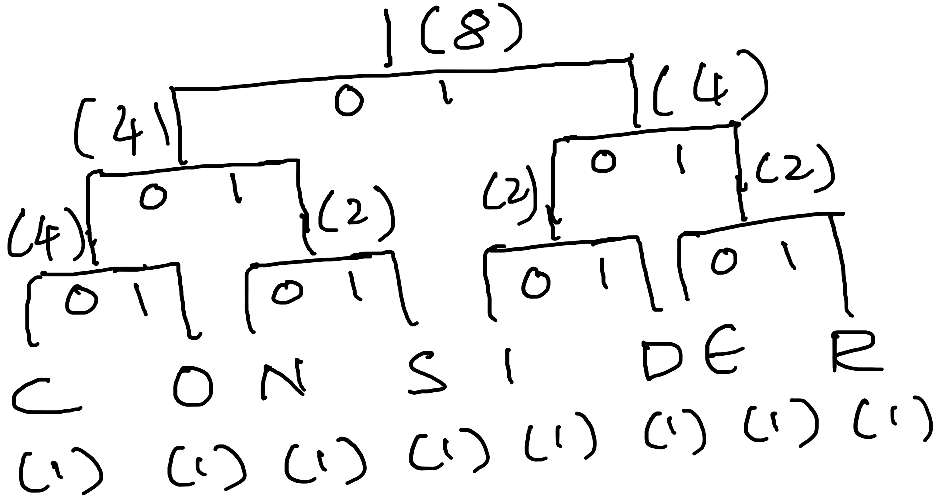
ECE462 – Lecture 10

Example: Find a Huffman code for the word "CONSIDER"

Information Source {C,O,N,S,I,D,E,R}

Frequency Count 1 1 1 1 1 1 1 1

Huffman Tree



Average H of bits = $\frac{24}{8}$ = 3

Entropy of the source n =8*$\frac{1}{8}$ $\log_2 8$ =3

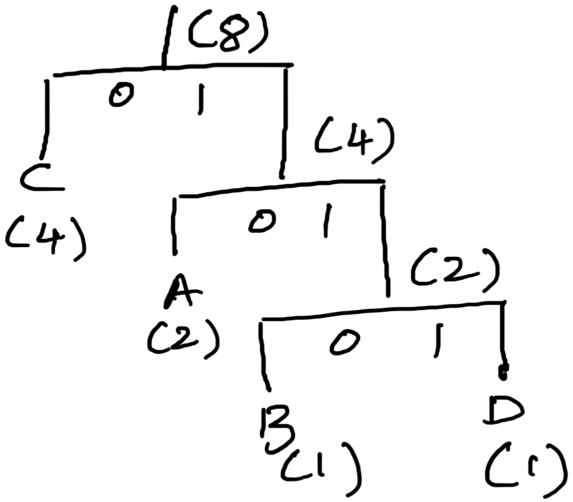
Symbol	#	code	total bits
C	1	000	3
O	1	001	3
N	1	010	3
S	1	011	3
I	1	100	3
D	1	101	3
E	1	111	3
R	1		3
	8		24

- As expected in a source with equiprobable outcomes there is no benefit from Huffman coding compares to fixed length coding.
- For Huffman coding to have an advantage compared to fixed length code at least one of the source elements must have higher or lower probability than the rest of the elements

Example: Consider the Information Source {A,B,C,D}

Frequency Count 2 1 4 1

Huffman Tree



Order elements : C A B D

count : 4 2 1 1

Symbol	Count#	code	total bits
C	4	0	4
A	2	1 0	4
B	1	1 1 0	3
D	1	1 1 1	3
	8		14

Average Information code rate (bits)

$\frac{14}{8} = \frac{7}{4} = 1.75$

(C) (A) (B,D)

Entropy of source : $\eta = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + \frac{2}{8}\log_2(8)$ = 0.5+0.5+0.75 =1.75 bits

he Huffman code achieves the entropy of the sources when all probabilities $\{P_i\}$ of all symbols {elements} of the source are power of $\frac{1}{2}$

Average Information
code rate (bits)

$$\frac{14}{8} = \frac{7}{4} = 1.75$$

Student Name: Student Number:

University of Toronto
Faculty of Applied Science and Engineering

MIDTERM EXAMINATION
ECE462H1S, Multimedia Systems
Time allowed: 90 minutes
March 9, 2006
Examiner: D. Hatzinakos

Exam type A
Non-programmable Calculators are allowed

---

1) Suppose a data source produces ones and zeros independently with probabilities $P(0)=7/12$ and $P(1)=5/12$. What is the binary entropy of this source? (3 points)

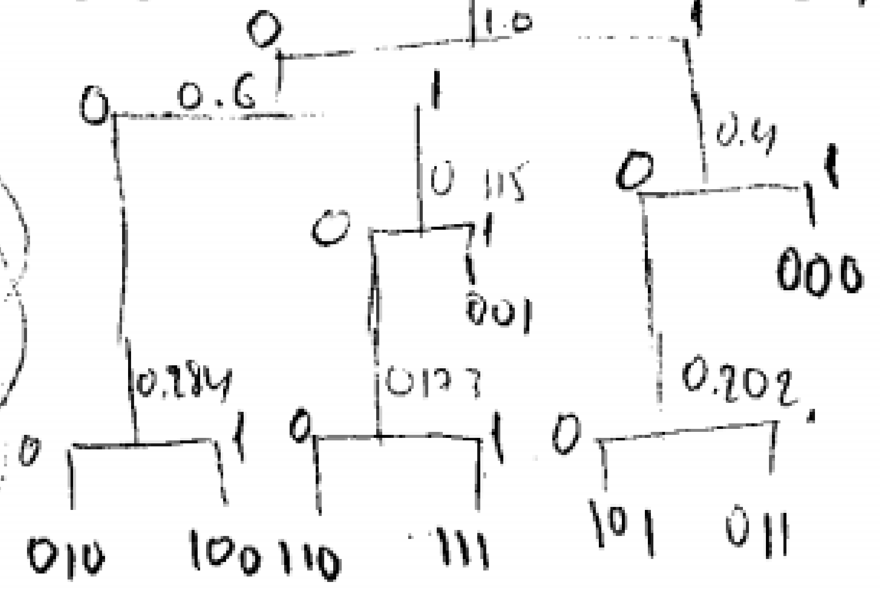
$$H = \frac{7}{12}\log_2\frac{12}{7} + \frac{5}{12}\log_2\frac{12}{5} = 0.9799 \text{ bits}$$

