

信息论基础

李 莹

liying2009@ecust.edu.cn

课程信息

- 教材及主要参考书:

 《信息论基础教程》第3版 李梅, 李亦农

北京邮电大学出版社, 2015年8月

 《信息论——基础理论与应用》第4版, 傅祖芸

电子工业出版社, 2015年2月

 T. M. Cover, Elements of Information Theory

- 考核: 平时成绩 30% (作业、考勤)
期末考试 70% (闭卷)

第一章：绪论

一、什么是信息

二、通信系统模型

三、信息论的研究内容

四、信息论的形成和发展

第一章：绪论

一、什么是信息

二、通信系统模型

三、信息论的研究内容

四、信息论的形成和发展

1. 概述

2. 信息的通俗概念

3. 信息的狭义概念（香农信息）

1. 概述

组成客观世界的三大基本要素：

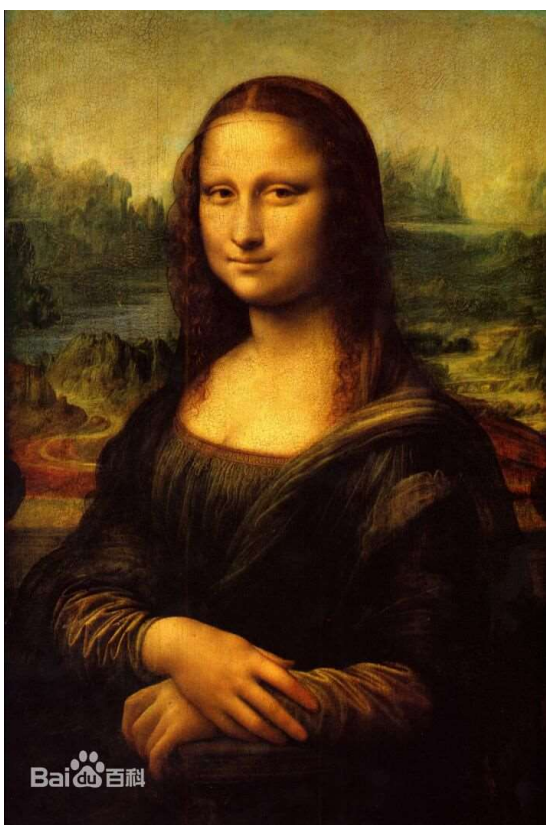
- 物质
- 能量
- 信息

没有物质什么都不存在，没有能量什么都不会发生，没有信息什么都没有意义。

——美国学者欧廷格

大漠孤烟直，
长河落日圆。

——王维



向左转弯



向右转弯

蚂蚁的气味



2. 信息的通俗概念

信息的通俗概念：消息就是信息。

用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式，把客观物质运动和主观思维活动的状态表达出来，就称为消息。

