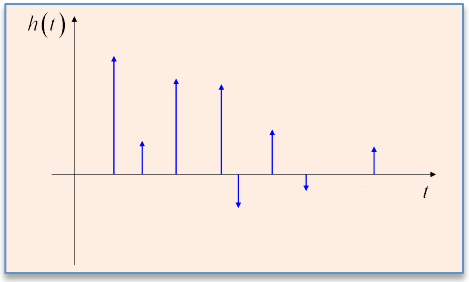
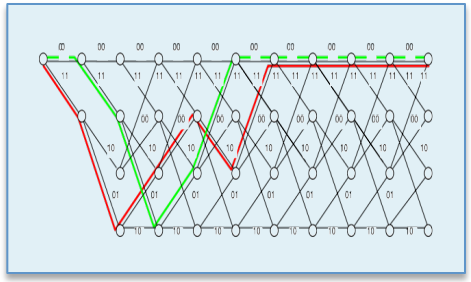
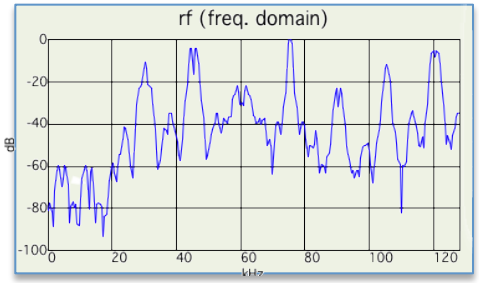
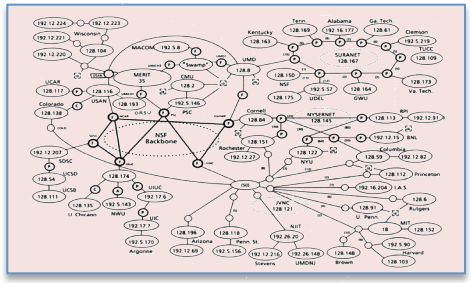








INTRODUCTION TO EECS II

DIGITAL COMMUNICATION SYSTEMS

6.02 Fall 2014
Lecture #1

- Digital vs. analog communication
- The birth of modern digital communication
- Information and entropy
- Binary codes

6.02 Course Staff

Katrina LaCurts
George Verghese

Vincent Chan
Srini Devadas
Yury Polyanskiy
Victor Zue

Ali AlShehab
Max Dunitz
Ellen Finch
Ameesh Goyal
Ravi Netravali
Quan Nguyen
Evangelos Taratoris
Pratiksha Thaker

+ LAs and Graders!

Course Ethos*

Great material …
a direct and tangible line
of development from
200 years ago to systems
of importance today
(and tomorrow!), including
many links to MIT

Lots to learn (and teach) …
collaboratively

Individual effort …
we have to be seeing
your own work
on anything submitted
for evaluation, with
all collaboration
fully acknowledged

*Animating principles

Lectures and Recitations

… are your best and most efficient entry to the subject. Even the stuff that's confusing or not clear in lecture/recitation is useful in letting you know what to get clarified. So please attend!!

Your "contract" in this subject is not with notes or slides, but with the staff, and specifically with what's developed in lectures and recitations.

Digital vs. Analog Communication

- ANALOG Communicating a continuous-time waveform (e.g.: acoustic speech; voltage from a microphone).
  - Analog electronics
  - Fidelity to the waveform

- DIGITAL Communicating a message comprising a discrete-time sequence of symbols from some source alphabet (e.g.: written text; Morse telegraphy; computer communication)
  - Often coded onto some other sequence of symbols that's adapted to the communication channel, e.g., binary digits, 0 and 1.
  - Often involving analog communication across the physical channel
  - Fidelity to the message
  - Well suited to taking advantage of ever increasing computational power, storage, interconnectedness

Samuel F.B. Morse
1791-1872

UNITED STATES PATENT OFFICE.

SAMUEL F. B. MORSE, OF NEW YORK, N. Y.

IMPROVEMENT IN THE MODE OF COMMUNICATING INFORMATION BY SIGNALS BY THE APPLICATION OF ELECTRO-MAGNETISM.

Specification forming part of Letters Patent No. **1,647**, dated June 20, 1840.

To all whom it may concern:

Be it known that I, the undersigned, SAMUEL F. B. MORSE, of the city, county, and State of New York, have invented a new and useful machine and system of signs for transmitting intelligence between distant points by the means of a new application and effect of electro-magnetism in producing sounds and signs, or either, and also for recording permanently by the same means, and application, and effect of electro-magnetism, any signs thus produced and representing intelligence, transmitted as before named between distant points; and I denominate said invention the "American Electro-Magnetic Telegraph," of which the following is a full and exact description, to wit:

It consists of the following parts—first, of a circuit of electric or galvanic conductors from any generator of electricity or galvanism and of electro-magnets at any one or more points in said circuit; second, a system of signs by which numerals, and words repre-

of any generator of electricity or galvanism, to one or more electro-magnets placed at any point or points in said circuit, the magnetic power thus concentrated in such magnet or magnets is used for the purposes of producing sounds and visible signs, and for permanently recording the latter at any and each of said points at the pleasure of the operator and in the manner hereinafter described—that is to say, by using the system of signs which is formed of the following parts and variations, viz:

