Principles of Communications
ECS 332

Dr. Prapun Suksompong

prapun@siit.tu.ac.th

Source Coding

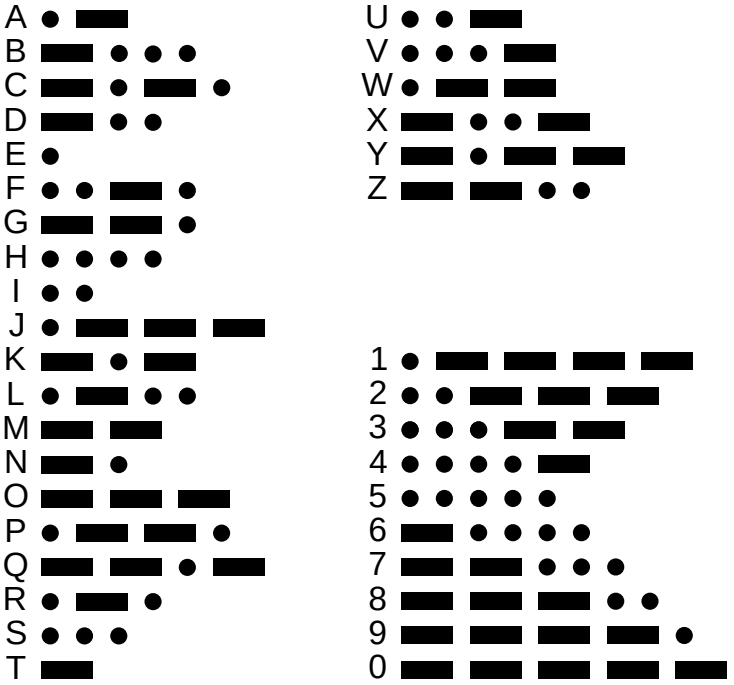
Office Hours:
BKD 3601-7

Monday	**14:40-16:00**
Friday	**14:00-16:00**

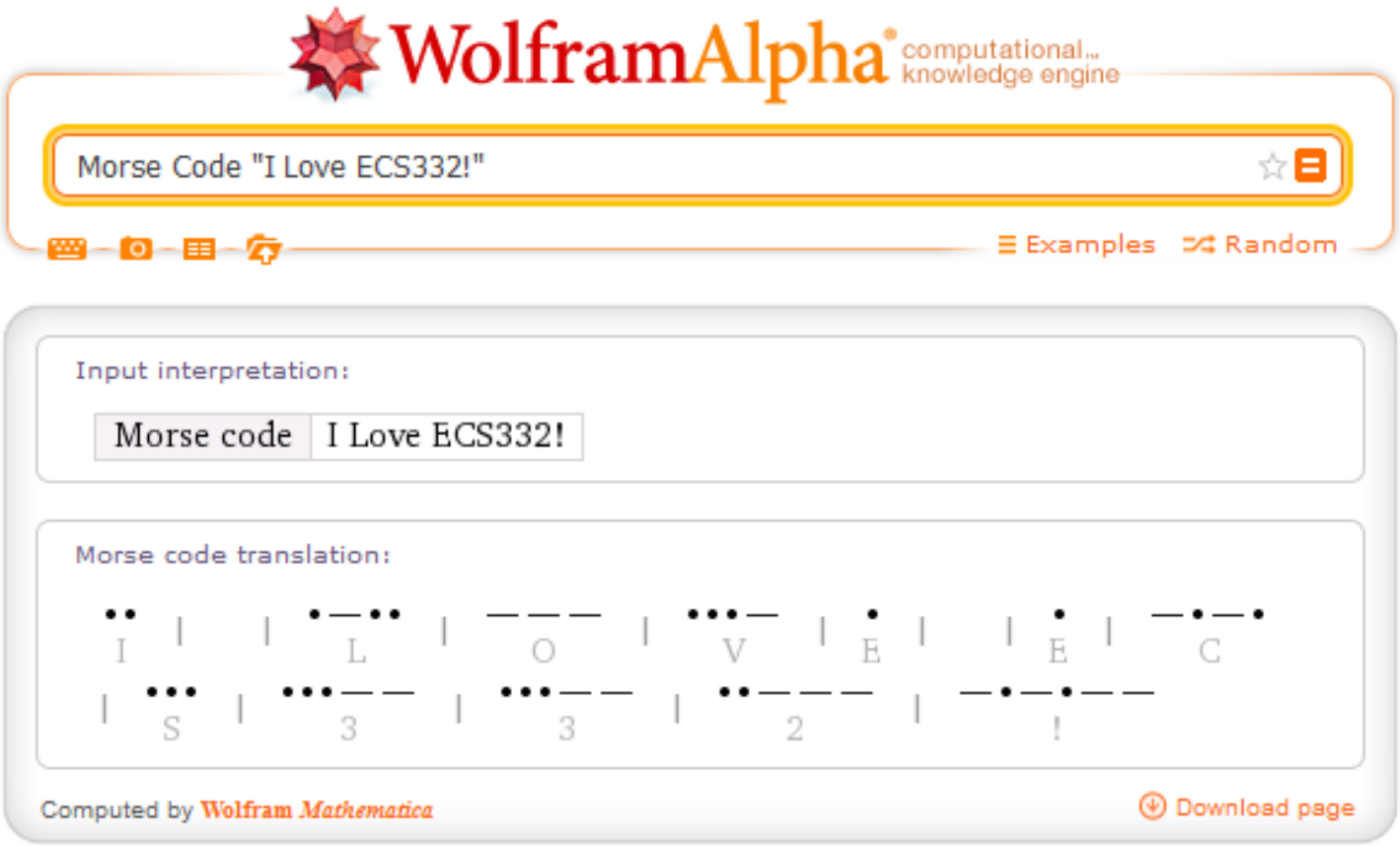
