3

Hamming codes

3.1 BINARY REPRESENTATION OF NUMBERS

Recall that our ordinary 'decimal' number system is the denary number system which means that the number system uses 10 as base. In this system we know that the number 947 actually equals: $9 \times 10^2 + 4 \times 10^1 + 7$. Similarly

$$\begin{aligned} 3\,450\,671 &= 3 \times 10^6 + 4 \times 10^5 + 5 \times 10^4 + 0 \times 10^3 + 6 \times 10^2 + 7 \times 10 + 1 \\ &= 3 \times 10^6 + 4 \times 10^5 + 5 \times 10^4 + 6 \times 10^2 + 7 \times 10 + 1 \end{aligned}$$

We can in fact define a number system with any number r, $1 \leq r \leq 9$, as base. In such a number system each place will be occupied by one of the numbers, $0, 1, \ldots, r-1$. In view of this, the number system with base $r = 1$ is not of any interest. With the coming into existence of computers the number system using 2 as base has gained considerable importance. Such a system of numbers is called a **binary number system** or a binary representation of numbers. Observe that any place in this representation of the numbers is occupied by either 0 or 1. A number *abcde* in this system represents the number

$$a \times 2^4 + b \times 2^3 + c \times 2^2 + d \times 2 + e$$

For example

$$101101 = 1 \times 2^5 + 1 \times 2^3 + 1 \times 2^2 + 1$$

i.e. the number

$$32 + 8 + 4 + 1 = 45$$

in the denary system. In this way, it is quite easy to give denary representation of a number, binary representation of which is given. The reverse process is also equally simple and given any number in the denary system, we can get its binary representation. We explain this procedure through an example.

Example 3.1
Consider the number 67. Observe that

$$67 = 64 + 2 + 1 = 2^6 + 2^1 + 1$$

and so binary representation of the number 67 is 1000011. Similarly

$$43 = 32 + 11 = 32 + 8 + 2 + 1 = 2^5 + 2^3 + 2^1 + 1$$

and its binary representation is 101011.

Next consider the number 243. We have

$$\begin{aligned} 243 &= 128 + 115 \\ &= 128 + 64 + 51 \\ &= 128 + 64 + 32 + 19 \\ &= 128 + 64 + 32 + 16 + 2 + 1 \\ &= 2^7 + 2^6 + 2^5 + 2^4 + 2^1 + 1 \end{aligned}$$

and so the binary representation of 243 is 11110011. Observe that putting one or more zeros on the left of the binary representation of a number does not alter the number. Also, observe that a number which is at most 2^m can be represented by a string of 0s and 1s of length $m + 1$, and a string of 0s and 1s of length $m + 1$ will always represent a number at most 2^m. Moreover, a number has only one 1 (and the rest 0s) in its binary representation iff it is a power of 2.

We next describe a recursion algorithm for converting a number into a binary number.

Let a number n be given.

1. Divide by 2 with r_0 as remainder and n_1 the quotient.
2. Divide n_1 by 2. Let r_1 be the remainder and n_2 the quotient.

Continue the process until the quotient becomes 0. Let the remainders obtained in succession be $r_0, r_1, \ldots, r_k$. Then, the binary representation of the number n is $r_k r_{k-1} \cdots r_1 r_0$. We now illustrate this process with an example.

Example 3.2
Consider the number 483.

$$\begin{array}{rl} 2\underline{)483} & \\ 2\underline{)241} & r_0 = 1 \\ 2\underline{)120} & r_1 = 1 \\ 2\underline{)60} & r_2 = 0 \\ 2\underline{)30} & r_3 = 0 \\ 2\underline{)15} & r_4 = 0 \\ 2\underline{)7} & r_5 = 1 \\ 2\underline{)3} & r_6 = 1 \\ 2\underline{)1} & r_7 = 1 \\ 0 & r_8 = 1 \end{array}$$

Hence, in the binary system, 483 is represented by 111100011.

