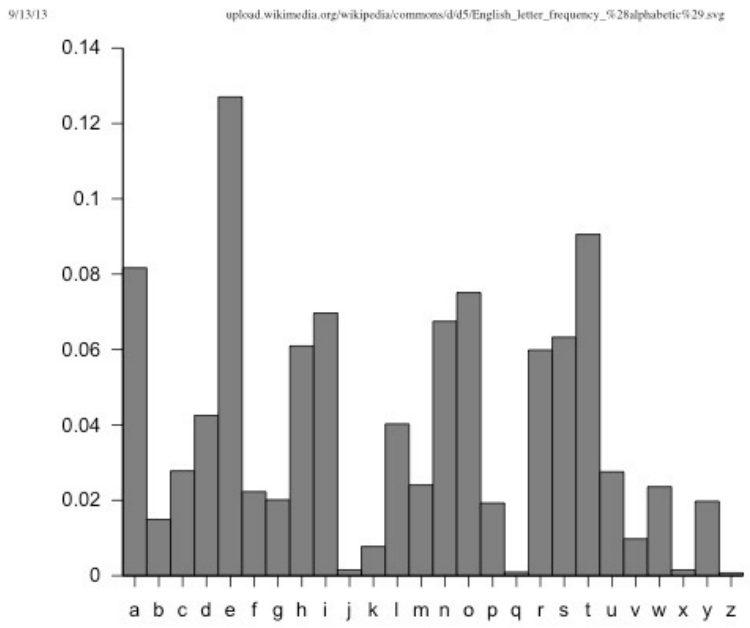
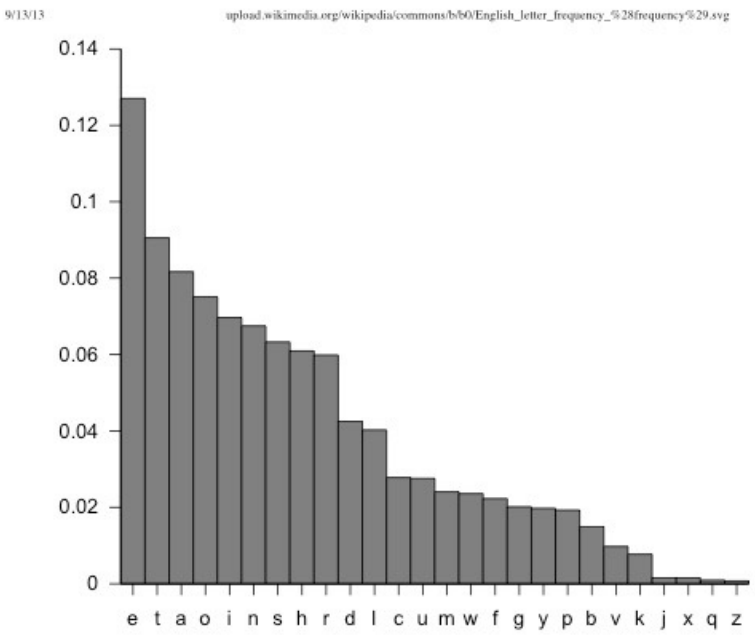
Letter frequency in English

source: Wikipedia

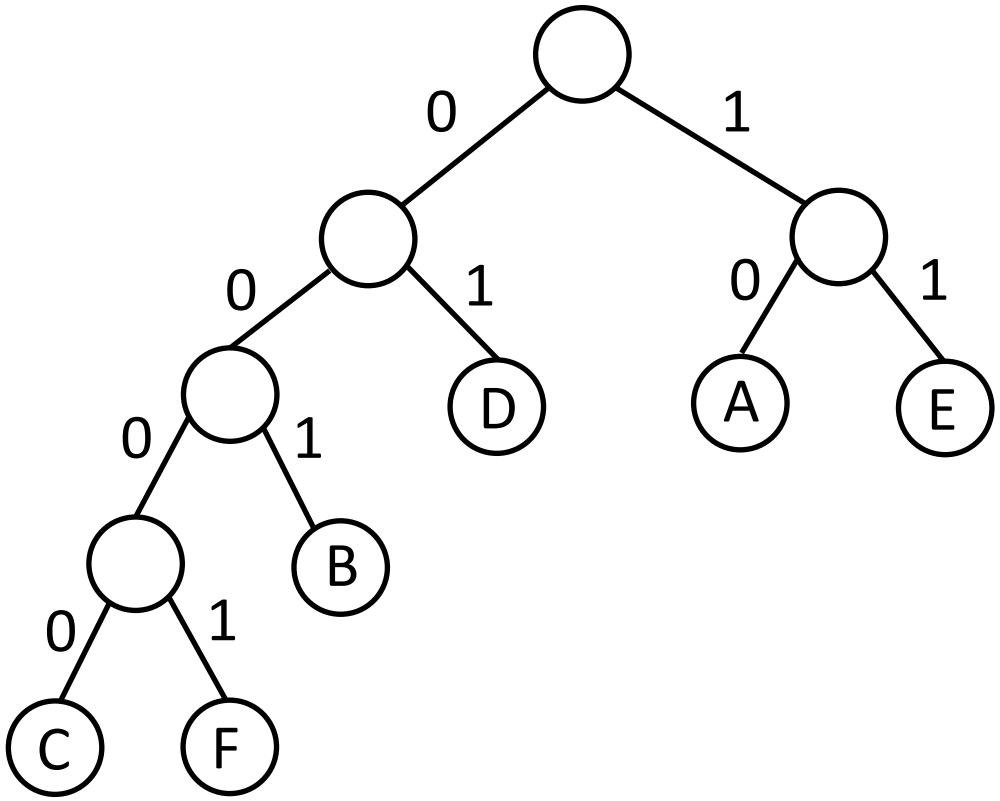
http://en.wikipedia.org/wiki/Letter_frequency





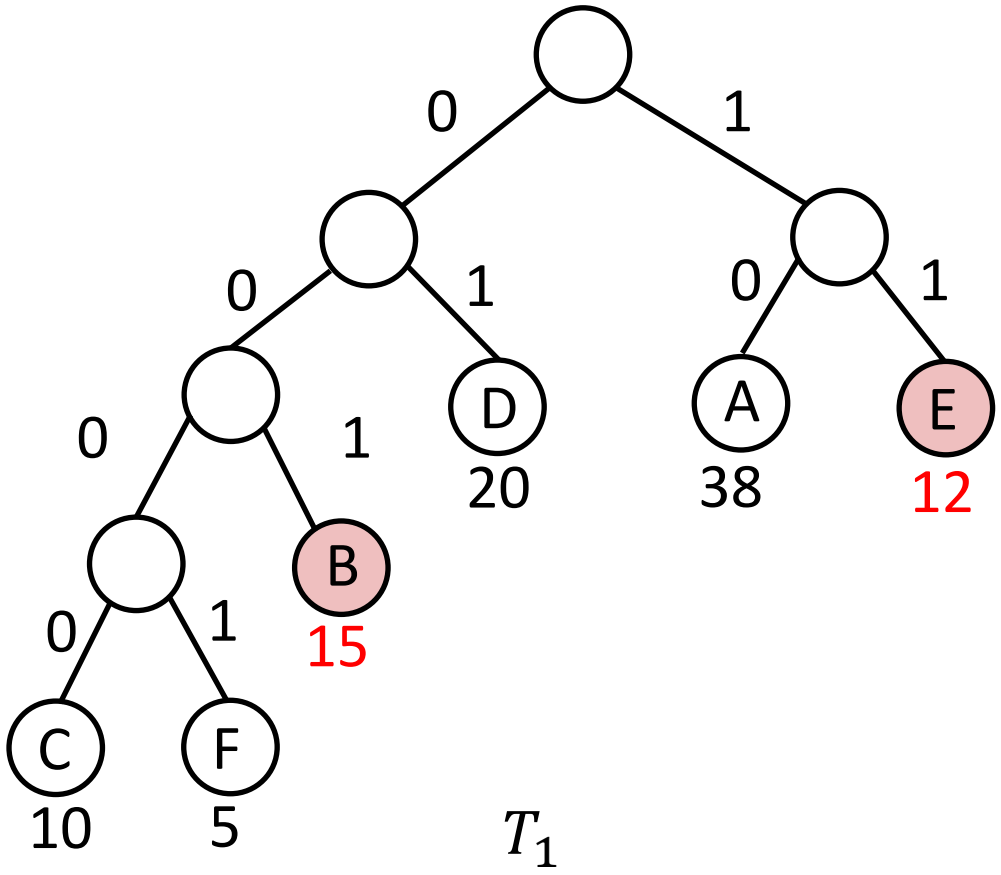
Example of binary tree representation of prefix code