Signs of numerals consist, first, of ten dots or punctures, made in measured distances of equal extent from each other, upon paper or any substitute for paper, and in number corresponding with the numeral desired to be represented. Thus one dot or puncture for the numeral 1, two dots or punctures for the numeral 2, three of the same for 3, four for 4, five for 5, six for 6, seven for 7, eight for 8, nine for 9, and ten for 0, as particularly represented on the annexed drawing marked

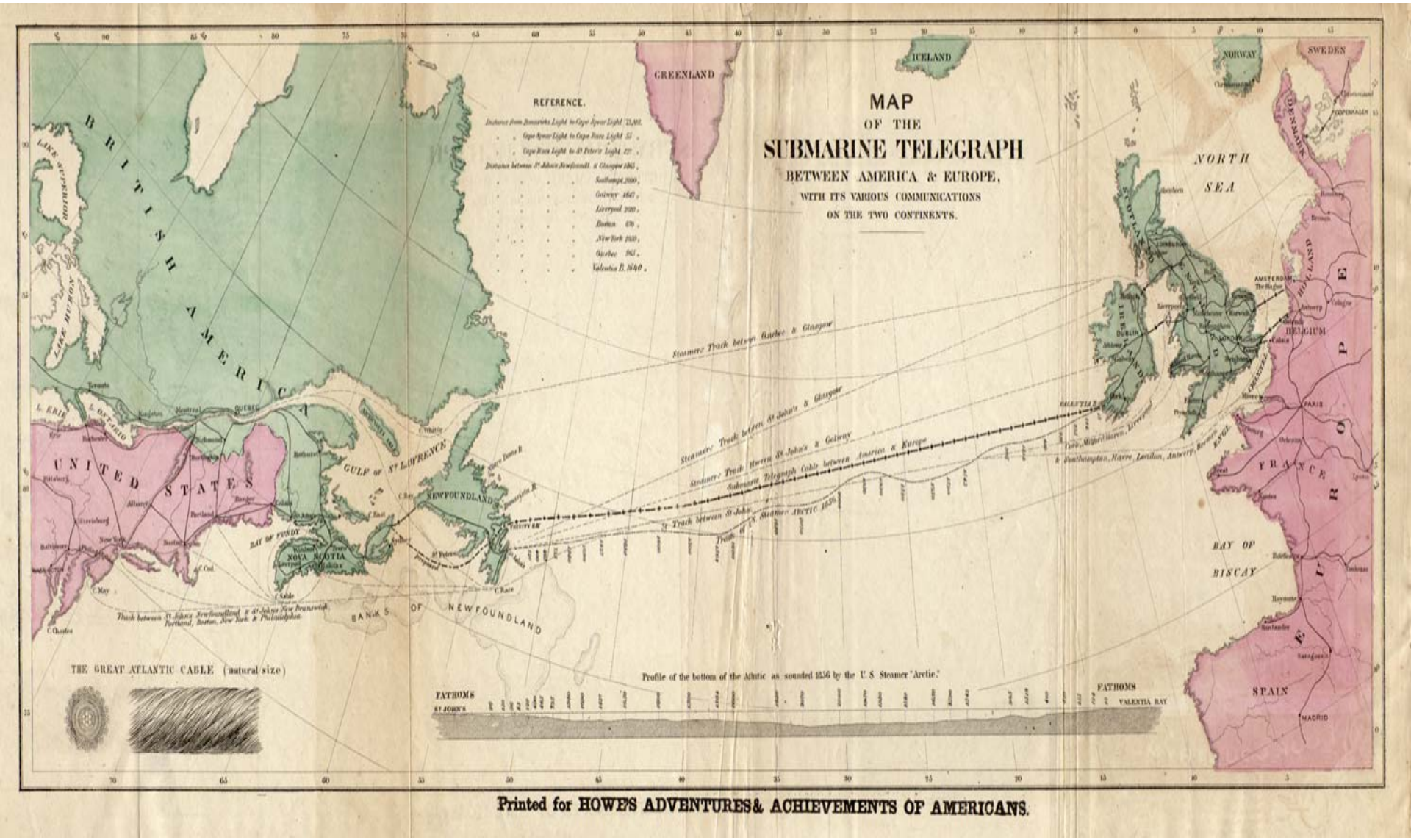
Samuel F.B. Morse

- Invented (1832 onwards, patent #1,647 in 1840) the most practical form of electrical telegraphy, including keys, wire arrangements, electromagnets, marking devices, relays, …, and Morse code!

