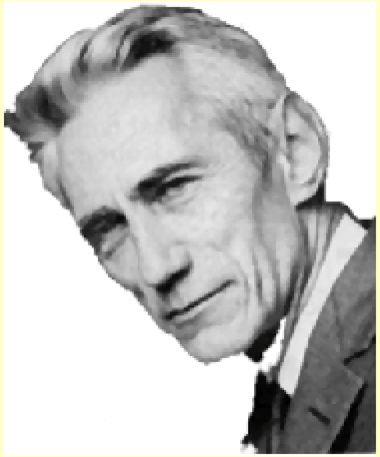
Theory of Communication

Information Measures:

- ***Channel capacity:*** C ***bit/sec***
- ***Source entropy:*** H ***Sh/cycle***
- ***Source rate:*** r ***cycle/sec***

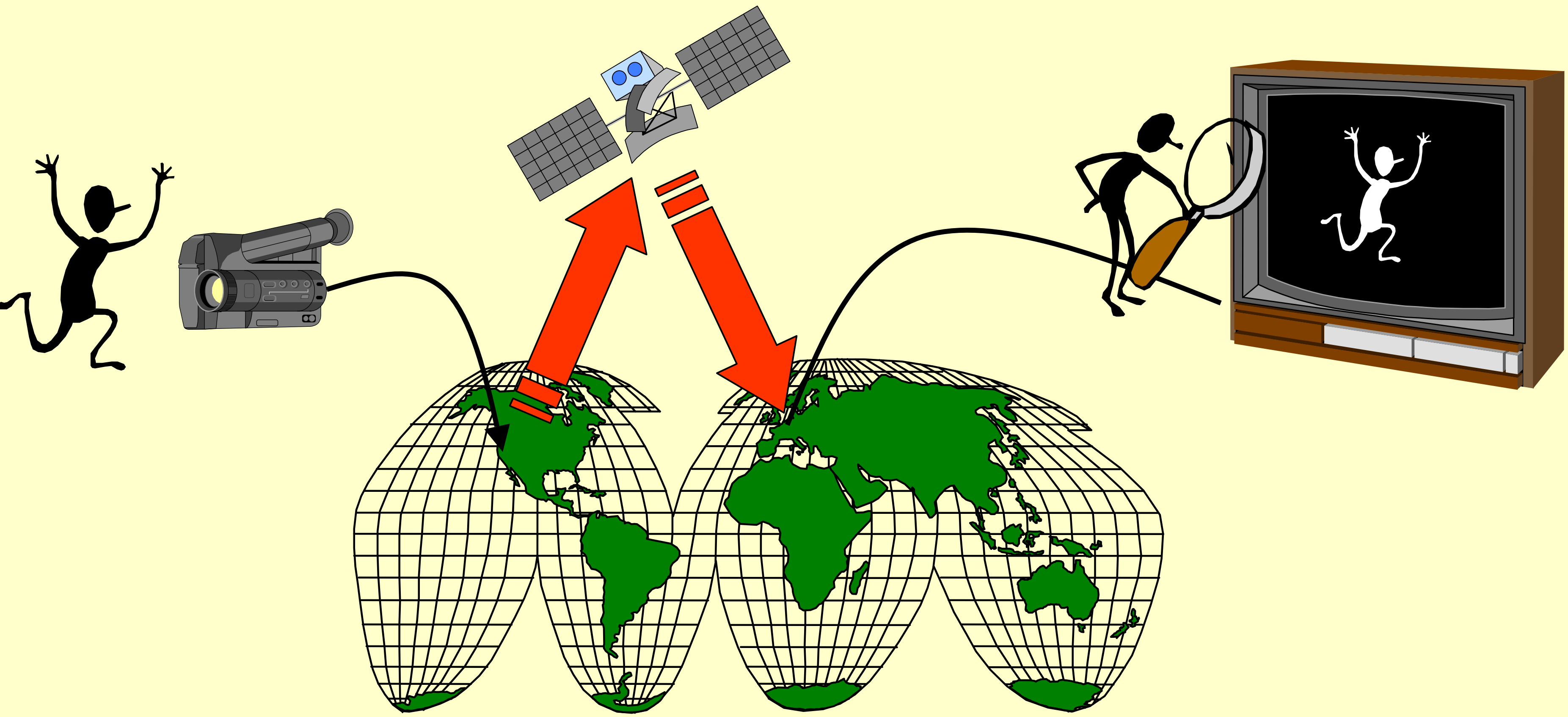
Theorem (Shannon): Communication, without error, is

- possible, when C>Hr;
- not possible, when C<Hr.

