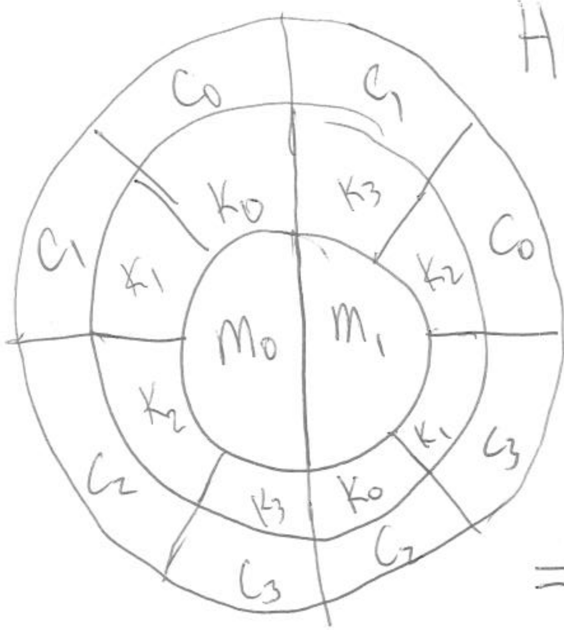
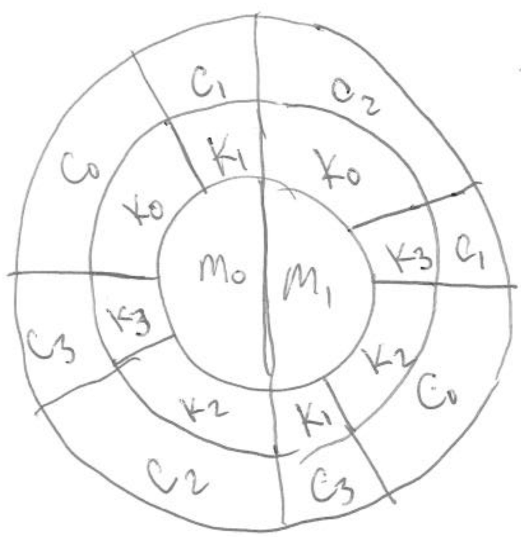


$$
\begin{aligned}
H(K|C) &= P(C=c_0)H(K|C=c_0) \\
&\quad + P(C=c_1)H(K|C=c_1) \\
&\quad + P(C=c_2)H(K|C=c_2) \\
&\quad + P(C=c_3)H(K|C=c_3) \\
&= \cancel{4}\,1
\end{aligned}
$$

$$
\begin{aligned}
&= H(K)+H(M)-H(C) \\
&= \log_2 4 + 1 - \log_2 4 \\
&= \underline{1}
\end{aligned}
$$



$H(K|C) = H(K)+H(M)-H(C)$

but $H(K)=H(C)$

and $H(M)=1$

$\Rightarrow H(K|C) = 1$

$$H(\text{next letter} \mid \text{first letter} = A) = P(\text{next}=A)\log_2\left(\frac{1}{P(\text{next}=A)}\right) + P(\text{next}=B)\log_2\left(\frac{1}{P(\text{next}=B)}\right) + P(\text{next}=C)\log_2\left(\frac{1}{P(\text{next}=C)}\right) + P(\text{next}=D)\log_2\left(\frac{1}{P(\text{next}=D)}\right)$$

$$= \frac{4}{10}\log_2(10/4) + \frac{1}{10}\log_2(10) + \frac{1}{2}\log_2(2)$$

$$H(\text{first} \mid \text{1st letter} = D) = \cancel{5/11 \log_2 11/5 + 6/11 \log_2(11/6)}$$

