Entropy based data compression algorithms

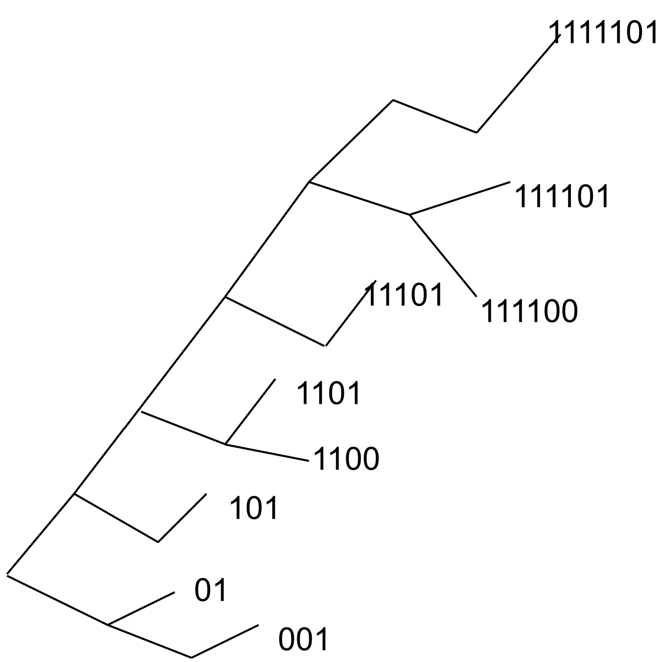
Shannon-Fano, Huffman,
Shannon-Fano-Elias codes

Shannon – Fano code

$$p_1 = 0.49; p_2 = 0.14; p_3 = 0.14; p_4 = 0.07; p_5 = 0.07; p_6 = 0.04; p_7 = 0.02; p_8 = 0.02; p_9 = 0.01$$

$$l_1 = \left\lceil ld\frac{1}{p_1} \right\rceil = \lceil 1.029 \rceil = 2;\ l_2 = \left\lceil ld\frac{1}{p_2} \right\rceil = \lceil 2.836 \rceil = 3;\ l_3 = \left\lceil ld\frac{1}{p_3} \right\rceil = \lceil 2.836 \rceil = 3;\ l_4 = \left\lceil ld\frac{1}{p_4} \right\rceil = \lceil 3.836 \rceil = 4;\ l_5 = \left\lceil ld\frac{1}{p_5} \right\rceil = \lceil 3.836 \rceil = 4;$$

$$l_6 = \left\lceil ld\frac{1}{p_6} \right\rceil = \lceil 4.64 \rceil = 5;\ l_7 = \left\lceil ld\frac{1}{p_7} \right\rceil = \lceil 5.64 \rceil = 6;\ l_8 = \left\lceil ld\frac{1}{p_8} \right\rceil = \lceil 5.64 \rceil = 6; l_9 = \left\lceil ld\frac{1}{p_9} \right\rceil = \lceil 6.64 \rceil = 7$$



Szimb.	Kódszó
X_1	01
X_2	001
X_3	101
X_4	1100
X_5	1101
X_6	11101
X_7	111100
X_8	111101
X_9	1111101

Efficiency

$$p_1 = 0.49; p_2 = 0.14; p_3 = 0.14; p_4 = 0.07; p_5 = 0.07; p_6 = 0.04; p_7 = 0.02; p_8 = 0.02; p_9 = 0.01$$

$$l_1 = \left\lceil ld\frac{1}{p_1} \right\rceil = \lceil 1.029 \rceil = 2;\; l_2 = \left\lceil ld\frac{1}{p_2} \right\rceil = \lceil 2.836 \rceil = 3;\; l_3 = \left\lceil ld\frac{1}{p_3} \right\rceil = \lceil 2.836 \rceil = 3;\; l_4 = \left\lceil ld\frac{1}{p_4} \right\rceil = \lceil 3.836 \rceil = 4;\; l_5 = \left\lceil ld\frac{1}{p_5} \right\rceil = \lceil 3.836 \rceil = 4;$$

$$l_6 = \left\lceil ld\frac{1}{p_6} \right\rceil = \lceil 4.64 \rceil = 5;\; l_7 = \left\lceil ld\frac{1}{p_7} \right\rceil = \lceil 5.64 \rceil = 6;\; l_8 = \left\lceil ld\frac{1}{p_8} \right\rceil = \lceil 5.64 \rceil = 6; l_9 = \left\lceil ld\frac{1}{p_9} \right\rceil = \lceil 6.64 \rceil = 7$$

$$L = 0.49 * 2 + 0.28 * 3 + 0.14 * 4 + 0.04 * 5 + 0.04 * 6 + 0.01 * 7 =$$
$$= 0.98 + 0.84 + 0.56 + 0.2 + 0.24 + 0.07 = 2.89$$

