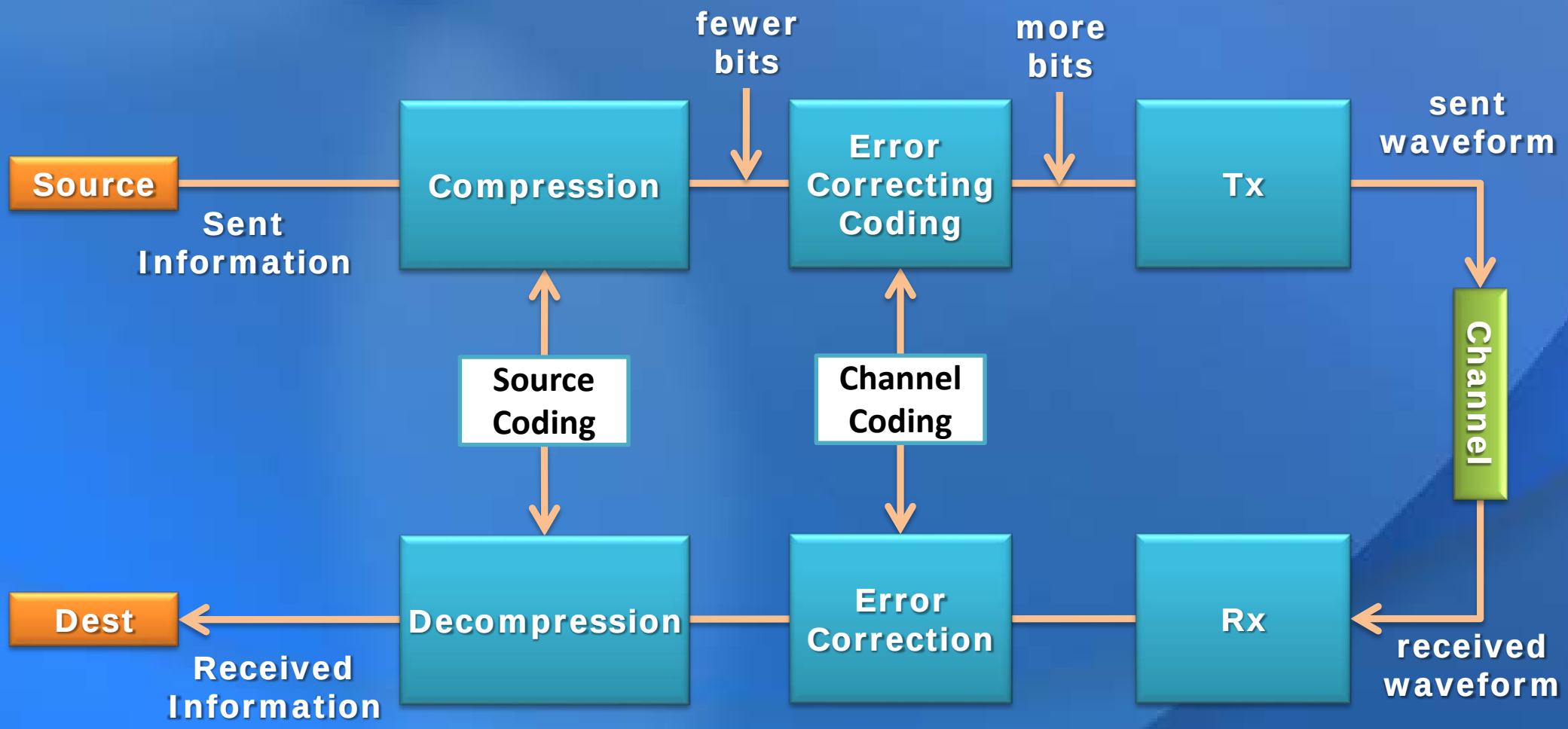
Source Coding

Communication System



Lossless vs. Lossy

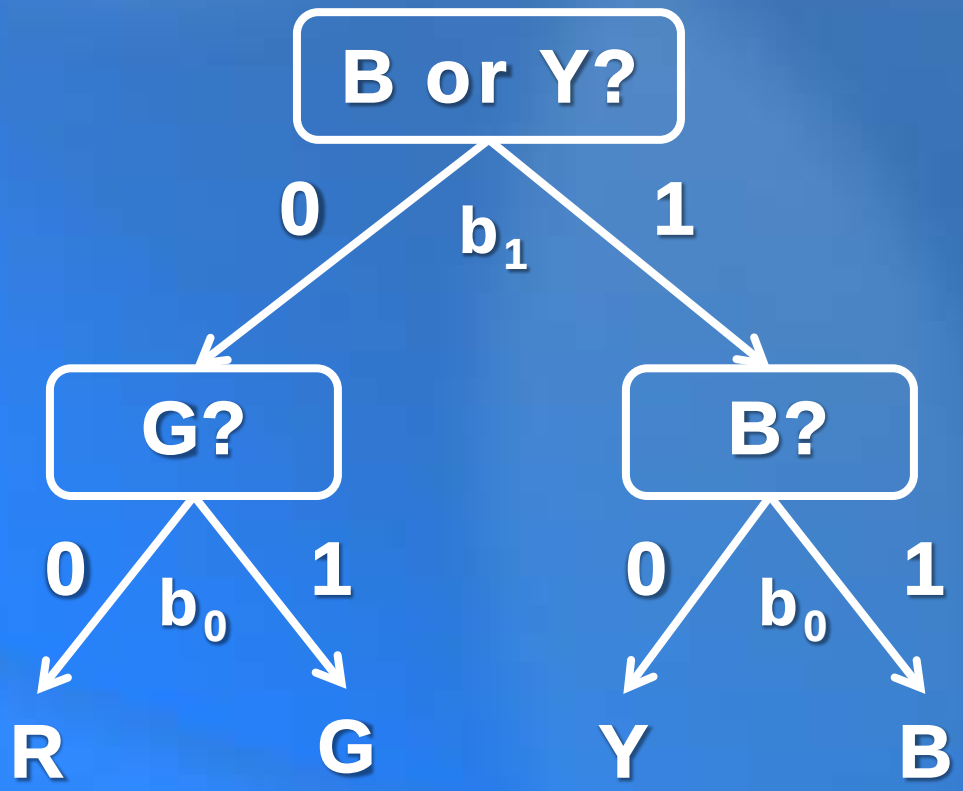
