ELEC3203 Digital Coding and Transmission

Half of the unit:

- **Information Theory**
- **MODEM** (modulator and demodulator)

Professor Sheng Chen: Building 53, Room 4005
E-mail: sqc@ecs.soton.ac.uk

Lecture notes from: Course Office (ECS Student Services)
or Download from:
http://www.ecs.soton.ac.uk/~sqc/ELEC3203/

Other half: **CODEC** (source coding, channel coding) by Prof. Lajos Hanzo / Dr. Rong Zhang

What's New (Should be Old) in Communications

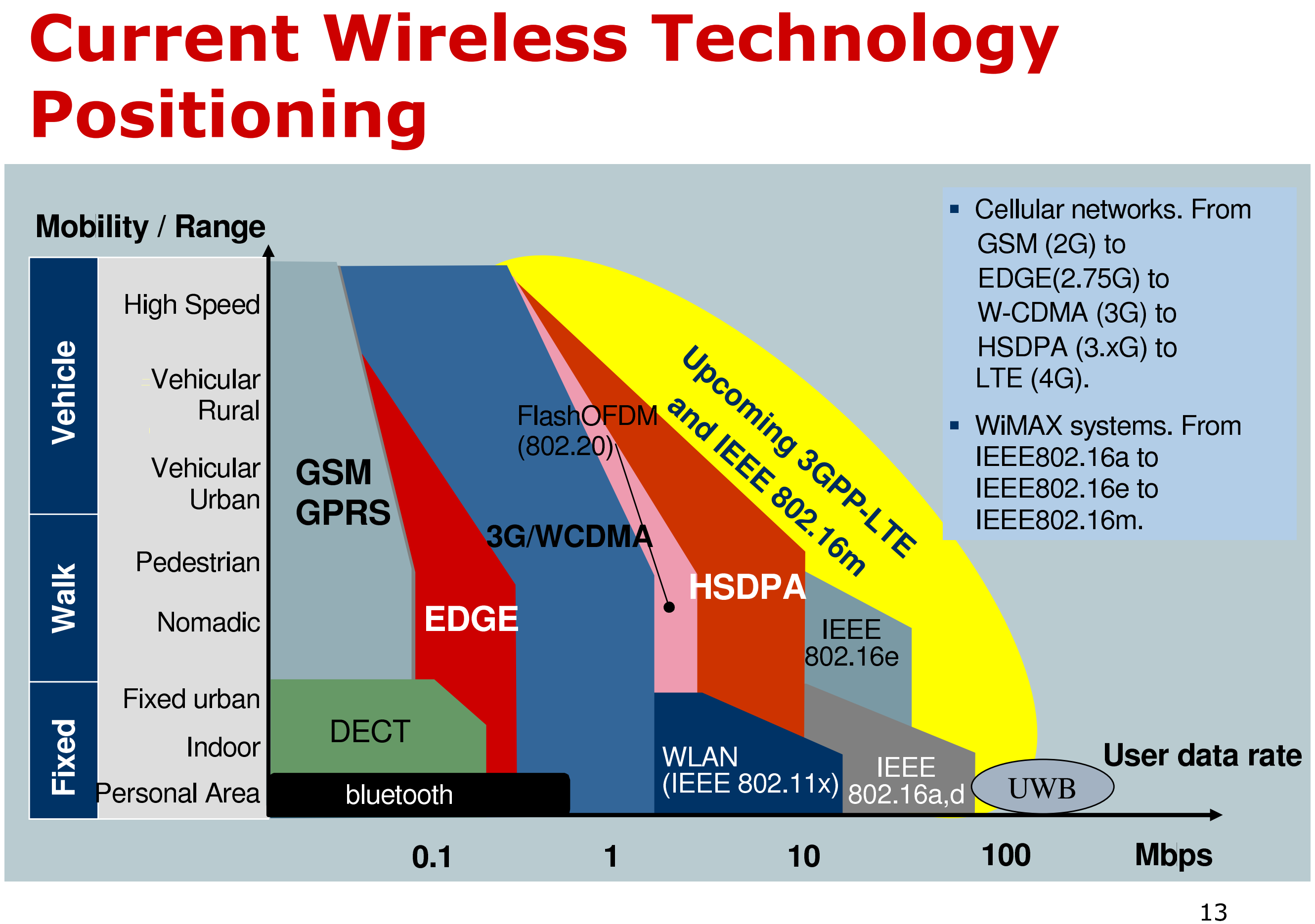
- Imagine a few scenarios:
  - In holiday, use your smart phone to take video/pictures and share them online instantaneously
  - Anywhere anytime with your smart device/phone connected
  - Live in two real-life worlds, physical world and digital virtual world – What percentage of your life is living in digital virtual world ?
- Do you know these words:

  CDMA, multicarrier, OFDM, space-time processing, turbo coding, WiFI, WiMAX,

  LTE-Advanced, massive MIMO, device-to-device communication
- In this introductory course, we should go through some A B C ... of **Digital Communication**, enabling technology of our information society

Wireless and Mobile Networks

- A view around 2008: 2G, 3G, "future" B3G or 4G



Some improved 2G, HSCSD: high-speed circuit switched data, GPRS: general packet radio service, EDGE: enhanced data rates for GSM evolution. Also, HIPERLAN: high performance radio local area network

- 4G now deployed and "future" is B4G or 5G, push both **mobility** and **rate** axises further

How You Define Communications

- Aim of **tele**communications: to communicate information between geographically separated locations via a communications channel of adequate quality (at certain rate reliably)

- Since our world becomes ever digital – digital economy, etc, the transmission is based on digital data, which something is obtained from analogue quantities by

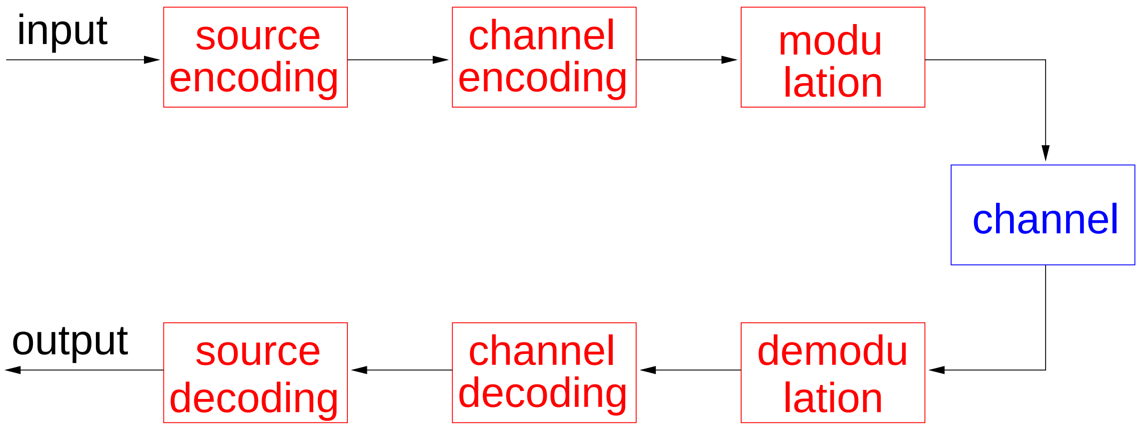
1. sampling (Nyquist: sampling with at least twice the maximum frequency), and
2. quantisation (introduction of quantisation noise through rounding off)

- Transmitting at certain rate requires certain spectral **bandwidth**

- Here channel means whole system, which has certain **capacity**, the maximum rate that can be used to transmit information through the system **reliably**

General Transmission Scheme

- A **digital transmission** scheme generally involves:



- Input/output are considered digital (if analogue ⇒ sampled/quantised)
- Vertical partition: **CODEC**, **MODEM**, **channel** (transmission medium)
- Horizontal partition: Your 3G mobile phone, for example, contains a pair of **transmitter** and **receiver**,
  - together called **transceiver**, consisting of a CODEC and MODEM

What is Information

- Generic question: what is **information**? How to measure it (**unit**)?

  As communication is concerned with conveying information from "source" (transmitter) to "destination" (receiver), let us start with source

- Generic digital source is characterised by:
  - Source alphabet (message or symbol set): $m_1, m_2, \cdots, m_q$
  - Probability of occurrence (symbol probabilities): $p_1, p_2, \cdots, p_q$
    e.g. binary equiprobable source $m_1 = \mathbf{0}$ and $m_2 = \mathbf{1}$ with $p_1 = 0.5$ and $p_2 = 0.5$
  - Symbol rate (symbols/s or Hz)
  - Probabilistic interdependence of symbols (correlation of symbols, e.g. does m_i tell us nothing about m_j or something? or nothing?)

- At a specific symbol interval, symbol m_i is transmitted correctly to receiver
  - What is *amount of information* conveyed from transmitter to receiver?
  - The answer:

$$I(m_i) = \log_2 \frac{1}{p_i} = -\log_2 p_i \quad \text{(bits)}$$

Concept of Information

- Forecast: tomorrow, rain in three different places:

  1. Raining season in a tropical forest
  2. Somewhere in England
  3. A desert where rarely rains

- **Information content** of an event is connected with **uncertainty** or inverse of probability. The more unexpected (smaller probability) the event is, the more information it contains

- **Information theory** (largely due to Shannon)

  – Measure of information
  – Information capacity of channel
  – coding as a means of utilising channel capacity

A note: Shannon's information theory is a foundation to our information society/digital world, as Einstein's relativity theory to modern physics

Shannon Limit

- We know different communication system designs achieve different performance levels and we also know system performance is always limited by the available *signal power*, the inevitable *noise* and the need to limit *bandwidth*

  What is the ultimate performance **limit** of communication systems, underlying only by the fundamental physical nature?

