Principle of Digital Communication

---Lecture 1 Introduction

How can you get your marks?

Attendances 16%
Homeworks 14% (2)
Quizes 20% (2)
Final report 50% (1)
---including doing simulations and theoretical analysis

Contact me

Dr. Zhenzhen Gao
http://gr.xjtu.edu.cn/web/zhenzhen.gao
Email: zhenzhengao@xjtu.edu.cn
Address: No. 1 West building, Room No. 446

About the course

It’s a major course that you take as a graduate student in the communication area.

I suppose you have taken these prior courses:
Information theory
stochastic process
principle of communications
Signals and systems

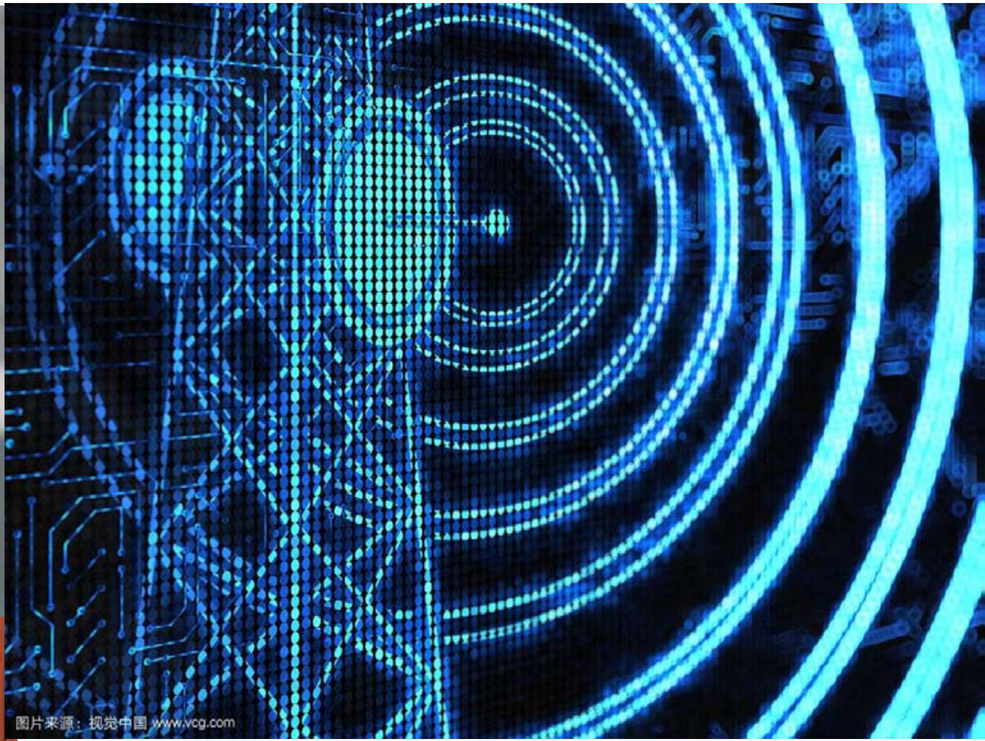
The following books are used as the main material for this course.
The first one is the lecture notes of MIT.
The book has unique insights in digital communications.
The course is given by Prof. G. Gallager, who is a student of Claude Shannon.

You can get more details including videos from MIT open courses
http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-450-principles-of-digital-communications-i-fall-2006/video-lectures/

The second one is organized in a more conventional way
More specific things can be found in this book.

I will use both of them to make this course
have some depth and also easy to understand.

Lecture 1 Introduction to digital communications



Why do we call it digital communication?

Digital communication systems, by definition, are communication systems that use such a digital sequence as an interface between the source and the channel input (and similarly between the channel output and final destination).

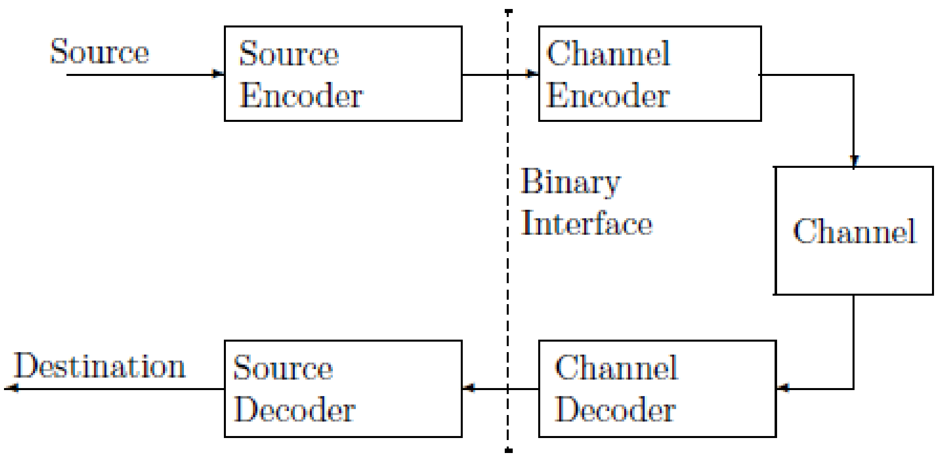


Fig. 1.1 Placing a **binary interface** between source and channel.

To view all communication sources, e.g., speech waveforms, image waveforms, and text files, as being representable by binary sequences

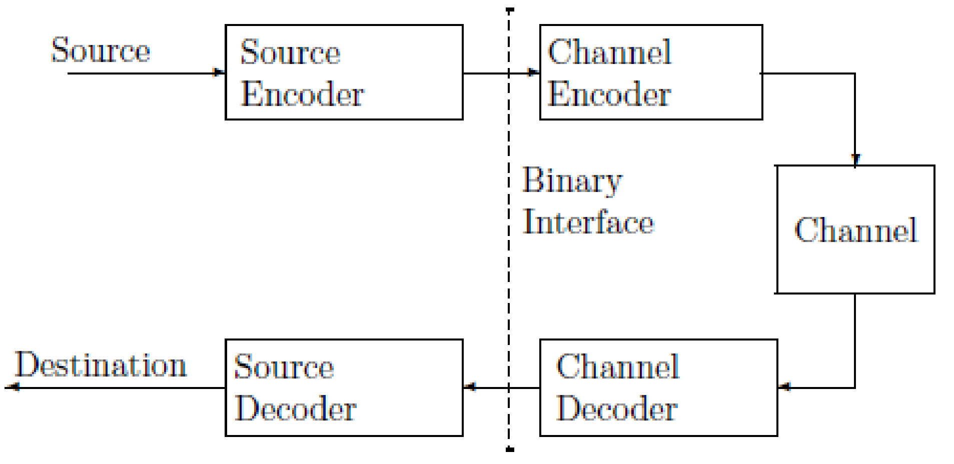


Fig. 1.1 Placing a **binary interface** between source and channel.

