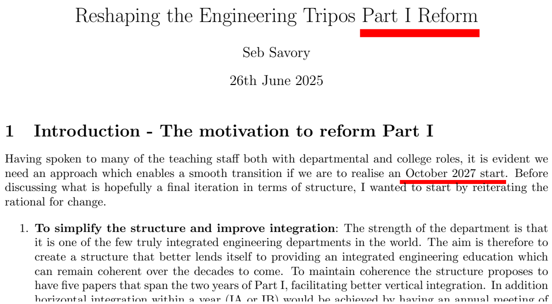
UNIVERSITY OF CAMBRIDGE
Department of Engineering
Probabilistic Systems, Information and Inference Lab

Shannon's Coding Theorem for a Simple Channel

Jossy Sayir, Ψ^2

Div F Conference, 7 October 2025

Reform of Undergraduate Engineering Tripos



Reshaping the Engineering Tripos Part I Reform

Seb Savory

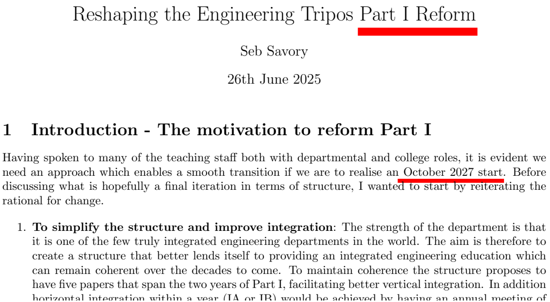
26th June 2025

1 Introduction - The motivation to reform Part I

Having spoken to many of the teaching staff both with departmental and college roles, it is evident we need an approach which enables a smooth transition if we are to realise an October 2027 start. Before discussing what is hopefully a final iteration in terms of structure, I wanted to start by reiterating the rational for change.

1. **To simplify the structure and improve integration**: The strength of the department is that it is one of the few truly integrated engineering departments in the world. The aim is therefore to create a structure that better lends itself to providing an integrated engineering education which can remain coherent over the decades to come. To maintain coherence the structure proposes to have five papers that span the two years of Part I, facilitating better vertical integration. In addition horizontal integration within a year (IA or IB) would be achieved by having an annual meeting of

Reform of Undergraduate Engineering Tripos



Reshaping the Engineering Tripos Part I Reform

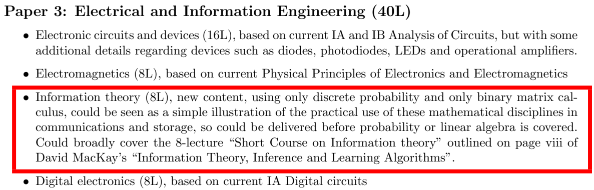
Seb Savory

26th June 2025

1 Introduction - The motivation to reform Part I

Having spoken to many of the teaching staff both with departmental and college roles, it is evident we need an approach which enables a smooth transition if we are to realise an October 2027 start. Before discussing what is hopefully a final iteration in terms of structure, I wanted to start by reiterating the rational for change.

1. **To simplify the structure and improve integration**: The strength of the department is that it is one of the few truly integrated engineering departments in the world. The aim is therefore to create a structure that better lends itself to providing an integrated engineering education which can remain coherent over the decades to come. To maintain coherence the structure proposes to have five papers that span the two years of Part I, facilitating better vertical integration. In addition horizontal integration within a year (IA or IB) would be achieved by having an annual meeting of



Paper 3: Electrical and Information Engineering (40L)

- Electronic circuits and devices (16L), based on current IA and IB Analysis of Circuits, but with some additional details regarding devices such as diodes, photodiodes, LEDs and operational amplifiers.
- Electromagnetics (8L), based on current Physical Principles of Electronics and Electromagnetics
- Information theory (8L), new content, using only discrete probability and only binary matrix calculus, could be seen as a simple illustration of the practical use of these mathematical disciplines in communications and storage, so could be delivered before probability or linear algebra is covered. Could broadly cover the 8-lecture “Short Course on Information theory” outlined on page viii of David MacKay’s “Information Theory, Inference and Learning Algorithms”.
- Digital electronics (8L), based on current IA Digital circuits

