

# INTRODUCTION TO DIGITAL COMMUNICATIONS (DC I)

DIRK DAHLHAUS

COMMUNICATIONS LABORATORY [COMLAB]

Winter Semester 2018-2019

# Organisation

- time:
  - winter semester (WS), annually, Thursday: 12:00-14:00 /Friday: 8:00-10:00
  - summer semester (SS), annually, Wednesday: 8:00-10:00/Thursday: 14:00-16:00
- place: WS: (-1319/-1319), SS: (1332/-1418).
- workload: 60 hours course attendance, 120 hours self-study
- language: English, oral exam (30 minutes either in English or in German)
- exercises: are integrated in the lecture (on demand)
- regular attendance of the lecture *and* the exercises is mandatory to pass the exam
- the lecture is based on the books (well-known in communications)
  - John G. Proakis, „Digital Communications“, McGraw-Hill, 4th ed., ISBN 0-07-118183-0
  - Papoulis S. U. Pillai, „Probability, Random Variables, and Stochastic Processes“, McGraw-Hill, 4th ed., ISBN 0071226613
- upon passing the exam and after attending Introduction to Digital Communications Lab, you obtain **6 credit points for lecture, exercises and lab**

# Table of Contents

## 1 Introduction

- Mathematical Models for Communication Channels
- Generalized Functions

## 2 Probability

- Fundamentals in Probability Theory
- Useful Probability Distributions
- Central Limit Theorem

## 3 Fourier Transform

- Properties of FT
- Sampling Theorem

## 4 Stochastic Processes

- Gaussian Processes
- Response of an LTI System to a Stationary Input Signal
- Discrete-Time (Stochastic) Signals
- Cyclostationary Processes

## 5 Communication Signals and Systems

- Representation of Band-pass Signals and Systems
- Orthogonal Expansions of Signals
- Representation of Digitally Modulated Signals

## 6 Optimum Receivers for the AWGN Channel

- Characterization of Thermal Noise
- Signal Space Representation
- Optimum Detection

# Table of Contents

## 1 Introduction

- Mathematical Models for Communication Channels
- Generalized Functions

## 2 Probability

- Fundamentals in Probability Theory
- Useful Probability Distributions
- Central Limit Theorem

## 3 Fourier Transform

- Properties of FT
- Sampling Theorem

## 4 Stochastic Processes

- Gaussian Processes
- Response of an LTI System to a Stationary Input Signal
- Discrete-Time (Stochastic) Signals
- Cyclostationary Processes

## 5 Communication Signals and Systems

- Representation of Band-pass Signals and Systems
- Orthogonal Expansions of Signals
- Representation of Digitally Modulated Signals

## 6 Optimum Receivers for the AWGN Channel

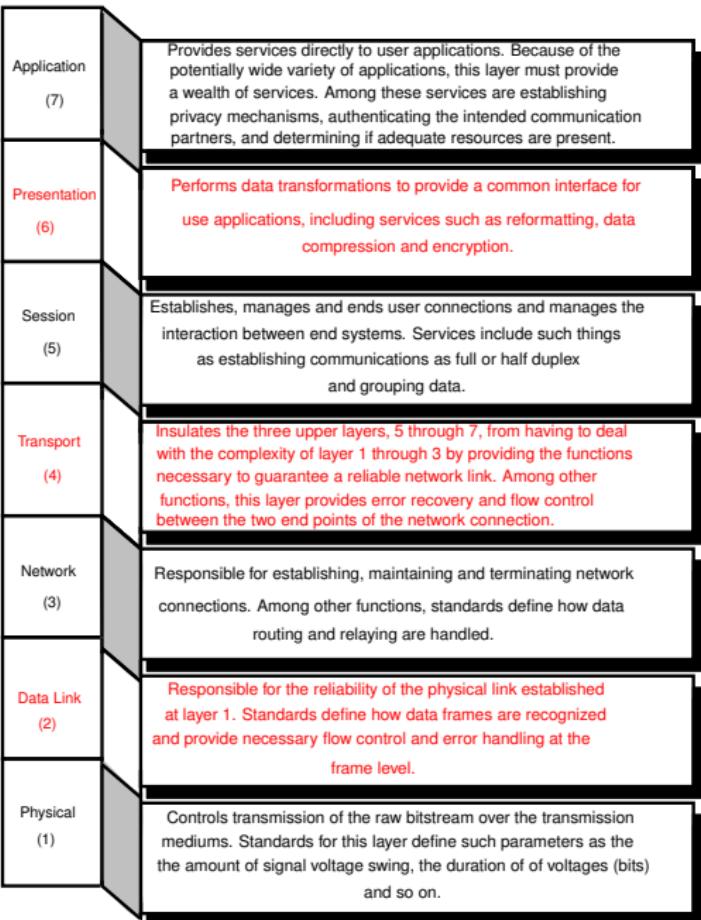
- Characterization of Thermal Noise
- Signal Space Representation
- Optimum Detection

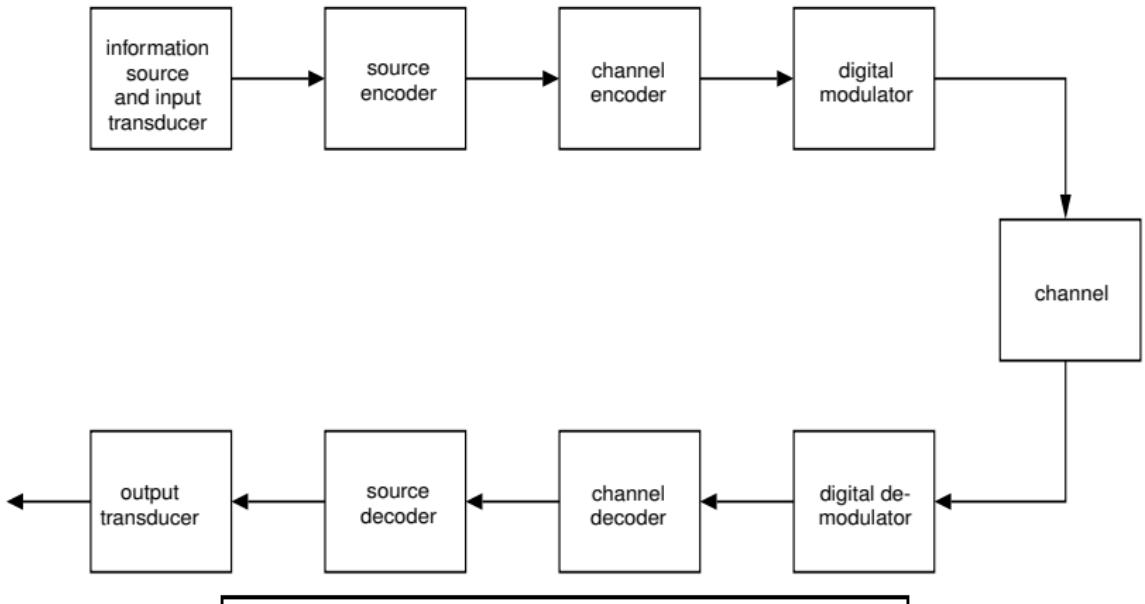
# Introduction

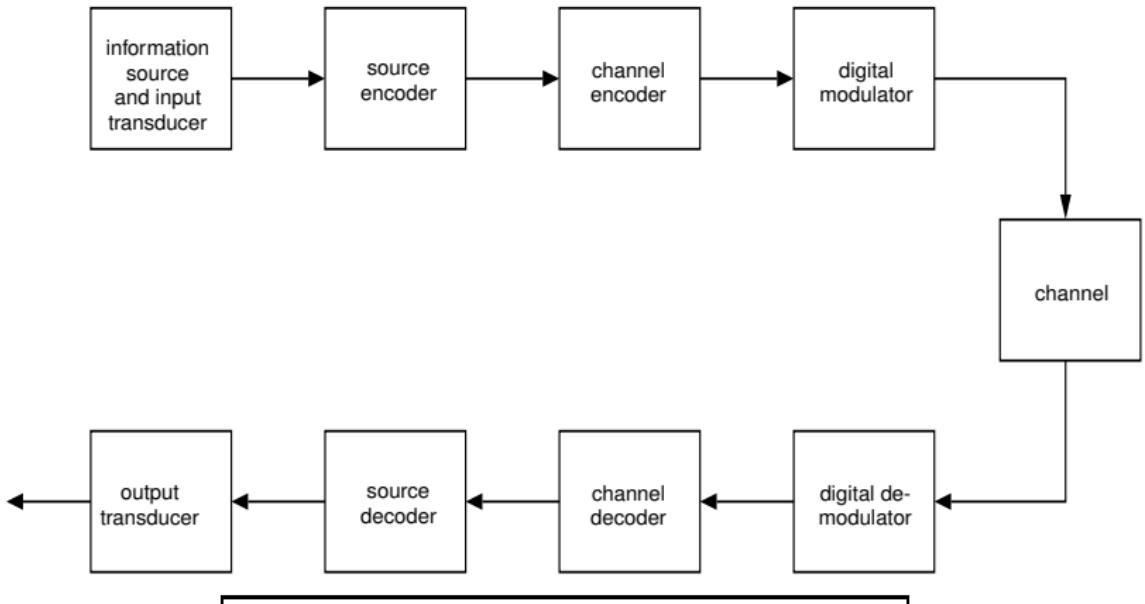
- What is *Communications* all about?
- Examples:
  - wireless programming of a cardiac pace maker
  - playing a compact disc (CD) or digital versatile disc (DVD)
  - reading bar codes (super market) and radio frequency identification (RFID) tags (ski lift etc.)
  - mobile communications in cellular radio systems
  - broadband communications (back bone for internet traffic, data base access, etc.)
  - satellite and deep space communications
- Are there any commonalities to these examples which allow them to be classified in a uniform way?

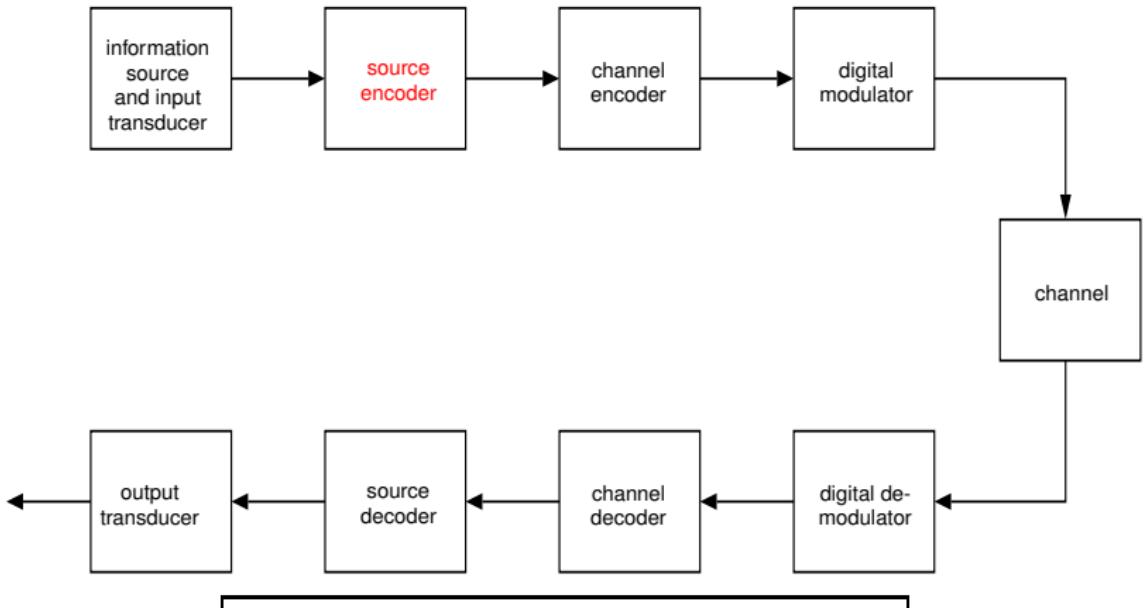
# Overview

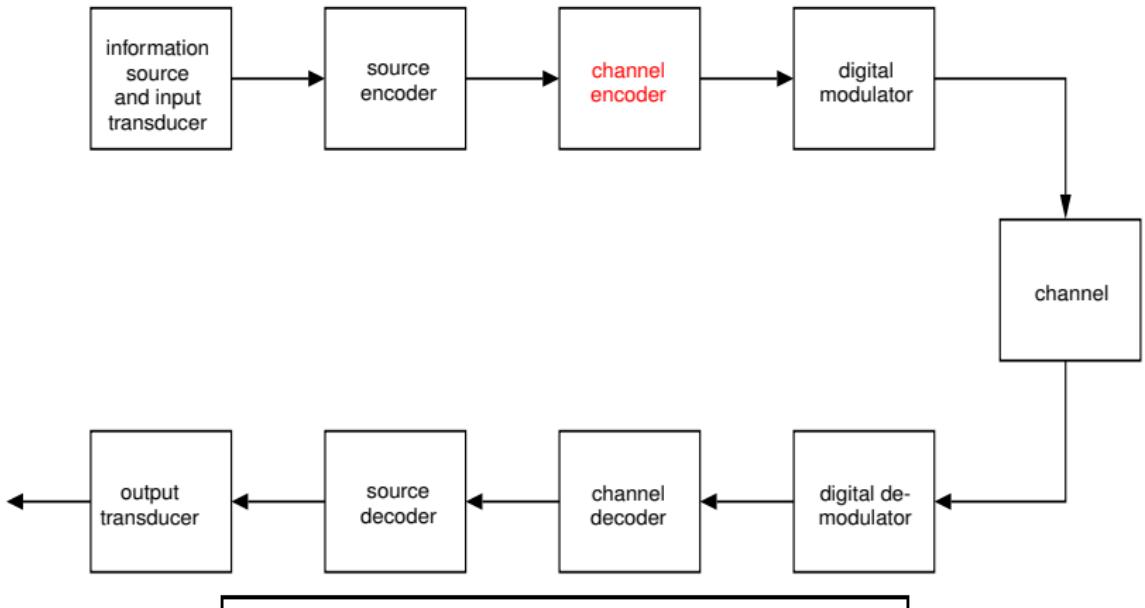
- Open Systems Interconnection (OSI) Reference Model for standardization of different approaches in communication systems
- in the OSI model, a layer offers a certain service to higher layers
- here, consider exclusively the so-called *physical layer* (PHY), i.e. the transmission from a certain information source or transmitter (TX) to an information drain or receiver (RX)

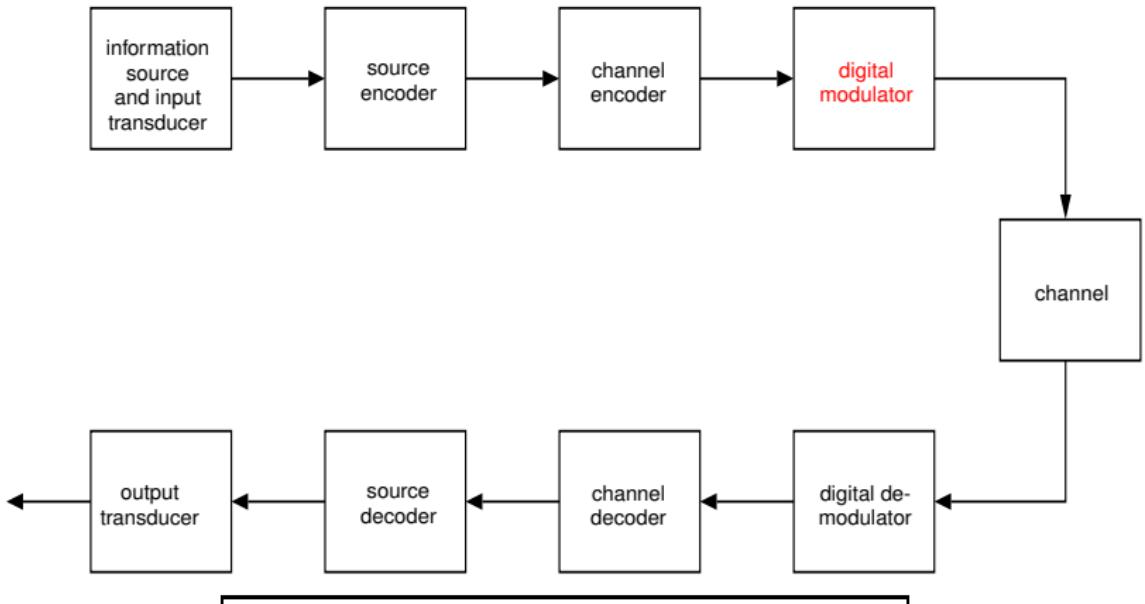


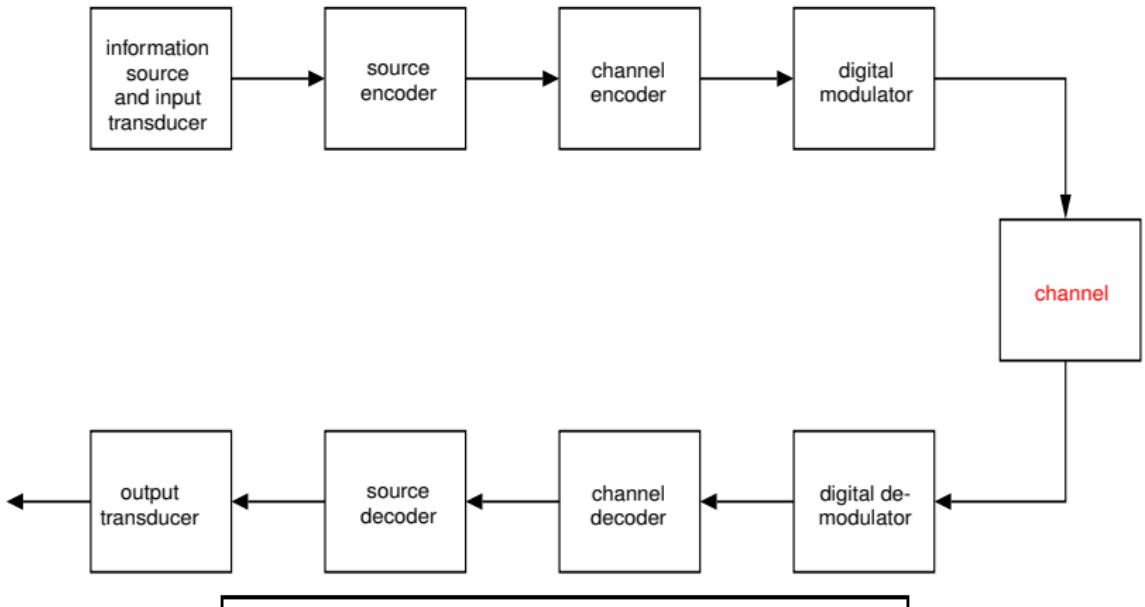


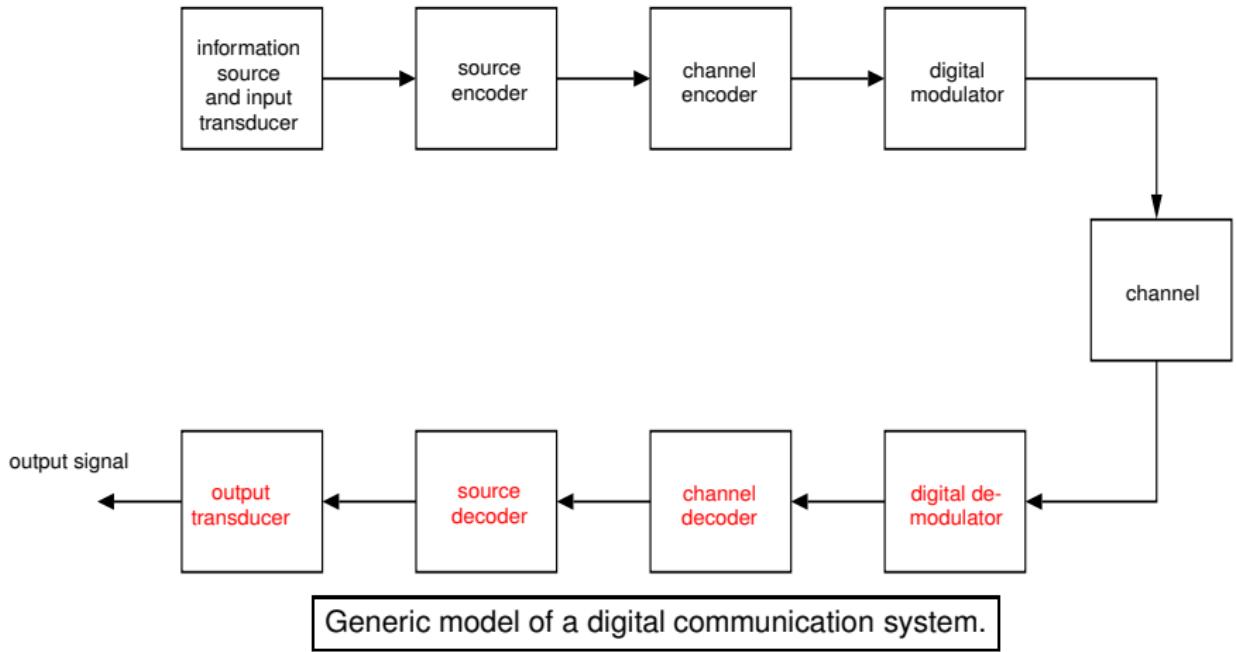


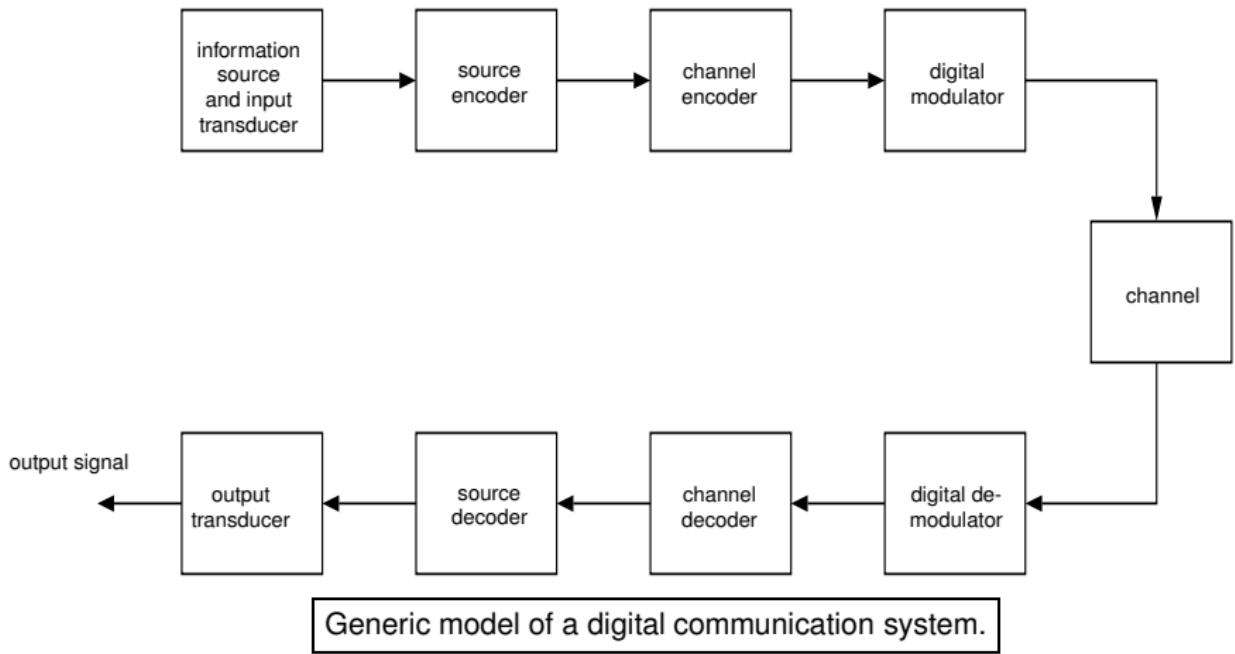




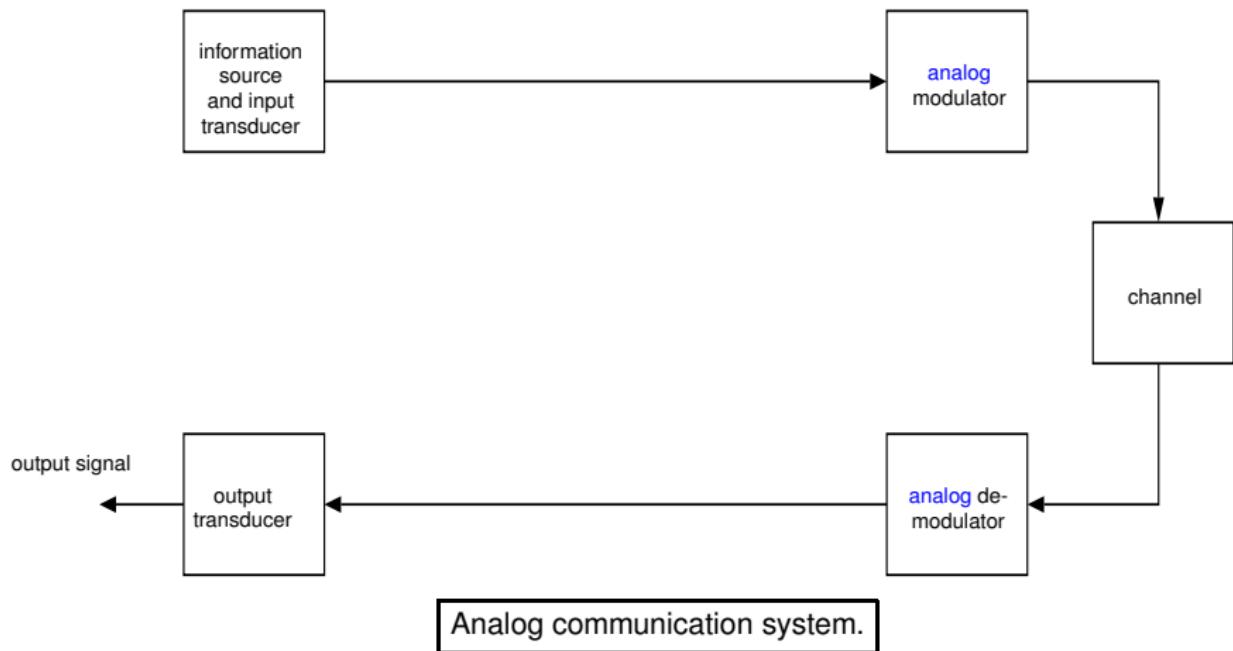






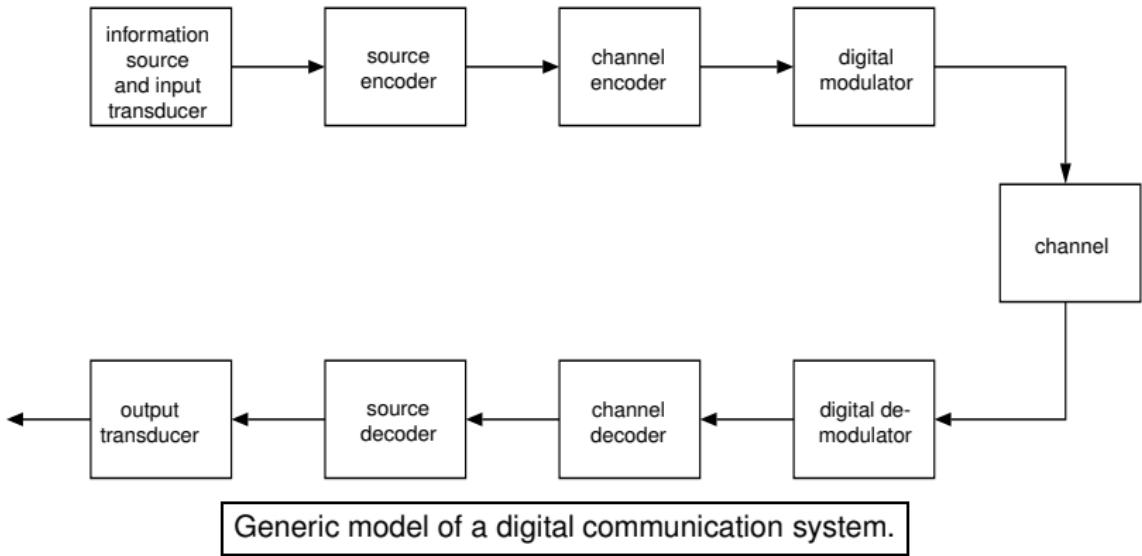


Are there alternatives to **digital** communications?



Are there alternatives to **digital** communications?

⇒ Yes, so-called **analog** communications using a simple, but possibly inefficient (missing source coding) and non-robust (missing channel coding) *transmitter* (TX) and *receiver* (RX) (= transceiver or TRX)



Obviously, the channel is the key element for the design of communication systems w.r.t.

- offered data rates
- performance (transmission quality)

There are two extremes:

- **perfect** channel (arbitrarily high data rates without errors at the decoder output)
- **useless** channel (no error-free message decoding is possible)

True transmission channels are neither perfect nor useless, but operate in between the two extreme cases

Thus, the following questions arise:

- Is there a **generic model** for describing transmission channels?
- What are the limitations of the model?
- Which tools are required for a proper description of the model?

Before answering the questions above, we consider a number of channels typically encountered in digital communication systems:

- ① wireline channels in telephone networks
- ② fibre-optic channels in broadband backbone networks
- ③ wireless channels in (mobile) radio systems
- ④ acoustic channels in underwater digital communication systems
- ⑤ storage channels in data storing systems

⇒ common feature of these channels: **linearity** of the transmission medium

Examples of relations of input signals  $s(t)$  and output signals  $r(t)$  of transmission media:

- **delay:** for an arbitrary non-negative real value  $t_0$

$$r(t) = s(t - t_0)$$

- **scaling:** for an arbitrary real value  $\alpha$

$$r(t) = \alpha s(t)$$

- **delay dispersion:** with  $\alpha_i \in \mathbb{R}$  and  $t_i \in \mathbb{R}^+$

$$r(t) = \sum_{i=1}^N \alpha_i s(t - t_i)$$

- **time-variant delay dispersion:** with  $\alpha_i(t) : \mathbb{R} \rightarrow \mathbb{R}$  and  $t_i(t) : \mathbb{R} \rightarrow \mathbb{R}^+$

$$r(t) = \sum_{i=1}^N \alpha_i(t) s(t - t_i(t))$$

⇒ common feature of these channels: **linearity** of the transmission medium

# Mathematical Models for Communication Channels

A **linear system** can be interpreted as a **linear mapping**  $\mathcal{L}$  from a vector space  $\mathcal{S}$  to a vector space  $\mathcal{R}$ . Here,  $\mathcal{S}$  and  $\mathcal{R}$  represent the vector spaces of (real-valued or complex-valued) input and output signals in  $\mathbb{R}$

A linear mapping is defined by the two relations:

$$\begin{aligned}\mathcal{L}(s_1(t) + s_2(t)) &= \mathcal{L}s_1(t) + \mathcal{L}s_2(t) \quad \forall \quad s_1(t), s_2(t) \in \mathcal{S}, \\ \mathcal{L}(\alpha s_1(t)) &= \alpha \mathcal{L}s_1(t) \quad \forall \quad s_1(t) \in \mathcal{S} \text{ and } \alpha \in \mathbb{C}\end{aligned}$$

How can  $\mathcal{L}$  be characterized in concrete mathematical terms?

To this end, we use the following identity (with  $\delta(t)$  denoting Dirac's delta function):

$$s(t) = \int_{\mathbb{R}} \delta(t - u)s(u)du$$

The output signal  $r(t)$  of the channel for an excitation with input signal  $s(t)$  results to (all subsequent integrals are assumed to exist)

$$\begin{aligned} r(t) &= \mathcal{L}s(t) \\ &= \mathcal{L} \int_{\mathbb{R}} \delta(t-u)s(u)du \\ &= \int_{\mathbb{R}} \mathcal{L}(\delta(t-u))s(u)du \\ \Rightarrow r(t) &= \int_{\mathbb{R}} k(t;u)s(u)du \end{aligned}$$

What is the meaning of the expression  $k(t; u)$ ?

- so-called **kernel** of the linear system
- Since  $k(t; u) = \mathcal{L}(\delta(t-u))$  it is obvious that  $k(t; u)$  represents the **system response to the input signal  $\delta(t-u)$**

Analogy in linear algebra with vectors

$$\begin{aligned}\mathbf{s} &= [s_1, \dots, s_M]^T \\ \mathbf{r} &= [r_1, \dots, r_N]^T\end{aligned}$$

and  $(N \times M)$ -dimensional matrix  $\mathbf{K} = \{K\}_{N,M}$  of suitable finite dimensions  $M$  and  $N$ :

$$r(t) = \int_{\mathbb{R}} k(t; u)s(u)du \quad \Leftrightarrow \quad \mathbf{r} = \mathbf{K}\mathbf{s}.$$

Observations:

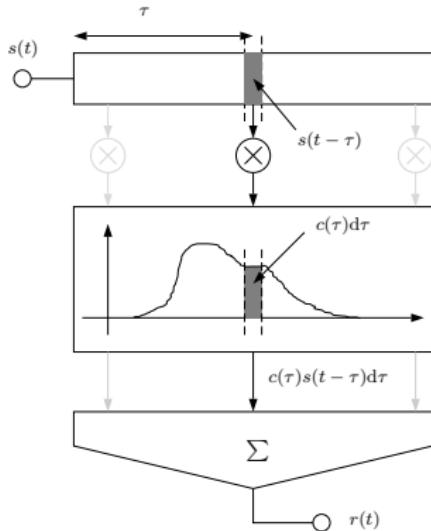
- Obviously, the linear operator (matrix)  $\mathbf{K}$  corresponds to  $k(t; u)du$
- The variable  $t$  corresponds to the *row* index of the matrix  $\mathbf{K}$ , the variable  $u$  to the column index
- The matrix-vector product  $\mathbf{K}\mathbf{s}$  (sum of products  $K_{t,u}s_u$ ) corresponds to  $\int_{\mathbb{R}} k(t; u)s(u)du$  (integral of products  $k(t; u)du$   $s(u)$ )

For **time-invariant** systems, we have  $k(t; u) = k(t - u)$  and the resulting integral represents a *convolution*:

$$r(t) = k(t) \star s(t) = \int_{\mathbb{R}} k(t - u)s(u)du = s(t) \star k(t)$$

Upon substitution  $\tau = t - u \Rightarrow$  interpret  $c(\tau) = k(\tau)$  as a transversal filter:

$$r(t) = \int_{\mathbb{R}} c(\tau)s(t - \tau)d\tau$$



Interpretation of convolutional integral as transversal filter.

Since  $r(t) = c(t - u)$  for  $s(t) = \delta(t - u) \Rightarrow c(t)$  represents the **channel impulse response**

Can we find an intuition for the meaning of the convolution integral?

**Step 1:** Consider the Fourier transformation  $\mathcal{F}$  of the time signal

$$s(t) \xrightarrow{\quad\bullet\quad} \mathcal{F}\{s(t)\} = S(f) = \int_{\mathbb{R}} s(t) \exp(-j2\pi ft) dt$$

and the Fourier inverse  $\mathcal{F}^{-1}$  of

$$S(f) \xrightarrow{\quad\bullet\quad} \mathcal{F}^{-1}\{S(f)\} = s(t) = \int_{\mathbb{R}} S(j\omega) \exp(j2\pi ft) df,$$

Here, we have considered the definition of FT as a function of the frequency  $f$  rather than the quantity  $j\omega$  (angular frequency variable) where all integrals are assumed to exist including possible Dirac delta functions.

An important example is the complex exponential time function and its Fourier transform defined by

$$\exp(j2\pi f_0 t) \xrightarrow{\quad\bullet\quad} \delta(f - f_0).$$

**Step 2:** Consider the Fourier transform of the convolutional integral given by

$$R(f) = \mathcal{F}\{r(t)\} = \mathcal{F}\{c(t) * s(t)\} = \mathcal{F}\{c(t)\} \mathcal{F}\{s(t)\} = C(f)S(f).$$

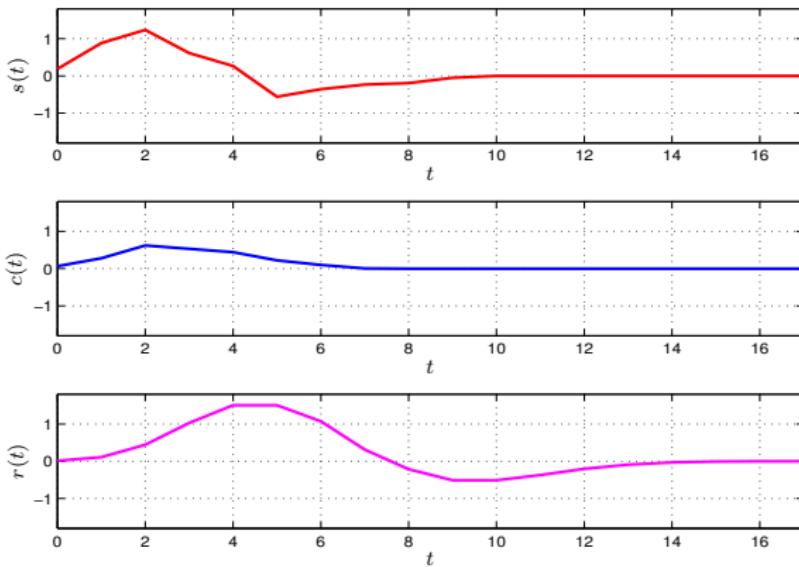
Obviously, the time signal  $\sigma(f_0; t) = \exp(j2\pi f_0 t)$  is an *eigenfunction* of the channel. Clearly, if  $\sigma(f_0; t)$  is the channel input signal, the output is proportional to  $\sigma(f_0; t)$  as follows from

$$R(f)|_{s(t)=\sigma(f_0; t)} = C(f)\delta(f - f_0) \bullet\!\!\!-\!\!\!\circ r(t) = C(f_0) \exp(j2\pi f_0 t).$$

Here,  $C(f_0)$  is the *eigenvalue* of the linear time-invariant (LTI) channel for the eigenfunction  $\sigma(f_0; t)$  and is known as the *transfer function* of the channel

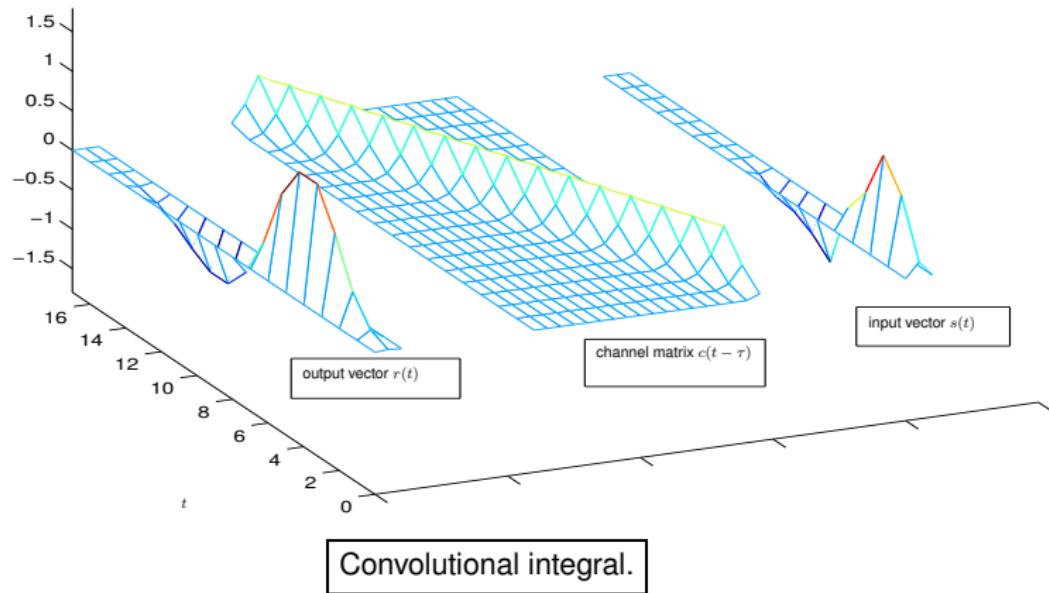
Nota Bene (N.B.): considerations of eigenfunctions and eigenvalues are completely analogous to the corresponding notions in linear algebra, i.e. eigenvectors and eigenvalues

Example: convolution of signals with finite temporal support, usual representation



Convolutional integral.

## Example: convolution of signals, vector-matrix representation



**Step 3:** Finally, since we can interpret

$$s(t) = \int_{\mathbb{R}} S(f) \exp(j2\pi ft) df$$

as the linear superposition of complex exponentials  $\exp(j2\pi ft)$  weighted by complex numbers  $S(f)df$ , the convolution can be interpreted as follows:

For arbitrary  $c(t)$ , each frequency component  $S(f)$  of  $s(t)$  is weighted by a complex number  $C(f)$  and the resulting components  $R(f) = C(f)S(f)$  are linearly superimposed to give the output signal  $r(t) = \int_{\mathbb{R}} R(f) \exp(j2\pi ft) df$ .

This property makes the Fourier transform a powerful tool in the characterization of LTI systems

For **time-variant** channels, the time-invariant channel impulse response  $c(t)$  is not sufficient for the channel description. Instead, upon substituting  $\tau = t - u$ , we obtain in analogy with the time-invariant case

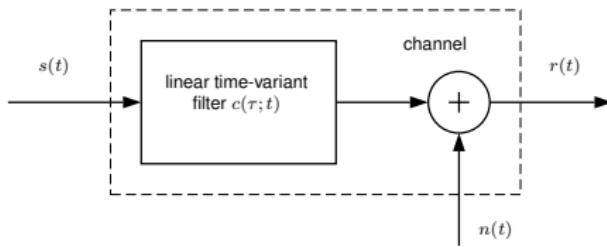
$$\begin{aligned} r(t) &= \int_{\mathbb{R}} k(t; u)s(u)du \\ &= \int_{\mathbb{R}} k(t; t - \tau)s(t - \tau)d\tau \\ &= \int_{\mathbb{R}} c(\tau; t)s(t - \tau)d\tau \end{aligned}$$

with  $c(\tau; t) = k(t; t - \tau)$  denoting the **time-variant channel impulse response** at time  $t$  to a Dirac pulse applied at time  $t - \tau$ . The last integral expression is a direct extension of the usual convolution integral (arising in the time-invariant case) to a time-variant system behaviour

Some remarks on the aforementioned linear channel models:

- The channel impulse responses (CIRs) above (either time-variant or time-invariant) are assumed to represent a **real-valued or complex-valued function** defined on  $\mathbb{R}$  (in the case of  $c(\tau)$ ) or  $\mathbb{R}^2$  (in the case of  $c(\tau; t)$ )
- If the received signal was indeed given by  $r(t) = k(t) \star s(t)$ , the transceiver design would be rather simple and close to the case of the perfect channel  $k(t) = \delta(t)$  (except for spectral zeros, we could, for example, equalize the channel in the frequency domain by multiplying  $R(f)$  by  $1/C(f)$  and taking the inverse Fourier transform afterwards)
- In most cases, however, we do not know about the exact form of the CIR, but still have to design the transceiver such that its performance complies with the specifications for the system at hand. The CIR is subject to *random* effects such as fading, scattering from objects and Doppler effects in wireless systems, random disturbances in front-ends, drift effects etc., so that, in general, the CIR is rather to be characterized as a **stochastic process**

- Apart from the transmission channel itself, **the receiver is a source of noise** which limits the information rate and performance and thus has to be taken into account in the transceiver design. It is one of the major contributions of information theory (initiated by the work of Shannon in the 1940s) to design schemes to overcome the impact of random disturbances in a digital communication system
- A simple standard approach encountered in practical systems to model noise in receivers is **additive noise**



Linear time-variant channel with additive noise model.