To convert that binary sequence into a form suitable for transmission over particular physical media such as cable, twisted wire pair, optical fiber, or electromagnetic radiation through space.

By today, with digital cameras, digital video, digital voice, etc., the idea of digitizing any kind of source is commonplace. The notion of a binary interface before channel transmission is almost as commonplace.

For example, we all refer to the speed of our internet connection in bits per second.

why digital communication systems are now standard？

Why digital communication systems are now standard?

- Digital hardware has become so cheap, reliable, and miniaturized, that digital interfaces are quite practical.

- A standardized binary interface between source and channel simplifies networking, which now reduces to sending binary sequences through the network.

- A standardized binary interface between source and channel simplifies implementation and understanding, since source coding/decoding can be done independently of the channel and vice versa.

source/channel separation theorem

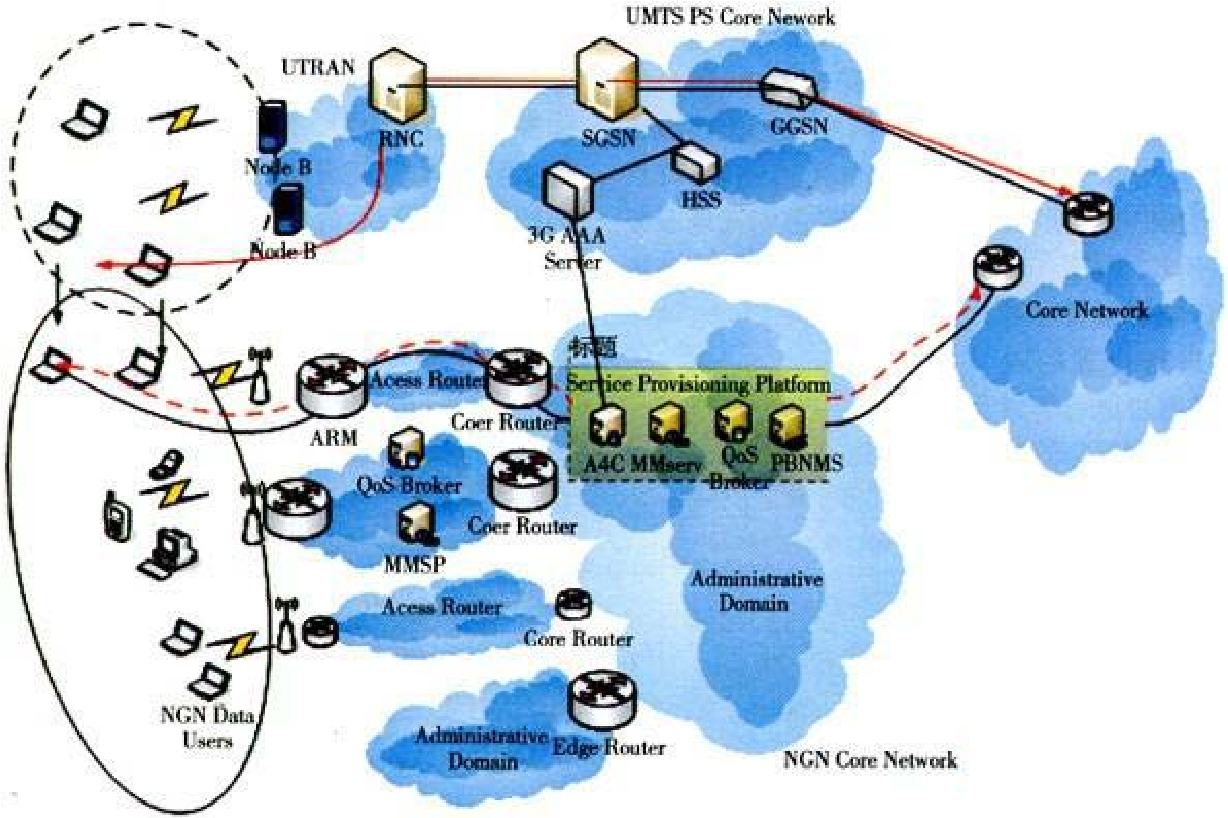
(stated for stationary memoryless sources and channels by Claude Shannon in 1948)

A direct part that states that if the minimum achievable source coding rate of a given source is strictly below the capacity of a channel, then the source can be reliably transmitted through the channel by appropriate encoding and decoding operations;

→ reliable transmission can be accomplished by separate source and channel coding, where the source (resp., channel) encoder and decoder need not take into account the channel (resp., source) statistics.

A converse part stating that if the source coding rate is strictly greater than capacity, then reliable transmission is impossible.

→ either reliable transmission is possible by separate source-channel coding or it is not possible at all.



Such complex networks need to be based on some simple architectural principles in order to be understood, managed, and maintained.

Two such fundamental architectural principles are standardized interfaces and layering.

1.1 Standardized interfaces and layering

A standardized interface allows the user or equipment on one side of the interface to ignore all details about the other side of the interface except for certain specified interface characteristics.

The idea of layering in communication systems is to break up communication functions into a string of separate layers as illustrated in Figure 1.2.