Morse code

(wired and wireless)

- **Telegraph network**
- Samuel **Morse**, 1838
- A sequence of on-off tones (or , lights, or clicks)



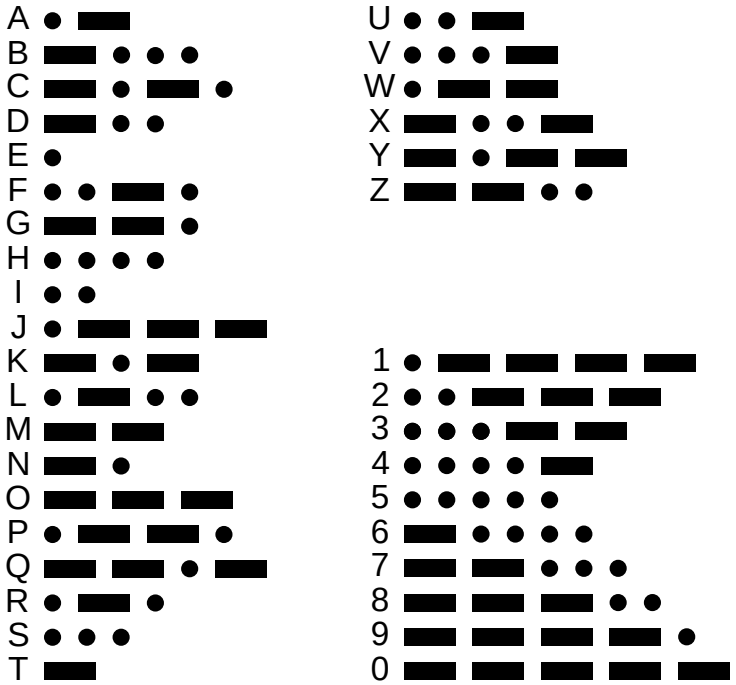
Example



[http://www.wolframalpha.com/input/?i=Morse+Code+%22I+Love+ECS332!%22]

