Consider the noisy typewriter channel, mapping

$$\mathcal{A}_Z = \{\texttt{A}, \texttt{B}, \texttt{C}, \ldots, \texttt{Y}, \texttt{Z}, -\}$$

with $|\mathcal{A}_Z| = 27$, into $\mathcal{A}_Y = \mathcal{A}_Z$, where each letter is mapped with equal probabilities into the preceding, the following or the same letter (p. 41-44 of the notes). Design an efficient code by which to *reliably* send symbols from $A_X = A_Z$ through the channel (i.e., you should be able to send and retrieve a text using the 27 symbols with no error).

Solution: We can base an efficient and reliable code by using a subset of non-confusable inputs, e.g.,

$$|S| = \{\texttt{B}, \texttt{E}, \texttt{H}, \texttt{K}, \texttt{N}, \texttt{Q}, \texttt{T}, \texttt{W}, \texttt{Z}\}$$

with $|S| = 9$. Since there are 81 sequences of 2 symbols from $|S|$,

$$|S^{(2)}| = 81$$

and $81 > 27$, we can just select *any* subset of 27 pairs of symbols from S, i.e.,

$$S' \subset S^{(2)}$$

with $|S'| = 27$, to reliably encode the input. For instance, we can divide the 27 letters into 3 groups of 9 letters,

$$g_1 = \{\texttt{A}, \texttt{B}, \ldots, \texttt{I}\}, \quad g_2 = \{\texttt{J}, \texttt{K}, \ldots, \texttt{R}\}, \quad g_3 = \{\texttt{S}, \texttt{T}, \ldots, -\}$$

and use the first letter of the code to encode for the group, e.g.,

$$g_1 \mapsto \texttt{B}, \quad g_2 \mapsto \texttt{E}, \quad g_3 \mapsto \texttt{H}$$

and the second letter of the code to encode for the specific letter within the group,

$$\{\texttt{A}, \texttt{B}, \ldots, \texttt{I}\} \mapsto \{\texttt{B}, \texttt{E}, \ldots, \texttt{Z}\}$$

$$\{\texttt{J}, \texttt{K}, \ldots, \texttt{R}\} \mapsto \{\texttt{B}, \texttt{E}, \ldots, \texttt{Z}\}$$

$$\{\texttt{S}, \texttt{T}, \ldots, -\} \mapsto \{\texttt{B}, \texttt{E}, \ldots, \texttt{Z}\}$$

arriving at the two-letter code $E : A_X \to S^{(2)}$:

A$\mapsto$BB, B$\mapsto$BE, C$\mapsto$BH, D$\mapsto$BK, E$\mapsto$BN, F$\mapsto$BQ, G$\mapsto$BT, H$\mapsto$BW, I$\mapsto$BZ,

J$\mapsto$EB, K$\mapsto$EE, L$\mapsto$EH, M$\mapsto$EK, N$\mapsto$EN, O$\mapsto$EQ, P$\mapsto$ET, Q$\mapsto$EW, R$\mapsto$EZ,

S$\mapsto$HB, T$\mapsto$HE, U$\mapsto$HH, V$\mapsto$HK, W$\mapsto$HN, X$\mapsto$HQ, Y$\mapsto$HT, Z$\mapsto$HW, $-\mapsto$HZ,

The average length of this code is 2, irrespective of the input distribution. The rate is

$$R = \frac{K}{2} = \frac{\log 27}{2} = \frac{3 \log 3}{2}$$

We know that the channel capacity is

$$C = \max_{Z} I[Z:Y] = 2\log 3 > R = \frac{3}{2}\log 3$$

which is more than the channel capacity. When using the channel at itys capacity, we should use 1.5 letters to encode a single letter (not 2). We know that we can achieve the capacity by encoding sequences of messages $\mathcal{A}_X^{(m)}$ into non-confusable sequences of inputs strings $\mathcal{A}_Z^{(n)}$, and in the limit $n \to \infty$, we should communicate information at the channel capacity. In fact, for this simple case it is sufficient to consider $m = 2$, $n = 3$. For $n = 3$, we can consider the set of sequences of 3 non-confusable input symbols,

$$z^{(3)} \in S^{(3)}$$

with $|S^{(3)}| = 9^3 = 729$. These sequences are sufficient to reliably encode $27^2 = 729$ symbols $x^{(2)} \in \mathcal{A}_X^{(2)}$. In other words, we can map a *pair* of letters $X^{(2)} \in \mathcal{A}_X$ from the 27-letter alphabet into a *triple* of letters from $S^{(3)}$, e.g.

AA$\mapsto$BBB, AB$\mapsto$BBE, AC$\mapsto$BBH, ... $-$U $\mapsto$ ZZT, $-$Z $\mapsto$ ZZW, $--$ $\mapsto$ ZZZ

This is (n, K) block code with $n = 3$ and $K = \log 729 = 6\log 3$. Hence the rate is

$$R = \frac{K}{n} = 2\log 3 = C$$

achieving the channel capacity. In this code, we use only 1.5 letters to encode a letter.