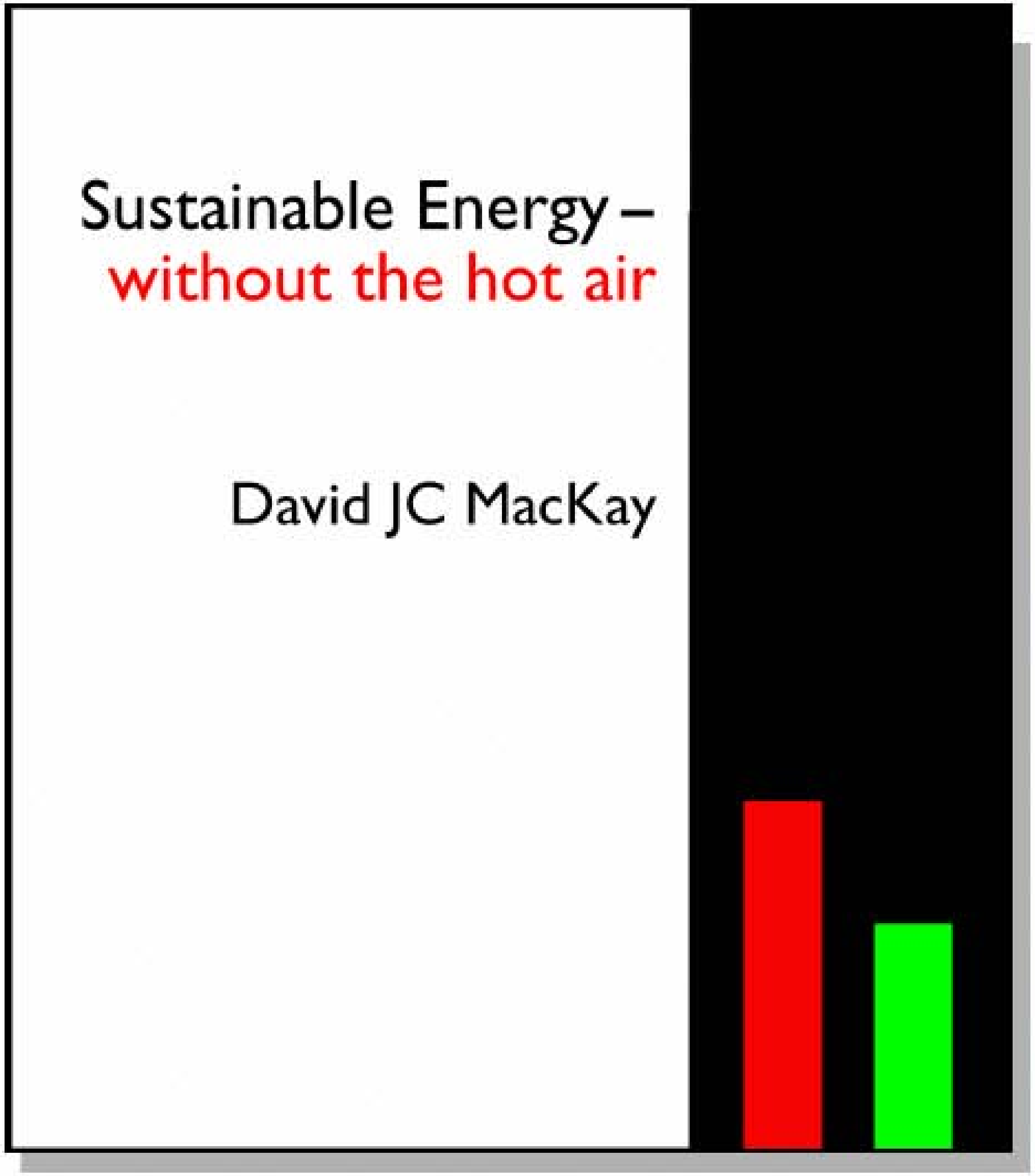
Sunday, 30 August 2009

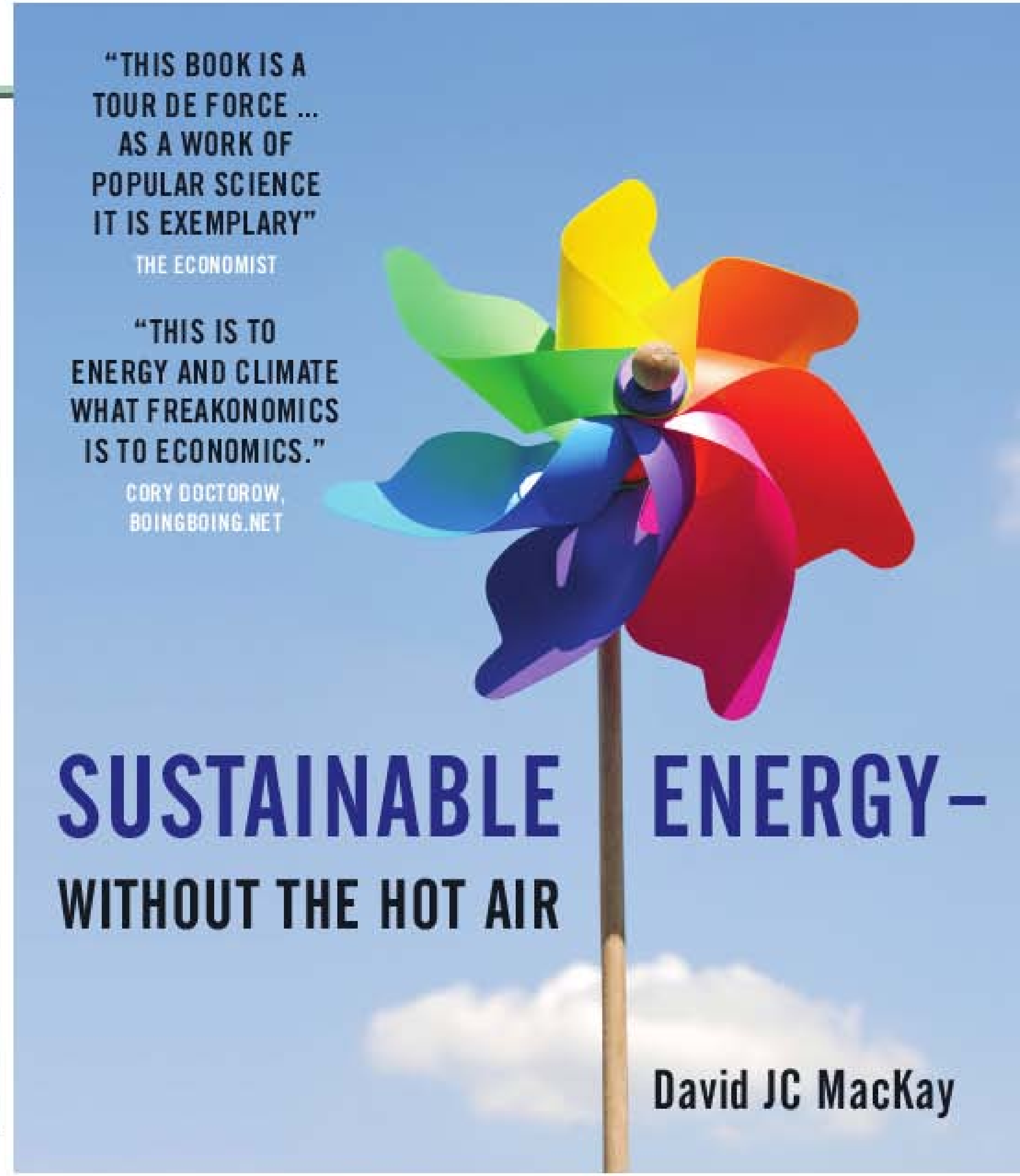
SHARE | PRINT | EMAIL | AAA TEXT SIZE

Ed Miliband, the Energy and Climate Change Secretary, is set to appoint Cambridge University Professor David MacKay as his chief scientific officer this week.

The move will be a boon to the British energy sector: industry leaders from Royal Dutch Shell EDF Energy and QinetiQ have all praised Professor MacKay's hugely successful book, Sustainable Energy: Without the Hot Air. Companies looking to get involved in the Government's nuclear roll-out programme will be particularly hopeful that his appointment will quash some of the political arguments against the plans.

Another book





www.withouthotair.com