- Again, as in the case of **random effects in the transmission medium**, the noise  $n(t)$  represents a random process and the received signals in a practical system are instances (sometimes also termed realizations) of the corresponding processes
- Summarizing the comments above, most channels that we will deal with lead to a model of the received signal defined by

$$r(t) = \int_{\mathbb{R}} c(\tau; t) s(t - \tau) d\tau + n(t)$$

with  $c(\tau; t)$  and  $n(t)$  representing **stochastic processes** with properties which are to be chosen for the system at hand

- Other sources of disturbances include interference from other communication links (in particular in wireless transmission), cross-talk (telephone lines), amplifier saturation effects leading to non-linear channel properties, intermodulation, synchronization errors etc. The aforementioned model represents a reasonable **trade-off between simplicity and accuracy** and can often be extended to include further channel properties which are critical for the overall transceiver design
- Thus, in the following, we will consider concepts from **probability** and **stochastic processes** in order to characterize the properties of the received signals, to get familiar with standard probability distributions and to establish basic approaches for characterizing the performance including bounding techniques

# Generalized functions

- An important class of the linear time-invariant systems can be represented in the physical-technical relation using linear ordinary **differential equations** (ODE) with constant coefficients. The output signal for a specific excitation at the input can be determined through finding the solution of the corresponding equation. The following introduction to the important mathematical basics are taken from the book of T. Frey, M. Bossert, *Signal- und Systemtheorie*, 2. edition, Vieweg und Teubner, ISBN 978-3-8351-0249-1, 2008
- When considering the step type signals, there is a problem here since such signals are not differentiable at some points (e.g. ramp function  $tu(t)$  at  $t = 0$ ) or discontinuous (unit step function  $u(t)$  at  $t = 0$ )
- This problem can be solved based on extending the function notion, where we introduce the generalized functions or distributions.

### Introduction to generalized functions

- The starting point in this introduction to generalized functions is the unit step function

$$u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0, \end{cases}$$

which shows a discontinuity at  $t = 0$  and therefore not differentiable.  
Nevertheless, we will try to assign this function some kind of derivative.

- To that end, we consider the set of functions

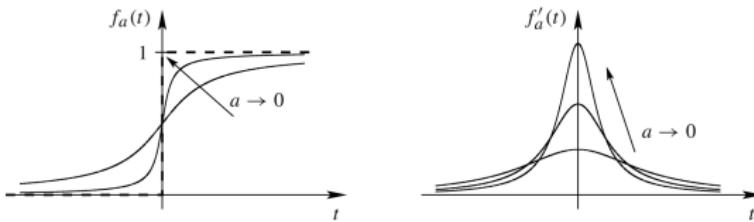
$$f_a(t) = \frac{1}{\pi} \left[ \arctan \left( \frac{t}{a} \right) + \frac{\pi}{2} \right], \quad a > 0$$

which has the limit of one for increasing  $t$  in the positive direction and zero in the negative direction

## Introduction to generalized functions

- When  $a$  becomes smaller, the transition becomes steeper, where at the limiting case  $a \rightarrow 0$  the unit step function reads:

$$u(t) = \lim_{a \rightarrow 0} \frac{1}{\pi} \left[ \arctan \left( \frac{t}{a} \right) + \frac{\pi}{2} \right].$$



The unit step function and dirac impulse function as a limit.

- We have then represented the discontinuous function as a limit of continuous and differentiable function, which in turn can be differentiated. This allows writing the derivative of the unit step function as:

$$\frac{d}{dt} u(t) = \frac{d}{dt} \lim_{a \rightarrow 0} \frac{1}{\pi} \left[ \arctan \left( \frac{t}{a} \right) + \frac{\pi}{2} \right] = \lim_{a \rightarrow 0} \frac{1}{\pi} \cdot \frac{a}{a^2 + t^2} = \lim_{a \rightarrow 0} f'_a(t).$$

Where we have assumed that the limiting operator and the differentiation can be interchanged

## Introduction to generalized functions

- Based on the figure we can see that the last equation represents an impulse, which becomes narrower and higher for the decreasing  $a$ . At the limiting case  $a \rightarrow 0$  this impulse vanishes for  $t \neq 0$  while for  $t = 0$  it diverges:

$$\lim_{a \rightarrow 0} f'_a(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}$$

- The integration of this impulse is still constant:

$$\int_{-\infty}^{\infty} f'_a(t) dt = 1$$

- With this procedure (i.e. the representation of the step type signals as a limit parameterized functions), we can in principle solve the differential equations of non-differentiable signals. However, this procedure is a bit inconvenient.

## Introduction to generalized functions

- Therefore, we will first assign the impulse  $\lim_{a \rightarrow 0} f'_a(t)$  a symbolic notation. We denote it as continuous **Dirac Impulse  $\delta(t)$** . Later on, we will notice that this impulse represents the continuous correspondence of the time-discrete dirac impulse  $\delta[k]$ . Clearly, we can interpret the dirac impulse as a limit:

$$\delta(t) = \lim_{a \rightarrow 0} f'_a(t) = \lim_{a \rightarrow 0} \frac{1}{\pi} \cdot \frac{a}{a^2 + t^2}$$

- However, we have thereby limited ourselves, since the limit of other functions can also be utilized in this representation. For example, the following representation

$$\delta(t) = \lim_{f \rightarrow \infty} \frac{\sin(2\pi ft)}{\pi t}$$

can be likewise chosen, which has an application in the next chapter in relation to Fourier transformation.

### Introduction to generalized functions

- A general representation can be obtained when characterizing the impulse by its **impact**. In this case, with the help of a test function  $\phi(t)$  the equations can be summarized to the following implicit definition:

$$\int_{-\infty}^{\infty} \delta(t) \cdot \phi(t) dt = \phi(0), \quad \forall \phi(t) \in C_0^\infty(\mathbb{R}^n)$$

$C_0^\infty(\mathbb{R}^n)$  represents the space of the infinitely often differentiable and beyond a finite interval vanishing functions, so that the test functions are arbitrarily easy-going. Clearly, the relation between the last equation and the properties of  $f'_a(t)$  can be justified as follows: The multiplication of the test function  $\phi(t)$  by the dirac impulse  $\delta(t)$  sifts all the values to  $\phi(0)$ . Shifting this value before the integral and considering  $\int_{-\infty}^{\infty} f'_a(t) dt = 1$ , we get the equation above.

## Introduction to generalized functions

- We have characterized *indirectly* the dirac impulse by the relation before, namely by its *impact on the test functions*. In this case, we do not consider the (conventional) functions, which are directly characterized as mapping of the real numbers into a mapping domain, but rather the **generalized functions** or **distributions**.
- Distributions represent an extension of the function notion and include therefore also all conventional functions. In the general form, the definition of a distribution  $\psi(t)$  reads

$$\int_{-\infty}^{\infty} \psi(t) \cdot \phi(t) dt = \langle \psi(t), \phi(t) \rangle = F[\phi(t)], \quad \forall \phi(t) \in C_0^\infty(\mathbb{R}^n)$$

$F[\phi(t)]$  represents a linear functional relation, which assigns every test function  $\phi(t)$  a real or complex number.

**Example: characterizing the unit step function as distribution**

- The unit step function  $u(t)$  is characterized by the linear functional

$$F[\phi(t)] = \int_0^\infty \phi(t) dt$$

as distribution  $\psi(t)$  to

$$\int_{-\infty}^\infty \psi(t) \cdot \phi(t) dt = \int_0^\infty \phi(t) dt.$$

- In the following, we see how powerful the concept of the distribution is, where it is possible to assign every distribution a derivative, also when they are as conventional functions not differentiable. The calculation rule is obtained if we formally call the derivative of distribution  $\psi'(t)$  and integrate by parts:

$$\begin{aligned} \langle \psi'(t), \phi(t) \rangle &= \int_{-\infty}^\infty \psi'(t) \cdot \phi(t) dt = \underbrace{[\psi(t) \cdot \phi(t)]_{-\infty}^\infty}_{=0} - \int_{-\infty}^\infty \psi(t) \cdot \phi'(t) dt \\ &= -\langle \psi(t), \phi'(t) \rangle \end{aligned}$$

### Example: characterizing the unit step function as distribution

- At that, we have used the assumption, that the test function vanishes beyond a finite interval, i.e. that  $\phi(\pm\infty) = 0$  holds
- With the help of this rule we can also reduce the difficult problem of differentiating generalized functions to the finding the derivative of conventional differentiable functions. Based on the assumption, that the test functions are arbitrarily often differentiable, this implies that every distribution is arbitrarily often differentiable.

### Example: the Derivative of the unit step function

- The derivative of the distribution  $\psi(t) = u(t)$  (unit step function) with

$$\langle \psi(t), \phi(t) \rangle = \int_0^\infty \phi(t) dt$$

yields based on the last example

$$\langle \psi'(t), \phi(t) \rangle = -\langle \psi(t), \phi'(t) \rangle = - \int_0^\infty \phi'(t) dt = -[\phi(t)]_0^\infty = \phi(0).$$

This corresponds to the dirac impulse, i.e.  $u'(t) = \delta(t)$  (see the equality of the distributions below)

- The explanations in this section have shown, how introducing of the distributions has solved the problem of finding the derivative of the step type functions in a mathematically correct way. The dirac impulse plays in that a major role, which represents the derivative of the unit step function. We can understand the dirac impulse in the following as a symbolic function and therefore handle it as in conventional functions, because in most cases the calculation rules of the distributions coincide with those of the functions.

## Special continuous signals

### 1. Dirac impulse

- An important signal in the theory of continuous systems is the dirac impulse:

The dirac impulse  $\delta(t)$  is characterized implicitly as distribution by

$$\int_{-\infty}^{\infty} \delta(t) \cdot \phi(t) dt = \phi(0),$$

where  $\phi(t)$  must be continuous and arbitrarily often differentiable

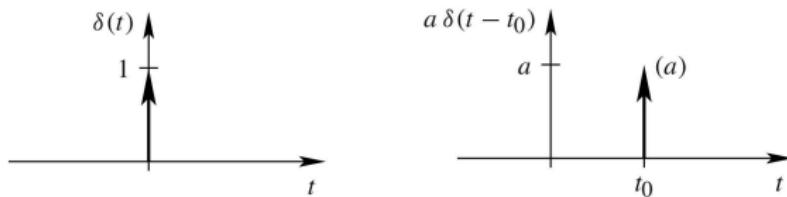
- It is the continuous correspondence to the time discrete dirac impulse and hence shows the corresponding properties as well as the same great system-theoretical meaning
- The dirac impulse owns the following elementary properties:

$$\delta(t) = 0, \quad t \neq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(t) dt = 1$$

## Special continuous signals

- The dirac impulse can be figuratively represented as an arrow with length corresponding to the scaling. The latter refers also to the **weight** of the dirac impulse and is written when necessary in parentheses next to the pulse to be uniquely identified.
- In analogy to the discrete, the dirac impulse owns the **sampling or sifting property**, which directly results from the last both equations:

$$x(t) \cdot \delta(t - t_0) = x(t_0) \cdot \delta(t - t_0) \quad \text{resp.} \quad \int_{-\infty}^{\infty} x(t) \cdot \delta(t - t_0) dt = x(t_0).$$



Figurative representation of the dirac impulse.

## Special continuous signals

- The signal  $x(t)$  must be continuous at the point  $t = t_0$ . In particular, it holds for  $x(t) = t$ , that  $t \cdot \delta(t) = 0$
- The dirac impulse represents the identity element of the (continuous) convolution, which has a central role in relation to linear systems. This implies:

$$x(t) = x(t) * \delta(t) = \int_{-\infty}^{\infty} x(\tau) \cdot \delta(t - \tau) d\tau.$$

- In the calculation of the time scaled dirac impulse, the following rule is considered:

$$\delta(at) = \frac{1}{|a|} \cdot \delta(t) \quad \text{resp.} \quad \int_{-\infty}^{\infty} \delta(at) dt = \frac{1}{|a|}$$

- The relations result directly from evaluating the integral

$$\int_{-\infty}^{\infty} \delta(at) dt = \int_{-\infty}^{\infty} \delta(\tau) \frac{1}{|a|} d\tau = \frac{1}{|a|},$$

where the absolute value comes from interchanging the integral limits for the negative values of  $a$ . As a special case, it leads to the relation  $\delta(t) = \delta(-t)$ , from which it follows, that the dirac impulse represents a even distribution.

### Special continuous signals

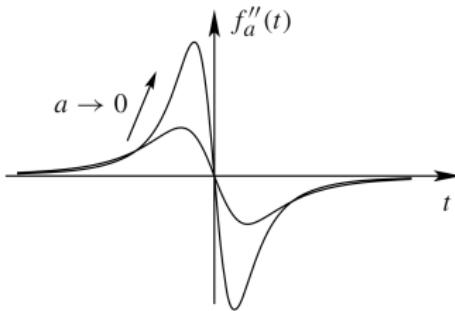
- Since every distribution is infinitely often differentiable, this means obviously that we can differentiate the dirac impulse. When applying the rule  $\langle \psi'(t), \phi(t) \rangle = -\langle \psi(t), \phi'(t) \rangle$ , we get

$$\langle \psi'(t), \phi(t) \rangle = -\langle \psi(t), \phi'(t) \rangle = \phi'(0)$$

which is however beyond any intuition. In contrast, the representation as a limit provides an interpretation

$$\delta'(t) = \lim_{a \rightarrow 0} f_a''(t) = \lim_{a \rightarrow 0} -\frac{1}{\pi} \cdot \frac{2at}{(a^2 + t^2)^2}$$

### Special continuous signals



Representing the derivative of dirac impulse as a limit.

- Based on the figure, we can recognize that it is about a point-symmetric double impulse. The area below each of the two individual impulses goes to infinity, so that they are not dirac impulse.
- Finally, it can be noticed that the derivative of the dirac impulse, like all other derivatives of it, exist and hence we can expect that these distributions have no practical meaning as signals.

## Special continuous signals

## 2. Unit step function

- We have previously characterized the unit step function as a conventional function, also it is represented in the example as a distribution, from which the following relation for dirac impulse results:

$$u(t) = \int_{-\infty}^t \delta(\tau) d\tau \quad \text{resp.} \quad \delta(t) = \frac{du(t)}{dt} = u'(t).$$

- The derivative of the time scaled unit step function can be obtained from the chain rule of the differential calculus as:

$$\frac{d}{dt} u(at) = u'(at) \cdot \frac{d}{dt}[at] = \delta(at) \cdot a = \frac{1}{|a|} \delta(t) \cdot a = \text{sign}(a) \cdot \delta(t).$$

- As a special case for  $a = -1$ , the derivative of the time mirrored unit step function

$$\frac{d}{dt} u(-t) = -\delta(t).$$

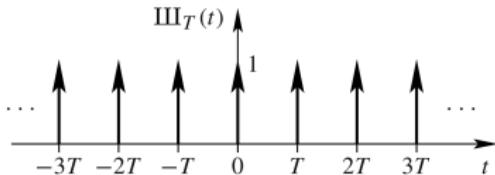
- It is observed that the unit step function is undefined at the point  $t = 0$  according to the definition of the distribution. However, we have chosen the value to be 0.5, which is consistent to the representation of the distribution as a limit by the equation  $u(t) = \lim_{a \rightarrow 0} \frac{1}{\pi} [\arctan(\frac{t}{a}) + \frac{\pi}{2}]$

## Special continuous signals

### 3. Impulse comb

- For representing the periodic signals as well as the relation between the time-discrete and the continuous systems, the following signal is important:  
**Impulse comb:**

$$\Xi_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) \quad \text{resp.} \quad \Xi(t) = \sum_{k=-\infty}^{\infty} \delta(t - k)$$



Figurative representation of the impulse comb.

### Special continuous signals

- Due to the property  $\delta(at) = \frac{1}{|a|} \cdot \delta(t)$ , the following relation between the two representations holds

$$\Xi_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) = \sum_{k=-\infty}^{\infty} \frac{1}{|T|} \delta\left(\frac{t}{T} - k\right) = \frac{1}{|T|} \Xi\left(\frac{t}{T}\right).$$

- Other names for the impulse comb are dirac impulse comb, dirac impulse train oder Scha function. The figurative representation of the impulse comb comes according to the dirac representation again as arrows with the corresponding length.

## The derivative at jump and bend discontinuities

- As an motivation to the introduction to the generalized functions, we have mentioned at the beginning of this section the derivative of the step type signals. Finding the derivative is mathematically possible using the formal representation as a distribution. However, a simple and a clear solution is possible through decomposing such a non-differentiable function in the conventional way into a differentiable part and the residual part is described by the elementary distributions
- A function  $f(t)$  with a jump discontinuity at  $t = t_0$ , which has the left resp. the right side limit

$$a = \lim_{t \rightarrow t_0^-} f(t) \quad \text{resp.} \quad b = \lim_{t \rightarrow t_0^+} f(t)$$

can be decomposed into a continuous part  $f_s(t)$  and a correspondingly scaled and delayed unit step function:

$$f(t) = f_s(t) + (b - a) \varepsilon(t - t_0)$$

### The derivative at jump and bend discontinuities

- As a derivative of the overall function we can write:

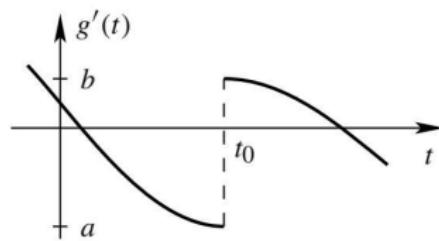
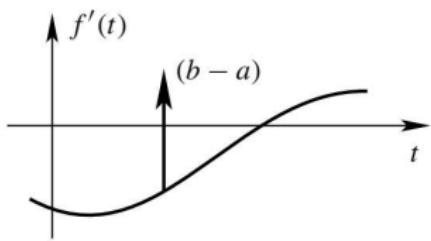
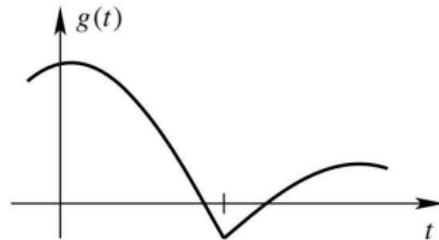
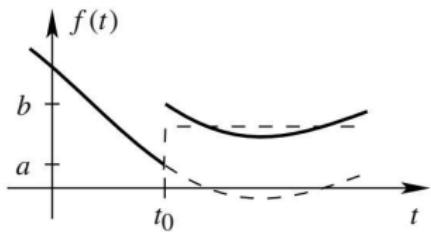
$$f'(t) = f'_s(t) + (b - a) \delta(t - t_0),$$

i.e. at the jump discontinuity point, a dirac impulse appears in the derivative, whose weight corresponds to the height of the jump. Here, we have at first assumed that  $f_s(t)$  is differentiable at  $t = t_0$ , i.e. shows no bend discontinuities.

- The derivative at the bend discontinuity point brings however no difficulty. When the function  $g(t)$  has a bend discontinuity at the point  $t = t_0$ , then the limits of the left and right side of the derivative will be different at this point:

$$a = \lim_{t \rightarrow t_0^-} g'(t) \quad \neq \quad b = \lim_{t \rightarrow t_0^+} g'(t)$$

### The derivative at jump and bend discontinuities



The derivative of the functions with jump and bend discontinuities.

### The derivative at jump and bend discontinuities

- The derivative at the bend discontinuity point shows a step, whose height corresponds to the difference between the limits of the right and the left side derivative. Formally, one could again decompose into a conventionally differentiable part and a ramp function with a slope of  $(b - a)$ , which leads to a height of the step  $(b - a)$
- It is observed, that a step usually appears in causal signals at  $t = 0$ , which yields a dirac impulse in the derivative. Formally, we obtain this result based on representing the signal as a product with the step function and applying the product rule of differential calculus:

$$\begin{aligned}\frac{d}{dt}[x(t) \cdot u(t)] &= \frac{d}{dt}[x(t)] \cdot u(t) + x(t) \cdot \frac{d}{dt}[u(t)] \\ &= x'(t) \cdot u(t) + x(t) \cdot \delta(t) = x'(t) \cdot u(t) + x(0) \cdot \delta(t).\end{aligned}$$

## Distributions

- A distribution  $\psi(t)$  is implicitly characterized by the following relation:

$$\int_{-\infty}^{\infty} \psi(t) \cdot \phi(t) dt = \langle \psi(t), \phi(t) \rangle = F[\phi(t)], \quad \forall \phi(t) \in C_0^\infty(\mathbb{R}^n)$$

- $\phi(t)$  is called the test function and  $C_0^\infty(\mathbb{R}^n)$  originates all functions  $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ , which are
  - continuous
  - arbitrarily often differentiable
  - vanishes beyond a finite interval, i.e.  $\phi(t) = 0 \quad \forall |t| > t_0$

## Distributions

- $F[\phi(t)]$  represents a continuous linear functional, i.e. a mapping  $F : C_0^\infty(\mathbb{R}^n) \rightarrow \mathbb{R}$  with the following properties

- linearity:  $F[a_1\phi_1(t) + a_2\phi_2(t)] = a_1F[\phi_1(t)] + a_2F[\phi_2(t)]$
- Continuity:  $\lim_{n \rightarrow \infty} \phi_n(t) = \phi(t) \Rightarrow \lim_{n \rightarrow \infty} F[\phi_n(t)] = F[\phi(t)],$

where  $\phi_n(t)$  represents a sequence of the functions, that converges point-wise (for all  $t$ ) to the function  $\phi(t)$ . This definition is chosen to match the well-known continuity definition of the functions, which states that a function  $f(x)$  is continuous at  $x$ , if the following holds for all the converging sequence at this point  $x_n$ :

$$\lim_{n \rightarrow \infty} x_n = x \Rightarrow \lim_{n \rightarrow \infty} f(x_n) = f(x)$$

- Due to the assumed linearity of the functional  $F[\phi(t)]$ , the definition of the distribution itself is linear, i.e. it applies:

$$\langle \psi(t), a_1\phi_1(t) + a_2\phi_2(t) \rangle = a_1 \langle \psi(t), \phi_1(t) \rangle + a_2 \langle \psi(t), \phi_2(t) \rangle$$

## Distributions

- Since a distribution is implicitly characterized by its impact on a test function, we can not in general directly represent it. Representing  $f(t)$  as a conventional function applies:

$$\int_{-\infty}^{\infty} \psi(t) \cdot \phi(t) dt = \int_a^b f(t) \cdot \phi(t) dt, \quad \forall \phi(t),$$

where  $\psi(t) = f(t)$  in the interval  $[a; b]$ , and the distribution therefore can be represented as a function. Moreover, we can recognize that the distribution notion encloses the function notion.

## Equality of distributions

- Because of characterizing the distribution by its impact on a test function, the pointwise comparison between two distributions is not possible as in functions. That is why we characterize again the equality through the test functions:

$$\psi_1(t) = \psi_2(t) \Leftrightarrow \langle \psi_1(t), \phi(t) \rangle = \langle \psi_2(t), \phi(t) \rangle, \quad \forall \phi(t)$$

### Even and odd distributions

- A distribution  $\psi(t)$  is even (odd), if the following holds for every odd (even) test functions  $\phi(t)$ :

$$\int_{-\infty}^{\infty} \psi(t) \cdot \phi(t) dt = 0.$$

### Calculation rule

- The summarized calculation rules in the next table hold for distributions. They result directly from characterization by the integral with the help of elementary integration rules
- For example, the time shifted distribution reads

$$\begin{aligned}\langle \psi(t - t_0), \phi(t) \rangle &= \int_{-\infty}^{\infty} \psi(t - t_0) \cdot \phi(t) dt \\ &= \int_{-\infty}^{\infty} \psi(t') \cdot \phi(t' + t_0) dt' = \langle \psi(t), \phi(t + t_0) \rangle\end{aligned}$$

## Calculation rules for distributions

### Operations with distributions:

$$\text{time shift} \quad \langle \psi(t - t_0), \phi(t) \rangle = \langle \psi(t), \phi(t + t_0) \rangle$$

$$\text{scaling} \quad \langle \psi(at), \phi(t) \rangle = \frac{1}{|a|} \langle \psi(t), \phi\left(\frac{t}{a}\right) \rangle$$

$$\text{differentiation} \quad \langle \psi^{(k)}(t), \phi(t) \rangle = (-1)^k \langle \psi(t), \phi^{(k)}(t) \rangle$$

### Joining distributions and functions:

$$\text{addition} \quad \langle \psi_1(t) + \psi_2(t), \phi(t) \rangle = \langle \psi_1(t), \phi(t) \rangle + \langle \psi_2(t), \phi(t) \rangle$$

$$\text{product with function} \quad \langle g(t) \cdot \psi(t), \phi(t) \rangle = \langle \psi(t), g(t) \cdot \phi(t) \rangle$$

$$\text{convolution} \quad \langle \psi_1(t) * \psi_2(t), \phi(t) \rangle = \langle \psi_1(t), \langle \psi_2(\tau), \phi(\tau + t) \rangle \rangle$$

## Generalized derivatives

- Since every distribution is infinitely often differentiable, it can help in differentiating functions, which are not differentiable in the classical sense, because they show jump and bend discontinuities. By using the formal representation and the derivative of the function as a distribution, we however lose every intuition. For this reason, we decompose the function into a differentiable part and a residual part, which contains bend and jump discontinuities and can be represented by the elementary distributions. The differentiation is carried out separately for both parts, which is allowed because of the linearity of the distributions
- For representing the residual part, the ramp function is used for the bend discontinuity and the step function for the jump discontinuity. When representing them as distributions, we obtain:

$$\text{Ramp function } \rho(t): \langle \rho(t), \phi(t) \rangle = \int_0^\infty t \cdot \phi(t) dt$$

$$\text{Step function } u(t): \langle u(t), \phi(t) \rangle = \int_0^\infty \phi(t) dt = \langle \rho'(t), \phi(t) \rangle$$

$$\text{Dirac impulse } \delta(t): \langle \delta(t), \phi(t) \rangle = \phi(0) = \langle u'(t), \phi(t) \rangle$$

and therefore:  $\frac{d}{dt}\rho(t) = \rho'(t) = u(t)$  and  $\frac{d}{dt}u(t) = u'(t) = \delta(t)$

### Generalized limits

- Specific limits, which do not exist in the classical sense, can be represented with the help of distributions

#### Limit of the complex exponential function

- The limit  $\lim_{\omega \rightarrow \infty} e^{j\omega t}$  does not exist in the classical sense. The representation and the evaluation in the sense of the distribution leads to

$$\left\langle \lim_{\omega \rightarrow \infty} e^{j\omega t}, \phi(t) \right\rangle = \int_{-\infty}^{\infty} \left( \lim_{\omega \rightarrow \infty} e^{j\omega t} \right) \cdot \phi(t) dt,$$

where the integral exists because of the limit of the exponential function and the properties of the test function

## Limit of the complex exponential function

- From that with the help of the integration by parts we get

$$\lim_{\omega \rightarrow \infty} \int_{-\infty}^{\infty} \underbrace{e^{j\omega t}}_{u'} \cdot \underbrace{\phi(t)}_v dt = \lim_{\omega \rightarrow \infty} \left( \left[ \frac{1}{j\omega} e^{j\omega t} \phi(t) \right]_{-\infty}^{\infty} - \frac{1}{j\omega} \int_{-\infty}^{\infty} e^{j\omega t} \frac{d\phi(t)}{dt} dt \right) = 0,$$

where the two terms cancel each other due to the factors  $1/j\omega$  appearing in the limit.

- The comparison to the distribution  $\psi(t) \equiv 0$ , which is defined by  $\langle \psi(t), \phi(t) \rangle = 0$ , results in the generalized limit

$$\lim_{\omega \rightarrow \infty} e^{j\omega t} = 0$$

- Considering the real and the imaginary parts separately, we obtain directly:

$$\lim_{\omega \rightarrow \infty} \cos(\omega t) = 0 \quad \text{and} \quad \lim_{\omega \rightarrow \infty} \sin(\omega t) = 0$$

## Limit of the complex exponential function

- In many cases, the generalized limits lead to the dirac impulse, therefore the limits can be used for the representation of it. For example, the limit of a rectangular pulse

$$\lim_{T \rightarrow 0} \frac{1}{T} \text{rect}\left(\frac{1}{T}\right)$$

can be carried out by representing as distribution

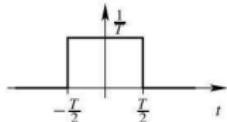
$$\left\langle \lim_{T \rightarrow 0} \frac{1}{T} \text{rect}\left(\frac{1}{T}\right), \phi(t) \right\rangle = \int_{-\infty}^{\infty} \lim_{T \rightarrow 0} \frac{1}{T} \text{rect}\left(\frac{1}{T}\right) \cdot \phi(t) dt = \lim_{T \rightarrow 0} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \phi(t) dt = \phi(0)$$

i.e. this as a generalized limit corresponds to dirac impulse  $\delta(t)$

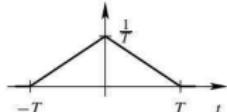
- In a similar way, other generalized limits can also be used for representing the dirac impulse. The table in the next slide shows a corresponding selection. All these limit underlying functions show the two properties, that they firstly vanish in the limit for all  $t \neq 0$  (sifting property) and secondly the value of each of the integrals is one independently from the value of the parameter.

## Representing the dirac impulse by means of the generalized limits

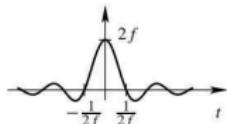
rectangular function  $\delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \text{ rect}\left(\frac{t}{T}\right)$



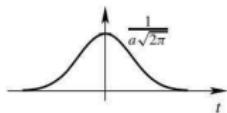
triangular function  $\delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \Lambda\left(\frac{t}{T}\right)$



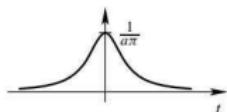
sinc function  $\delta(t) = \lim_{f \rightarrow \infty} \frac{\sin(2\pi f t)}{\pi t}$



Gaussian function  $\delta(t) = \lim_{a \rightarrow 0} \frac{1}{a \sqrt{2\pi}} e^{-\frac{t^2}{2a^2}}$



rational function  $\delta(t) = \lim_{a \rightarrow 0} \frac{1}{\pi} \frac{a}{a^2+t^2}$



## Integration of the complex exponential function

- With the help of distribution theory, the improper integral is solved by the complex exponential functions:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{j2\pi ft} df &= \lim_{F \rightarrow \infty} \int_{-F}^F e^{j2\pi ft} df = \lim_{F \rightarrow \infty} \left[ \frac{1}{j2\pi t} e^{j2\pi ft} \right]_{-F}^F \\ &= \lim_{F \rightarrow \infty} \frac{1}{j2\pi t} (e^{j2\pi F t} - e^{-j2\pi F t}) = \lim_{F \rightarrow \infty} \frac{\sin(2\pi F t)}{\pi t} \end{aligned}$$

- According to the table before, this represents a generalized limit which can be described by dirac impulse, so that we can obtain the following solution:

$$\int_{-\infty}^{\infty} e^{j2\pi ft} df = \delta(t)$$

- For the infinite sum of complex exponential function it applies:

$$\sum_{n=-\infty}^{\infty} e^{j2\pi nt} = \Xi(t)$$

The proof can be carried out simply with the help of Fourier Transformation, which will be later intensively considered

# Table of Contents

- 1 Introduction
  - Mathematical Models for Communication Channels
  - Generalized Functions
- 2 Probability
  - Fundamentals in Probability Theory
  - Useful Probability Distributions
  - Central Limit Theorem
- 3 Fourier Transform
  - Properties of FT
  - Sampling Theorem
- 4 Stochastic Processes
  - Gaussian Processes
  - Response of an LTI System to a Stationary Input Signal
  - Discrete-Time (Stochastic) Signals
  - Cyclostationary Processes
- 5 Communication Signals and Systems
  - Representation of Band-pass Signals and Systems
  - Orthogonal Expansions of Signals
  - Representation of Digitally Modulated Signals
- 6 Optimum Receivers for the AWGN Channel
  - Characterization of Thermal Noise
  - Signal Space Representation
  - Optimum Detection

# Fundamentals in Probability Theory

- **Probability:**

First, we will recap some **fundamentals in probability theory** including the **axioms of Kolmogorov**, **random variables**, **functions of random variables**, **moments** and **typical probability density functions** encountered mainly in wireless communication systems. **Bounding techniques** are introduced which serve for deriving bounds on the tail probability of random variables to be applied e.g. to the tails of Gaussian random variables arising in standard detection problems.

- **Stochastic Processes:**

The stochastic processes to be investigated describe **fading processes in wireless communications**, **noise processes in receivers** or **parameter variations in transceivers**. Important concepts include **stationarity issues**, **correlation functions**, **power density functions**, **linear systems** and continuous as well as discrete-time **stochastic signals and systems**.

- Assume we are given a set  $S$  of elementary events  $A_i, i = 1, \dots, n$  defined by

$$S = \{A_1, \dots, A_n\}.$$

$S$  is also termed **sample space**.

- As an example, for tossing a die with the usual outcomes, we would have  
 $S = \{1, 2, 3, 4, 5, 6\}.$
- From  $S$ , we could form a total of  $2^n$  subsets and assign probabilities  $P(A)$  to each subset  $A$ . The latter represents a certain **event**.
- For example, the subset  $A = \{1, 3, 5\}$  characterizes the event of tossing a face with an odd number of dots on it.
- The (elementary) events can be related to each other by usual set operations such as e.g. intersections ( $A_1 \cap A_2$ ) and unions ( $A_1 \cup A_2$ ).
- The expression  $A_1 \subset A_2$  indicates that the subset  $A_1$  is contained in  $A_2$ , i.e.  $A_1$  is a **subset** of  $A_2$  and  $A_2$  is a **superset** of  $A_1$ .

The probabilities  $P(\cdot)$  are defined by the **axioms of Kolmogorov** holding for any events  $A$  and  $B$ :

①  $P(A) \geq 0$

②  $P(S) = 1$

③ if  $A \cap B = \emptyset$  then  $P(A \cup B) = P(A) + P(B)$

Direct consequences from the axioms are (cf. Papoulis, Sect. 2.2):

- $P(\emptyset) = 0$
- for arbitrary events  $A$  and  $B$ , we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

## Joint events and joint probabilities

- consider joint events  $(A, B)$ , e.g. tossing two dice with possible outcomes  $S = (i, j), i, j = 1, \dots, 6$
- main difference of joint events as compared to single ones results from **marginalization**, i.e. the consideration of single events resulting from corresponding unions of joint events
- all properties of single events  $A_i$  and  $B_j$  for  $i = 1, \dots, n$  and  $j = 1, \dots, m$  defined by the axioms carry over to joint events, e.g.
  - $0 \leq P(A_i, B_j) \leq 1$  for all  $i, j$
  - for **mutually exclusive**  $B_j, j = 1, \dots, m$ , we have

$$\sum_{j=1}^m P(A_i, B_j) = P(A_i)$$

- for **mutually exclusive joint events**, we have  $\sum_{i=1}^n \sum_{j=1}^m P(A_i, B_j) = 1$

## Conditional probabilities

- consider joint events  $(A, B)$ ; the function

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

defines a probability measure for  $P(B) > 0$  which is the **conditional probability of  $A$  given the occurrence of  $B$**

- the joint probability can be expressed by the conditional probabilities according to

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$

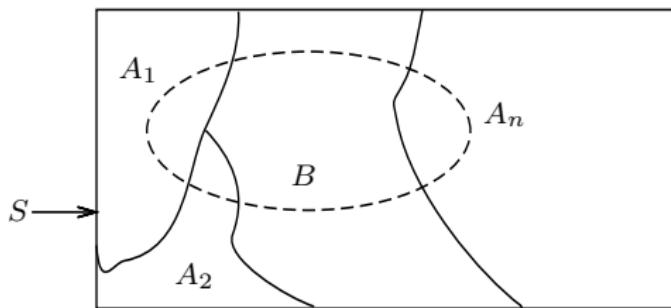
⇒ thus we have a rule for **exchanging** the arguments in the conditional probability rule

- Example for the die experiment: what is the probability of **tossing a one** (= event  $A$ ) given that the **face showing up is an odd number** (= event  $B$ )?  
⇒  $P(A|B) = \frac{1}{6}/\frac{1}{2} = \frac{1}{3}$ ; this result is intuitive, since we assume the probability of  $A$  to be  $1/3$  once we have tossed an odd number.

## Total probability and Bayes' theorem

- **Total probability theorem:** If  $\mathbf{A} = [A_1, \dots, A_n]$  is a partition of  $S$  and  $B$  is an arbitrary event, then

$$P(B) = P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)$$



- **Bayes' theorem:** inserting the total probability theorem into the definition of the conditional probability and exploiting  $P(A|B)P(B) = P(B|A)P(A)$ , we obtain Bayes' theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)}$$

Ex.: We have four boxes containing the following **good** and **defective** components: box 1: **1900 g, 100 d**, box 2: **300 g, 200 d**, box 3: **900 g, 100 d**, box 4: **900 g, 100 d**.  
We select a box at random and draw a component out of it.

- What is the probability for the component to be defective?

Let  $B_i$  the event of selecting box  $i$  and  $D$  the event of drawing a defective component. From

$$P(B_1) = P(B_2) = P(B_3) = P(B_4) = 1/4$$

and

$$P(D|B_1) = \frac{100}{2000} = 0.05$$

$$P(D|B_2) = \frac{200}{500} = 0.4$$

$$P(D|B_3) = \frac{100}{1000} = 0.1$$

$$P(D|B_4) = \frac{100}{1000} = 0.1$$

We obtain

$$P(D) = 0.05 \times \frac{1}{4} + 0.4 \times \frac{1}{4} + 0.1 \times \frac{1}{4} + 0.1 \times \frac{1}{4} = 0.1625.$$

- The drawn component is defective. What is the probability that it came from box 2?  
We want to calculate  $P(B_2|D)$ . Since

$$P(D) = 0.1625, \quad P(D|B_2) = 0.4, \quad P(B_2) = 0.25,$$

We find

$$P(B_2|D) = \frac{P(D|B_2)P(B_2)}{P(D)} = 0.4 \times \frac{0.25}{0.1625} = 0.615.$$

- A comment on **assumptions** and **deductions**:  
the numbers for  $P(D|B_i)$  have been assumed based on our experience with the different boxes, as well as the probability  $P(B_i)$  of choosing box  $i$ ; once we have fixed these numbers, we **deduce** that  $P(D) = 0.1625$  and  $P(B_2|D) = 0.615$

## Statistical independence

- Two events A and B are called **independent** if

$$P(A, B) = P(A)P(B).$$

- As a consequence, we have for the conditional probability of A given B

$$P(A|B) = \frac{P(A, B)}{P(B)} = P(A).$$

That is, no matter whether or not  $B$  has happened, the probability of event  $A$  does not change.

- Ex. (lottery jackpot): People having won a lottery drawing have exactly the same probability to win again as any other one playing the lottery, given that subsequent drawings are independent. Also, people *waiting* for their number to be drawn (since it has never been drawn before) also could have any other valid combination of numbers without a reduction in probability to win in the next drawing.

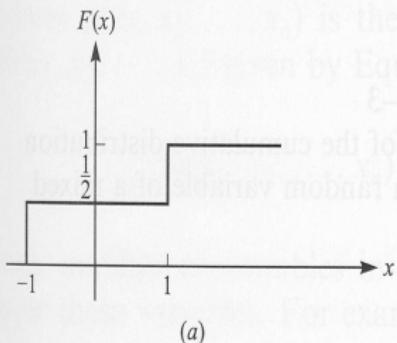
## Random variables and cumulative distribution function

- Associate a number  $X = X(s)$  with the outcomes  $s$  of an experiment in a sample space  $S$ . Thus, the domain of  $X(s)$  is  $S$ , the range in the case of a real-valued  $X$  is  $\mathbb{R}$ . This function is called a **random variable (RV)**.
- We can distinguish between **discrete** and **continuous** RVs depending on the type of sample space  $S$ .
- Suppose we are given a RV  $X$ . Consider the event  $\{X \leq x\}$ , where  $x$  is an arbitrary real number. We write the probability of this event as

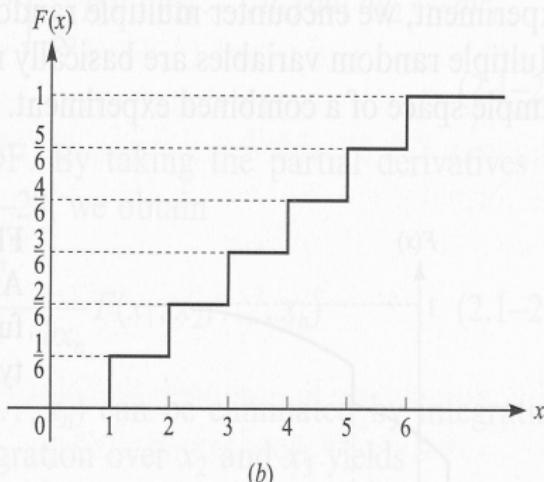
$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

- The function  $F(x)$  is called **probability distribution function** or **cumulative distribution function (CDF)** and has certain properties (e.g. non-decreasing for increasing  $x$ ). For discrete RVs,  $F(x)$  has usually jumps for those  $x$  representing the outcomes for the underlying sample space.

## Random variables and cumulative distribution function



(a)

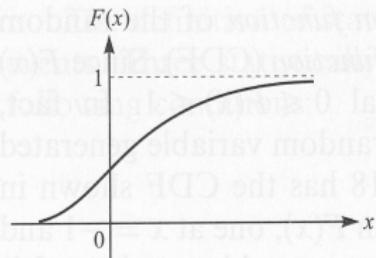


(b)

**FIGURE 2.1–1**

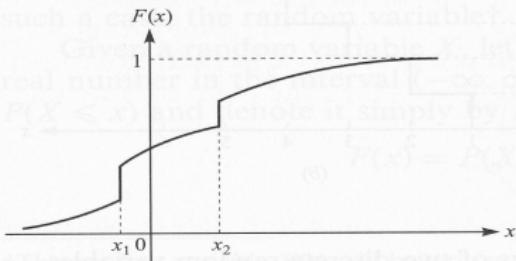
Examples of the cumulative distribution functions of two discrete random variables.