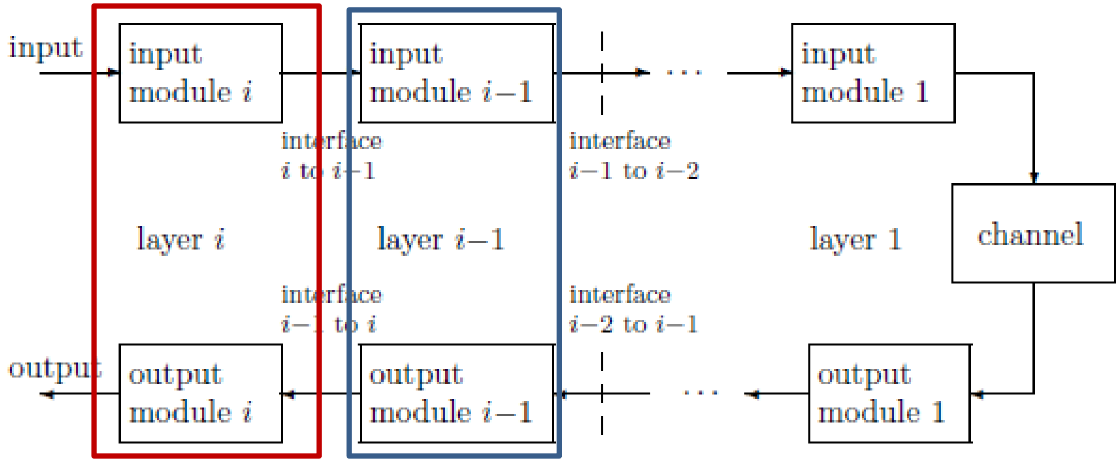
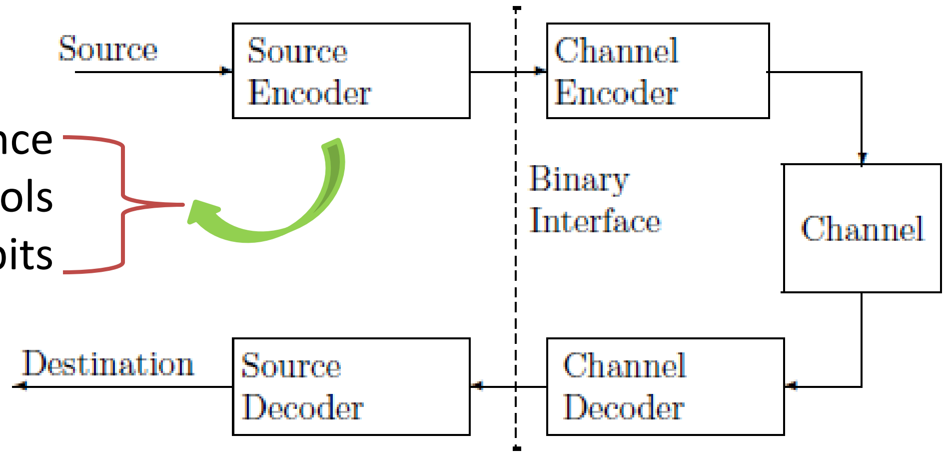


Fig. 1.2 Layers and interfaces

As an example, the source coding/decoding layer for a waveform source can be split into 3 layers as shown in Figure 1.3

Analog waveform to sequence
Sequence of numbers to symbols
Symbols to bits



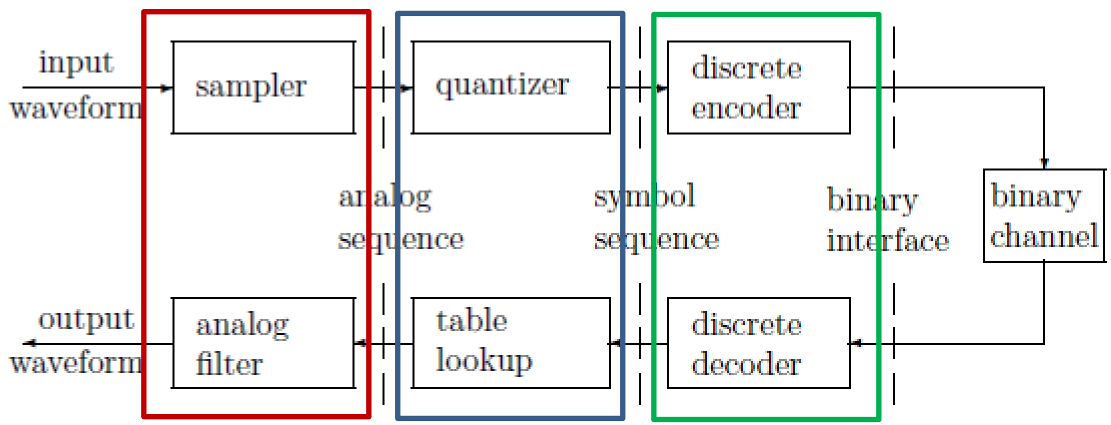
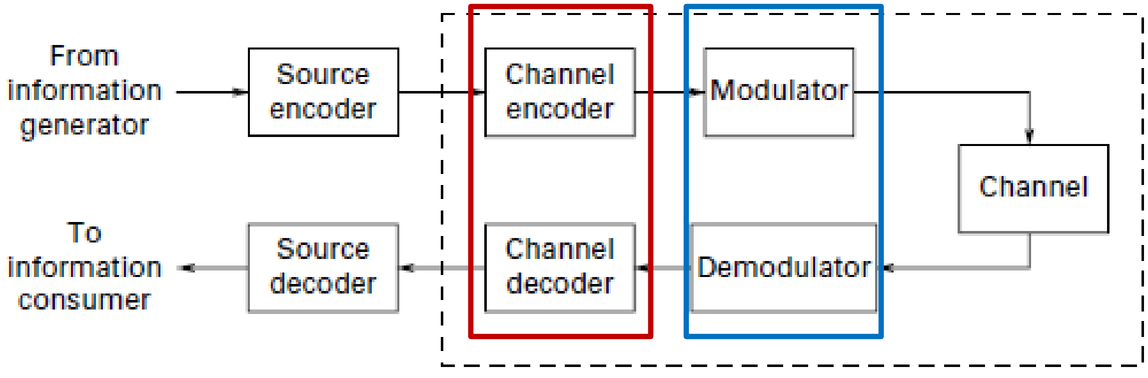
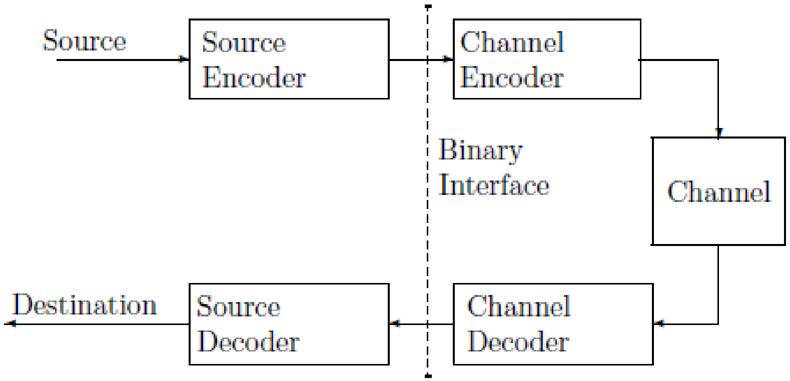