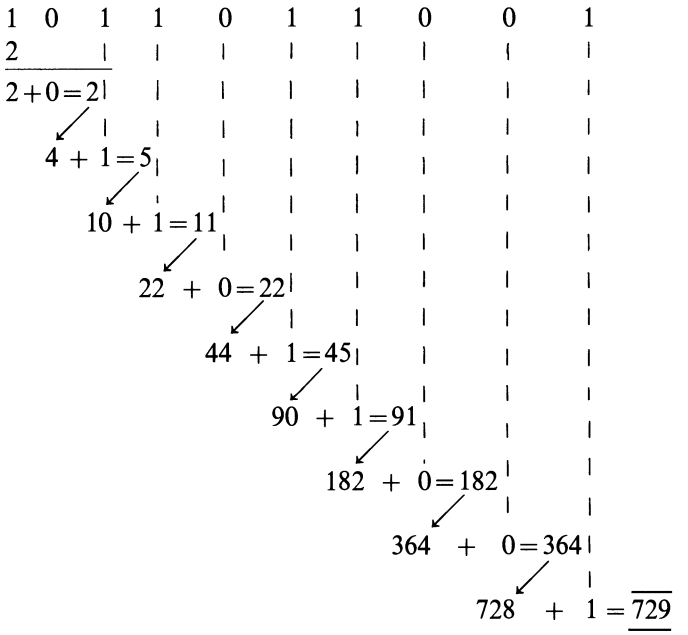
The algorithm for the conversion of a number from binary system to denary system is equally simple. Let a number n be represented in the binary system as $r_k r_{k-1} \cdots r_1 r_0$. To convert it into the decimal system:

1. Multiply r_k by 2 and add r_{k-1} to obtain $2r_k + r_{k+1}$.
2. Multiply $2r_k + r_{k-1}$ by 2 and add r_{k-2} to obtain

$$2(2r_k + r_{k-1}') + r_{k-2} = 2^2 r_k + 2r_{k-1} + r_{k-2}$$

Continue this process to exhaust all the r_i's. The result so obtained is the number n in the decimal representation. The coefficient of r_k will be 2^k and so on.

Example 3.3
Consider a binary number 1011011001. We then obtain the decimal representation of this number as follows.



The arrow at each step indicates multiplication of the number above it by 2.

3.2 HAMMING CODES

Hamming codes are single error, correcting codes. Given any positive integer r, we can always construct an $(m = 2^r - r - 1, n = 2^r - 1)$ code which corrects each single error that might occur and corrects no other errors. These codes are defined by the following procedure.

Procedure for forming a Hamming code

Step 1

Choose a positive integer r. Then the code words of the code will have $n = 2^r - 1$ digits and the message words will have $m = 2^r - r - 1$ digits. Thus, in each code word, there are r check digits. The number of check symbols present in any code word of a block code is called the **redundancy** of the code. So, the code being constructed has redundancy r.

Step 2

In each code word $b = (b_1, b_2, \ldots, b_n)$ use $b_{2^0}, b_{2^1}, \ldots, b_{2^{r-1}}$ as the r check digits and place the $2^r - r - 1$ message digits in the remaining b_j positions but in their original order. For example if we take $r = 3$, then in the code word $b = (b_1, \ldots, b_7)$ corresponding to the message word $a = (a_1, a_2, a_3, a_4)$, b_1, b_2, b_4 are the check symbols and $b_3 = a_1, b_5 = a_2, b_6 = a_3, b_7 = a_4$. Again, if we take $r = 4$, then in the code word $b = (b_1, \ldots, b_{15})$ corresponding to the message word $a = (a_1, \ldots, a_{11})$ b_1, b_2, b_4, b_8 are the check symbols and $b_3 = a_1, b_5 = a_2, b_6 = a_3, b_7 = a_4$, $b_9 = a_5$, $b_{10} = a_6$, $b_{11} = a_7$, $b_{12} = a_8$, $b_{13} = a_9$, $b_{14} = a_{10}$, $b_{15} = a_{11}$.

Step 3

Form a matrix $\mathbf{M}$ of $2^r - 1$ rows and r columns in which the ith row is the binary representation of the number i. For example:

(a) if $r = 2$ then $\mathbf{M}$ is the 3×2 matrix with

$$\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

(b) if $r = 3$, then $\mathbf{M}$ is the 7×3 matrix with

$$\mathbf{M}' = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

(c) if $r = 4$, then $\mathbf{M}$ is the 15×4 matrix with

$$\mathbf{M}' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Step 4

Form the matrix equation $\mathbf{bM} = 0$ which gives r linear equations in the r unknowns $b_1, b_2, \ldots, b_{2^{r-1}}$. For example:

(a) if $r = 2$, then

$$\mathbf{bM} = 0 \Rightarrow b_2 + b_3 = 0 \quad \text{and} \quad b_1 + b_3 = 0$$

(b) if $r = 3$, then

$$\mathbf{bM} = 0 \Rightarrow b_4 + b_5 + b_6 + b_7 = 0 \quad b_2 + b_3 + b_6 + b_7 = 0$$

and

$$b_1 + b_3 + b_5 + b_7 = 0$$

(c) while if $r = 4$, then

$$\mathbf{bM} = 0 \Rightarrow b_8 + b_9 + b_{10} + b_{11} + b_{12} + b_{13} + b_{14} + b_{15} = 0$$

$$b_4 + b_5 + b_6 + b_7 + b_{12} + b_{13} + b_{14} + b_{15} = 0$$

$$b_2 + b_3 + b_6 + b_7 + b_{10} + b_{11} + b_{14} + b_{15} = 0$$

$$b_1 + b_3 + b_5 + b_7 + b_9 + b_{11} + b_{13} + b_{15} = 0$$

Observe that, in each of the above examples, the r linear equations are such that every one of them contains exactly one unknown variable. We shall prove this in the general case. In view of this, the r linear equations give a unique solution for the r unknown quantities so that the code word is uniquely determined by the given message word.

Step 5

To encode a message word, place the message digits in the correct b_j positions and then make the check digits $b_{2^i}, 0 \leq i \leq r-1$ satisfy the r linear equations as in Step 4 above. There will be exactly one b_{2^i} in each equation and so the r equations lead to a unique solution for $b_{2^0}, b_{2^1}, \ldots, b_{2^{r-1}}$.

Lemma 3.1

In each of the r linear equations as in Step 4 above there is exactly one b_{2^i} present.

Proof

Suppose that b_{2^i} occurs in the equation which is obtained by multiplying $\mathbf{b} = (b_1 \cdots b_n)$ with the kth column of the matrix $\mathbf{M}$. Then the 2^ith entry in the kth column of $\mathbf{M}$ is 1, i.e. the $(2^i, k)$th entry of $\mathbf{M}$ is 1. This entry is in the 2^ith row of $\mathbf{M}$ and 2^ith row of $\mathbf{M}$ is the binary representation of the number 2^i which has only the $(i+1)$th entry from the right equal to 1. Hence this entry occurs in the $(r-i)$th column and so $k = r - i$. Thus if b_{2^i} and b_{2^j} both occur in the equation which is obtained by multiplying $\mathbf{b}$ with the kth column of $\mathbf{M}$, then we have $k = r - i$ and $k = r - j$. Hence $i = j$ and at most one check symbol b_{2^i} occurs in any equation.

On the other hand, it is clear from the above reasoning that a given check symbol b_{2^i} occurs in the equation obtained by multiplying $\mathbf{b}$ with the $(r-i)$th column of $\mathbf{M}$.

The code as obtained by the constructive procedure (Steps 1–5 above) is called the $(2^r - r - 1, 2^r - 1)$ Hamming code.

Proposition 3.1
Hamming code is a group code.

Proof
Let $b = b_1 b_2 \cdots b_n$ be the code word corresponding to the message word $a = a_1 a_2 \cdots a_m$ and $b' = b'_1 b'_2 \cdots b'_n$ be the code word corresponding to the message word $a' = a'_1 a'_2 \cdots a'_m$. Let $\mathbf{M}$ be the $(2^r - 1) \times r$ matrix in which the ith row is the binary representation of the number i. Then $\mathbf{bM} = 0 = \mathbf{b'M}$ and, therefore, $(\mathbf{b} + \mathbf{b'})\mathbf{M} = 0$. Also the entries in $b + b'$ at the positions other than the $2^0, 2^1, \ldots, 2^{r-1}$ positions are $a_i + a'_i$ in the order in which they occur in $a + a'$. Also b_{2^i} is given by the equation when $\mathbf{b}$ is multiplied by the $(r - i)$th column of $\mathbf{M}$ and similarly for b'_{2^i}. Thus $b + b'$ is the code word corresponding to the message word $a + a'$ and $a \to b$ where b is the code word in the Hamming code corresponding to the message word a is a group homomorphism $\mathbb{B}^m \to \mathbb{B}^n$. Hence the Hamming code is a group code.