Morse code: Key Idea

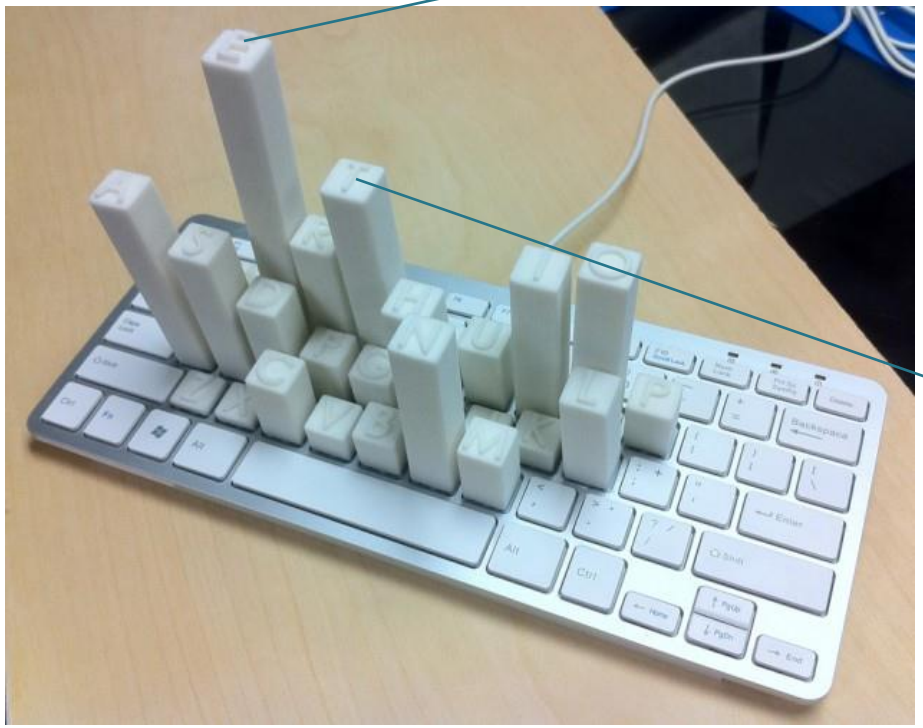
Frequently-used characters (e,t) are mapped to short codewords.

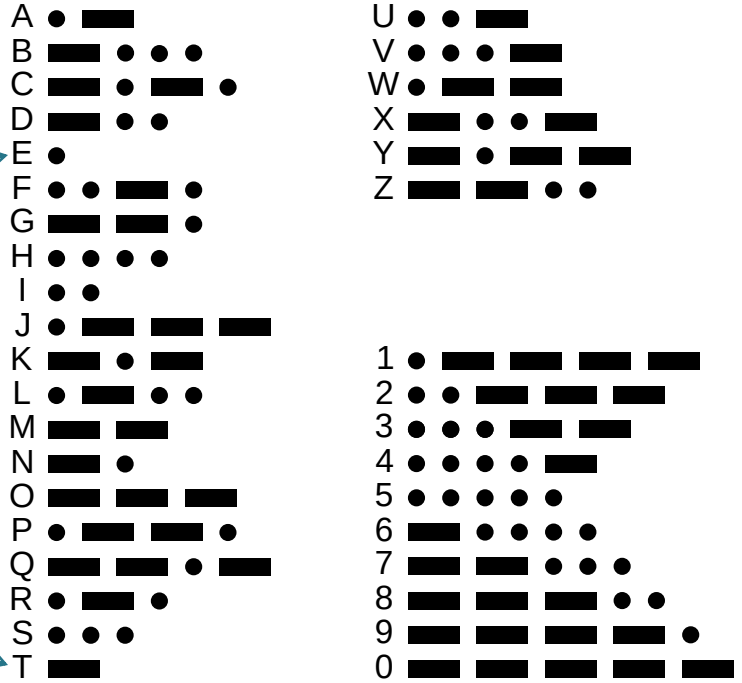


Basic form of compression.