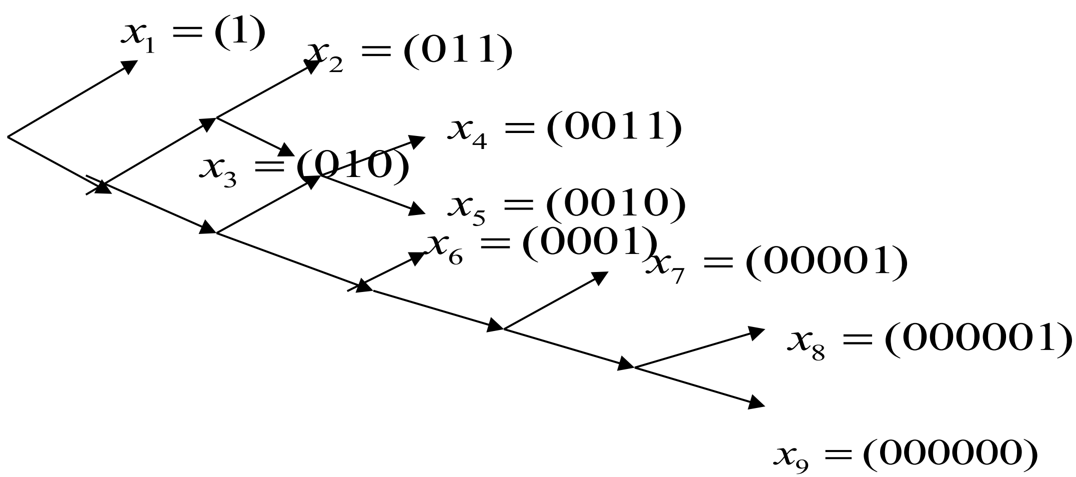
$$H(X) = \sum_{i=1}^{9} p_i ld\left(\frac{1}{p_i}\right) = 2.314$$

$$Efficiency = \frac{H(X)}{L} = 0.08 = 80\%$$

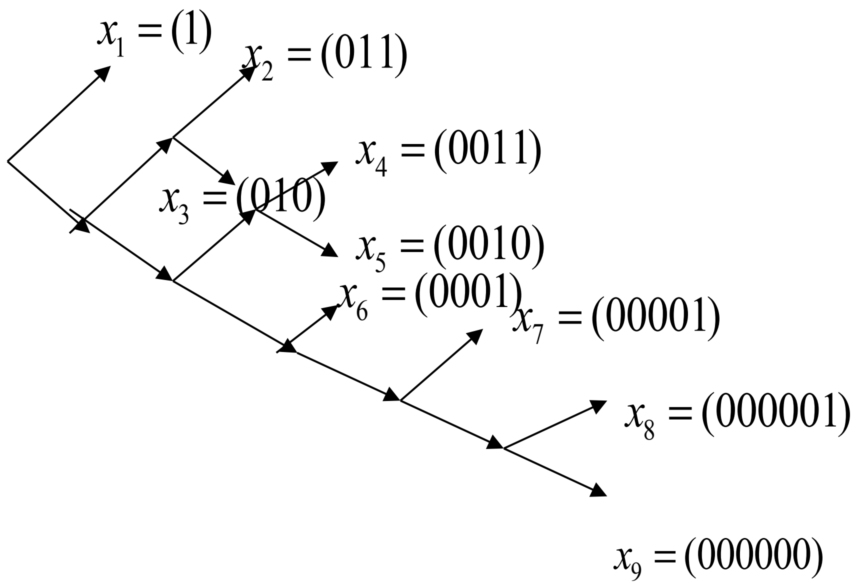
Solution by Huffman coding

$p_1 = 0.49;$	$0.49;$	$0.49;$	$0.49;$	$0.49;$	$0.49;$	$0.49;$	$0.49;$
$p_2 = 0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.28;$	$0.51;$
$p_3 = 0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.14;$	$0.23;$	
$p_4 = 0.07;$	$0.07;$	$0.07;$	$0.07;$	$0.14;$	$0.23;$		
$p_5 = 0.07;$	$0.07;$	$0.07;$	$0.07;$	$0.09;$			
$p_6 = 0.04;$	$0.04;$	$0.04;$	$0.09;$				
$p_7 = 0.02;$	$0.02;$	$0.05;$					
$p_8 = 0.02;$	0.03						
$p_9 = 0.01$							

