Error Correcting Codes
&
Probability Propagation

David J.C. MacKay
Department of Physics, Cavendish laboratory, University of Cambridge

http://wol.ra.phy.cam.ac.uk/mackay/

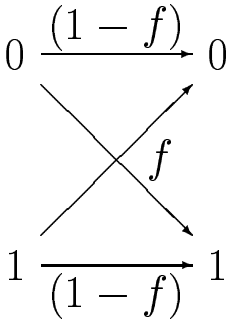
The central problem of information theory:

To achieve reliable communication over an unreliable channel.

Unreliable channels include satellite links, telephone lines, disc drives.

An idealized noisy channel:

Binary symmetric channel, noise level $f = 7.5\%$

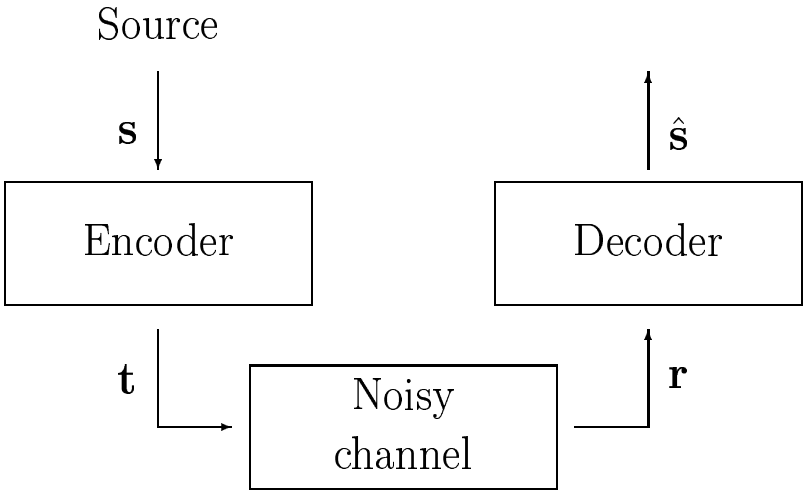


7.5% of bits are flipped

How to achieve reliable communication?

We would like to achieve virtually error-free communication. *e.g.*, an error probability of $\sim 10^{-15}$ per bit.

The 'system' solution for achieving reliable communication

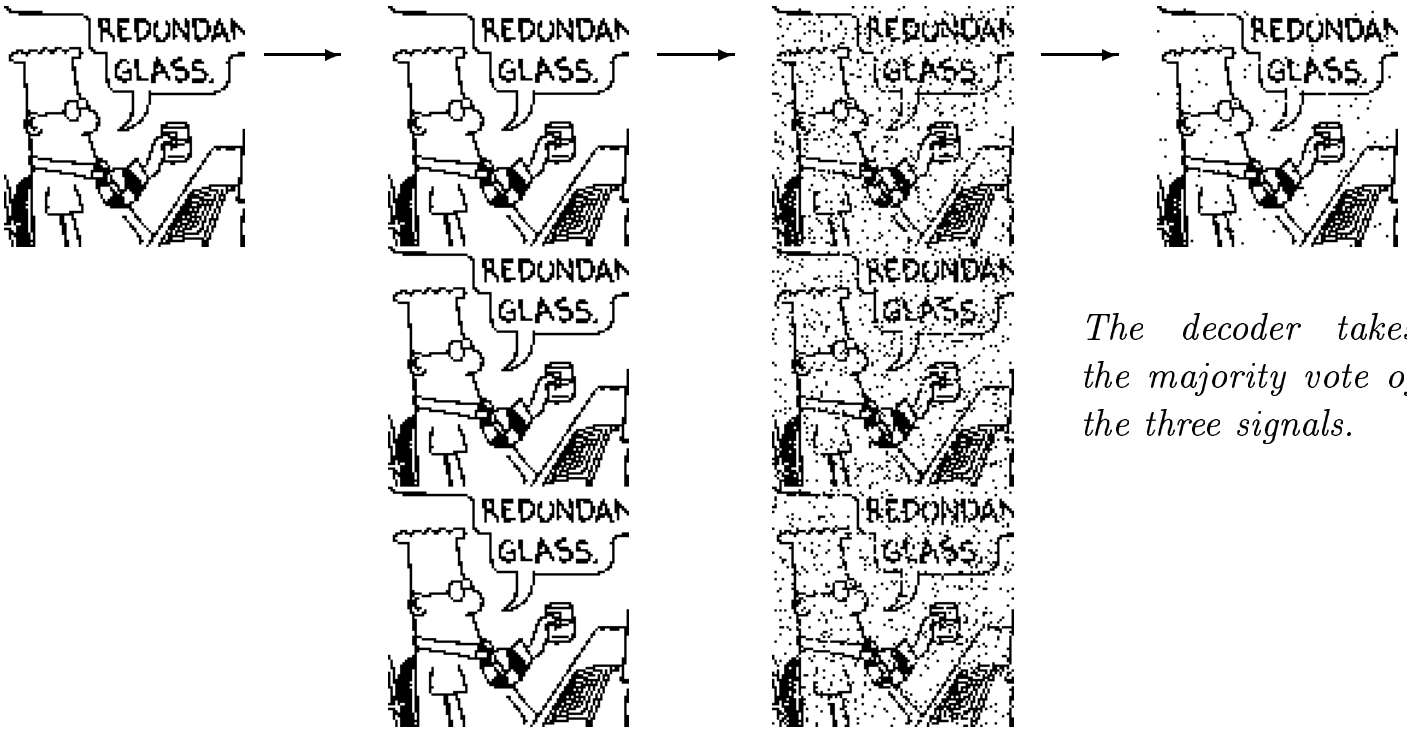


The encoding system introduces redundancy in some systematic way into the transmitted vector $\mathbf{t}$. The decoding system makes use of this known redundancy to deduce, given the received vector $\mathbf{r}$, what the noise introduced by the channel *and* the original source vector probably were.