## Random variables and cumulative distribution function



**FIGURE 2.1–2**

An example of the cumulative distribution function of a continuous random variable.



**FIGURE 2.1–3**

An example of the cumulative distribution function of a random variable of a mixed type.

## Probability density function

- If the derivative of  $F(x)$  exists, we can write

$$F(x) = \int_{-\infty}^x p(u)du, \quad -\infty < x < \infty,$$

where

$$p(u) = \frac{dF(x)}{dx}, \quad -\infty < x < \infty,$$

denotes the so-called **probability density function (PDF)**. Since  $F(x)$  is nondecreasing,  $p(x)$  is non-negative.

- For the case of a discrete RV  $X$  that can take values from the set  $\{x_1, \dots, x_n\}$ , the PDF results to

$$p(x) = \sum_{i=1}^n P(X = x_i) \delta_{x-x_i}.$$

The Dirac impulses lead to the jumps in the CDF of  $X$ . Thus, a continuous RV has a continuous CDF.

## Probability density function

- The PDF can be used to find the probability that a RV lies in a certain interval.  
From

$$\begin{aligned} P(X \leq x_2) &= P(X \leq x_1) + P(x_1 < X \leq x_2) \\ F(x_2) &= F(x_1) + P(x_1 < X \leq x_2) \end{aligned}$$

with  $x_1 \leq x_2$ , we have

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x)dx.$$

- multiple RVs: the **joint CDF** for the two RVs  $X_1$  and  $X_2$  is defined as

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(u_1, u_2)du_1 du_2$$

where  $p(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2)$  denotes the **joint PDF of  $X_1$  and  $X_2$** .

### Conditional PDF and statistical independence

- **Conditional PDF:** For a given joint PDF  $p(x_1, x_2)$  and  $p(x_2) \neq 0$ , the **conditional PDF** given  $X_2 = x_2$  is defined as

$$P(x_1 | x_2) = \frac{p(x_1, x_2)}{p(x_2)}.$$

- **Statistically independent RVs:** For a given joint PDF  $p(x_1, x_2)$  and a corresponding joint CDF  $F(x_1, x_2)$ , the RVs  $X_1$  and  $X_2$  are called **statistically independent**, if we can factorize both the joint PDF and the joint CDF according to

$$\begin{aligned} F(x_1, x_2) &= F(x_1)F(x_2) \\ p(x_1, x_2) &= p(x_1)p(x_2). \end{aligned}$$

Function  $Y = g(X)$  of a random variable  $X$ : one-to-one mapping

In many cases, we are given the

PDF of a random variable  $X$

and want to find the

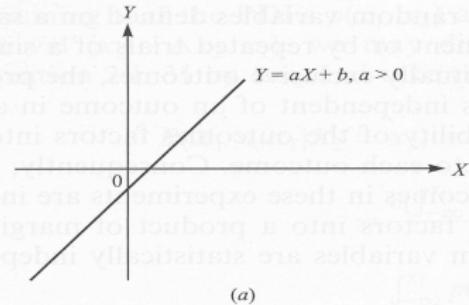
PDF of  $Y = g(X)$ ,

where  $g(X)$  is function of  $X$ . The mapping  $g$  from  $X$  to  $Y$  may be one-to-one or many-to-one.

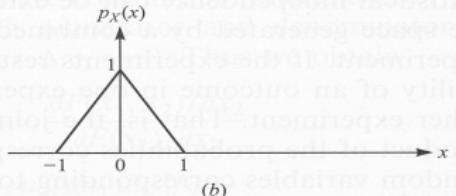
**Example:**  $Y = aX + b$  with  $a > 0$  and  $b \in \mathbb{R}$  (one-to-one). We obtain

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} p_X(x) dx \\ &= F_X\left(\frac{y-b}{a}\right) = F_X(g^{-1}(y)). \end{aligned}$$

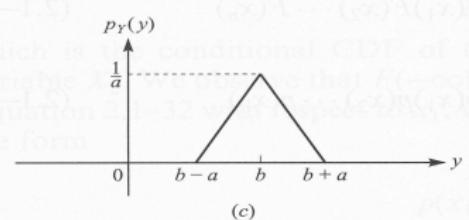
## Linear transformation of $X$



(a)



(b)



(c)

**FIGURE 2.1-4**

A linear transformation of a random variable  $X$  and an example of the corresponding PDFs of  $X$  and  $Y$ .

**Function  $Y = g(X)$  of a random variable  $X$ : many-to-one mapping**

In the general case of a many-to-one mapping  $Y = g(X)$ , we resort to the so-called

**Fundamental theorem:**

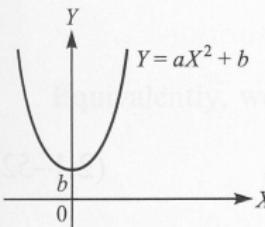
For a given random variable  $X$  with PDF  $p_X(x)$ , the function  $Y = g(X)$  has a PDF defined by

$$p_Y(y) = \sum_{i=1}^n \frac{p_X(x_i)}{|g'(x_i)|} = \sum_{i=1}^n \frac{p_X(x_i(y))}{|g'(x_i(y))|}$$

with  $g'(x) = \partial g / \partial x$  and roots  $y = g(x_1) = g(x_2) = \dots = g(x_n)$ .

**Function  $Y = g(X)$  of a random variable  $X$ : many-to-one mapping**

**Example:**  $Y = aX^2 + b$  with  $a > 0$  and  $b \in \mathbb{R}$  (**many-to-one**).



**FIGURE 2.1-5**  
A quadratic transformation of the random variable  $X$ .

- We have the two roots  $x_1 = -x_2 = \sqrt{(y-b)/a}$  for  $y \geq a(\min_{x \in \mathbb{R}} |x|)^2 + b$ .
- From  $g'(x) = 2ax$ , we find

$$p_Y(y) = \frac{p_X\left(\sqrt{(y-b)/a}\right)}{2a\sqrt{(y-b)/a}} + \frac{p_X\left(-\sqrt{(y-b)/a}\right)}{2a\sqrt{(y-b)/a}}.$$

One-to-one mapping between multidimensional random variables

- Assume we are given  $n$  functions  $Y_i = g_i(X_1, X_2, \dots, X_n)$ ,  $i = 1, 2, \dots, n$  and a joint PDF  $p_X(x_1, x_2, \dots, x_n)$ . Furthermore, the domain  $X_1, X_2, \dots, X_n \in R_X$  is mapped to  $Y_1, Y_2, \dots, Y_n \in R_Y$ .
- From

$$\int \int \dots \int_{R_Y} p_Y(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n \\ = \int \int \dots \int_{R_X} p_X(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

and the substitution

$$X_i = g_i^{-1}(Y_1, Y_2, \dots, Y_n) \equiv g_i^{-1}, \quad i = 1, 2, \dots, n,$$

We obtain directly

$$p_Y(y_1, y_2, \dots, y_n) = p_X(g_1^{-1}, g_2^{-1}, \dots, g_n^{-1}) |J|$$

One-to-one mapping between multidimensional random variables

- Here,  $J$  denotes the Jacobian of the transformation defined by the determinant

$$J = \begin{vmatrix} \frac{\partial g_1^{-1}}{\partial y_1} & \frac{\partial g_2^{-1}}{\partial y_1} & \cdots & \frac{\partial g_n^{-1}}{\partial y_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_1^{-1}}{\partial y_n} & \frac{\partial g_2^{-1}}{\partial y_n} & \cdots & \frac{\partial g_n^{-1}}{\partial y_n} \end{vmatrix}$$

- Example:**  $\mathbf{Y} = \mathbf{A}\mathbf{X}$  with non-singular  $(n \times n)$ -dimensional matrix  $\mathbf{A}$  and  $n$ -dimensional column vectors  $\mathbf{Y}$  and  $\mathbf{X}$  (often used in bit-error-rate performance analysis of receivers). From  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$  and  $\mathbf{B} = \{b_{ij}\} = \mathbf{A}^{-1}$ , we have  $J = 1/\det \mathbf{A}$  and obtain

$$p_Y(y_1, y_2, \dots, y_n) = p_X \left( \sum_{j=1}^n b_{1j} y_j, \sum_{j=1}^n b_{2j} y_j, \dots, \sum_{j=1}^n b_{nj} y_j \right) \frac{1}{|\det \mathbf{A}|}.$$

# Statistical Averages of Random Variables

- Assume we are given a PDF  $p_X(x)$  of a RV  $X$  and a function  $Y = g(X)$  resulting in a PDF  $p_Y(y)$  of  $Y$ . We want to characterize the **average** behavior of  $Y$  by a number defined as the **expected value** or **expectation of  $Y$**  defined by

$$m_Y = E[Y] = \int_{\mathbb{R}} y p_Y(y) dy = E[g(X)] = \int_{\mathbb{R}} g(X) p_X(x) dx.$$

- The expectation is a **linear operator** and can thus be exchanged (for suitable assumptions on convergence of the involved operations) with other linear operators such as sums, integration and derivative, for example

$$E[\text{SUM}[X_1, X_2]] = E[X_1 + X_2] = E[X_1] + E[X_2] = \text{SUM}[E[X_1], E[X_2]].$$

- With  $Y = (X - m_X)^n$ , the  **$n^{\text{th}}$  central moment** is defined by

$$E[Y] = E[(X - m_X)^n] = \int_{\mathbb{R}} (X - m_X)^n p_X(x) dx.$$

- For  $n = 2$ , the **variance of  $X$**  defined by

$$\sigma_X^2 = E[(X - m_X)^2] = \int_{\mathbb{R}} (X - m_X)^2 p_X(x) dx = E[X^2] - m_X^2$$

provides a **measure of the dispersion of the RV  $X$ .**

- For **jointly distributed RVs  $X_i$  and  $X_j$**  consider the **correlation** (neglecting index  $X_i X_j$  in the joint PDF) **between  $X_i$  and  $X_j$**

$$E[X_i X_j] = \int_{\mathbb{R}} \int_{\mathbb{R}} x_i x_j p(x_i, x_j) dx_i dx_j$$

and their **covariance**

$$\mu_{ij} = E[(X_i - m_i)(X_j - m_j)] = E[X_i X_j] - m_i m_j.$$

### Definitions

- Two RVs  $X_i$  and  $X_j$  are **uncorrelated** if  $E[X_i X_j] = E[X_i]E[X_j] = m_i m_j$  and thus their covariance  $\mu_{ij} = 0$ .
- If two RVs  $X_i$  and  $X_j$  are independent, they are also **uncorrelated**. Independence is, however, not a necessary condition for zero correlation.
- Two RVs  $X_i$  and  $X_j$  are **orthogonal** if  $E[X_i X_j] = 0$ . Obviously, two uncorrelated RVs with either one being a zero-mean RV are orthogonal.
- All joint central and non-central moment definitions can be extended to higher orders as well as to more than two RVs.

# Characteristic Functions of Random Variables

- It is well-known that the use of the Fourier transform in a linear system can simplify the calculation of explicit expressions, e.g. the output signal for a given input signal and a given transfer function.
- Similarly, we can **simplify the calculation of the PDF of a sum of mutually independent RVs**. To this end, we introduce the **characteristic function**  $\psi(j\nu)$  of a RV  $X$  as the (modified) **Fourier transform of  $p(x)$**  according to

$$\begin{aligned}\psi(j\nu) &\equiv E\left[e^{j\nu X}\right] = \int_{\mathbb{R}} e^{j\nu x} p(x) dx \\ p(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \psi(j\nu) e^{-j\nu x} d\nu.\end{aligned}$$

- Before calculating the aforementioned PDF of a sum of mutually independent RVs, we consider the relation between the characteristic function and moments of a RV  $X$ .

- Evaluating

$$\frac{d\psi(j\nu)}{d\nu} = j \int_{\mathbb{R}} x e^{j\nu x} p(x) dx$$

at  $\nu = 0$ , we obtain

$$E[X] = m_X = -j \left. \frac{d\psi(j\nu)}{d\nu} \right|_{\nu=0}.$$

- Repeating the differentiation process yields

$$E[X^n] = (-j)^n \left. \frac{d^n \psi(j\nu)}{d\nu^n} \right|_{\nu=0}.$$

- The characteristic function can be expressed using a Taylor expansion about  $\nu = 0$  according to

$$\psi(j\nu) = \sum_{n=0}^{\infty} \left[ \frac{d^n \psi(j\nu)}{d\nu^n} \right]_{\nu=0} \frac{\nu^n}{n!} = \sum_{n=0}^{\infty} E[X^n] \frac{(j\nu)^n}{n!}.$$

- Suppose we are to calculate the PDF of  $Y = \sum_{i=1}^n X_i$ , where  $X_i$  are statistically independent RVs with  $p(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$ .
- Using the characteristic function of  $Y$ , we can simplify the problem according to

$$\begin{aligned}\psi_Y(j\nu) &= E[e^{j\nu Y}] = E\left[e^{j\nu \sum_{i=1}^n X_i}\right] \\ &= \int_{\mathbb{R}^n} \left(\prod_{i=1}^n e^{j\nu x_i}\right) p(x_1, \dots, x_n) dx_1 \dots dx_n = \prod_{i=1}^n \psi_{X_i}(j\nu).\end{aligned}$$

- If the  $X_i$  are, in addition to the statistical independence, identically distributed, each with characteristic function  $\psi_X(j\nu)$ , we have

$$\psi_Y(j\nu) = [\psi_X(j\nu)]^n$$

and calculating the Fourier inverse of  $\psi_Y(j\nu)$  is often much less difficult than the  $n$ -fold convolution required for calculating  $p_Y(y)$  directly.

- When working with  $n$ -dimensional RVs, it is sometimes appropriate to define an  **$n$ -dimensional characteristic function** by

$$\begin{aligned}\psi(j\nu_1, j\nu_2, \dots, j\nu_n) &= E\left[e^{j \sum_{i=1}^n \nu_i X_i}\right] \\ &= \int_{\mathbb{R}^n} e^{j \sum_{i=1}^n \nu_i x_i} p(x_1, \dots, x_n) dx_1 \dots dx_n.\end{aligned}$$

- Of special interest is the case  $n = 2$ :

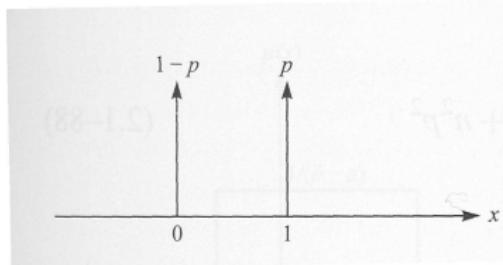
$$\psi(j\nu_1, j\nu_2) = \int_{\mathbb{R}^2} e^{j(\nu_1 x_1 + \nu_2 x_2)} p(x_1, x_2) dx_1 dx_2.$$

- As above, partial derivatives of  $\psi(j\nu_1, j\nu_2)$  can be used to generate the joint moments. For example, we obtain

$$E[X_1 X_2] = - \left. \frac{\partial^2 \psi(j\nu_1, j\nu_2)}{\partial \nu_1 \partial \nu_2} \right|_{\nu_1 = \nu_2 = 0}.$$

Useful probability distributions: **Binomial distribution**

- Let  $X$  be a **discrete** RV with  $X \in \{0, 1\}$  and  $P(X = 0) = 1 - p$  and  $P(X = 1) = p$ .



**FIGURE 2.1–6**  
The probability distribution function of  $X$ .

- Now suppose that

$$Y = \sum_{i=1}^n X_i$$

where  $X_i, i = 1, 2, \dots, n$  are **statistically independent and identically distributed (i.i.d.)** RVs with a PDF as shown in the figure.

What is the **PDF of  $Y$ ?**

## Useful probability distributions: Binomial distribution

- Obviously, the range of  $Y$  is the set of integers  $\{0, 1, \dots, n\}$ .
- From basic combinatorics, we find

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

with

$$\binom{n}{k} = \frac{n!}{k!(n - k)!}$$

denoting the binomial coefficient.

- Consequently, the PDF of the binomial distribution is given by

$$p(y) = \sum_{k=0}^n P(Y = k) \delta(y - k) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta(y - k).$$

## Useful probability distributions: Binomial distribution

- The CDF of the binomial distribution is given by

$$F(y) = P(Y \leq y) = \sum_{k=0}^{\lfloor y \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$

with  $\lfloor y \rfloor$  denoting the largest integer  $m$  such that  $m \leq y$ .

- The first two moments of  $Y$  are given by

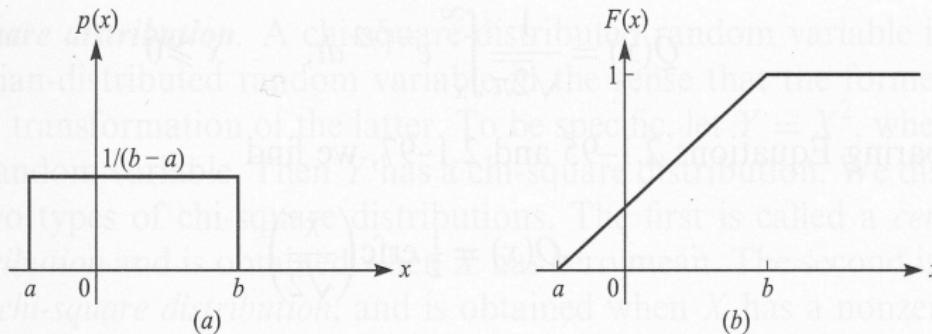
$$\begin{aligned} E[Y] &= np \\ E[Y^2] &= np(1-p) + n^2 p^2 \\ \sigma_Y^2 &= np(1-p). \end{aligned}$$

- The characteristic function of  $Y$  (sum of i.i.d. RVs) is given by

$$\psi_Y(j\nu) = \psi_X^n(j\nu) = (1 - p + pe^{j\nu})^n.$$

Useful probability distributions: **Uniform distribution**

- The PDF and CDF of a uniformly distributed (continuous) RV  $X$  with  $X \in [a, b]$  is shown below.



**FIGURE 2.1-7**

The PDF and CDF of a uniformly distributed random variable.

## Useful probability distributions: Uniform distribution

- The first two moments of  $X$  are

$$\begin{aligned}E[X] &= \frac{1}{2}(a + b) \\E[X^2] &= \frac{1}{3}(a^2 + b^2 + ab) \\\sigma_X^2 &= \frac{1}{12}(a - b)^2.\end{aligned}$$

- The characteristic function of  $X$  (sum of i.i.d. RVs) is given by

$$\psi_X(j\nu) = \frac{e^{j\nu b} - e^{j\nu a}}{j\nu(b - a)}.$$

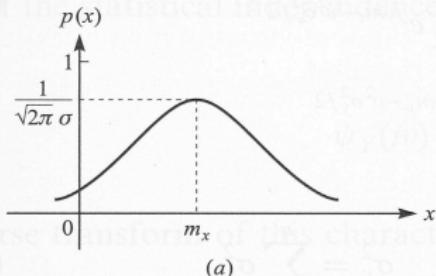
- A uniformly distributed RV implemented e.g. in a computer program can be used to generate other distributions (cf. transformation method below).

Useful probability distributions: Gaussian (normal) distribution

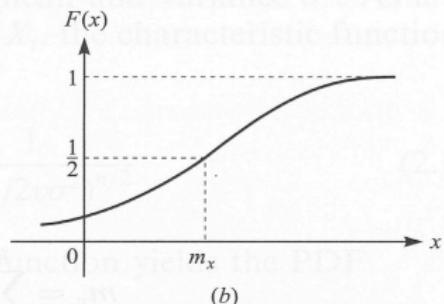
- The PDF of a Gaussian (or normally) distributed RV  $X$  is

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m_X)^2}{2\sigma^2}},$$

where  $m_X = E[X]$  and  $\sigma^2 = \sigma_X^2$  is the variance of  $X$ .



(a)



(b)

**FIGURE 2.1–8**

The PDF and CDF of a Gaussian-distributed random variable.

Useful probability distributions: Gaussian (normal) distribution

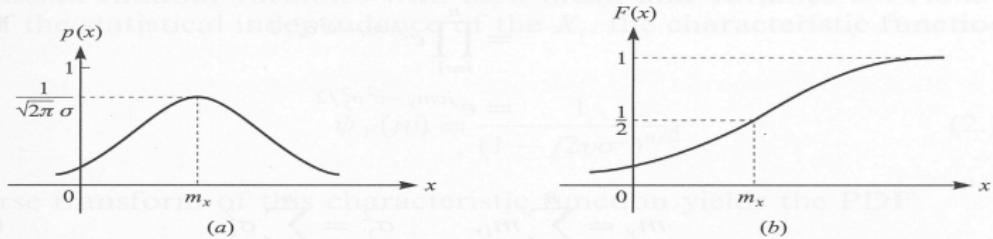
- The CDF of a Gaussian (or normally) distributed RV  $X$  is

$$\begin{aligned}
 F(x) &= \int_{-\infty}^x p(u)du \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(u-m_X)^2}{2\sigma^2}} du \\
 &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{x-m_X}{\sqrt{2\sigma}}} e^{-t^2} dt \\
 &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^0 e^{-t^2} dt + \frac{1}{\sqrt{\pi}} \int_0^{\frac{x-m_X}{\sqrt{2\sigma}}} e^{-t^2} dt \\
 &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x-m_X}{\sqrt{2\sigma}}\right),
 \end{aligned}$$

where the **error function** is defined as

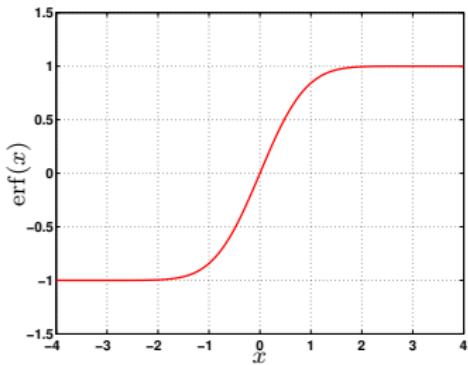
$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

## Gaussian (normal) distribution



**FIGURE 2.1-8**

The PDF and CDF of a Gaussian-distributed random variable.



### Gaussian (normal) distribution

- Sometimes, the CDF is expressed by the **complementary error function**

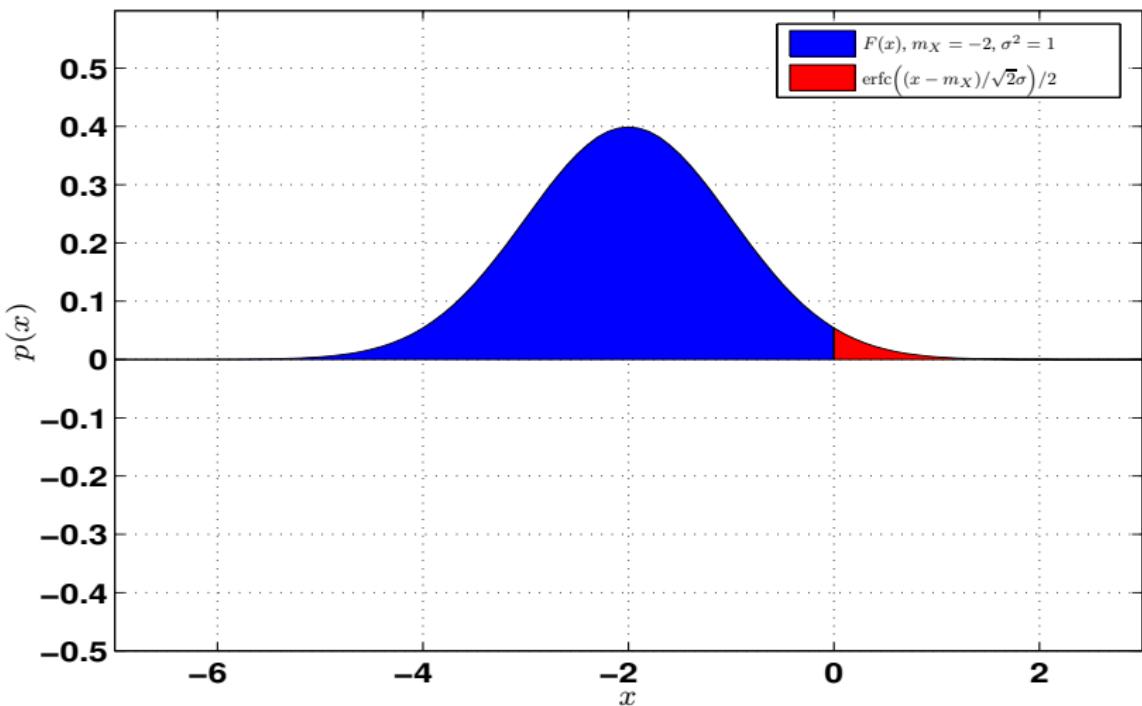
$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt = 1 - \operatorname{erf}(x)$$

according to

$$F(x) = 1 - \frac{1}{2} \operatorname{erfc}\left(\frac{x - m_X}{\sqrt{2}\sigma}\right).$$

- The complementary error function serves for characterizing the **area under the tail of the Gaussian PDF** for  $x > m_X$ .

### Gaussian (normal) distribution



### Gaussian (normal) distribution

- The central moments of a Gaussian RV are

$$E[(X - m_X)^k] \equiv \mu_k = \begin{cases} 1 \cdot 3 \cdot \dots \cdot (k-1)\sigma^k & \text{for even } k \\ 0 & \text{for odd } k. \end{cases}$$

- The ordinary moments can be expressed by the central moments as

$$E[X^k] = \sum_{i=0}^k \binom{k}{i} m_X^i \mu_{k-i}.$$

- The multivariate Gaussian distribution  $\mathcal{N}(\mathbf{m}_X; \mathbf{M})$  is defined by

$$p(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{M}}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{m}_X)^T \mathbf{M}^{-1} (\mathbf{x} - \mathbf{m}_X) \right)$$

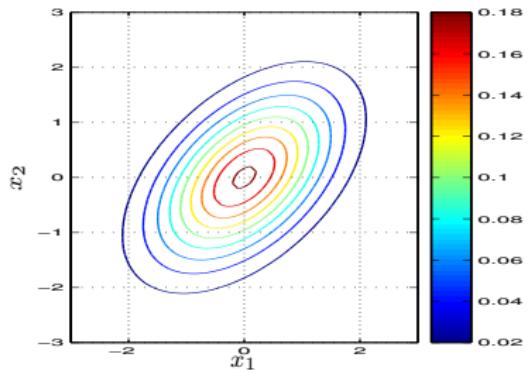
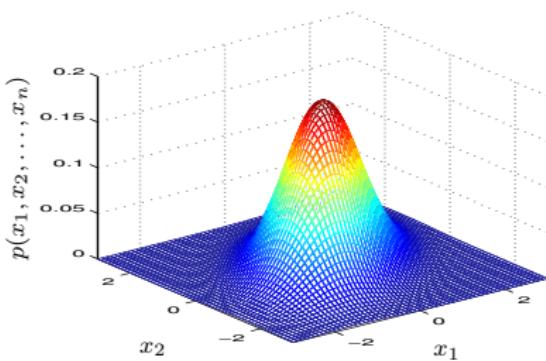
with  $\mathbf{m}_X = [E[X_1], E[X_2], \dots, E[X_n]]^T$  and the covariance matrix

$$\mathbf{M} = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T].$$

## Multivariate Gaussian (normal) distribution

We denote a multivariate Gaussian random vector  $\mathbf{X}$  by  $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_\mathbf{X}; \mathbf{M})$  (reads:  $\mathbf{X}$  is multivariate Gaussian distributed with expected value  $\mathbf{m}_\mathbf{X}$  and covariance matrix  $\mathbf{M}$ ).

**Example:**  $\mathbf{m}_\mathbf{X} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$



## Multivariate Gaussian (normal) distribution

- The  $n$ -dimensional Fourier transform yields the characteristic function

$$\psi(j\mathbf{v}) = \exp\left(j\mathbf{m}_X^T \mathbf{v} - \frac{1}{2}\mathbf{v}^T \mathbf{M} \mathbf{v}\right)$$

- For a **diagonal covariance matrix**  $\mathbf{M}$ , the components of  $\mathbf{X}$  are independent (by consideration of either  $p(\mathbf{x})$  or  $\psi(j\mathbf{v})$ ). Thus, **if components of a joint Gaussian distribution are uncorrelated, they are also independent** (not true, in general, for other types of distribution)
- Linear transformation**

Consider a linear transformation of a Gaussian distributed random vector  $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_X; \mathbf{M})$  according to

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

with a non-singular  $(n \times n)$ -dimensional matrix  $\mathbf{A}$ .

## Multivariate Gaussian (normal) distribution

**Linear transformation** As shown above, the jacobian of this transformation is  $J = 1/\det \mathbf{A}$ . Since  $\mathbf{X} = \mathbf{A}^{-1}(\mathbf{Y} - \mathbf{b})$ , we have

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{M}} |\det \mathbf{A}|} \times \\ &\quad \exp\left(-\frac{1}{2}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \mathbf{m}_X)^T \mathbf{M}^{-1} (\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \mathbf{m}_X)\right) \\ &= \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{Q}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{m}_Y)^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{m}_Y)\right) \end{aligned}$$

with

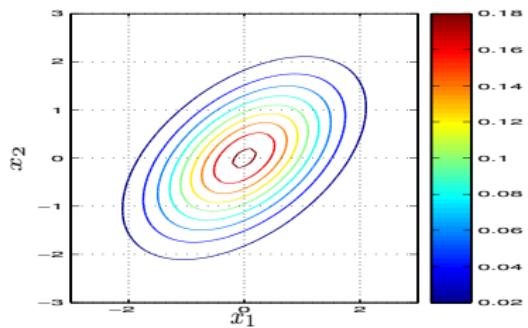
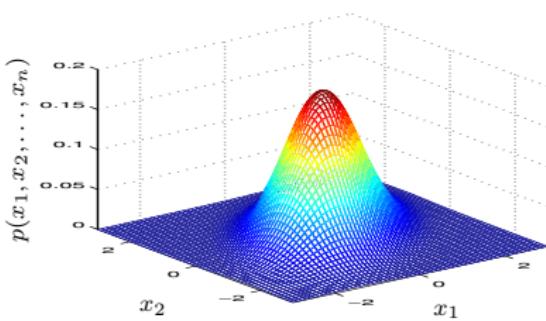
$$\begin{aligned} \mathbf{m}_Y &= \mathbf{A}\mathbf{m}_X + \mathbf{b} \\ \mathbf{Q} &= \mathbf{A}\mathbf{M}\mathbf{A}^T. \end{aligned}$$

Linearly transformed jointly Gaussian RVs are jointly Gaussian distributed.

### Multivariate Gaussian (normal) distribution

**Example:** Reconsider the example

$$\mathbf{m}_X = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$



Sometimes it makes sense to use a linear transformation to **decorrelate** the Gaussian components  $\Rightarrow$  the transformation matrix  $\mathbf{A}$  should thus be chosen so that

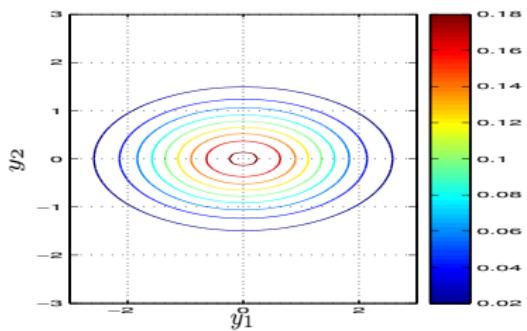
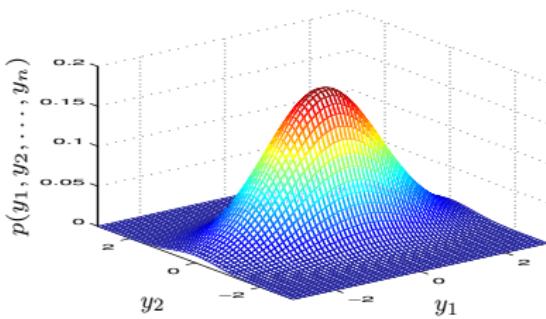
$$\mathbf{A}\mathbf{M}\mathbf{A}^T = \mathbf{D} \text{ is a diagonal matrix.}$$

## Multivariate Gaussian (normal) distribution

Since  $\mathbf{M}$  is a symmetric matrix, we can choose the transformation matrix to be orthogonal ( $\mathbf{A}^T = \mathbf{A}^{-1}$ ) consisting of the eigenvectors of  $\mathbf{M}$  given by

$$\mathbf{A} = \sqrt{\frac{1}{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A}\mathbf{M}\mathbf{A}^T = \mathbf{D} = \begin{bmatrix} 3/2 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

Resulting PDF of decorrelated Gaussian random vector  $\mathbf{Y} = \mathbf{AX}$ :



### Transformation method for generation of a RV with prescribed PDF

- Assume we are given a uniformly distributed RV  $X$  with

$$p(x)dx = \begin{cases} dx & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and  $\int_{\mathbb{R}} p(x)dx = 1$ , implemented by a **random number generator (RNG)** on a computer. Can we use this RNG to simulate a **Gaussian** RV  $Y$ ?

- We extend the problem by looking at a generic RV  $Y$  to be simulated having an **invertible CDF**  $F(y)$  and PDF  $p(y) = dF(y)/dy$ . From the fundamental theorem, we have

$$|p(y)dy| = |p(x)dx| \Leftrightarrow p(y) = p(x) \left| \frac{dx}{dy} \right|.$$

## Transformation method for generation of a RV with prescribed PDF

- Since  $p(x) = 1$  for  $0 < x < 1$ , we set

$$p(y) = p(x) \left| \frac{dx}{dy} \right| = \left| \frac{dx}{dy} \right| = \frac{dx}{dy} \quad \text{for } dx/dy > 0.$$

- This is a differential equation with solution  $x = F(y) \Leftrightarrow y = F^{-1}(x)$ .

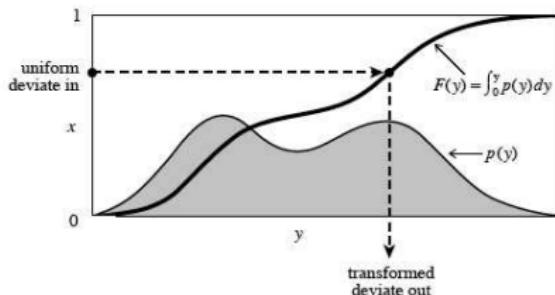


Fig. 7.2.1 in [1]: Transformation method for generating a random deviate  $y$  from a known probability distribution  $p(y)$ . The indefinite integral of  $p(y)$  must be known and invertible. A uniform deviate  $x$  is chosen between 0 and 1. Its corresponding  $y$  on the definite-integral curve is the desired deviate.

**Interpretation:** Since  $F(y)$  is the area under the PDF left to  $y$ , choose a uniform random  $X$ , find the value  $Y$  that has that fraction  $X$  of probability area to its left and return the value of that  $Y$ .

From [1]: Numerical Recipes in FORTRAN 77: The Art of Scientific Computing (ISBN 0-521-43064-X).

## Central $\chi^2$ (chi-square) distribution

- Suppose we are given  $n$  i.i.d. Gaussian RVs  $X_i, i = 1, 2, \dots, n$ , with  $X_i \sim \mathcal{N}(0; \sigma^2)$  and consider a transformation  $Y = \sum_{i=1}^n X_i^2$
- The RV  $Y$  has a **central chi-square distribution with  $n$  degrees of freedom** and we write  $Y \sim \chi_n^2$
- For  $n = 1$ , we have  $Y = X_1^2$ . From the many-to-one mapping example with  $a = 1$  and  $b = 0$ , we obtain immediately

$$p_Y(y) = \frac{u(y)}{\sqrt{2\pi y}\sigma} e^{-\frac{y}{2\sigma^2}} \quad \text{with the char. function } \psi(j\nu) = \frac{1}{(1 - j2\nu\sigma^2)^{1/2}},$$

where  $u(p)$  is the unit-step function defined by

$$u(p) = \begin{cases} 1 & \text{for } p \geq 0 \\ 0 & \text{for } p < 0 \end{cases}.$$

## Central $\chi^2$ (chi-square) distribution

- For  $n \geq 1$ , we thus have

$$\psi_Y(j\nu) = \frac{1}{(1 - j2\nu\sigma^2)^{n/2}} \quad \text{with PDF} \quad p_Y(y) = \frac{u(y)}{\sigma^n 2^{n/2} \Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}.$$

- Here,  $\Gamma(p)$  is the gamma function defined as

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$$

$$\Gamma(p) = (p-1)! \quad \text{for a positive integer } p$$

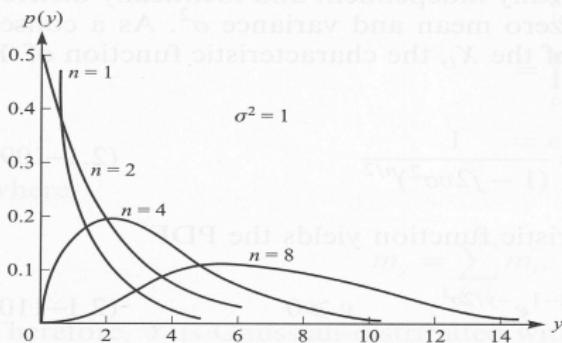
$$\Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(p+1) = p \Gamma(p)$$

## Central $\chi^2$ (chi-square) distribution

The first two moments of  $Y$  are

$$\begin{aligned} E[Y] &= n\sigma^2 \\ E[Y^2] &= 2n\sigma^4 + n^2\sigma^4 \\ \sigma_Y^2 &= 2n\sigma^4. \end{aligned}$$



**FIGURE 2.1-9**

The PDF of a chi-square-distributed random variable for several degrees of freedom.

## Non-central $\chi^2$ (chi-square) distribution

- Suppose we are given  $n$  independent Gaussian RVs  $X_i, i = 1, 2, \dots, n$ , with  $X_i \sim \mathcal{N}(m_i; \sigma^2)$  and consider a transformation  $Y = \sum_{i=1}^n X_i^2$
- The RV  $Y$  has a non-central chi-square distribution with  $n$  degrees of freedom.
- For  $n = 1$ , we have  $Y = X_1^2$ . From frame #49 with  $a = 1$  and  $b = 0$ , we obtain immediately

$$p_Y(y) = \frac{u(y)}{\sqrt{2\pi y}\sigma} e^{-\frac{y+m_1^2}{2\sigma^2}} \cosh\left(\frac{\sqrt{y}m_1}{\sigma^2}\right).$$

- The characteristic function results to

$$\psi(j\nu) = \frac{e^{jm_1^2\nu/(1-j2\nu\sigma^2)}}{(1-j2\nu\sigma^2)^{1/2}}.$$

## Non-central $\chi^2$ (chi-square) distribution

- For  $n \geq 1$ , we thus have

$$\psi_Y(j\nu) = \frac{\exp\left(j\frac{\nu \sum_{i=1}^n m_i^2}{1-j2\nu\sigma^2}\right)}{(1-j2\nu\sigma^2)^{n/2}}$$

- Taking the Fourier inverse of  $\psi_Y(j\nu)$ , we obtain

$$p_Y(y) = \frac{u(y)}{2\sigma^2} \left(\frac{y}{s^2}\right)^{(n-2)/4} e^{-\frac{s^2+y}{2\sigma^2}} I_{\frac{n}{2}-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right)$$

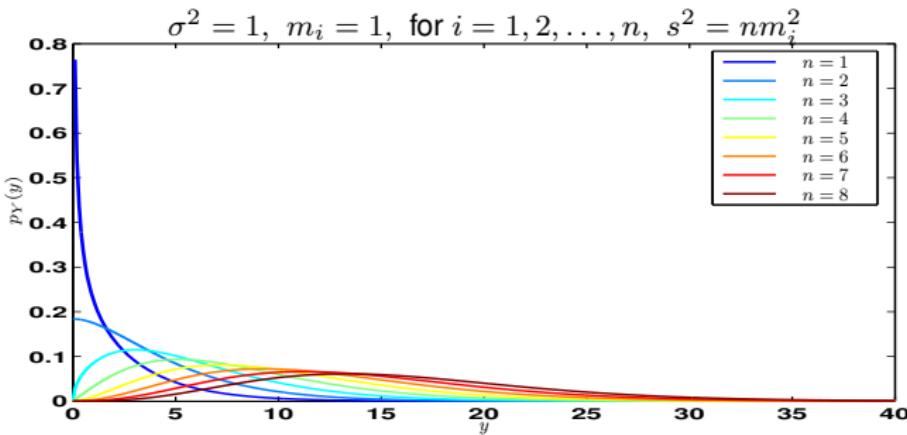
with the noncentrality parameter  $s^2 = \sum_{i=1}^n m_i^2$  and the  $\alpha^{\text{th}}$ -order modified Bessel function of the first kind

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\alpha+2k}}{k!\Gamma(\alpha+k+1)}, \quad x \geq 0.$$

## Non-central $\chi^2$ (chi-square) distribution

The first two moments of  $Y$  are

$$\begin{aligned} E[Y] &= n\sigma^2 + s^2 \\ E[Y^2] &= 2n\sigma^4 + 4\sigma^2 s^2 + (n\sigma^2 + s^2)^2 \\ \sigma_Y^2 &= 2n\sigma^4 + 4\sigma^2 s^2. \end{aligned}$$

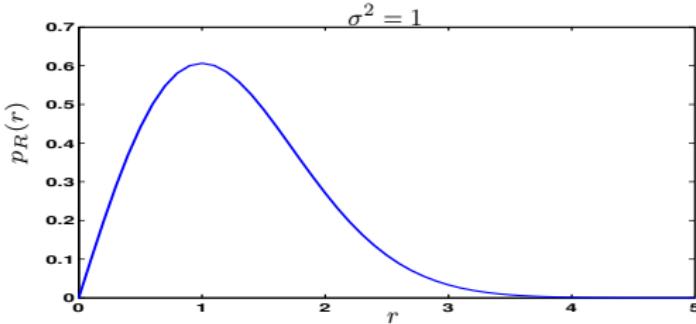


## Rayleigh distribution

- Suppose  $Y = X_1^2 + X_2^2$  with i.i.d.  $X_1, X_2 \sim \mathcal{N}(0; \sigma^2)$ , so that  $Y$  has a central chi-square distribution with **two degrees of freedom**  $p_Y(y) = \frac{1}{2\sigma^2} e^{-\frac{y}{2\sigma^2}} u(y)$
- Then,  $R = \sqrt{Y} = \sqrt{X_1^2 + X_2^2}$  has a **Rayleigh PDF** (use fundamental theorem)

$$p_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r).$$

- $R$  can be interpreted as the length of a jointly Gaussian distributed random vector in  $\mathbb{R}^2$ .