消息中包含信息，消息是信息的载体。

2. 信息的通俗概念

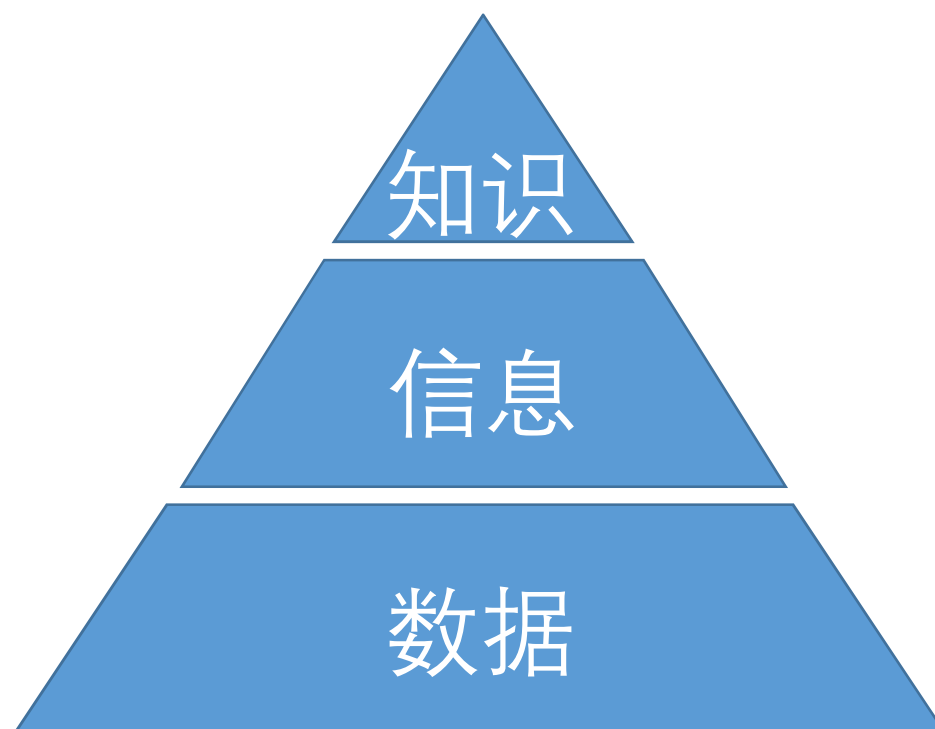
信号是表示消息的物理量，包括电信号、光信号等。

信号中携带着消息，信号是消息的载体。



2. 信息的通俗概念

数据、信息、知识



3. 信息的狭义概念（香农信息）

香农信息：信息是对事物运动状态或存在方式的不确定性的描述。

通信的过程就是消除不确定性的过程。

通信的基本问题是在一点（信宿）精确或近似恢复另一点（信源）所选择的消息。—— 香农

3. 信息的狭义概念（香农信息）

例1：

甲袋红、白球各50个，乙袋红、白、蓝、黑球各25个。比较从甲袋中取出一个球是红球的事件和从乙袋中取出一个球是红球的事件发生的难易程度，也就是事件发生的不确定性。

3. 信息的狭义概念（香农信息）

例2：

上海地区十月份可能出现的天气包括：晴、阴、雨、雪。比较天气预报为“晴”和天气预报为“雪”，给人们带来的信息量。

结论：不确定性的与事件发生的概率有关。

3. 信息的狭义概念（香农信息）

不确定性的的大小与事件发生的概率有关



不确定性是概率的函数



因此，信息量可以表示为概率的函数。

3. 信息的狭义概念（香农信息）

信息与概率的关系：

- 事件发生的概率越大，该事件包含的信息量越小；
- 如果一个事件发生的概率为1，那么它包含的信息量为0；
- 两个相互独立事件所提供的信息量应等于它们各自提供的信息量之和。