Fig. 1.3 Breaking the source coding/decoding layer into 3 layers for a waveform source



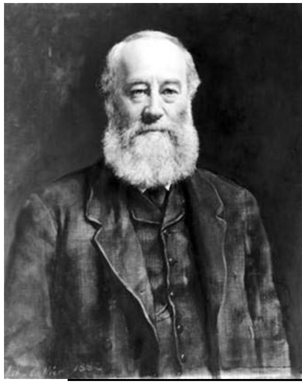
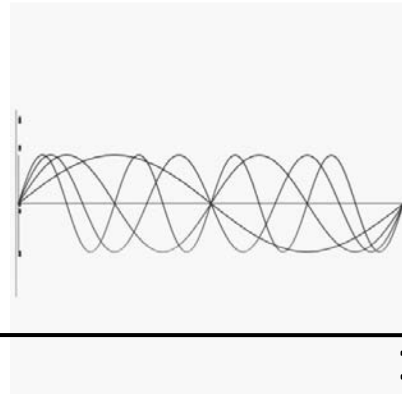
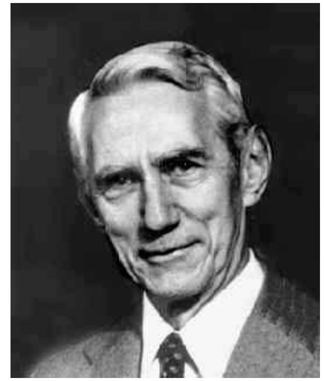
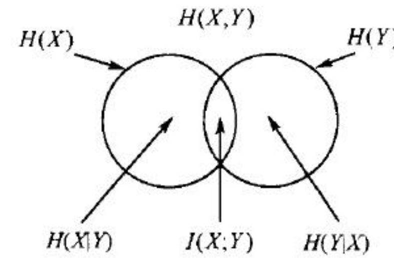


The channel coding/decoding layer can also be split into several layers, but there are a number of ways to do this. For example, binary error-correction coding/decoding can be used as an outer layer with modulation and demodulation as an inner layer.

1.2 Communication sources

The source might be discrete, i.e., it might produce a sequence of discrete symbols, such as letters from the English or Chinese alphabet, binary symbols from a computer file, etc.
Alternatively, **the source might produce an analog waveform**, such as a voice signal from a microphone, the output of a sensor, etc.

Whatever the nature of the source, the output from the source will be modeled as **a sample function of a random process**.
It is not obvious why the inputs to communication systems should be modeled as random, and in fact this was not appreciated before Shannon developed information theory in 1948.

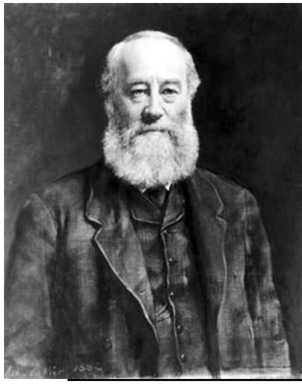





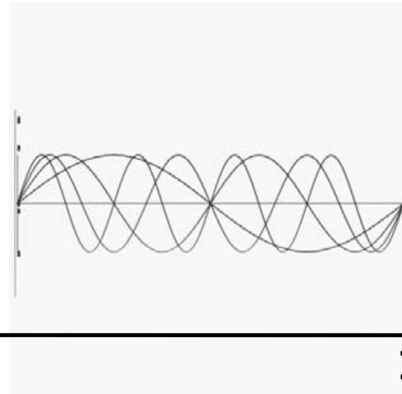
1948

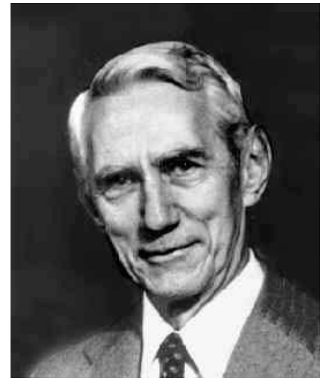
The study of communication before 1948 was based on Fourier analysis; basically one studied the effect of passing sine waves through various kinds of systems and components and viewed the source signal as a superposition of sine waves.

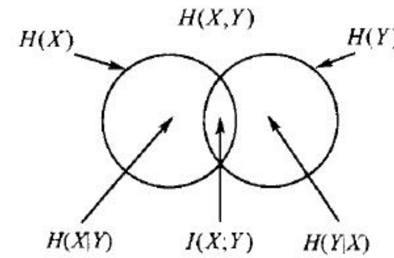
Shannon's view, however, was that if the recipient knows that a sine wave of a given frequency is to be communicated, why not simply regenerate it at the output rather than send it over a long distance?

Or, if the recipient knows that a sine wave of unknown frequency is to be communicated, why not simply send the frequency rather than the entire waveform?