$\Gamma = \{A, B, C, D, E, F\}$



Symbol	Codeword
A	10
B	001
C	0000
D	01
E	11
F	0001

Average depth, given code and frequencies

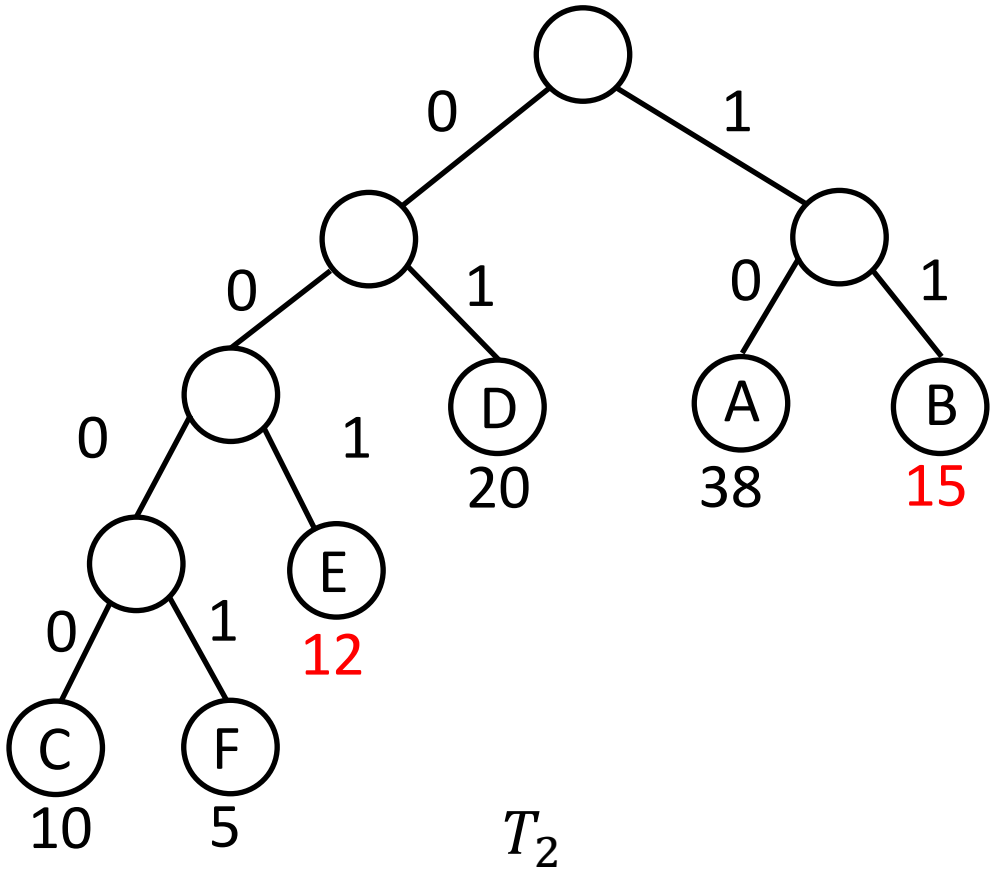


T_1

Symbol	Frequency	Codeword
A	0.38	10
B	0.15	001
C	0.10	0000
D	0.20	01
E	0.12	11
F	0.05	0001

$$AD(T_1) = 0.10\times4 + 0.05\times4 + 0.15\times3 + 0.20\times2 + 0.38\times2 + 0.12\times2 = 2.45$$

Improved code

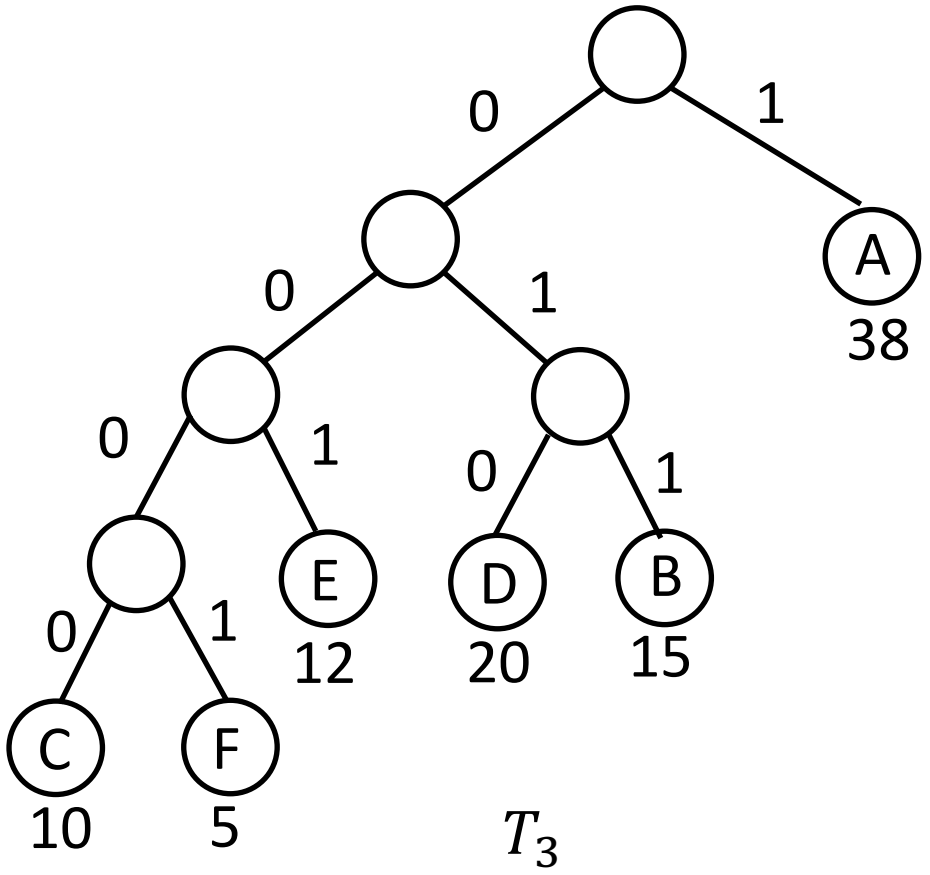


T_2

Symbol	Frequency	Codeword
A	0.38	10
B	0.15	11
C	0.10	0000
D	0.20	01
E	0.12	001
F	0.05	0001

$$AD(T_2) = 0.10\times4 + 0.05\times4 + 0.12\times3 + 0.20\times2 + 0.38\times2 + 0.15\times2 = 2.42$$