Lossless data compression

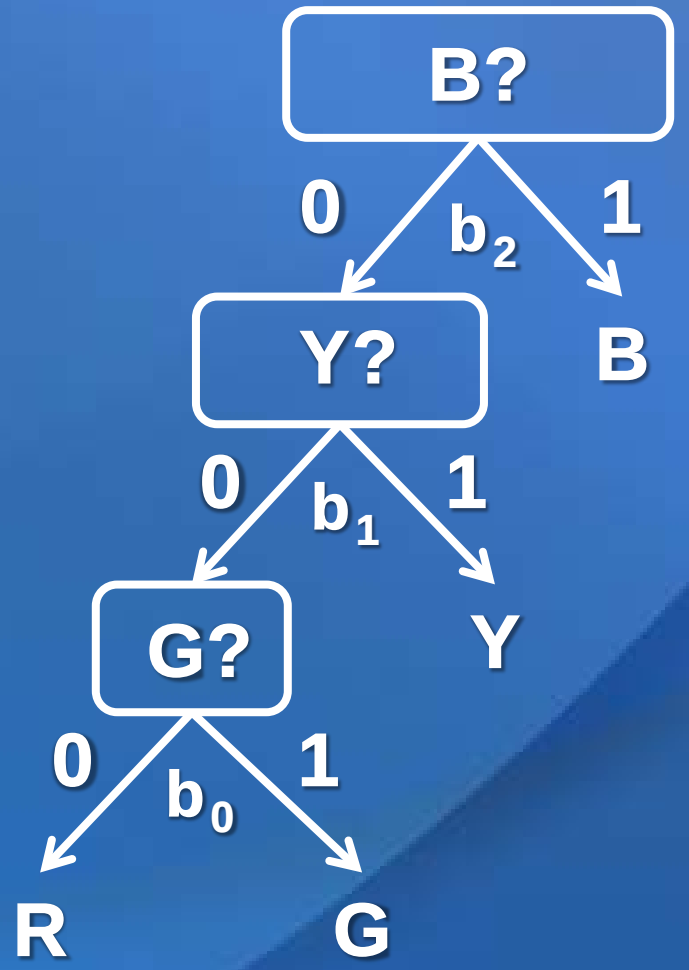
- OUT exactly same as IN
- Huffman coding