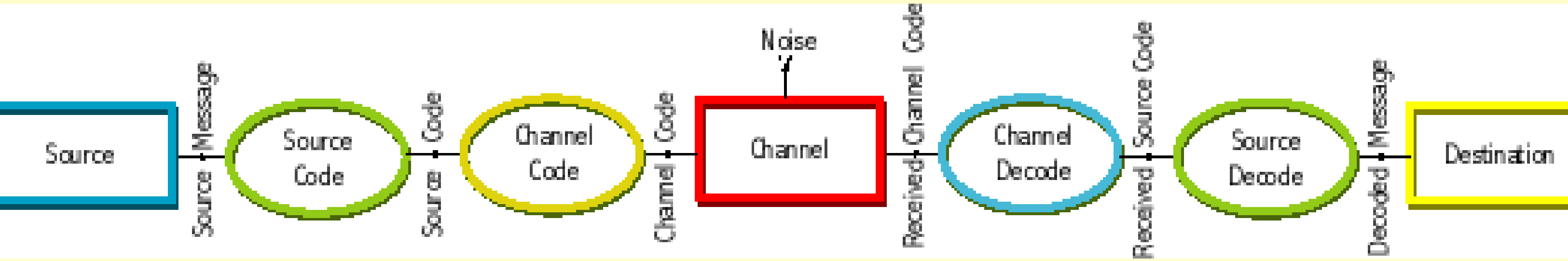


Shannon 1948

Application



Shannon's Proof



Independent problems:

1. Entropy Code for the Source
2. Error Control Code for the Channel

Entropy Coding

Source $S = s_0\ s_1\ s_2\ ..s_N\ \ldots$

- **Stochastic model:** $Pr(s_0 \ldots s_N) = p_N$
- **Memory-less:** $Pr(s_0 \ldots s_N) = \prod Pr(s_k)$
- **Entropy:** $H(S) = \sum p_N \log(1/\ p_N)$