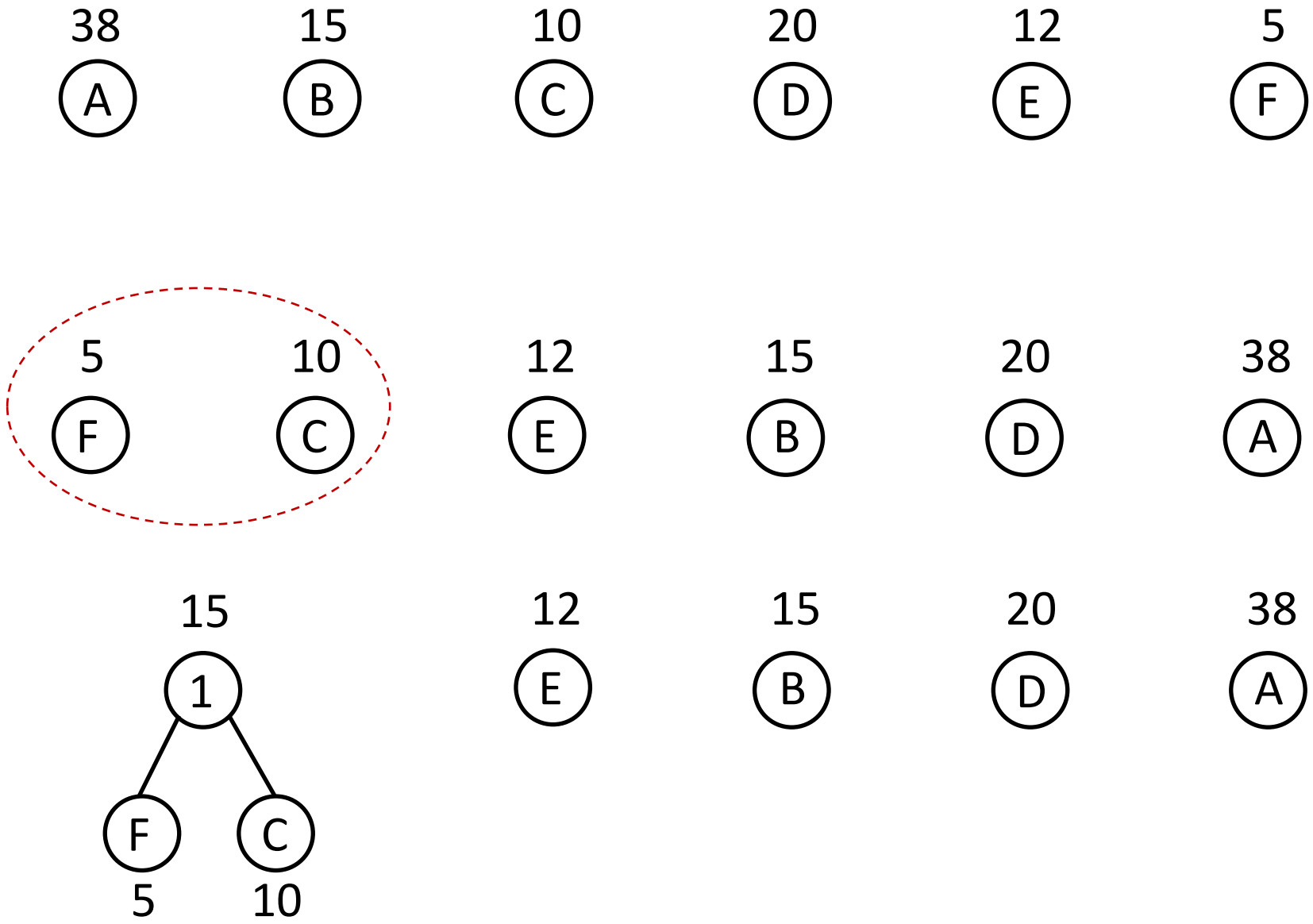
Optimal code

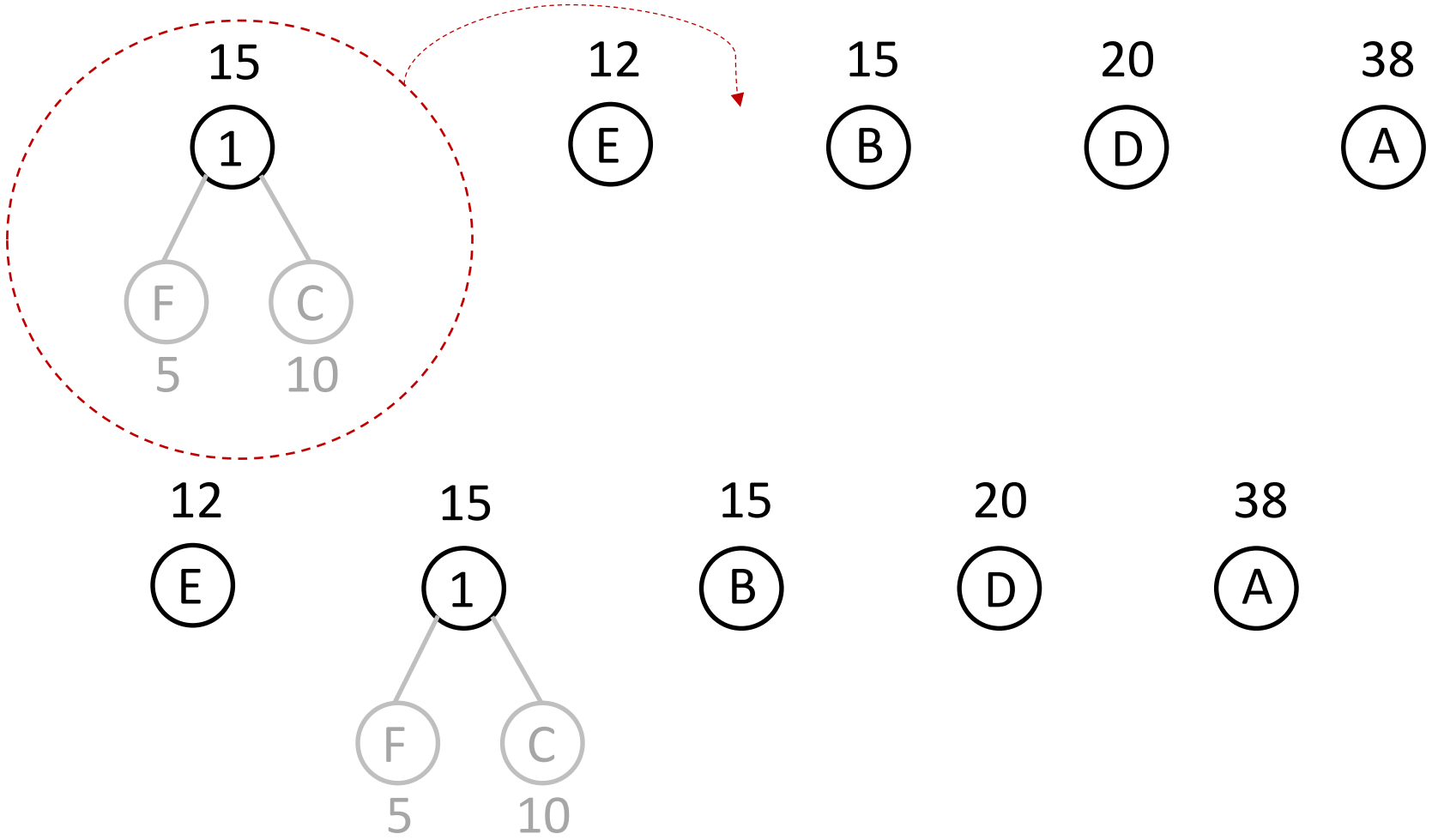


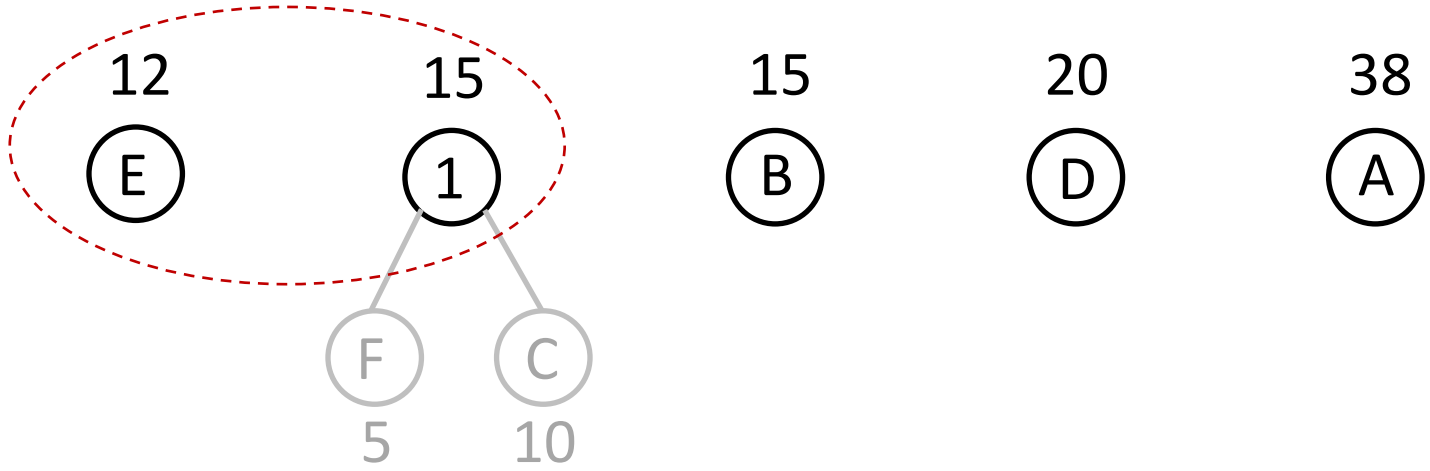
T_3

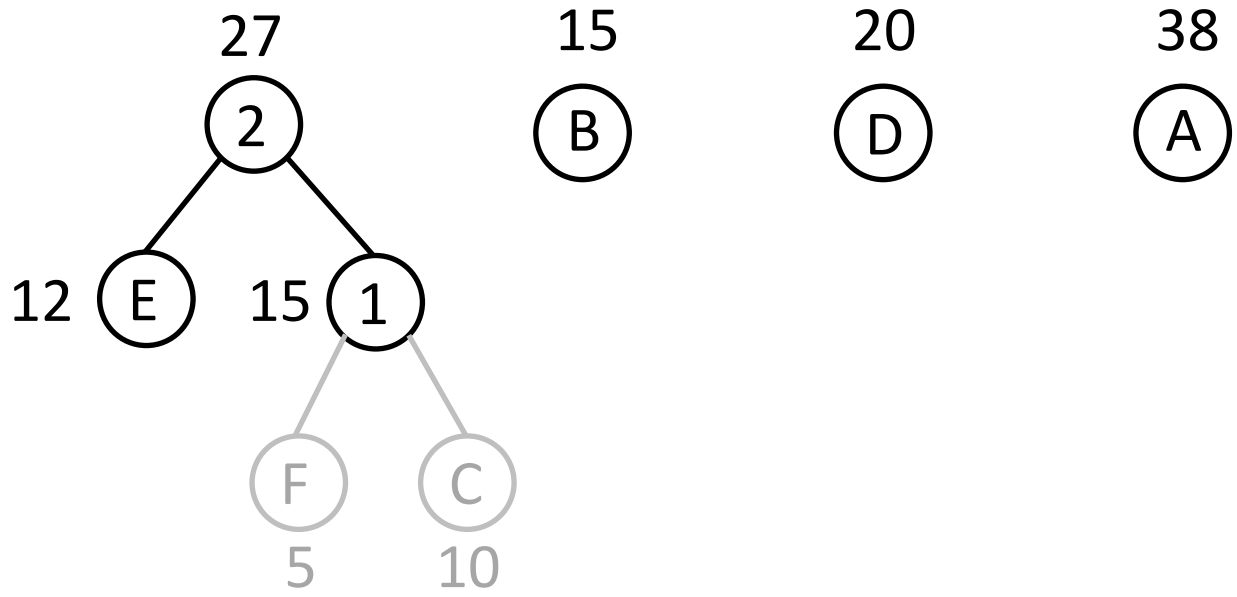
Symbol	Frequency	Codeword