A simple encoder: add redundancy by repetition

$\mathbf{s}$ ENCODER $\mathbf{t}$ CHANNEL $\mathbf{r}$ DECODER $\hat{\mathbf{s}}$

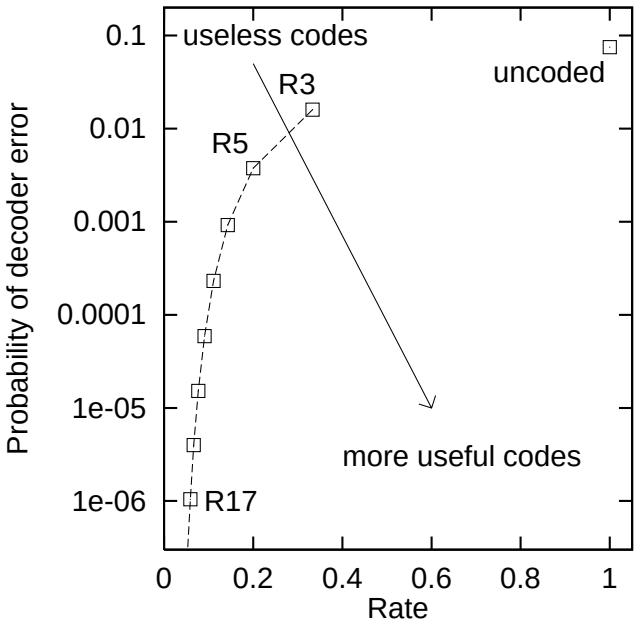
$f = 7.5\%$



The decoder takes the majority vote of the three signals.

Good news: only 1.6% of decoded bits are in error
Bad news: rate of communication reduced to 1/3

We can further reduce the error probability by repeating more times. With 17 repetitions ('R17'), the error probability is reduced to 10^{-6}. But the rate has fallen to 1/17.



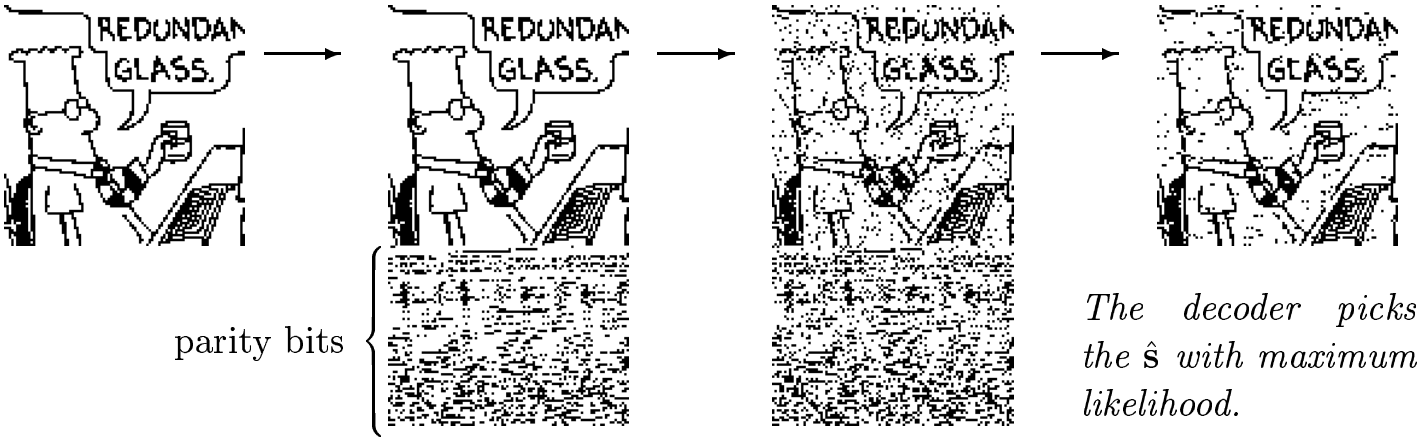
A more complex encoder: The (7,4) Hamming code

Every *four* source bits are protected with *three* **parity** bits.