Theorem 3.1
The minimum distance of any Hamming code is 3.

Proof
Since Hamming code is a group code, it is enough to prove that the weight of any non-zero code word is at least 3 and there is at least one code word of weight exactly 3. Again, all the message symbols of a message word occur somewhere in the corresponding code word. Therefore for the first assertion, it is enough to prove that the weight of a non-zero code word which corresponds to a message word of weight at most 2 has weight at least 3.

Case (i)
Let a be a message word with $\mathrm{wt}(a) = 1$ and $b = b_1 b_2 \cdots b_n$ be the corresponding code word. Suppose that the non-zero message symbol occurs in the ith position in b. Then $i \neq 2^j$ for any j and in the binary representation of i there are at least two non-zero entries. Let these be in the sth and tth position from the left.

Let $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_r$ denote the columns of the $(2^r - 1) \times r$ matrix $\mathbf{M}$ in which jth row is the binary representation of j. Consider the equations $\mathbf{bM}_s = 0$ and $\mathbf{bM}_t = 0$ out of the r linear equations $\mathbf{bM} = 0$. In the equation $\mathbf{bM}_s = 0$, there is exactly one check symbol and every other symbol is a message symbol. Let $b_2 k$ be the check symbol present in this equation. Every message symbol except b_i being 0, we have $b_2 k + b_i = 0$ (by the choice of s) and so $b_2 k \neq 0$. Similarly, we find that the check symbol occurring in the equation $\mathbf{bM}_t = 0$ is non-zero. Every one of the r linear equations $\mathbf{bM} = 0$ contains only

one check symbol and check symbols in different equations are different. Thus, the two non-zero check symbols in the equations $\mathbf{bM}_s = 0$ and $\mathbf{bM}_t = 0$ are different. Therefore $\text{wt}(b) \geq 3$.

Case (ii)
Let a be a message word with $\text{wt}(a) = 2$ and let $b = b_1 b_2 \cdots b_n$ be the corresponding code word. Suppose that the two non-zero message symbols occur in the ith and jth positions. Then i and j are not powers of 2. Again $i \neq j$ implies that binary representations of i and j differ in at least one place. Suppose that these differ from each other in the sth position from the left. We may, therefore, suppose that the sth position (from the left) of i is 1 and that of j is 0. (We can reverse the roles of i and j otherwise.) Let b_{2^k} be the check symbol which appears in the equation $\mathbf{bM}_s = 0$. Since every other symbol in this equation is a message symbol and

$$b_\ell = 0 \forall \ell \text{ except } \ell = i \text{ and } \ell = j$$

and the ith entry in $\mathbf{M}_s$ is 1 while the jth entry in $\mathbf{M}_s$ is 0, this equation reduces to $b_{2^k} + b_i = 0$, i.e. $b_{2^k} \neq 0$. Thus, in b, there is at least one check symbol which is non-zero and $\text{wt}(b) \geq 3$.

Let $b = b_1 b_2 \cdots b_n$ be the code word corresponding to the message word $a = a_1 a_2 \cdots a_m$, where $a_1 = 1$ and $a_i = 0$ for $2 \leq i \leq m$. In the equation $\mathbf{bM}_s = 0$, the check symbol takes the value 1 iff the third entry in the column $\mathbf{M}_s$ is 1. Thus the number of non-zero check symbols in b equals the number of non-zero entries in the third row of $\mathbf{M}$. But the third row of $\mathbf{M}$ has exactly two non-zero entries. Hence, the number of non-zero check symbols in b is 2 and, so, $\text{wt}(b) = 3$. This proves that the minimum distance of the Hamming code is 3.

Corollary
Hamming code is a single error correcting and double error detecting code.

Proof
This follows from the above theorem and Theorems 1.1 and 1.3.

Remark
A Hamming code being a group code, it follows that error vectors that go undetected are precisely those which equal some code word.

Exercise 3.1
Compute all the code words of the (4, 7) Hamming code.

Remarks
The $(2^r - 1) \times r$ matrix $\mathbf{M}$ in the construction of Hamming code is of rank r. Also $\mathbf{bM} = 0$ iff $\mathbf{M}^t \mathbf{b}^t = 0$. Thus $\mathbf{M}' = \mathbf{M}^t$ is a parity check matrix of the