## Rice distribution

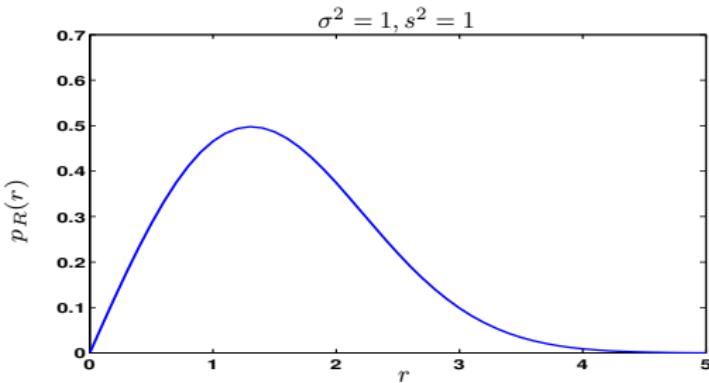
- Suppose  $Y = X_1^2 + X_2^2$  with independent  $X_i \sim \mathcal{N}(m_i; \sigma^2)$  for  $i = 1, 2$ , so that  $Y$  has a non-central chi-square distribution with **two degrees of freedom**

$$p_Y(y) = \frac{1}{2\sigma^2} e^{-\frac{y+s^2}{2\sigma^2}} I_0\left(\sqrt{y}\frac{s}{\sigma^2}\right) u(y)$$

- Then,  $R = \sqrt{Y} = \sqrt{X_1^2 + X_2^2}$  has a **Rice PDF** (use fundamental theorem)

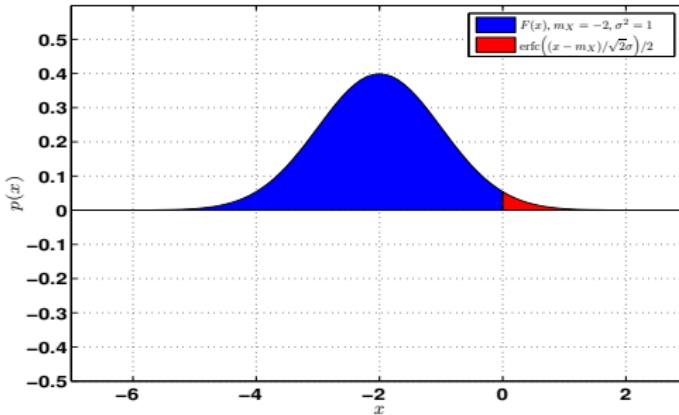
$$p_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+s^2}{2\sigma^2}} I_0\left(\frac{rs}{\sigma^2}\right) u(r).$$

- Again,  $R$  can be interpreted as the length of a jointly Gaussian distributed random vector in  $\mathbb{R}^2$



## Bounds on Tail Probability

- Remember that for the Gaussian PDF, we have explicitly defined the function  $\text{erfc}(x)$  to evaluate the so-called tail probability



- For the general case of an arbitrary PDF, this function cannot be defined or is unknown. Thus, we try to find upper bounds for the tail probability instead of finding exact expressions for it
- It is highly desired to characterize the corresponding bound using moments instead of using the complete characteristics of the PDF itself

## The Chebyshev inequality

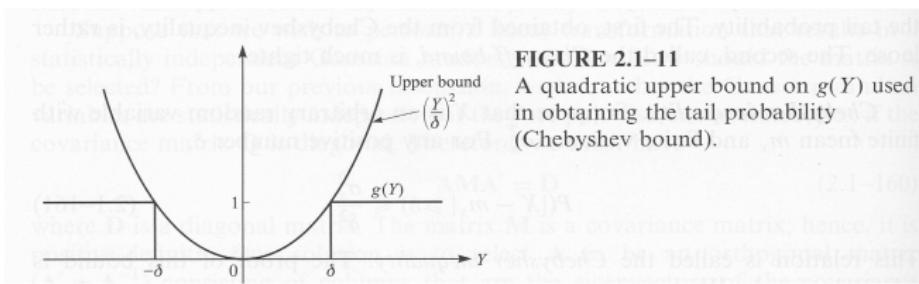
- For a zero-mean RV  $Y$ , we want to upper-bound the probability  $P(|Y| \geq \delta)$ . To this end, define a function

$$g(Y) = \begin{cases} 1 & \text{for } |Y| \geq \delta \\ 0 & \text{for } |Y| < \delta \end{cases}.$$

Thus, we can formulate the probability to be bounded as

$$E[g(Y)] = P(|Y| \geq \delta).$$

- As a second step, suppose that we upper-bound  $g(Y)$  by  $g(Y) \leq (Y/\delta)^2$



### The Chebyshev inequality

- As a result, we have

$$E[g(Y)] = P(|Y| \geq \delta) \leq E[(Y/\delta)^2] = \frac{E[Y^2]}{\delta^2} = \frac{\sigma_Y^2}{\delta^2}$$

and thus the Chebyshev inequality

$$P(|Y| \geq \delta) \leq \frac{\sigma_Y^2}{\delta^2}$$



Once we know the variance of  $Y$ , we can upper-bound the tail probability



For many applications, the Chebyshev inequality is extremely loose resulting from the looseness of the quadratic  $(Y/\delta)^2$  in overbounding  $g(Y)$

- Which other bound can be used to improve the Chebyshev inequality?
- Which other bound can be used to upper-bound a single tail probability?

### The Chernoff bound

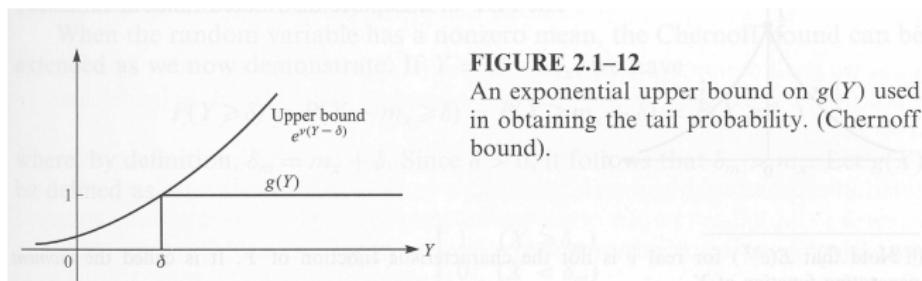
- We want to upper-bound the probability  $P(Y \geq \delta)$ . To this end, define the function

$$g(Y) = \begin{cases} 1 & \text{for } Y \geq \delta \\ 0 & \text{for } Y < \delta. \end{cases}$$

Thus, we can formulate the probability to be bounded as

$$E[g(Y)] = P(Y \geq \delta).$$

- Suppose that we upper-bound  $g(Y)$  by  $g(Y) \leq e^{\nu(Y-\delta)}$  for  $\nu \geq 0$



### The Chernoff bound

- Similar to the Chebyshev inequality,  $E[g(Y)] = P(Y \geq \delta) \leq E[e^{\nu(Y-\delta)}]$  and we choose the non-negative value  $\nu$  which minimizes  $E[e^{\nu(Y-\delta)}]$
- Necessary condition:** Let  $\nu = \hat{\nu}$  satisfy

$$\begin{aligned} \frac{d}{d\nu} E[e^{\nu(Y-\delta)}] &= E\left[\frac{d}{d\nu} e^{\nu(Y-\delta)}\right] = E[(Y-\delta)e^{\nu(Y-\delta)}] \\ &= e^{-\nu\delta} (E[Ye^{\nu Y}] - \delta E[e^{\nu Y}]) = 0 \\ \Rightarrow E[Ye^{\nu Y}] &= \delta E[e^{\nu Y}] \text{ with } \nu \geq 0. \end{aligned}$$

- Then the **Chernoff bound** reads  $P(Y \geq \delta) \leq e^{-\hat{\nu}\delta} E[e^{\hat{\nu}Y}]$



Exponential decay with  $\delta$ .



Requires solution of  $E[Ye^{\nu Y}] = \delta E[e^{\nu Y}]$

# The Weak Law of Large Numbers

- Suppose we are given i.i.d. RVs  $X_i, i = 1, 2, \dots, n$  with finite mean  $m_X$  and finite variance  $\sigma_X^2$ . Consider the sample mean

$$Y = \frac{1}{n} \sum_{i=1}^n X_i.$$

- It can be easily shown that  $E[Y] = m_Y = m_X$  and  $E[(Y - m_X)^2] = \sigma_Y^2 = \sigma_X^2/n$ .
- Since  $Z = Y - m_Y$  is a zero-mean RV and thus  $E[Z^2] = \sigma_Z^2 = \sigma_Y^2$ , we can apply the Chebyshev inequality and obtain

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - m_X\right| \geq \delta\right) = P(|Y - m_Y| \geq \delta) \leq \frac{\sigma_Y^2}{\delta^2} = \frac{\sigma_X^2}{n\delta^2}.$$

- Weak law of large numbers:  $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - m_X\right| \geq \delta\right) = 0$ .

# Central Limit Theorem

- Suppose we are given i.i.d. RVs  $X_i, i = 1, 2, \dots, n$  with finite mean  $m_X$  and finite variance  $\sigma_X^2$ .
- Normalize RVs for  $i = 1, 2, \dots, n$  according to

$$U_i = \frac{X_i - m_X}{\sigma_X},$$

where the  $U_i$  have zero mean and unit variance.

- Consider the RV

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i,$$

where again  $Y$  have zero mean and unit variance.

- What is the PDF of  $Y$  for  $n \rightarrow \infty$ ?

- Consider the characteristic function (CHF) of  $Y$  given by

$$\psi_Y(j\nu) = E \left[ \exp \left( \frac{j\nu \sum_{i=1}^n U_i}{\sqrt{n}} \right) \right] = \prod_{i=1}^n \psi_{U_i} \left( \frac{j\nu}{\sqrt{n}} \right) = \left( \psi_U \left( \frac{j\nu}{\sqrt{n}} \right) \right)^n.$$

- Expand  $\psi_U \left( \frac{j\nu}{\sqrt{n}} \right)$  in a Taylor series around  $\nu = 0$  and exploit  $E[U] = 0$  and  $E[U^2] = 1$ :

$$\begin{aligned}\psi_U \left( \frac{j\nu}{\sqrt{n}} \right) &= 1 + j \frac{\nu}{\sqrt{n} 1!} E[U] - \frac{\nu^2}{n 2!} E[U^2] - j \frac{\nu^3}{(\sqrt{n})^3 3!} E[U] + \dots \\ &= 1 - \frac{\nu^2}{2n} + \frac{1}{n} R(\nu, n)\end{aligned}$$

with  $R(\nu, n) \xrightarrow{n \rightarrow \infty} 0$ .

- As a result, we have for the CHF of  $Y$

$$\psi_Y(j\nu) = \left[ 1 - \frac{\nu^2}{2n} + \frac{1}{n} R(\nu, n) \right]^n.$$

- Take the natural logarithm of  $\psi_Y(j\nu)$

$$\ln \psi_Y(j\nu) = n \ln \left[ 1 - \frac{\nu^2}{2n} + \frac{1}{n} R(\nu, n) \right]$$

and exploit, that for small values of  $x$ , we have

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots$$

- As a consequence

$$\ln \psi_Y(j\nu) = n \left[ -\frac{\nu^2}{2n} + \frac{R(\nu, n)}{n} - \frac{1}{2} \left( -\frac{\nu^2}{2n} + \frac{R(\nu, n)}{n} \right)^2 + \dots \right].$$

- As a final step, we let  $n \rightarrow \infty$  and obtain

$$\ln \psi_Y(j\nu) = -\nu^2/2 \quad \Leftrightarrow \quad \lim_{n \rightarrow \infty} \psi_Y(j\nu) = e^{-\nu^2/2}.$$

- Consider a Gaussian RV  $X \sim \mathcal{N}(m_X; \sigma_X^2)$  and calculate its CHF:

$$\begin{aligned}\psi_X(j\nu) &= E[e^{j\nu X}] = \frac{1}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{j\nu x - \frac{(x-m_X)^2}{2\sigma_X^2}} dx \\ &= e^{j\nu m_X - \nu^2 \sigma_X^2 / 2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m_X-j\nu\sigma_X^2)^2}{2\sigma_X^2}} dx \\ &= e^{j\nu m_X - \nu^2 \sigma_X^2 / 2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma_X^2}} dx = e^{j\nu m_X - \nu^2 \sigma_X^2 / 2}.\end{aligned}$$

- Identifying the blue and the red expressions for  $m_X = 0$  and  $\sigma_X^2 = 1$  with each other, we conclude that  $Y$  has an asymptotic distribution given by  $\mathcal{N}(0; 1)$  independently of  $p_U(u)$

The sum of i.i.d. RVs with finite mean and variance is asymptotically Gaussian distributed

- The CLT has different versions where one version treats the case of independent RVs having different PDFs
- It is important to check the existence of the first two moments as indicated in the formulation of the CLT
- A counter example with non-existing expectation and variance is the (standard) Cauchy distribution

$$p_X(x) = \frac{1}{\pi(1+x^2)},$$

where the CLT cannot be applied without further assumptions on the convergence properties of the involved integrals

# Table of Contents

## 1 Introduction

- Mathematical Models for Communication Channels
- Generalized Functions

## 2 Probability

- Fundamentals in Probability Theory
- Useful Probability Distributions
- Central Limit Theorem

## 3 Fourier Transform

- Properties of FT
- Sampling Theorem

## 4 Stochastic Processes

- Gaussian Processes
- Response of an LTI System to a Stationary Input Signal
- Discrete-Time (Stochastic) Signals
- Cyclostationary Processes

## 5 Communication Signals and Systems

- Representation of Band-pass Signals and Systems
- Orthogonal Expansions of Signals
- Representation of Digitally Modulated Signals

## 6 Optimum Receivers for the AWGN Channel

- Characterization of Thermal Noise
- Signal Space Representation
- Optimum Detection

# Fourier Transform

- Until now, we have considered **random variables** with corresponding PDFs, CDFs and basic properties being relevant for the subsequent treatment of communication systems.
- The **Fourier transform (FT)** has been mentioned briefly before in the context of
  - ① mathematical models for communication channels (applied to signals in time domain)
  - ② characteristic functions to be used for characterization of random variables.
- Subsequently, we will consider some aspects of the FT for characterizing **deterministic signals and channels**.
- Upon discussion of basic properties of signals for the deterministic case, we will extend the scope to the case of **stochastic processes** where we will make use of the FT in a probabilistic (rather than deterministic) framework.

- Assume we are given a signal  $s(t)$  and want to find a **spectral representation in a frequency domain** similar to a spectrum of light.

$$\begin{aligned} S(f) &= \mathcal{F}\{s(t)\} & = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt \\ s(t) &= \mathcal{F}^{-1}\{S(f)\} = \mathcal{F}^{-1}\{\mathcal{F}\{s(t)\}\} & = \int_{-\infty}^{\infty} S(f) \exp(j2\pi ft) df. \end{aligned}$$

where all integrals are assumed to exist (cf. mathematical literature for details on convergence).

- The functions  $s(t)$  and  $S(f)$  form a **Fourier pair** which we denote as  $s(t) \longleftrightarrow S(f)$ . Even if  $s(t) \in \mathbb{R}$ , we have, in general,  $S(f) \in \mathbb{C}$ .
- N.B.** The FT  $S(f)$  provides a spectral characterization of  $s(t)$  without a temporal localization of the frequency contributions (i.e., different from notes in a music piece). Time-frequency characterization approaches (rather than the usual FT) are needed to provide short-term spectra.

## Properties of FT

Below, we will discuss some properties of the FT  $S(f)$  of  $s(t)$ , where we define

$$s(t) \circ\bullet S(f) = \Re\{S(f)\} + j\Im\{S(f)\} = M(f) + jN(f)$$

and  $M(f)$  and  $N(f)$ , resp., denote the real and imaginary parts, resp., of  $S(f)$ . If not stated otherwise, we assume  $s(\infty) = s(-\infty) = 0$ .

1. **Limiting value of  $S(f)$  (lemma of Riemann-Lebesgue)** Under mild conditions on  $s(t)$ , we have

$$S(\infty) = S(-\infty) = 0.$$

This property seems plausible since for increasing  $|f|$ , the exponential  $\exp(-j2\pi ft) = \cos(2\pi ft) - j\sin(2\pi ft)$  is a fast oscillating function so that the contributions  $s(t) \exp(-j2\pi ft)$  in  $S(f)$  cancel out.

2. **Linearity** Let  $s_i(t) \circ\bullet S_i(f)$  for  $i = 1, 2$  and consider arbitrary constants  $a_i$ , for  $i = 1, 2$ . Then we have

$$a_1 s_1(t) + a_2 s_2(t) = s(t) \circ\bullet S(f) = a_1 S_1(f) + a_2 S_2(f).$$

3. **Conjugate complex time function** From the definition of  $S(f)$ , we immediately obtain

$$s^*(t) \circ\longmapsto S^*(-f)$$

where  $s(t)^*$  denotes the conjugate complex function of  $s(t)$ , i.e.

$$s^*(t) = \Re\{s(t)\} - j\Im\{s(t)\}.$$

4. **Real function  $s(t)$**  For  $s(t) \in \mathbb{R}$ , i.e.  $s(t) = s^*(t) \forall t \in \mathbb{R}$ , we have (cf. exercises)  
 $\forall f$

$$\begin{aligned} S(-f) &= S^*(f) && \text{or equivalently} \\ M(-f) &= M(f) && \text{and} && N(-f) = -N(f). \end{aligned}$$

Thus, for real-valued  $s(t)$ , the real part of  $S(f)$  is an even function, and the imaginary part of  $S(f)$  is an odd function.

5. **Imaginary function  $s(t)$**  For  $s(t) = -s^*(t) \forall t \in \mathbb{R}$ , we have  $\forall f$  (to be shown in analogy with property 4)

$$\begin{aligned} S(-f) &= -S^*(f) && \text{or equivalently} \\ M(-f) &= -M(f) && \text{and} \quad N(-f) = N(f). \end{aligned}$$

Thus, for imaginary-valued  $s(t)$ , the real part of  $S(f)$  is an odd function, and the imaginary part of  $S(f)$  is an even function.

6. **Conjugate complex frequency function** From the definition of  $S(f)$ , we immediately obtain

$$s^*(-t) \circ\bullet S^*(f).$$

7. **Reflection at origin** From the definition of  $S(f)$ , we immediately obtain

$$s(-t) \circ\bullet S(-f).$$

8. **Time shift** For  $s(t - t_0)$  with  $t_0$  representing a real constant, we have

$$s(t - t_0) \xrightarrow{\text{FT}} e^{-j2\pi f t_0} S(f).$$

This follows from

$$\mathcal{F}\{s(t - t_0)\} = \int_{\mathbb{R}} s(t - t_0) e^{-j2\pi f t} dt = e^{-j2\pi f t_0} \int_{\mathbb{R}} s(t) e^{-j2\pi f t} dt = e^{-j2\pi f t_0} S(f).$$

Alternatively, we can directly consider the definition of the inverse Fourier transform

$$s(t - t_0) = \int_{\mathbb{R}} S(f) e^{j2\pi f(t-t_0)} df = \int_{\mathbb{R}} [S(f) e^{-j2\pi f t_0}] e^{j2\pi f t} df.$$

Obviously, a time shift does not change the value of

$|\mathcal{F}\{s(t)\}| = |\mathcal{F}\{s(t - t_0)\}| = |S(f)|$ , but just adds a linear phase term  $e^{-j2\pi f t_0}$ .

9. Frequency shift For  $S(f - f_0)$  with  $f_0$  representing a real constant, we have

$$S(f - f_0) \xrightarrow{\quad} e^{j2\pi f_0 t} s(t).$$

This follows from the definition of the Fourier transform

$$S(f - f_0) = \int_{\mathbb{R}} s(t) e^{-j2\pi(f-f_0)t} dt = \int_{\mathbb{R}} [s(t) e^{j2\pi f_0 t}] e^{-j2\pi f t} dt.$$

Similar to property 8, a frequency shift does not change the value of  $|s(t)| = |e^{j2\pi f_0 t} s(t)|$ . However, the signal  $s(t)$  is multiplied by  $e^{j2\pi f_0 t}$  which is important in modulation and demodulation processes. Here,  $f_0$  represents the so-called **carrier frequency**.

**Example:** modulation of a sine wave resulting in a modulated signal

$$s(t) \cos(2\pi f_0 t) \xrightarrow{\quad} \frac{1}{2}(S(f - f_0) + S(f + f_0)); \text{ note the shift of } S(f) \text{ by } \pm f_0.$$

10. Time scaling For a real constant  $a \neq 0$ , we have

$$s(at) \xrightarrow{\quad} \frac{1}{|a|} S\left(\frac{f}{a}\right).$$

This follows from

$$\begin{aligned} \int_{\mathbb{R}} s(at)e^{-j2\pi ft} dt &= \begin{cases} \frac{1}{a} \int_{\mathbb{R}} s(t)e^{-j2\pi \frac{f}{a}t} dt & \text{for } a > 0 \\ -\frac{1}{a} \int_{\mathbb{R}} s(t)e^{-j2\pi \frac{f}{a}t} dt & \text{for } a < 0. \end{cases} \\ &= \frac{1}{|a|} S\left(\frac{f}{a}\right). \end{aligned}$$

Thus, for a time compression  $|a| > 1$ , we obtain a frequency expansion, while for a time expansion  $|a| < 1$ , we have a frequency compression. This effect is a consequence of the **uncertainty principle** being inherent in the Fourier transform.

11. **Temporal derivative** Assuming  $s(t)$  to be continuous in  $\mathbb{R}$  and the Fourier transforms of  $s(t)$  and  $s'(t) = ds(t)/dt$  to exist, we have

$$s'(t) = \frac{ds(t)}{dt} \xrightarrow{\text{FT}} j2\pi f S(f).$$

This follows from the integration by parts

$$\mathcal{F}\{s'(t)\} = \int_{\mathbb{R}} s'(t) e^{-j2\pi f t} dt = s(t) e^{-j2\pi f t} \Big|_{-\infty}^{\infty} + j2\pi f \int_{\mathbb{R}} s(t) e^{-j2\pi f t} dt = j2\pi f S(f).$$

**Example:** A standard example is the voltage-current relation of a time-invariant linear inductance  $L$  defined in the time domain by

$$u(t) = L \frac{di(t)}{dt}.$$

Taking the Fourier transform of both sides yields with  $u(t) \xrightarrow{\text{FT}} U(f)$  and  $i(t) \xrightarrow{\text{FT}} I(f)$  the well-known relation  $U(f) = j2\pi f L I(f)$ .

12. **Temporal integration** Assuming  $s(t)$  and  $g(t) = \int_{-\infty}^t s(t)dt$  to have Fourier transforms, the integration  $g(t) = \int_{-\infty}^t s(t)dt$  can be written as a convolution between  $s(t)$  and the unit-step function

$$\begin{aligned} u(t) &= \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{2} & \text{for } t = 0 \\ 1 & \text{for } t > 0 \end{cases} \\ &= \frac{1}{2} + \frac{1}{2}\text{sign}(t), \end{aligned}$$

where  $\text{sign}(t)$  is the signum function. We obtain as a consequence

$$\int_{-\infty}^t s(t)dt \xrightarrow{\text{---}} \left( \frac{1}{2}\delta(f) + \frac{1}{j2\pi f} \right) S(f).$$

**Example:** A standard example is the voltage-current relation of a time-invariant linear capacitance  $C$  defined in the time domain by

$$u(t) = \frac{1}{C} \int_{-\infty}^t i(t)dt.$$

Taking again the Fourier transform of both sides yields the well-known relation  $U(f) = \frac{1}{j2\pi f C} I(f)$ , for DC component  $I(0)$  equals to zero.

13. **Convolution in time domain** Assume we are given two pairs of functions  $s_1(t) \circlearrowleft S_1(f)$  and  $s_2(t) \circlearrowleft S_2(f)$ . Consider the **convolution** which we mentioned already in the context of linear time-invariant systems:

$$s(t) = s_1(t) * s_2(t) = \int_{\mathbb{R}} s_1(\tau) s_2(t - \tau) d\tau = \int_{\mathbb{R}} s_1(t - \tau) s_2(\tau) d\tau = s_2(t) * s_1(t).$$

Note first, that the convolution is **commutative**. For the Fourier transforms, we obtain

$$\mathcal{F}\{s(t)\} = S(f) = S_1(f) \cdot S_2(f).$$

This follows from

$$\begin{aligned} S(f) &= \int_{\mathbb{R}} \int_{\mathbb{R}} s_1(\tau) s_2(t - \tau) d\tau e^{-j2\pi f t} dt = \int_{\mathbb{R}} \left[ s_1(\tau) \int_{\mathbb{R}} s_2(t - \tau) e^{-j2\pi f t} dt \right] d\tau \\ &= \int_{\mathbb{R}} s_1(\tau) e^{-j2\pi f \tau} S_2(f) d\tau = S_1(f) \cdot S_2(f). \end{aligned}$$

14. **Convolution in frequency domain** Assume we are given two pairs of functions  $s_1(t) \circ\bullet S_1(f)$  and  $s_2(t) \circ\bullet S_2(f)$ . Consider the convolution in the frequency domain defined by

$$S(f) = S_1(f) \star S_2(f) = \int_{\mathbb{R}} S_1(\nu) S_2(f - \nu) d\nu = \int_{\mathbb{R}} S_1(f - \nu) S_2(\nu) d\nu = S_2(f) \star S_1(f).$$

From property 12, we know that the convolution is commutative. For the inverse Fourier transforms, we obtain

$$s(t) = s_1(t) \cdot s_2(t) \circ\bullet S_1(f) \star S_2(f).$$

The derivation is completely analogous to property 12 and will not be repeated here.

**15. Parseval's equation** Rewriting property 13, we first have

$$\int_{\mathbb{R}} s_1(\tau) s_2(t - \tau) d\tau = \int_{\mathbb{R}} S_1(f) S_2(f) e^{j2\pi f t} df,$$

from which we can derive certain interesting special cases. Upon setting  $t = 0$ , we have

$$\int_{\mathbb{R}} s_1(t) s_2(-t) dt = \int_{\mathbb{R}} S_1(f) S_2(f) df.$$

From property 6, we know that  $s^*(-t) \longleftrightarrow S^*(f)$ . Replacing  $s_2(t)$  by  $s_2^*(-t)$  requires to replace  $S_2(f)$  by  $S_2^*(f)$ , so that we have

$$\int_{\mathbb{R}} s_1(t) s_2^*(t) dt = \int_{\mathbb{R}} S_1(f) S_2^*(f) df.$$

15. Parseval's equation (cont.) Finally, upon setting  $s(t) = s_1(t) = s_2(t)$  and thus  $S(f) = S_1(f) = S_2(f)$ , we obtain Parseval's equation

$$\int_{\mathbb{R}} |s(t)|^2 dt = \int_{\mathbb{R}} |S(f)|^2 df.$$

The integrals define the **energy of  $s(t)$** . Since the right hand side represents an integral of a non-negative function depending on the frequency  $f$ , we can interpret  $|S(f)|^2$  as a **spectrum** or, more precisely, as the **energy spectral density of  $s(t)$** .

**N.B.** The **energy spectral density of  $s(t)$**  defined above for a **deterministic** function  $s(t)$  should not be confused with the **power spectral density of a stationary stochastic process** to be defined below, where the considered function represents a **random** variable for any  $t$ .

**Example #1**

We want to calculate the Fourier transform of the **rectangular function**

$$s(t) = \text{rect}(t/T) = \begin{cases} 1 & \text{for } |t/T| < 1 \\ 1/2 & \text{for } |t/T| = 1 \\ 0 & \text{for } |t/T| > 1. \end{cases}$$

Upon insertion into the definition of  $S(f)$ , we obtain

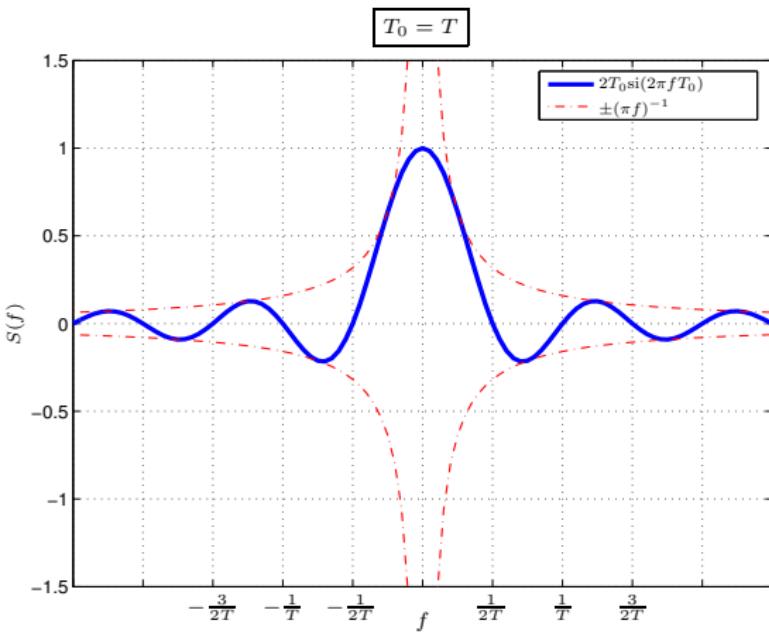
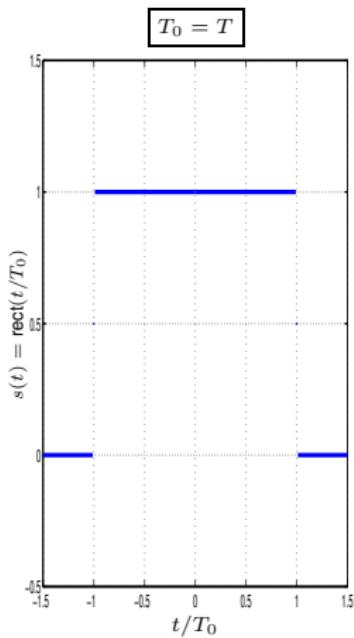
$$S(f) = \int_{-T}^T e^{-j2\pi f t} dt = \frac{e^{-j2\pi f T} - e^{j2\pi f T}}{-j2\pi f} = \frac{\sin(2\pi f T)}{\pi f} = 2T \text{si}(2\pi f T),$$

where we have used the function

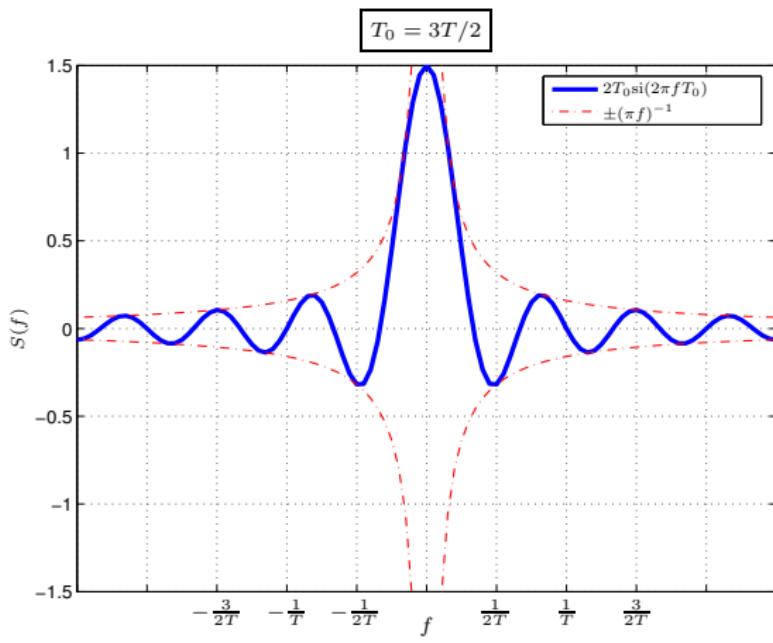
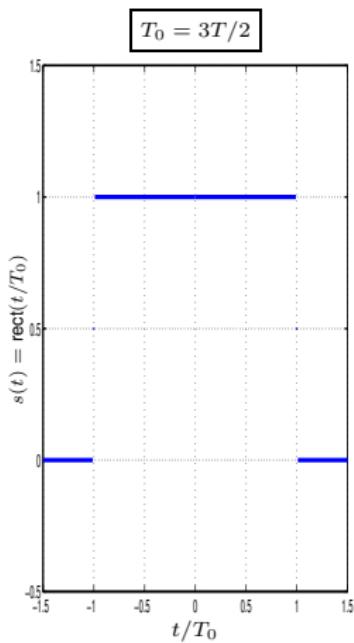
$$\text{si}(t) = \frac{\sin(t)}{t} \quad \text{for } t \neq 0, \quad \text{si}(0) = 1.$$

This function is a.k.a. **sinc-function** denoted as  $\text{sinc}(t)$ .

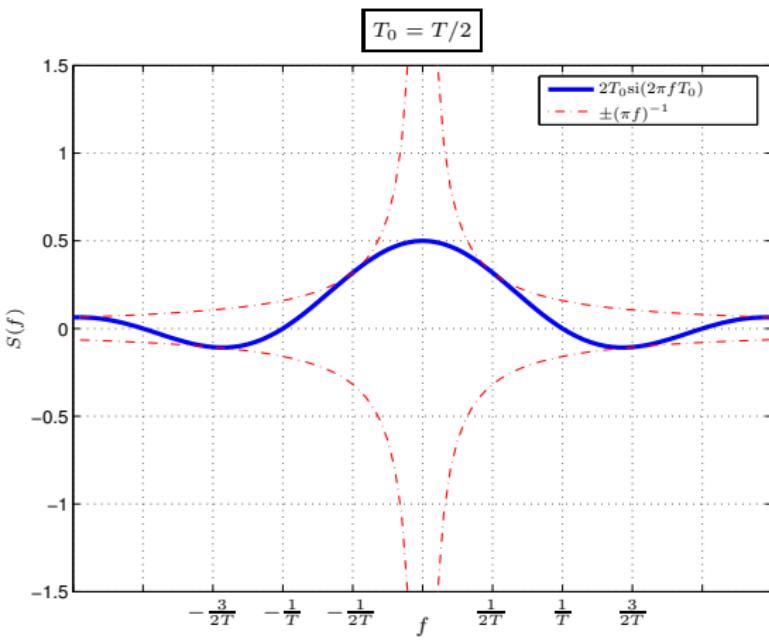
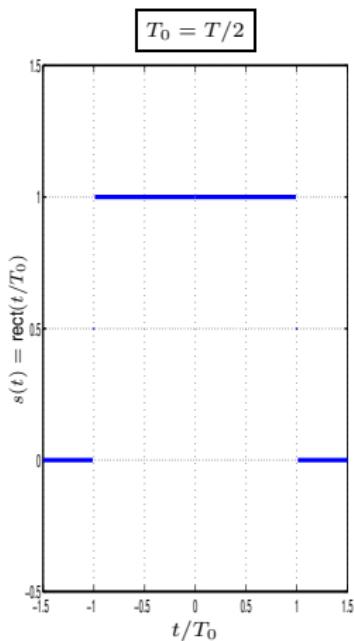
## Example #1 (cont.)



## Example #1 (cont.)



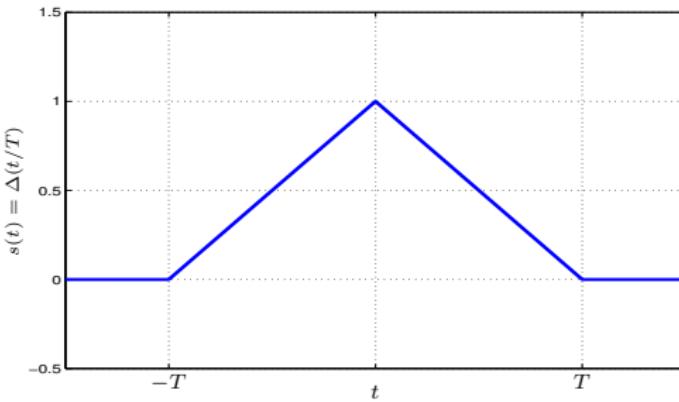
## Example #1 (cont.)



**Example #2**

We want to calculate the Fourier transform of the triangular function of width  $2T$

$$s(t) = \Delta(t/T) = \frac{1}{T} \text{rect}\left(\frac{2t}{T}\right) * \text{rect}\left(\frac{2t}{T}\right).$$



**Example #2 (cont.)**

From property 10 (time scaling), we have

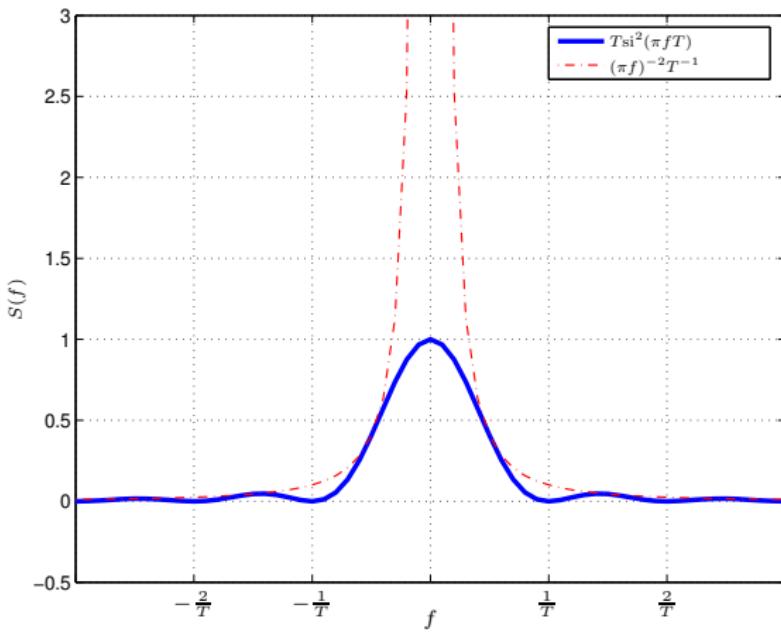
$$\text{rect}(2t/T) \circ\bullet \frac{1}{2}2T \text{si}\left(\frac{2\pi fT}{2}\right) = T \text{si}(\pi fT).$$

From property 13, we know that a convolution in the time domain leads to the product of the corresponding Fourier transforms. Since we convolve two identical functions, we immediately get

$$\Delta(t/T) \circ\bullet S(f) = \frac{1}{T} (T \text{si}(\pi fT))^2 = T \text{si}^2(\pi fT).$$

## Example #2 (cont.)

$$\Delta(t/T) \circ \bullet S(f) = \frac{1}{T} (\text{Ts}(i\pi f T))^2 = T \text{si}^2(\pi f T).$$



### Dirac's delta function

- Until now, we have mainly considered conventional functions. For example, we have claimed that (cf. frame #99)  $s(\infty) = s(-\infty) = 0$ .
- However, the class of conventional functions is often too restricted to be useful in many applications of communication systems. Therefore, we also consider so-called **generalized functions** which are defined as limits of sequences of functions (cf. math literature for further details).
- In particular, we will make use of the well-known **Dirac delta function** (or distribution) in order to simplify the derivation of otherwise complicated functions using formal rules for calculation.
- Here, we consider Dirac's delta function  $s(t) = \delta(t)$  and related Fourier transforms. We get

$$\mathcal{F}\{\delta(t - t_0)\} = \int_{\mathbb{R}} \delta(t - t_0) e^{-j2\pi f t} dt = e^{-j2\pi f t_0},$$

so that we have  $\delta(t - t_0) \xrightarrow{\text{FT}} e^{-j2\pi f t_0}$  and for  $t_0 = 0$ ,  $\delta(t) \xrightarrow{\text{FT}} 1$ .

### Dirac's delta function (cont.)

- Thus, if all frequencies provide the same amplitude  $|S(f)| = \text{const.}$ , the time-domain signal will be a Dirac delta impulse. Clearly, the signal energy in this case is infinite.
- Similarly, if we consider a Dirac delta impulse in the frequency domain at  $f = f_0$ , we have

$$\mathcal{F}^{-1}\{\delta(f - f_0)\} = \int_{\mathbb{R}} \delta(f - f_0) e^{j2\pi f t} df = e^{j2\pi f_0 t},$$

so that  $e^{j2\pi f_0 t} \circledast \delta(f - f_0)$  and for  $f_0 = 0$ , we have  $1 \circledast \delta(f)$ .

- To find the Fourier transform of trigonometric functions, we use Euler's equations:

$$\cos(2\pi f_0 t) = \frac{(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t})}{2}$$

$$\sin(2\pi f_0 t) = \frac{(e^{j2\pi f_0 t} - e^{-j2\pi f_0 t})}{2j}.$$

### Dirac's delta function (cont.)

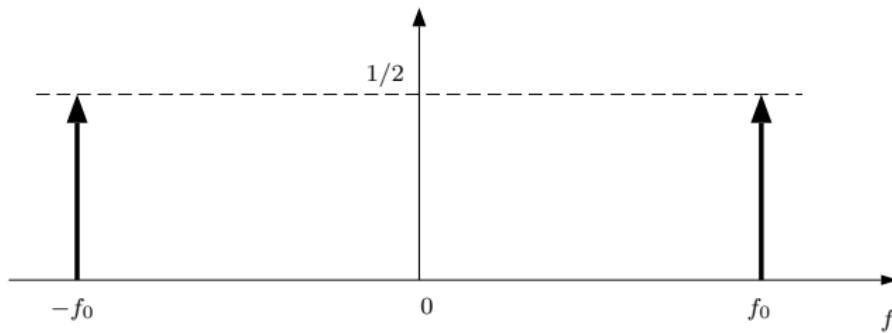
- As a consequence of Euler's equations, we have

$$\cos(2\pi f_0 t) \circ\bullet \frac{1}{2}(\delta(f - f_0) + \delta(f + f_0))$$

$$\sin(2\pi f_0 t) \circ\bullet \frac{1}{2j}(\delta(f - f_0) - \delta(f + f_0)) = \frac{j}{2}(\delta(f + f_0) - \delta(f - f_0)).$$

- Consider for example  $\mathcal{F}\{\cos(2\pi f_0 t)\}$ :

$$S(f) = \frac{1}{2}(\delta(f - f_0) + \delta(f + f_0))$$



## General periodic functions

- It is well-known that we can represent a periodic function  $s(t) = s(t + kT)$  by a **Fourier series**, where  $k \in \mathbb{Z}$  and  $T$  is the period of  $s(t)$ . Thus, we can write

$$s(t) = \sum_{n=-\infty}^{\infty} S_n e^{jn2\pi F t}$$

with  $F = 1/T$  denoting the frequency spacing of the individual oscillations contributing to  $s(t)$ .

