

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)



A	● —	U	● ● —
B	— ● ● ●	V	● ● ● —
C	— — ● — ●	W	● — — —
D	— — ● ●	X	— ● ● — —
E	●	Y	— — — — ●
F	● ● — — ●	Z	— — — ● ●
G	— — — ●		
H	● ● ● ●		
I	● ●		
J	● — — — —		
K	— ● ● — —	1	● — — — — —
L	● — — ● ●	2	● ● — — — —
M	— — — —	3	● ● ● — — —
N	— — ●	4	● ● ● ● — —
O	— — — — —	5	● ● ● ● ●
P	● — — — — ●	6	— ● ● ● ● ●
Q	— — — ● — —	7	— — — ● ● ●
R	● — — — ●	8	— — — — ● ●
S	● ● ● ●	9	— — — — — ●
T	— — — —	0	— — — — — —



Example

The image shows a screenshot of the WolframAlpha website. At the top, the WolframAlpha logo is displayed with the tagline "computational... knowledge engine". Below the logo is a search bar containing the text "Morse Code 'I Love ECS332!'". To the right of the search bar are icons for a star and a menu. Below the search bar are several navigation icons and the text "Examples Random".

The main content area is divided into two sections. The first section is titled "Input interpretation:" and contains a text box with the text "Morse code I Love ECS332!". The second section is titled "Morse code translation:" and displays the Morse code for the input text. The Morse code is shown in two rows, with the letters and symbols aligned vertically. The first row shows the letters I, L, O, V, E, E, C. The second row shows the letters S, 3, 3, 2, !.

At the bottom of the page, there is a footer that says "Computed by Wolfram Mathematica" and a "Download page" link with a download icon.

Morse code: Key Idea

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

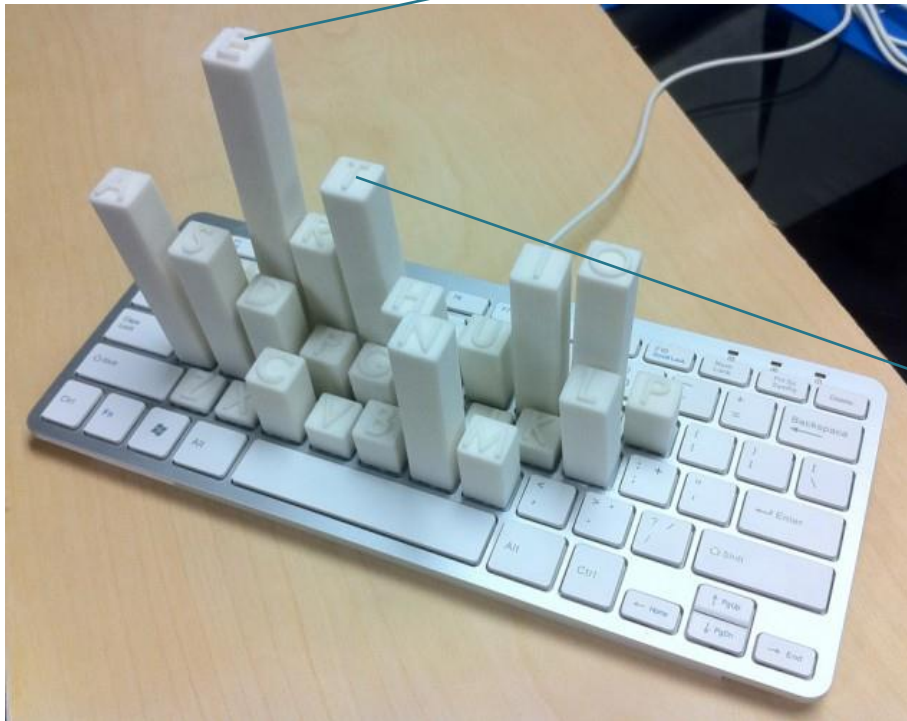
A	• —	U	• • —
B	— • • •	V	• • • —
C	— • — •	W	• — —
D	— • •	X	— • • —
E	•	Y	— • — —
F	• • — •	Z	— — • •
G	— — •		
H	• • • •		
I	• •		
J	• — — —		
K	— • — —		
L	• — • •		
M	— —		
N	— •		
O	— — —		
P	• — — •		
Q	— — • —		
R	• — •		
S	• • •		
T	—		
		1	• — — — —
		2	• • — — —
		3	• • • — —
		4	• • • • —
		5	• • • • •
		6	— • • • •
		7	— — • • •
		8	— — — • •
		9	— — — — •
		0	— — — — —

Basic form of compression.