Code $C = c_0\ c_1\ c_2\ ..c_N\ \ldots$

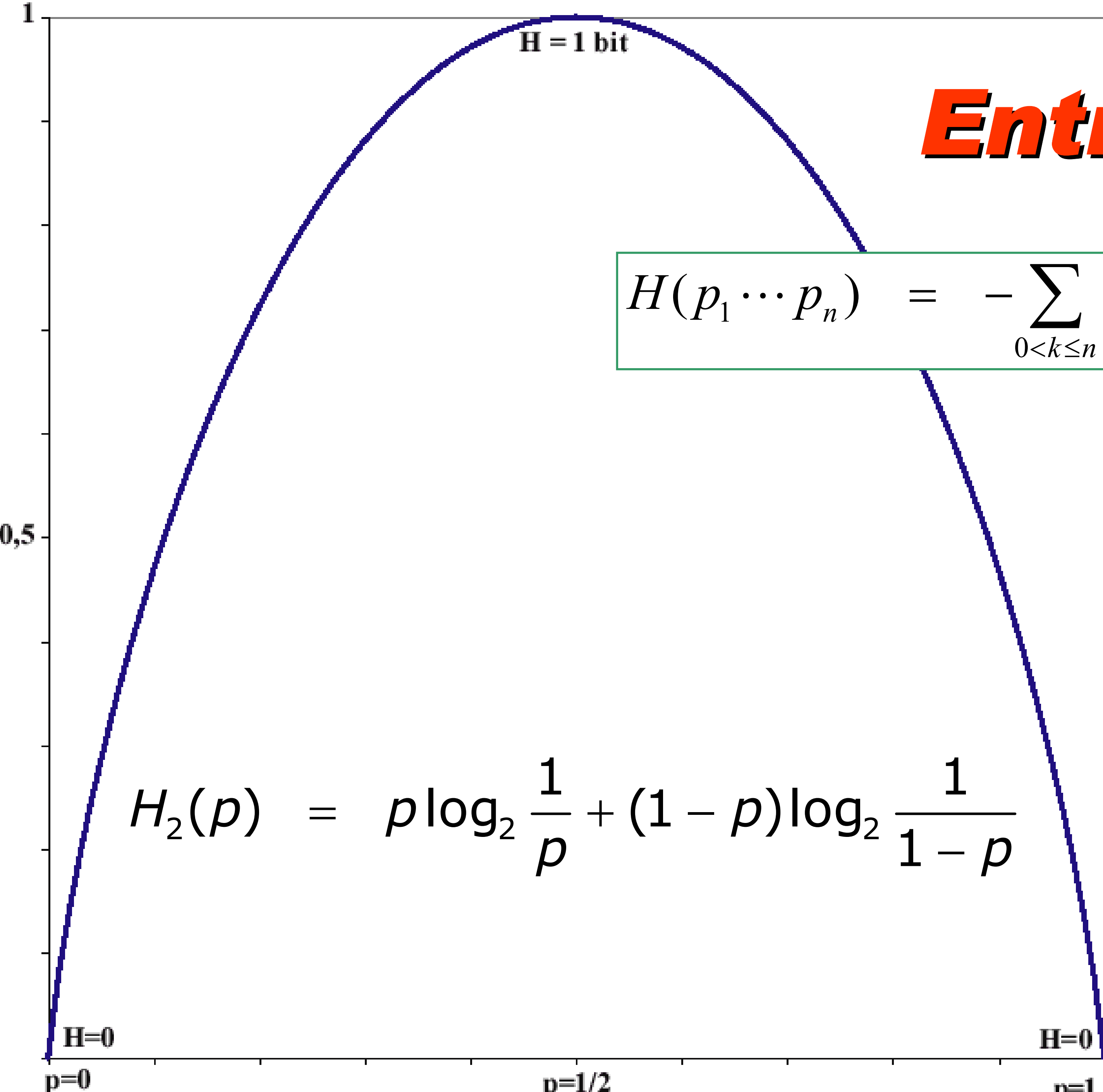
Average code length $|C_N| = 1/N \sum_{k<N} p_k\ |c_k|$

Theorem (Shannon) At all cycle N:

- **For all codes:** $|C_N| >= H(S)$
- **There exists codes:** $|C_N| < H(S) + 1/N$