- Worked tirelessly to establish the technology

- After initial struggles, telegraphy was quickly adopted and widely deployed
  - Trans-Atlantic cable attempts 1857 (16 hours to send 98 words from Queen Victoria to President Buchanan!), 1858, 1865, finally success in 1866 (8 words/ minute)
  - Trans-continental US in 1861 (effectively ended the Pony Express)
  - Trans-Pacific 1902

- Telegraphy transformed communication (trans-Atlantic time from 10 days by ship to minutes by telegraph) and commerce, also spurred major developments in EE theory & practice (Henry, Kelvin, Heaviside, Pupin, …)



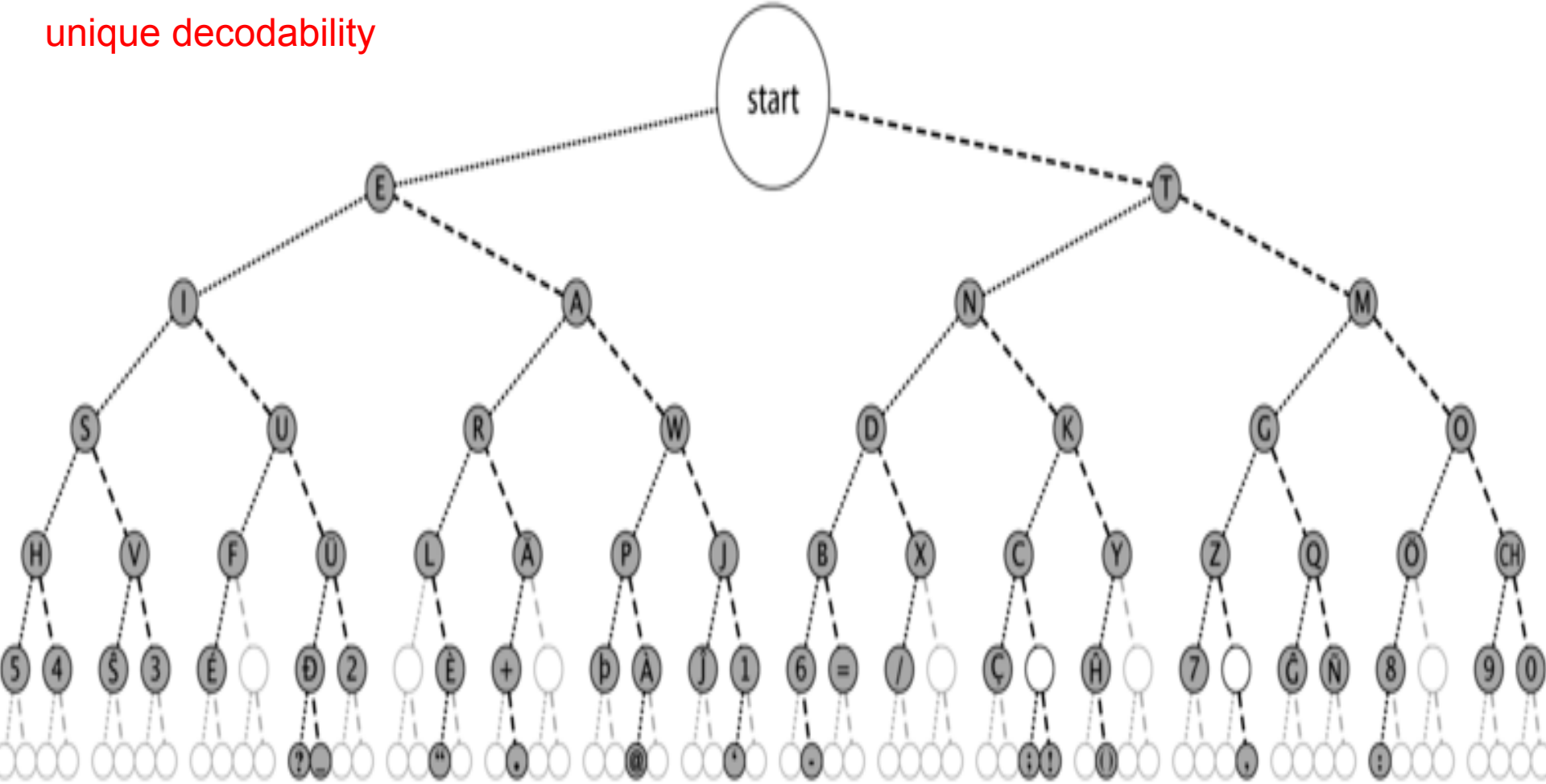
Route of the 1858 trans-Atlantic cable

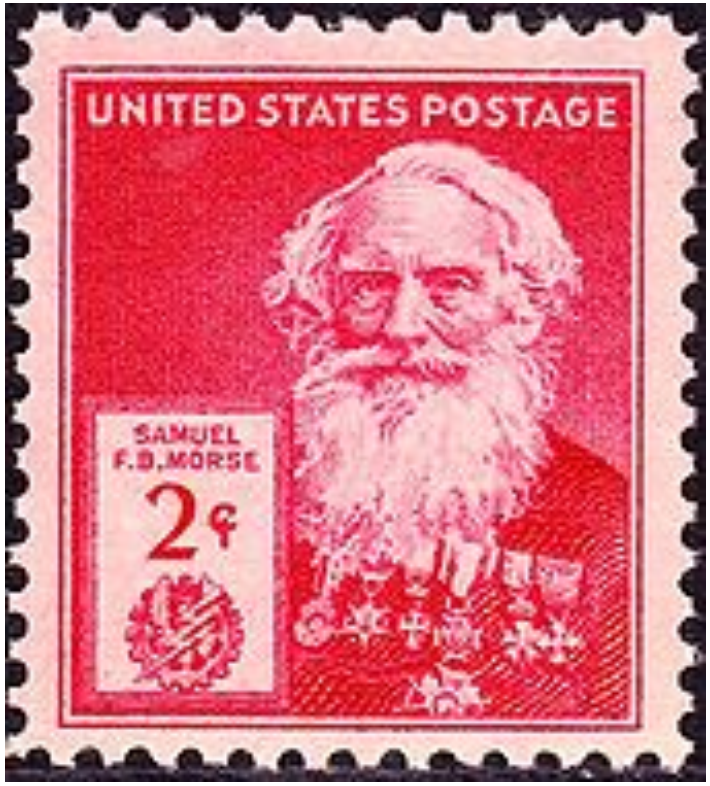