$$= 20/26 \log_2(26/20) + \frac{6}{26}\log_2(26/6)$$

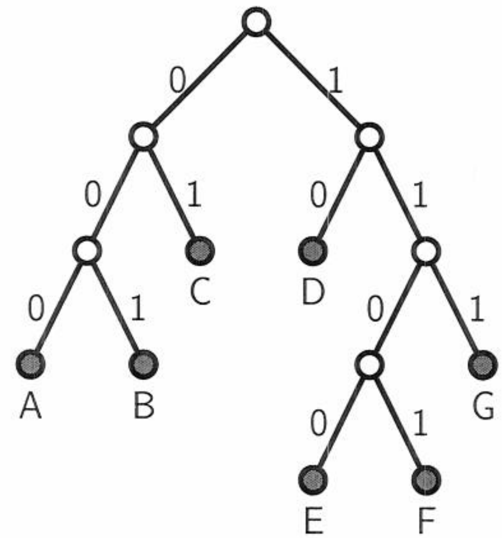
$$H(\text{word} \mid \text{at least one letter } D) = P(\text{first letter } D \text{ \& not } 2^{nd})\, H(\text{word} \mid \text{first letter } D) + P(2^{nd} \text{ letter } D \text{ \& not } 1^{st})\, H(\text{word} \mid 2\text{nd letter } D) + P(\text{both letters } D)\, H(\text{word} \mid \text{both letters } D)$$

$$= \frac{3}{10}\left(\frac{1}{10}\log_2 10 + \frac{9}{10}\log_2 10/9\right) + P(\text{word} = AD \text{ or } BD) \cdot H(\text{word} \mid \text{word} = AD \text{ OR } BD)$$

$$= \frac{3}{10}\left(\frac{1}{10}\log_2 10 + \frac{9}{10}\log_2 10/9\right) + \left(\frac{4}{10}\cdot\frac{5}{10} + \frac{1}{10}\cdot\frac{6}{10}\right) \cdot \left(\frac{20/100}{26/100} \cdot \log_2(26/20) + \frac{6/100}{26/100}\log_2(26/6)\right)$$

A Comma-Free Binary Code

Definition: A binary code is *comma-free* if no prefix of the code of a letter is the code of another letter.