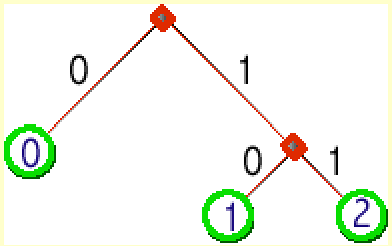
Entropy

$$H(p_1 \cdots p_n) \;=\; -\sum_{0<k\leq n} p_k \log_2 p_k$$



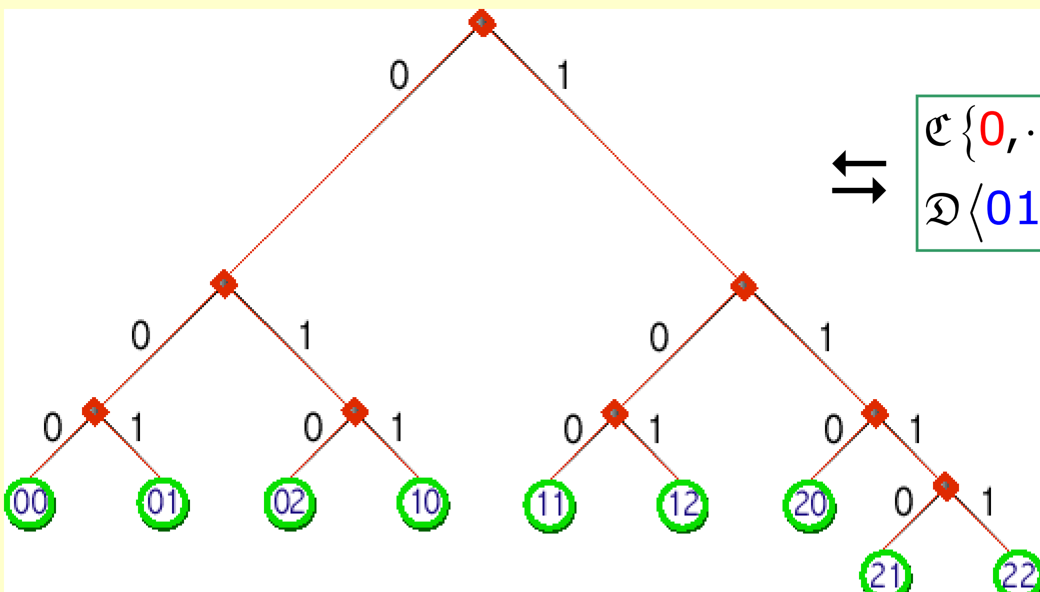
$$H_2(p) \;=\; p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$$

Decision Tree $\leftrightarrows$ Prefix Code



$\leftrightarrows$

$$\mathfrak{C}\{0,1,2\} = \{0,10,11\}$$
$$\mathfrak{D}\langle 0110010\rangle = \langle 02001\rangle$$