Elements of Information Theory

- Shannon's Source Coding Theorem
- Entropy is the lower bound for compressing a data source

- Shannon's Channel Coding Theorem
- Capacity is the supremum of information rates at which data can be transmitted with arbitrary reliability over a channel

Elements of Information Theory

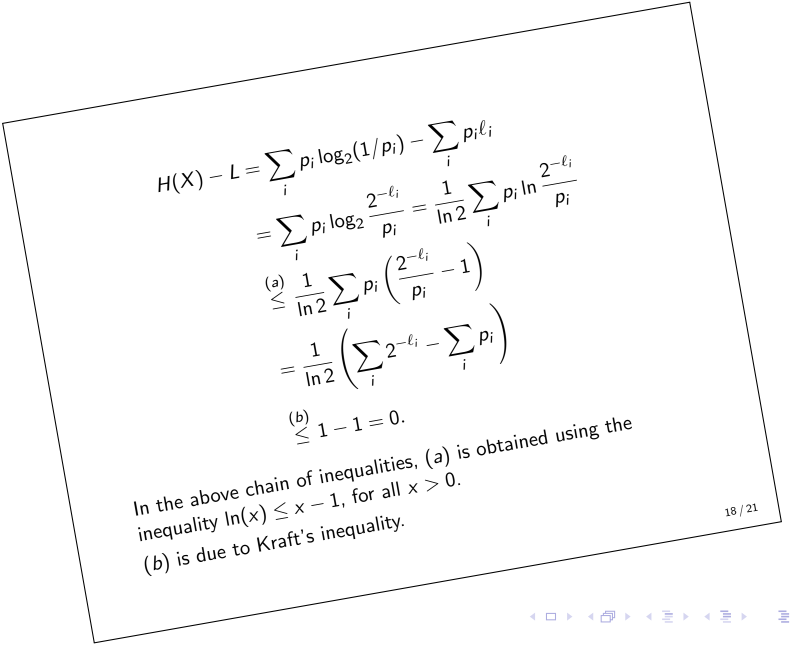
- Shannon's Source Coding Theorem
- Entropy is the lower bound for compressing a data source

- Shannon's Channel Coding Theorem
- Capacity is the supremum of information rates at which data can be transmitted with arbitrary reliability over a channel