某个消息的不确定性（含有的信息量）可以表示为：

$$I(x_i) = \log \frac{1}{p(x_i)} = -\log p(x_i)$$

$$x_1 \quad p(x_1) = \frac{1}{5} \quad I(x_1) = -\log \frac{1}{5} = \log 5$$

$$x_2 \quad p(x_2) = \frac{4}{5} \quad I(x_2) = -\log \frac{4}{5} = \log 1.25$$

3. 信息的狭义概念（香农信息）

香农信息的优点：

- 有明确的数学表达式，定量化
- 与人们直观理解的信息含义一致
- 不考虑受信者主观感受的不同，认为同一消息对任何受信者，所得信息量相同。

香农信息的局限：

- 没有考虑受信者的主观特性和主观意义

第一章：绪论

一、什么是信息

二、通信系统模型

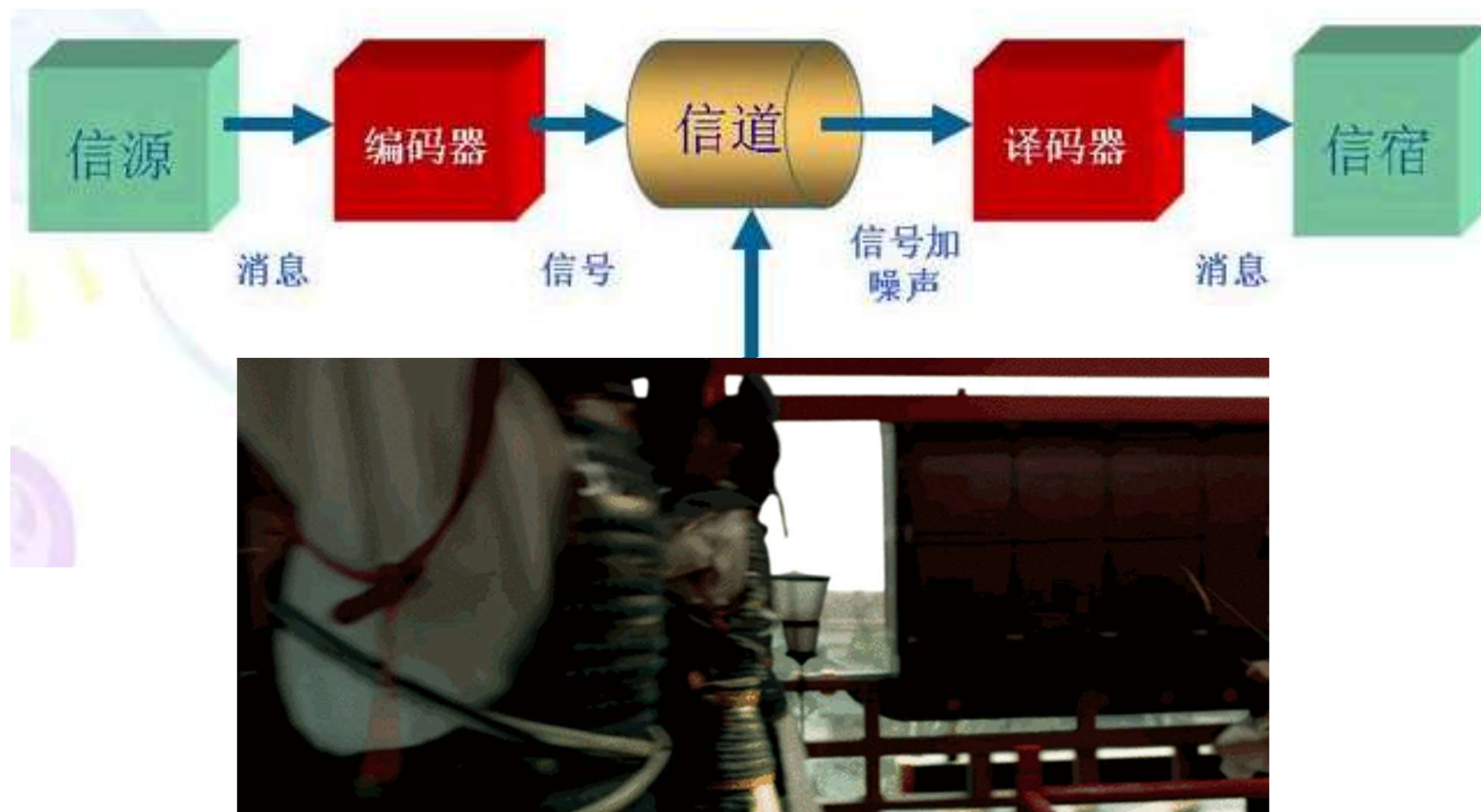
1. 通信系统模型

2. 提高通信系统的性能指标的措施

三、信息论的研究内容

四、信息论的形成和发展

1. 通信系统模型



1. 通信系统模型

- 信源
- 编码器
- 信道
- 译码器
- 信宿

通信系统的基本要求

- 有效性
 - 多、快：单位时间内传递的信息量最大
- 可靠性
 - 好：在干扰情况下失真率和差错率最小
- 其它要求
 - 适应性
 - 经济性
 - 实用性
 - 保密性

1. 通信系统模型

1) 信源

研究内容：

- 信源发出的消息的统计特性
 - 离散信源、连续信源、波形信源
 - 有记忆信源和无记忆信源
 - 平稳信源和非平稳信源
- 信源产生信息的速率

熵率

1. 通信系统模型

2) 编码器

- 编码器的功能：将消息变成适合信道传输的信号
- 编码器包括：
 - 信源编码器
 - 信道编码器
 - 调制器



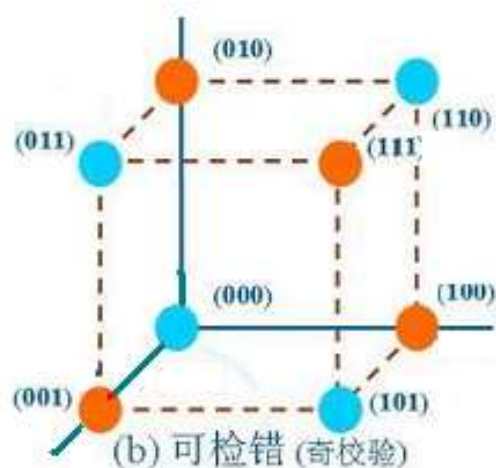
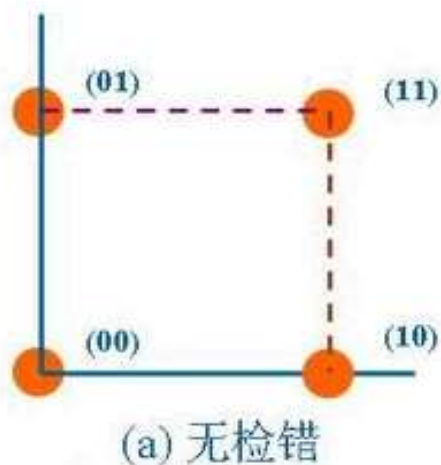
1. 通信系统模型