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

s ENCODER **t** CHANNEL **r** DECODER $\hat{\mathbf{s}}$

$f = 7.5\%$



The decoder picks the $\hat{\mathbf{s}}$ *with maximum likelihood.*

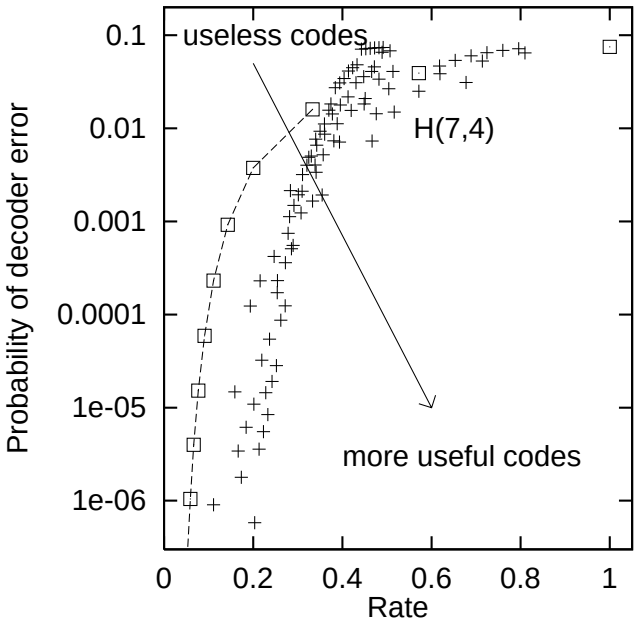
4% of decoded bits are in error

rate of communication is 4/7

In theory, what could the best codes achieve?

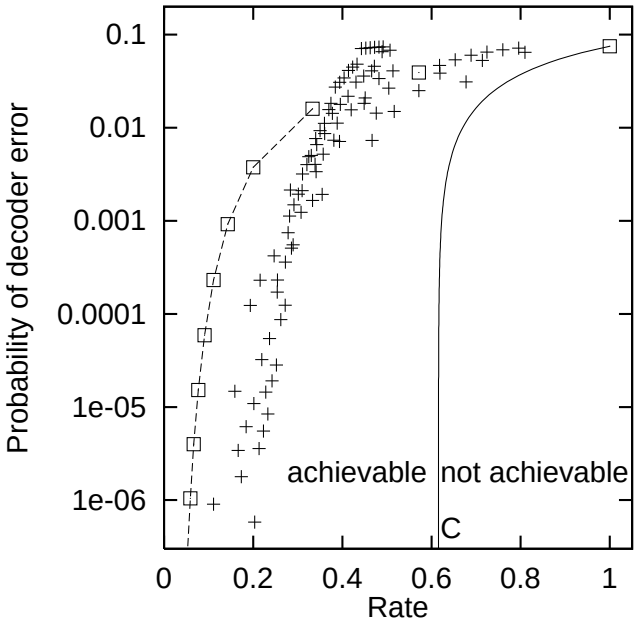
We would like *small* error probability and *large* rate.

The repetition codes, the Hamming (7,4) code, and related textbook codes (+) perform as shown here.



One might guess that the error probability can only be made very small by making the rate very small.

However, in 1948 Shannon proved the remarkable result that, for any given channel, the boundary between achievable and nonachievable points meets the R axis at a *non-zero* value $R = C$.



The practical challenge is to create error-correcting codes that get close to what Shannon proved is possible.

Low density parity check codes

Our codes are defined in terms of a *very sparse matrix*, $\mathbf{H}$, *e.g.*,

$$\mathbf{H} \quad = \quad \begin{array}{|cccccccccccc|}
\hline
 & 1 & & & & 1 & & 1 & & & 1 & \\
 & & 1 & & & & 1 & & 1 & & & \\
 & 1 & & & 1 & & 1 & & & & & \\
 & & 1 & & 1 & & & & & & & \\
 & 1 & & & & & 1 & & & & & 1 \\
1 & & & & 1 & & & 1 & & & & \\
 & & 1 & & & & 1 & & & & & \\
 & 1 & & & 1 & & & & & & & \\
1 & & & & & 1 & & & & & & \\
 & & 1 & & 1 & & & & & & & \\
 & 1 & & & & & & & & & & \\
1 & & & & 1 & & & & & & & \\
\hline
\end{array}$$

which is known to the encoder and decoder.

The encoding method and the decoding problem

We use encodings, $\mathbf{t}$ which satisfy $\mathbf{Ht} = 0 \pmod 2$. These consist of a source message $\mathbf{s}$ followed by appropriately chosen parity checks. The received vector is $\mathbf{r} = \mathbf{t} + \mathbf{n} \pmod 2$, where $\mathbf{n}$ is the noise. The receiver knows $\mathbf{H}$ and can compute $\mathbf{z} = \mathbf{Hr} = \mathbf{Ht} + \mathbf{Hn} = \mathbf{Hn}$. The decoding problem is then to find the sparsest vector $\mathbf{x}$ satisfying the equation

$$\mathbf{Hx} = \mathbf{z} \pmod 2,$$

this $\mathbf{x}$ being the best guess for $\mathbf{n}$. If we can find $\mathbf{n}$, we can find $\mathbf{t}$, and from that the original message.

$\mathbf{H}$ is very sparse, and the $\mathbf{x}$ we are trying to find is sparse, so this problem doesn't sound intractable.

History

These codes were first studied in 1962 by Gallager, but were then generally forgotten by the coding theory community.

Theoretical result:

Low density parity check codes, in spite of their simple construction, are very good codes, *given an optimal decoder*.

Practical results:

We have developed two decoding strategies.