Both of these and a few more things are currently taught in our 3rd year course

3F7: Information Theory and Coding

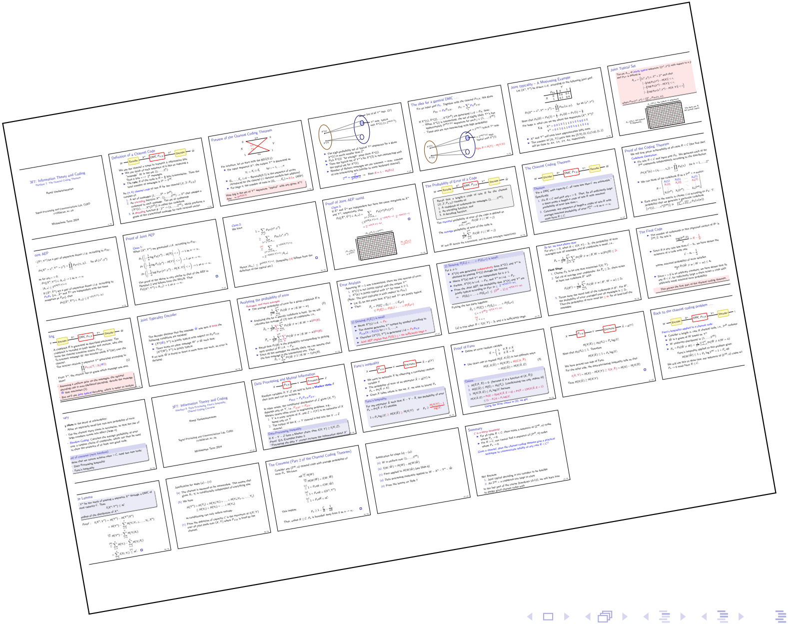
3F7 proof of the Source Coding Theorem



$$\begin{aligned} H(X) - L &= \sum_i p_i \log_2(1/p_i) - \sum_i p_i \ell_i \\ &= \sum_i p_i \log_2 \frac{2^{-\ell_i}}{p_i} = \frac{1}{\ln 2} \sum_i p_i \ln \frac{2^{-\ell_i}}{p_i} \\ &\overset{(a)}{\leq} \frac{1}{\ln 2} \sum_i p_i \left(\frac{2^{-\ell_i}}{p_i} - 1 \right) \\ &= \frac{1}{\ln 2} \left(\sum_i 2^{-\ell_i} - \sum_i p_i \right) \\ &\overset{(b)}{\leq} 1 - 1 = 0. \end{aligned}$$

In the above chain of inequalities, (a) is obtained using the inequality $\ln(x) \leq x - 1$, for all $x > 0$. (b) is due to Kraft's inequality.

3F7 proof of the Channel Coding Theorem



The issue with channel coding

- a general proof of the channel coding theorem is quite long
- it has a direct and a converse part
- it requires advanced concepts such as typicality, maximum likelihood decision, union and other bounds...

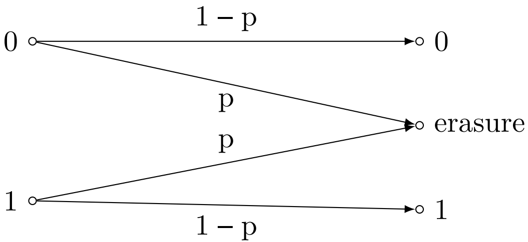
The issue with channel coding

- a general proof of the channel coding theorem is quite long
- it has a direct and a converse part
- it requires advanced concepts such as typicality, maximum likelihood decision, union and other bounds...

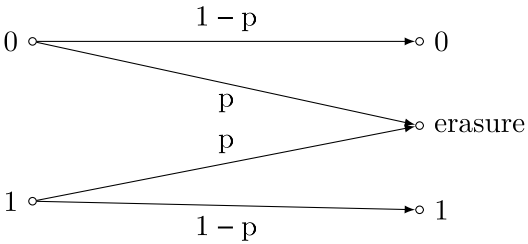
It's impractical for a 1st year course because:

- the maths are too advanced
- it's hard to get a solid intuition from the proof

The binary erasure channel

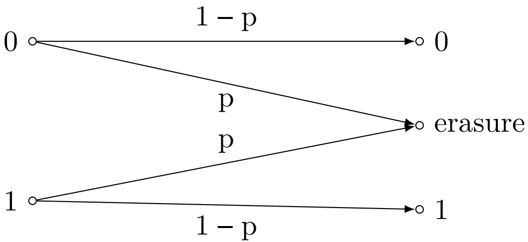


The binary erasure channel



- Capacity: $C = 1 - p$

The binary erasure channel



- Capacity: $C = 1 - p$
- For any rate $R = 1 - p - \varepsilon$ with $\varepsilon > 0$, arbitrary reliability can be achieved

Random linear coding

$$[\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_\mathrm{N}] = [\mathrm{u}_1, \ldots, \mathrm{u}_\mathrm{K}] \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- this is binary arithmetic (over GF(2)): $1 + 1 = 0$
- rate $\mathrm{R} = \mathrm{K}/\mathrm{N}$
- $\mathrm{K} < \mathrm{N}$, codeword longer than info, code adds redundancy
- code design: intelligent (e.g. Hamming code) or random coding: just pick a random binary matrix