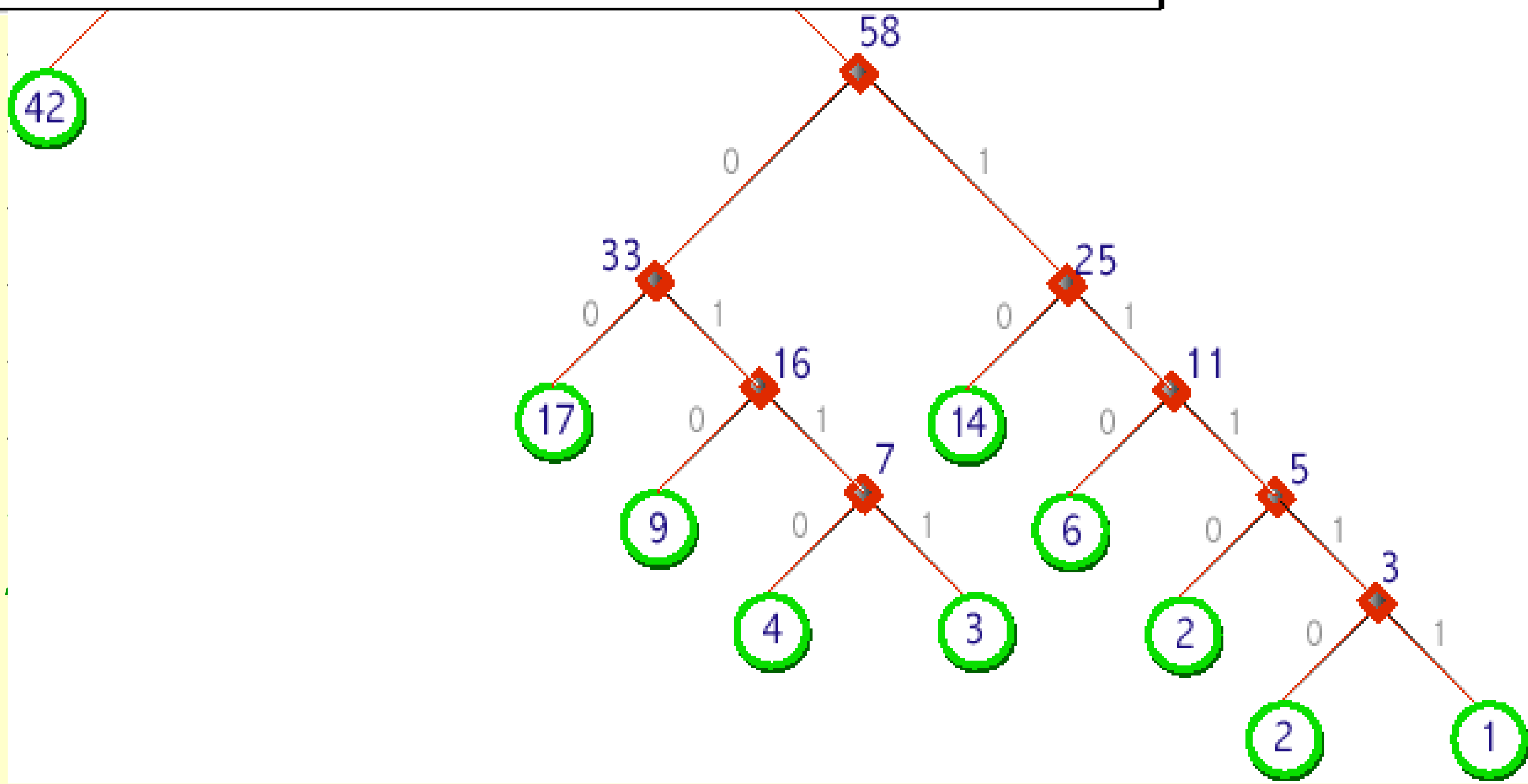
$\leftrightarrows$

$$\mathfrak{C}\{0,\cdots 6,7,8\} = \{000,\cdots 110,1110,1111\}$$
$$\mathfrak{D}\langle 011000101\rangle = \langle 305\rangle$$

Huffman's Algorithm

N	1	2	3	4	5	6	7	8	9	>9
$100 P_N$	42	17	9	6	4	3	2	2	1	14
huff(N)	1	3	4	4	5	5	5	6	6	3
bits(N)	1.3	2.6	3.4	4.1	4.6	5.1	5.5	5.8	6.1	2.9

42 17
42 17
42 17
42 17
42 17
42 17
42 25
42 33
58 42
100



Source Coding

Kraft's Inequality

The code lengths $l_k = |c_k|$ of a prefix code $(c_1 \cdots c_N)$ satisfy:

$$\sum_{1 \le k \le N} 2^{-l_k} \le 1.$$

Huffman's Algorithm

Let $(c_1 \cdots c_N) = H_u(q_1 \cdots q_N)$ be the code constructed by Huffman's algorithm, for input distribution $\sum_{1 \le k \le N} q_k \le 1$.

If $l_k = \log_2 \frac{1}{q_k}$ is an integer, then $|c_k| \le l_k$ for $1 \le k \le N$.

Gibbs Inequality

$\sum_{1 \le k \le N} q_k \le 1$ and $\sum_{1 \le k \le N} p_k = 1$ implies:

$$\sum_{1 \le k \le N} p_k \log_2 \frac{1}{p_k} \le \sum_{1 \le k \le N} p_k \log_2 \frac{1}{q_k}.$$

Shannon's Source Theorem

For any prefix code $(c_1 \cdots c_N)$ and distribution $\sum_{1 \le k \le N} p_k = 1$:

$$H = \sum_{1 \le k \le N} p_k \log_2 \frac{1}{p_k} \le \sum_{1 \le k \le N} p_k |c_k| = |C|.$$

The Huffman code $C = H_u(S^M)$ of an order M extension of the source S has length:

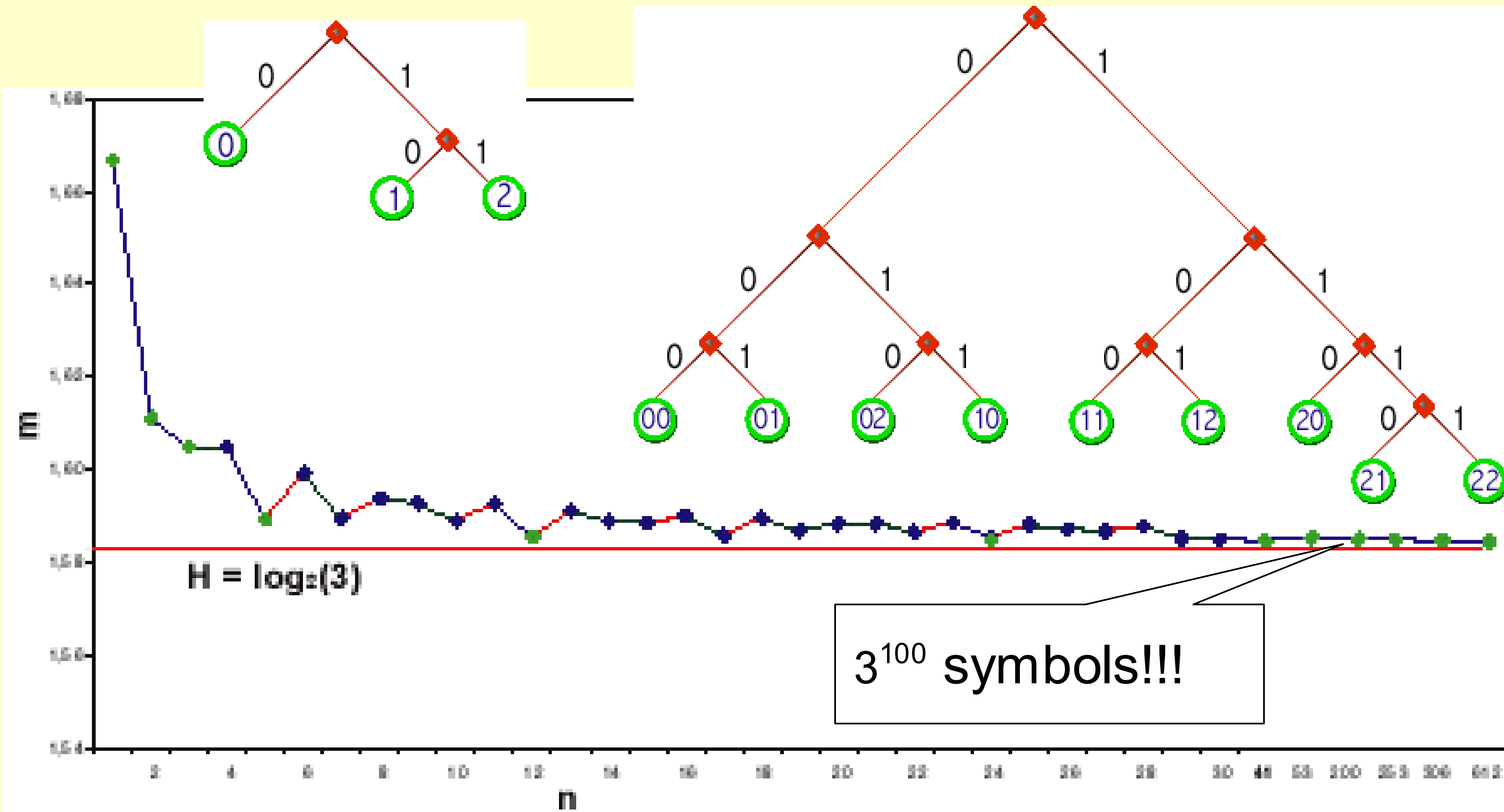
$$|C| \le H + \frac{1}{M}.$$

Ternary uniform source

Entropy: $H(1/3,1/3,1/3) = \log_2(3) = 1.5849...$

Huffman: $H_u(1/3,1/3,1/3) = 5/3 = 1.(6)$

Source Extension: $H_u(1/9,\dots,1/9) = 29/18 = 1.6(1)$



Entropy Coding

Reversible

- *Loss-less compression:* $D(C(m))=m$
- *Asymptotically optimal:* $|C|=H(S)_{+\varepsilon}$