1: Mean field methods – inspired by statistical physics and neural networks

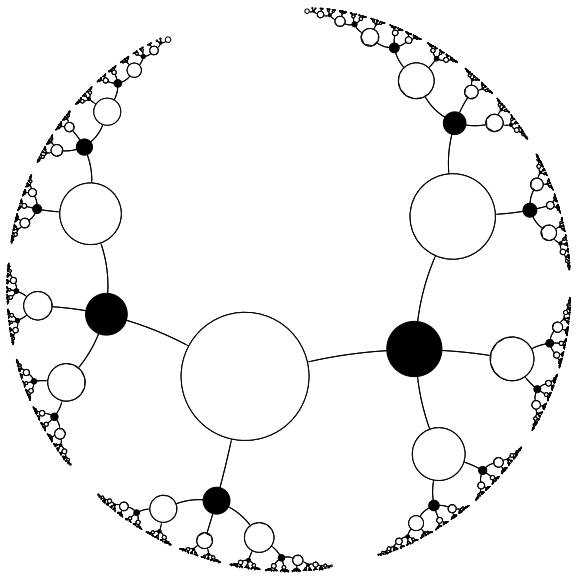
2: Iterative probabilistic decoding – from artificial intelligence. This is the better of the two methods.

Iterative probabilistic decoding

We attempt to solve the decoding problem

$$\mathbf{Hx} = \mathbf{z} \pmod 2$$

by a message-passing algorithm called *probability propagation*. The components of $\mathbf{x}$ and $\mathbf{z}$ can be thought of as nodes in a graph whose edges are defined by the 1s in $\mathbf{H}$.



White nodes represent bits, x_l;
black nodes represent checks, z_m;
each edge corresponds to a 1 in $\mathbf{H}$.

On each iteration, a probability ratio is propagated along each edge in the graph, and each node x_l updates its probability that it should actually be in state 1.

If the graph were *cycle-free* then this probability propagation algorithm would generate the correct answer.

It is not cycle-free, but the algorithm still performs extremely well.

THE ENCODER

We demonstrate a large code that encodes $K = 10000$ source bits into $N = 20000$ transmitted bits.

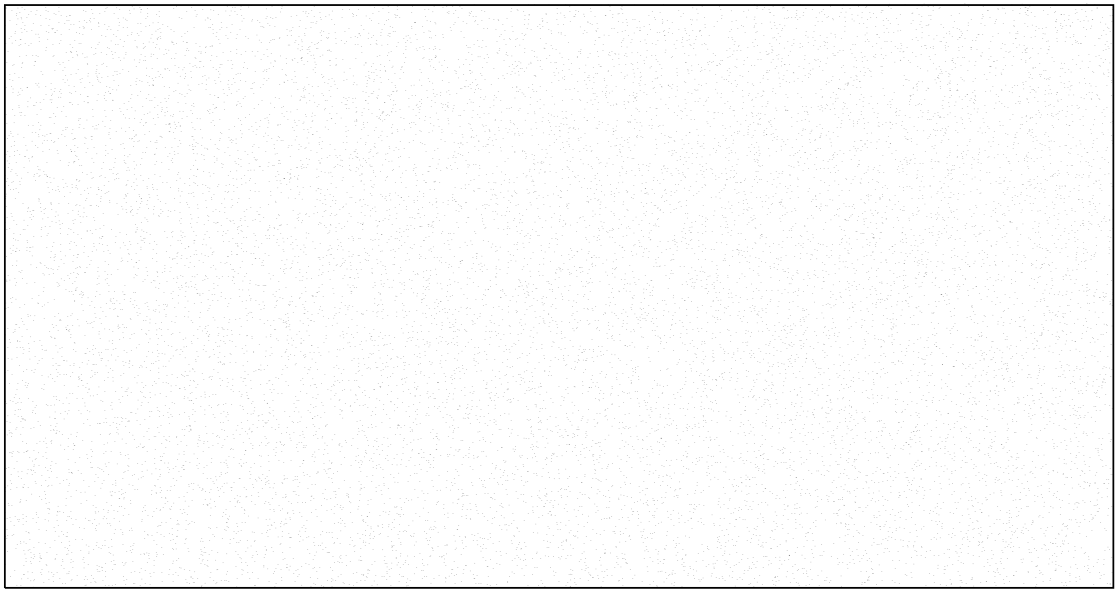
Each parity bit depends on about 5000 source bits.

The encoder is derived from a very sparse 10000×20000 matrix $\mathbf{H}$ with three 1s per column.

TRANSMITTED:

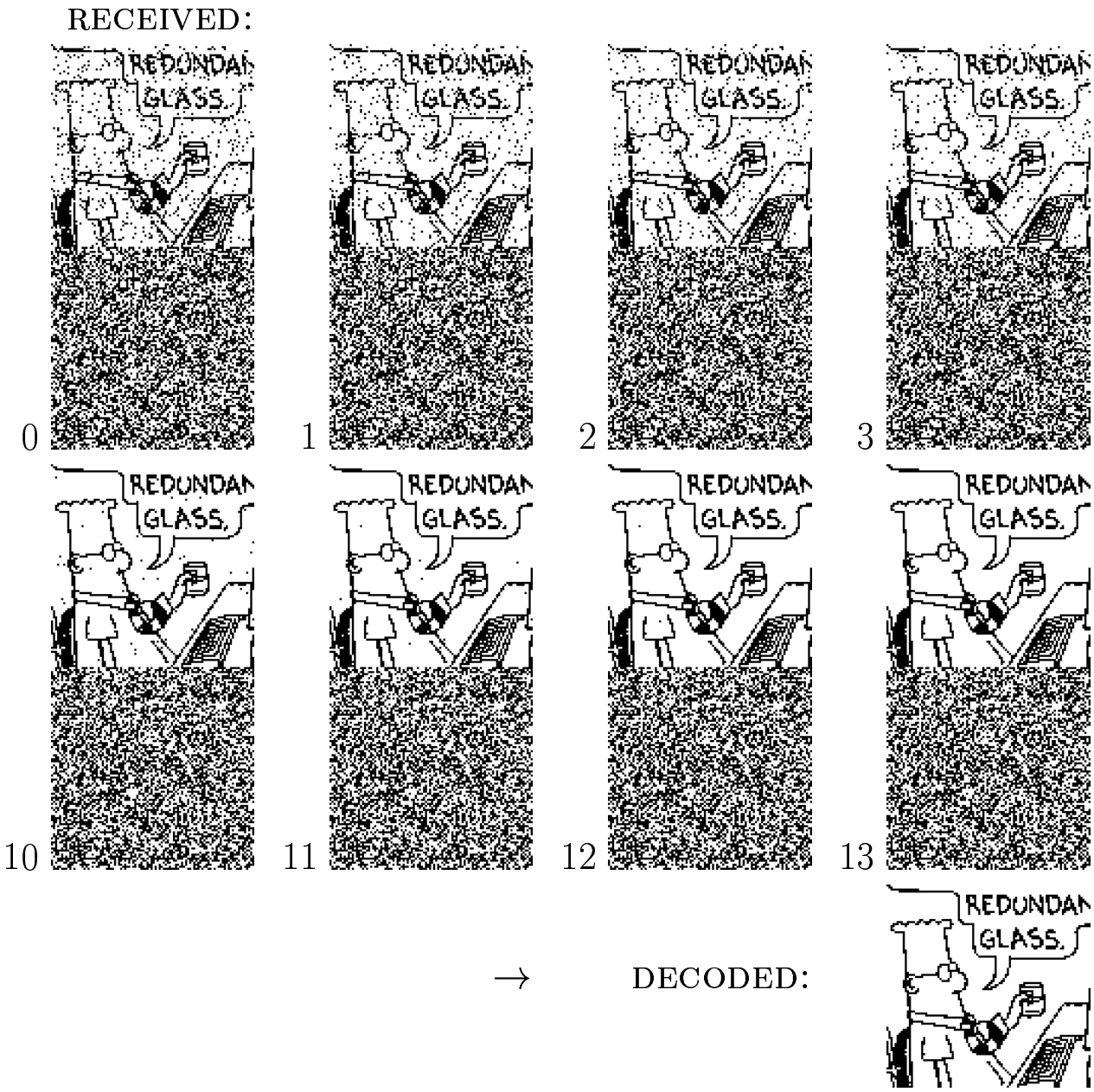
parity bits

$\mathbf{H} =$



Iterative decoding

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

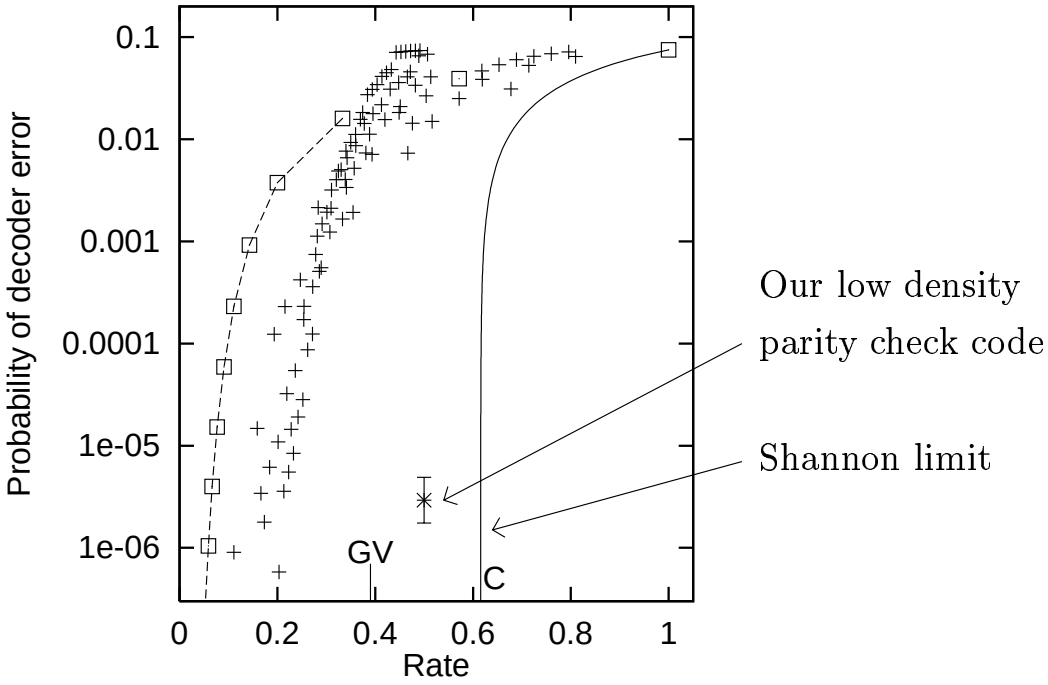


This final decoding is error free.