Matrix inversion decoding

$$[0, ?, 1, ?, ?, ?, 1, 0] = [\mathrm{u}_1, \mathrm{u}_2, \mathrm{u}_3] \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Matrix inversion decoding

$$[0, ?, 1, ?, ?, ?, 1, 0] = [\mathrm{u}_1, \mathrm{u}_2, \mathrm{u}_3] \begin{bmatrix} 0 & \cancel{1} & 1 & \cancel{0} & \cancel{0} & \cancel{1} & 0 & 1 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 1 & 0 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 0 & 1 \end{bmatrix}$$

Matrix inversion decoding

$$[0, ?, 1, ?, ?, ?, 1, 0] = [u_1, u_2, u_3] \begin{bmatrix} 0 & \cancel{1} & 1 & \cancel{0} & \cancel{0} & \cancel{1} & 0 & 1 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 1 & 0 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 0 & 1 \end{bmatrix}$$

$$[0, 1, 1, 0] = [u_1, u_2, u_3] \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Matrix inversion decoding

$$[0, ?, 1, ?, ?, ?, 1, 0] = [u_1, u_2, u_3] \begin{bmatrix} 0 & \cancel{1} & 1 & \cancel{0} & \cancel{0} & \cancel{1} & 0 & 1 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 1 & 0 \\ 1 & \cancel{0} & 0 & \cancel{1} & \cancel{1} & \cancel{0} & 0 & 1 \end{bmatrix}$$

$$[0, 1, 1, 0] = [u_1, u_2, u_3] \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- there are L non-erased positions, where $L \sim B(N, 1 - p)$ is a random variable
- does the resulting $K \times L$ matrix have rank K?
- if yes: the information can be decoded
- if no: the information cannot be decoded for sure

Probability of success given L

- assume random coding, and L non-erased symbols
- probability of success = probability that a random $K \times L$ matrix has rank K
- if $L < K$, P(success)$= 0$
- if $L \geq K$, this is a counting problem
- there are $2^{K \times L}$ binary matrices. Which ones of them have full rank?

Counting full rank matrices

$$1 \times \mathrm{L} \quad \overbrace{\begin{bmatrix} - & - & - & \dots & - \end{bmatrix}}^{\mathrm{L}}$$

Counting full rank matrices

$$1 \times \mathrm{L} \qquad \overbrace{[\,- \quad - \quad - \quad \dots \quad -\,]}^{\mathrm{L}} \qquad 2^{\mathrm{L}} - 1$$

Counting full rank matrices

$$1 \times \mathrm{L} \qquad \overbrace{[- \quad - \quad - \quad \dots \quad -]}^{\mathrm{L}} \qquad 2^{\mathrm{L}} - 1$$

$$2 \times \mathrm{L} \qquad [- \quad - \quad - \quad \dots \quad -]$$

Counting full rank matrices

$$\begin{array}{ccc} 1\times \mathrm{L} & \overbrace{[\,-\quad -\quad -\quad \dots\quad -\,]}^{\mathrm{L}} & 2^{\mathrm{L}}-1 \\ 2\times \mathrm{L} & [\,-\quad -\quad -\quad \dots\quad -\,] & 2^{\mathrm{L}}-2 \end{array}$$

Counting full rank matrices

$$\begin{array}{lcl}
1 \times \mathrm{L} & \overbrace{[- \quad - \quad - \quad \dots \quad -]}^{\mathrm{L}} & 2^{\mathrm{L}} - 1 \\
2 \times \mathrm{L} & [- \quad - \quad - \quad \dots \quad -] & 2^{\mathrm{L}} - 2 \\
3 \times \mathrm{L} & [- \quad - \quad - \quad \dots \quad -] &
\end{array}$$

