Solutions to Final Exam

Question 1

(a) A message source generates two different messages with probabilities 0.125 and 0.875. What is the information rate of this source in bits per message?

(b) If FEC were used to transmit this information over a Binary Symmetric Channel (BSC) with an error rate (p) of 0.2 (or 0.05), would it be possible to transmit this information without errors? Why or why not?

Solution

(a) The information rate (or "entropy") of a source is defined as:

$$H = \sum_i (-\log_2(P_i) \times P_i) \quad \text{bits/message}$$
$$= -\log_2(0.125) \cdot 0.125 - \log_2(0.8755) \cdot 0.875$$
$$\boxed{\approx 0.544 \text{ bits/message}}$$

(b) The maximum rate at which information can be transmitted with an arbitrarily low error rate over a channel is its capacity. For a BSC the capacity is:

$$C = 1 - (-p \log_2 p - (1-p)\log_2(1-p))$$
$$= 1 - (-0.2 \log_2(0.2) - 0.8 \log_2(0.8))$$
$$\boxed{\approx 0.278 \text{ bits/message}}$$

In this case the capacity is less than the information rate and it will not be possible to transmit the information over this channel without errors.

or:

$$C = 1 - (-p \log_2 p - (1-p)\log_2(1-p))$$
$$= 1 - (-0.05 \log_2(0.05) - 0.95 \log_2(0.05))$$
$$\boxed{\approx 0.714 \text{ bits/message}}$$

In this case the capacity is greater than the information rate and it will be possible to transmit the information over this channel without errors.

Question 2

What is the UTF-8 encoding for the Unicode glyph "Arabic-Indic digit seven", ٧, that has a Unicode code point of `U+0667` (or "Arabic-Indic digit three", ٣, that has the Unicode code point of `U+0663`)?

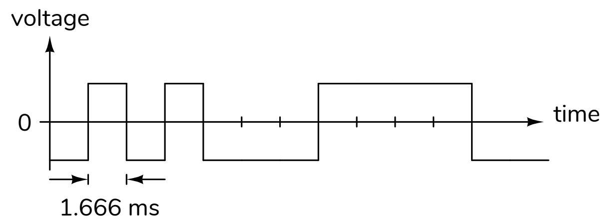
Solution

The code points `U+0663` and `U+0667` in binary are `110 0110 0011` and `110 0110 0111`. According to Table 1 of Lecture 1 both can be UTF-8 encoded into two bytes as follows:

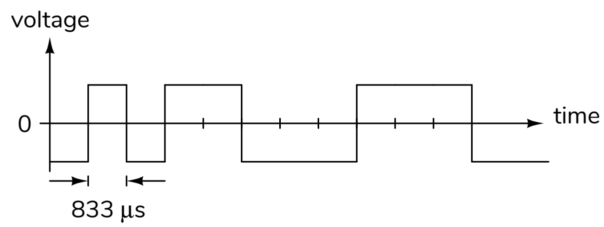
- For `U+0663` = `110 0110 0011`, *yyyyy* is `11001` and *xxxxxx* is `100011`. The UTF-8 encoding is thus `1101 1001` and `1010 0011` which is `0xD9, 0xA3`.
- For `U+0667` = `110 0110 0111`, *yyyyy* is `11001` and *xxxxxx* is `100111`. The UTF-8 encoding is thus `1101 1001` and `1010 0111` which is `0xD9, 0xA7`.

Question 3

The following waveform shows the waveform received over an asynchronous serial ("RS-232") interface configured for 8 data bits and even parity.



or:



(a) What is the baud rate?

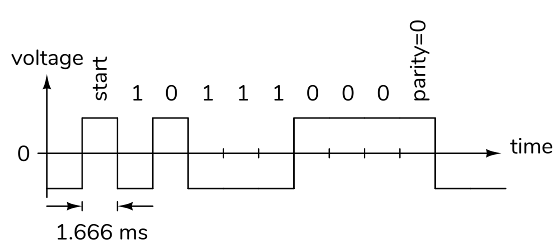
(b) What value was received? Give your answer as a hexadecimal number.

(c) Are any errors indicated?