Morse code: Key Idea

Frequently-used characters are mapped to short codewords.

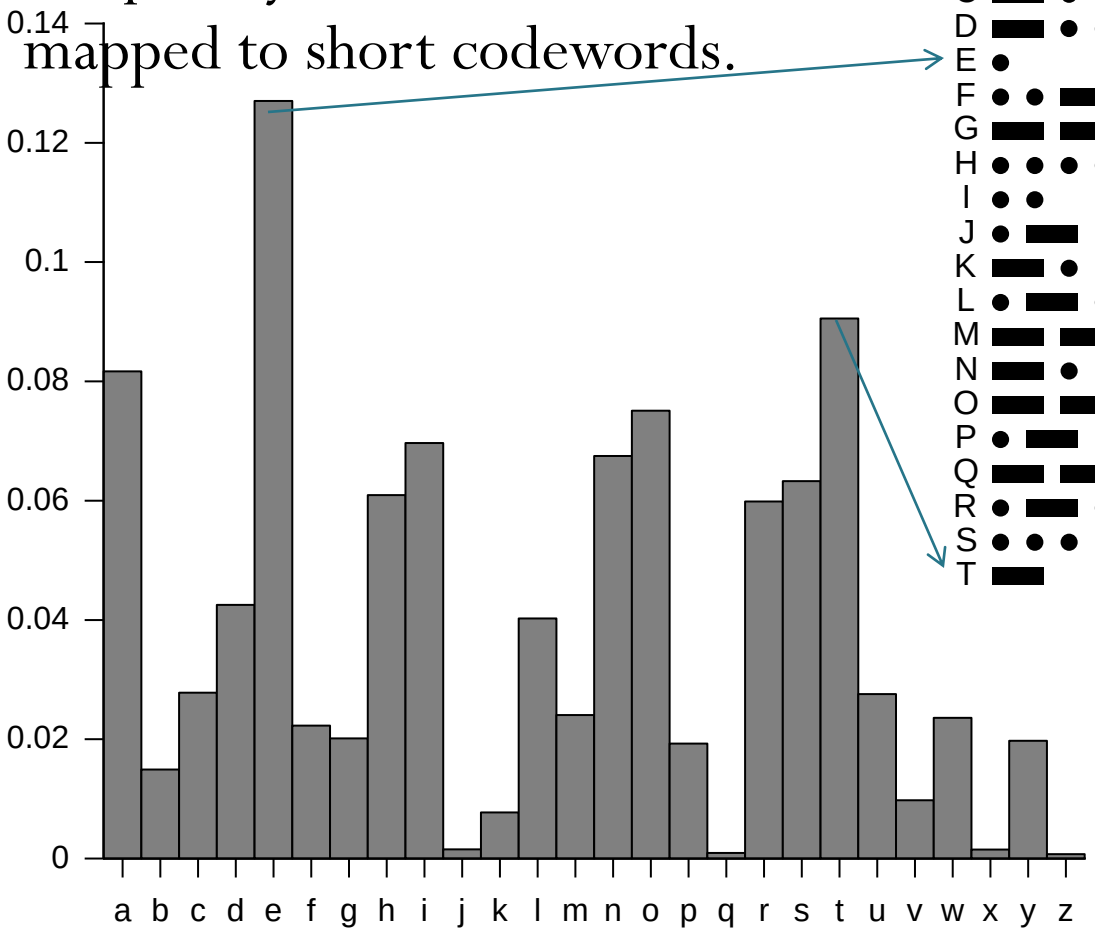


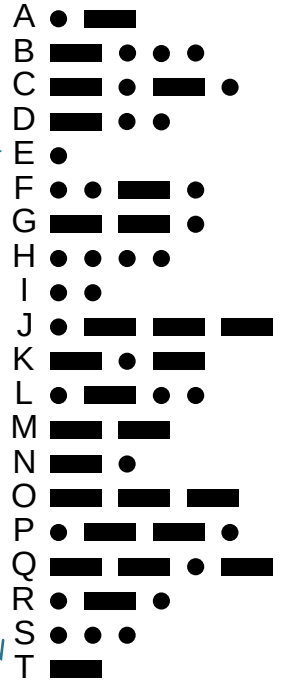


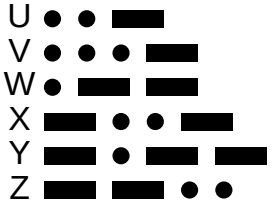
Relative frequencies of letters in the English language