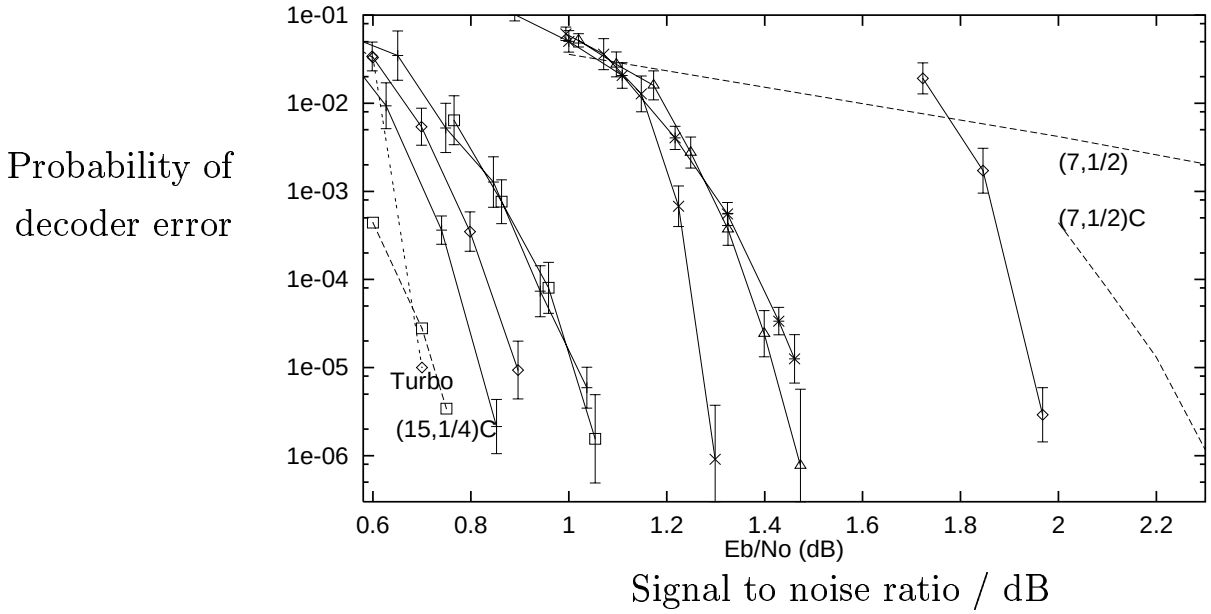
In the case of an unusually noisy transmission, the decoding algorithm fails to find a valid decoding. For this code and a channel with $f = 7.5\%$, such failures happen about once in every 100,000 transmissions.

These codes outperform textbook codes by a substantial margin.

Results for binary symmetric channel with $f = 7.5\%$



Results for Gaussian channel



Shannon limit is off the left side of the figure.

Solid lines: low density parity check codes.

Dotted lines (right): textbook codes as used in satellites and the Voyager spacecraft.

Dotted lines (left): state of the art codes: 'Turbo' codes and 'Galileo' code (15,1/4).

Improving Gallager Codes

– Clump bits and checks together

Also known as...

Generalize to other finite fields $GF(q)$

Our original work on low density parity check codes used ordinary binary arithmetic, known as '$GF(2)$'. The addition and multiplication tables for $GF(2)$ are:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

We can also define error-correcting codes using the addition and multiplication tables of other finite fields, for example $GF(4)$:

$+$	0	1	A	B
0	0	1	A	B
1	1	0	B	A
A	A	B	0	1
B	B	A	1	0

$\cdot$	0	1	A	B
0	0	0	0	0
1	0	1	A	B
A	0	A	B	1
B	0	B	1	A

We define low density parity check matrices using elements of $GF(4)$, and translate our binary messages into $GF(4)$ using, for example:

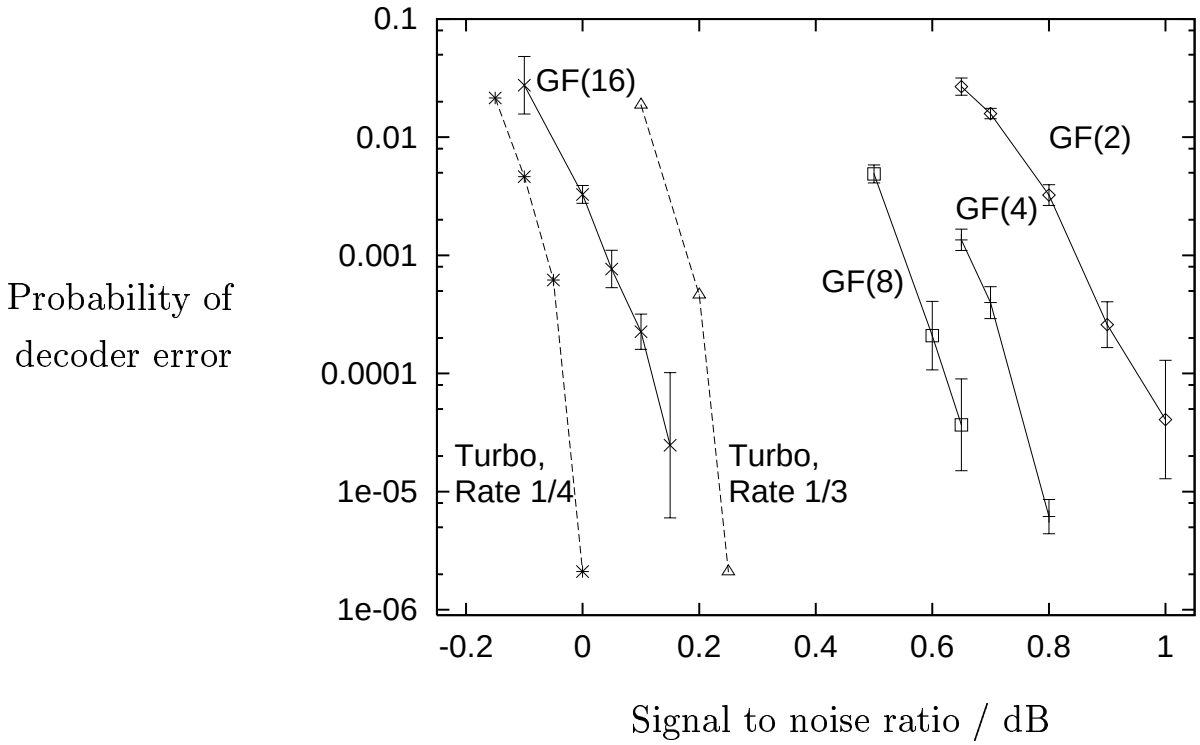
binary	$\leftrightarrow$	$GF(4)$
00	$\leftrightarrow$	0
01	$\leftrightarrow$	1
10	$\leftrightarrow$	A
11	$\leftrightarrow$	B

Results

The resulting codes over $GF(4)$, $GF(8)$, $GF(16)$, ..., when decoded with the iterative probabilistic decoder, perform nearly one decibel better.

The computational cost for working in $GF(q)$ scales as q^2.

Results for Gaussian channel



Shannon limit is off the left side of the figure.

Solid lines: low density parity check codes.

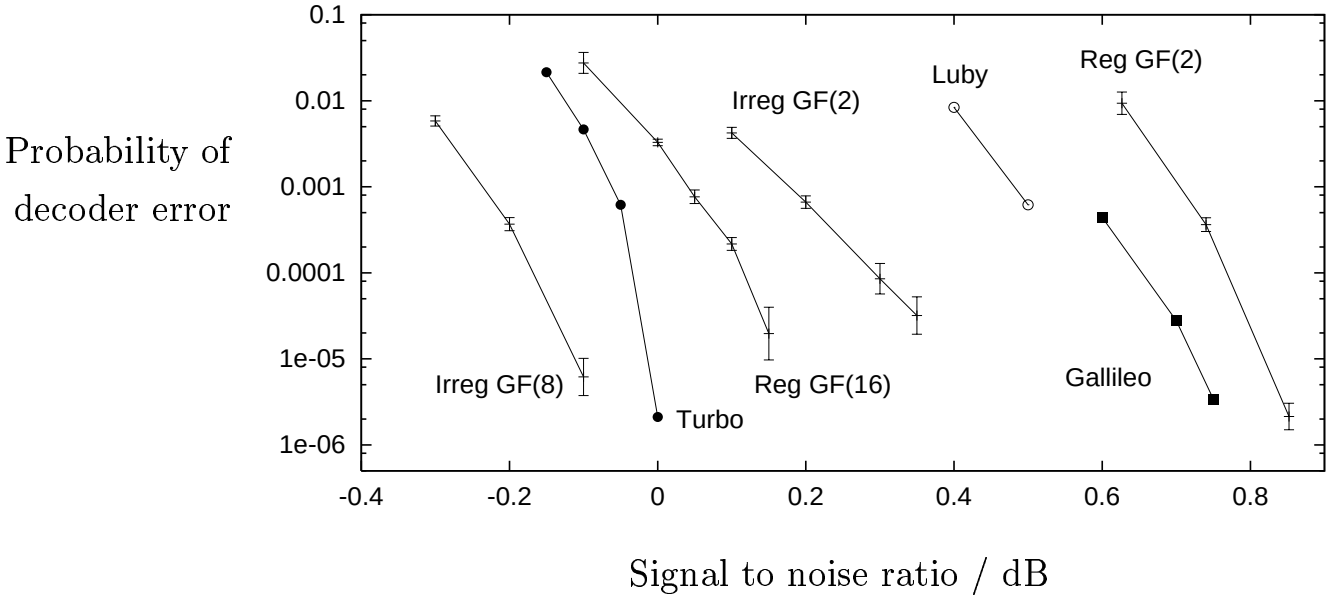
Dotted lines: JPL's latest 'Turbo' codes.

Improving Gallager Codes II

– Make the graph *irregular*

Luby *et al* showed that irregular constructions work better. The best graphs have irregular node connectivity and regular check connectivity.

Combining these two ideas (irregular graphs, and grouping nodes together), Matthew Davey produced the best known code of rate 0.25.



Conclusion

The state of the art solution to the communication problem is:

Combine a simple, **pseudo-random** code with
an approximate **probability-based** decoder.

http://wol.ra.phy.cam.ac.uk/mackay/

References

Berlekamp, E. R., McEliece, R. J., and van Tilborg, H. C. A. (1978) On the intractability of certain coding problems. *IEEE Transactions on Information Theory* **24** (3): 384–386.