Morse code: Key Idea

Frequently-used characters are mapped to short codewords.



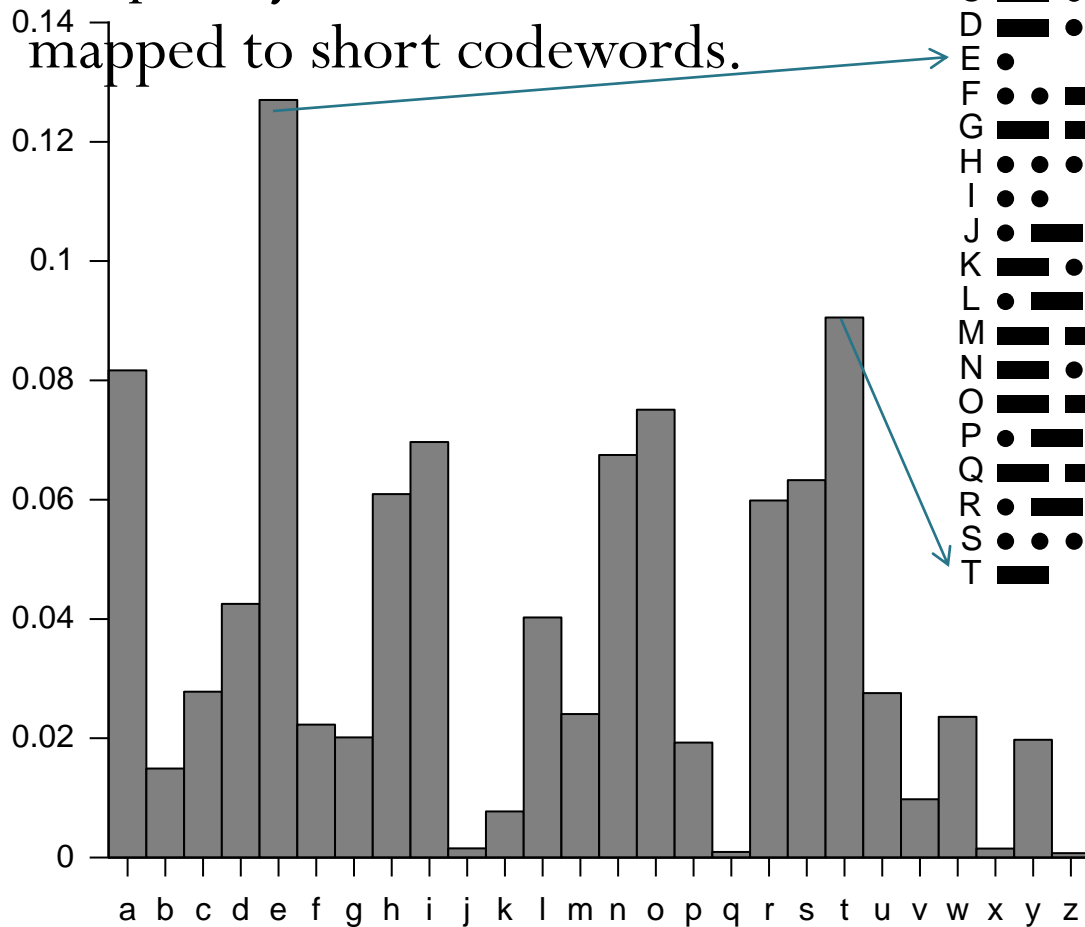
A	● —	U	● ● —
B	— ● ● ●	V	● ● ● —
C	— ● — ●	W	● — — —
D	— ● ●	X	— ● ● —
E	●	Y	— ● — —
F	● ● — ●	Z	— — ● ●
G	— — ●		
H	● ● ● ●		
I	● ●		
J	● — — —		
K	— ● — —		
L	● — ● ●		
M	— — —		
N	— ●		
O	— — — —		
P	● — — ●		
Q	— — ● —		
R	● — ● ●		
S	● ● ●		
T	—		
		1	● — — — —
		2	● ● — — —
		3	● ● ● — —
		4	● ● ● ● —
		5	● ● ● ● ●
		6	— ● ● ● ●
		7	— — ● ● ●
		8	— — — ● ●
		9	— — — — ●
		0	— — — — —

Relative frequencies
of letters in the
English language



Morse code: Key Idea

Frequently-used characters are mapped to short codewords.