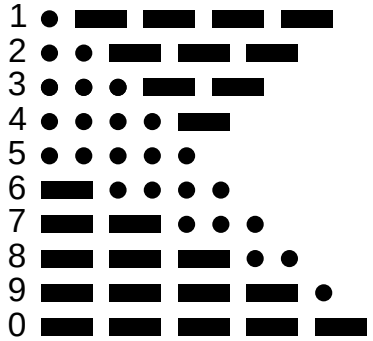
Morse code: Key Idea

Frequently-used characters are mapped to short codewords.









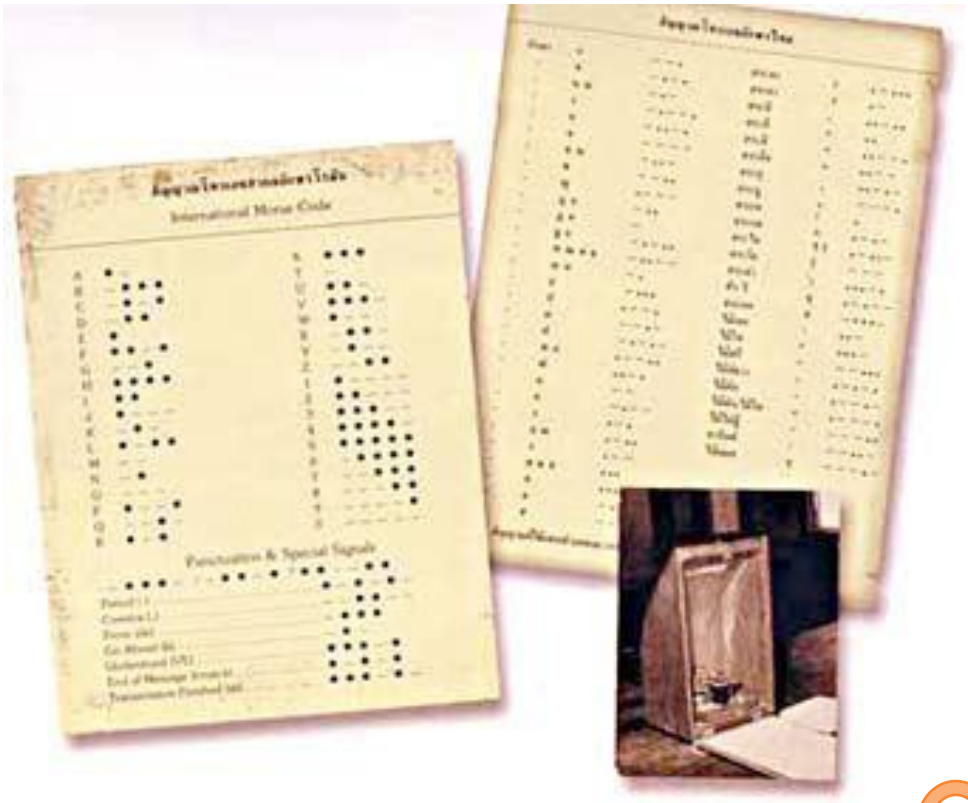
Relative frequencies of letters in the English language