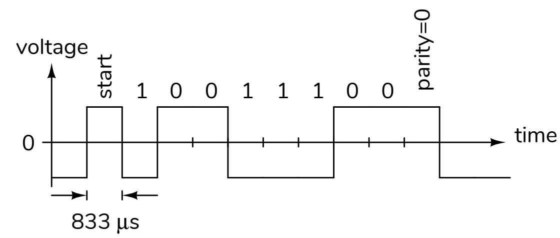
Solution

(a) The baud rate is the inverse of the shortest possible time between signal transitions, in this case one bit period. This is $1/1.666\times10^{-3}\,\mathrm{s} = \boxed{600\ \mathrm{Hz}}$ or $1/833\times10^{-6}\,\mathrm{s} = \boxed{1200\ \mathrm{Hz}}$.

(b) The bits are transmitted lsb-first with a `1` encoded as a negative voltage. As shown below there is a start bit at the start and a parity bit at the end. The bits in msb-first order are `0001 1101` or $\boxed{\text{0x1D}}$ for the first example and `0011 1001` or $\boxed{\text{0x39}}$ for the second.



or:



(c) Even parity means there should be an even number of ones, including the parity bit. In the both cases there are four ones (even number) and so $\boxed{\text{no error is indicated}}$.

Question 4

A noise signal has Gaussian probability distribution. Its average voltage is 1 V_{DC} and the RMS (AC) voltage is 1 (or 0.5) V_{RMS}. What is the probability that the signal will be negative (less than 0 V)?

Solution

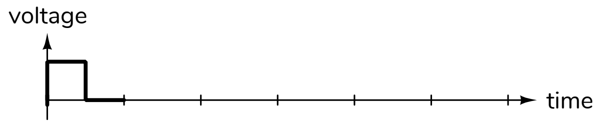
In this case the mean is $\mu = 1$, the standard deviation is $\sigma = 1$ (or 0.5) and the threshold is $v = 0$. The normalized threshold is

$$
\begin{aligned}
t &= \frac{v-\mu}{\sigma} \\
&= \frac{0-1}{1} = -1 \quad \text{or} \\
&= \frac{0-1}{0.5} = -2
\end{aligned}
$$

Using Figure 1 from Lecture 3 (“Noise”) or a calculator can we find the cumulative probability that the noise is less than the normalized threshold. For $t = -1$ this is $\boxed{\text{about 15\% (0.159)}}$ and for $t = -2$ this is $\boxed{\text{about 2\% (0.0228)}}$.

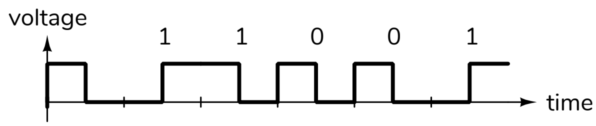
Question 5

Draw the waveform that would be used to transmit the sequence of bits `1`, `1`, `0`, `0`, `1` using *differential* Manchester coding using the convention that a 1 is transmitted as a different symbol. Draw your waveform starting after the symbol drawn below. The symbol shown below is not included in the sequence of bits given above.



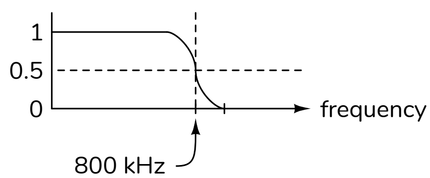
Solution

A differential line code encodes data as the difference between successive symbols. For Manchester the two symbols are high-low and low-high. The diagram below shows the encoding of the bit sequence above.

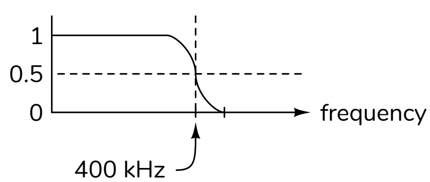


Question 6

A channel has the frequency response shown below:



or:



(a) What is the maximum symbol rate that can be transmitted over this channel without ISI?

(b) Assuming the symbol rate calculated above, how many bits per second could be transmitted if one of 8 (or 4) different symbols were transmitted in each symbol interval?

Solution

(a) The maximum symbol rate that can be transmitted without ISI over a channel that has a symmetrical frequency response about f is $2f$. In this case, for $f = 800$ kHz the maximum symbol rate would be $\boxed{1.6\,\text{MHz}}$ and for $f = 400$ kHz the maximum symbol rate would be $\boxed{800\,\text{kHz}}$.