A	0.38	1
B	0.15	011
C	0.10	0000
D	0.20	010
E	0.12	001
F	0.05	0001

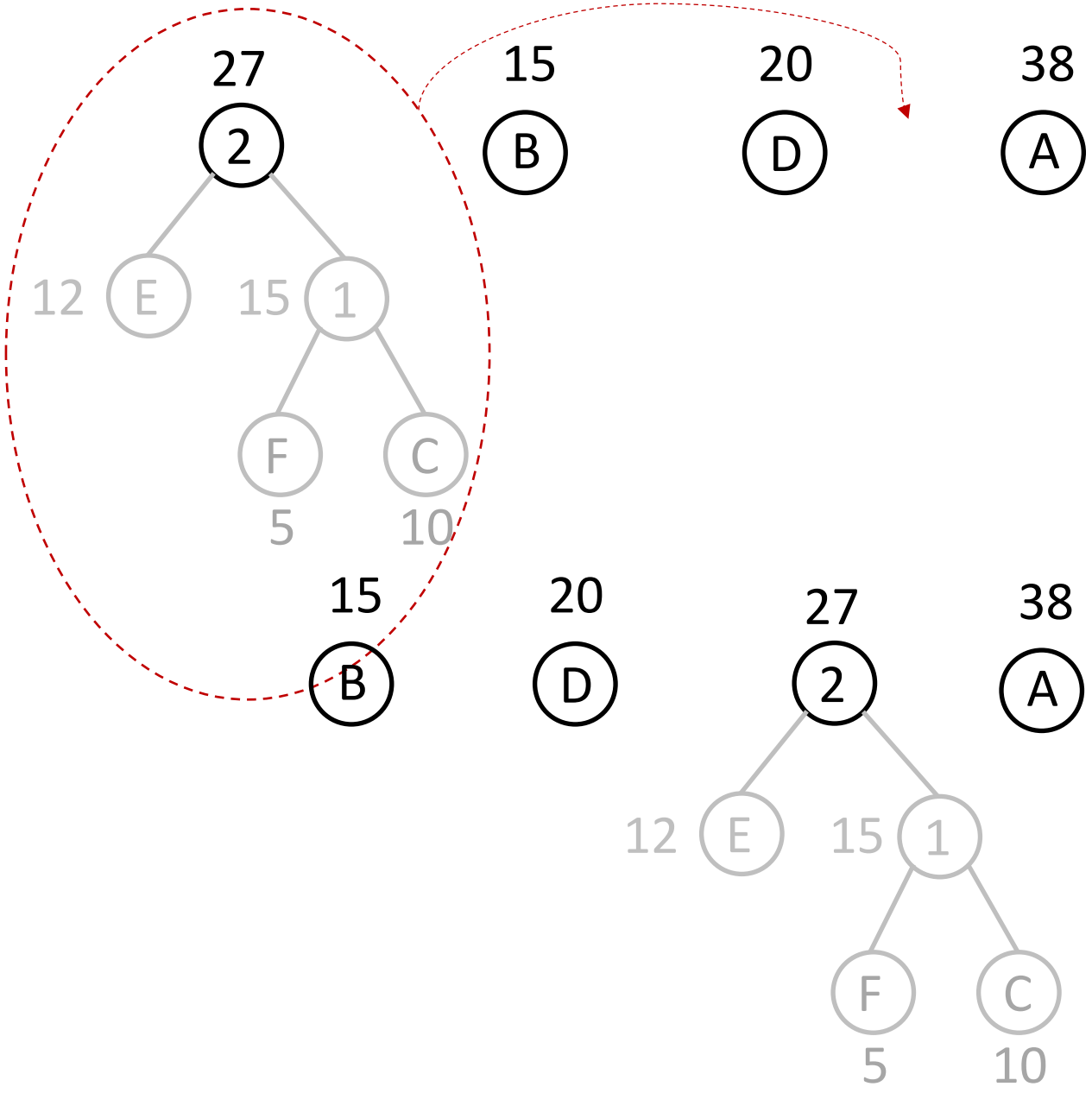
$$AD(T_3) = 0.10\times4 + 0.05\times4 + 0.12\times3 + 0.20\times3 + 0.15\times3 + 0.38\times1 = 2.39$$