Counting full rank matrices

$$1 \times L \quad \overbrace{[\,-\quad -\quad -\quad \dots\quad -\,]}^{L} \quad 2^L - 1$$

$$2 \times L \quad [\,-\quad -\quad -\quad \dots\quad -\,] \quad 2^L - 2$$

$$3 \times L \quad [\,-\quad -\quad -\quad \dots\quad -\,] \quad 2^L - 4$$

Counting full rank matrices

$$\begin{array}{ccc}
1 \times \mathrm{L} & \overbrace{[- \quad - \quad - \quad \dots \quad -]}^{\mathrm{L}} & 2^{\mathrm{L}} - 1 \\
2 \times \mathrm{L} & [- \quad - \quad - \quad \dots \quad -] & 2^{\mathrm{L}} - 2 \\
3 \times \mathrm{L} & [- \quad - \quad - \quad \dots \quad -] & 2^{\mathrm{L}} - 4 \\
 & \vdots &
\end{array}$$

In conclusion

$$\#(\text{full rank } \mathrm{K} \times \mathrm{L} \text{ matrix}) = \prod_{\ell=0}^{\mathrm{K}-1} (2^{\mathrm{L}} - 2^{\ell})$$

Probability of full rank matrix

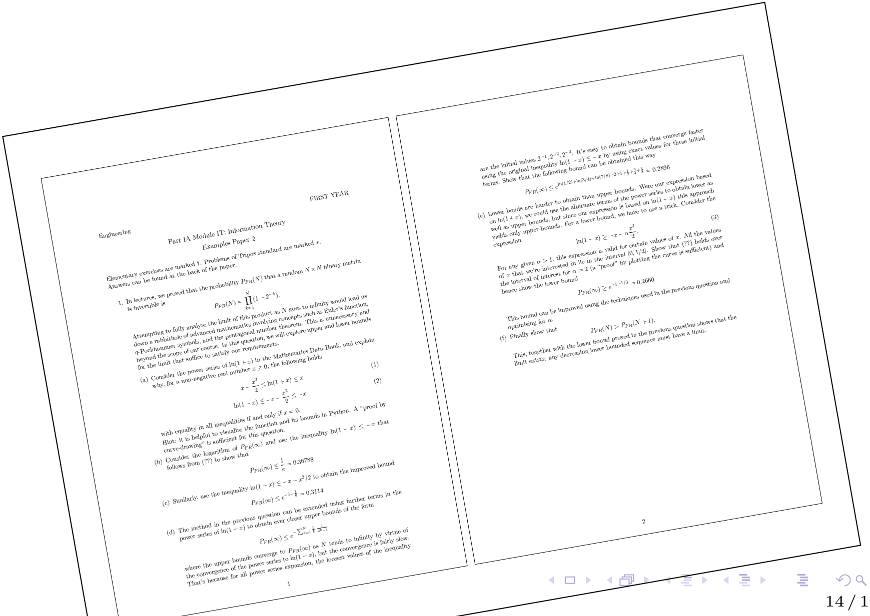
$$\begin{aligned}
\mathrm{P}(\text{full rank } \mathrm{K} \times \mathrm{L} \text{ random matrix}) &= \frac{\prod_{\ell=0}^{\mathrm{K}-1}(2^{\mathrm{L}} - 2^{\ell})}{2^{\mathrm{K}\times\mathrm{L}}} \\
&= \prod_{\ell=0}^{\mathrm{K}-1}(1 - 2^{\ell-\mathrm{L}}) \\
&= \prod_{\ell=\mathrm{L}-\mathrm{K}+1}^{\mathrm{L}}(1 - 2^{-\ell})
\end{aligned}$$

The magic number

If L = K,

$$\begin{aligned} \mathrm{P}(\text{full rank } \mathrm{K} \times \mathrm{K} \text{ random matrix}) &= \prod_{\ell=1}^{\mathrm{K}} (1 - 2^{-\ell}) \\ &= \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{7}{8} \cdot \frac{15}{16} \cdot \frac{31}{32} \cdots \\ &= 0.288788095086602\ldots \end{aligned}$$

Proving convergence and bounds on the magic number



FIRST YEAR

Engineering

Part IA Module IT: Information Theory

Examples Paper 2

Elementary exercises are marked †. Problems of Tripos standard are marked ∗. Answers can be found at the back of the paper.