(b) With 8 different symbols we can transmit $\log_2(8) = 3$ bits per symbol so the bit rate would be $3 \times 1.6 \times 10^6 = \boxed{4.8\,\text{Mbps}}$. With 4 different symbols we can transmit $\log_2(4) = 2$ bits per symbol so the bit rate would be $2 \times 800 \times 10^3 = \boxed{1.6\,\text{Mbps}}$

Question 7

A system uses differential data transmission over two conductors labelled `D+` and `D-`. There are two possible output conditions:

- When transmitting a **`1`**, the voltage on D_+ is 1 V (relative to ground) and the voltage on D_- is 0 V (relative to ground).
- When transmitting a **`0`**, D_+ is 0 V (relative to ground), and D_- is 1 V (relative to ground).

(a) What is the common mode voltage in the two states?

(b) What are the differential voltages ($D_+ - D_-$) in the two states?

Solution

(a) The common-mode voltage is the average of the two voltages. In this case it's $\frac{1+0}{2} = \boxed{0.5}$ V.

(b) The differential voltage, $D_+ - D_-$, is $1-0 = \boxed{1}$ V when transmitting a 1 and $0-1 = \boxed{-1}$ V when transmitting a 0.

Question 8

A message consisting of the bits **`1010`** (or **`1001`**) is to be protected using a CRC with a generator polynomial of **`1111`**.

(a) What is the length of the CRC in bits?

(b) What is the value of the CRC computed using the simple method described in the lecture notes?

(c) What message would be transmitted (data plus CRC)?

(d) What are the values of k and n?

Solution

(a) The length of the CRC is one less than the number of bits in the generator polynomial, in this case $\boxed{3}$.

(b) The CRC is the remainder after dividing by the generator polynomial:

```
      --------
1111 |1010000
      1111
      ----
       1010
       1111
       ----
        1010
        1111
        ----
         1010
         1111
         ----
          101
```

`101` for the first case and

```
      --------
1111 |1001000
      1111
      ----
       1100
       1111
       ----
        0110
        ----
         1100
         1111
         ----
          011
```

`011` for the second case.

(c) The message transmitted would be the CRC appended to the data bits: `1010 101` (or `1001 011`).

(d) The number of data bits is $\boxed{k = 4}$. The total number of bits is $\boxed{n = 7}$.

Question 9

A code has the following three codewords:

- `0000000`
- `1010101`
- `1111111`

(a) What is the minimum distance of this code?

(b) How many errors can be detected?

(c) How many errors can be corrected?

(d) If the codeword `0001111` is received, what codeword was most likely transmitted?

Solution

(a) The Hamming distance between the first codeword and the other two are 4 and 7 respectively. The Hamming distance between the second and third codewords is 3. Thus the minimum Hamming distance of this code is $d_{\min} = \boxed{3}$.

(b) $d_{\min} - 1 = \boxed{2}$ errors can be detected.

(c)
$$\left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor = \left\lfloor \frac{2}{2} \right\rfloor = \boxed{1}$$
errors can be corrected.

(d) If the codeword `0001111` is received, the distances from each of the codewords is 4, 4 and 3 respectively so the last codeword, `1111111` was the most likely transmitted.

Question 10

(a) A maximal-length PRBS sequence includes 128 (or 256) ones (1's) in each period. What is the period of the PRBS in bits?

(b) How many *pairs* are required for a bidirectional 1000-BASE-T (or 10-BASE-T) Ethernet link to operate?

(c) What is the most appropriate type of ARQ for a communication system when the delay is short relative to the frame duration?

Solution

(a) A m-sequence of length $2^n - 1$ has 2^{n-1} ones. Thus $n = \log_2(128) + 1 = 8$ and the period is $2^n - 1 = \boxed{255 \text{ bits}}$ (or $n = \log_2(256) + 1 = 9$ and the period is $2^9 - 1 = \boxed{511 \text{ bits}}$).

(b) 1000-BASE-T Ethernet uses all four pairs (in both directions) so the $\boxed{\text{four pairs are required}}$. 10-BASE-T uses one pair in each direction so $\boxed{\text{two pairs are required}}$.

(c) If the delay is short then $\boxed{\text{stop-and-wait ARQ}}$ does not limit the throughput and is simpler than the other alternatives.