รหัสมอร์สภาษาไทย

รหัสสัญญาณโทรเลขมอร์สภาษาไทย

ที่เริ่มใช้เมื่อ ๑ พฤศจิกายน ๒๔๕๕

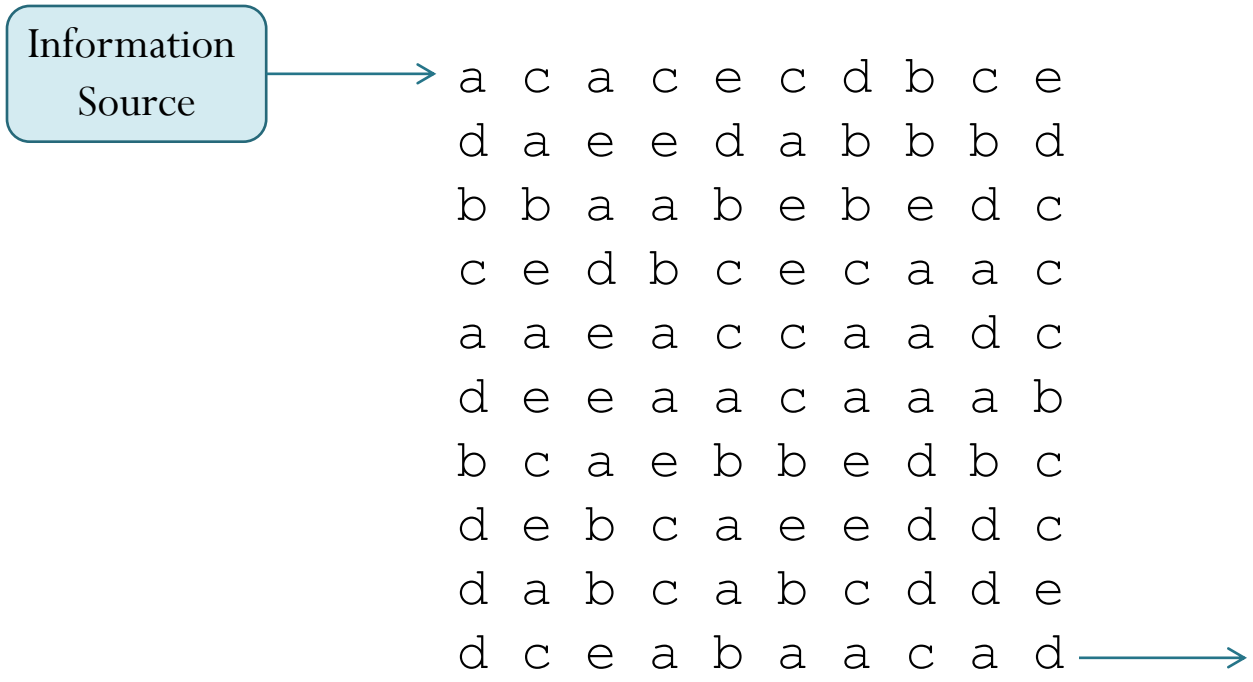
๑.	ก		๒๖.	ศ	ษ	ส	
๒.	ข		๒๗.	ห			
๓.	ค	ฆ	๒๘.	อ			
๔.	ง		๒๙.	ฮ			
๕.	จ		๓๐.	ฤ			
๖.	ฉ		๓๑.	ฯ ฤๅ			
๗.	ช	ฌ	๓๒.	ว			
๘.	ซ		๓๓.	ะ			
๙.	ญ		๓๔.	ั			
๑๐.	ฐ	ฏ	๓๕.	า			
๑๑.	ฑ	ฒ	๓๖.	ำ			
๑๒.	ฎ	ฐ	๓๗.	ิ			
๑๓.	ฏ	ต	ท	ธ	ฒ	๓๘.	ี
๑๔.	ณ	ฌ	๓๙.	ึ			
๑๕.	บ		๔๐.	ื			
๑๖.	ป		๔๑.	ุ	ู		
๑๗.	ผ		๔๒.	เ			
๑๘.	ฝ		๔๓.	แ			
๑๙.	พ	ภ	๔๔.	โ			
๒๐.	ฟ		๔๕.	ใ			(ไม้ยก)
๒๑.	ม		๔๖.	ไ			(ไม้ไต)
๒๒.	ย		๔๗.	็			(การันต์)
๒๓.	ร		๔๘.	่			(ไม้เอก)
๒๔.	ล	ฬ	๔๙.	้			
๒๕.	ว		๕๐.	ๆ			(ไม้ยมก)