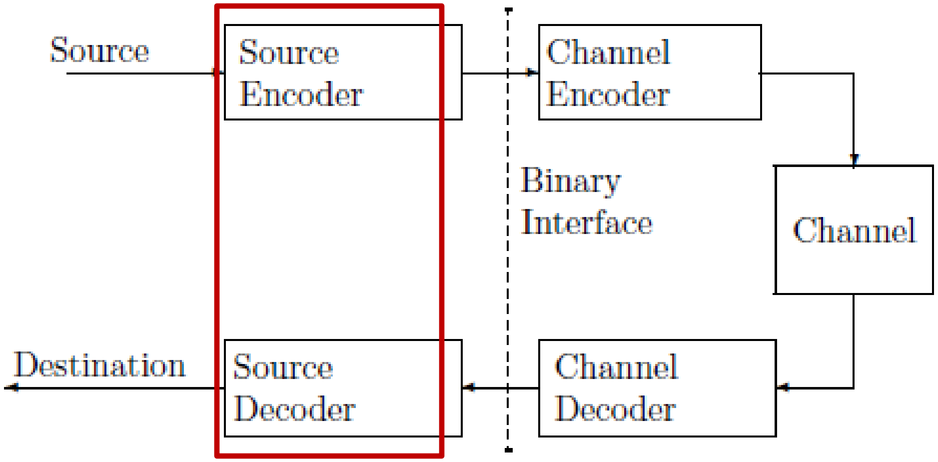


1948

The essence of Shannon's viewpoint is that the set of possible source outputs, rather than any particular output, is of primary interest.

Tthe communication system must be designed to communicate whichever one of these possible source outputs actually occurs. The objective of the communication system then is to transform each possible source output into a transmitted signal in such a way that these possible transmitted signals can be best distinguished at the channel output. A probability measure is needed on this set of possible source outputs to distinguish the typical from the atypical.

1.2.1 Source coding



The source encoder has the function of converting the input from its original form into a sequence of bits.

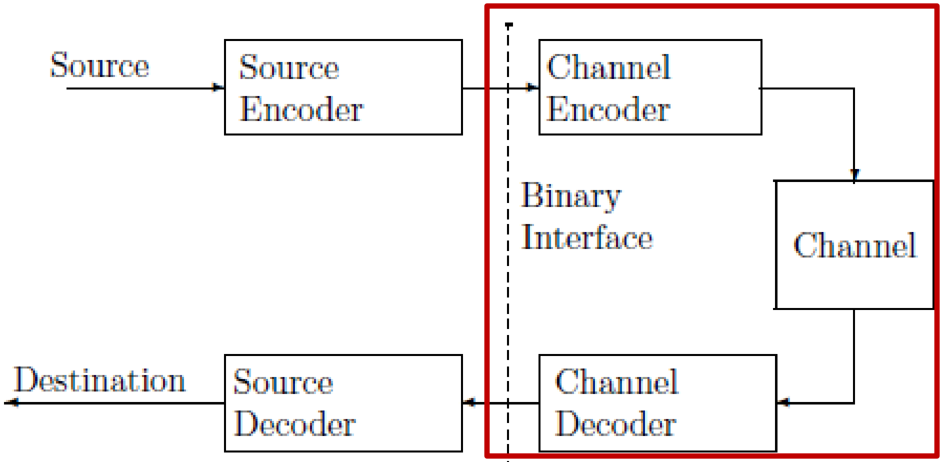
As discussed before, the major reasons for this almost universal conversion to a bit sequence are as follows: inexpensive digital hardware, standardized interfaces, layering, and the source/channel separation theorem.

The most straightforward approach to analog source coding is called analog to digital (A/D) conversion.
The source waveform is first **sampled** at a sufficiently high rate.
Each sample is then **quantized** sufficiently finely for adequate reproduction.
The problems of sampling and quantization are largely separable.

Beyond the basic objective of conversion to bits, the source encoder often has the further objective of **doing this as efficiently as possible**— i.e., transmitting as few bits as possible, subject to the need to reconstruct the input adequately at the output. In this case source encoding is often called data compression.

Coding for discrete sources is introduced in Chapter 2 of Prof. Gallager’s book. Quantization is introduced in Chapter 3. These are not in the interest of this course.

1.3. Communication channels



In general, a channel is viewed as that part of the communication system between source and destination that is given and **not under the control of the designer**.

It is more common to view the coupling devices (the amplifiers, antennas, etc) as part of the channel, since their design is quite separable from that of the rest of the channel encoder. This is another example of layering.

noise waveform

a waveform input

a waveform output

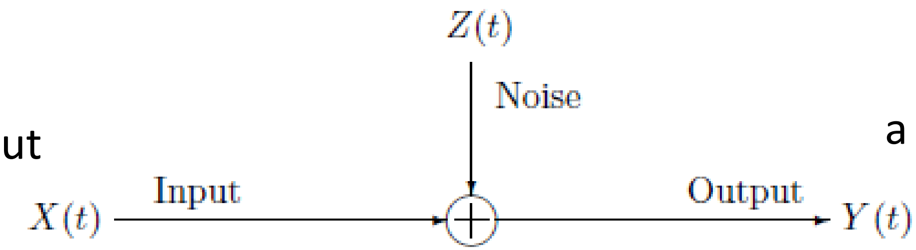


Fig. 1.4 An additive white Gaussian noise (AWGN) channel.

1. Each of these waveforms are viewed as random processes.
2. The noise Z(t) is often modeled as white Gaussian noise.
3. The input is usually constrained in power and bandwidth.

The additional required ingredient for noise to be called additive is that its probabilistic characterization does not depend on the input.