- 信源编码器：

- 去除信源消息中的冗余度，提高传输的有效性。

- 信道编码器：

- 将信源编码后的符号加上冗余符号，提高传输的可靠性。



1. 通信系统模型

- 调制器：

- 功能：将信道编码后的符号变成适合信道传输的信号
- 目的：提高传输效率

1. 通信系统模型

3) 信道

- 狭义信道

从发送端到接收段的媒介或通道

如电缆、光缆、无线电波、磁盘等等

- 广义信道

媒介和调制器、转换器等

1. 通信系统模型

研究内容：

信道的统计特性

- 无噪声信道、有噪声信道
- 离散信道、连续信道、波形信道
- 有记忆信道和无记忆信道
- 恒参信道（平稳信道）和随参信道（非平稳信道）
- 单用户信道和多用户信道

信道传输信息的最高速率

信道容量

1. 通信系统模型

4) 译码器

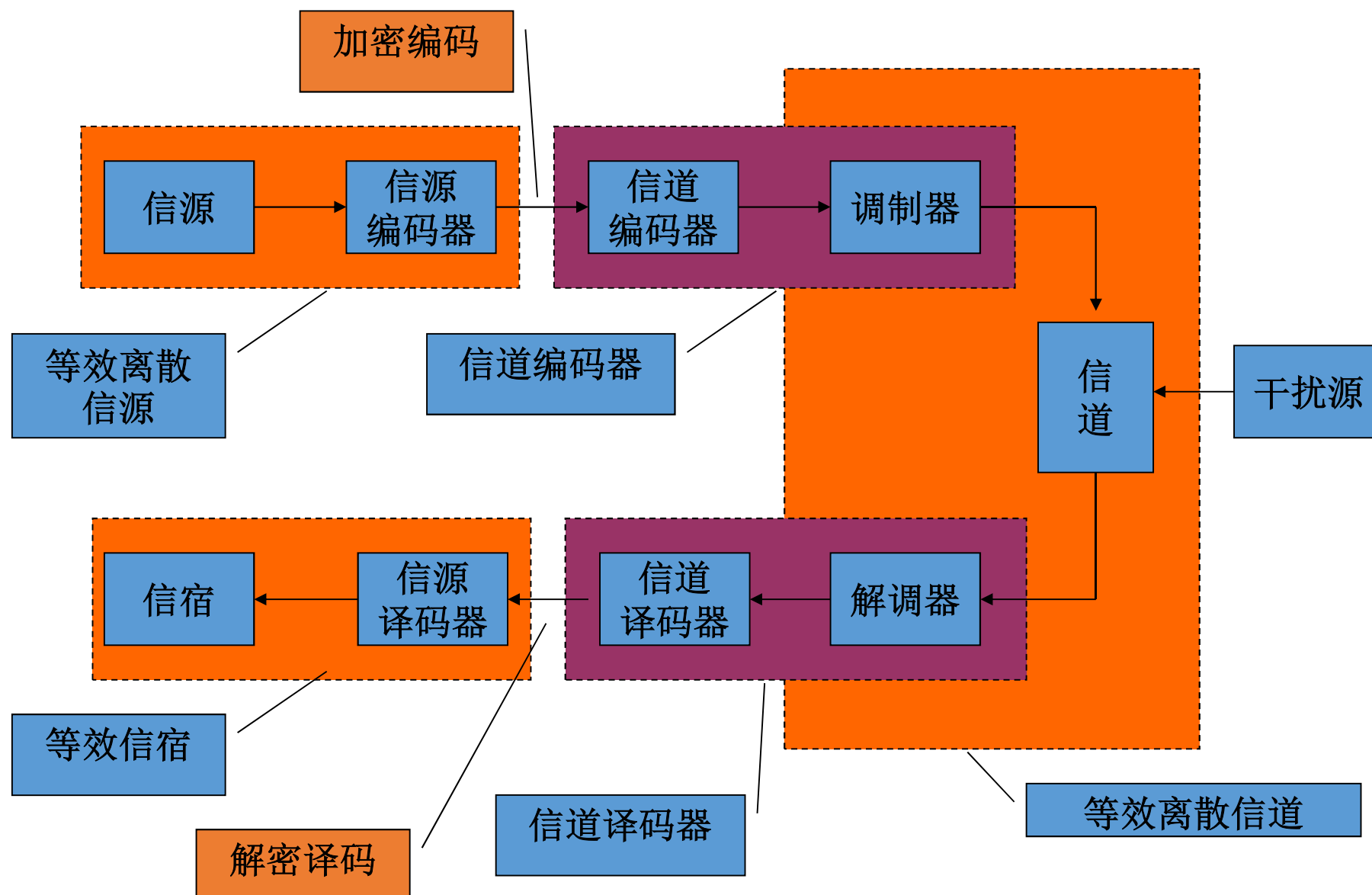
- 译码器的功能：从接收到的信号中恢复消息。
- 包括：
 - 解调器
 - 信道译码器
 - 信源译码器