A=000	B=001	C=01
D=10	E=1100	F=1101
G=111		

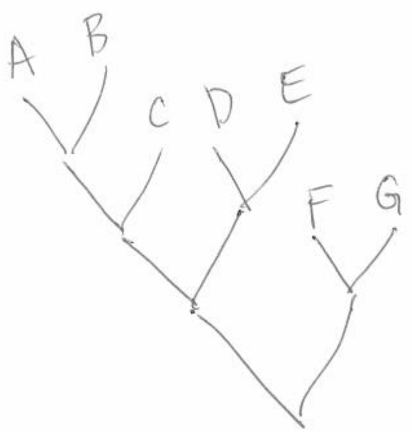
FEED

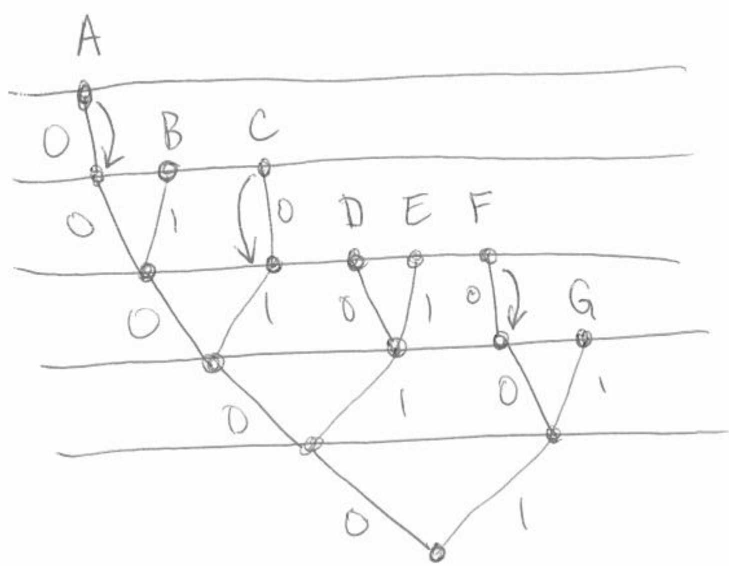
1101 1100 110010

File length $= 3N_A + 3N_B + 2N_C + 2N_D + 4N_E + 4N_F + 3N_G$

Tree from Heights

α	A	B	C	D	E	F	G
p_α	$\frac{2}{64}$	$\frac{4}{64}$	$\frac{6}{64}$	$\frac{10}{64}$	$\frac{13}{64}$	$\frac{13}{64}$	$\frac{16}{64}$
$\lceil log_2(\frac{1}{p_\alpha}) \rceil$	5	4	4	3	3	3	2



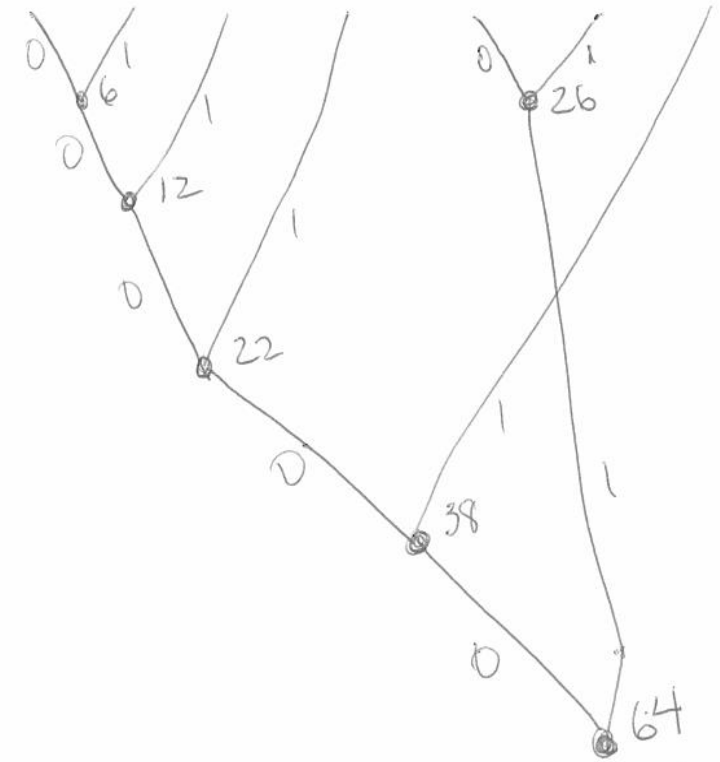


tree is not complete

but drop down all branches with single branch + now it is complete

Huffman Code

A	B	C	D	E	F	G
2	4	6	10	13	13	16



Bits per character =
$168/64 = 2.625$

Expected file length with this code

$2\cdot5 + 4\cdot5 + 4\cdot6 + 3\cdot10 + 2\cdot16 + 2\cdot13 + 2\cdot13$

$10 + 20 + 24 + 30 + 32 + 26 + 26 = 168$

110
52

Expected Code Length

Theorem 2 *The best possible expected code length (bits per letter) is*

$$\boxed{H = \sum_{i=1}^{n} p_i \log_2 1/p_i}$$

Proof.

Letter frequenices $N_1, N_2, \ldots, N_k$ ($N = \sum_{i=1}^{k} N_i$)

Code lengths $h_1, h_2, \ldots, h_k$ (from a binary tree)

$p_i = N_i/N$ and $q_i = 1/2^{h_i}$

$$\begin{aligned}
\text{File length} &= \sum_{i=1}^{k} N_i h_i \\
&= \sum_{i=1}^{k} N_i \log_2 2^{h_i} \qquad \frac{N_i}{N}\cdot N \\
&= N \sum_{i=1}^{k} p_i \log_2 1/q_i \\
&\geq N \sum_{i=1}^{k} p_i \log_2 1/p_i = NH
\end{aligned}$$