A ● —
 B — ● ● ●
 C — ● — ●
 D — ● ●
 E ●
 F ● ● — ●
 G — — ●
 H ● ● ● ●
 I ● ●
 J ● — — — ●
 K — ● — —
 L ● — ● ●
 M — — —
 N — ●
 O — — — —
 P ● — — ● ●
 Q — — — ● —
 R ● — ● ●
 S ● ● ●
 T —

U ● ● —
 V ● ● ● —
 W ● — — —
 X — ● ● — —
 Y — ● — — —
 Z — — ● ●

1 ● — — — —
 2 ● ● — — —
 3 ● ● ● — —
 4 ● ● ● ● —
 5 ● ● ● ● ●
 6 — ● ● ● ●
 7 — — ● ● ●
 8 — — — ● ●
 9 — — — — ●
 0 — — — — —

Relative frequencies of letters in the English language

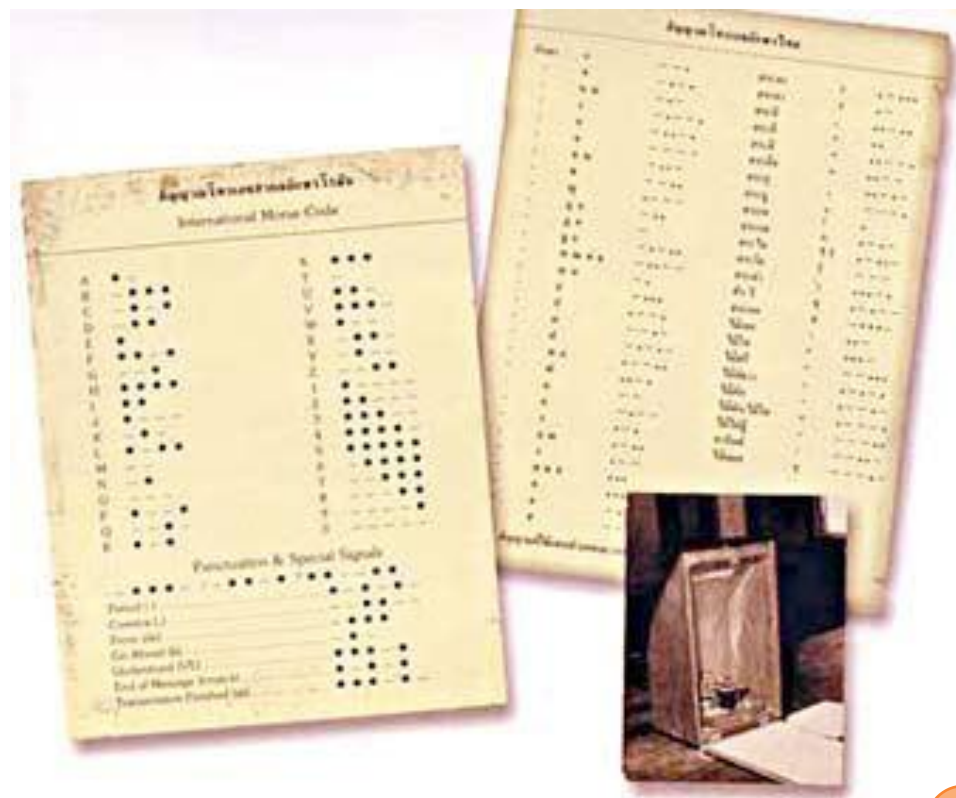


รหัส摩อรัสมษาไทย



รหัสสัญญาณโทรเลขของประเทศไทย
ที่เริ่มใช้เมื่อ 1 พฤศจิกายน 2455

1. ---	ก	26. ---	ค ข ฌ
2. ---	ข	27. ---	ท
3. ---	ค ฌ	28. ---	ด
4. ---	ด	29. ---	ธ
5. ---	ง	30. ---	ถ
6. ---	ฉ	31. ---	ฬ
7. ---	ช ฌ	32. ---	ง
8. ---	ช	33. ---	-
9. ---	ญ	34. ---	-
10. ---	ค ฌ	35. ---	-
11. ---	ค ฌ	36. ---	-
12. ---	ด ฌ	37. ---	-
13. ---	ท ก ฌ ฌ	38. ---	-
14. ---	ท ฌ	39. ---	-
15. ---	น	40. ---	-
16. ---	บ	41. ---	-
17. ---	ป	42. ---	-
18. ---	ผ	43. ---	-
19. ---	พ ก	44. ---	-
20. ---	ม	45. ---	ไฟตก
21. ---	น	46. ---	ไฟโท
22. ---	น	47. ---	การันตี
23. ---	ง	48. ---	ไฟตก
24. ---	ค ฌ	49. ---	-
25. ---	ง	50. ---	ไฟตก