1. 通信系统模型

5) 信宿

- 信宿是消息传送的对象（人或机器）。
- 香农信息论不研究信宿。



2. 提高通信系统性能指标的措施

提高有效性：（数据压缩）

➤信源编码：无失真信源编码和限失真信源编码

提高可靠性：（可靠传输）

➤信道编码

第一章：绪论

一、什么是信息

二、通信系统模型

三、信息论的研究内容

四、信息论的形成和发展

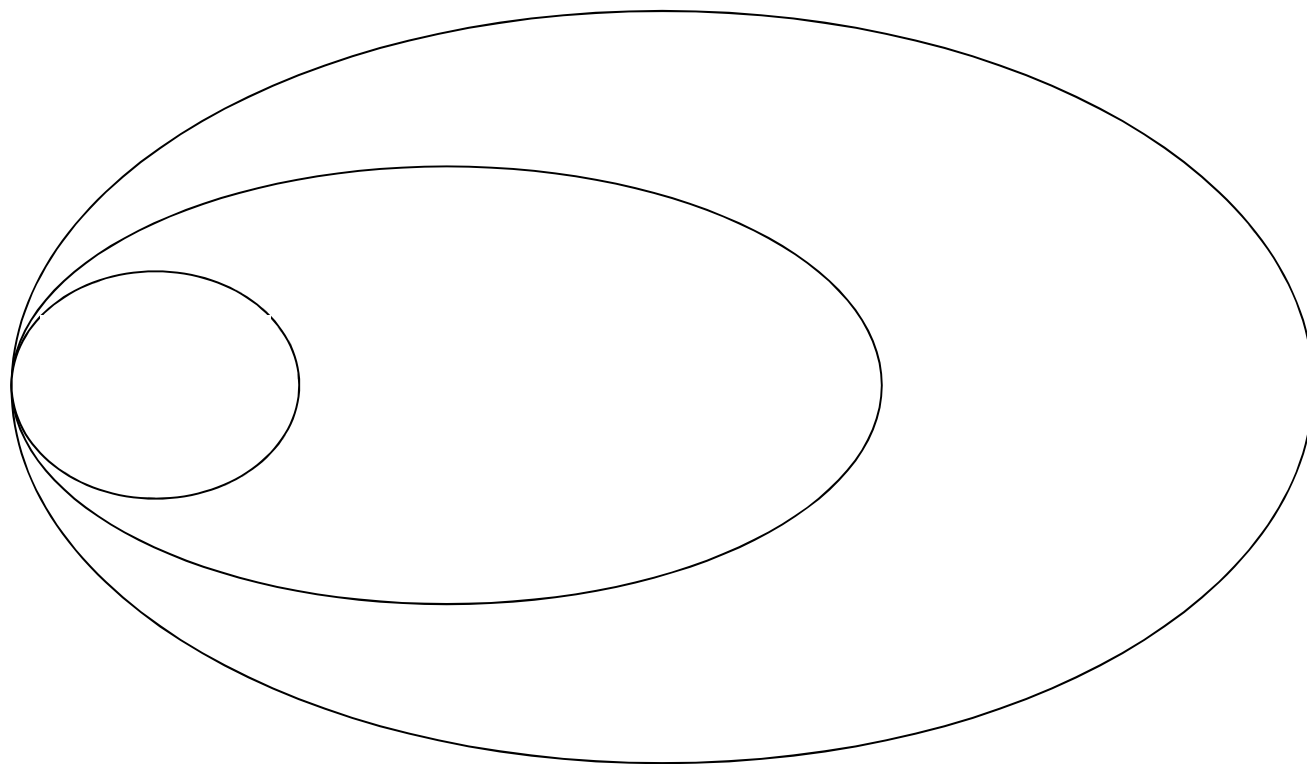
1. 信息论研究的主要问题

2. 什么是信息论

3. 信息论的应用

1. 信息论研究的主要问题

- 狭义信息论：又称香农信息论。
- 一般信息论：也叫工程信息论。
- 广义信息论



1. 信息论研究的主要问题

- 1) 什么是信息？如何度量信息？
- 2) 怎样确定信源输出信息的速率？
- 3) 对于一个信道，它传输信息的最高速率（信道容量）是多少？

1. 信息论研究的主要问题

4) 无失真信源编码，所需要的最少码符号数是多少？

香农第一定理：如果编码后的信源序列的编码信息率不小于信源的熵，那么一定存在一种无失真信源编码方法；否则，不存在这样的一种无失真信源编码方法。

1. 信息论研究的主要问题

5) 在有噪声信道中，有没有可能实现几乎无差错的传输信息？

香农第二定理：如果信道的信息传输率小于信道容量，那么总可以找到一种编码方式，使得当编码序列足够长时传输差错任意小；否则，不存在使差错任意小的信道编码方式。

1. 信息论研究的主要问题

6) 如果信源编码时，允许一定的失真，那么信源编码所需的最少码符号数又是多少？

香农第三定理：对于任意的失真度 $D \geq 0$ ，只要码字足够长，那么总可以找到一种编码方法，使编码后的编码信息率 $\geq R(D)$ ，而码的平均失真度 $d \leq D$ 。

2. 什么是信息论

狭义信息论（**Shannon**经典信息论）

香农信息论的内容可用一句话概况为

“一个概念、三个定理”