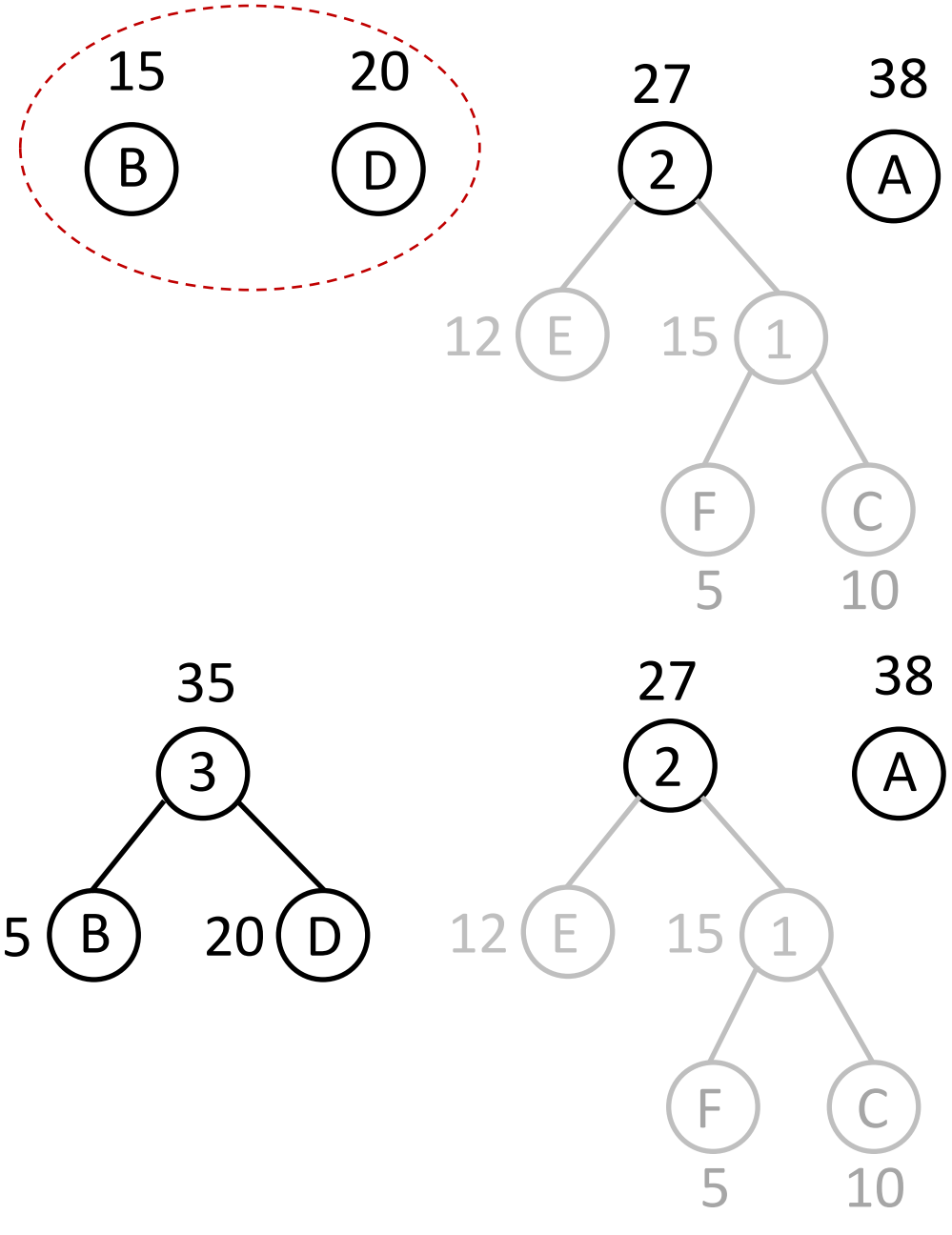

38
A
15
B
10
C
20
D
12
E
5
F
5
F
10
C
12
E
15
B
20
D
38
A
15
1
F
C
5
10
12
E
15
B
20
D
38
A

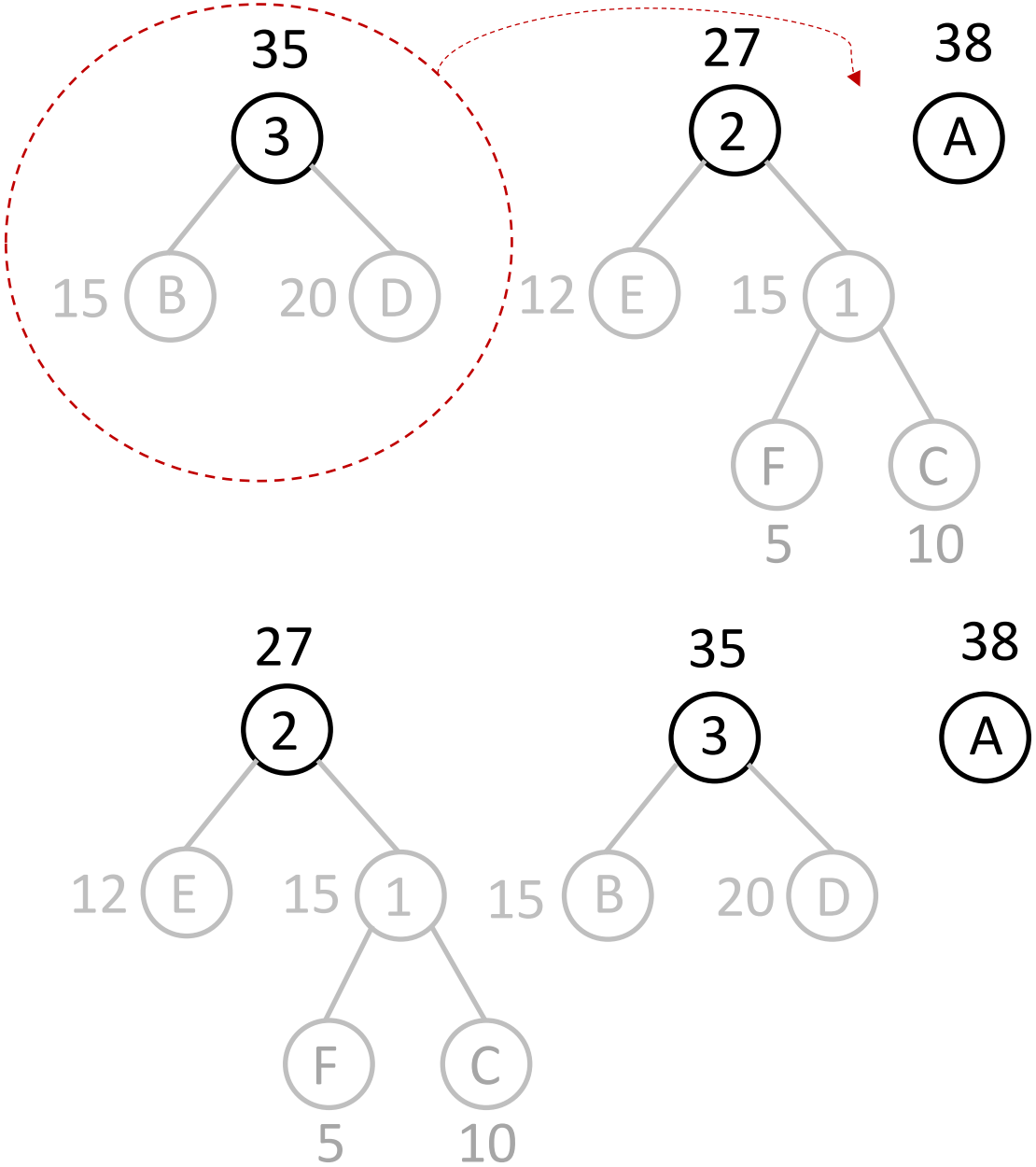

15
1
F
C
5
10
12
E
15
B
20
D
38
A
12
E
15
1
F
C
5
10
15
B
20
D
38
A

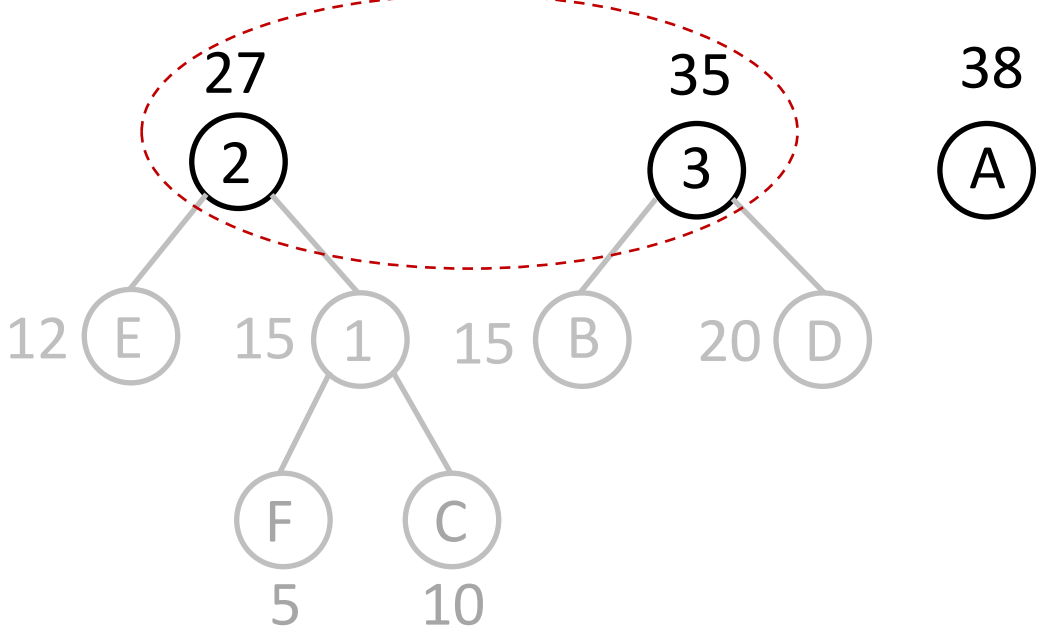

12
15
15
20
38
E
1
B
D
A
F
C
5
10