International Morse Code

1. A dash is equal to three dots.
2. The space between parts of the same letter is equal to one dot.
3. The space between two letters is equal to three dots.
4. The space between two words is equal to seven dots.

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	▬	▬	▬	▬	▬

Need a Space symbol for instantaneous and unique decodability




UNITED STATES POSTAGE
SAMUEL
F.B.MORSE
2¢

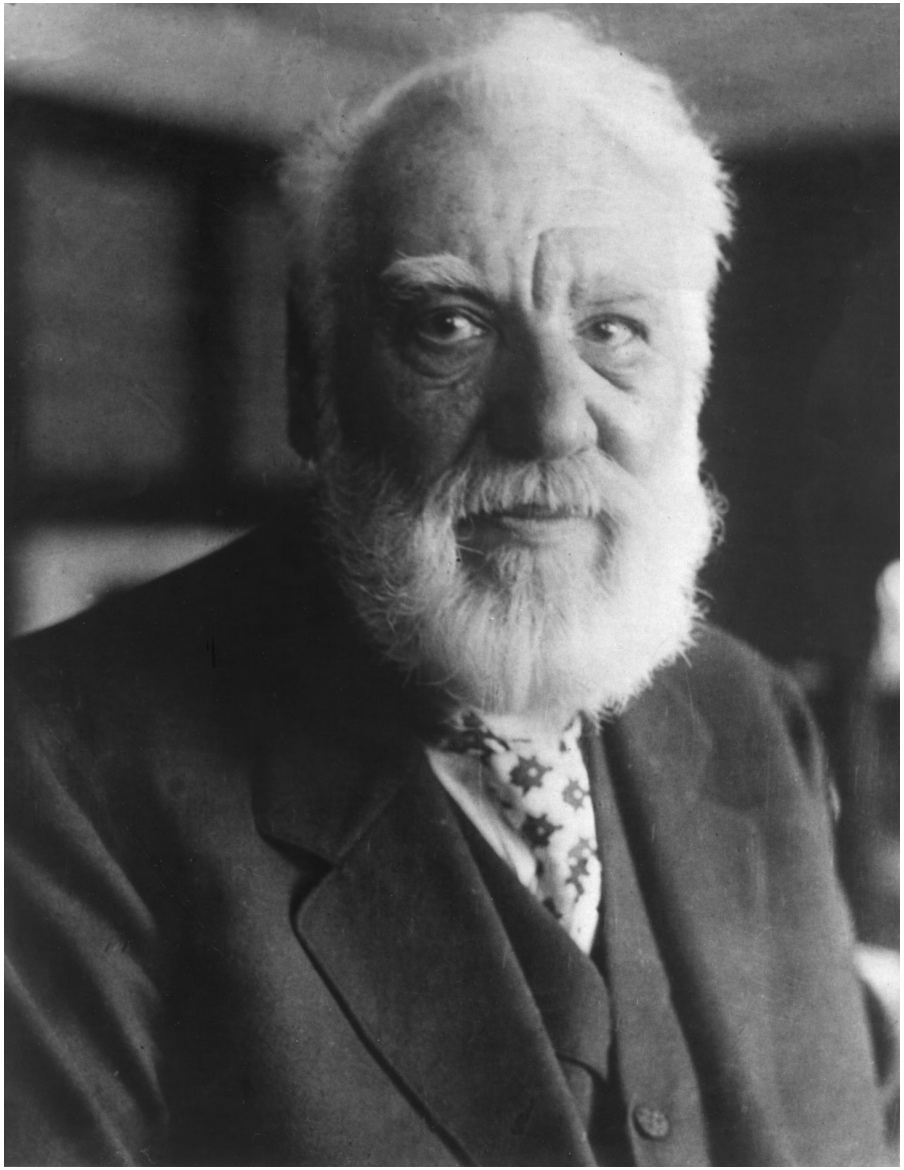
Fast-forward 100 years

- Via
  - Telephone ("Improvement in Telegraphy", patent # 174,456, Bell 1876)
  - Wireless telegraphy (Marconi 1901)
  - AM radio (Fessenden 1906)
  - FM radio (Armstrong 1933)
  - Television broadcasting by the BBC (1936)

- Mostly back to analog for a while!