(1) 信源信息的度量——信息熵

(2) 无失真信源编码——香农第一定理

该定理解决信源无损压缩极限的理论问题，未给出普遍的编码方法。

(3) 有噪信道编码——香农第二定理

该定理解决信息传输极限的理论问题，未给出普遍的信道编码方法。

(4) 保真度准则下信源编码——香农第三定理

该定理解决有损压缩极限的理论问题，未给出编码方法。

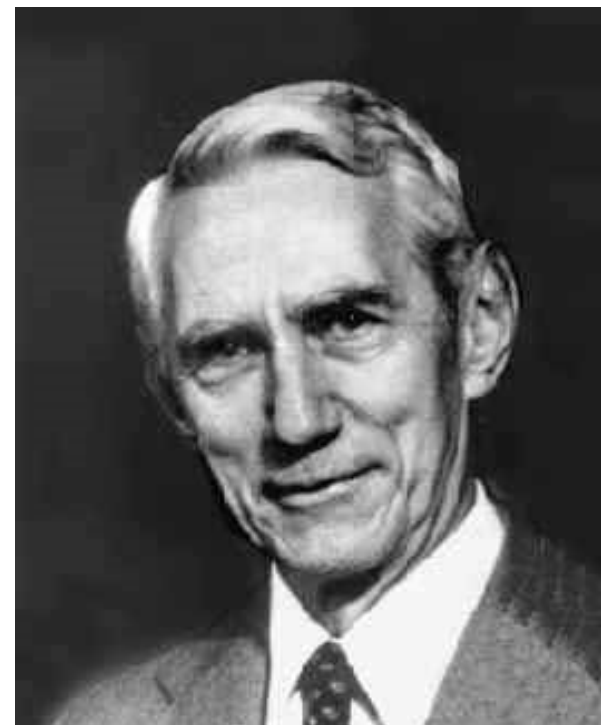
2. 什么是信息论

- 信息论是通信的数学基础，它以概率论为主要数学工具，详细研究了通信中的各个环节，以定理的形式给出了信源编码、信道编码的理论极限，为各种具体的通信技术提供了理论上的指导。

- 信息论创立的标志：香农于1948年发表的论文：

—— **A Mathematical Theory of Communication**

（通信的数学理论）



2. 什么是信息论

信息论的特点

以概率论、随机过程为基本研究工具。

研究的是通信系统的整个过程，而不是单个环节，并以编、译码器为重点。

关心的是最优系统的性能和怎样达到这个性能（并不具体设计系统）。

要求信源为随机过程，不研究信宿。

2. 信息论的应用

- 信息论帮助通信工程师从全局的观点观察和设计通信系统。
- 信息论是从事信息通信系统研究和开发的必备的知识。
- 香农信息论的目标是研究通信系统的信息传递，而不是帮助人们理解信息含义。香农信息论有它的局限性。

2. 信息论的应用

通信的基本问题是在一点精确地或近似地恢复另一点（信源）所选择的消息。通常，这些消息是有含义的，但是这些语义方面的问题与通信问题无关，而重要的方面是实际消息是从一个可能的消息集合中选择出的一条消息。

—— 香农

2. 信息论的应用

- 信息论的应用举例

- 语音信号压缩（G.711, GSM, Vocoder ...）
- 计算机文件压缩
- 模拟话路中数据传输速率的提高
- 其他（音频信号压缩 MP3、图象信号的压缩JPEG, MPEG等）

