Compression

In compression we map source symbols to labels, the latter are often binary. In this question we will look at the vanila version, where each source symbol is mapped to a set of binary values, before proceeding to the next symbol. Suppose the source symbols take values $\{a, b, c, d\}$ and we map them to the binary labels $\{1, 10, 100, 1000\}$ (in the order given).

(a) For the sequence *abbacaddabbcaabbacadabbca*, write the encoded sequence.

(b) Suggest a source probability distribution for which a compression of 1.5 bits per source symbol (on average) can be achieved, but all symbols should have positive probabilities. Is there a distribution such that 1 bit per source symbol (on average) suffices to represent the given IID sequence.

(c) Suppose d never occurs, and the source $\{a, b, c\}$ is mapped to $\{0, 01, 11\}$. Do you think we can find the source sequence from the encoded bits.

Error Correction-I

A length 7 binary vector is given, which is known to contain at most one non-zero value. Suppose you wish to build a logic circuit to display the position of the non-zero entry using a seven segment BCD display.
(a) Suppose 3 XOR gates, each taking 4 binary inputs are available to you. Explain the feasibility of a design.
(b) How about length 15 binary vector with at most one non-zero element, but you are given four $8-$input XOR gates. (Note: display is HEX now)

Error Correction-II

Consider a binary $m \times n$ matrix H with $m < n$. A set $\mathcal{C}$, which is the collection of all binary vectors $\bar{x}$ of length n such that $H\bar{x} = 0$ is called a *linear code*. While communicating, each different value/voltage-level/messagei/symbol will be mapped to separate code sequence in $\mathcal{C}$. Once it reaches the receiver, it can figure out the message by the reverse map, provided too many bits in the codeword transmitted is not flipped by noise and other effects.

(a) Show that $\mathcal{C}$ is closed under XOR (addition), i.e. if $\bar{x} \in \mathcal{C}$ and $\bar{y} \in \mathcal{C}$, then $\bar{x} \oplus \bar{y} \in \mathcal{C}$.

(b) Let the columns of H are given by the 3 bit binary representation of the numbers $(1, 2, 3, 4, 5, 6, 7)$. Thus H is a 7×3 binary matrix. For $\mathcal{C} = \{\bar{x} : H\bar{x} = 0\}$, what is the minimum number of bits at which two distinct code sequences differ.

(c) Use the above answer to find how many errors we can tolerate while transmitting a codeword. Any suggestion to correct the errors?