2) Generate a Huffman code for blocks of 3 bits (ones or zeros) from the source of question 1. What will be the codewords and the average codeword length for the code? (4 points)

First note that since 0 and 1 are produced independ-ently then for a block of three bits, $P(b_1 b_2 b_3) = P(b_1)P(b_2)P(b_3)$

Then

	Probability
000	(0.1985)
001	(0.1418)
010	(0.1118)
011	(0.1013)
100	(0.1418)
101	(0.1013)
110	(0.1013)
111	(0.0723)



Codewords

000 → 11
001 → 011
010 → 000
011 → 101
100 → 001
101 → 100
110 → 0100
111 → 0101

Average code length = 1·2·0.1985 + 3·3·0.1418 + 2·3·0.1013 + 1·4·0.1013 + 4·0.0723 = 2.97 bits

Example

QUESTION 4 (final exam 2014)

Let us design a Huffman Code for a source that puts out letters from an alphabet A={a1,a2,a3,a4,a5} with P(a1)=P(a3)=0.2, P(a2)=0.4 and P(a4)=P(a5)=0.1.

a) what is the entropy of the source?
b) Design a Huffman code , where sorting the alphabet you put combination of symbols as far down in the list as possible. Generate the code words of the code and provide the average code rate.
c) Design another Huffman code , where sorting the alphabet you put combination of symbols as far up in the list as possible. Generate the code words of this code and provide the average code rate.
d) Encode the sequence a2a1a3a2a1a2 using both codes derived above. Suppose now that was an error in the channel and the first bit received as a 0 instead of a 1. Decode the received sequence of bits. How many characters are received in error before the first correctly decoded character?
e) Repeat the last part with an error in the third bit.
f) Comment on any differences observed in the operation of the to codes.

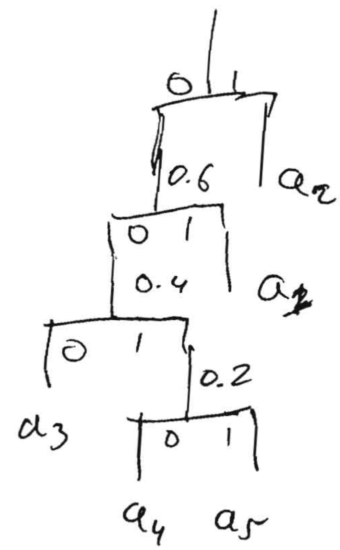
To be solved in class

a) $H(S) = 2 \cdot 0.2 \cdot \log_2 \frac{1}{0.2} + 0.4 \log_2 \frac{1}{0.4} + 2 \cdot 0.1 \log_2 \frac{1}{0.1} = \dots = 2.1$

$$\boxed{\text{Note: } \log_2 x = \frac{\log x}{\log 2}}$$

b) a_2 a_1 a_3 a_4 a_5

0.4 0.2 0.2 0.1 0.1



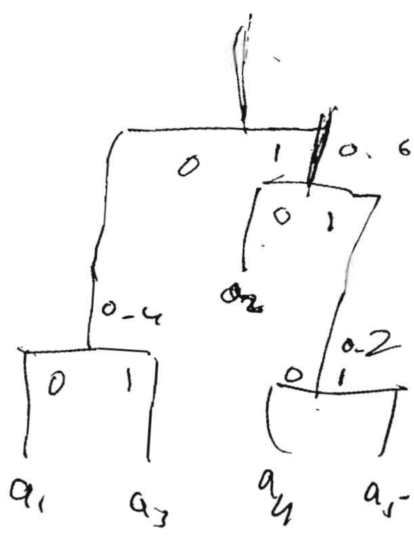
S_0

		Code	rate
a_2	0.4	1	0.4
a_1	0.2	01	0.4
a_3	0.2	000	0.6
a_4	0.1	0010	0.4
a_5	0.1	0011	0.4

average rate 2.2

c) a_2 a_1 a_3 a_4 a_5

0.4 0.2 0.2 0.1 0.1



			rate
a_2	0.4	10	0.8
a_1	0.2	00	0.4
a_3	0.2	01	0.4
a_4	0.1	110	0.3
a_5	0.1	111	0.3

rate 2.2

d) Encoding sequence

$a_2\ a_1\ a_3\ a_2\ a_1\ a_2$ — code 1 — 1010001011

$a_2\ a_1\ a_3\ a_2\ a_1\ a_2$ — code 2 — 10000/100010

received sequences (error in first bit) decoded sequence

code (1) 0010001011 → $a_4\ a_4\ a_2\ a_2$

code (2) 0000011 00010 → $a_1\ a_1\ a_3\ a_2\ a_1\ a_2$

e) received sequence (error in third bit)

decoded sequence

code 1 → 1000000 1011 → $a_2\ a_3\ a_4\ a_2\ a_2$

code 2 → 10100110 0010 → $a_2\ a_2\ a_3\ a_2\ a_1\ a_2$

Thus, second code more efficient than the first code in general