The channel capacity for a band-limited additive white Gaussian noise channel is perhaps the most famous result in information theory. If the input power is limited to P, the bandwidth limited to W, and the noise power per unit bandwidth is N_0, then the capacity (in bits per second) is

$$C = \mathrm{W} \log_2 \left(1 + \frac{P}{N_0 \mathrm{W}} \right)$$

In a somewhat more general model, called a linear Gaussian channel, the input waveform X(t) is first filtered in a linear filter with impulse response h(t), and then independent white Gaussian noise Z(t) is added, as shown in Figure 1.5, so that the channel output is

$$Y(t) = X(t) * h(t) + Z(t),$$

where "*" denotes convolution.

$$Y(t) = \int_{-\infty}^{\infty} X(t-\tau)h(\tau)\,d\tau + Z(t)$$

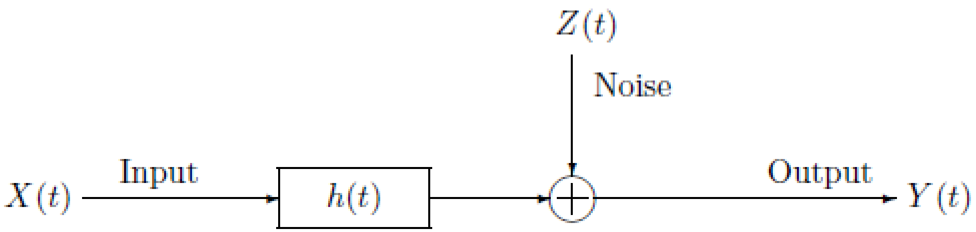


Fig. 1.5 Linear Gaussian channel model

The linear Gaussian channel is often a good model for wireline communication and for line-of-sight wireless communication.

1.3.1 Modulation

Modulation in analog communication

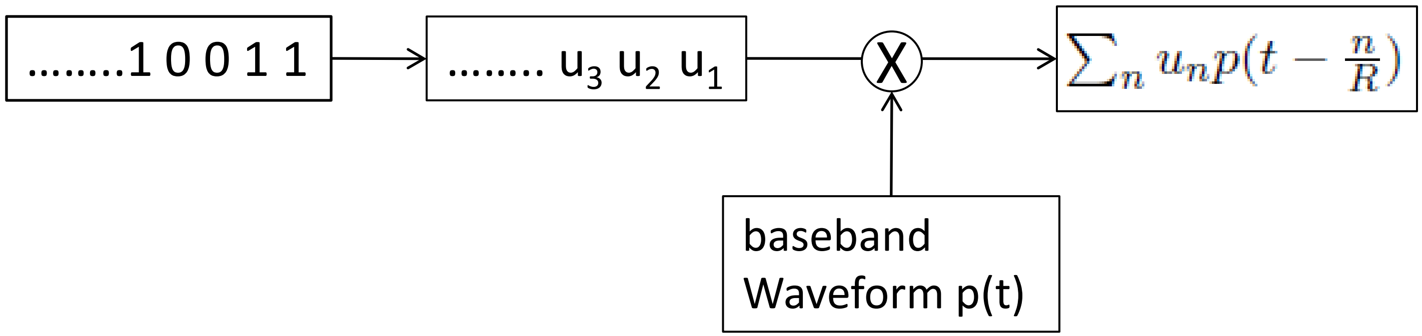
The terminology comes from the days of analog communication where modulation referred to the process of combining a lowpass signal waveform with a high frequency sinusoid, thus placing the signal waveform in a frequency band appropriate for transmission and regulatory requirements.

The analog signal waveform could modulate the amplitude, frequency, or phase of the sinusoid, but in any case, the original waveform (in the absence of noise) could be retrieved by demodulation.

Modulation in digital communication

As digital communication has increasingly replaced analog communication, modulation/demodulation is to map the binary sequence at the source/channel interface into a baseband waveform.

The resulting baseband waveform is converted to bandpass by the same type of procedure used for analog modulation, and the conversion of baseband to passband and back will be referred to as frequency conversion.



The modulated waveform can be a real baseband waveform, such as the simple 2-PAM scheme. The modulated waveform can be a complex baseband waveform (which is then modulated up to an appropriate passband as a real waveform).

It is also possible to choose M different signal pulses, $p_1(t)$,... ,$p_M(t)$. This includes frequency shift keying, pulse position modulation, phase modulation, and a host of other strategies.

Even with the trivially simple modulation scheme, there are a number of interesting questions, such as

1. **How to choose the elementary waveform p(t) so as to satisfy frequency constraints?**
2. **How to reliably detect the binary digits from the received waveform in the presence of noise and intersymbol interference?**