Berrou, C., and Glavieux, A. (1996) Near optimum error correcting coding and decoding: Turbo-codes. *IEEE Transactions on Communications* **44**: 1261–1271.

Chung, S.-Y., Urbanke, R. L., and Richardson, T. J., (1999) LDPC code design applet. http://truth.mit.edu/~sychung/gaopt.html.

Davey, M. C., and MacKay, D. J. C. (1998) Low density parity check codes over GF(q). *IEEE Communications Letters* **2** (6): 165–167.

Davey, M. C., and MacKay, D. J. C. (2000) Watermark codes: Reliable communication over insertion/deletion channels. In *Proceedings 2000 IEEE International Symposium on Information Theory*, p. 477.

Dietterich, T., and Bakiri, G., (1991) Error-correcting output codes: A general method for improving multiclass inductive learning programs. In Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), pages 572–577. AAAI Press, 1991.

Divsalar, D., Jin, H., and McEliece, R. J. (1998) Coding theorems for 'turbo-like' codes. In *Proceedings of the 36th Allerton Conference on Communication, Control, and Computing, Sept. 1998*, pp. 201–210, Monticello, Illinois. Allerton House.

Forney, Jr., G. D. (1966) *Concatenated Codes*. Cambridge, Mass.: MIT Press.

Frey, B. J. (1998) *Graphical Models for Machine Learning and Digital Communication*. Cambridge MA.: MIT Press.

Gallager, R. G. (1962) Low density parity check codes. *IRE Trans. Info. Theory* **IT-8**: 21–28.

Gallager, R. G. (1963) *Low Density Parity Check Codes*. Number 21 in Research monograph series. Cambridge, Mass.: MIT Press.

Gallager, R. G. (1968) *Information Theory and Reliable Communication*. New York: Wiley.

Golomb, S. W., Peile, R. E., and Scholtz, R. A. (1994) *Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111*. New York: Plenum Press.

Luby, M. G., Mitzenmacher, M., Shokrollahi, M. A., and Spielman, D. A. (1998) Improved low-density parity-check codes using irregular graphs and belief propagation. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, p. 117.

MacKay, D. J. C. (1995) Free energy minimization algorithm for decoding and cryptanalysis. *Electronics Letters* **31** (6): 446–447.

MacKay, D. J. C., (1997) Iterative probabilistic decoding of low density parity check codes. Animations available on world wide web. http://wol.ra.phy.cam.ac.uk/mackay/codes/gifs/.

MacKay, D. J. C. (1999) Good error correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory* **45** (2): 399–431.

MacKay, D. J. C., and Davey, M. C. (2000) Evaluation of Gallager codes for short block length and high rate applications. In *Codes, Systems and Graphical Models*, ed. by B. Marcus and J. Rosenthal, volume 123 of *IMA Volumes in Mathematics and its Applications*, pp. 113–130. New York: Springer-Verlag.

MacKay, D. J. C., and Neal, R. M. (1995) Good codes based on very sparse matrices. In *Cryptography and Coding. 5th IMA Conference*, ed. by C. Boyd, number 1025 in Lecture Notes in Computer Science, pp. 100–111. Berlin: Springer.

MacKay, D. J. C., and Neal, R. M. (1996) Near Shannon limit performance of low density parity check codes. *Electronics Letters* **32** (18): 1645–1646. Reprinted *Electronics Letters*, **33**(6):457–458, March 1997.

MacKay, D. J. C., Wilson, S. T., and Davey, M. C. (1998) Comparison of constructions of irregular Gallager codes. In *Proceedings of the 36th Allerton Conference on Communication, Control, and Computing, Sept. 1998*, pp. 220–229, Monticello, Illinois. Allerton House.

McEliece, R. J. (1977) *The Theory of Information and Coding: A Mathematical Framework for Communication*. Reading, Mass.: Addison-Wesley.

McEliece, R. J., MacKay, D. J. C., and Cheng, J.-F. (1998) Turbo decoding as an instance of Pearl's 'belief propagation' algorithm. *IEEE Journal on Selected Areas in Communications* **16** (2): 140–152.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann.

Shannon, C. E. (1948) A mathematical theory of communication. *Bell Sys. Tech. J.* **27**: 379–423, 623–656.

Spielman, D. A. (1996) Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory* **42** (6.1): 1723–1731.

Swanson, L. (1988?) A new code for Galileo. In *Proc. 1988 IEEE International Symposium Information Theory*, pp. 94–95.

Tanner, R. M. (1981) A recursive approach to low complexity codes. *IEEE Transactions on Information Theory* **27** (5): 533–547.

Urbanke, R., Richardson, T., and Shokrollahi, A., (1999) Design of provably good low density parity check codes. Submitted.

Wiberg, N., (1996) *Codes and Decoding on General Graphs*. Dept. of Electrical Engineering, Linköping, Sweden dissertation. Linköping studies in Science and Technology. Dissertation No. 440.

Wiberg, N., Loeliger, H.-A., and Kötter, R. (1995) Codes and iterative decoding on general graphs. *European Transactions on Telecommunications* **6**: 513–525.