$p_1 = 0.49;$	0.49;	0.49;	0.49;	0.49;	0.49;	0.49;	0.49;	1
$p_2 = 0.14;$	0.14;	0.14;	0.14;	0.14;	0.14;	0.28;	0.51;	
$p_3 = 0.14;$	0.14;	0.14;	0.14;	0.14;	0.14;	0.23;		
$p_4 = 0.07;$	0.07;	0.07;	0.07;	0.14;	0.23;			
$p_5 = 0.07;$	0.07;	0.07;	0.07;	0.09;				
$p_6 = 0.04;$	0.04;	0.04;	0.09;					
$p_7 = 0.02;$	0.02;	0.05;						
$p_8 = 0.02;$	0.03							
$p_9 = 0.01$								



The coding LUT



Szimb.	Kódszó
X_1	1
X_2	011
X_3	010
X_4	0010
X_5	0011
X_6	0001
X_7	00001
X_8	000001
X_9	000000

Efficiency

$$L^{HUFF} = 0.49*1 + 0.14*3 + 0.14*3 + 0.07*4 + 0.07*4 + 0.04*4 + 0.02*5 + 0.02*6 + 0.01*6 = 2.33$$

$$L^{SF} = 0.49*2 + 0.28*3 + 0.14*4 + 0.04*5 + 0.04*6 + 0.01*7 = 2.89$$

$f_s = 160 \text{ MHz} = 1,6*10^8$ esetén

$$R_{Huffman} = 372,8 \text{ Mbps} \qquad R_{SF} = 462 \text{ Mbps}$$

$$\boxed{\text{Diff. dataspeed} = 1,6*10^8\left(2.89 - 2.33\right) = 89,6*10^6 \approx 90 \text{ Mbps}}$$

Solution by Shannon-Fano-Elias codes

Prob. Mass func.	Distribution	Mod. Distr.	Bin. conversion	Codeword
$p_1 = 0.49;$	$F(1) = 0;$	$\bar{F}(1) = 0.245;$	$0.001\overline{11110101110000101000};$	001;
$p_2 = 0.14;$	$F(2) = 0.49;$	$\bar{F}(2) = 0.56;$	$0.1\overline{000111101101110000101};$	1000;
$p_3 = 0.14;$	$F(3) = 0.63;$	$\bar{F}(3) = 0.7;$	$0.10\overline{1100};$	1011;
$p_4 = 0.07;$	$F(4) = 0.77;$	$\bar{F}(4) = 0.805;$	$0.110\overline{01110000101000111101};$	11001;
$p_5 = 0.07;$	$F(5) = 0.84;$	$\bar{F}(5) = 0.875;$	$0.1110;$	11100;
$p_6 = 0.04;$	$F(6) = 0.91;$	$\bar{F}(6) = 0.93;$	$0.111\overline{01110000101000111101};$	111011;
$p_7 = 0.02;$	$F(7) = 0.95;$	$\bar{F}(7) = 0.96;$	$0.\overline{11110101110000101000};$	1111010;
$p_8 = 0.02;$	$F(8) = 0.97;$	$\bar{F}(8) = 0.98;$	$0.1\overline{11110101110000101000};$	1111101;
$p_9 = 0.01$	$F(9) = 0.99;$	$\bar{F}(9) = 0.995;$	$0.111\overline{11110101110000101000};$	11111110;

SFE code (cont’)

Szimb.	Kódszó
X_1	001
X_2	1000
X_3	1011
X_4	11001
X_5	11100
X_6	111011
X_7	1111010
X_8	1111101
X_9	11111110

$$L^{SFE} = 0.49*3 + 0.14*4 + 0.14*4 + 0.07*5 + 0.07*5 + 0.04*6 + 0.02*7 + 0.02*7 + 0.01*8 = 3.89$$

$$L^{SF} = 2.89 \qquad L^{HUFF} = 2.33$$

Data speed

$$L^{HUFF} = 2.33 \qquad L^{SF} = 2.89 \qquad L^{SFE} = 3.89$$

$$\Downarrow$$

$$R_{Huffman} = 372,8 \text{ Mbps} \qquad R_{SF} = 462 \text{ Mbps} \qquad R_{SFE} = 622,4 \text{ Mbps}$$

Without data compression 640 Mbps !!!!

Conclusion: small inprovement in „*L*” matters a lot in data speed !!!