- Bell Labs galaxy of researchers, who also set the stage for a return to digital
  - Nyquist, Bode, Hartley, …

Alexander
Graham Bell
1847–1922



Harry Nyquist
1889–1976

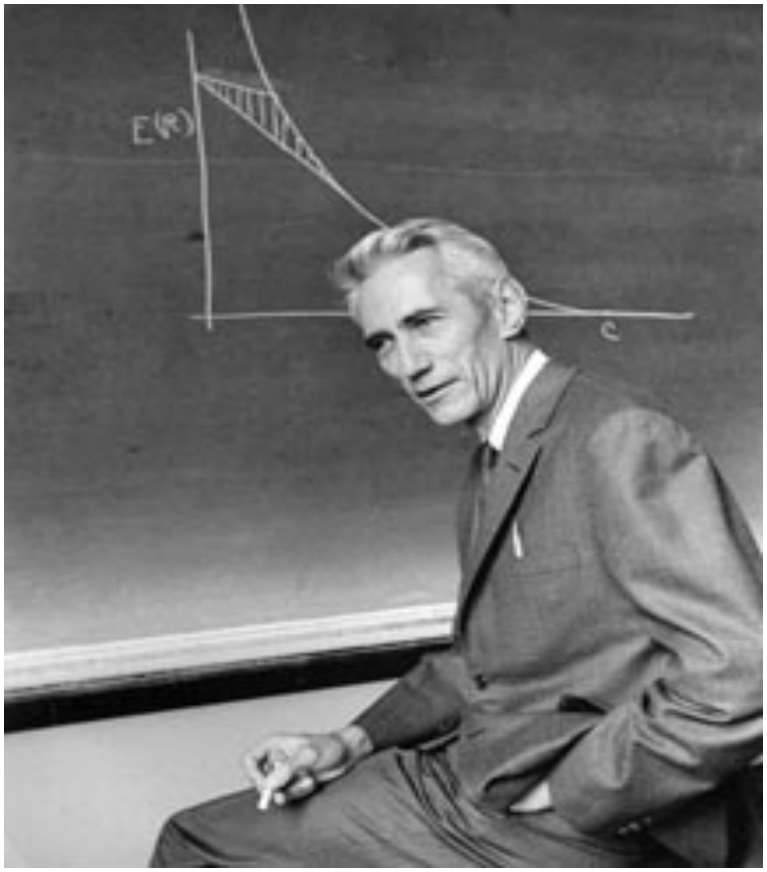
Prolific and seminal work from Bell Labs, including
"Certain topics in telegraph transmission theory" 1928

Claude E. Shannon, 1916-2001

1937 Masters thesis, EE Dept, MIT
A symbolic analysis of relay and switching circuits
Introduced application of Boolean algebra to logic circuits, and vice versa.
Very influential in digital circuit design.
"Most important Masters thesis of the century"

1940 PhD, Math Dept, MIT
An algebra for theoretical genetics
To analyze the dynamics of Mendelian populations.

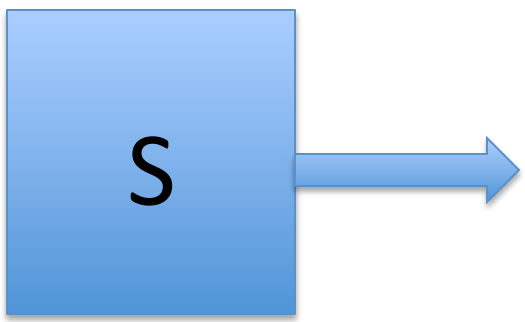
Joined Bell Labs in 1940.



"A mathematical theory of cryptography" 1945/1949
"A mathematical theory of communication" 1948