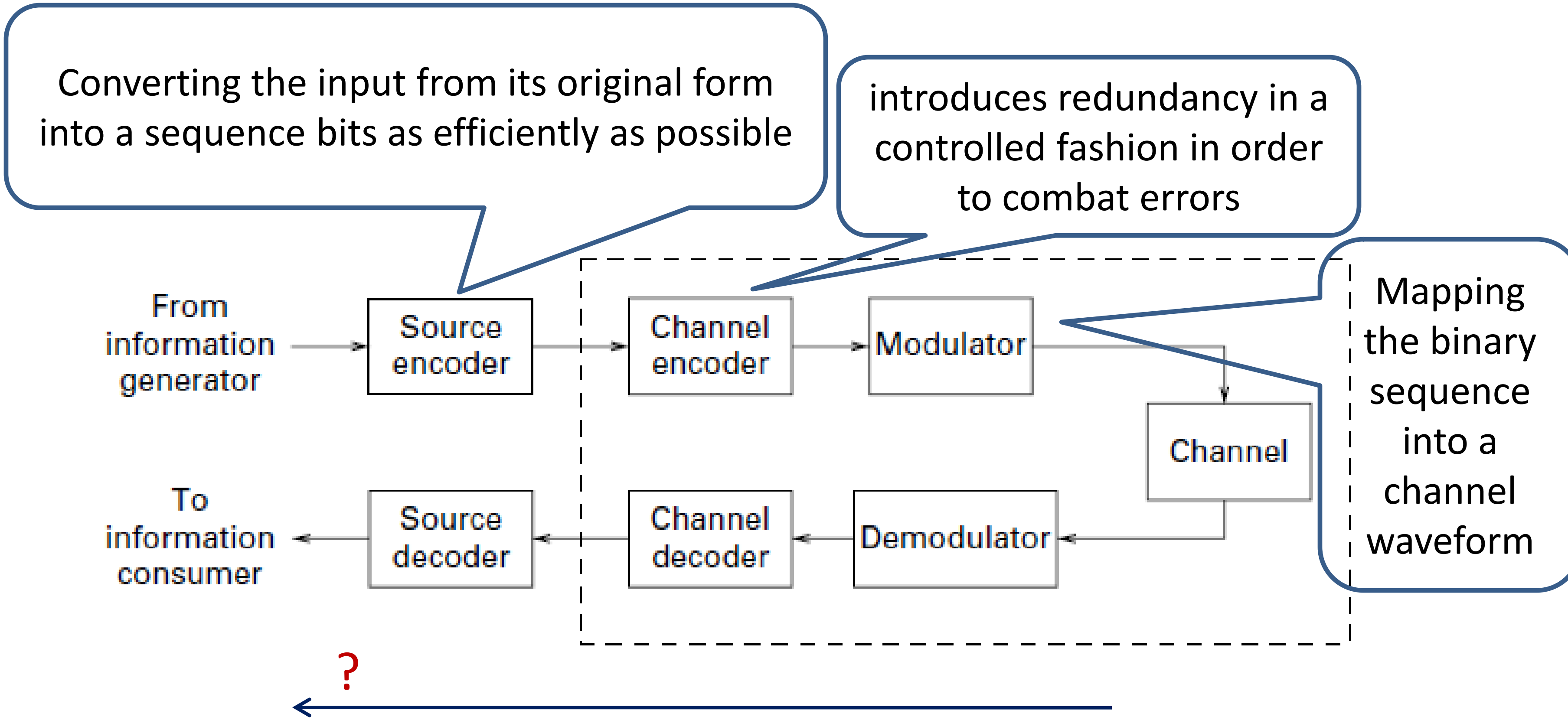
1.3.2 Error correction

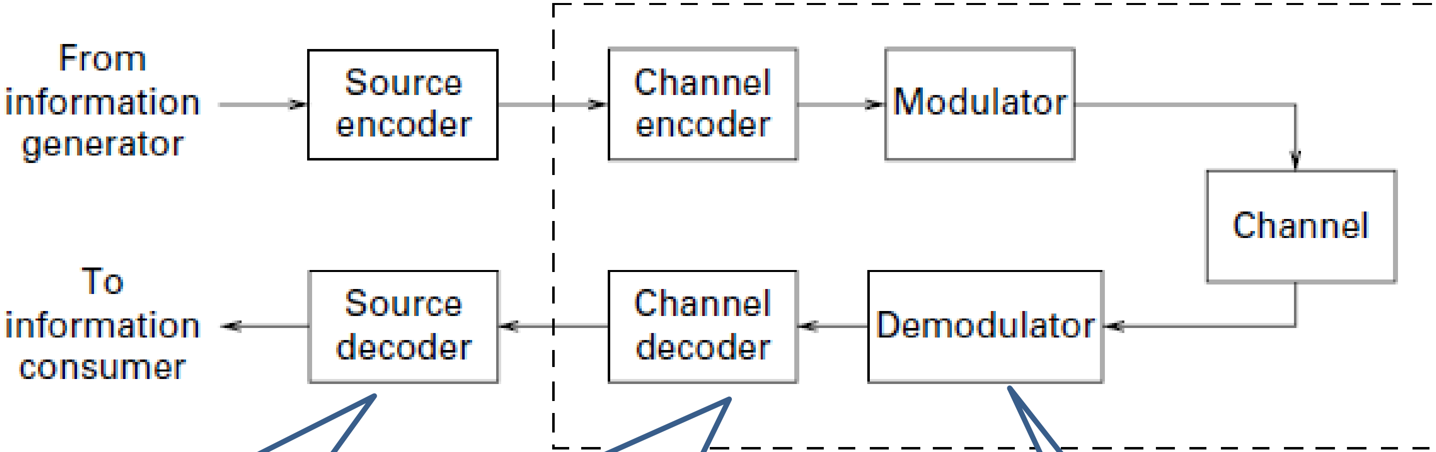
Frequently the error probability incurred with simple modulation and demodulation techniques is too high. One possible solution is to separate the channel encoder into two layers, first an error-correcting code, and then a simple modulator.

As a very simple example, the bit rate into the channel encoder could be reduced by a factor of 3, and then each binary input could be repeated 3 times before entering the modulator. If at most one of the 3 binary digits coming out of the demodulator were incorrect, it could be corrected by majority rule at the decoder, thus reducing the error probability of the system at a considerable cost in data rate.

What Shannon showed was that more sophisticated coding schemes can achieve arbitrarily low error probability at any data rate below a value known as the channel capacity. Conversely, it is not possible to achieve low error probability at rates above the channel capacity.
In the past few years channel coding schemes have been developed that can closely approach this channel capacity.

A brief proof of this channel coding theorem is given in Chapter 8 of Prof. Gallager's book.
However, error correction is not the major topic in this course.


Converting the input from its original form into a sequence bits as efficiently as possible
introduces redundancy in a controlled fashion in order to combat errors
Mapping the binary sequence into a channel waveform
From information generator
Source encoder
Channel encoder
Modulator
Channel
Demodulator
Channel decoder
Source decoder
To information consumer
?



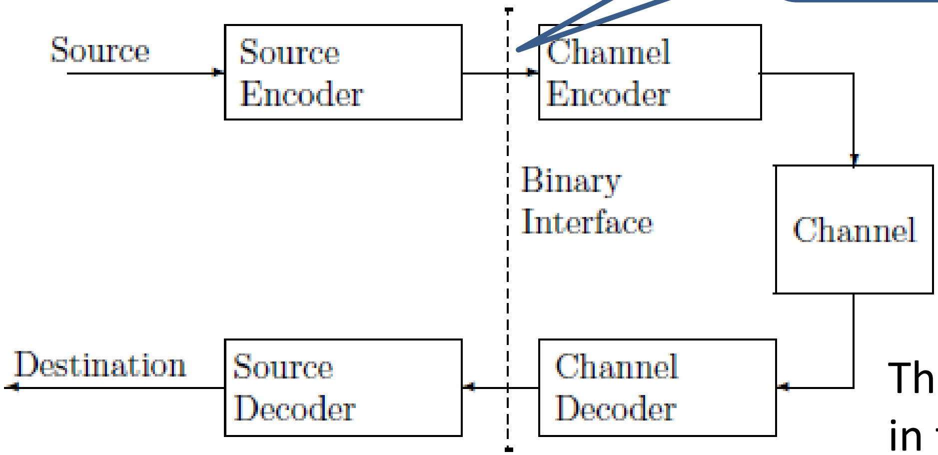
converts the estimated information bits produced by the channel decoder into a format that can be used by the end user

exploits the redundancy of the channel code to improve the estimates from the demodulator,
to produce an estimate of the sequence of information bits

