Error-correcting codes
Lecture notes - Chapters 1 & 47

- Cambridge University Press
- 640 pages, 35 pounds
- 28 pounds at CUP bookshop
- Also available free online

www.inference.phy.cam.ac.uk/mackay/itila/
Purpose: reliable communication over unreliable channels

eg, Binary symmetric channel

\[ f = 0.1 \]
Q: A file of $N = 10000$ bits is stored on this disc drive (with $f = 0.1$), then read. Roughly how many bits are flipped?

\[ \text{[ ]} \pm \text{[ ]} \]
Q: To make a successful business selling 1 Gigabyte disc drives, how small does the flip probability $f$ need to be?
Solutions:

- Physical solution

- System solution
System solution

Source

Encoder

Noisy channel

Decoder

s → t → r

s → ŝ

adds redundancy

does inference
Repetition code 'R3'

\[ f = 10\% \]
Repetition code 'R3'

\[ f = 10\% \]
Performance of repetition codes
### Block codes

#### (7,4) Hamming Code

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>0000 000</td>
</tr>
<tr>
<td>0001</td>
<td>0001 011</td>
</tr>
<tr>
<td>0010</td>
<td>0010 111</td>
</tr>
<tr>
<td>0011</td>
<td>0011 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0100</td>
<td>0100 110</td>
</tr>
<tr>
<td>0101</td>
<td>0101 101</td>
</tr>
<tr>
<td>0110</td>
<td>0110 001</td>
</tr>
<tr>
<td>0111</td>
<td>0111 010</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1000 101</td>
</tr>
<tr>
<td>1001</td>
<td>1001 110</td>
</tr>
<tr>
<td>1010</td>
<td>1010 010</td>
</tr>
<tr>
<td>1011</td>
<td>1011 001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>s</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100</td>
<td>1100 011</td>
</tr>
<tr>
<td>1101</td>
<td>1101 000</td>
</tr>
<tr>
<td>1110</td>
<td>1110 100</td>
</tr>
<tr>
<td>1111</td>
<td>1111 111</td>
</tr>
</tbody>
</table>

![Diagram of the Hamming code](http://emeagwali.com/)
(7,4) Hamming Code

\[ H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1
\end{bmatrix} \]

\[ M = 3 \]

\[ N = 7 \]

Valid transmissions \( t \) satisfy

\[ H t = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} \mod 2 \]
(7,4) Hamming Code

\[
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 \\
\end{bmatrix}
\]

\[M = 3\]  \[N = 7\]

Valid transmissions \(t\) satisfy

\[
Ht = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} \mod 2
\]

Received signal \(r = t + n\)

Syndrome \(z = Hr = Hn\).

Syndrome decoder \(z \rightarrow \hat{n}\).
What's achievable?
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 $\epsilon > 0$,

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

Non-constructive proof
If average weight of all babies is $< \epsilon$, there must be (at least!) one baby with weight $< \epsilon$. 
Shannon proved his noisy-channel theorem for any discrete memoryless channel
Channels with erasures

Binary erasure channel

8-ary erasure channel

Packet-erasing channel
There are other noisy channels...

Redundant Glass.

\[
\begin{array}{c}
0 \\ (1-f) \\ 1 \\
\end{array}
\quad \frac{(1-f)}{f} \quad \frac{1}{(1-f)} \\
\begin{array}{c}
0 \\
\end{array}
\]

Redundant Glass.

Insertions and Deletions

Synchronization errors
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)
Idea

Decoding problems, such as

\[ P_1(x) = \frac{1}{Z_1} e^{\beta \left[ x_1 x_2 x_3 x_5 + x_2 x_3 x_4 x_6 + x_1 x_3 x_4 x_7 \right] + \sum_{n=1}^{N} b_n x_n} \]

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

Solve the decoding problem

\[ \max_x P_1(x) \]

using variational methods?
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 noisy observations of \( As \) modulo 2, where \( A \) is a binary matrix. This binary vector is replaced by a vector of probabilities, optimised by free energy minimisation. Experiments on the inference of the state of a linear feedback shift register indicate that this algorithm supersedes the Meier and Staffelbach polynomial algorithm.
For which codes are approximate message-passing methods effective?
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

\[ H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \]

Valid transmissions \( t \) satisfy

\[ Ht = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mod 2 \]

Received signal \( r = t + n \)

Syndrome \( z = Hr = Hn \).

Syndrome decoder \( z \rightarrow \hat{n} \).
Low-density parity-check code

\[ \mathbf{H} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
\end{bmatrix} \]

Gallager 1962; MacKay & Neal 1995
Decoding by the sum-product algorithm

Probabilities

Likelihood ratios
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 $10000 \times 20000$ matrix $\mathbf{H}$ with three 1s per column.

TRANSMITTED:

parity bits

\[
\mathbf{H} = \begin{bmatrix} \end{bmatrix}
\]
Low Density Parity Check Code (f = 7.5%)  

Iterative probabilistic decoding

- low-density parity-check code
- Shannon limit

![Graph showing probability of decoder error vs rate with GV and C markers.](image)
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:**
Low Density Parity Check Code

Iterative probabilistic decoding
BSC: $f=7.5\%$

- low-density parity-check code
- Shannon limit

![Graph showing probability of decoder error vs. rate with annotations]

- GV
- C
Dependence on blocklength and column weight

(a)

(y-axis: 1, 0.1, 0.01, 0.001, 0.0001, 1e-05, 1e-06)

(x-axis: 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5)

N=816
N=408
N=96
(N=204)
N=204

(b)

(y-axis: 1, 0.1, 0.01, 0.001, 0.0001, 1e-05)

(x-axis: 1, 1.5, 2, 2.5, 3, 3.5, 4)

j=3
j=4
j=5
j=6
'Density evolution' on infinite graphs

Entropy versus iteration number

Identifies the **threshold** for sum-product decoding
Beyond simple LDPC codes - Gaussian channel results
Beyond simple LDPC codes - Gaussian channel results
Two codewords

Maximum tolerable noise-level for a bounded-distance decoder

Shannon's maximum noise-level
Convolutional codes

(a)   Octal name

(1, 353)\_8

(b)   (247, 371)\_8

(c)   (1, \(\frac{247}{371}\)) \_8
`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

\[
\binom{N}{Nf} \approx 2^{NH_2(f)}
\]

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

\[
\begin{aligned}
\text{TOTAL NUMBER OF TYPICAL OUTPUTS} & \quad \leq \\
2^K \times 2^{NH_2(f)} & \quad 2^N
\end{aligned}
\]

i.e.,

\[
K + NH_2(f) \leq N
\]

i.e.,

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