MIT faculty
1956-1978

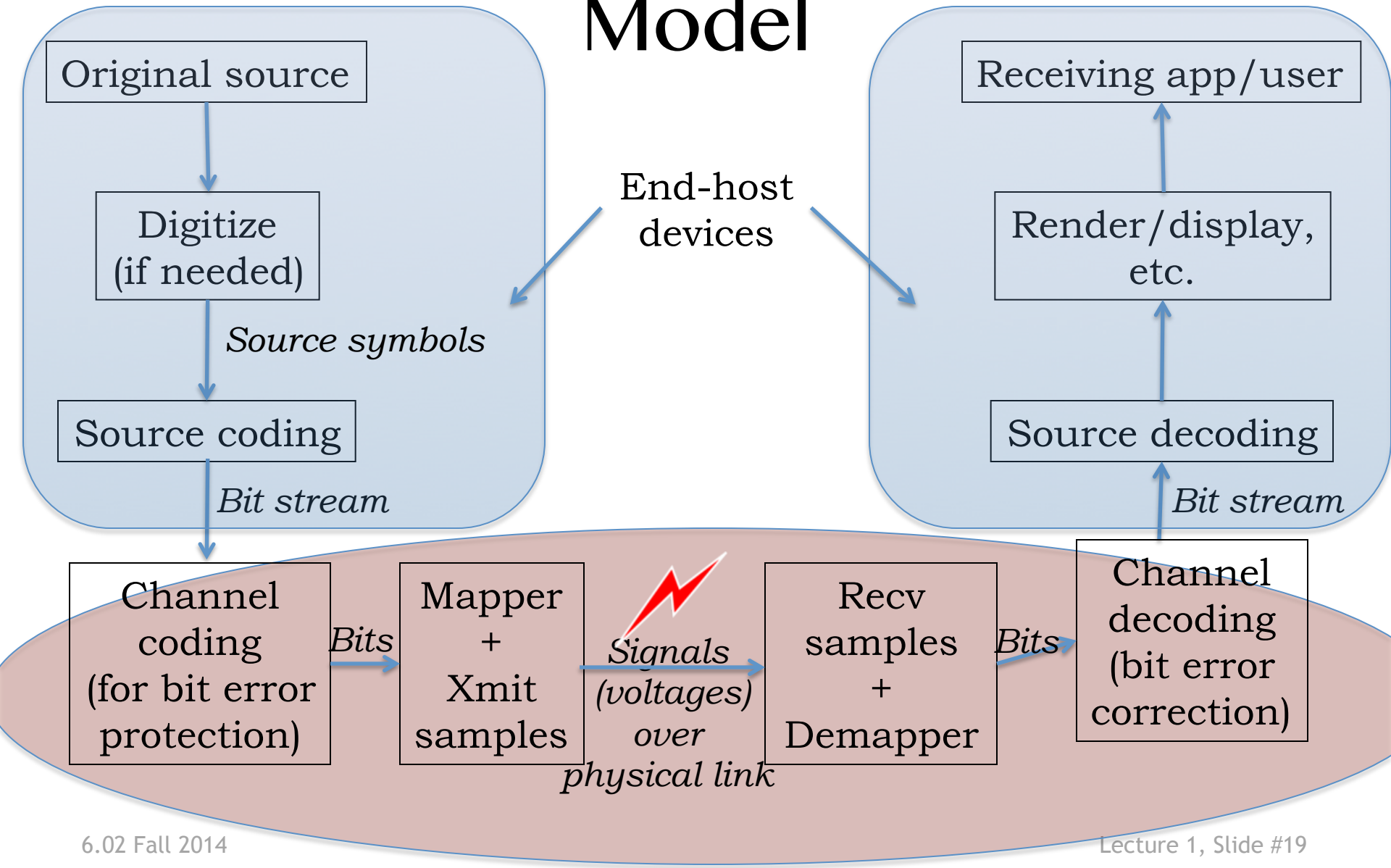
Stationary Discrete Memoryless Probabilistic Source



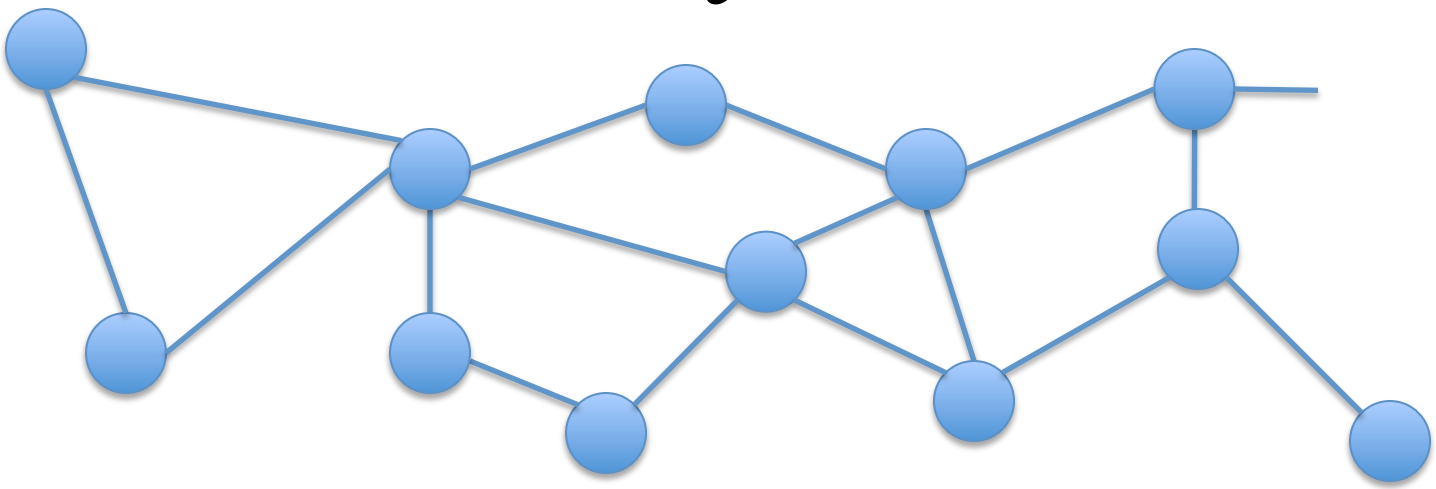
$\ldots, s_8, s_2, s_{11}, s_1, s_4, s_{13}, s_2, \ldots$

Emitting symbols sequentially in time, with the symbol at each time chosen independently of the choices at other times, but using the same probability distribution, i.e., an independent, identically distributed or i.i.d. symbol stream

Single Link Communication Model



6.02 Syllabus



Point-to-point communication channels (transmitter→receiver):

- Measuring and appropriately encoding information BITS
- Transmission on physical channels SIGNALS
- Noise, bit errors, error correction
- Sharing a channel

Multi-hop networks:

- Packet switching, efficient routing PACKETS
- Reliable delivery on top of a best-efforts network

Probabilistic Models

- Universe **U** of elementary outcomes e_1, e_2, … . One and only one outcome in each experiment or run of the model.