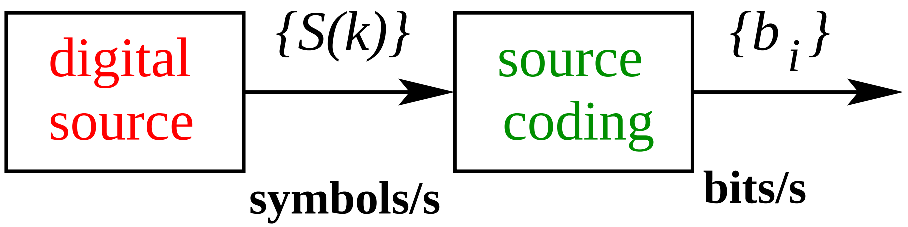
- Shannon's information theory addresses this question

  Shannon's theorem: If the rate of information from a source does not exceed the capacity of a communication channel, then there exists a coding technique such that the information can be transmitted over the channel with arbitrarily small probability of error, despite the presence of noise

- This is "existing" theory, not tell how to so it. As we get closer to capacity, we either need too long delay (infinite code length) or too higher processing power

  In 1992, two French Electronics professors developed practical *turbo coding*, which allows get close to Shannon limit (transmit information at capacity rate), with finite delay and power

Source and Source Coding



- Digital source transmits sequence of symbols $\{S(k)\}$, where k is time index, and is characterized by:

  1. Symbol set or source alphabet: $\mathcal{S} = \{m_i, 1 \leq i \leq q\}$
  2. Probability of occurring for m_i: p_i, $1 \leq i \leq q$
  3. Symbol rate R_s [symbols/s]
  4. Source memory or interdependency in $\{S(k)\}$
     e.g. what is probability of $S(k) = m_j$ given previous $S(k-1) = m_i$?

- Source coding maps each $S(k)$ to a code word or bit pattern, and hence convert sequence of symbols $\{S(k)\}$ [symbols/s] into sequence of bits $\{b_i\}$ [bits/s]

- So source coding is concerned with coding each $m_i \in \mathcal{S}$ with a unique code word or bit pattern

Information Content

- We first consider memoryless source, i.e. source with independent symbols: $m_1, m_2, \cdots, m_q$, and probability of occurrence: $p_1, p_2, \cdots, p_q$

- Definition of information: amount of information in ith symbol m_i is defined by

$$I(m_i) = \log_2 \frac{1}{p_i} = -\log_2 p_i \quad \text{(bits)}$$

Note the unit of information: bits !

- Properties of information
  - Since probability $0 \le p_i \le 1$, $I(m_i) \ge 0$: information is nonnegative
  - If $p_i > p_j$, $I(m_i) < I(m_j)$: the lower the probability of a source symbol, the higher the information conveyed by it
  - $I(m_i) \to 0$ as $p_i \to 1$: symbol with probability one carries no information
  - $I(m_i) \to \infty$ as $p_i \to 0$: symbol with probability zero carries infinite amount of information (but it never occurs)

Physical Interpretation

- Information content of a symbol or message is equal to minimum number of binary digits required to encode it and, hence, has a unit of bits
  - Binary equiprobable symbols: $m_1, m_2 \leftrightarrow \mathbf{0}, \mathbf{1}$, minimum of one binary digit (one bit) is required to represent each symbol
    Equal to information content of each symbol: $I(m_1) = I(m_2) = \log_2 2 = 1$ bit
  - Four equiprobable symbols: $m_1, m_2, m_3, m_4 \leftrightarrow \mathbf{00}, \mathbf{01}, \mathbf{10}, \mathbf{11}$ minimum of two bits is required to represent each symbol
    Equal to information content of each symbol: $I(m_1) = I(m_2) = I(m_3) = I(m_4) = \log_2 4 = 2$ bits
  - In general, q equiprobable symbols m_i, $1 \leq i \leq q$, minimum number of bits to represent each symbol is $\log_2 q$
    Equal to information content of each symbol: $I(m_i) = \log_2 q$ bits

- Use $\log_2 q$ bits for each symbol is called Binary Coded Decimal
  - Equiprobable case: m_i, $1 \leq i \leq q$, are equiprobable $\Rightarrow$ BCD is good
  - Non-equiprobable case? or what is best source coding?