第一章：绪论

一、什么是信息

二、通信系统模型

三、信息论的研究内容

四、信息论的形成和发展

1. 技术背景

2. 香农的主要工作

1. 技术背景

人类通讯技术的发展

- 公元前**1775**年，希腊文字发明
- 公元前**1400**年，中国甲骨文
- 公元前**8**世纪，烽火通讯（“烽火戏诸侯”）——光、无线



1. 技术背景

近现代的通讯技术

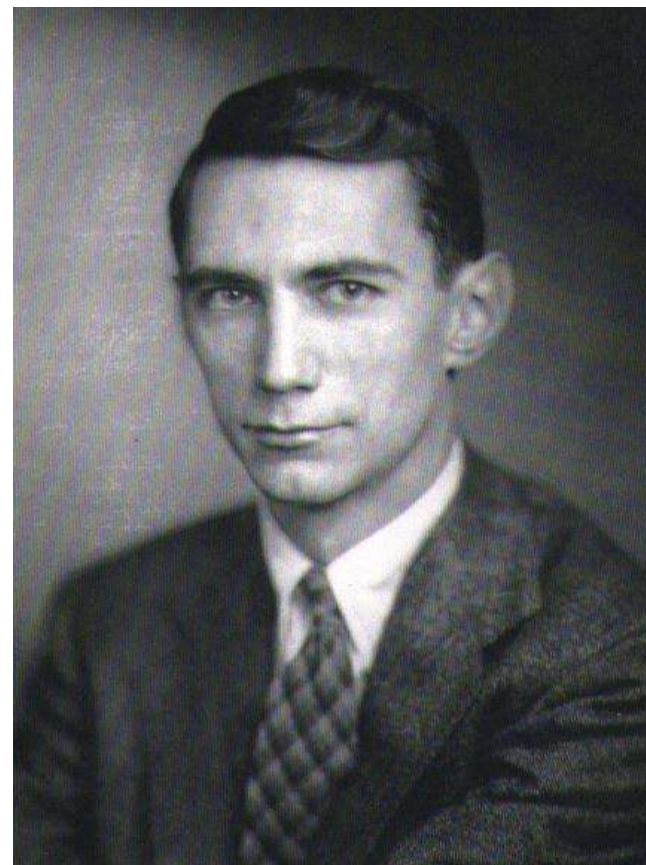
- 19世纪初，**Morse**电码发明
- 1839年，英国开始商业运行
- 1844年，美国开始商业运行
- 1850-1866年，洲际海底电缆
- 1877年，电话诞生
- 1927年，**NBC**开播两套无线节目
- 1939年，电视广播开始
- 1946年，宾夕法尼亚大学制造第一台电子计算机
- 1957年，苏联发射人造卫星**Sputnik**
- 1969年，**ARPANet**开始运营
- 1977年，旅行者从木星发回信号
- 1981年，蜂窝电话商用
- 1991年，**World Wide Web** 发明

1. 技术背景

- 电报 (**Morse, 1838**) → 考虑字母出现频率的编码方案
 - 电话 (**Bell, 1876**)
 - 无线电报(**Marconi,1887**)
 - 调幅广播 (**1900's 早期**)
 - 单边带调制 (**Carson, 1922**)
 - 电视 (**1925-1927**)
 - 调频广播 (**Armstrong, 1936**)
 - 脉冲编码调制 (**Reeves, 1937-1939**)
 - 声码器 (**Dudley, 1939**)
 - 扩频通信 (**1940's**) → 以降低信号还原度为代价，传输带宽可以小于信号带宽
- 信息传送速率与可靠性是相互依存的两个自由度

2. 香农

Claude Elwood Shannon (1916-2001) 1916年4月30日出生于美国密西根州，是一位美国数学工程师。他在通信技术与工程方面的创造性工作，为计算机与远程通信奠定了坚实的理论基础；作为信息论的创始人，人们尊崇香农为**信息论及数字通信时代**的奠基之父，被认为是**20世纪最伟大的科学家之一**。



2. 香农的主要工作

1936年香农在密西根大学获得数学与电气工程学士学位，然后进入MIT念研究生。

1938年，22岁的Shannon在MIT获得了电气工程的硕士学位，并且完成了一篇名为《A Symbolic Analysis of Relay and Switching Circuits》(继电器与开关电路的符号分析)的硕士学位论文。奠定了数字电路的数学基础，开启了通向电路逻辑的大门。

1940年，Shannon在MIT获得了数学博士学位，他的博士论文题目是《An Algebra for Theoretical Genetics》(理论遗传学的代数学)。

1941年，Shannon进入了AT&T的贝尔实验室作了一名数学研究员，在贝尔实验室工作到1972年，从24岁到55岁，整整31年。1956年他当了MIT的访问教授，1958年成为正式教授，1978年退休。