- Events A, B, C, … are subsets of **U**. We say event A has occurred if the outcome of the experiment lies in A.

- Events form an "algebra" of sets, i.e., A *or* B (union, also written A+B) is an event, A *and* B (intersection, also written AB or A,B) is an event, *not* A (complement, also written A^c) is an event. So **U** and the null set **0** are also events.

- Probabilities are defined on events, such that $0 \le P(A) \le 1$, P(**U**)=1, and P(A+B)=P(A)+P(B) if A and B are mutually exclusive, i.e. if AB=**0**. More generally, P(A+B)=P(A)+P(B)-P(AB).

Conditional Probability and Independence

- Events A, B, C, D, E, etc., are said to be (mutually) independent if the joint probability of every combination of these events factors into product of individual probabilities, so P(ABCDE)=P(A)P(B)P(C)P(D)P(E), P(ABCD)=P(A)P(B)P(C)P(D), P(ADE)=P(A)P(D)P(E), etc.

- Conditional probability P(A, given that B has occurred):

$$P(A|B) = P(AB)/P(B)$$
$$= P(B|A)P(A)/P(B) \quad \text{(the key to Bayes's rule)}$$

Random Variables, Expectation, Variance

- A(real-valued) random variable R is a mapping from elementary outcomes e_i to real numbers r_i --- or equivalently, a numerically specified outcome of a probabilistic experiment.

- The expected value or mean value of a random variable is the probability-weighted average of the (numerical) outcomes: $E[R] = \sum r_i P(r_i)$

- If $G=f(R)$, the $E[G] = \sum f(r_i)P(r_i) = \sum g_i P(r_i)$

- The variance is the expected squared deviation from the mean, so

$$var(R) = \sum (r_i - E[R])^2 P(r_i)$$

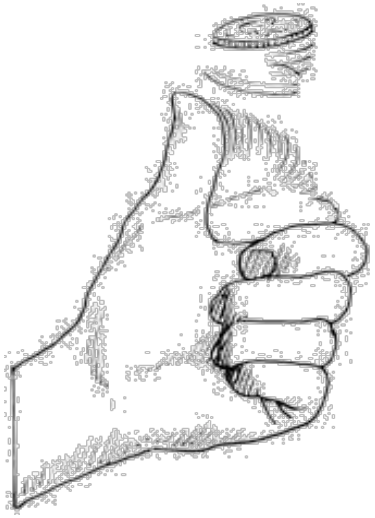
Measuring Information

Shannon's (and Hartley's) definition of the information obtained on being told the outcome s_i of a probabilistic experiment S:

$$I(S = s_i) = \log_2\left(\frac{1}{p_S(s_i)}\right)$$

where $p_S(s_i)$ is the probability of the event $S = s_i$.

The unit of measurement (when the log is base-2) is the **bit** (**b**inary **i**nformation uni**t** --- not the same as binary digit!).



1 bit of information corresponds to $p_S(s_i) = 0.5$. So, for example, when the outcome of a *fair* coin toss is revealed to us, we have received 1 bit of information.

"Information is the resolution of uncertainty"
Shannon

Properties of Information Definition

- A lower-probability outcome, i.e., a more uncertain outcome, yields higher information

- A highly informative outcome does not necessarily mean a more valuable outcome, only a more surprising outcome, i.e., there's no intrinsic value being assessed (can think of information as degree of surprise)

- Often used fact: The information in independent events is additive.

  (This is what motivates using a logarithmic measure in the first place.)

Expected Information or Average Uncertainty = Entropy

Consider a discrete random variable S, which may represent the set of possible symbols to be transmitted at a particular time, taking possible values $s_1, s_2, \ldots, s_N$, with respective probabilities .

$$p_S(s_1), p_S(s_2), \ldots, p_S(s_N)$$

The *entropy* $H(S)$ of S is the expected (or mean or average) value of the information obtained by learning the outcome of S :

$$H(S) = \sum_{i=1}^{N} p_S(s_i) I(S = s_i) = \sum_{i=1}^{N} p_S(s_i) \log_2 \left(\frac{1}{p_S(s_i)} \right)$$