Information of Memoryless Source

Source emitting symbol sequence of length N, $\{S(k)\}_{k=1}^{N}$. Memoryless implies that each message emitted $S(k)$ is independent of the previous messages $S(k-i)$, $i \geq 1$

Assume that N is large, so symbol m_i appears $p_i \cdot N$ times in the sequence

- Information contribution from ith symbol m_i

$$I_i = (p_i \cdot N) \cdot \log_2 \frac{1}{p_i}$$

- Total information of symbol sequence of length N

$$I_{total} = \sum_{i=1}^{q} I_i = \sum_{i=1}^{q} p_i \cdot N \cdot \log_2 \frac{1}{p_i}$$

- Average information per symbol (entropy) is

$$\frac{I_{total}}{N} = \sum_{i=1}^{q} p_i \cdot \log_2 \frac{1}{p_i} \quad \text{(bits/symbol)}$$

Entropy and Information Rate

- Memoryless source **entropy** is defined as the average information per symbol:

$$H = \sum_{i=1}^{q} p_i \cdot \log_2 \frac{1}{p_i} = - \sum_{i=1}^{q} p_i \cdot \log_2 p_i \qquad \text{(bits/symbol)}$$

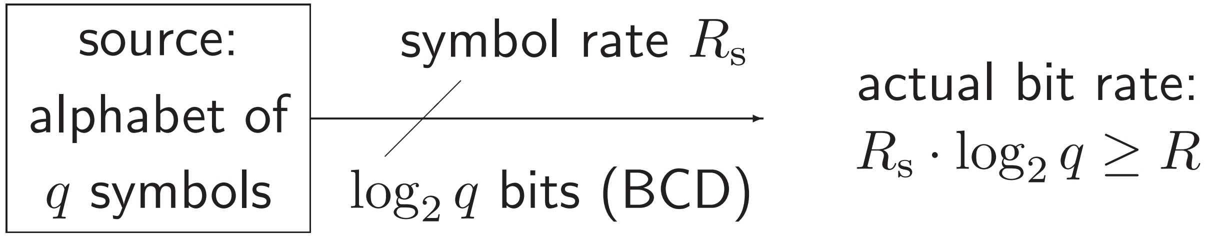
- Memoryless source emitting at rate of R_{s} symbols/s has **information rate**:

$$R = R_{\mathrm{s}} \cdot H \qquad \text{(bits/sec)}$$

- Entropy, a fundamental physical quantity of the source, quantifies average information conveyed per symbol
  - Entropy does not concern with symbol rate, i.e. it is defined by physical quantities 1., 2. and 4. of source
- Information rate, a fundamental physical quantity of the source, tells you how many bits/s information the source really needs to send out
  - Information rate is defined by all 1., 2., 3. and 4. physical quantities of source

A Look at BCD Source Coding

- If each symbol is encoded by $\log_2 q$ bits, i.e. binary coded decimal, average output bit rate is $R_b = R_\mathrm{s} \cdot \log_2 q$



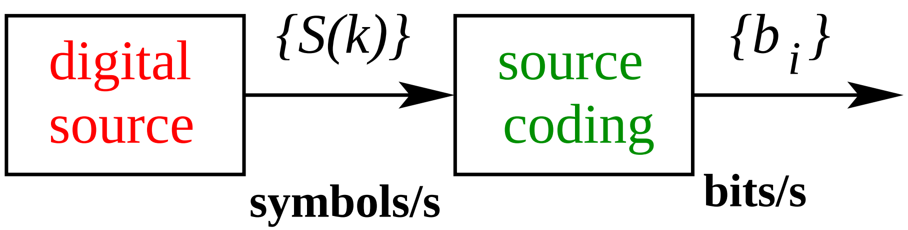
- Note information rate R is always smaller or equal to the average output bit rate R_b of the source[1] !

- So bit rate you sent is R_b [bits/s], but what you would really need to send were R [bits/s], and $R < R_b$ in general
  - It appears wasting – ideally we would like to send at what is really necessary
  - This because later we will learnt transmitting at certain rate R_b [bits/s], we need certain level of resource (bandwidth and power)

[1] Hint: $H \leq \log_2 q$

Summary

- Overview of a digital communication system: system building blocks
- Appreciation of information theory
- Information content of a symbol, properties of information
- Memoryless source with independent symbols: entropy and source information rate



  - Entropy, a fundamental physical quantity of the source, quantifies average information conveyed per symbol
  - Information rate, a fundamental physical quantity of the source, tells you how many bits/s of information the source really needs to send out
  - For memoryless source, binary coded decimal source coding is generally "waste"