Ex. DMS (1)

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

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

Information
Source

a c a c e c d b c e
d a e e d a b b b d
b b a a b e b e d c
c e d b c e c a a c
a a e a c c a a d c
d e e a a c a a a b
b c a e b b e d b c
d e b c a e e d d c
d a b c a b c d d e
d c e a b a a c a d

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 Coding in MATLAB (2)

codebook =

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

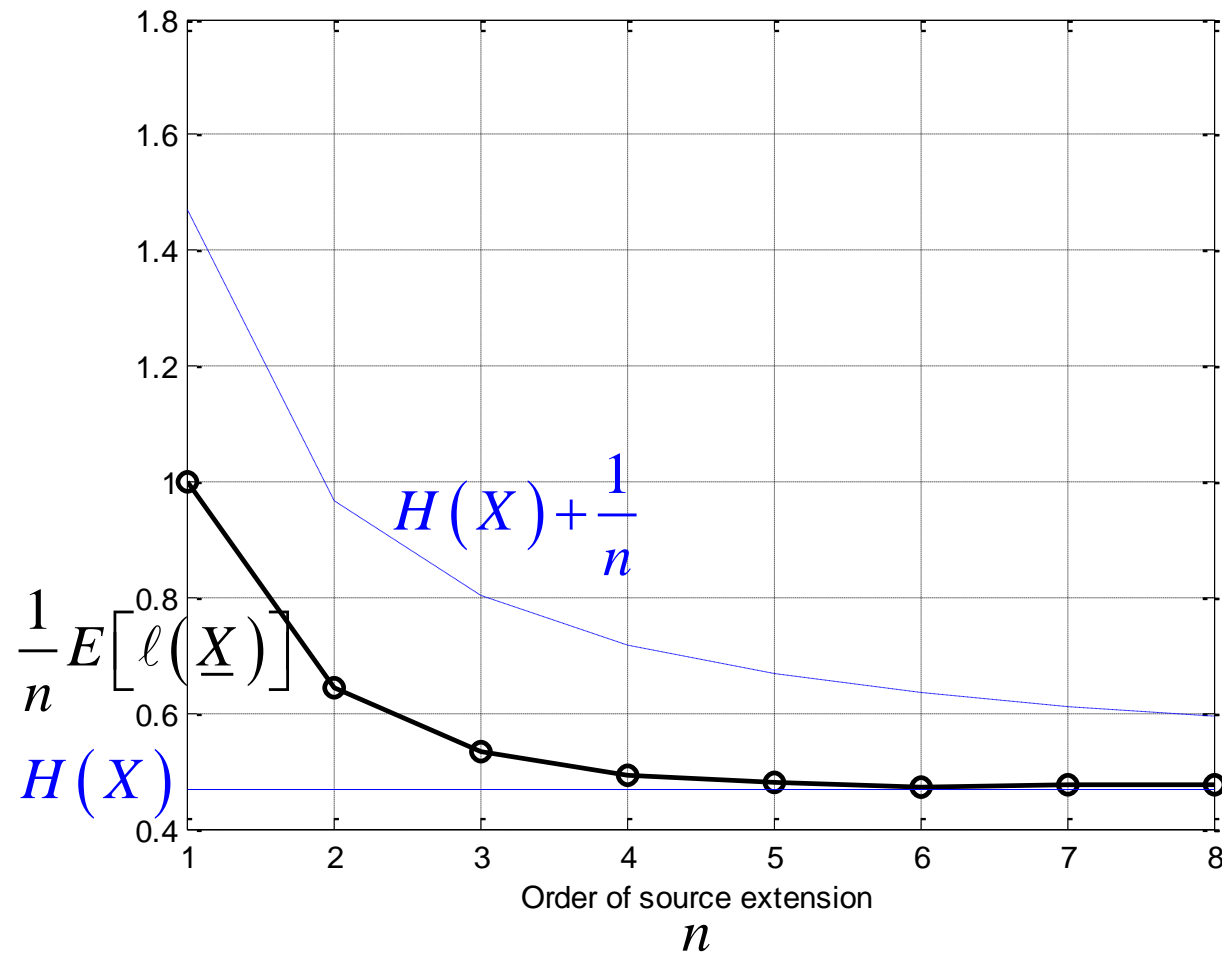
sourceString =

```
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)