Ex. DMS (1)

$$\mathcal{S}_X = \{a, b, c, d, e\}$$

$$p_X(x) = \begin{cases} 1/5, & x \in \{a, b, c, d, e\} \\ 0, & \text{otherwise} \end{cases}$$



Approximately 20% are letter 'a's

Ex. DMS (2)

$$\mathcal{S}_X = \{1, 2, 3, 4\}$$

$$p_X(x) = \begin{cases} 1/2, & x = 1, \\ 1/4, & x = 2, \\ 1/8, & x \in \{3, 4\} \\ 0, & \text{otherwise} \end{cases}$$

Information Source →

```
2 1 1 2 1 4 1 1 1 1
1 1 4 1 1 2 4 2 2 1
3 1 1 2 3 2 4 1 2 4
2 1 1 2 1 1 3 3 1 1
1 3 4 1 4 1 1 2 4 1
4 1 4 1 2 2 1 4 2 1
4 1 1 1 1 2 1 4 2 4
2 1 1 1 2 1 2 1 3 2
2 1 1 1 1 1 1 2 3 2
2 1 1 2 1 4 2 1 2 1
```

→

Approximately 50% are number '1's

Shannon–Fano coding

- Proposed in Shannon's "A Mathematical Theory of Communication" in 1948
- The method was attributed to Fano, who later published it as a technical report.
- Should not be confused with
  - Shannon coding, the coding method used to prove Shannon's noiseless coding theorem, or with
  - Shannon–Fano–Elias coding (also known as Elias coding), the precursor to arithmetic coding.

Huffman Code

- MIT, 1951
- Information theory class taught by Professor Fano.
- Huffman and his classmates were given the choice of
  - a term paper on the problem of finding the most efficient binary code.

  or
  - a final exam.
- Huffman, unable to prove any codes were the most efficient, was about to give up and start studying for the final when he hit upon the idea of using a frequency-sorted binary tree and quickly proved this method the most efficient.
- Huffman avoided the major flaw of the suboptimal Shannon-Fano coding by building the tree from the bottom up instead of from the top down.

Huffman Coding in MATLAB (1)