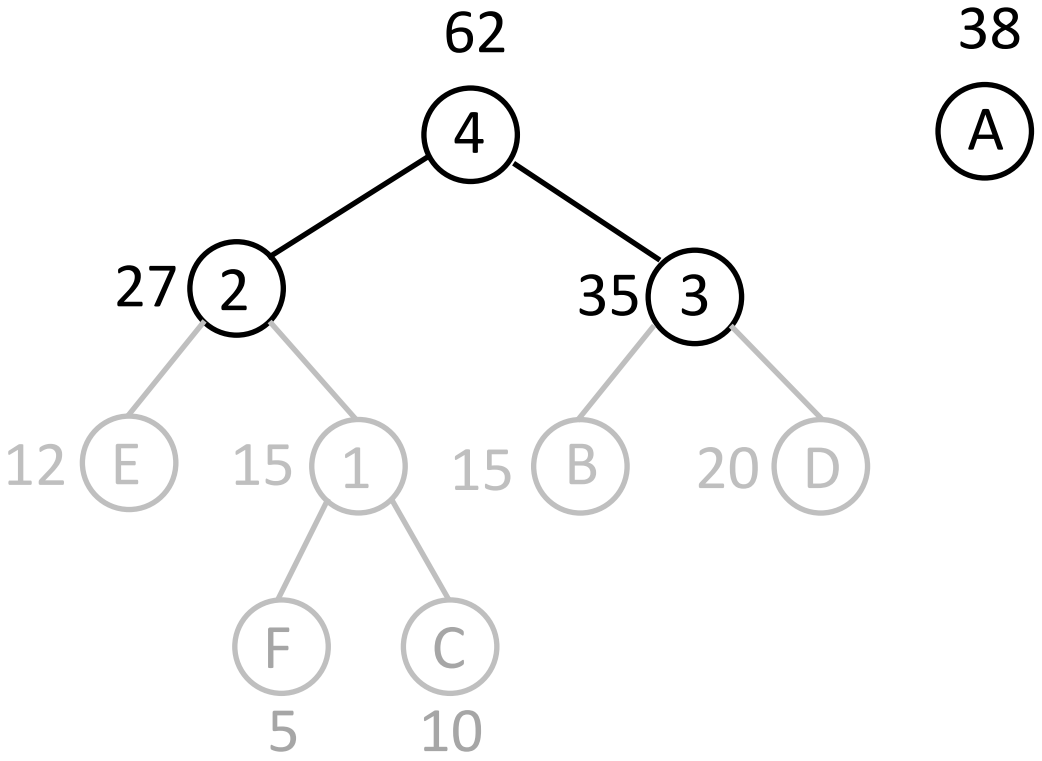

27
15
20
38
2
B
D
A
12
E
15
1
F
C
5
10

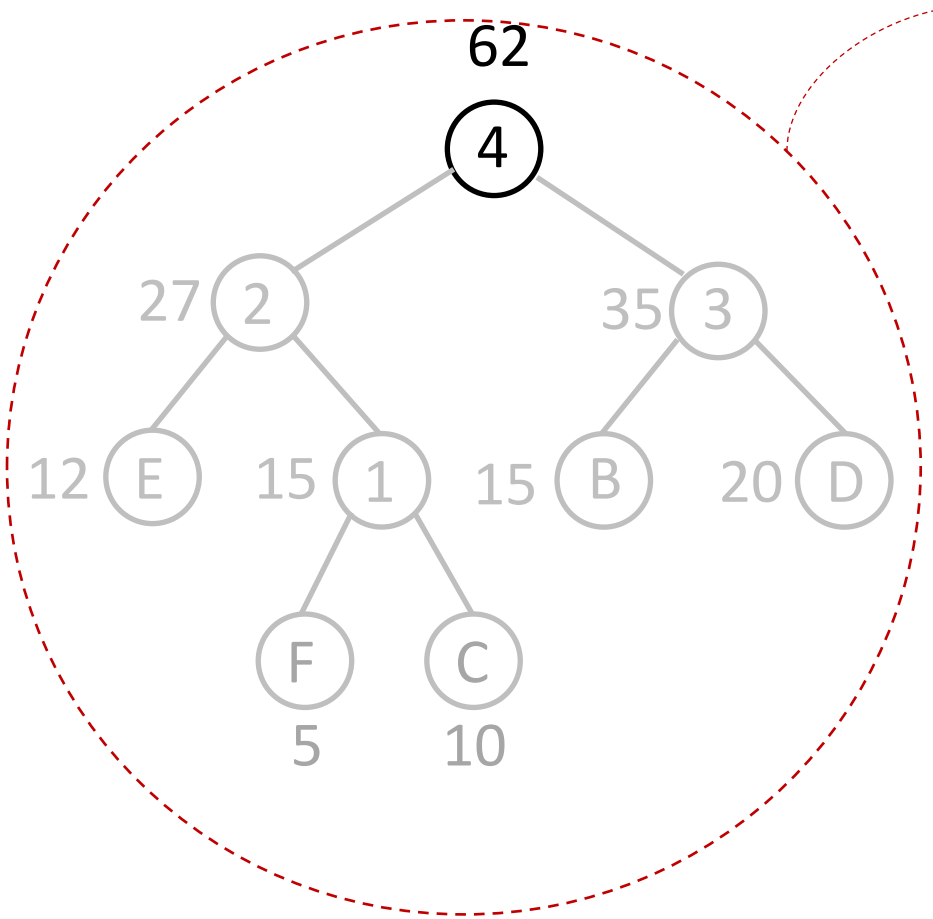

27
2
12
E
15
1
F
5
C
10
15
B
20
D
38
A
15
B
20
D
27
2
38
A
12
E
15
1
F
5
C
10


15
20
B
D
27
38
2
A
12
E
15
1
F
C
5
10
35
3
15
B
20
D
27
38
2
A
12
E
15
1
F
C
5
10


35
3
15
B
20
D
27
2
12
E
15
1
F
5
C
10
38
A
27
2
12
E
15
1
F
5
C
10
35
3