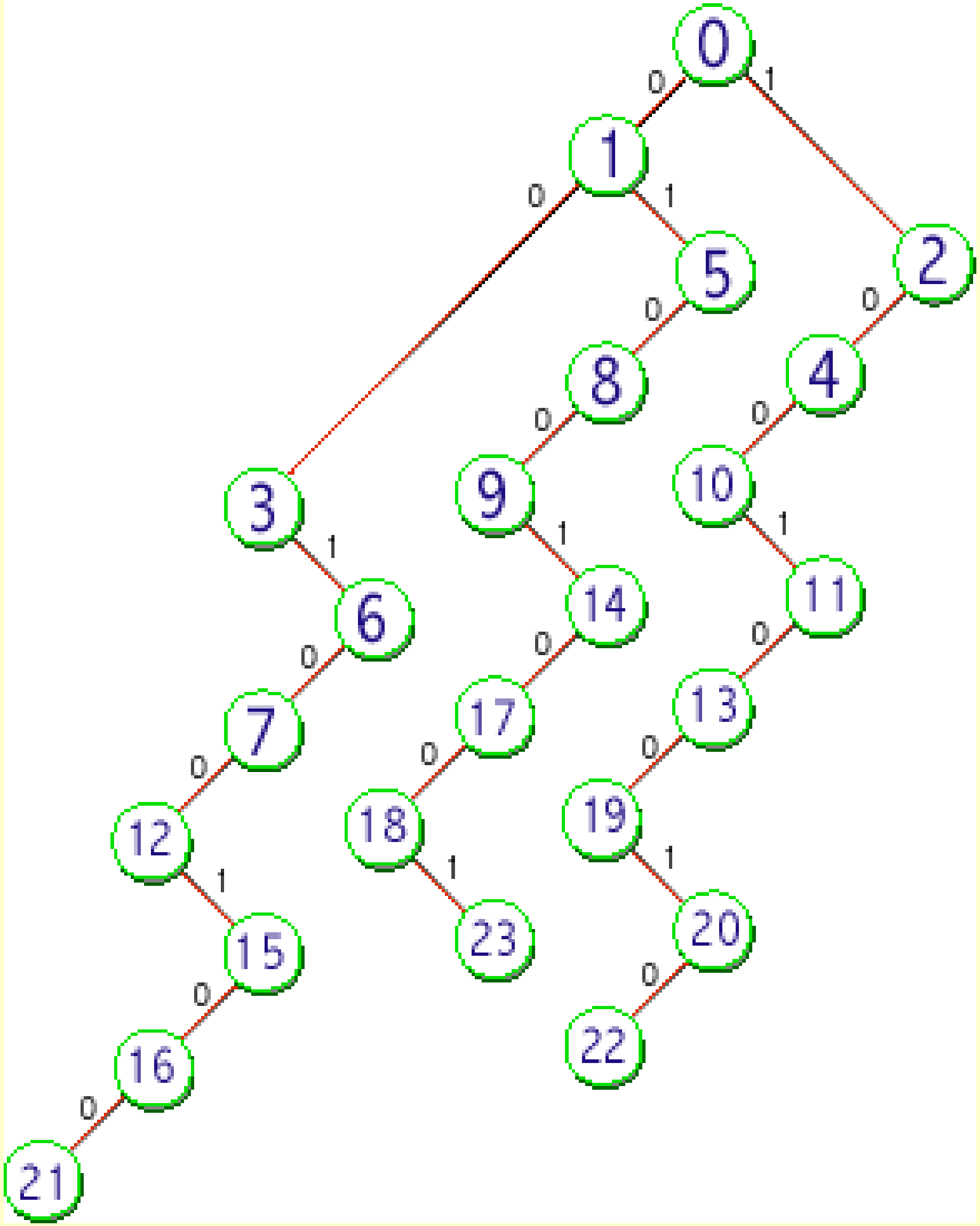
Stochastic model

- *Huffman code:* **block coding**
- *Arithmetic code:* continuous coding

Adaptive model

- *LZW* (Lempel, Ziv, Welsh)

LZW



Arithmetic Code