When all the $p_S(s_i)$ are *equal* (with value $1/N$), then

$$H(S) = \log_2 N \qquad \text{or} \qquad N = 2^{H(S)}$$

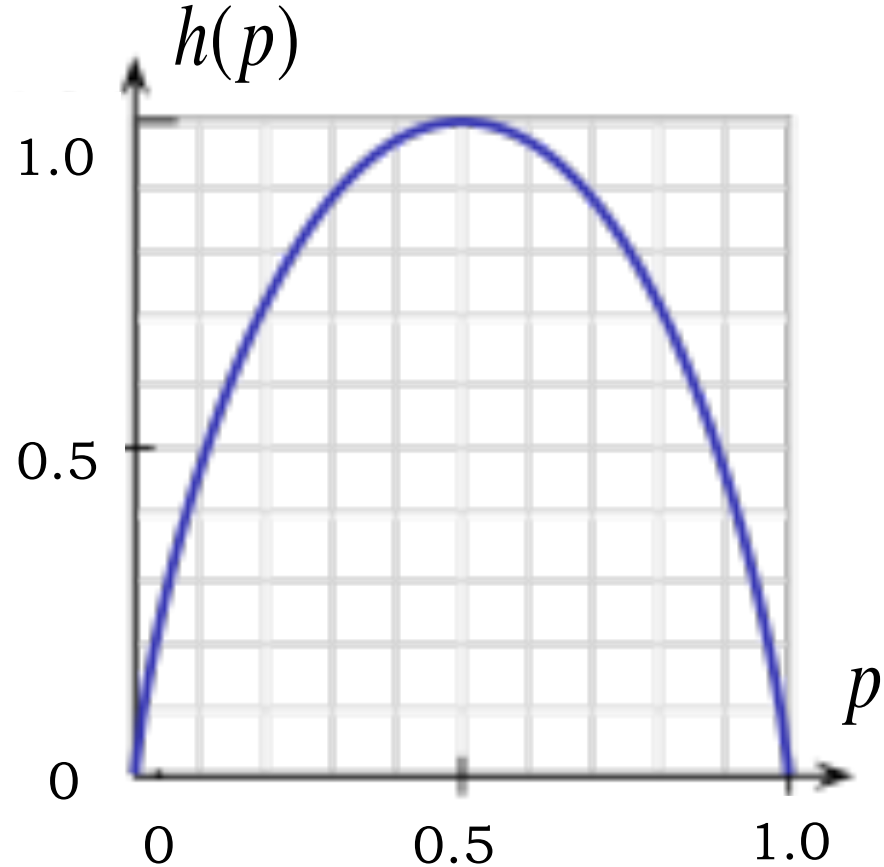
This is the maximum attainable value!

e.g., Binary entropy function $h(p)$

Heads (or C=1) with probability p

Tails (or C=0) with probability $1-p$

So the maximum information carried on average by a binary digit (0 or 1) e.g., reporting the result of a coin toss, is 1 bit, and is obtained with a fair coin.



$$H(C) = -p\log_2 p - (1-p)\log_2(1-p) = h(p)$$

Binary Encoding

- A binary encoding maps source symbols to binary strings – codewords
  - Important for channels that can only be in one of two states
  - So preceding example is important

- We want the sequence of symbols to be uniquely recoverable in real time from the binary stream (i.e., concatenation of codewords) that encodes it
  - Prefix-free codes, representable as leaves of a coding tree (graph)

- We could also map blocks of symbols to (longer) codewords

Significance of Entropy

Entropy (in bits) tells us the average amount of information (in bits) that must be delivered in order to resolve the uncertainty about the outcome of a trial. This is a lower bound on the number of binary digits that must, on the average, be used to encode our messages:

$$H \leq L$$

Confusingly, a binary digit is also referred to as a "bit!"

If we send fewer binary digits on average, the receiver will have some uncertainty about the outcome described by the message.

If we send more binary digits on average, we're wasting the capacity of the communications channel by sending binary digits we don't have to.

Achieving the entropy lower bound is the "gold standard" for an encoding (at least from the viewpoint of information compression).

Fixed-length Encodings

An obvious choice for binary encoding of equally probable outcomes is to choose a fixed-length code that has enough sequences to encode the necessary information

- 96 printing characters → 7-"bit" ASCII
- Unicode characters → UTF-16
- 10 decimal digits → 4-"bit" BCD (binary coded decimal)

Fixed-length codes have some advantages:

- They are "random access" in the sense that to decode the nth message symbol one can decode the nth fixed-length sequence without decoding sequences 1 through n-1.
- Table lookup suffices for encoding and decoding

Variable-length Encodings

- Map more probably symbols to shorter codewords
- Or map blocks of symbols to codewords, with more probably block mapped to shorter codewords

Connection to (Binary) Coding

- Suppose P(C=1) = 1/1024 in a coin toss, i.e., very small probability of getting Heads, typically only one in 1024 trials. Then

$$h(p) = (1/1024)\log_2(1024/1) + (1023/1024)\log_2(1024/1023)$$

$= .0112$ bits of uncertainty or information per trial on average

- So using 1000 binary digits (C=0 or 1) to encode the results of 1000 tosses of this particular coin seems inordinately wasteful, i.e., 1 binary digit per trial. Can we get closer to an average of .0112 binary digits/trial? i.e., can we use closer to 11.2 binary digits on average to encode the results of 1000 trials with this coin?

- Yes!