1. In lectures, we proved that the probability $P_{FR}(N)$ that a random $N \times N$ binary matrix is invertible is

$$P_{FR}(N) = \prod_{k=1}^{N}(1-2^{-k}).$$

Attempting to fully analyse the limit of this product as N goes to infinity would lead us down a rabbithole of advanced mathematics involving concepts such as Euler's function, q-Pochhammer symbols, and the pentagonal number theorem. This is unnecessary and beyond the scope of our course. In this question, we will explore upper and lower bounds for the limit that suffice to satisfy our requirements.

(a) Consider the power series of $\ln(1+z)$ in the Mathematics Data Book, and explain why, for a non-negative real number $x \geq 0$, the following holds

$$x - \frac{x^2}{2} \leq \ln(1+x) \leq x \qquad (1)$$

$$\ln(1-x) \leq -x - \frac{x^2}{2} \leq -x \qquad (2)$$

with equality in all inequalities if and only if $x = 0$. Hint: it is helpful to visualise the function and its bounds in Python. A "proof by curve-drawing" is sufficient for this question.

(b) Consider the logarithm of $P_{FR}(\infty)$ and use the inequality $\ln(1-x) \leq -x$ that follows from (??) to show that

$$P_{FR}(\infty) \leq \frac{1}{e} = 0.36788$$

(c) Similarly, use the inequality $\ln(1-x) \leq -x - x^2/2$ to obtain the improved bound

$$P_{FR}(\infty) \leq e^{-1-\frac{1}{6}} = 0.3114$$

(d) The method in the previous question can be extended using further terms in the power series of $\ln(1-x)$ to obtain ever closer upper bounds of the form

$$P_{FR}(\infty) \leq e^{-\sum_{k=1}^{N}\frac{1}{k}\cdot\frac{1}{2^k-1}}$$

where the upper bounds converge to $P_{FR}(\infty)$ as N tends to infinity by virtue of the convergence of the power series to $\ln(1-x)$, but the convergence is fairly slow. That's because for all power series expansion, the loosest values of the inequality

are the initial values $2^{-1}, 2^{-2}, 2^{-3}$. It's easy to obtain bounds that converge faster using the original inequality $\ln(1-x) \leq -x$ by using exact values for these initial terms. Show that the following bound can be obtained this way

$$P_{FR}(\infty) \leq e^{\ln(1/2)+\ln(3/4)+\ln(7/8)-2+1+\frac{1}{2}+\frac{3}{4}+\frac{7}{8}} = 0.2896$$

(e) Lower bouds are harder to obtain than upper bounds. Were our expression based on $\ln(1+x)$, we could use the alternate terms of the power series to obtain lower as well as upper bounds, but since our expression is based on $\ln(1-x)$ this approach yields only upper bounds. For a lower bound, we have to use a trick. Consider the expression

$$\ln(1-x) \geq -x - \alpha\frac{x^2}{2}. \qquad (3)$$

For any given $\alpha > 1$, this expression is valid for certain values of x. All the values of x that we're interested in lie in the interval $[0, 1/2]$. Show that (??) holds over the interval of interest for $\alpha = 2$ (a "proof" by plotting the curve is sufficient) and hence show the lower bound

$$P_{FR}(\infty) \geq e^{-1-1/3} = 0.2660$$

This bound can be improved using the techniques used in the previous question and optimising for α.

(f) Finally show that

$$P_{FR}(N) > P_{FR}(N+1).$$

This, together with the lower bound proved in the previous question shows that the limit exists: any decreasing lower bounded sequence must have a limit.

The magic number

If $L > K$,

$$\begin{aligned} P &= \prod_{\ell=L-K+1}^{K} (1 - 2^{-\ell}) \\ &= 0.288788095086602 \cdot \frac{2}{1} \cdot \frac{4}{3} \cdots (1 - 2^{L-K})^{-1} \end{aligned}$$

- this goes to 1 exponentially as $L - K$ increases
- for a given target reliability ε, there exists an L for which

$$P(\text{full rank } K \times L \text{ random matrix}) > 1 - \varepsilon$$

for all $N > L$.