2. 香农的主要工作

1948年，发表《通信的数学理论》。

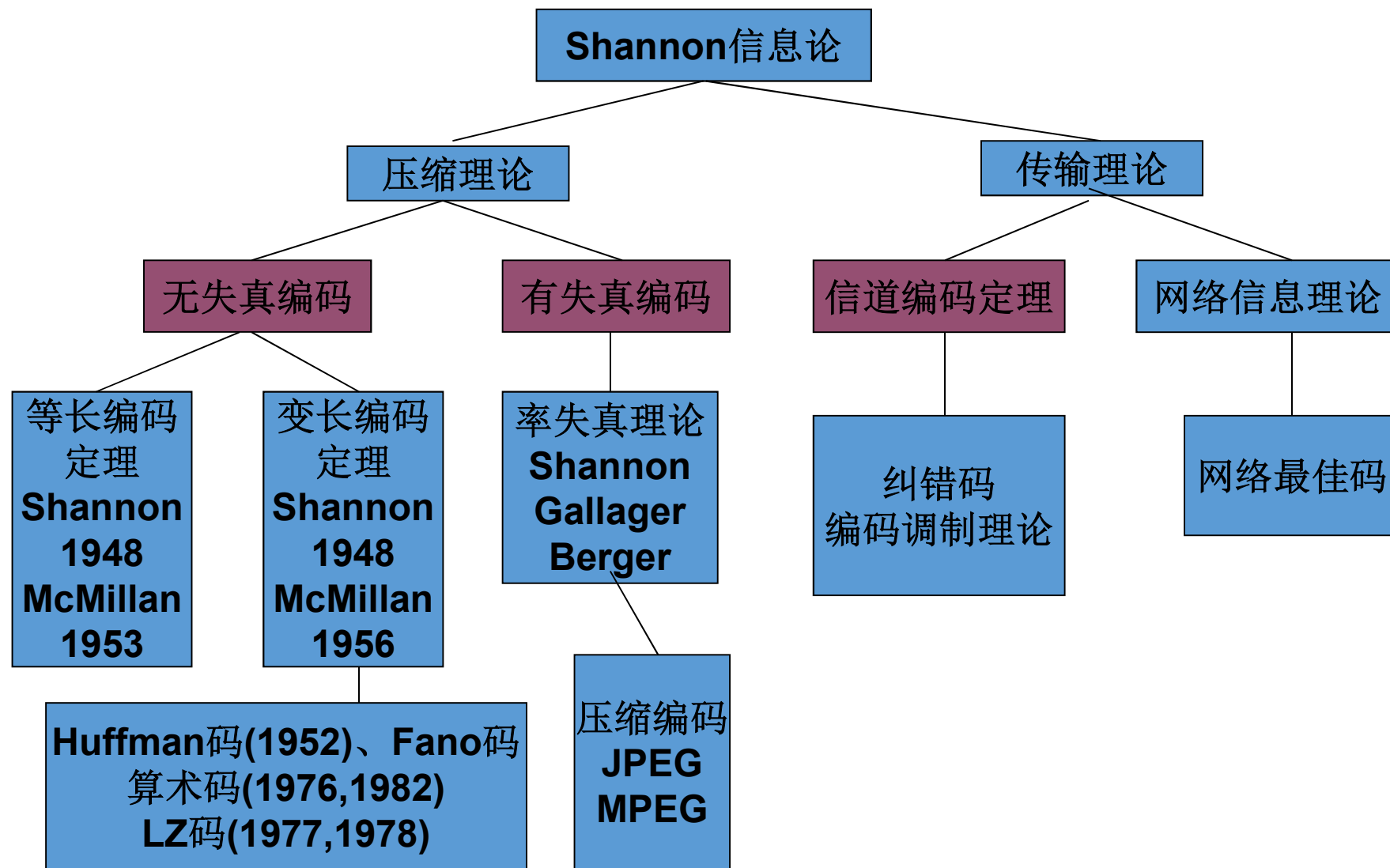
《A Mathematical Theory of Communication》

1949年，发表《噪声下的通信》。

1956年，发表《噪声信道的零差错容量》。

1959年，发表《在保真度准则下的离散信源编码定理》。

1961年，发表《双路通信系统》。



狭义信息论的科学体系

本课程约定的符号表示

- 大写字母等表示随机变量 X, Y, X_1, \dots, X_N
- 小写字母等表示随机变量的具体取值 x, y, x_1, \dots, x_N
- 大写黑体字母 \mathbf{X}, \mathbf{Y} 等表示多维随机变量，也就是随机矢量
- 小写黑体字母 \mathbf{x}, \mathbf{y} 等表示随机矢量的具体取值