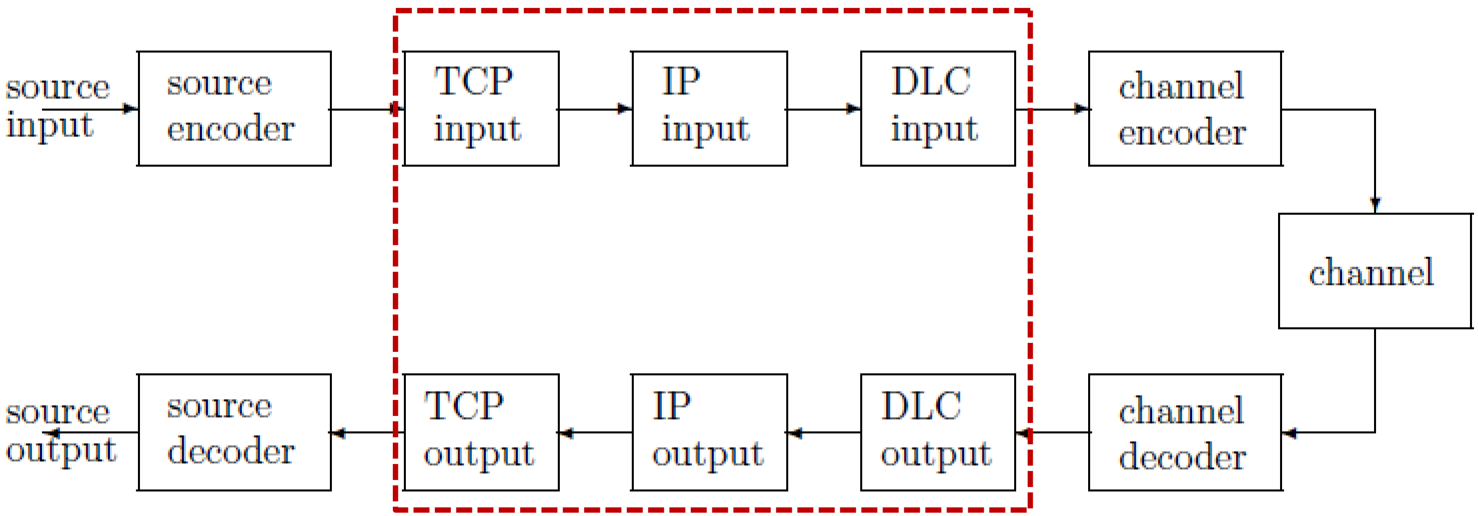
Synchronization, channel equalization, to produce estimated decisions on the transmitted symbols

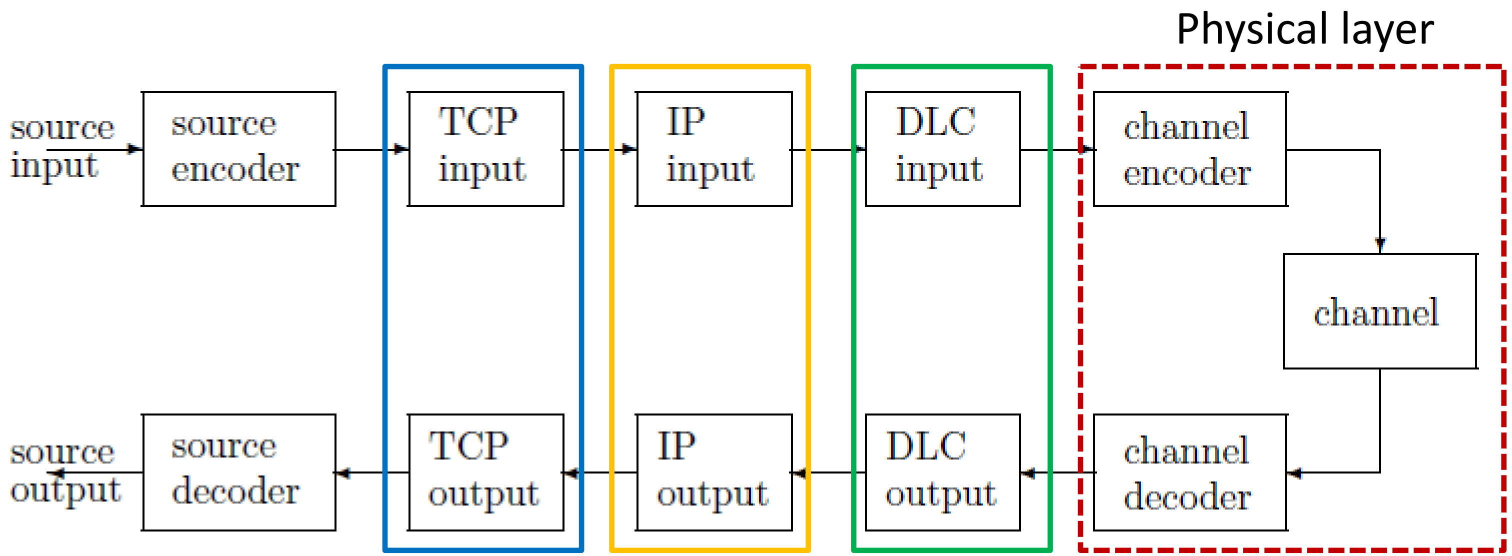
1.4. Digital interface

- Unequal rates
- Errors
- Networks



The replacement of the binary interface in the Figure with 3 layers in an oversimplified view of the internet





TCP(transport control protocol) module **associated with each source/destination pair**; this is responsible for end-to-end error recovery and for slowing down the source when the network becomes congested.

IP (internet protocol) module **associated with each node in the network**; these modules work together to route data through the network and to reduce congestion.

DLC (data link control) module **associated with each channel**; this accomplishes rate matching and error recovery on the channel