Lossy data compression

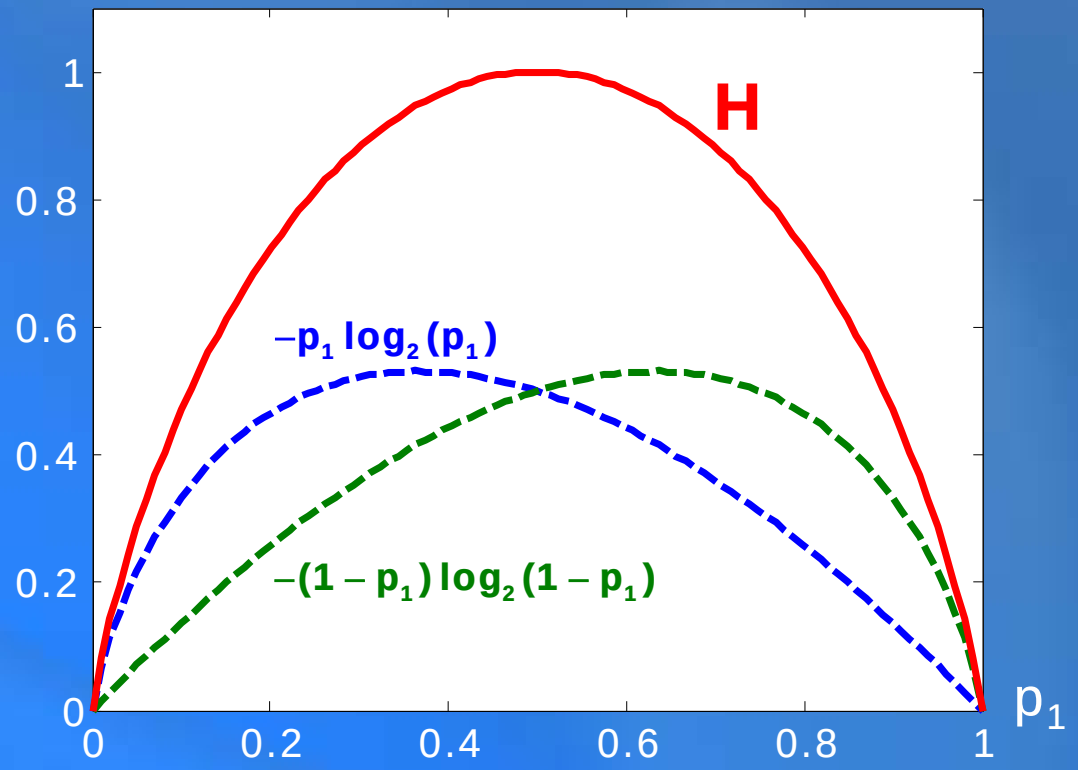
- OUT “close” or “similar” to IN
- MP3 coding

The Guessing Game





Entropy

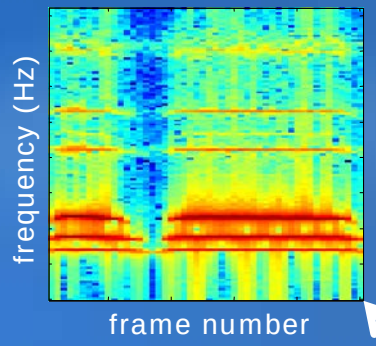


If $K=2$, $H = -(1-p_1)\log_2(1-p_1) - p_1 \log_2(p_1)$

$$H = -\sum_{k=0}^{K-1} p_k \log_2(p_k)$$

- A measure of the average information contained in a source randomly emitting K symbols.
- A lower bound on the average code length.

MP3 Coding



loss of information happens here

lossless compression

Input

Short-time Frequency Analysis

Non-Uniform Quantization

Huffman Coding

Output

16 bits per channel
× 2 channels
× 44.1kHz sampling rate
=
~1.41Mbps

Psycho Acoustic Model

~128kbps