- The coefficients in the Fourier series are defined by

$$S_n = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-jn2\pi F t} dt.$$

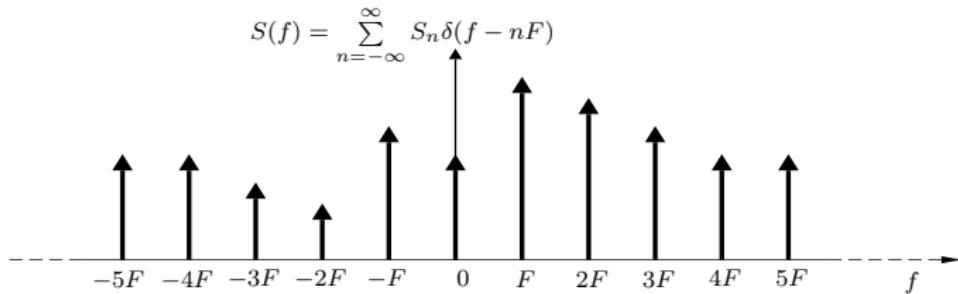
- Since  $e^{j2\pi f_0 t} \xrightarrow{\text{FT}} \delta(f - f_0)$ , we can calculate the **Fourier transform of a periodic  $s(t)$** .

### General periodic functions (cont.)

- We obtain

$$S(f) = \mathcal{F}\{s(t)\} = \sum_{n=-\infty}^{\infty} S_n \delta(f - nF).$$

- As a result, the spectrum of a periodic signal is non-zero only at  $f = nF$  and looks like a comb.



### General periodic functions (cont.)

- Of particular interest in communications is the case of the so-called **impulse train**

$$s(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT).$$

- The Fourier series coefficient of  $s(t)$  are given by

$$S_n = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-jn2\pi F t} dt = \frac{1}{T} \quad \forall n,$$

so that the Fourier series reads

$$s(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{jn2\pi F t}.$$

- As above, since  $e^{j2\pi f_0 t} \xrightarrow{\text{---}} \delta(f - f_0)$ , we can calculate  $\mathcal{F}\{s(t)\}$ .

### General periodic functions (cont.)

- We obtain

$$S(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nF).$$

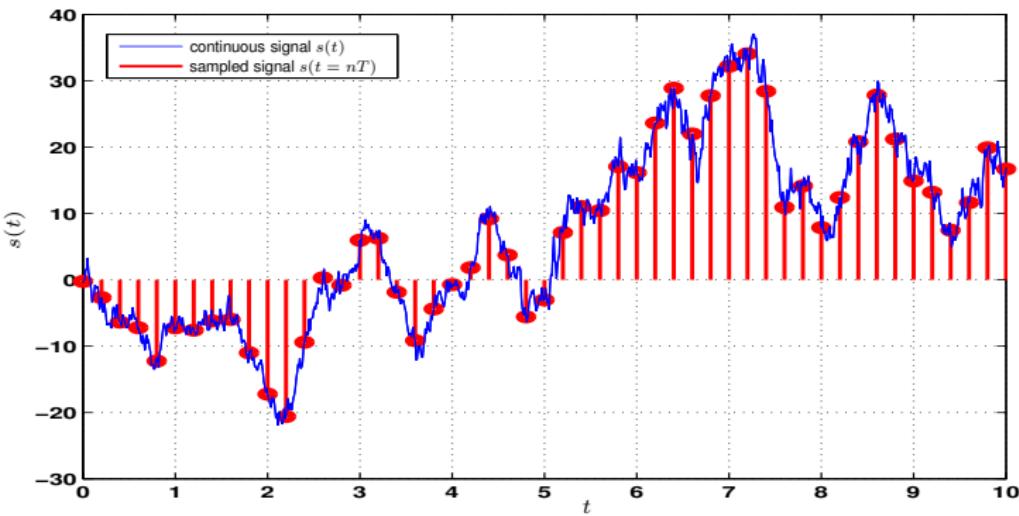
Thus, we finally have

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) \longleftrightarrow \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nF).$$

- As a result: a time-domain impulse train has a Fourier transform which is another impulse train in the frequency domain.
- Interpretation** Clearly, the periodic function  $s(t)$  has a comb spectrum  $S(f)$ . Since, however,  $s(t)$  is itself of comb type,  $S(f)$  has to be periodic. Thus, both  $s(t)$  and  $S(f)$  are periodic and of comb type, namely impulse trains, in accordance with our above derivation.

# Sampling Theorem

- In today's digital transmission systems, we are dealing with a signal representation in terms of **signal samples** rather than of **time-continuous signals**.
- The signal samples are taken either at the transmitter (prior to transmission of corresponding waveforms) or at the receiver (upon sampling the received signals) with a certain rate  $1/T$  where  $T$  is the time interval between subsequent samples.



- The sampling reduces the information about the signal  $s(t)$  to discrete points in time  $t = nT$  with samples  $s(nT)$ . Can we find a **spectral representation** of the signal samples in order to characterize the train of samples in the frequency domain?
- Before we treat the problem in greater detail, consider the following thought experiment. It is known that a periodic (in time domain) function  $s(t) = s(t + nT)$  for  $n \in \mathbb{Z}$  has a comb spectrum

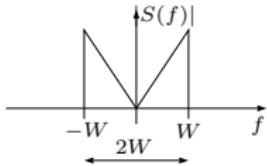
$$S(f) = \mathcal{F}\{s(t)\} = \sum_{n=-\infty}^{\infty} S_n \delta(f - nF) = \sum_{n=-\infty}^{\infty} S_n \delta\left(f - \frac{n}{T}\right).$$

If we represent the train of signal samples in the time domain in the form

$$u(t) = \sum_{n=-\infty}^{\infty} s(nT) \delta(t - nT),$$

is it true that we obtain a **periodic Fourier transform**?

- Consider a signal  $s(t)$  with FT  $S(f) = 0$  for  $|f| > W$ , where  $2W$  is the overall bandwidth occupied by  $s(t)$ .



- Due to the sifting property of the Dirac delta function, we have

$$u(t) = \sum_{n=-\infty}^{\infty} s(nT)\delta(t - nT) = s(t) \sum_{n=-\infty}^{\infty} \delta(t - nT).$$

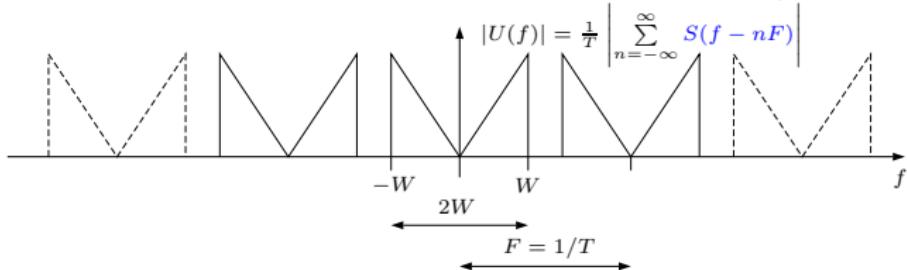
- Now we can use the result  $\sum_{k=-\infty}^{\infty} \delta(t - kT) \bullet \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nF)$  with  $F = 1/T$  and property 14 of the FT to calculate  $u(t) \bullet U(f)$  according to

$$\begin{aligned} U(f) &= \mathcal{F} \left\{ s(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) \right\} = S(f) * \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nF) \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} S(f - nF) \end{aligned} \quad \boxed{U(f) \text{ as a function of } S(f)}.$$

- Obviously,  $U(f) = U(f + kF)$  for  $k \in \mathbb{Z}$  is indeed a **periodic function** and we can interpret the time domain samples  $s(-nT)$  as the coefficients of a Fourier series of  $U(f)$ . This follows from

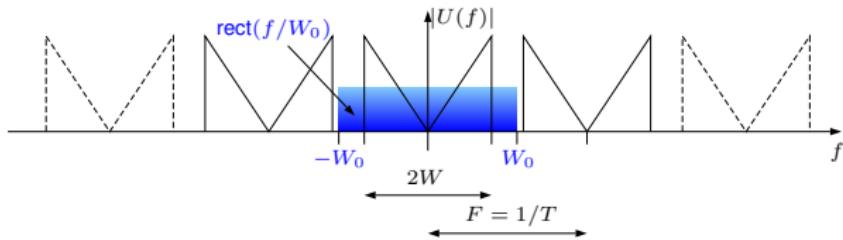
$$\begin{aligned} U(f) &= \mathcal{F} \left\{ \sum_{n=-\infty}^{\infty} s(nT) \delta(t - nT) \right\} = \sum_{n=-\infty}^{\infty} s(nT) \exp(-j2\pi nTf) \\ &= \sum_{n=-\infty}^{\infty} s(-nT) \exp(j2\pi nTf) \end{aligned} \quad \boxed{U(f) \text{ as a function of } s(nT)}.$$

- We obtain the following spectrum of the sampled signal for  $F = 1/T > 2W$ :



- Obviously, for  $F = 1/T > 2W$ , the individual spectral components  $S(f - nF)$  do not overlap. In this case, we can restore  $S(f)$  by multiplying  $U(f)$  by a rectangular function  $T_{\text{rect}}(f/W_0)$  with  $W < W_0 < 1/T - W$ :

$$\begin{aligned} S(f) &= TU(f)\text{rect}(f/W_0) \\ \Leftrightarrow s(t) &= \mathcal{F}^{-1}\{TU(f)\text{rect}(f/W_0)\}. \end{aligned}$$



The inverse FT provides

$$\begin{aligned}
 s(t) &= \mathcal{F}^{-1}\{TU(f)\text{rect}(f/W_0)\} = T \int_{-W_0}^{W_0} U(f)e^{j2\pi ft} df \\
 &= T \int_{-W_0}^{W_0} \sum_{n \in \mathbb{Z}} s(nT)e^{j2\pi f(t-nT)} df = T \sum_{n \in \mathbb{Z}} s(nT) \frac{e^{j2\pi f(t-nT)}}{j2\pi(t-nT)} \Big|_{-W_0}^{W_0} \\
 &= \sum_{n \in \mathbb{Z}} \frac{Ts(nT)}{\pi(t-nT)} \sin(2\pi W_0(t-nT)) = 2W_0T \sum_{n \in \mathbb{Z}} s(nT)\text{si}(2\pi W_0(t-nT)).
 \end{aligned}$$

- A simple choice for  $W_0$  for reconstruction of  $s(t)$  from the samples  $s(nT)$  is obviously

$$W_0 = \frac{1}{2T}.$$

- In this case, we find the so-called **sampling theorem for low-pass signals** (a.k.a. Nyquist theorem or Shannon-Kotelnikov theorem):

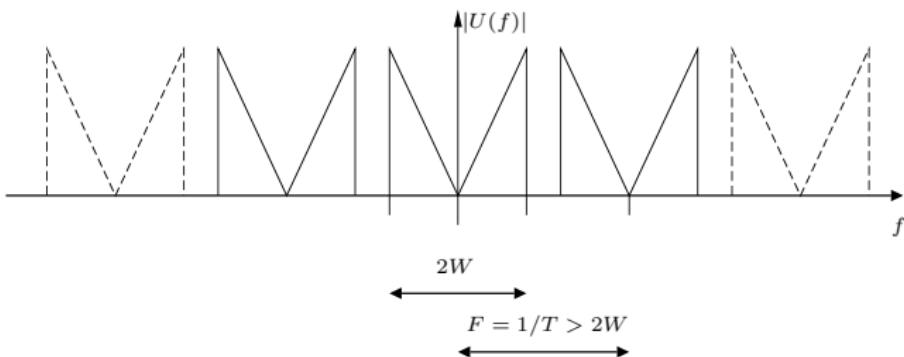
For a given bandlimited signal  $s(t)$  with a cut-off-frequency  $W$ , i.e.  $S(f) = 0$  for  $|f| > W$ , the signal is uniquely defined by samples taken at a frequency  $1/T$  which is larger than  $2W$ . The signal can be reconstructed from the samples by

$$s(t) = \sum_{n \in \mathbb{Z}} s(nT) \text{si}\left(\pi \frac{t - nT}{T}\right).$$

- If the signal does not contain singularities at the band edges, we can choose  $1/T = 2W$  and find

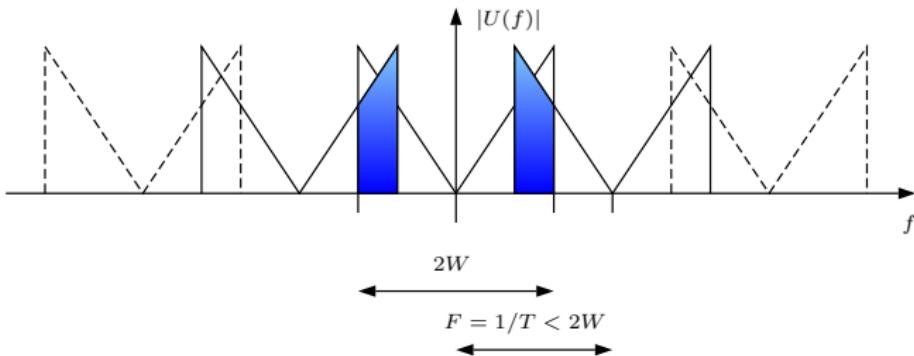
$$s(t) = \sum_{n \in \mathbb{Z}} s\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right).$$

When  $1/T > 2W$



there is no aliasing from adjacent frequency-shifted versions  $S(f - \frac{n}{T})$  within  $|f| \leq \frac{1}{2T}$  and the original signal can be uniquely represented by the samples taken at instances  $t = nT$  for  $n \in \mathbb{Z}$ .

What happens in the case  $1/T < 2W$ ?



The blue parts denote the so-called **aliasing**, where aliases are erroneously interpreted as being part of the original spectrum  $S(f)$  at the corresponding frequency.

# Table of Contents

## 1 Introduction

- Mathematical Models for Communication Channels
- Generalized Functions

## 2 Probability

- Fundamentals in Probability Theory
- Useful Probability Distributions
- Central Limit Theorem

## 3 Fourier Transform

- Properties of FT
- Sampling Theorem

## 4 Stochastic Processes

- Gaussian Processes
- Response of an LTI System to a Stationary Input Signal
- Discrete-Time (Stochastic) Signals
- Cyclostationary Processes

## 5 Communication Signals and Systems

- Representation of Band-pass Signals and Systems
- Orthogonal Expansions of Signals
- Representation of Digitally Modulated Signals

## 6 Optimum Receivers for the AWGN Channel

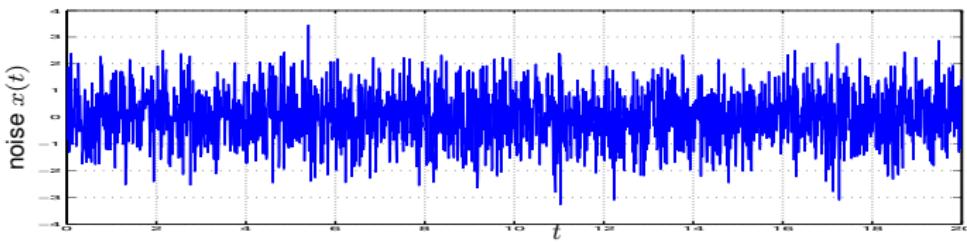
- Characterization of Thermal Noise
- Signal Space Representation
- Optimum Detection

# Stochastic Processes

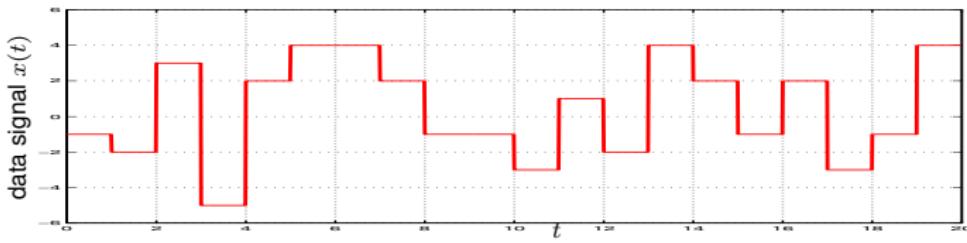
- In communication systems, we encounter **random functions of time**. Among these, we have
  - signals generated from information sources (e.g. audio signals from a CD, data signals for transmitting internet traffic, sensor signals)
  - transfer functions of communication channels (e.g. radio channel, fibre-optical channel)
  - noise generated in the front-end amplifier of a receiver (e.g. broadband thermal noise)
  - interference (e.g. intracell interference (multiple access, multiuser interference) and intercell interference in a UMTS cellular radio system)
- Random functions of time  $t$  are called **stochastic processes** and extend the notion of a **random variable** to random variables indexed by a parameter  $t$
- The stochastic process is denoted e.g. as  $X(t)$  where the parameter  $t$  is continuous, whereas  $X$  may be either continuous or discrete, depending on the characteristics of the source generating the process (e.g. continuous audio signal and discrete data signal)

- In the same way as a random variable  $Y$  takes values  $y$ , a stochastic process  $X(t)$  can take certain time functions  $x(t)$  called **sample functions** of the stochastic process. A measurement of a noise voltage generated by a single resistor can be interpreted as a sample function  $x(t)$  of the underlying  $X(t)$
- The set of all possible sample functions represents the **sample space of  $X(t)$** . In general, the number of sample functions in this ensemble is extremely large, often it is infinite as is the cardinality of most continuous random variables
- If we consider a set of time instants  $t_1 > t_2 > t_3 > \dots > t_n$  with positive integer  $n$ , the corresponding samples  $X(t_1), X(t_2), \dots, X(t_n)$  represent  $n$  random variables forming a **random vector** which is described by its **joint PDF**  $p(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ , where  $X_{t_i} \equiv X(t_i)$  for  $i = 1, \dots, n$  take values  $x_{t_1}$
- Here, we will use the short notation  $p(x_{t_1}, x_{t_2}, \dots, x_{t_n})$  to indicate  $p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_{t_1}, x_{t_2}, \dots, x_{t_n})$

- Continuous noise process  $X(t) = x(t)$  with  $X(t) \in \mathbb{R}$ :



- Discrete-valued data signal  $X(t) = x(t)$  with  $X(t) \in \{-5, -4, \dots, 3, 4\}$ :



- In many applications, we encounter stochastic processes whose properties are independent of  $t$  (e.g. the noise example on previous slide)
- The aforementioned time-independence of the properties of  $X(t)$  can be formulated as follows:

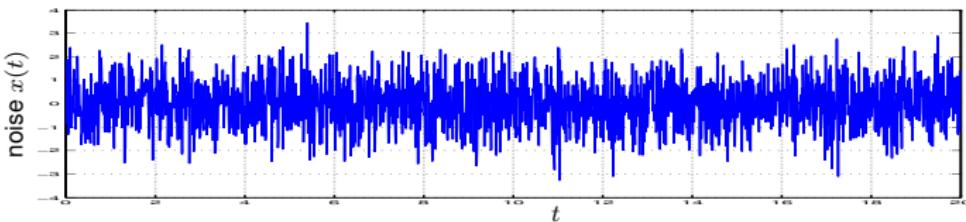
Consider  $n$  samples of the process  $X(t)$  defined by  $X_{t_i}$  for  $i = 1, \dots, n$ . Now shift the sampling instants by  $t$  to obtain another set of  $n$  samples of the process  $X(t)$  defined by  $X_{t_i+t} = X(t_i + t)$  for  $i = 1, \dots, n$ . If the joint PDFs of both random vectors are identical, i.e.

$$p(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = p(x_{t_1+t}, x_{t_2+t}, \dots, x_{t_n+t}) \quad \text{for all } t \text{ and } n ,$$

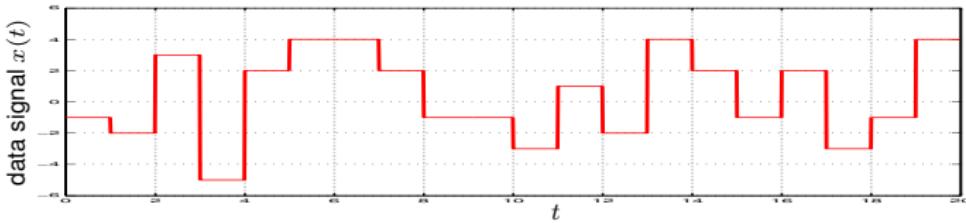
the stochastic process  $X(t)$  is said to be **stationary in the strict sense**

- The statistics of a stationary stochastic process are invariant to any translation in time, as we originally stated the properties of the corresponding processes (independence of  $t$ )

- Continuous noise process  $X(t) = x(t)$  with  $X(t) \in \mathbb{R}$ : can  $X(t)$  be a stationary process?



- Discrete-valued data signal  $X(t) = x(t)$  with  $X(t) \in \{-5, -4, \dots, 3, 4\}$  with  $X(t)$  being constant within  $t \in [n, n + 1)$ ,  $n \in \mathbb{Z}$  and independent data levels for consecutive  $n$ : can  $X(t)$  be a stationary process?



# Statistical Averages

## Moments of $X(t)$

- Just as we have defined **statistical averages for random variables**, we can define **statistical averages for stochastic processes** called **ensemble averages**
- Let  $X(t)$  define a stochastic process and let  $X_{t_i} \equiv X(t_i)$  as above. The  $n^{\text{th}}$  moment of  $X_{t_i}$  is defined as

$$E[X_{t_i}^n] = \int_{-\infty}^{\infty} x_{t_i}^n p(x_{t_i}) dx_{t_i}$$

- In general, the value of the  $n^{\text{th}}$  moment will depend on the time instant  $t_i$  if  $p(x_{t_i})$  depends on  $t_i$
- When the process is stationary, however,  $p(x_{t_i}) = p(x_{t_i+t})$  for all  $t$ . Hence, the PDF is independent of  $t$  so that the  $n^{\text{th}}$  moment is independent of time

### Joint second-order moments

- Next we consider the two random variables  $X_{t_i} \equiv X(t_i)$  for  $i = 1, 2$ .
- The correlation between  $X_{t_1}$  and  $X_{t_2}$  is measured by the joint (second-order) moment

$$E[X_{t_1} X_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} x_{t_2} p(x_{t_1}, x_{t_2}) dx_{t_1} dx_{t_2}.$$

- Since  $E[X_{t_1} X_{t_2}]$  depends, in general, on  $t_1$  and  $t_2$ , we denote it by  $\phi(t_1, t_2) = E[X_{t_1} X_{t_2}]$  and call it the **autocorrelation function of  $X(t)$** .
- When the process is stationary, the joint PDF of the pair  $(X_{t_1}, X_{t_2})$  is identical to the joint PDF of the pair  $(X_{t_1+t}, X_{t_2+t})$  for any arbitrary  $t$ . As a result, the autocorrelation function of  $X(t)$  does not depend on the specific time instants  $t_1$  and  $t_2$ , but depends rather on the **time difference  $\tau = t_1 - t_2$** .

**Joint second-order moments (cont.)**

- Thus, for a stationary process, the autocorrelation function is

$$E[X_{t_1}X_{t_2}] = \phi(t_1, t_2) = \phi(t_1 - t_2) = \phi(\tau),$$

where we have  $\tau = t_1 - t_2$  as above or, equivalently,  $t_2 = t_1 - \tau$

- Rewriting the equation above and exploiting the stationarity

$E[X_{t_1}X_{t_2}] = E[X_{t_1+\tau}X_{t_2+\tau}]$ , we obtain

$$\phi(\tau) = E[X_{t_1}X_{t_1-\tau}] = E[X_{t_1+\tau}X_{t_1}] = E[X_{t_1}X_{t_1+\tau}] = \phi(-\tau)$$

The autocorrelation function  $\phi(\tau)$  of a stationary process is an even function

- Choosing  $\tau = 0$ , we observe that

$\phi(0) = E[X_t^2]$  denotes the average power in the process  $X(t)$ .

Clearly, the average power as a second moment of  $X_t$  is independent of  $t$

### Joint second-order moments (cont.)

- There exist **nonstationary processes** where
  - still  $E[X_t]$  is independent of  $t$  (and thus constant) and
  - the autocorrelation function satisfies the condition  $\phi(t_1, t_2) = \phi(t_1 - t_2)$
- Such a process is called **wide-sense stationary**
- As a consequence, wide-sense stationarity is obviously a less stringent condition than strict-sense stationarity:
  - Any strict-sense stationary process is also wide-sense stationary
  - Wide-sense stationary processes are, in general, not strict-sense stationary. If, however, the statistics of  $X(t)$  are completely described by first-order and second-order moments and  $X(t)$  is wide-sense stationary, it is also strict-sense stationary. An important example is a Gaussian process to be discussed below
- In the following, **stationarity** is always to be understood in the **wide sense**

**Joint second-order moments (cont.)**

- Related to the **autocorrelation function** is the **autocovariance function** of a stochastic process defined as

$$\mu(t_1, t_2) = E[(X_{t_1} - m(t_1))(X_{t_2} - m(t_2))] = \phi(t_1, t_2) - m(t_1)m(t_2),$$

where  $m(t) = E[X_t]$ .

- When the process is (wide-sense) stationary, the autocovariance function simplifies to

$$\mu(t_1, t_2) = \mu(t_1 - t_2) = \mu(\tau) = \phi(\tau) - m^2,$$

where  $\tau = t_1 - t_2$  and  $m = E[X_t]$ .

- For a zero-mean stochastic process, i.e.  $E[X_t] \equiv 0$  for all  $t$ , the autocorrelation function equals its autocovariance function, i.e.

$$\phi(t_1, t_2) = \mu(t_1, t_2).$$

- Remark: joint second-order moments can be extended to **joint higher-order moments** in a straightforward manner

# Gaussian Processes

- Suppose that  $X(t)$  is a Gaussian random process, so that at time instants  $t = t_i$ ,  $i = 1, 2, \dots, n$ , the random variables  $X_{t_i}$ ,  $i = 1, 2, \dots, n$ , forming the random vector

$$\mathbf{X} = [X_{t_1}, X_{t_2}, \dots, X_{t_n}]^T$$

are jointly Gaussian distributed with **mean vector**

$$\mathbf{m}_\mathbf{X} = [E[X_{t_1}], E[X_{t_2}], \dots, E[X_{t_n}]]^T$$

and the **covariance matrix**

$$\mathbf{M} = E[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{X} - \mathbf{m}_\mathbf{X})^T] = \{\mu(t_i, t_j)\} \quad \text{for } i, j = 1, 2, \dots, n.$$

- The multivariate PDF of  $\mathbf{X}$  is given as (cf. slide #69)

$$p(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{M}}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{m}_\mathbf{X})^T \mathbf{M}^{-1} (\mathbf{x} - \mathbf{m}_\mathbf{X}) \right).$$

- If the Gaussian process is stationary, then  $E[X_t] = m$  and  $\mu(t_i, t_j) = \mu(t_i - t_j)$ . Thus, the mean vector has identical components
- If we have equidistant sampling with  $t_{i+1} - t_i = t_\Delta$  for all  $i$ , the covariance matrix  $\mathbf{M}$  has a Toeplitz structure, i.e. identical values on diagonals
- Gaussian stochastic processes are examples of processes depending only on first and second order moment functions. If a Gaussian distributed  $X(t)$  is thus wide-sense stationary, it is also strict-sense stationary
- Important examples of Gaussian stochastic processes are
  - **fading processes** modelling the impulse response of mobile radio channels
  - **broadband white Gaussian noise processes** to be discussed below
- The Gaussianity of the processes results, in most cases, from the central limit theorem. In fading channels, we have a superposition of independent multipath components and in thermal noise processes the additive superposition of independently moving particles

# Averages for Joint Stochastic Processes

- Let  $X(t)$  and  $Y(t)$  denote two stochastic processes and let  $X_{t_i} \equiv X(t_i)$ ,  $i = 1, 2, \dots, n$ , and  $Y_{t'_j} \equiv Y(t'_j)$ ,  $j = 1, 2, \dots, m$ , represent the random variables at times  $t_1 > t_2 > t_3 > \dots > t_n$  and  $t'_1 > t'_2 > t'_3 > \dots > t'_m$ , respectively
- The two processes are characterized statistically by their joint PDF

$$p(x_{t_1}, x_{t_2}, \dots, x_{t_n}, y_{t'_1}, y_{t'_2}, \dots, y_{t'_m})$$

for any set of time instants  $t_1, t_2, \dots, t_n, t'_1, t'_2, \dots, t'_m$  and for any positive integer values of  $n$  and  $m$

- The cross-correlation function of  $X(t)$  and  $Y(t)$ , denoted by  $\phi_{xy}(t_1, t_2)$  is defined as the joint moment

$$\phi_{xy}(t_1, t_2) = E[X_{t_1} Y_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} y_{t_2} p(x_{t_1}, y_{t_2}) dx_{t_1} dy_{t_2}$$

- The cross-covariance is

$$\mu_{xy}(t_1, t_2) = \phi_{xy}(t_1, t_2) - m_x(t_1)m_y(t_2)$$

- When the processes are jointly and individually stationary, we have

①  $\phi_{xy}(t_1, t_2) = \phi_{xy}(t_1 - t_2)$  and  $\mu_{xy}(t_1, t_2) = \mu_{xy}(t_1 - t_2)$

②  $\phi_{xy}(-\tau) = E[X_{t_1} Y_{t_1+\tau}] = E[X_{t_1-\tau} Y_{t_1}] = E[Y_{t_1} X_{t_1-\tau}] = \phi_{yx}(\tau)$

- The stochastic processes  $X(t)$  and  $Y(t)$  are said to be statistically independent if and only if

$$p(x_{t_1}, x_{t_2}, \dots, x_{t_n}, y'_{t_1}, y'_{t_2}, \dots, y'_{t_m}) = p(x_{t_1}, x_{t_2}, \dots, x_{t_n})p(y'_{t_1}, y'_{t_2}, \dots, y'_{t_m})$$

for all choices of  $t_i$  and  $t'_i$  and for all positive integers  $n$  and  $m$

- The processes are said to be uncorrelated if  $\phi_{xy}(t_1, t_2) = E[X_{t_1}]E[Y_{t_2}]$  and thus the cross-covariance is  $\mu_{xy}(t_1, t_2) = 0$

- A complex-valued stochastic process  $Z(t)$  is defined as

$$Z(t) = X(t) + jY(t),$$

where  $X(t)$  and  $Y(t)$  are stochastic processes with a joint PDF

- The joint PDF of the random variables  $Z_{t_i} \equiv Z(t_i)$  for  $i = 1, 2, \dots, n$  is given by the joint PDF of the components  $(X_{t_i}, Y_{t_i})$ ,  $i = 1, 2, \dots, n$ . Thus, the PDF characterizing  $Z_{t_i}$ ,  $i = 1, 2, \dots, n$ , is

$$p(x_{t_1}, x_{t_2}, \dots, x_{t_n}, y'_{t_1}, y'_{t_2}, \dots, y'_{t_m})$$

- The complex-valued stochastic process  $Z(t)$  is encountered in the representation of narrowband band-pass noise in terms of its equivalent low-pass components as will be discussed in the chapter on modulation techniques

- An important property of such a process is its **autocorrelation function** defined as

$$\begin{aligned}\phi_{zz}(t_1, t_2) &= \frac{1}{2} E[Z_{t_1} Z_{t_2}^*] \\ &= \frac{1}{2} E[(X_{t_1} + jY_{t_1})(X_{t_2} - jY_{t_2})] \\ &= \frac{1}{2} [\phi_{xx}(t_1, t_2) + \phi_{yy}(t_1, t_2) + j(\phi_{yx}(t_1, t_2) - \phi_{xy}(t_1, t_2))],\end{aligned}$$

where  $\phi_{xx}(t_1, t_2)$  and  $\phi_{yy}(t_1, t_2)$  are the autocorrelation functions of  $X(t)$  and  $Y(t)$ . Furthermore,  $\phi_{xy}(t_1, t_2)$  and  $\phi_{yx}(t_1, t_2)$  denote the cross-correlation functions .

- Note that the factor 1/2 in the definition of the autocorrelation function of a complex-valued stochastic process is an **arbitrary but mathematically convenient normalization factor** (cf. chapter on modulation techniques).

- When the processes  $X(t)$  and  $Y(t)$  are jointly and individually stationary, the autocorrelation function of  $Z(t)$  becomes

$$\phi_{zz}(t_1, t_2) = \phi_{zz}(t_1 - t_2) = \phi_{zz}(\tau),$$

where again  $t_2 = t_1 - \tau$ .

- Taking the complex conjugate on both sides, we have

$$\phi_{zz}^*(\tau) = \frac{1}{2} E[Z_{t_1}^* Z_{t_1-\tau}] = \frac{1}{2} E[Z_{t_1+\tau}^* Z_{t_1}] = \phi_{zz}(-\tau),$$

so that  $\boxed{\phi_{zz}(\tau) = \phi_{zz}^*(-\tau)}$ .

- Now suppose that  $Z(t) = X(t) + jY(t)$  and  $W(t) = U(t) + jV(t)$  are two complex-valued stochastic processes. The cross-correlation function of  $Z(t)$  and  $W(t)$  is defined as

$$\begin{aligned}\phi_{zw}(t_1, t_2) &= \frac{1}{2}E[Z_{t_1} W_{t_2}^*] \\ &= \frac{1}{2}E[(X_{t_1} + jY_{t_1})(U_{t_2} - jV_{t_2})] \\ &= \frac{1}{2}[\phi_{xu}(t_1, t_2) + \phi_{yv}(t_1, t_2) + j(\phi_{yu}(t_1, t_2) - \phi_{xv}(t_1, t_2))].\end{aligned}$$

- When  $X(t)$ ,  $Y(t)$ ,  $U(t)$  and  $V(t)$  are pairwise stationary, the above cross-correlation functions become functions of  $\tau = t_1 - t_2$  and

$$\phi_{zw}^*(\tau) = \frac{1}{2}E[Z_{t_1}^* W_{t_1-\tau}] = \frac{1}{2}E[Z_{t_1+\tau}^* W_{t_1}] = \phi_{wz}(-\tau).$$

# Response of an LTI System to a Stationary Input Signal

- Consider a **linear time-invariant (LTI) system** characterized by its impulse response  $h(t)$  or, equivalently, by its frequency response  $H(f)$  with

$$h(t) \circ\bullet H(f).$$

- The system output may be expressed in terms of the convolutional integral

$$y(t) = \int_{\mathbb{R}} h(\tau)x(t - \tau)d\tau.$$

- Suppose now that  $x(t)$  is a sample function of a **stationary stochastic process  $X(t)$** . We wish to determine **the mean** and **the autocorrelation function of the output  $y(t)$** .
- From the linearity of both the convolution and the expectation, we obtain

$$E[Y(t)] = \int_{\mathbb{R}} h(\tau)E[X_{t-\tau}]d\tau = m_X \int_{\mathbb{R}} h(\tau)d\tau = m_X H(0) = m_y = \text{const.}$$

- The autocorrelation function of  $y(t)$  is

$$\begin{aligned}
 \phi_{yy}(t_1, t_2) &= \frac{1}{2} E[Y_{t_1} Y_{t_2}^*] \\
 &= \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} h(\beta) h^*(\alpha) E[X(t_1 - \beta) X^*(t_2 - \alpha)] d\alpha d\beta \\
 &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(\beta) h^*(\alpha) \phi_{xx}(t_1 - t_2 + \alpha - \beta) d\alpha d\beta.
 \end{aligned}$$

- Obviously, the expectation of  $y(t)$  is constant and the autocorrelation function of  $y(t)$  is a function of  $\tau = t_1 - t_2$  so that the output process is also stationary. We obtain

$$\phi_{yy}(\tau) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(\beta) h^*(\alpha) \phi_{xx}(\tau + \alpha - \beta) d\alpha d\beta.$$

- Can we simplify the expression for  $\phi_{yy}(\tau)$ ?

- The expression

$$\phi_{yy}(\tau) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(\beta) h^*(\alpha) \phi_{xx}(\tau + \alpha - \beta) d\alpha d\beta.$$

represents a two-fold convolution of  $\phi_{xx}(\tau)$  with  $h(\tau)$  and  $h^*(-\tau)$ , i.e.

$$\phi_{yy}(\tau) = \phi_{xx}(\tau) \star h(\tau) \star h^*(-\tau).$$

- From property 13 and property 6 of the Fourier transform, we obtain with  $\phi_{yy}(\tau) \circ \bullet \Phi_{yy}(f)$  and  $\phi_{xx}(\tau) \circ \bullet \Phi_{xx}(f)$

$$\Phi_{yy}(f) = \Phi_{xx}(f) H(f) H^*(f) = \Phi_{xx}(f) |H(f)|^2.$$

- As a consequence, we have

$$\phi_{yy}(\tau) = \int_{\mathbb{R}} \Phi_{xx}(f) |H(f)|^2 e^{j2\pi f \tau} df.$$

- In particular, the average output power of the stationary process  $Y_t$  is given by

$$\phi_{yy}(0) = E[|Y_t|^2] = \int_{\mathbb{R}} \Phi_{xx}(f) |H(f)|^2 df \geq 0.$$

- Since we can choose  $H(f)$  to be extremely narrowband, we conclude that we must have  $\boxed{\Phi_{xx}(f) \geq 0}$  for all  $f \in \mathbb{R}$ .
- We conclude, in view of  $\phi_{xx}(\tau) \circlearrowleft \Phi_{xx}(f)$  and

$$\phi_{xx}(0) = E[|X_t|^2] = \int_{\mathbb{R}} \Phi_{xx}(f) df$$

The Fourier transform  $\Phi_{xx}(f) = \int_{\mathbb{R}} \phi_{xx}(\tau) e^{-j2\pi\tau f} d\tau \geq 0$  of the autocorrelation function  $\phi_{xx}(\tau)$  of a stationary stochastic process  $X(t)$  is the power density spectrum of  $X(t)$ . In particular,  $\phi_{xx}(0) = E[|X_t|^2] = \int_{\mathbb{R}} \Phi_{xx}(f) df$ .

- Example: Suppose that the low-pass filter illustrated in Fig. 2.2-1 is excited by a stochastic process  $X(t)$  having a power density spectrum

$$\Phi_{xx}(f) = \frac{1}{2}N_0 \quad \text{for all } f.$$

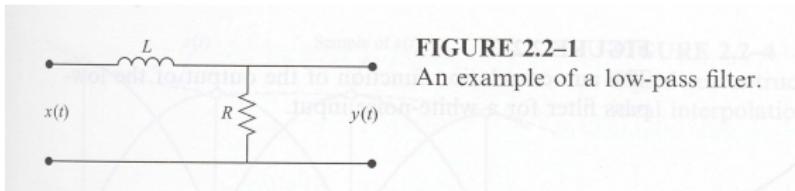


FIGURE 2.2-1

An example of a low-pass filter.

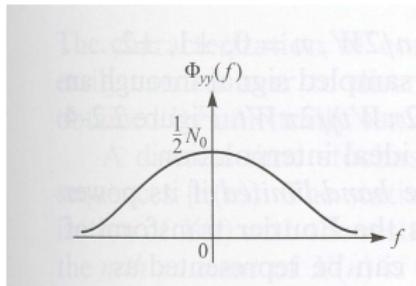
- The so-called **white-noise process  $X(t)$**  has infinite power due to the constant spectral power density  $N_0/2$  over an infinite bandwidth. Note that due to the limited bandwidth of the low-pass filter, we can expect the output signal power to be finite.

- The transfer function of the low-pass filter is

$$H(f) = \frac{R}{R + j2\pi f L} = \frac{1}{1 + j2\pi f L/R}.$$

- The spectral power density of the output process is

$$\Phi_{yy}(f) = \frac{N_0}{2} \frac{1}{1 + (2\pi L/R)^2 f^2}.$$

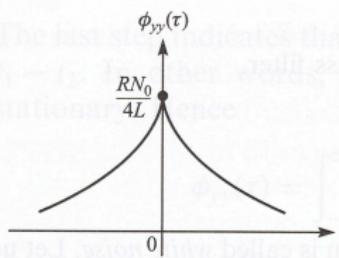


**FIGURE 2.2-2**

The power density spectrum of the low-pass filter output when the input is white noise.

- Using the inverse Fourier transform, we obtain

$$\phi_{yy}(\tau) = \int_{\mathbb{R}} \frac{N_0}{2} \frac{1}{1 + (2\pi f/L)^2} e^{j2\pi f\tau} df = \frac{RN_0}{4L} e^{-(R/L)|\tau|}.$$



**FIGURE 2.2-3**

The autocorrelation function of the output of the low-pass filter for a white-noise input.

- In particular, we obtain as expected

$$\phi_{yy}(0) = E[|Y_t|^2] = \frac{RN_0}{4L} < \infty.$$

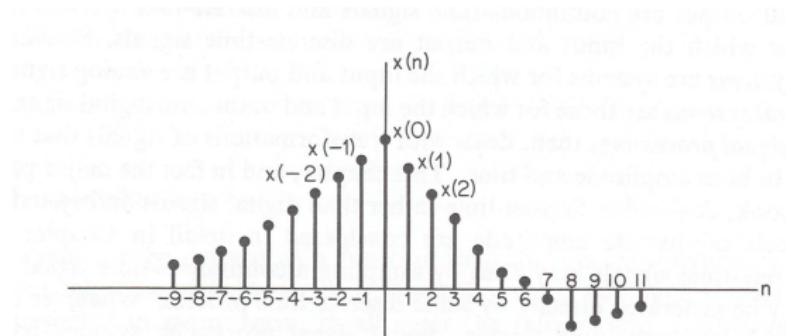
- Show as an exercise that for a stationary input signal  $X(t)$ , the cross-correlation function between  $Y(t)$  and  $X(t)$  with  $\tau = t_1 - t_2$  is given by

$$\phi_{yx}(t_1, t_2) = \phi_{yx}(\tau) = \frac{1}{2} E[Y_{t_1} X_{t_2}^*] = \phi_{xx}(\tau) \star h(\tau) \circ \bullet \Phi_{yx}(f) = \Phi_{xx}(f) H(f).$$

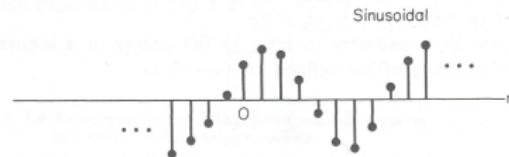
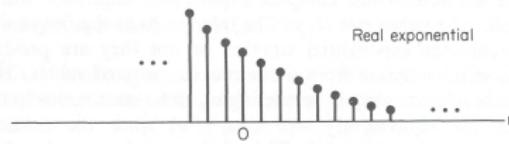
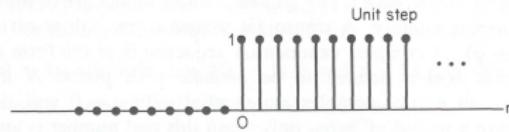
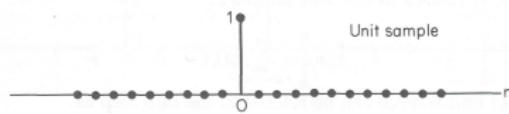
- As a result, we can devise an approach for measuring  $h(\tau)$ :
  - In the ideal case, use a white noise process as input signal and obtain a measurement being proportional to  $h(\tau)$  by the cross-correlation between input and output signals.
  - In a practical system, a sufficiently broadband random pulse can be used as input signal. The cross-correlation of the received signal at the channel output and the broadband pulse is a scaled version of  $h(\tau)$

# Discrete-Time (Stochastic) Signals

- In many applications, we are confronted with **purely discrete-time signal processing** instead of the previously considered **continuous-time signal processing**
- In discrete-time signal processing, we are dealing with **sequences depending on an integer index  $n$**  rather than **functions depending on time  $t$**



## Examples of discrete-time signals:



- Examples for discrete-time signal processing in communication systems include
  - pulse shaping of signals prior to transmission
  - filtering of signal samples obtained at a receiver
- In the following, we will consider basic properties of **discrete-time signals and systems  $x(n)$**  with  $n$  representing the discrete-time index. These properties correspond to those we considered for the **continuous-time case**, i.e. for functions  $x(t)$  of time  $t$ , namely
  - representation of discrete-time signals
  - transmission over linear time-invariant channels
  - Fourier transform (frequency domain representation)
  - sampling theorem for band-limited stochastic processes
  - discrete-time stochastic signals

## Representation of discrete-time signals

- A discrete-time signal is often represented as a sequence of samples

$$\{x(n)\} \quad \text{with } -\infty < n < \infty.$$

- Basic to the representation of a discrete-time signal  $x(n)$  is the Kronecker Delta corresponding to Dirac's Delta function in the continuous-time case:

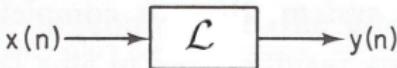
$$\delta(n) = \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{for } n \neq 0. \end{cases}$$

- Similar to the continuous-time case, we can use the sifting property of  $\delta(n)$  according to

$$x(n) = \sum_{n \in \mathbb{Z}} x(k)\delta(n - k).$$

- The functions  $\{\delta(n - k), k \in \mathbb{Z}\}$  represent a **basis** in  $\ell^2$  (the inner product vector space of square summable sequences) and  $x(k)$  are the **components** of  $x(n)$  in that basis. Thus, the representation above can be used to separate the information contained in  $x(n)$  into **time-dependent unit vectors**  $\delta(n - k)$  and **coefficients** being constant w.r.t.  $t$ .