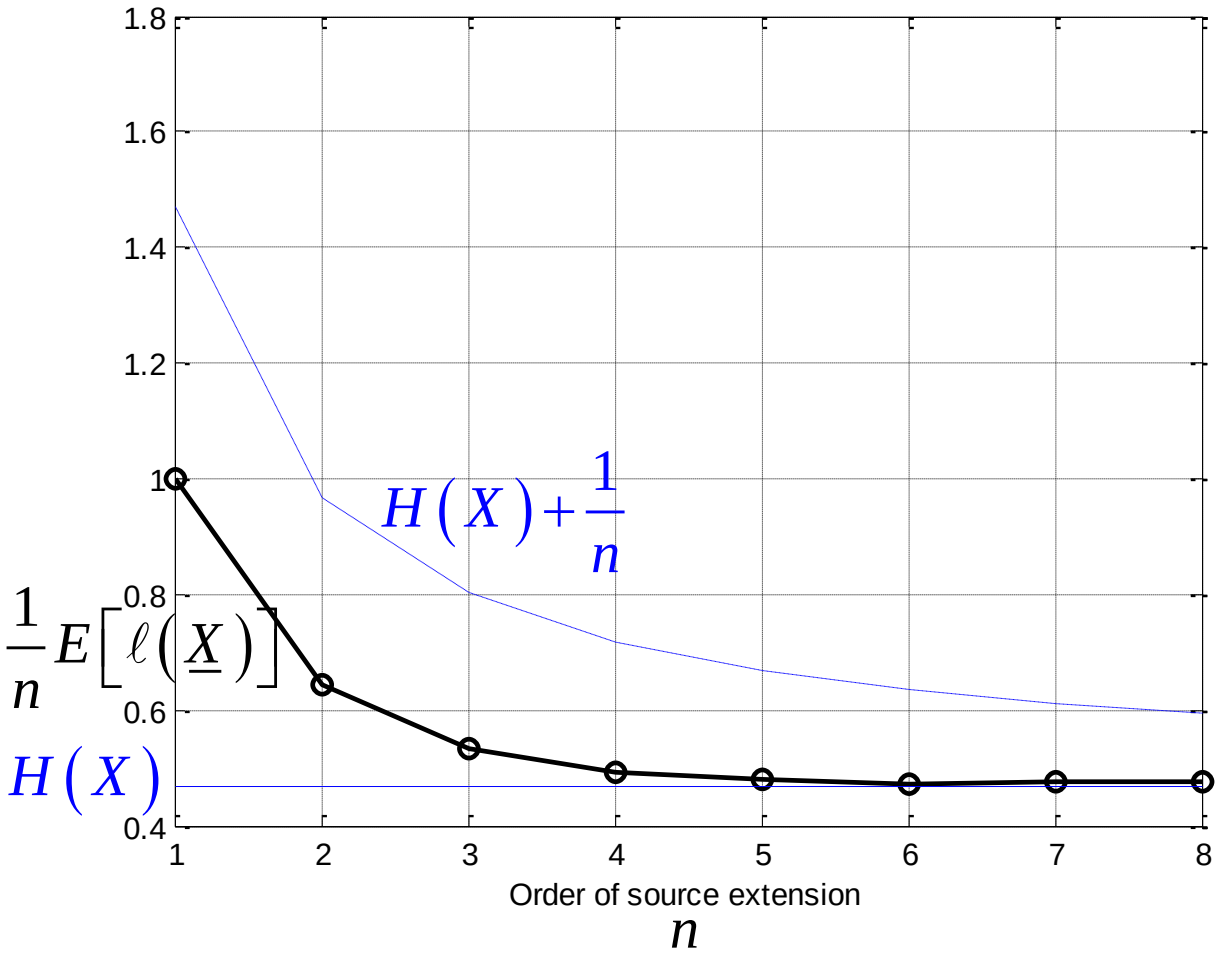
```
pX = [0.5 0.25 0.125 0.125];          % pmf of X
SX = [1:length(pX)];                  % Source Alphabet
[dict,EL] = huffmandict(SX,pX);       % Create codebook

%% Pretty print the codebook.
codebook = dict;
for i = 1:length(codebook)
   codebook{i,2} = num2str(codebook{i,2});
end
codebook

% Try to encode some random source string
n = 5; % Number of source symbols to be generated
sourceString = randsrc(1,10,[SX; pX]) % Create data using pX
encodedString = huffmanenco(sourceString,dict) % Encode the data
```


[Huffman_Demo_Ex1]

Huffman Coding in MATLAB (2)

```
codebook =

    [1]    '0'
    [2]    '1  0'
    [3]    '1  1  1'
    [4]    '1  1  0'


sourceString =

     1     4     4     1     3     1     1     4     3     4


encodedString =

     0     1     1     0     1     1     0     0     1     1     1     0     0     1     1     0     1     1     1     1     1     0
```

Huffman Coding: Source Extension



Huffman Coding: Uniform pmf

(no source extension)