Comparative analysis

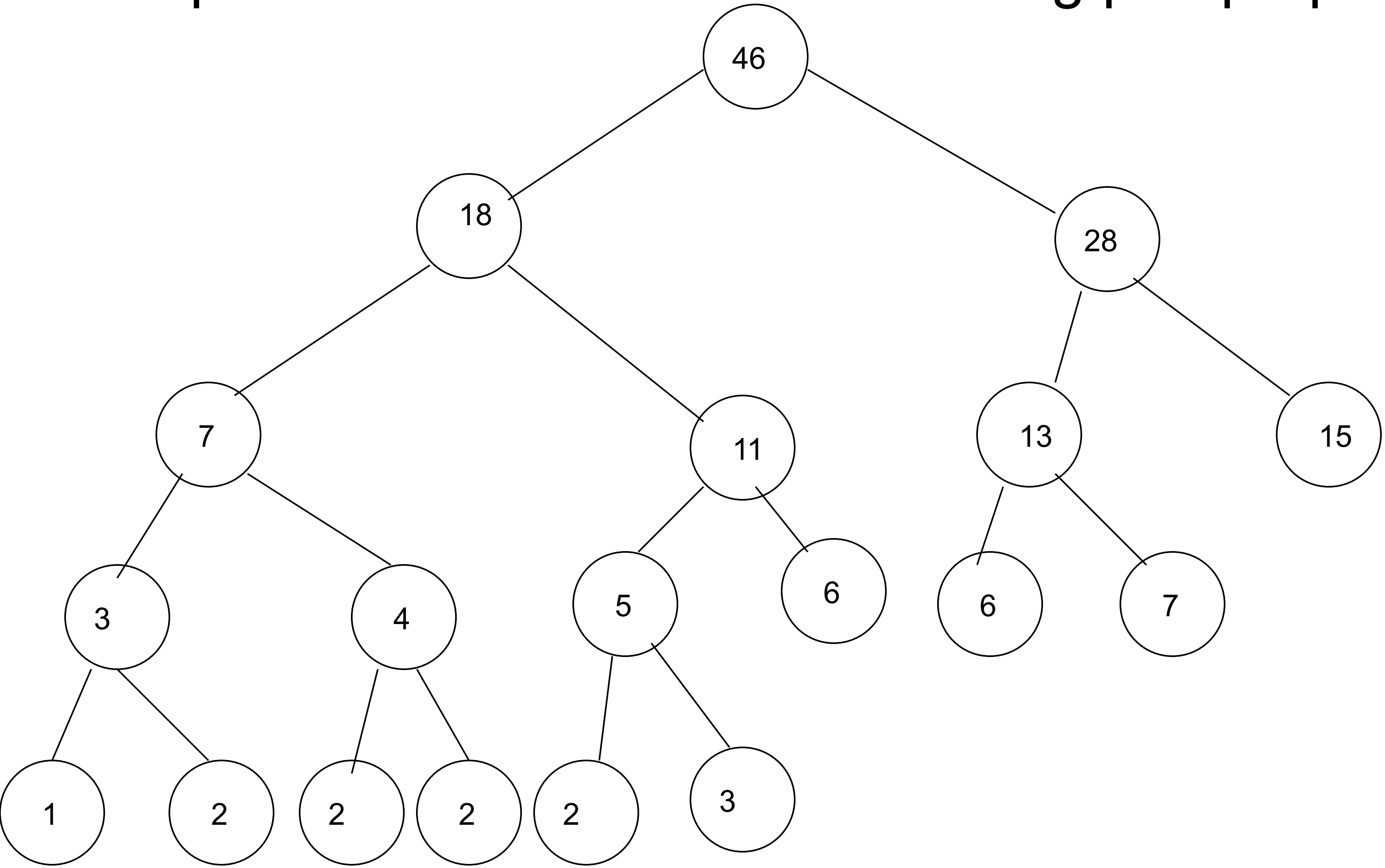
$$f_s = 160 \text{ MHz} = 1,6 * 10^8$$

performance

Kód	**Telj. kép**	**Átl. hossz**	**Adatátvseb**	**Alg. Kompl.**
Huffman	*L_opt*	2.33	372,8 Mbps	Ker.+ bin. fa
SF	$H(X)<L<H(X)+1$	2.89	462,4 Mbps	bin. fa
SFE	$H(X)<L<H(X)+2$	3.89	622,4 Mbps	Bin. konv.

alg. simplicity

Adaptive Huffman codes – the sibling pair property



1,2,2,2,2,3; 3,4,5,6,6,7; 7,11,13,15; 18,28; 46 monotone increasing sequence from bottom layer to top layer

Every parent node is the sum of children nodes

Construction of adaptive Huffman codes