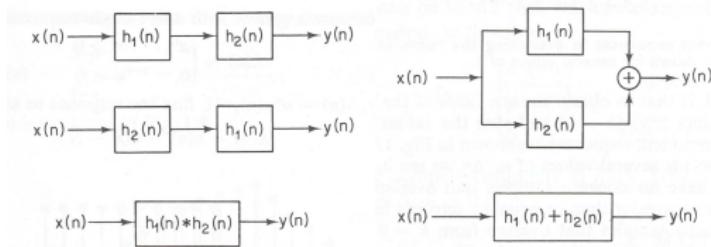
## Transmission of discrete-time signals over linear time-invariant channels



- Consider a **linear time-invariant mapping  $\mathcal{L}$**  as in the continuous-time case (cf. slide #16) with input signal  $x(n)$  and output signal  $y(n) = \mathcal{L}x(n)$ . Let  **$h(n)$  be the (impulse) response of the channel** for an input signal  $\delta(n)$ . Then, we have the **convolution sum**  $y(n) = x(n) \star h(n)$  defined by

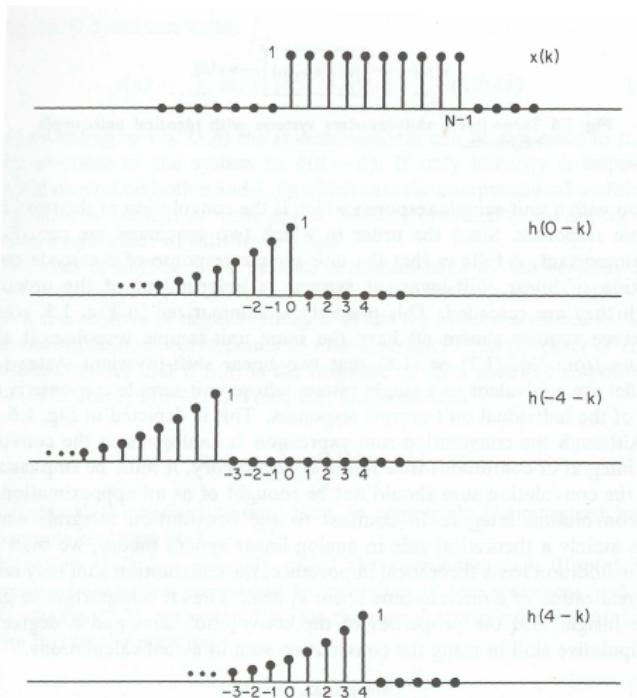
$$y(n) = \mathcal{L}x(n) = \mathcal{L} \sum_{n \in \mathbb{Z}} x(k)\delta(n - k) = \sum_{n \in \mathbb{Z}} x(k)\mathcal{L}\delta(n - k) = \sum_{n \in \mathbb{Z}} x(k)h(n - k).$$

- Equivalent systems with identical impulse responses:  
concatenation (left) and parallel channels (right)



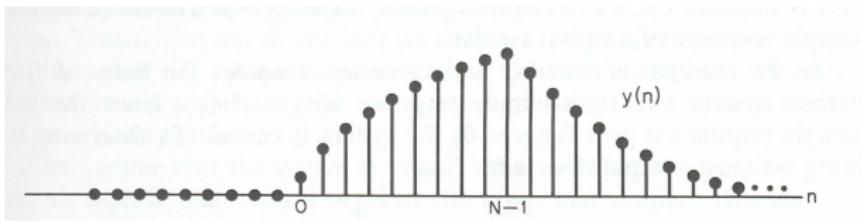
## Transmission of discrete-time signals over linear time-invariant channels

Example of a convolution:



## Transmission of discrete-time signals over linear time-invariant channels

Example of a convolution (cont.):



## Fourier transform (frequency domain representation)

- The Fourier transform in the case of a discrete-time signal  $x(n)$  is simply the Fourier series with  $x(n)$  representing the series coefficients. Thus, we have the **periodic Fourier transform**

$$X(f) = X(f + k) = \sum_{n \in \mathbb{Z}} x(n) \exp(-j2\pi f n), \quad k \in \mathbb{Z},$$

and the **Fourier inverse**

$$x(n) = \int_{-1/2}^{1/2} X(f) \exp(j2\pi f n) df.$$

- With complete analogy to the continuous-time case, we obtain for the Fourier transform of a convolution sum in the frequency domain

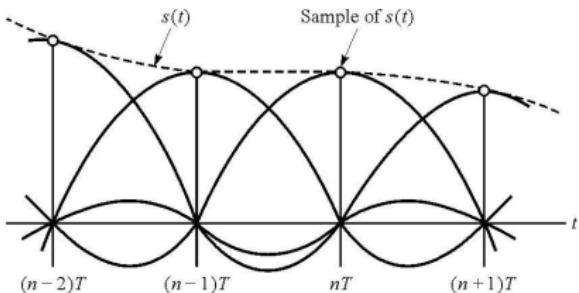
$$y(n) = h(n) \star x(n) \circ \bullet Y(f) = H(f)X(f).$$

## Sampling theorem for band-limited stochastic processes

- Recall that a band-limited **deterministic signal  $s(t)$**  with Fourier transform  $S(f) = 0$  for  $|f| > W$  sampled at the **Nyquist rate  $f_N = 2W$  samples/s** can be reconstructed from its samples by the interpolation formula

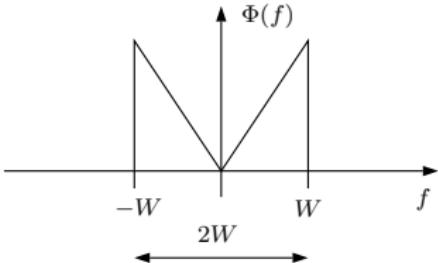
$$s(t) = \sum_{n \in \mathbb{Z}} s\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right).$$

- Equivalently, the original signal  $s(t)$  can be reconstructed by passing the impulse train of sampled signals through an ideal low-pass filter with impulse response  $h(t) = \text{si}(2\pi Wt)$ .



### Sampling theorem for band-limited stochastic processes (cont.)

- In view of the application cases of the stochastic processes discussed before, an important question arises: Can we **extend the sampling theorem to stochastic processes** and, if so, how and to **what class of signals?**
- To make use of band-limitation in the derivation of the corresponding theorem, we consider a stationary stochastic process  $X(t)$  whose **power density spectrum (PDS)  $\Phi(f)$**  satisfies  $\Phi(f) = 0$  for  $|f| > W$ . Such a process will be called **band-limited**



## Sampling theorem for band-limited stochastic processes (cont.)

- As we know,  $\Phi(f)$  is the power density spectrum of  $X(t)$  and we have  $\Phi(f) \bullet\!\!-\!\!\circ \phi(\tau)$ . Thus, we can apply the sampling theorem to  $\phi(\tau)$  and obtain

$$\phi(\tau) = \sum_{n \in \mathbb{Z}} \phi\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(\tau - \frac{n}{2W}\right)\right).$$

- Sampling theorem** ( $\stackrel{\text{i.q.m.}}{=} \dots$  equality in the quadratic mean)

A band-limited stationary stochastic process  $X(t)$  can be represented by

$$X(t) \stackrel{\text{i.q.m.}}{=} \sum_{n \in \mathbb{Z}} X\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right),$$

which means that

$$E \left[ \left| X(t) - \sum_{n \in \mathbb{Z}} X\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right) \right|^2 \right] = 0.$$

## Sampling theorem for band-limited stochastic processes (cont.)

**Proof:**

Here, we consider only real-valued processes, while the proof for complex-valued processes is analogous. Since  $\Phi(f) = 0$  for  $|f| > W$ , the function

$$\tilde{\Phi}(f) = \Phi(f) \exp(-j2\pi fa) \xrightarrow{\text{---o}} \tilde{\phi}(\tau) = \phi(\tau - a)$$

with  $a \in \mathbb{R}$  is also band-limited. Applying the Nyquist theorem to  $\tilde{\phi}(\tau)$ , we obtain

$$\phi(\tau - a) = \sum_{n \in \mathbb{Z}} \phi\left(\frac{n}{2W} - a\right) \text{si}\left(2\pi W\left(\tau - \frac{n}{2W}\right)\right).$$

Let us define

$$\hat{X}(t) = \sum_{n \in \mathbb{Z}} X\left(\frac{n}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right).$$

We have to show that

$$E \left[ \left( X(t) - \hat{X}(t) \right) \left( X(t) - \sum_{m \in \mathbb{Z}} X\left(\frac{m}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{m}{2W}\right)\right) \right) \right] = 0.$$

## Sampling theorem for band-limited stochastic processes (cont.)

First, we have

$$E\left[\left(X(t) - \hat{X}(t)\right)X\left(\frac{m}{2W}\right)\right] = \phi\left(t - \frac{m}{2W}\right) - \sum_{n \in \mathbb{Z}} \phi\left(\frac{n-m}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right).$$

The r.h.s. is zero if we choose  $a = m/2W$  above. Since this is true for any  $m$ , we find that

$$E\left[\left(X(t) - \hat{X}(t)\right) \sum_{m \in \mathbb{Z}} X\left(\frac{m}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{m}{2W}\right)\right)\right] = 0.$$

Similarly, we find

$$E\left[\left(X(t) - \hat{X}(t)\right)X(t)\right] = \phi(0) - \sum_{n \in \mathbb{Z}} \phi\left(\frac{n}{2W} - t\right) \text{si}\left(2\pi W\left(t - \frac{n}{2W}\right)\right).$$

Choosing  $a = \tau = t$ , the r.h.s. is again zero. Hence, we have

$$\begin{aligned} E\left[\left|X(t) - \hat{X}(t)\right|^2\right] &= E\left[\left(X(t) - \hat{X}(t)\right) \left(X(t) - \sum_{m \in \mathbb{Z}} X\left(\frac{m}{2W}\right) \text{si}\left(2\pi W\left(t - \frac{m}{2W}\right)\right)\right)\right] \\ &= 0 - 0 = 0. \end{aligned}$$

- The characterization of continuous-time stochastic signals given above can be easily carried over to discrete-time stochastic signals
- A discrete-time stochastic process  $X(n)$  consists of an ensemble of sample sequences  $\{x(n)\}$ . The characterization of  $X_n = X(n)$  is completely analogous to  $X(t)$  except that now  $n$  is an integer (time) variable
- The  $m^{\text{th}}$  moment of  $X(n)$  is defined by

$$E[X_n^m] = \int_{\mathbb{R}} x_n^m p(x_n) dx_n$$

and the autocorrelation sequence is

$$\phi(n, k) = \frac{1}{2} E[X_n X_k^*] = \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} x_n x_k^* p(x_n, x_k) dx_n dx_k.$$

- Similarly, the **autocovariance sequence** is  $\mu(n, k) = \phi(n, k) - \frac{1}{2}E[X_n]E[X_k^*]$ .
- For a **stationary process**, we have

$$\phi(n, k) = \phi(n - k) \quad \text{and}$$

$$\mu(n, k) = \mu(n - k) = \phi(n - k) - \frac{1}{2}|m_x|^2$$

with  $m_x = E[X(n)]$ .

- For a **stationary process** as input signal to an LTI system with impulse response  $h(t)$  and  $y(n) = h(n) \star x(n)$ , we further have

$$E[y(n)] = \sum_{k \in \mathbb{Z}} h(k)E[x(n - k)] = m_x \sum_{k \in \mathbb{Z}} h(k) = m_x H(0) = \text{const.}$$

- The autocorrelation sequence for the output process is given by

$$\begin{aligned}\phi_{yy}(k) &= \frac{1}{2} E[y^*(n)y(n+k)] \\ &= \frac{1}{2} \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} h^*(i)h(j) E[x^*(n-i)x(n+k-j)] \\ &= \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} h^*(i)h(j) \phi_{xx}(k-j+i). \\ &= \phi(k) \star h(k) \star h^*(-k).\end{aligned}$$

- Applying the rules for the Fourier transform being similar to the ones in the continuous-time case, we obtain

$$\Phi_{yy}(f) = \Phi_{xx}(f) |H(f)|^2,$$

which is formally identical to the result in the continuous-time case. However, note that the PSDs  $\Phi_{yy}(f)$  and  $\Phi_{xx}(f)$  as well as the frequency response  $H(f)$  are periodic in  $f$  with period 1

# Cyclostationary Processes

- In dealing with signals carrying digital information, we encounter stochastic processes that have periodic statistical averages.
- Consider

$$X(t) = \sum_{n \in \mathbb{Z}} a_n g(t - nT),$$

where  $\{a_n\}$  is a discrete-time sequence of random variables with mean  $m_a = E[a_n] \forall n$  and autocorrelation sequence  $\phi_{aa}(k) = \frac{1}{2}E[a_n * a_{n+k}]$ .

- The function  $g(t)$  is a deterministic pulse used to transmit the symbols  $a_n$  at a rate of  $1/T$ .
- We have

$$E[X(t)] = \sum_{n \in \mathbb{Z}} E[a_n]g(t - nT) = m_a \sum_{n \in \mathbb{Z}} g(t - nT) = E[X(t - kT)] \quad \text{for } k \in \mathbb{Z}.$$

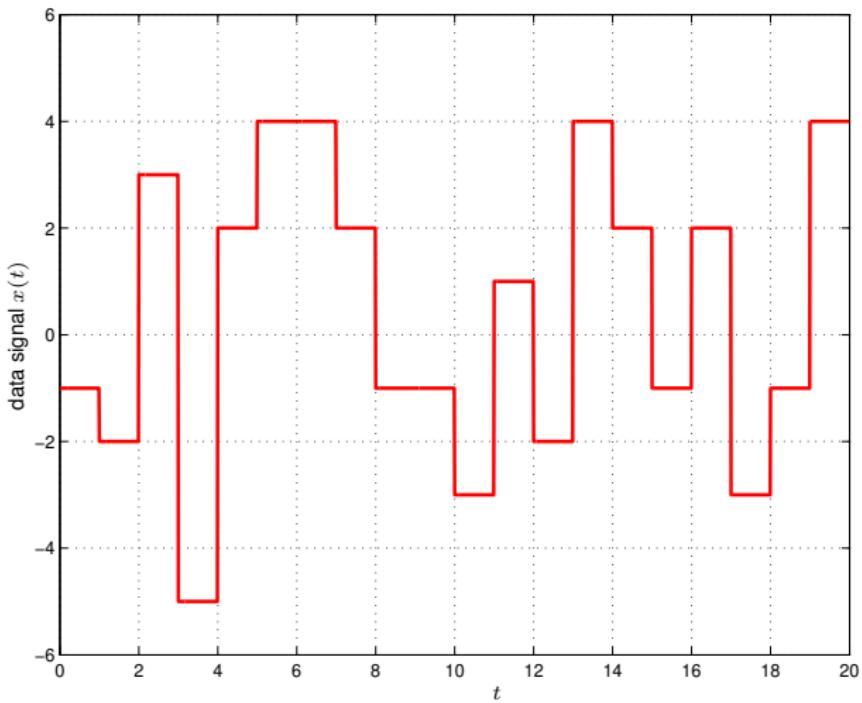
⇒ The mean is periodic in  $t$ .

- The autocorrelation function is given by

$$\begin{aligned}\phi_{xx}(t + \tau, t) &= \frac{1}{2} E[X(t + \tau) X^*(t)] \\ &= \frac{1}{2} \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} E[a_n^* a_m] g^*(t - nT) g(t + \tau - mT) \\ &= \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \phi_{aa}(m - n) g^*(t - nT) g(t + \tau - mT) \\ &= \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \phi_{aa}(m - n) g^*(t - (n - k)T) g(t + \tau - (m - k)T) \\ &= \phi_{xx}(t + kT + \tau, t + kT).\end{aligned}$$

⇒ The autocorrelation function is periodic in  $t$ .

- For  $k \in \mathbb{Z}$ , a stochastic process with both  $E[X(t)] = E[X(t + kT)]$  and  $\phi_{xx}(t + \tau, t) = \phi_{xx}(t + kT + \tau, t + kT)$  is called cyclostationary or periodically stationary.

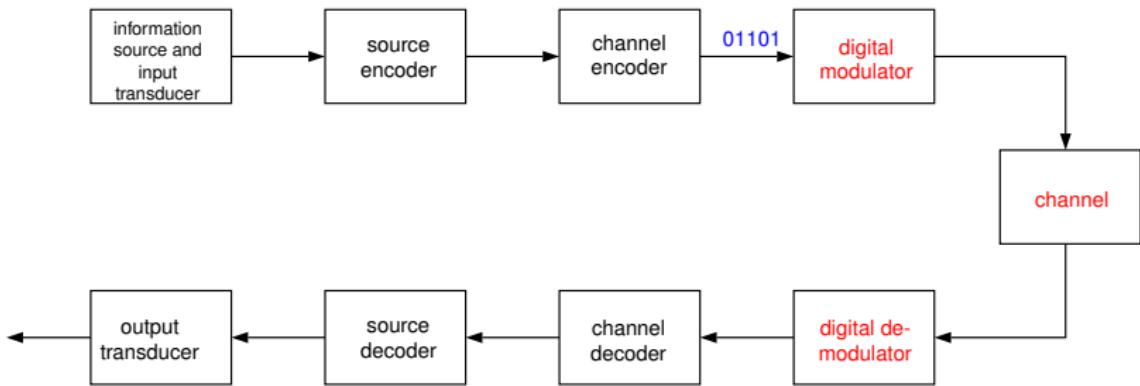


# Table of Contents

- 1 Introduction
  - Mathematical Models for Communication Channels
  - Generalized Functions
- 2 Probability
  - Fundamentals in Probability Theory
  - Useful Probability Distributions
  - Central Limit Theorem
- 3 Fourier Transform
  - Properties of FT
  - Sampling Theorem
- 4 Stochastic Processes
  - Gaussian Processes
  - Response of an LTI System to a Stationary Input Signal
  - Discrete-Time (Stochastic) Signals
  - Cyclostationary Processes
- 5 Communication Signals and Systems
  - Representation of Band-pass Signals and Systems
  - Orthogonal Expansions of Signals
  - Representation of Digitally Modulated Signals
- 6 Optimum Receivers for the AWGN Channel
  - Characterization of Thermal Noise
  - Signal Space Representation
  - Optimum Detection

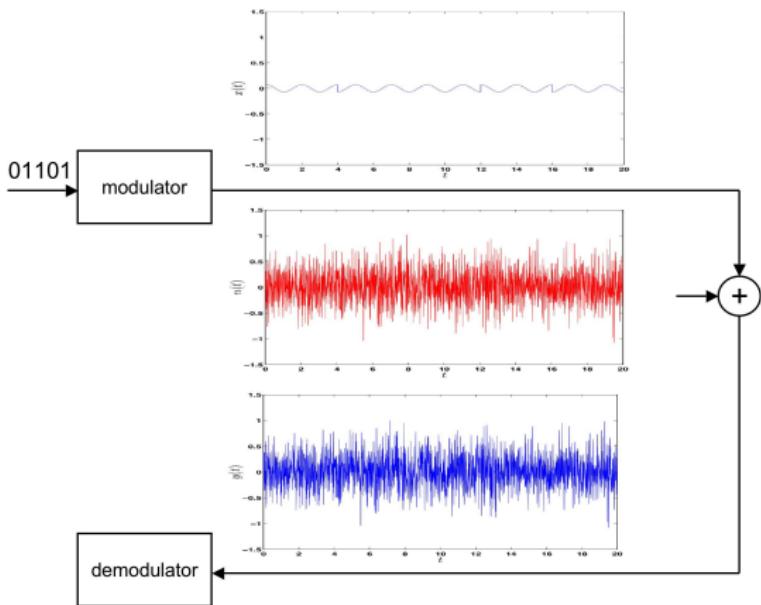
# Communication Signals and Systems

Reconsider the generic model of a digital communication system:



- Below, we will characterize the communication signals and the transmission channel representing the **inner part** of the communication system
- The task is to transmit the sequence of digital symbols  $\dots 01101 \dots$  at the channel encoder output over the channel
- The encoding and decoding modules will be discussed later on

Example: Transmission of a **binary phase-shift keying (BPSK)** signal over an **additive white Gaussian noise (AWGN)** channel



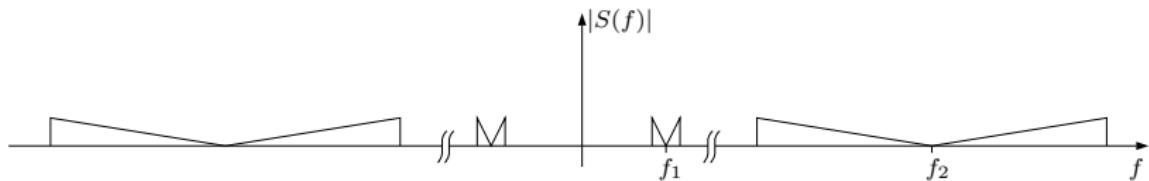
A number of questions come up in the characterization of signals and systems:

- How do we map the digital sequence of zeros and ones to signal waveforms? ⇒ modulation
- Which kind of signal waveforms do we consider?
- What are the properties of the resulting signals?
- How can we describe the transmission channel?
- What are suitable signals for a given transmission channel? ⇒ robustness and implementation effort
- How can we characterize noise and/or interference from other information sources?
- How can we demap the received signal to restore the transmitted digital sequence?
- How do we characterize the performance of the overall transmission?
- What are the assumptions on the transceiver and the channel to carry out the transmission in the aforementioned framework?

# Representation of Band-pass Signals and Systems

- For the representation of transmitted signals, we will consider exclusively real-valued signals, i.e.  $s(t) \in \mathbb{R}$  for  $t \in \mathbb{R}$ , where  $t$  represents time and  $s(t)$  can be e.g. a voltage or a current signal defined in the front-end of a transmitter.
- In communication systems, the transmitter and the receiver involved in a transmission are often called **peers** and we have different ways of communication among the peers (e.g. broadcast or peer-to-peer transmission).
- In general, the available Fourier spectrum will be used simultaneously by different groups of peers and care must be taken in order to avoid interference between the different links.
- A common way of sharing the Fourier spectrum among a group of transceivers is to split up the available spectrum into non-overlapping intervals (called **frequency bands**) and assign a band to a group of peers for information exchange within this band.

- From the uncertainty principle being inherent in the Fourier transform, we know that high-rate data signals will cover a correspondingly large frequency interval (**broadband transmission**), while low-rate data signals are essentially non-zero in a narrow frequency interval (**narrowband transmission**).
- As a result, broadband (resp. narrowband) signals are typically transmitted in higher (resp. lower) frequency bands located around a center frequency  $f_2$  (resp.  $f_1$ ).



- Clearly, a combination of broadband and narrowband transmissions is possible in higher frequency bands, but usually not in lower frequency bands.

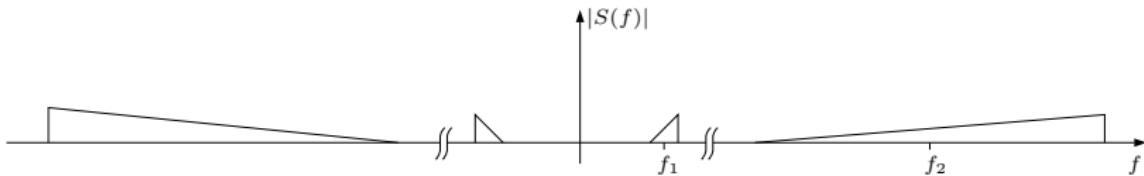
- Signals  $s(t)$  with a frequency spectrum contained in a certain frequency band are called **band-pass signals**. This expression is due to the interpretation of such a signal as the output of a band-pass filter being fed with a broadband signal having non-zero spectral content beyond the pass-band of the filter.
- Except for the general need to split up the frequency domain into non-overlapping bands, there are also **technical requirements** arising from the transmission over different physical media (i.e. the transmission channel) which suggest to consider the **transmission of a signal spectrum around a certain center (or carrier) frequency  $f_c$** .
- **Wireless communications** are an important example for the concept of band-pass signals. Here, for a given size of the employed antenna, the ability of the antenna to radiate (and receive) signals efficiently depends on the wavelength  $\lambda$  of the corresponding waves.

- Typically, the antenna size is in the order of magnitude of the wavelengths to be radiated efficiently. As an example, consider a hand-held mobile phone which has a typical size of about 15 cm. From  $c = \lambda f_c$  with  $c$  denoting the speed of light, we have

$$f_c = \frac{3 \cdot 10^8 \text{ m/s}}{0.15 \text{ m}} = 2 \text{ GHz}.$$

This is a typical carrier frequency for wireless communications (cellular radio, wireless local area networks etc.), which is located in the so-called **microwave range** (microwave ovens operate at a frequency of about 2.4 GHz).

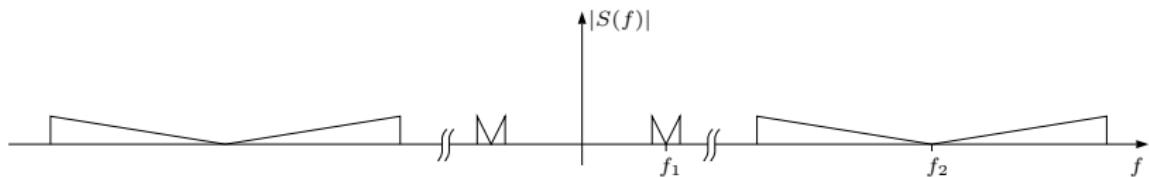
- Note that the spectra in the different frequency bands do not have to have symmetry properties w.r.t. carrier frequency as opposed to the case  $f_c = 0$ .



- A special case of a band-pass signal arises if we assume that we are given a real-valued **low-pass (or baseband) signal**  $s_\ell(t)$   $\xrightarrow{\quad}$   $S_\ell(f)$  which is shifted to the carrier frequency band by

$$s(t) = s_\ell(t) \cos(2\pi f_c t) \xrightarrow{\quad} S(f) = \frac{1}{2}(S_\ell(f - f_c) + S_\ell(f + f_c)).$$

Now, the symmetry properties of  $S_\ell(f)$  w.r.t.  $f = 0$  carry over to symmetry properties of  $S(f)$  w.r.t.  $f = \pm f_c$ .



- For the representation of transmitted signals, we will consider exclusively real-valued signals, i.e.  $s(t) \in \mathbb{R}$  for  $t \in \mathbb{R}$ .

- From the properties of the Fourier transform, it is known that for  $s(t) \in \mathbb{R}$ , we have  $S(-f) = S^*(f)$ .
- Therefore, the complete information about  $s(t)$  is contained in  $S(f)$  for  $f \geq 0$ .
- In view of the aforementioned special case, we define the spectral content of  $S(f)$  for  $f \geq 0$  (scaled by a factor 2 to compensate for the factor 1/2 introduced by the multiplication of a baseband signal by  $\cos(2\pi f_c t)$ )

$$S_+(f) = 2u(f)S(f) = (1 + \text{sign}(f))S(f),$$

where  $S(f)$  is the Fourier transform of  $s(t)$ ,  $u(f)$  is the unit step function and  $\text{sign}(f)$  is the signum function defined by

$$\text{sign}(f) = \begin{cases} 1 & \text{for } f > 0 \\ 0 & \text{for } f = 0 \\ -1 & \text{for } f < 0. \end{cases}$$

- In the time-domain, we obtain the so-called **analytic signal**  $s_+(t)$  or **pre-envelope** of  $s(t)$ , being, in general, **complex-valued** and defined by

$$s_+(t) = \int_{\mathbb{R}} S_+(f) e^{j2\pi f t} df = \mathcal{F}^{-1} \{2u(f)\} \star \mathcal{F}^{-1} \{S(f)\}.$$

- Since

$$\mathcal{F}^{-1} \{2u(f)\} = \mathcal{F}^{-1} \{1 + \text{sign}(f)\} = \mathcal{F}^{-1} \{1\} + \mathcal{F}^{-1} \{\text{sign}(f)\} = \delta(t) + \frac{j}{\pi t},$$

we have

$$s_+(t) = \left( \delta(t) + \frac{j}{\pi t} \right) \star s(t) = s(t) + \frac{j}{\pi t} \star s(t) = s(t) + j\hat{s}(t),$$

where

$$\hat{s}(t) = \frac{1}{\pi t} \star s(t) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{s(\tau)}{t - \tau} d\tau.$$

- Thus, the signal  $\hat{s}(t)$  can be viewed as the output of a filter with impulse response

$$h(t) = \frac{1}{\pi t}, \quad -\infty < t < \infty$$

when excited by the input signal  $s(t)$ . Such a filter is called [Hilbert transformer](#) whose frequency response is obviously

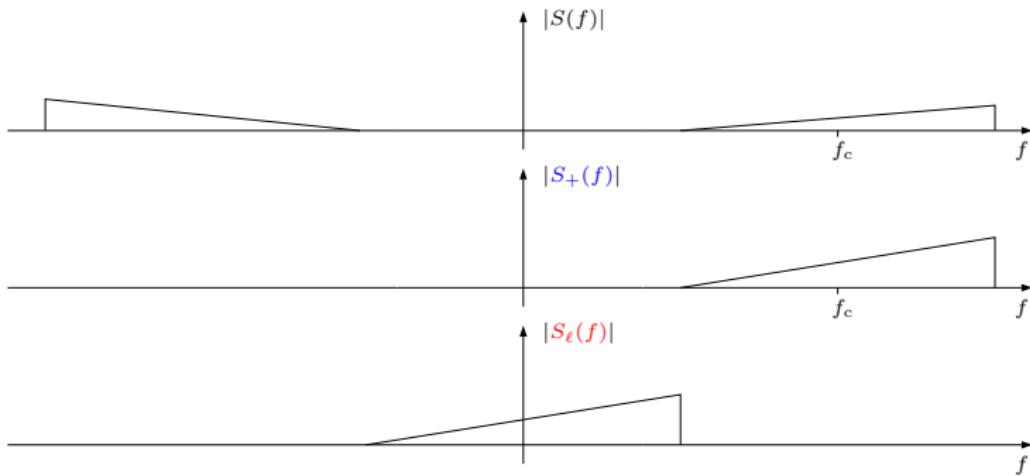
$$H(f) = \begin{cases} -j & \text{for } f > 0 \\ 0 & \text{for } f = 0 \\ j & \text{for } f < 0. \end{cases}$$

Thus, the Hilbert transformer is a  $\pi/2$ -phase shifter, which together with the factor  $j$  provides the frequency response  $\text{sign}(f)$ .

- Note that the Hilbert transformer in its general form is a non-causal filter (i.e.,  $h(t) \neq 0$  for  $t < 0$ ).
- Since  $s(t) \in \mathbb{R}$  and  $s_+(t) = s(t) + j\hat{s}(t) = s(t) + j\frac{1}{\pi t} * s(t)$ , we obtain that [the imaginary part of an analytic signal is the Hilbert transform of the real part](#).

Clearly, the analytic signal  $s_+(t)$  is, in general, a special case of a band-pass signal. We can define an **equivalent low-pass representation** by performing a **frequency translation** of  $S_+(f)$  according to

$$S_\ell(f) = S_+(f + f_c).$$



- The equivalent time-domain relation is

$$s_\ell(t) = s_+(t)e^{-j2\pi f_c t} = [s(t) + j\hat{s}(t)]e^{-j2\pi f_c t}$$

and thus  $s_+(t) = s_\ell(t)e^{j2\pi f_c t}$ .

- If we set  $s_\ell(t) = x(t) + jy(t)$  and equate real and imaginary parts , we obtain

$$\begin{aligned}s(t) &= x(t) \cos(2\pi f_c t) - y(t) \sin(2\pi f_c t) \\ \hat{s}(t) &= x(t) \sin(2\pi f_c t) + y(t) \cos(2\pi f_c t).\end{aligned}$$

- The first equation can be used to characterize the band-pass signal. The low-frequency signal components  $x(t)$  and  $y(t)$  may be viewed as amplitude modulations impressed on the carrier components  $\cos(2\pi f_c t)$  and  $\sin(2\pi f_c t)$ , respectively.
- Since the carrier components are in phase quadrature,  $x(t)$  and  $y(t)$  are called the quadrature components of the band-pass signal  $s(t)$ .

- An alternative representation results from

$$s(t) = \Re \left\{ [x(t) + jy(t)] e^{j2\pi f_c t} \right\} = \Re \left\{ s_\ell(t) e^{j2\pi f_c t} \right\},$$

where  $\Re \{\cdot\}$  denotes the real part of the complex-valued argument.

- The low-pass signal  $s_\ell(t)$  is usually called the **complex envelope** of the real-valued signal  $s(t)$  and can be viewed as the **equivalent low-pass signal**.
- A third representation can be obtained upon setting  $s_\ell(t) = a(t)e^{j\theta(t)}$  with

$$\begin{aligned} \text{the envelope of } s(t) \text{ defined by } a(t) &= \sqrt{x^2(t) + y^2(t)} \\ \text{the phase of } s(t) \text{ defined by } \theta(t) &= \tan^{-1} \left( \frac{y(t)}{x(t)} \right). \end{aligned}$$

- Then

$$s(t) = \Re \left\{ s_\ell(t) e^{j2\pi f_c t} \right\} = \Re \left\{ a(t) e^{j[2\pi f_c t + \theta(t)]} \right\} = a(t) \cos(2\pi f_c t + \theta(t)).$$

- The Fourier transform of  $s(t)$  is

$$S(f) = \int_{\mathbb{R}} \Re \left\{ s_\ell(t) e^{j2\pi f_c t} \right\} e^{-j2\pi f t} dt.$$

Using the identity  $\Re \{a\} = \frac{1}{2}(a + a^*)$ , we obtain finally

$$\begin{aligned} S(f) &= \frac{1}{2} \int_{\mathbb{R}} \left[ s_\ell(t) e^{j2\pi f_c t} + s_\ell^*(t) e^{-j2\pi f_c t} \right] e^{-j2\pi f t} dt \\ &= \frac{1}{2} [S_\ell(f - f_c) + S_\ell^*(-f - f_c)]. \end{aligned}$$

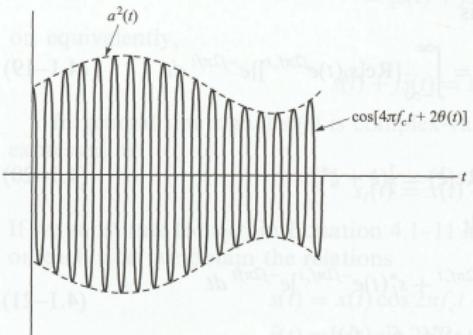
- Note that for the special case of a real-valued signal  $s_\ell(t)$  modulating a carrier signal  $\cos(2\pi f_c t)$ , we have  $S_\ell(f) = S_\ell^*(-f)$  and

$$S(f) = \frac{1}{2} [S_\ell(f - f_c) + S_\ell(f + f_c)]$$

- The energy in the signal  $s(t)$  is defined as

$$\begin{aligned}\mathcal{E} &= \int_{\mathbb{R}} s^2(t) dt = \int_{\mathbb{R}} a^2(t) \cos^2(2\pi f_c t + \theta(t)) dt \\ &= \frac{1}{2} \int_{\mathbb{R}} a^2(t)(1 + \cos(4\pi f_c t + 2\theta(t))) dt \approx \frac{1}{2} \int_{\mathbb{R}} a^2(t) dt = \frac{1}{2} \int |s_\ell(t)|^2 dt,\end{aligned}$$

where the approximation results from the integration of the product of the narrowband process  $a^2(t)$  and the carrier signal  $\cos(4\pi f_c t + 2\theta(t))$  providing negligible contributions to the integral.



- To characterize the transmission properties of a **linear band-pass system**, we assume that  $h(t)$  is real so that  $H(f) = H^*(-f)$ .
- Let us define the **equivalent low-pass frequency response** by

$$H_\ell(f - f_c) = \begin{cases} H(f) & \text{for } f > 0 \\ 0 & \text{for } f \leq 0 \end{cases}.$$

- As a consequence, we have

$$H_\ell^*(-f - f_c) = \begin{cases} 0 & \text{for } f \geq 0 \\ H^*(-f) & \text{for } f < 0 \end{cases}.$$

- Using  $H(f) = H^*(-f)$ , we obtain for  $f \in \mathbb{R} \setminus \{0\}$

$$H(f) = H_\ell(f - f_c) + H_\ell^*(-f - f_c),$$

which resembles the relation  $S(f) = \frac{1}{2}[S_\ell(f - f_c) + S_\ell^*(-f - f_c)]$  except for the factor 1/2.

- Taking the inverse Fourier transform, we obtain

$$h(t) = h_\ell(t)e^{j2\pi f_c t} + h_\ell^*(t)e^{-j2\pi f_c t} = 2\Re \left\{ h_\ell(t)e^{j2\pi f_c t} \right\},$$

where  $h_\ell(t)$  is the inverse Fourier transform of  $H_\ell(f)$ .

- Note that, in general, the impulse response  $h_\ell(t)$  of the equivalent low-pass system is complex-valued.
- Comparing the two equations

$$\begin{aligned} h(t) &= 2\Re \left\{ h_\ell(t)e^{j2\pi f_c t} \right\} \\ s(t) &= \Re \left\{ s_\ell(t)e^{j2\pi f_c t} \right\}, \end{aligned}$$

we conclude that the relations of the signals  $s(t)$  and  $s_\ell(t)$  and  $h(t)$  and  $h_\ell(t)$ , resp., are identical except for the factor 2.

- What is the response of a linear time-invariant (LTI) system to an input signal expressed in the equivalent low-pass domain (the latter being sometimes called complex baseband)?

- Assume we excite an LTI system with a narrowband band-pass signal  $s(t)$  whose equivalent low-pass signal w.r.t. a carrier frequency  $f_c$  is  $s_\ell(t)$ . Let the LTI system be described by the channel impulse response  $h(t)$  and its equivalent low-pass impulse response  $h_\ell(t)$ .
- Obviously, the output signal  $r(t)$  of the LTI system is again a band-pass signal which can be expressed as

$$r(t) = \int_{\mathbb{R}} s(\tau)h(t - \tau)d\tau = \Re \left\{ r_\ell(t)e^{j2\pi f_c t} \right\}$$

with  $r_\ell(t)$  denoting the equivalent low-pass signal of  $r(t)$ .

- In the Fourier domain, we obtain

$$\begin{aligned} R(f) &= S(f)H(f) = \frac{1}{2}[S_\ell(f - f_c) + S_\ell^*(-f - f_c)][H_\ell(f - f_c) + H_\ell^*(-f - f_c)] \\ &= \frac{1}{2}[S_\ell(f - f_c)H_\ell(f - f_c) + S_\ell^*(-f - f_c)H_\ell^*(-f - f_c)] \end{aligned}$$

- If we define the equivalent low-pass signal of  $R(f)$  in the Fourier domain by

$$R_\ell(f) = H_\ell(f)S_\ell(f),$$

we obtain  $R(f) = \frac{1}{2}[R_\ell(f - f_c) + R_\ell^*(-f - f_c)]$ .

- Clearly, in the time domain, we have upon taking the inverse Fourier transform

$$r_\ell(t) = s_\ell(t) \star h_\ell(t) = \int_{\mathbb{R}} s_\ell(\tau)h_\ell(t - \tau)d\tau.$$

- This result can be formulated as follows:

The response of an LTI system to a band-pass signal is again a band-pass signal, whose low-pass equivalent signal in the frequency domain results from the product of Fourier transforms of the low-pass equivalent input signal and the low-pass equivalent channel impulse response.

- In particular, we can ignore the frequency translation expressed by  $f_c$  in the characterization of modulated signals.

- The characterization of band-pass signals above treats the case of **deterministic signals**. Below, we extend the representation to sample functions of a **band-pass stationary stochastic process**.
  - Unlike in the deterministic case, we will derive the **relations between the correlation functions and power spectra of the (stochastic) band-pass signal and the correlation function and the power spectra of the equivalent (stochastic) low-pass signal**.
  - Suppose that  $n(t)$  is a sample function of a zero-mean stationary stochastic process with power spectral density  $\Phi_{nn}(f)$ . The process is said to be a **narrowband band-pass process** if the support of  $\Phi_{nn}(f)$  is centered around  $\pm f_c$  and much smaller than  $f_c$ .
  - The quantity  $f_c$  is usually termed the **carrier frequency** of  $n(t)$ .
  - We proceed as in the deterministic case, i.e. we first define the **analytic signal**  $n_+(t)$  which is the output of a filter with transfer function  $H(f) = 2u(f)$  and input signal  $n(t)$ .
- N.B.: With a slight abuse of notation, we use  $n(t)$  for both sample functions as well as the stochastic process itself.

