Solutions to Final Exam

Question 1

(a) The entropy of a source is given by

$$H = \sum_i (-\log_2(P_i) \times P_i) \quad \text{bits/message}$$

There were two version of this question with message probabilities $P_i = \{0.9, 0.05, 0.05\}$ and $P_i = \{0.8, 0.1, 0.1\}$. The results are $0.9\log_2(0.9) + 2 \times 0.05\log_2(0.05) = 0.57$ and $0.8\log_2(0.8) + 2 \times 0.1\log_2(0.1) = 0.92 = 0.92$

(b) If a source generates one message per second the average information rate will be the same as the source entropy with units of $\frac{\text{bits}}{\text{message}} \times \frac{\text{messages}}{\text{second}} = \frac{\text{bits}}{\text{seconds}}$). Since the channel capacity is, by definition, the maximum information rate that can be transmitted reliably over that channel, the required capacity is the information rate computed in (a).

(c) Assuming an AWGN channel we can compute the required SNR using the Shannon channel capacity theorem:

$$C = B\log_2(1 + \frac{S}{N})$$

by solving for $\frac{S}{N}$, the SNR, in dB:

$$\text{SNR} = 10\log\left(2^{\frac{C}{B}} - 1\right)$$

There were two values for B (1 Hz and 2 Hz) resulting in four possible answers:

C	B	SNR (dB)
0.92	1	-0.5
0.92	2	-4.2
0.57	1	-3.2
0.57	2	-6.6

Question 2

(a) n is the number of bits per codeword (7) and k is the number of data bits ($\log_2(8) = 3$).

(b) The most likely transmitted codeword is the one with the shortest distance to the received codeword. If the received codeword was `1100000` the most likely transmitted codeword is `1100010` at distance 1. It the received codeword was `0000011` the most likely transmitted codeword was `0001011`, also at distance 1.

(c) As computed above, the Hamming distance between the most likely transmitted codeword and the received codeword is 1.

(d) The bit most likely in error is the one that differs: the sixth bit if `1100000` was received and the fourth bit if `0000011` was received.

Question 3

The UTF-8 encoding of a string composed of a capital 'A', a space, and the Greek capital letter Omega is obtained by UTF-8 encoding each of the characters.

From the Unicode character tables, the code point for 'A' is 0x41, for a space, 0x20 and for 'Ω' 0x3A9. The first two can be represented as a single byte because their code points can be represented in 7 bits. The binary value of 0x3A9 is 011 1010 1001 and will have a 2-byte UTF-8 encoding because its value can be represented with 11 bits. The first 5 bits are the 'y' field (01110) and the second 6 bits are the 'x' field (101001). When these fields are mapped into the corresponding fields in the two bytes in the UTF-8 encoding shown in the second line of table 3-6 we get: first-byte prefix=110 y=01110 second-byte prefix=10 x=101001. When grouped into bytes we have 1100 1111 and 1010 1001 which are 0xcf and 0xa9. So the overall encoding would be:

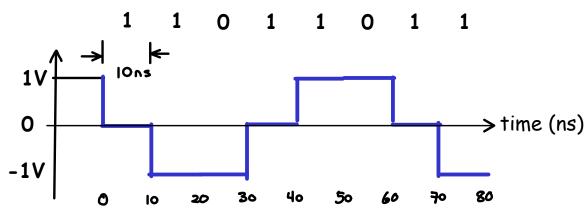
```
0x41 0x20 0xce 0xa9
```

The other version of the question was solved the same way except 'P' is 0x50, and a lower-case omega ('ω') is 0x03C9 = 011 1100 1001 which is encoded as 1100 1111 1000 1001 = 0xcf 0x89. So the overall encoded UTF-8 string is:

```
0x50 0x20 0xcf 0x89
```

Question 4

The MLT-3 line code changes level for each '1' bit. There are three levels with 100BASE-TX Ethernet using 0 and ± 1 V. Assuming the immediately preceding voltage level was a high (positive) level, the byte value 0xdb = 1101 1011 transmitted in LSB-first order would be:



With 4B5B encoding 10 bits would need to be transmitted instead of 8. The bit period would be 8 ns ($\frac{1}{125\times10^6}$) instead of 10 ns ($\frac{1}{100\times10^6}$). This was also marked correct.

Question 5

(a) The bytes labelled XX and YY are the IP header checksum field. The purpose is to detect transmission errors in the IP header.

(b) The IP checksum is computed according to the one's-complement checksum algorithm. There were two variants of this question, both of which gave the same 32-bit sum: 0x36339 Adding the MS 16 bits to the LS and inverting the bits gives the checksum: 0x9cc3 so XX is 9C and YY is C3.

(c) The value in the protocol field is 0x11, decimal 17 (UDP).

(d) The value in the destination address field is `FF FF FF FF` which in "dotted quad" notation is 255.255.255.255. This is the broadcast address so all hosts on the same IP network should receive this packet.

Question 6

If a communication system using NRZ signalling has a noise margin of v, then a noise voltage greater than v will cause an error. The probability of a Gaussian signal with zero mean ($\mu = 0$) and standard deviation σ) having a value less less than v is given by $P(t = \frac{v-\mu}{\sigma})$. Since we need the probability that the noise voltage is greater than v, we need to find $1 - P(t)$.

There were two version of this question ($v = 1, \sigma = 0.5$ and $v = 2, \sigma = 1$) both of which give $t = 2$. Using a calculator, the Logistic function approximation, or the graph in the lecture notes we find $P(2) \approx 0.98$ so the BER, the probability of an error, is about 2%.

Question 7

(a) A communication system using M-ary QAM at a symbol rate f_s transmits $f_s \log_2(M)$ bps. There were two version of this question with $M = 1024, f_s = 2$ MHz which results in 20 Mb/s and $M = 4096, f_s = 1$ MHz which results in 12 Mb/s.

(b) The velocity factor for a paper dielectric is $VF = \frac{1}{\sqrt{\varepsilon_r}} = \frac{1}{\sqrt{4}} = 0.5$. Thus the velocity of propagation is $VF \times c = 0.5 \times 3 \times 10^8 = 1.5 \times 10^8$ m/s.

(c) If a harmonic has a level that is -6 dB relative to the desired signal then its power is $10^{-6/10} = 0.25$ of the power of the desired signal. Similarly, a harmonic at -10 dB would have a power $10^{-10/10} = 0.1$ of the power of the desired signal. Assuming the powers of the two signals add then the total harmonic power is 0.35 relative to desired signal and the THD in dB is $10\log(\frac{0.35}{1}) = -4.5$ dB.