15
B
20
D
38
A

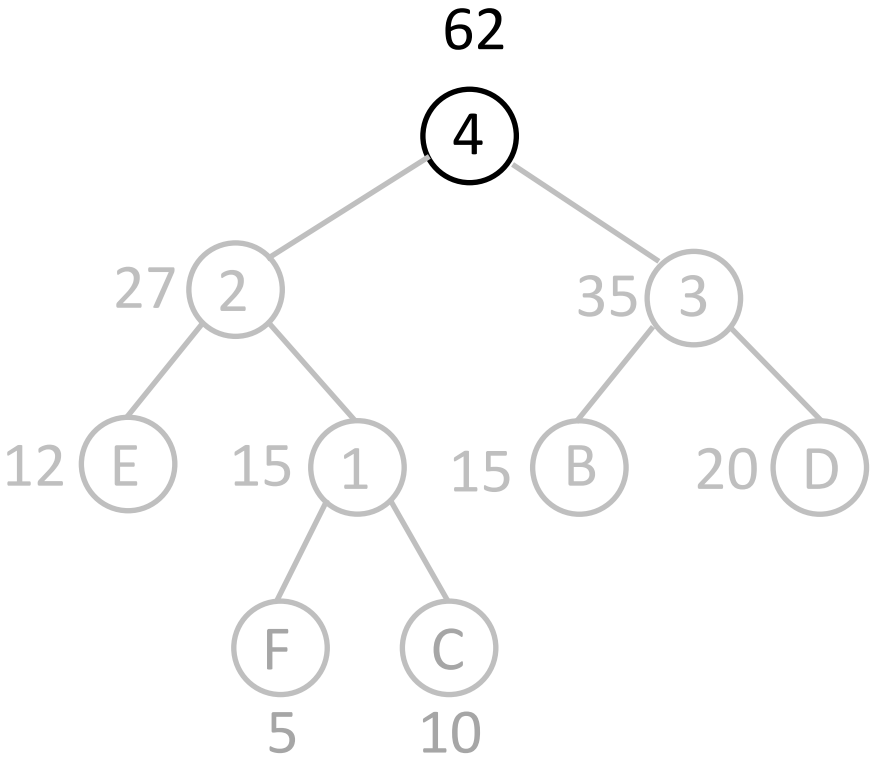

27
35
38
2
3
A
12
E
15
1
15
B
20
D
F
C
5
10

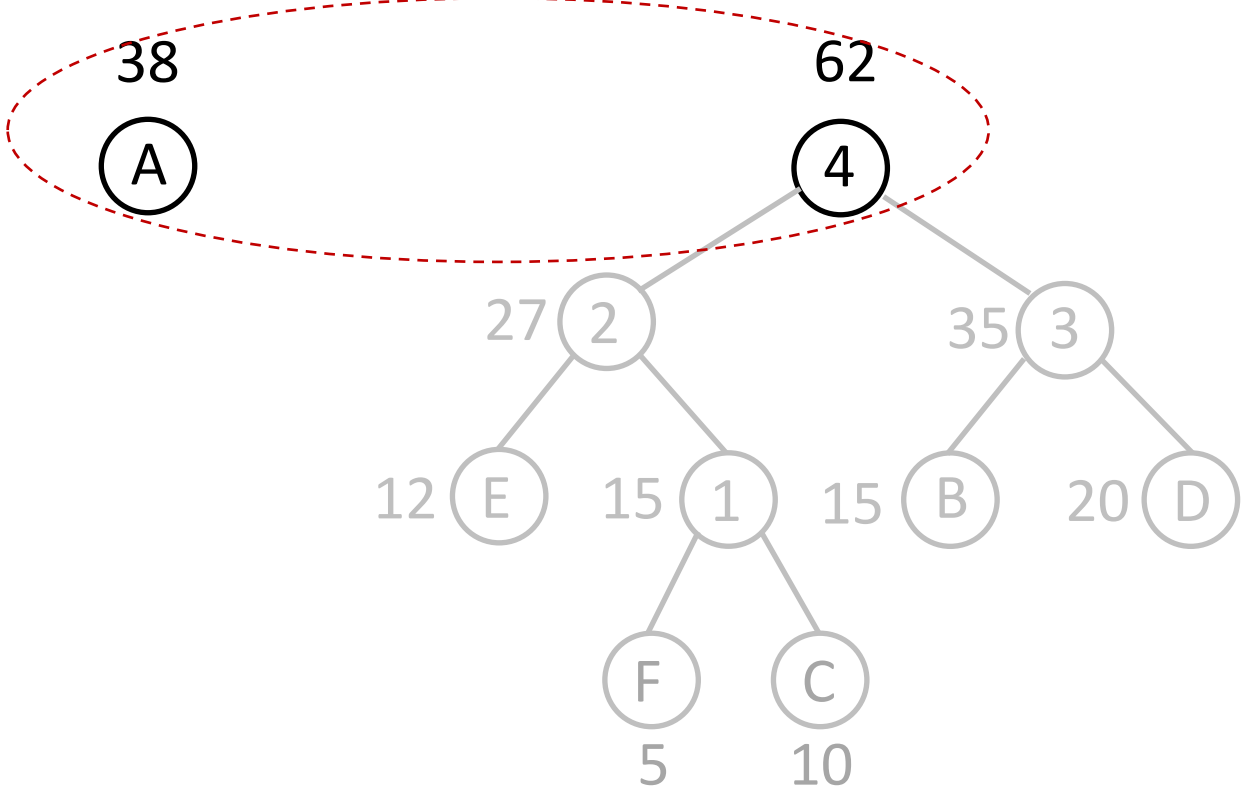

62
38
4
A
27
2
35
3
12
E
15
1
15
B
20
D
F
C
5
10

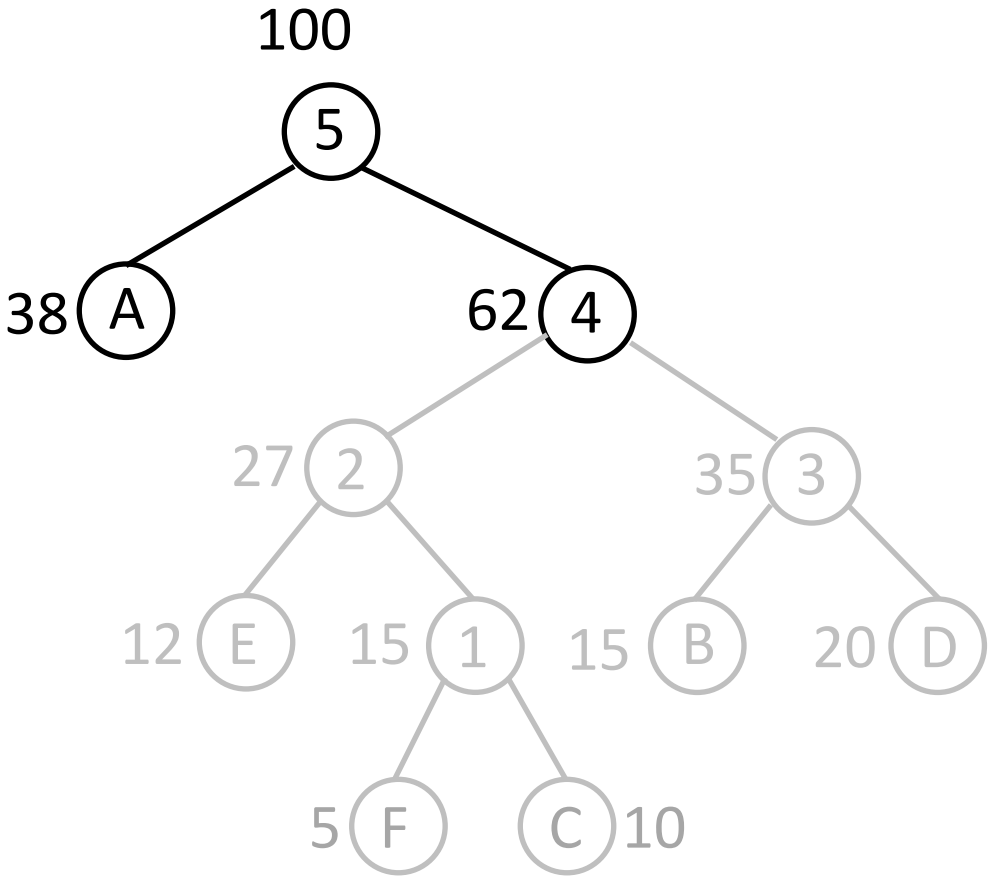

62
4
27
2
35
3
12
E
15
1
15
B
20
D
F
C
5
10

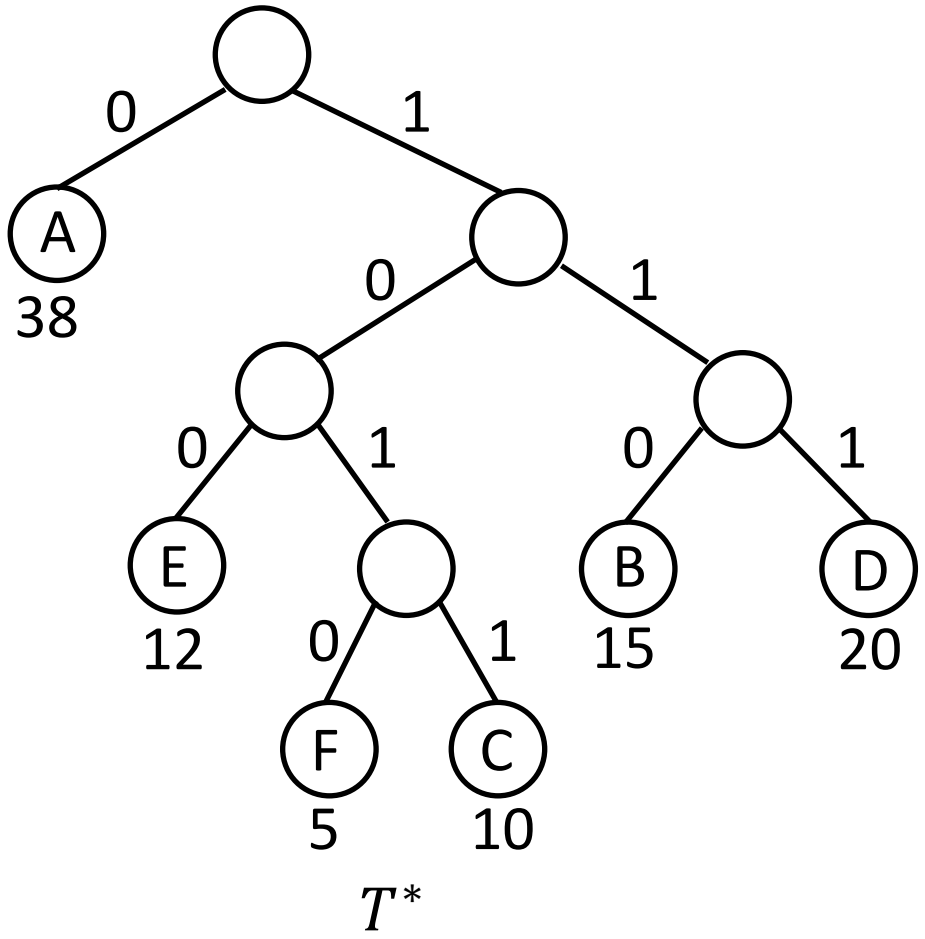
38
A

38
A


62
4
27
2
35
3
12
E
15
1
15
B
20
D
F
C
5
10


38
A
62
4
27
2
35
3
12
E
15
1
15
B
20
D
F
5
C
10


100
5
38
A
62
4
27
2
35
3
12
E
15
1
15
B
20
D
5
F
C
10



T^*

Symbol	Frequency	Codeword
A	.38	0
B	.15	110
C	.10	1011
D	.20	111
E	.12	011
F	.05	1010

$$AT(T^*) = 0.10\times4 + 0.05\times4 + 0.12\times3 + 0.20\times3 + 0.15\times3 + 0.38\times1 = 2.39$$