- Clearly, the power spectrum of  $n_+(t)$  is given by

$$\Phi_{n_+ n_+}(f) = 4u(f)\Phi_{nn}(f).$$

- Note that, as above, the original stochastic process is related to  $n_+(t)$  by  $n(t) = \Re\{n_+(t)\}$ .
- The complex envelope of  $n(t)$  is defined, in analogy with the deterministic case, as

$$z(t) = n_+(t)e^{-j2\pi f_c t} = x(t) + jy(t),$$

where  $x(t)$  and  $y(t)$  are the quadrature components of  $n(t)$ .

- Obviously, if  $z(t)$  is zero-mean, so are the quadrature components.
- The stationarity of  $n(t)$  implies that the autocorrelation and cross-correlation functions of  $x(t)$  and  $y(t)$  satisfy the following properties:

$$\begin{aligned}\phi_{xx}(\tau) &= \phi_{yy}(\tau) \\ \phi_{xy}(\tau) &= -\phi_{yx}(\tau).\end{aligned}$$

**Proof**

First, we have equality in the quadratic mean of the processes

$$n(t) = \Re \left\{ z(t) e^{j2\pi f_c t} \right\} = \Re \left\{ [x(t) + jy(t)] e^{j2\pi f_c t} \right\} = x(t) \cos(2\pi f_c t) - y(t) \sin(2\pi f_c t).$$

The autocorrelation function  $\Phi_{nn}(f)$  of  $n(t)$  is

$$\begin{aligned} E[n(t)n(t+\tau)] &= E \{ [x(t) \cos(2\pi f_c t) - y(t) \sin(2\pi f_c t)] \times \\ &\quad [x(t+\tau) \cos(2\pi f_c (t+\tau)) - y(t+\tau) \sin(2\pi f_c (t+\tau))] \} \\ &= \phi_{xx}(\tau) \cos(2\pi f_c t) \cos(2\pi f_c (t+\tau)) \\ &\quad + \phi_{yy}(\tau) \sin(2\pi f_c t) \sin(2\pi f_c (t+\tau)) \\ &\quad - \phi_{xy}(\tau) \sin(2\pi f_c t) \cos(2\pi f_c (t+\tau)) \\ &\quad - \phi_{yx}(\tau) \cos(2\pi f_c t) \sin(2\pi f_c (t+\tau)). \end{aligned}$$

Using the trigonometric identities for  $\cos A \cos B$ ,  $\sin A \cos B$  and  $\sin A \sin B$ , we obtain

**Proof (cont.)**

$$\begin{aligned}
 E[n(t)n(t+\tau)] &= \frac{1}{2}[\phi_{xx}(\tau) + \phi_{yy}(\tau)] \cos(2\pi f_c \tau) \\
 &\quad + \frac{1}{2}[\phi_{xx}(\tau) - \phi_{yy}(\tau)] \cos(2\pi f_c(2t + \tau)) \\
 &\quad - \frac{1}{2}[\phi_{yx}(\tau) - \phi_{xy}(\tau)] \sin(2\pi f_c \tau) \\
 &\quad - \frac{1}{2}[\phi_{yx}(\tau) + \phi_{xy}(\tau)] \sin(2\pi f_c(2t + \tau)).
 \end{aligned}$$

Since  $n(t)$  is a stationary process, the right-hand side must be independent of  $t$ . As a consequence, we find

$$\begin{aligned}
 \phi_{xx}(\tau) &= \phi_{yy}(\tau), \\
 \phi_{xy}(\tau) &= -\phi_{yx}(\tau).
 \end{aligned}$$

Inserting these equations in the above expression for  $E[n(t)n(t+\tau)]$ , we obtain

$$\phi_{nn}(\tau) = \phi_{xx}(\tau) \cos(2\pi f_c \tau) - \phi_{yx}(\tau) \sin(2\pi f_c \tau).$$

Note the formal similarity between the following equations

$$\begin{aligned}\phi_{nn}(\tau) &= \phi_{xx}(\tau) \cos(2\pi f_c \tau) - \phi_{yx}(\tau) \sin(2\pi f_c \tau) \\ n(t) &= x(t) \cos(2\pi f_c t) - y(t) \sin(2\pi f_c t),\end{aligned}$$

which express the band-pass process in terms of the quadrature components.

- The autocorrelation function of the equivalent low-pass process  $z(t)$  is defined as  $\phi_{zz}(\tau) = \frac{1}{2}E[z^*(t)z(t + \tau)]$ . Substituting  $z(t) = x(t) + jy(t)$ , we find

$$\phi_{zz}(\tau) = \frac{1}{2}[\phi_{xx}(\tau) + \phi_{yy}(\tau) - j\phi_{xy}(\tau) + j\phi_{yx}(\tau)].$$

- Exploiting  $\phi_{xx}(\tau) = \phi_{yy}(\tau)$  and  $\phi_{xy}(\tau) = -\phi_{yx}(\tau)$ , we find

$$\phi_{nn}(\tau) = \Re \left\{ \phi_{zz}(\tau) e^{j2\pi f_c \tau} \right\}.$$

- Thus, the autocorrelation function  $\phi_{nn}(\tau)$  of the band-pass stochastic process is uniquely determined from the autocorrelation function  $\phi_{zz}(\tau)$  and the carrier frequency  $f_c$ .

- The power density spectrum  $\Phi_{nn}(f)$  of the stochastic process  $n(t)$  is the Fourier transform of  $\phi_{nn}(\tau)$ :

$$\Phi_{nn}(f) = \int_{\mathbb{R}} \left\{ \Re \left\{ \phi_{zz}(\tau) e^{j2\pi f_c \tau} \right\} \right\} e^{-j2\pi f \tau} d\tau = \frac{1}{2} [\Phi_{zz}(f - f_c) + \Phi_{zz}(-f - f_c)].$$

- From the definition of the cross-correlation function, we have for the cross-correlation function of the quadrature components that

$$\phi_{yx}(\tau) = \phi_{xy}(-\tau).$$

Since, as shown above,  $\phi_{xy}(\tau) = -\phi_{yx}(\tau)$ , we conclude that

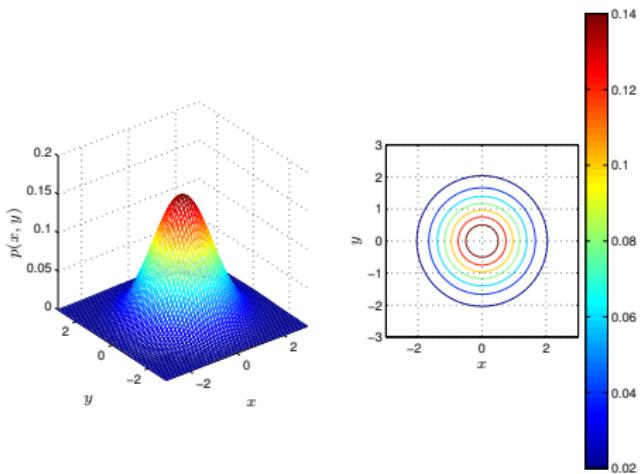
$$\phi_{xy}(\tau) = -\phi_{xy}(-\tau).$$

- Thus,  $\phi_{xy}(\tau)$  is an odd function of  $\tau$ . Consequently,  $\phi_{xy}(0) = 0$  and  $x(t)$  and  $y(t)$  are uncorrelated (for  $\tau = 0$  only).
- If  $\phi_{xy}(\tau) = 0$  for all  $\tau$ , then  $\phi_{zz}(\tau)$  is real and the power spectral density  $\Phi_{zz}(f)$  is symmetric about  $f = 0$ .

- If the quadrature components of a stationary stochastic process  $n(t)$  are jointly Gaussian distributed, they are independent for a time shift  $\tau = 0$  with PDF

$$p(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}},$$

where the variance  $\sigma^2$  is defined as  $\sigma^2 = \phi_{xx}(0) = \phi_{yy}(0) = \phi_{nn}(0)$ .



# Orthogonal Expansions of Signals

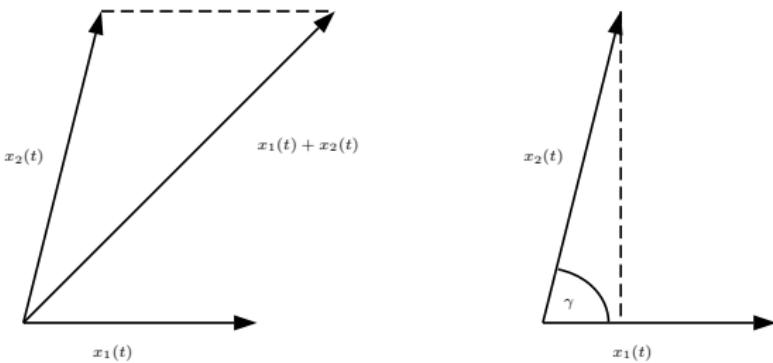
- Since in the description of signals and systems as well as the performance analysis, we make extensive use of concepts from linear algebra, we recap the most important topics here.
- We treat signals defined on some interval  $[a, b]$  as vectors and instantiate the corresponding operations for the signals.
- The **inner product** of two generally complex-valued signals  $x_1(t)$  and  $x_2(t)$  is denoted by

$$\langle x_1(t), x_2(t) \rangle = \int_a^b x_1(t)x_2^*(t)dt.$$

- The inner product induces a **norm**  $\|x(t)\|$  which corresponds to the **length of the vector  $x(t)$**  defined by

$$\|x(t)\| = \sqrt{\langle x(t), x(t) \rangle} = \sqrt{\int_a^b |x(t)|^2 dt}.$$

- Two signals  $x_1(t)$  and  $x_2(t)$  are **orthogonal** if their inner product is zero.
- A set of  $m$  signals are **orthonormal** if they are orthogonal and their norms are all unity.
- A set of  $m$  signals are **linearly independent** if no signal can be represented as a linear combination of the remaining signals.
- **Triangle inequality:**  $\|x_1(t) + x_2(t)\| \leq \|x_1(t)\| + \|x_2(t)\|$ .
- **Cauchy-Schwarz inequality:**  $|\langle x_1(t), x_2(t) \rangle| \leq \|x_1(t)\| \|x_2(t)\|$  with equality when  $x_2(t) = ax_1(t)$ , where  $a$  is any complex number.



- Suppose we are given a set of functions  $\{f_n(t), n = 1, 2, \dots, K\}$  which are orthonormal in the sense that

$$\langle f_n(t), f_m(t) \rangle = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n. \end{cases}$$

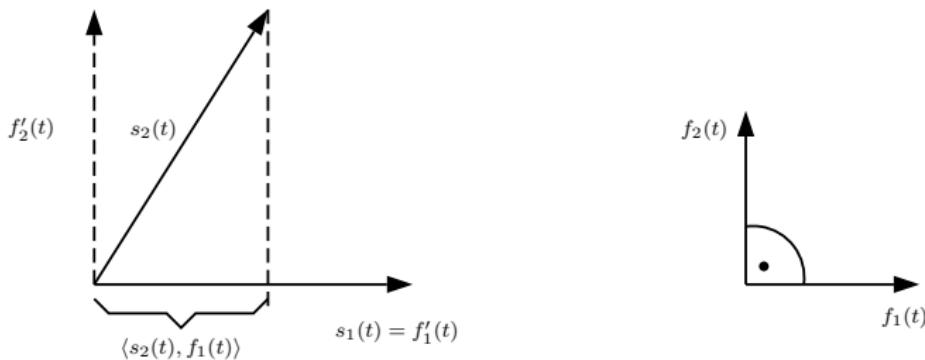
- Then, the expansion

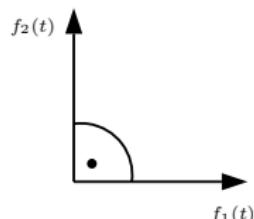
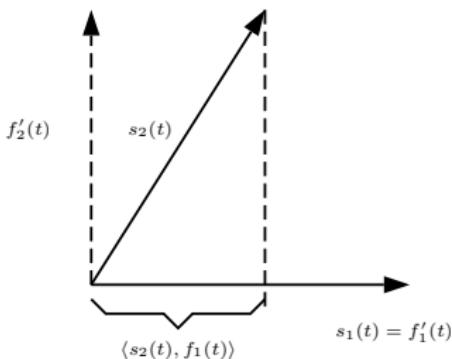
$$\hat{s}(t) = \sum_{k=1}^K s_k f_k(t)$$

with  $s_k = \langle s(t), f_k(t) \rangle$  for  $k = 1, \dots, K$  can be shown to provide the best approximation in the mean-square sense to  $s(t)$  in the space spanned by  $\{f_n(t), n = 1, 2, \dots, K\}$ .

- If the residual error energy  $\mathcal{E}_e = \|s(t) - \hat{s}(t)\|^2 = 0$  for every finite-energy signal, the set of orthonormal functions is said to be complete.
- Example: the set  $\left\{ \sqrt{1/T}, \sqrt{2/T} \cos(2\pi kt/T), \sqrt{2/T} \sin(2\pi kt/T) \right\}$  with  $k = 1, 2, \dots, \infty$  is complete in  $[a, b] = [0, T]$ .

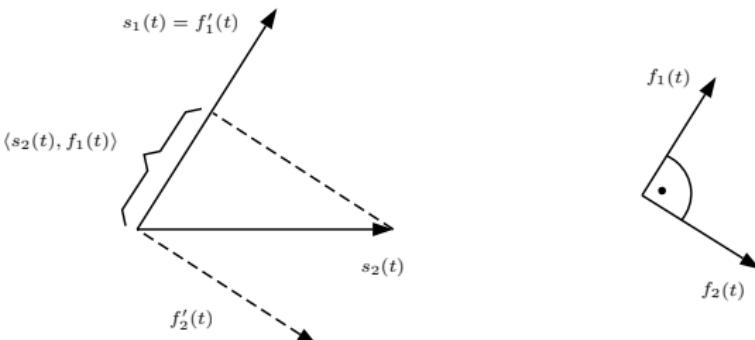
- Suppose we are given a set of functions  $\{s_i(t), i = 1, 2, \dots, M\}$  for  $t \in \mathbb{R}$  from which we want to construct a set of orthonormal waveforms which span a certain space of dimension  $N \leq M$ .
- The **Gram-Schmidt orthogonalization procedure** allows us to construct such a set. The scheme is completely analogous to the one defined for usual vectors.





$$\begin{aligned}f'_1(t) &= s_1(t) \\f_1(t) &= \frac{f'_1(t)}{\sqrt{\int_{\mathbb{R}} |f'_1(t)|^2 dt}} = \frac{f'_1(t)}{\sqrt{\varepsilon_1}} \\f'_2(t) &= s_2(t) - \langle s_2(t), f_1(t) \rangle f_1(t) \\f_2(t) &= \frac{f'_2(t)}{\sqrt{\int_{\mathbb{R}} |f'_2(t)|^2 dt}} = \frac{f'_2(t)}{\sqrt{\varepsilon_2}}\end{aligned}$$

define the unnormalized waveform  $f'_1(t)$   
normalize  $f'_1(t)$  to unit energy  
  
calculate unnormalized component  
being orthogonal to  $f_1(t)$   
normalize  $f'_2(t)$  to unit energy



$$\begin{aligned}
 f'_1(t) &= s_1(t) \\
 f_1(t) &= \frac{f'_1(t)}{\sqrt{\int_{\mathbb{R}} |f'_1(t)|^2 dt}} = \frac{f'_1(t)}{\sqrt{\varepsilon_1}} \\
 f'_2(t) &= s_2(t) - \langle s_2(t), f_1(t) \rangle f_1(t) \\
 f_2(t) &= \frac{f'_2(t)}{\sqrt{\int_{\mathbb{R}} |f'_2(t)|^2 dt}} = \frac{f'_2(t)}{\sqrt{\varepsilon_2}}
 \end{aligned}$$

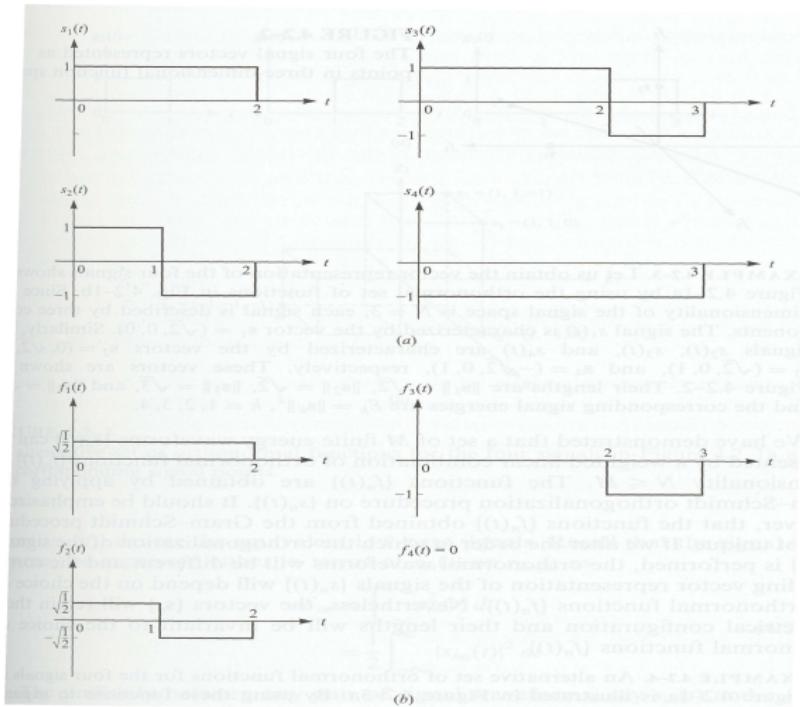
define the unnormalized waveform  $f'_1(t)$   
 normalize  $f'_1(t)$  to unit energy  
 calculate unnormalized component  
 being orthogonal to  $f_1(t)$   
 normalize  $f'_2(t)$  to unit energy

- The general scheme for constructing a basis  $\{f_k(t), k = 1, 2, \dots, N\}$  with  $N \leq M$  is thus given by

$$f_k(t) = \frac{s_k(t) - \sum_{i=1}^{k-1} \langle s_k(t), f_i(t) \rangle f_i(t)}{\sqrt{\int_{\mathbb{R}} \left| s_k(t) - \sum_{i=1}^{k-1} \langle s_k(t), f_i(t) \rangle f_i(t) \right|^2 dt}}.$$

- The orthogonalization process is continued until all the  $M$  signal waveforms  $\{s_i(t)\}$  have been exhausted and  $N \leq M$  orthonormal waveforms have been constructed.
- The dimensionality of the basis and thus of the signal space will be equal to  $M$  if all the signal waveforms  $\{s_i(t)\}$  are linearly independent, i.e. none of the signal waveforms is a linear combination of the other signal waveforms.

Example: Gram-Schmidt orthogonalization of  $M = 4$  linearly dependent waveforms  $\{s_k(t), k = 1, \dots, 4\}$



- Once we have constructed the basis of orthonormal waveforms  $\{f_k(t), k = 1, 2, \dots, N\}$ , we can express the  $M$  signals  $\{s_k(t), k = 1, 2, \dots, M\}$  as linear combinations of the  $\{f_k(t), k = 1, 2, \dots, N\}$ :

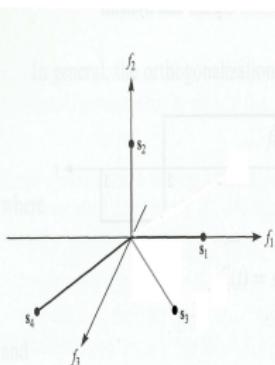
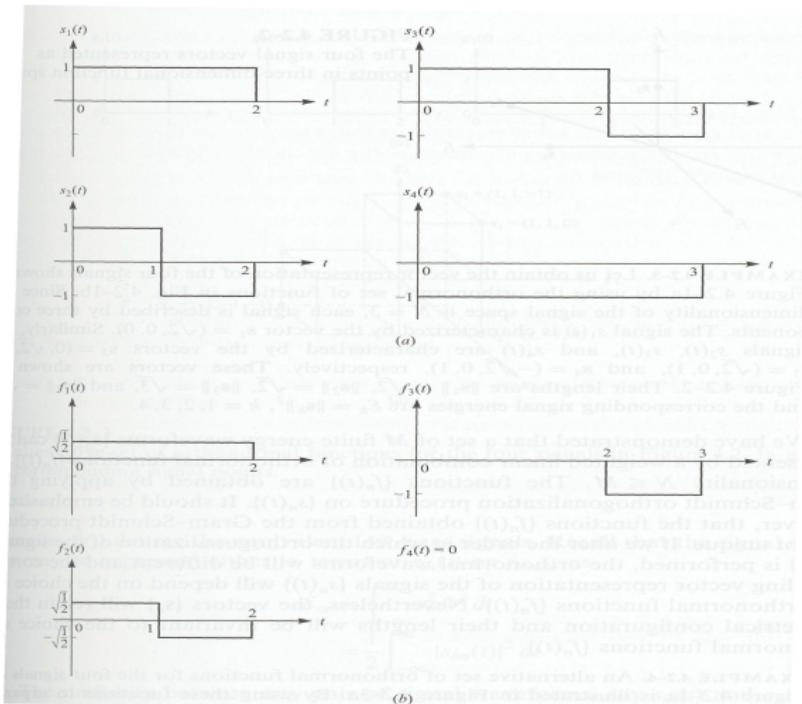
$$s_k(t) = \sum_{n=1}^N \langle s_k(t), f_n(t) \rangle f_n(t) = \sum_{n=1}^N s_{kn} f_n(t)$$

- Clearly, the energy of  $s_k(t)$  is given by

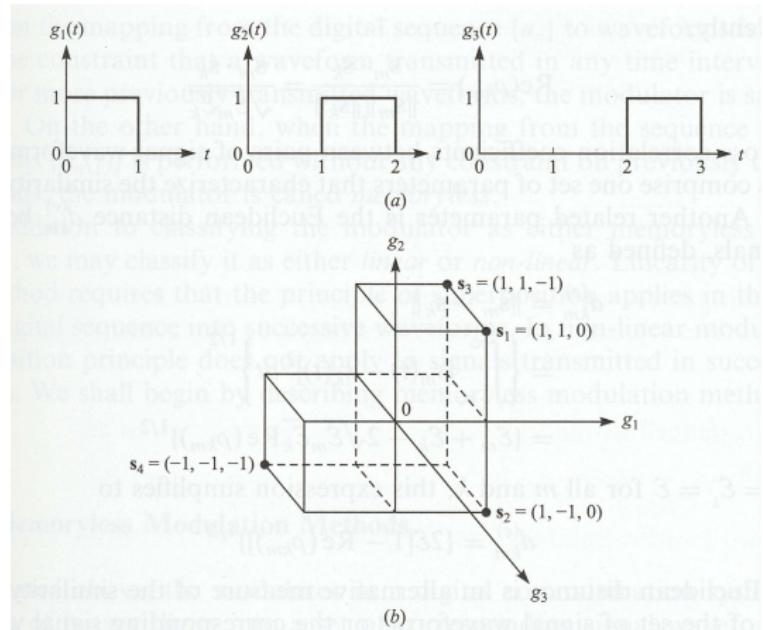
$$\mathcal{E}_k = \int_{\mathbb{R}} |s_k(t)|^2 dt = \sum_{n=1}^N |s_{kn}|^2 = \|s_k\|^2,$$

where the signal  $s_k(t)$  can be represented by the vector  $s_k = [s_{k1}, \dots, s_{kN}]$  and the energy of  $s_k(t)$  is just the squared length of the vector (i.e. its squared Euclidean distance from the origin).

Example: Gram-Schmidt orthogonalization of  $M = 4$  linearly dependent waveforms  $\{s_k(t), k = 1, \dots, 4\}$



Example: Representation of the  $M = 4$  linearly dependent waveforms  $\{s_k(t), k = 1, \dots, 4\}$  in an alternative basis  $\{g_i(t), i = 1, 2, 3\}$ .



**Problem:** Find the coefficients  $s_k = [s_{k1}, \dots, s_{k3}]$  for the signals  $\{s_k(t), k = 1, \dots, 4\}$  on page 252 using the basis  $\{g_i(t), i = 1, 2, 3\}$  shown above.

# Representation of Digitally Modulated Signals

- In the transmission of digital information over a communication channel, the modulator is the interface device that maps the digital information into analog waveforms that match the characteristics of the channel.
- The mapping is generally performed by taking blocks of  $k = \log_2 M$  binary digits (or bits) at a time from the information sequence  $\{a_n\}$  and selecting one of  $M = 2^k$  deterministic, finite energy waveforms  $\{s_m(t), m = 1, 2, \dots, M\}$  for transmission over the channel.
- When the mapping from the sequence  $\{a_n\}$  to the waveforms  $\{s_m(t)\}$  is performed without any constraint on previously transmitted waveforms, the modulator is called memoryless. If it depends, on the other hand, on previously transmitted symbols, it is said to have memory.
- In addition to classifying the modulator as either memoryless or having memory, we may classify it as either linear or non-linear. Linearity of a modulation method requires that the principle of superposition applies in the mapping of the digital sequence into successive waveforms. In non-linear modulation, the superposition principle does not apply to signals transmitted in successive time intervals.

## Memoryless modulation methods: PAM

- First, we consider **memoryless modulation methods**.
- The waveforms chosen by the modulator may differ in either amplitude or in phase or in frequency, or some combination of two or more signal parameters.
- We consider each of these signal types separately, beginning with **pulse amplitude modulation (PAM)**.
- In all cases, we assume that the sequence of bits at the input to the modulator occurs at a rate  $R$  bits/s.
- In digital PAM, the signal waveforms may be represented as

$$s_m(t) = \Re \left\{ A_m g(t) e^{j2\pi f_c t} \right\} = A_m g(t) \cos(2\pi f_c t), \quad m = 1, \dots, M, 0 \leq t \leq T,$$

where  $\{A_m, 1 \leq m \leq M\}$  denote the set of  $M$  possible amplitudes corresponding to  $M = 2^k$  possible  $k$ -bit blocks of symbols.

- The real-valued pulse  $g(t)$  influences the spectrum of the transmitted signal.

### Memoryless modulation methods: PAM

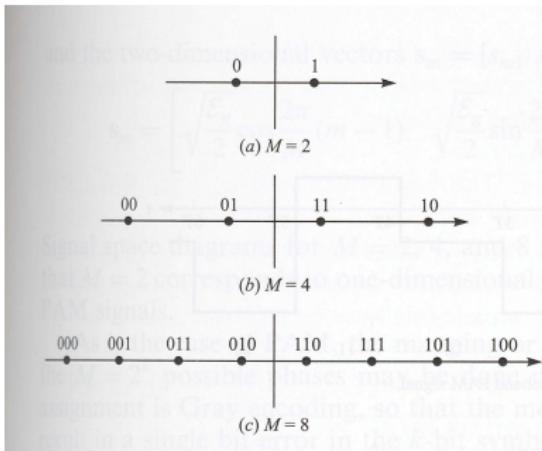
- The discrete values (or **levels**) are  $A_m = (2m - 1 - M)d$ ,  $m = 1, 2, \dots, M$ , where  $2d$  is the distance between adjacent signal amplitudes.
- The symbol rate for the PAM signal is  $R/k$ . This is the rate at which changes occur in the amplitude of the carrier to reflect the transmission of a new symbol
- The  $M$  PAM signals have energies

$$\mathcal{E}_m = \int_0^T |s_m(t)|^2 dt = \frac{1}{2} A_m^2 \int_0^T g^2(t) dt = \frac{1}{2} A_m^2 \mathcal{E}_g,$$

where  $\mathcal{E}_g$  denotes the energy of the pulse  $g(t)$ .

- Clearly, these signals are one-dimensional ( $N = 1$ ) and, hence, are represented by the general form  $s_m(t) = s_m f(t)$ , where  $f(t)$  is a unit-energy signal waveform defined as  $f(t) = \sqrt{\frac{2}{\mathcal{E}_g}} g(t) \cos(2\pi f_c t)$  and  $s_m = A_m \sqrt{\frac{1}{2} \mathcal{E}_g}$ .

## Memoryless modulation methods: PAM



**FIGURE 4.3-1**  
Signal space diagram for digital PAM signals.

- Mapping of  $k$  bits to  $M = 2^k$  symbol amplitudes can be done in a number of different ways.
- A reasonable way to implement this mapping is the so-called **Gray encoding** where adjacent amplitudes which can be detected erroneously at the receiver differ by only one bit.

### Memoryless modulation methods: PAM

- The distance between any pair of signal points is

$$d_{mn}^{(e)} = \sqrt{(s_m - s_n)^2} = \sqrt{\frac{1}{2}\mathcal{E}_g} |A_m - A_n| = d\sqrt{2\mathcal{E}_g} |m - n|.$$

- Thus, the minimum Euclidean distance between adjacent signal points is

$$d_{\min}^{(e)} = d\sqrt{2\mathcal{E}_g}.$$

- The PAM signal

$$s_m(t) = \Re \left\{ A_m g(t) e^{j2\pi f_c t} \right\} = A_m g(t) \cos(2\pi f_c t), \quad m = 1, \dots, M, 0 \leq t \leq T,$$

is a so-called double-sideband (DSB) signal and requires twice the channel bandwidth of the equivalent low-pass signal for transmission.

### Memoryless modulation methods: PAM

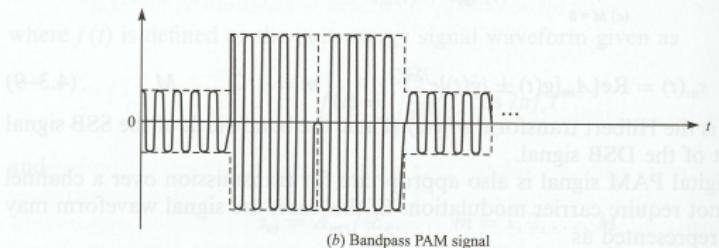
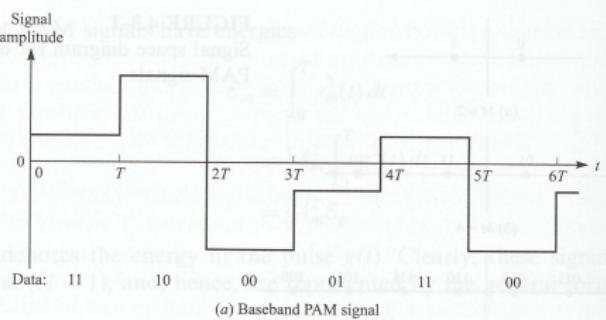
- Alternatively, we can use **single-sideband (SSB) PAM**, which has the representation (lower or upper sideband)

$$s_m(t) = \Re \left\{ A_m [g(t) \pm j\hat{g}(t)] e^{j2\pi f_c t} \right\}, \quad m = 1, \dots, M,$$

where  $\hat{g}(t)$  is the Hilbert transform of  $g(t)$ . Thus, the bandwidth of the SSB signal is half of that of the DSB signal.

- The digital PAM signal is also appropriate for **transmission over a channel that does not require carrier modulation**. In this case, the signal may be simply represented as  $s_m(t) = A_m g(t)$  for  $m = 1, 2, \dots, M$ . This is now called a **baseband signal**.
- In the special case of  $M = 2$  signals, we obtain **antipodal signals**  $s_1(t) = -s_2(t)$ .

## Memoryless modulation methods: PAM



**FIGURE 4.3-2**  
Baseband and band-pass PAM signals.

### Memoryless modulation methods: PSK

- In digital phase modulation, the  $M$  signal waveforms are represented as

$$\begin{aligned}
 s_m(t) &= \Re \left\{ g(t) e^{j2\pi(m-1)/M} e^{j2\pi f_c t} \right\}, \quad m = 1, 2, \dots, M, \quad 0 \leq t \leq T \\
 &= g(t) \cos \left( 2\pi f_c t + \frac{2\pi}{M}(m-1) \right) \\
 &= g(t) \cos \left( \frac{2\pi}{M}(m-1) \right) \cos(2\pi f_c t) - g(t) \sin \left( \frac{2\pi}{M}(m-1) \right) \sin(2\pi f_c t),
 \end{aligned}$$

where  $g(t)$  is the real-valued signal pulse and  $\theta_m = 2\pi(m-1)/M$ ,  $m = 1, 2, \dots, M$ , are the  $M$  possible phases of the carrier that convey the information.

- Digital phase modulation is usually called phase-shift keying (PSK). For  $M = 2$ , we have binary PSK (BPSK), for  $M = 4$  quaternary PSK (QPSK). For higher values of  $M$ , we usually use the value directly, like e.g. 8-PSK for  $M = 8$ .

### Memoryless modulation methods: PSK

- Note that the signal waveforms of an  $M$ -PSK signal set have equal energy, i.e.

$$\mathcal{E} = \int_0^T |s_m(t)|^2 dt = \frac{1}{2} \int_0^T g^2(t) dt = \frac{1}{2} \mathcal{E}_g.$$

The signal waveforms may be represented as a linear combination of two orthonormal signal waveforms  $f_1(t)$  and  $f_2(t)$ , i.e.

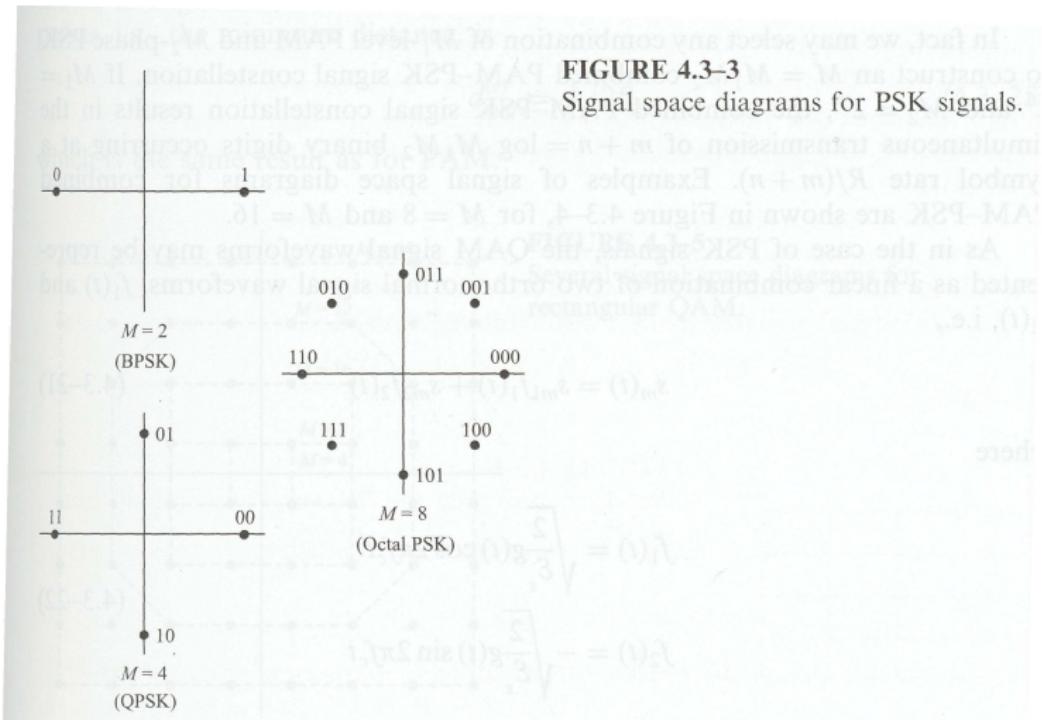
$$s_m(t) = s_{m_1} f_1(t) + s_{m_2} f_2(t), \quad \text{where}$$

$$f_1(t) = \sqrt{\frac{2}{\mathcal{E}_g}} g(t) \cos(2\pi f_c t) \quad f_2(t) = -\sqrt{\frac{2}{\mathcal{E}_g}} g(t) \sin(2\pi f_c t).$$

- The two-dimensional vectors  $\mathbf{s}_m = [s_{m_1} \ s_{m_2}]$  given by

$$\mathbf{s}_m = \left[ \sqrt{\frac{\mathcal{E}_g}{2}} \cos\left(\frac{2\pi}{M}(m-1)\right) \ \sqrt{\frac{\mathcal{E}_g}{2}} \sin\left(\frac{2\pi}{M}(m-1)\right) \right], m = 1, 2, \dots, M.$$

## Memoryless modulation methods: PSK



## Memoryless modulation methods: PSK

- The squared Euclidean distance between any pair of signal points is

$$d_{mn}^{(e)} = ||\mathbf{s}_m - \mathbf{s}_n||^2 = \frac{\mathcal{E}_g}{2} (I^2 + Q^2) \quad \text{with}$$

$$I^2 = \left( \cos \left( \frac{2\pi}{M}(m-1) \right) - \cos \left( \frac{2\pi}{M}(n-1) \right) \right)^2$$

$$Q^2 = \left( \sin \left( \frac{2\pi}{M}(m-1) \right) - \sin \left( \frac{2\pi}{M}(n-1) \right) \right)^2. \quad \text{Since}$$

$$I^2 + Q^2 = 1 - 2 \cos \left( \frac{2\pi}{M}(m-n) \right) + 1, \quad \text{we have}$$

$$d_{mn}^{(e)} = \sqrt{\mathcal{E}_g \left[ 1 - \cos \left( \frac{2\pi}{M}(m-n) \right) \right]}.$$

- Thus, the **minimum** Euclidean distance between adjacent signal points is

$$d_{\min}^{(e)} = \sqrt{\mathcal{E}_g \left[ 1 - \cos \left( \frac{2\pi}{M} \right) \right]}.$$

## Memoryless modulation methods: QAM

- The bandwidth efficiency of PAM/SSB can also be obtained by simultaneously impressing **two separate  $k$ -bit symbols from the information sequence  $\{a_n\}$**  on **two quadrature carriers  $\cos(2\pi f_{ct})$  and  $\sin(2\pi f_{ct})$** .
- The resulting modulation technique is called **quadrature PAM or QAM**.
- The corresponding waveforms may be expressed as

$$\begin{aligned}s_m(t) &= \Re \left\{ [A_{mc} + jA_{ms}]g(t)e^{j2\pi f_{ct}t} \right\}, \quad m = 1, 2, \dots, M, \quad 0 \leq t \leq T \\ &= A_{mc}g(t)\cos(2\pi f_{ct}) - A_{ms}g(t)\sin(2\pi f_{ct}),\end{aligned}$$

where  $A_{mc}$  and  $A_{ms}$  are the information-bearing signal amplitudes of the quadrature carriers and  $g(t)$  is the signal pulse.

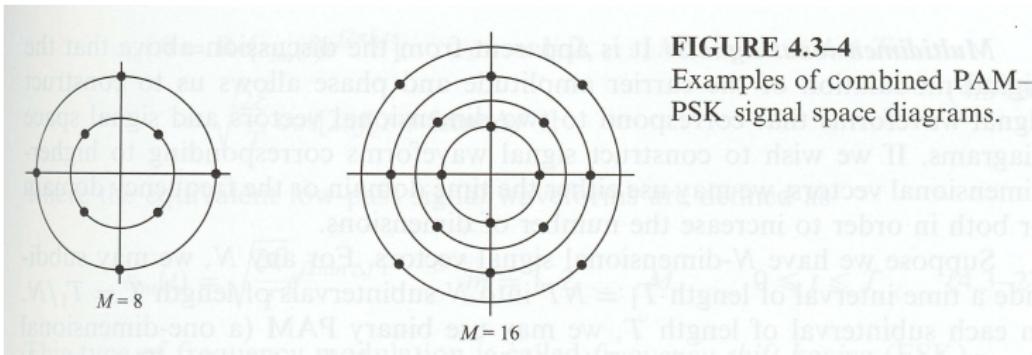
- Alternatively, we can denote the QAM signal waveforms as

$$s_m(t) = \Re \left\{ V_m e^{j\theta_m} g(t) e^{j2\pi f_{ct}t} \right\} = V_m g(t) \cos(2\pi f_{ct} + \theta_m)$$

where  $V_m = \sqrt{A_{mc}^2 + A_{ms}^2}$  and  $\theta_m = \tan^{-1}(A_{ms}/A_{mc})$ , i.e. QAM may be viewed as a **combined amplitude and phase modulation**.

## Memoryless modulation methods: QAM

- We may select any combination of  $M_1$ -level PAM and  $M_2$ -level PSK to construct an  $M = M_1 M_2$  combined PAM-PSK signal constellation.



- If  $M_1 = 2^n$  and  $M_2 = 2^m$ , the combined PAM-PSK signal constellation results in the simultaneous transmission of  $m + n = \text{ld}(M_1 M_2)$  binary digits occurring at a symbol rate  $R/(m + n)$ .

## Memoryless modulation methods: QAM

- Similarly to PSK signals, the QAM signal waveforms may be represented as a linear combination of two orthonormal signal waveforms,  $f_1(t)$  and  $f_2(t)$ , i.e.

$$s_m(t) = s_{m1} f_1(t) + s_{m2} f_2(t), \quad \text{where}$$

$$f_1(t) = \sqrt{\frac{2}{\mathcal{E}_g}} g(t) \cos(2\pi f_c t) \quad f_2(t) = -\sqrt{\frac{2}{\mathcal{E}_g}} g(t) \sin(2\pi f_c t).$$

- The two-dimensional vectors  $\mathbf{s}_m = [s_{m1} \ s_{m2}]$  are given by

$$\mathbf{s}_m = \begin{bmatrix} A_{mc} \sqrt{\frac{\mathcal{E}_g}{2}} & A_{ms} \sqrt{\frac{\mathcal{E}_g}{2}} \end{bmatrix}, \quad m = 1, 2, \dots, M,$$

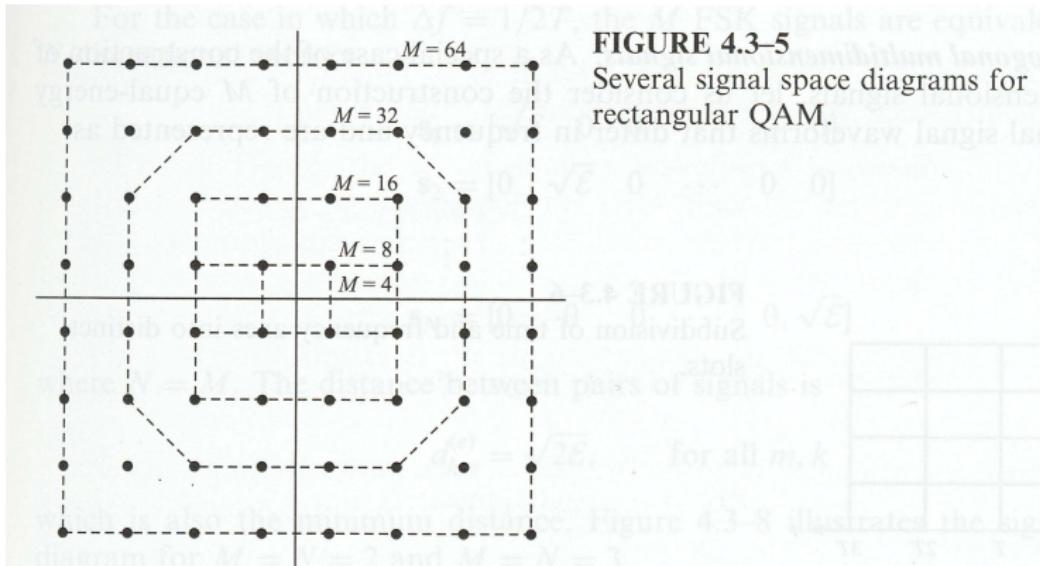
where again  $\mathcal{E}_g$  is the energy of the signal pulse  $g(t)$ .

- The Euclidean distance between any pair of signal vectors is

$$d_{mn}^{(e)} = \|\mathbf{s}_m - \mathbf{s}_n\| = \sqrt{\frac{\mathcal{E}_g}{2} \left[ (A_{mc} - A_{nc})^2 + (A_{ms} - A_{ns})^2 \right]}.$$

## Memoryless modulation methods: QAM

- Consider the case of rectangular signal space constellations:



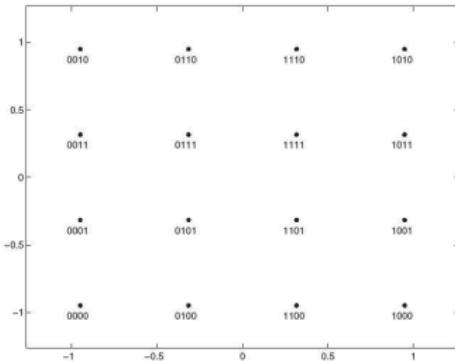
## Memoryless modulation methods: QAM

- Assuming the set of discrete values being given by

$$A_{mc} = A_{ms} = (2m - 1 - M)d \quad \text{for } m = 1, 2, \dots, M,$$

we conclude that the minimum distance between adjacent points in the signal constellation diagram is given by  $d_{\min}^{(e)} = \sqrt{\frac{\mathcal{E}_g}{2}[(2d)]^2} = d\sqrt{2\mathcal{E}_g}$ , which is the same result as for PAM.

- Gray encoding scheme for 16-QAM:

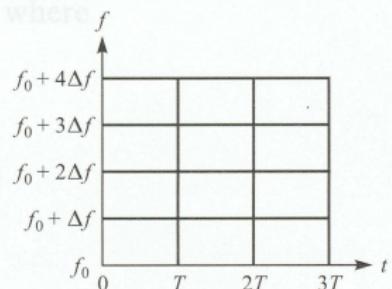


## Memoryless modulation methods: multidimensional signals

- Obviously, we can construct two-dimensional signal space constellations using the quadrature components of a certain carrier signals.
- Higher-dimensional signal waveforms can be constructed using either the time domain or the frequency domain or both.
- Suppose we have  $N$ -dimensional signal vectors. For any  $N$ , we may subdivide a time interval of length  $T_1 = NT$  into  $N$  subintervals of length  $T = T_1/N$ . In each subinterval of length  $T$ , we may use **binary PAM** (a one-dimensional signal) to transmit an element of the  $N$ -dimensional signal vector.
- If  $N$  is even, a time slot of length  $T$  may be used to simultaneously transmit two components of the  $N$ -dimensional vector by modulating the amplitude of quadrature carriers independently by the corresponding components.
- Thus, the  $N$ -dimensional signal vector is transmitted in  $\frac{1}{2}NT$  seconds, i.e.  $\frac{1}{2}N$  time slots).
- Similarly, we can use  $N/2$  frequency slots of bandwidth  $\Delta f$ .

### Memoryless modulation methods: multidimensional signals

- More generally, we may use both the time and the frequency domains jointly to transmit an  $N$ -dimensional signal vector.
- Example: subdivide time and frequency axes into 12 slots to transmit a 24-dimensional signal vector, e.g. using a 4-QAM signal in each slot.



**FIGURE 4.3-6**  
Subdivision of time and frequency axes into distinct slots.

## Memoryless modulation methods: orthogonal multidimensional signals

- As a special case of multidimensional signals, let us consider the construction of  **$M$  equal-energy orthogonal signal waveforms that differ in frequency** and are represented as

$$\begin{aligned}s_m(t) &= \Re \left\{ s_{\ell m}(t) e^{j2\pi f_c t} \right\}, \quad m = 1, 2, \dots, M, \quad 0 \leq t \leq T \\ &= \sqrt{\frac{2\mathcal{E}}{T}} \cos(2\pi f_c t + 2\pi m \Delta f t),\end{aligned}$$

where the equivalent low-pass signal waveforms are defined as

$$s_{\ell m}(t) = \sqrt{\frac{2\mathcal{E}}{T}} e^{j2\pi m \Delta f t}, \quad m = 1, 2, \dots, M, \quad 0 \leq t \leq T.$$

This type of **frequency modulation** is called **frequency-shift keying (FSK)**.

- How should we choose the parameter  $\Delta f$ ?

## Memoryless modulation methods: orthogonal multidimensional signals

- Remember that we defined the **inner product** of two generally complex-valued signals  $s_{\ell k}(t)$  and  $s_{\ell m}(t)$  by

$$\langle s_{\ell k}(t), s_{\ell m}(t) \rangle = \int_a^b s_{\ell k}(t) s_{\ell m}^*(t) dt,$$

which can be viewed as a measure of similarity between  $s_{\ell k}(t)$  and  $s_{\ell m}(t)$ .

- Since the inner product depends on the energies of  $s_{\ell k}(t)$  and  $s_{\ell m}(t)$ , we exploit the Cauchy-Schwarz inequality for normalization according to

$$\rho_{km} = \frac{\langle s_{\ell k}(t), s_{\ell m}(t) \rangle}{||s_{\ell k}(t)|| ||s_{\ell m}(t)||},$$

where  $\rho_{km}$  is the so-called **cross-correlation coefficient** being, in general, a complex quantity. Clearly, from the Cauchy-Schwarz inequality, we have that, independently of the signal energies,

$$|\rho_{km}| \leq 1.$$

## Memoryless modulation methods: orthogonal multidimensional signals

- Show as an exercise that the cross-correlation coefficient and the real-valued signals  $s_k(t) = \Re\{s_{\ell k}(t)e^{j2\pi f_c t}\}$  are related to each other according to

$$\Re\{\rho_{km}\} = \frac{\langle s_k(t), s_m(t) \rangle}{\|s_k(t)\| \|s_m(t)\|}.$$

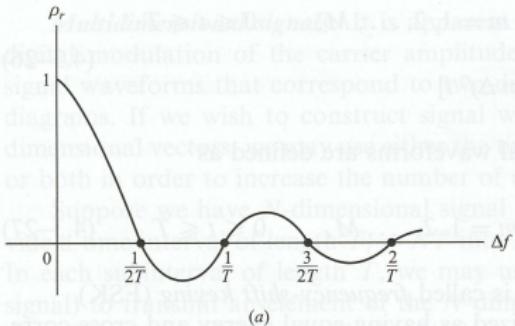
- The cross-correlation coefficient can be used to characterize the FSK signals. First, the signal waveforms have equal energy, where the cross-correlation coefficients are given by

$$\rho_{km} = \frac{2\mathcal{E}/T}{2\mathcal{E}} \int_0^T e^{j2\pi(k-m)\Delta f t} dt = \frac{\sin(\pi T(k-m)\Delta f)}{\pi T(k-m)\Delta f} e^{j\pi T(k-m)\Delta f},$$

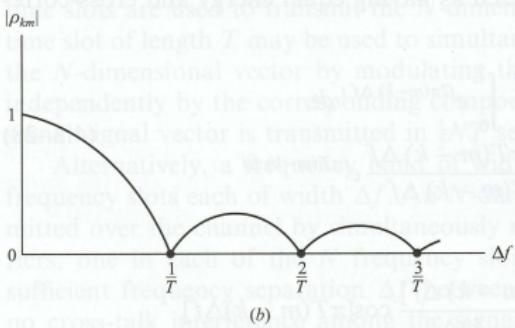
whose real parts are

$$\begin{aligned} \rho_r &\equiv \Re\{\rho_{km}\} = \frac{\sin(\pi T(k-m)\Delta f)}{\pi T(k-m)\Delta f} \cos \pi T(k-m)\Delta f = \frac{\sin(2\pi T(k-m)\Delta f)}{2\pi T(k-m)\Delta f} \\ &= \text{si}(2\pi T(k-m)\Delta f). \end{aligned}$$

## Memoryless modulation methods: orthogonal multidimensional signals

**FIGURE 4.3–7**

Cross-correlation coefficient as a function of frequency separation for FSK signals.



## Memoryless modulation methods: orthogonal multidimensional signals

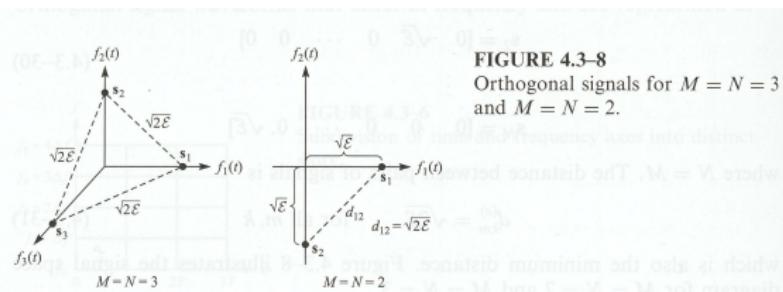
- Note from the figures, that  $|\rho_{km}| = 0$  for multiples of  $1/T$  whereas  $\Re \{\rho_{km}\} = 0$  for multiples of  $\frac{1}{2T}$ .
- We observe that  $\Re \{\rho_{km}\} = 0$  when  $\Delta f = \frac{1}{2T}$  and  $m \neq k$ .
- Since  $|m - k| = 1$  corresponds to adjacent frequency slots,  $\Delta f = \frac{1}{2T}$  represents the minimum frequency separation between adjacent signals for orthogonality of the  $M$  signals.
- For the case in which  $\Delta f = \frac{1}{2T}$ , the  $M$  FSK signals are equivalent to the  $M$ -dimensional vectors

$$\begin{aligned}\mathbf{s}_1 &= [\sqrt{\mathcal{E}} \quad 0 \quad 0 \quad \dots \quad 0 \quad 0] \\ \mathbf{s}_2 &= [0 \quad \sqrt{\mathcal{E}} \quad 0 \quad \dots \quad 0 \quad 0] \\ &\vdots \quad \vdots \\ \mathbf{s}_M &= [0 \quad 0 \quad 0 \quad \dots \quad 0 \quad \sqrt{\mathcal{E}}]\end{aligned}$$

- The distance between pairs of signals is identical to the minimum distance given by

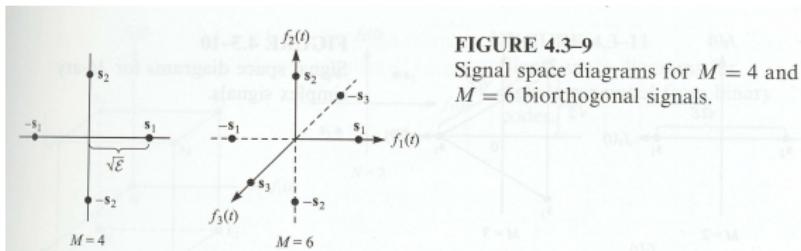
$$d_{km}^{(e)} = \sqrt{2\mathcal{E}}, \quad \text{for all } m, k.$$

## Memoryless modulation methods: orthogonal multidimensional signals



**FIGURE 4.3-8**  
Orthogonal signals for  $M = N = 3$   
and  $M = N = 2$ .

- A set of  $M$  biorthogonal signals can be constructed from  $\frac{1}{2}M$  orthogonal signals by simply including the negatives of the orthogonal signals. Thus, we require  $N = M/2$  dimensions for the construction of a set of  $M$  biorthogonal signals.



**FIGURE 4.3-9**  
Signal space diagrams for  $M = 4$  and  
 $M = 6$  biorthogonal signals.

## Simplex signals

- Suppose we have a set of  $M$  orthogonal waveforms  $\{s_m(t)\}$  or, equivalently, their vector representation  $\{\mathbf{s}_m\}$ .

- Their mean is

$$\bar{\mathbf{s}} = \frac{1}{M} \sum_{m=1}^M \mathbf{s}_m.$$

- Now let us construct another set of  $M$  signals by subtracting the mean from each of the  $M$  orthogonal signals. Thus,

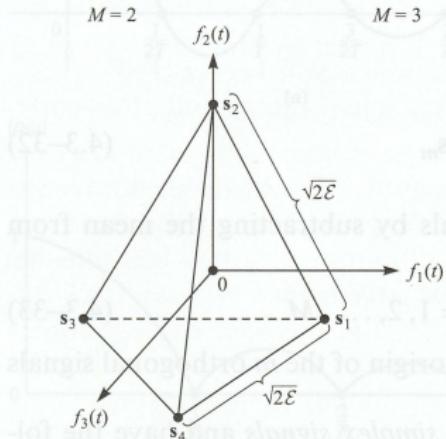
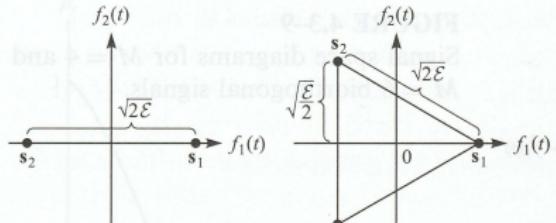
$$\mathbf{s}'_m = \mathbf{s}_m - \bar{\mathbf{s}}, \quad m = 1, 2, \dots, M.$$

- Obviously,  $\bar{\mathbf{s}}$  is the origin of the signal upon subtraction of  $\bar{\mathbf{s}}$  from  $\{s_m(t)\}$ .
- The resulting signal waveforms are called **simplex signals** and have the following properties which are to be shown as an exercise for all  $m$  and  $n$ :

$$\|\mathbf{s}'_m\|^2 = \mathcal{E} \left( 1 - \frac{1}{M} \right), \quad \Re \{ \rho_{mn} \} = \frac{\langle \mathbf{s}'_m, \mathbf{s}'_n \rangle}{\|\mathbf{s}'_m\| \|\mathbf{s}'_n\|} = -\frac{1}{M-1}, \quad d_{km}^{(e)} = \sqrt{2\mathcal{E}}.$$

- Note that the signal dimensionality is  $N = M - 1$ .

## Simplex signals



**FIGURE 4.3–10**  
Signal space diagrams for  $M$ -ary simplex signals.

## Signal waveforms from binary codes

- A set of  $M$  signaling waveforms can be generated from a set of  $M$  binary code words of the form

$$\mathbf{C}_m = [c_{m1} \ c_{m2} \ \cdots \ c_{mN}], \quad m = 1, 2, \dots, M,$$

where  $c_{mj} = 0$  or  $1$  for all  $m$  and  $j$ .

- Each component of a code word is mapped into an elementary BPSK waveform as follows:

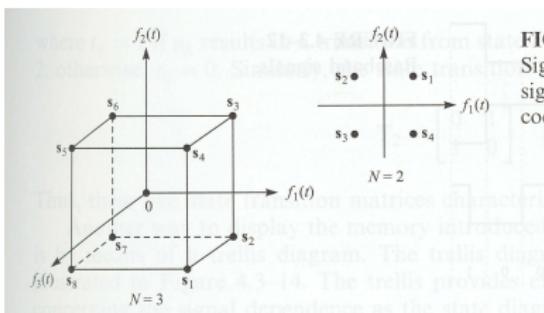
$$\begin{aligned} c_{mj} &= 1 \Rightarrow s_{mj}(t) = \sqrt{\frac{2\mathcal{E}_c}{T_c}} \cos(2\pi f_c t), \quad 0 \leq t \leq T_c \\ c_{mj} &= 0 \Rightarrow s_{mj}(t) = -\sqrt{\frac{2\mathcal{E}_c}{T_c}} \cos(2\pi f_c t), \quad 0 \leq t \leq T_c, \end{aligned}$$

where  $T_c = T/N$  and  $\mathcal{E}_c = \mathcal{E}/N$ . Thus, the  $M$  code words  $\{\mathbf{C}_m\}$  are mapped into a set of  $M$  waveforms  $\{s_m(t)\}$ .

- The waveforms can be represented in vector form with  $s_{mj} = \pm\sqrt{\mathcal{E}/N}$  as

$$\mathbf{s}_m = [s_{m1} \ s_{m2} \ \cdots \ s_{mN}], \quad m = 1, 2, \dots, M.$$

## Signal waveforms from binary codes



**FIGURE 4.3-11**  
Signal space diagrams for signals generated from binary codes.

- Note that there are  $2^N$  possible waveforms that can be constructed from the  $2^N$  possible binary code words. However, we may select only  $M < 2^N$  waveforms for the information transmission.
- Each of the  $M$  waveforms has energy  $\mathcal{E}$ . Clearly, the cross-correlation properties of the signal set depend on the selection of the  $M$  waveforms.
- Show as an exercise: **Adjacent signal points** have a cross-correlation coefficient  $\rho_r = \frac{\mathcal{E}(1-2/N)}{\mathcal{E}} = \frac{N-2}{N}$  and distance  $d^{(e)} = \sqrt{2\mathcal{E}(1 - \rho_r)} = 2\sqrt{\mathcal{E}/N}$ .

## Modulation with Memory

- Other modulation methods include, in particular, schemes **with memory**, where signaling in subsequent symbol periods depends on data symbol values in previous symbol intervals.
- The memory can be exploited to implement a spectral shaping and to take into account implementation issues in transceivers.
- Spectral shaping is concerned with a reduction of side lobes in the power density spectrum of the transmitted signal. Side lobes with a fast decay improve spectral efficiency since the spectrum adjacent to the signal spectrum at hand can be reused with less interference.
- Clearly, the signal space representation of a modulation scheme with memory is more complex than for a memoryless scheme.
- At the receiver, unlike in the case of memoryless modulation, we have to **detect a sequence of symbols** rather than an isolated symbol. The individual symbols can be thought of labels of the corresponding sequences or states of a so-called **trellis diagram**.
- These schemes being used in today's wireless systems will be treated in the lecture **Introduction to Information Theory and Coding**.

# Table of Contents

## 1 Introduction

- Mathematical Models for Communication Channels
- Generalized Functions

## 2 Probability

- Fundamentals in Probability Theory
- Useful Probability Distributions
- Central Limit Theorem

## 3 Fourier Transform

- Properties of FT
- Sampling Theorem

## 4 Stochastic Processes

- Gaussian Processes
- Response of an LTI System to a Stationary Input Signal
- Discrete-Time (Stochastic) Signals
- Cyclostationary Processes

## 5 Communication Signals and Systems

- Representation of Band-pass Signals and Systems
- Orthogonal Expansions of Signals
- Representation of Digitally Modulated Signals

## 6 Optimum Receivers for the AWGN Channel

- Characterization of Thermal Noise
- Signal Space Representation
- Optimum Detection

# Optimum Receivers for the AWGN Channel

- In the cases considered below, the transmitted signal  $s(t)$  propagates through a **linear time-invariant channel** characterized by its **channel impulse response  $h(t)$**  and is received by a certain **terminal** called the **receiver**.
- The **objective** of the receiver is to

detect the transmitted symbols with minimum probability of error.

- The channel output is given by

$$\rho(t) = h(t) \star s(t) = \int_{\mathbb{R}} h(\tau)s(t - \tau)d\tau,$$

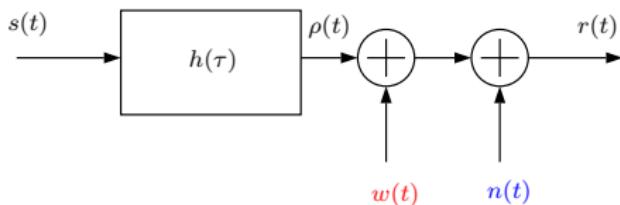
where we can consider both real-valued signals or signals in the complex baseband.

- Since the signal usually gets amplified prior to further signal processing, a **disturbance  $n(t)$**  is superimposed to  $\rho(t)$  which is generated by the receiver itself.
- Sometimes, in particular in wireless communications, we also encounter **interference  $w(t)$**  from other transmitters whose signal formats are either partially or completely unknown to the receiver.

- Thus, for a given transmitted signal  $s(t)$ , we have the following description of the received signal:

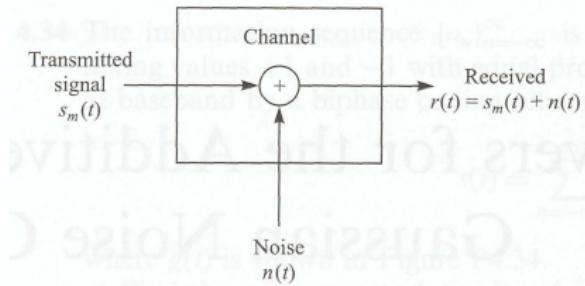
$w(t)$  ... interference from other information signals  
 (in the same frequency band at the same time)

$n(t)$  ... thermal noise generated in the receiver amplifier



- The main principles for the receiver design can be studied based on a simplified model. Therefore, in the following, we consider systems with ideal channels  $h(t) = \delta(t)$  and no interference, i.e.  $w(t) \equiv 0$ .

- Assuming that  $s(t) = s_m(t)$ , we obtain the following signal flow graph:



**FIGURE 5.1-1**

Model for received signal passed through an AWGN channel.

- Here,  $s_m(t)$  is one of  $M$  signal waveforms with  $m = 1, 2, \dots, M$ , where each waveform is transmitted within the symbol interval of duration  $T$  with  $0 \leq t \leq T$ , i.e.

$$r(t) = s_m(t) + n(t).$$

- How can we characterize the noise process  $n(t)$ ?

# Characterization of Thermal Noise

- The spectral power density of the thermal noise generated by a receiver's front-end amplifier in [W/Hz] can be approximated by

$$N_0(f) = h f \left[ \frac{1}{\exp\left(\frac{hf}{k_B T}\right) - 1} + 1 \right].$$

$h = 6.6256 \cdot 10^{-34}$  Js

...

Planck's constant

$k_B = 1.3805 \cdot 10^{-23}$  J/K

...

Boltzmann's constant

$f$

...

frequency in [Hz]

$T$

...

absolute temperature in [K]

- For sufficiently small frequencies ( $f < 10\text{GHz}$ ), we can use a Taylor series expansion to obtain the approximation  $\exp\left(\frac{hf}{k_B T}\right) \approx 1 + \frac{hf}{k_B T}$  and thus

$$N_0 \approx k_B T = \text{const.}$$

Further assumptions on the noise process:

- The complex baseband process  $z(t)$  is zero-mean and has uncorrelated quadrature components.
- Due to the central limit theorem, the quadrature components are **Gaussian distributed**.
- The spectral noise power density of the complex baseband process (in the frequency domain) is  $N_0$ . As a result, we obtain the autocorrelation function of the equivalent low-pass process  $z(t)$  as

$$\phi_{zz}(\tau) = \frac{1}{2} E[z^*(t)z(t + \tau)] \approx \frac{N_0}{2} \delta(\tau) = \phi_{xx}(\tau) = \phi_{yy}(\tau) = \phi_{nn}(\tau)$$

and the power spectral density of the **additive white Gaussian noise (AWGN)**  $n(t)$  as

$$\Phi_{nn}(f) = \frac{1}{2} N_0.$$

# Signal Space Representation

- The derivation of the optimum receiver minimizing the probability of a symbol error would require advanced concepts from measure theory. Parts of the derivation are treated in the lecture **Introduction to Signal Detection and Estimation**. Here, we will use certain heuristics in order to derive the optimum receiver.

## Correlator Demodulator

- To this end, we first represent the received signal in a signal space spanned by a set of  $K$  orthonormal basis functions  $\{f_k(t)\}$ , which contains all possible signals  $s_m(t)$  for  $1 \leq m \leq M$  and  $K \leq M$ .
- Can the noise process  $n(t)$  also represented in that signal space? Put differently, what is the **dimension of  $n(t)$ ?**
- In a  $K$ -dimensional space, the component of  $n(t)$  in the direction of  $f_k(t)$  for  $k = 1, \dots, K$  is given by  $n_k = \langle n(t), f_k(t) \rangle$ . Since the inner product is a linear transformation of  $n(t)$ , the quantities  $n_k$  for  $k = 1, 2, \dots, K < \infty$  are jointly Gaussian distributed and thus completely characterized by their **mean vector** and their **covariance matrix**.

We obtain

$$\begin{aligned} E[n_k] &= \int_0^T E[n(t)] f_k(t) dt = 0 \\ E[n_k n_m] &= \int_0^T \int_0^T E[n(t)n(\tau)] f_k(t) f_m(\tau) dt d\tau = \frac{1}{2} N_0 \int_0^T \int_0^T \delta(t - \tau) f_k(t) f_m(\tau) dt d\tau \\ &= \frac{1}{2} N_0 \int_0^T f_k(t) f_m(t) dt = \frac{1}{2} N_0 \delta_{km} \end{aligned}$$

with  $\delta_{km}$  denoting the Kronecker delta, i.e.  $\delta_{km} = 1$  if  $k = m$  and zero otherwise.

- We conclude that for arbitrary basis functions  $\{f_k(t)\}$  and arbitrary  $K$ , the elements of  $\{n_k\}$  are zero-mean uncorrelated (and thus mutually independent) Gaussian random variables with a common variance  $\sigma_n^2 = \frac{1}{2}N_0$ .
- Obviously,  $n(t)$  is of infinite dimension and, thus, we cannot represent it in a finite dimensional space spanned by the basis functions  $\{f_k(t)\}$  with  $k = 1, 2, \dots, K$ .
- This is another way of representing the infinite power of the process  $n(t)$  given by

$$E[n^2(t)] = \lim_{K \rightarrow \infty} \sum_{k=1}^K E[n_k^2] = \lim_{\tau \rightarrow 0} \frac{1}{2} N_0 \delta(\tau) = \infty,$$

where the limits do not exist in the mathematical sense, but represent a sloppy way of characterizing the infinite power of  $n(t)$ .

- Consider now the projection of  $r(t)$  onto the basis functions  $\{f_k(t)\}$ .

- Arranging the components in vectors, we have

$$\mathbf{r} = \mathbf{s}_m + \mathbf{n} \quad \text{for } m = 1, 2, \dots, M$$

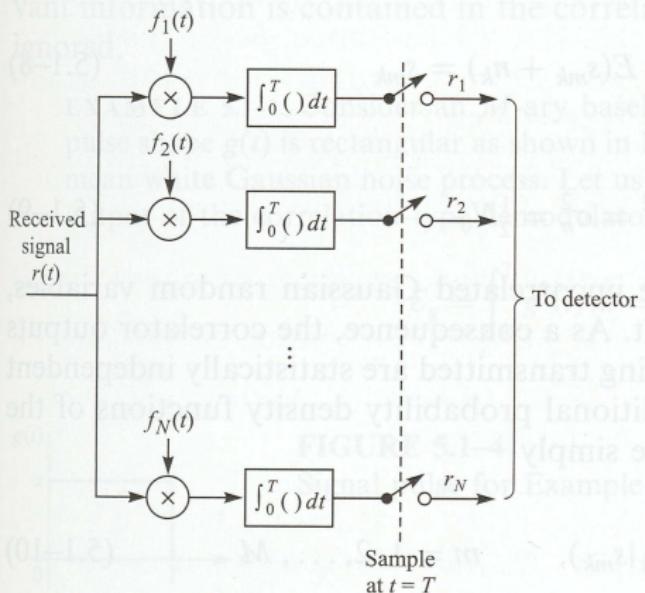
with

$$\begin{aligned}\mathbf{r} &= [r_1, \dots, r_K]^T \\ \mathbf{s}_m &= [s_{m1}, \dots, s_{mK}]^T \\ \mathbf{n} &= [n_1, \dots, n_K]^T.\end{aligned}$$

- Clearly (show as an exercise), the noise part  $n'(t) = n(t) - \sum_{k=1}^K n_k f_k(t)$  outside the signal space spanned by  $\{f_k(t)\}$  is uncorrelated with the  $k$  correlator outputs  $r_k$  for  $k = 1, 2, \dots, K$ . Thus

$$E[n'(t)r_k] = 0 \quad \text{for } k = 1, 2, \dots, K.$$

## Implementation using correlators:



**FIGURE 5.1-3**

Correlation-type demodulator.

Correlators<sup>1</sup>.

<sup>1</sup>In Fig 5.1-3  $N = K$

## Matched Filter Demodulator

- In many cases, where synchronisation (e.g. estimation of start/end of the interval  $[0, T]$ ) is carried out based on inner products, we have to calculate the quantities  $\int_0^T r(t)s_m(t - \tau)dt$  for several values of  $\tau$ .
- For  $\tau = 0$ , these quantities represent the demodulator outputs.
- While the values  $\int_0^T r(t)s_m(t - \tau)dt$  for  $\tau = \tau_1, \tau_2, \dots, \tau_K$  could be calculated by a number of  $K$  inner products or so-called **correlators**, a more efficient structure to calculate the quantities exists. This structure is based on our intuition that the sequence of correlation values resembles a sequence of values observed at the output of a filter.
- Instead of using a large number of correlators, we **consider a filter** with impulse response

$$h(t) = s_m(T - t), \quad 0 \leq t \leq T,$$

where  $h(t) = 0$  outside the interval  $t \in [0, T]$ . The **filter output** is

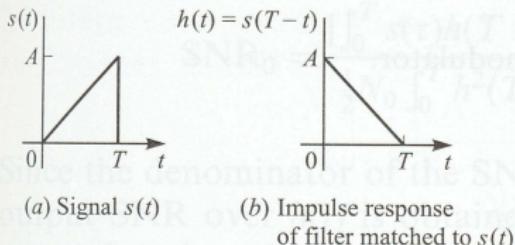
$$y(t) = \int_{t-T}^t r(\tau)h(t - \tau)d\tau = \int_{t-T}^t r(\tau)s_m(T - t + \tau)d\tau.$$

- Suppose the filter output is sampled at time  $t = T$ . In this case, we obtain

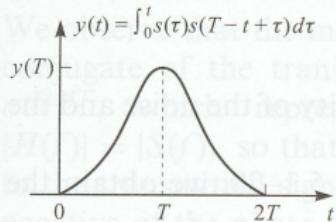
$$y(t) = \int_0^T r(\tau) s_m(\tau) d\tau = \langle r(t), s_m(t) \rangle,$$

i.e. the correlator output.

- Since the output of the filter is essentially the time-autocorrelation function of  $s(t)$ , which peaks at  $t = T$ , we call the filter **the matched filter to the signal  $s(t)$** .



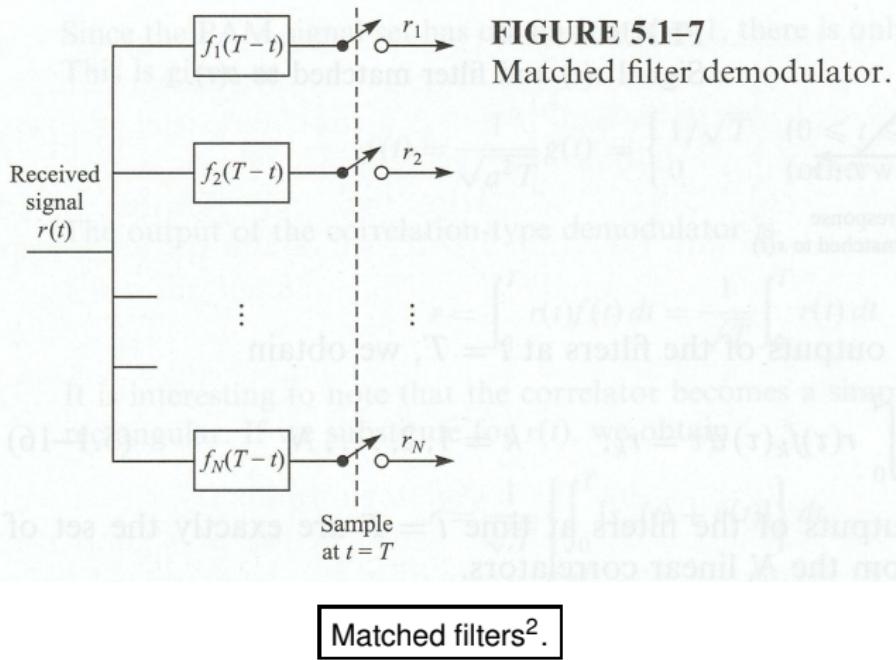
**FIGURE 5.1-5**  
Signal  $s(t)$  and filter matched to  $s(t)$ .

**FIGURE 5.1–6**

The matched filter output is the autocorrelation function of  $s(t)$ .

- If we want to calculate explicitly the components of  $r(t)$  in the signal space, we have to project  $r(t)$  onto the basis functions  $\{f_k(t)\}$ ,  $k = 1, 2, \dots, K$ , where  $K$  is the signal space dimension.
- Again, we can implement the inner products either as **correlators** or **filters matched to the basis functions  $\{f_k(t)\}$** .

## Implementation using matched filters:




---

<sup>2</sup>In Fig 5.1-7  $N = K$

# Optimum Detection

- Since  $n'(t)$  and  $\{r_k\}$  are Gaussian and uncorrelated, they are also statistically independent. Consequently,  $n'(t)$  does not contain any information that is relevant to the decision as to which signal waveform was transmitted.
- As a result, we can ignore  $n'(t)$  in the optimum decision rule and base the decision exclusively on the observation vector  $\mathbf{r}$ .
- From the viewpoint of the receiver, any of the  $M$  signals might be the one which was transmitted. As a consequence, we might say that the receiver has  $M$  hypotheses about the received signal, where each hypothesis is characterized by a conditional probability density function  $p(\mathbf{r} | \mathbf{s}_m)$ ,  $m = 1, 2, \dots, M$ , of the observables  $\mathbf{r}$  under the corresponding hypothesis  $\mathbf{s}_m$ .
- We write  $\mathbf{r} \sim p(\mathbf{r} | \mathbf{s}_m)$  to denote that  $\mathbf{r}$  is distributed according to  $p(\mathbf{r} | \mathbf{s}_m)$ .
- Thus, the optimum receiver design is based on choosing the hypothesis providing the least probability of error. The task is to separate optimally the set of possible observations into  $M$  sets where upon observation of any member of set  $m$  we decide in favor of hypothesis  $m$ .
- The decision process is also termed detection.

**Proposition** Consider  $M$  equiprobable hypotheses under which an observed random vector  $\mathbf{r} \in \Omega$  has density functions

$$\begin{aligned}\mathcal{H}_1 &: \mathbf{r} \sim p(\mathbf{r} | \mathbf{s}_1) \\ \mathcal{H}_2 &: \mathbf{r} \sim p(\mathbf{r} | \mathbf{s}_2) \\ &\vdots \\ \mathcal{H}_M &: \mathbf{r} \sim p(\mathbf{r} | \mathbf{s}_M).\end{aligned}$$

Then the following decision regions minimize error probability:

$$\Omega_m = \left\{ \mathbf{r} : p(\mathbf{r} | \mathbf{s}_m) = \max_{n=1, \dots, M} p(\mathbf{r} | \mathbf{s}_n) \right\} - \bigcup_{n=1}^{m-1} \Omega_n, \quad m = 1, \dots, M.$$

**Proof:** The probability of error  $P$  is given by

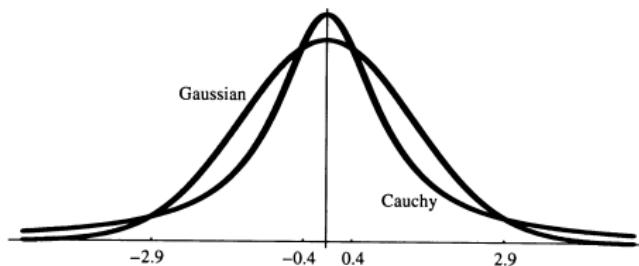
$$\begin{aligned} P &= 1 - \frac{1}{M} \sum_{n=1}^M P(\mathbf{r} \in \Omega_n | \mathbf{s}_n) \\ &= 1 - \frac{1}{M} \sum_{n=1}^M \int_{\Omega_n} p(\mathbf{r} | \mathbf{s}_n) d\mathbf{r} \\ &\geq 1 - \frac{1}{M} \int_{\Omega} \max_{n=1, \dots, M} p(\mathbf{r} | \mathbf{s}_n) d\mathbf{r}. \end{aligned}$$

Therefore, we have to choose

$$\Omega_m = \left\{ \mathbf{r} : p(\mathbf{r} | \mathbf{s}_m) = \max_{n=1, \dots, M} p(\mathbf{r} | \mathbf{s}_n) \right\} - \bigcup_{n=1}^{m-1} \Omega_n, \quad m = 1, \dots, M$$

to satisfy the lower bound with equality.

**Example:** Consider a binary hypothesis test between a Gaussian and Cauchy distributed scalar observable  $r$



$$H_1 : R \sim p(r | s_1) = \frac{1/\pi}{1+r^2},$$

$$H_2 : R \sim p(r | s_2) = \frac{1}{2\sqrt{\pi}} \exp(-r^2/4).$$

From the proposition  $\Rightarrow$

- choose  $H_1$ , if  $r \in \Omega_1 = \mathbb{R} \setminus \{(-2.92, -0.42) \cup (0.42, 2.92)\}$
- choose  $H_2$ , if  $r \in \Omega_2 = \{(-2.92, -0.42) \cup (0.42, 2.92)\}$ .

- For the AWGN channel, we have

$$\mathbf{r} = \mathbf{s}_m + \mathbf{n} \quad \text{for } m = 1, 2, \dots, M,$$

so that the conditional densities in the optimum decision rule, in view of the independent components of  $\mathbf{n}$ , are given by

$$p(\mathbf{r} | \mathbf{s}_m) = p_{\mathbf{n}}(\mathbf{r} - \mathbf{s}_m) = \prod_{k=1}^K p_{n_k}(r_k - s_{mk}).$$

- Note that  $p_{n_k}(n) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{n^2}{N_0}\right]$  so that

$$p(\mathbf{r} | \mathbf{s}_m) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{\|\mathbf{r} - \mathbf{s}_m\|^2}{N_0}\right].$$

- As a result, maximizing  $p(\mathbf{r} | \mathbf{s}_m)$  w.r.t.  $\mathbf{m}$  corresponds to minimizing  $\|\mathbf{r} - \mathbf{s}_m\|^2$ .
- We refer to this decision rule as minimum distance detection.
- Denote the quantity representing the detected index as

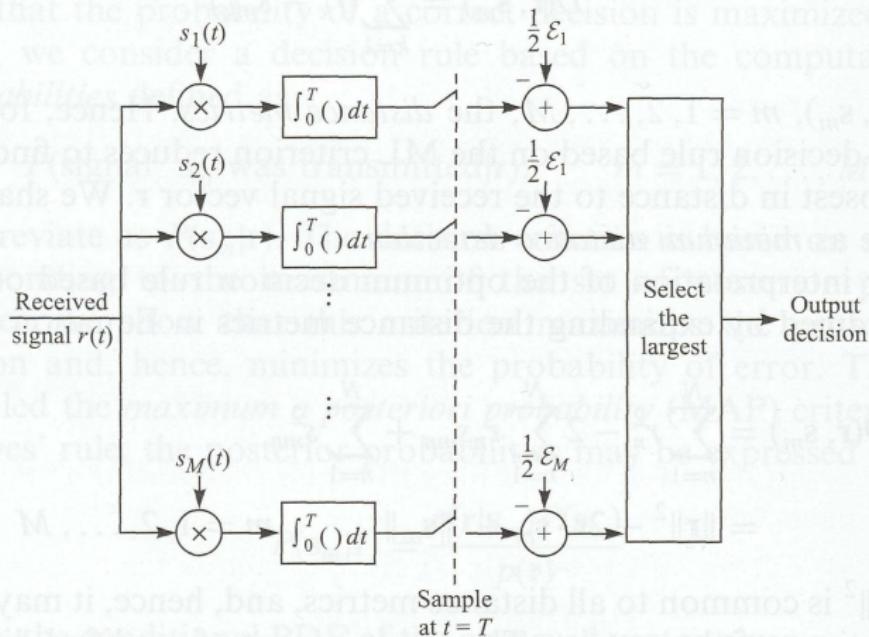
$$\begin{aligned}\hat{m} &= \arg \min_{m=1, \dots, M} \|\mathbf{r} - \mathbf{s}_m\|^2 = \arg \max_{m=1, \dots, M} \left\{ 2 \langle \mathbf{r}, \mathbf{s}_m \rangle - \|\mathbf{s}_m\|^2 \right\} \\ &= \arg \max_{m=1, \dots, M} \left\{ \langle \mathbf{r}, \mathbf{s}_m \rangle - \frac{1}{2} \|\mathbf{s}_m\|^2 \right\},\end{aligned}$$

where  $\arg \max_{m=1, \dots, M}$  denotes the index argument of the maximum.

- Obviously, the optimum decision rule first calculates the inner products  $\langle \mathbf{r}, \mathbf{s}_m \rangle = \int_0^T r(t)s_m(t)dt$ , i.e. demodulates the received signal, and then finds the optimum detection as a function of these inner products.



**FIGURE 5.1-2**  
Receiver configuration.

**FIGURE 5.1-9**

An alternative realization of the optimum AWGN receiver.

## Performance of optimum detection for binary equal-energy transmission

- Consider a binary transmission, i.e. a transmission of either of two waveforms, with equal energy  $\mathcal{E}_b$  and cross-correlation coefficient  $\rho_r = \Re\{\rho_{12}\}$ . Clearly,  $\mathcal{E}_b$  can be thought of as the **energy per bit at the receiver**.
- The model of the received signal is

$$r(t)|_{s(t)=s_i(t)} = s_i(t) + n(t), \quad i = 1, 2,$$

where  $s_i$  and  $n(t)$  denote the transmitted signal and the AWGN, respectively. As before, we assume

$$\phi_{nn}(\tau) = \frac{N_0}{2} \delta(\tau).$$

- Since the energies of both signals are identical, the optimum receiver computes the quantities

$$U_i = \langle r(t), s_i(t) \rangle, \quad i = 1, 2,$$

and chooses the argument of the maximum of  $U_i$  as the estimate of the transmitted symbol.

## Performance of optimum detection for binary equal-energy transmission

- Suppose that  $s_1(t)$  has been transmitted. The quantities  $U_i$ , as linear mappings of the Gaussian distributed stochastic process  $r(t)$ , are also Gaussian random variables with

$$\begin{aligned} U_1 &= \mathcal{E}_b + \langle n(t), s_1(t) \rangle \\ U_2 &= \mathcal{E}_b \rho_r + \langle n(t), s_2(t) \rangle. \end{aligned}$$

- The optimum decision can be formulated, upon calculating the Gaussian distributed decision variable

$$V = U_1 - U_2 = \mathcal{E}_b(1 - \rho_r) + \langle n(t), s_1(t) - s_2(t) \rangle \sim \mathcal{N}(\mu_V; \sigma_V^2),$$

as the comparison of  $V \stackrel{?}{<} 0$  and deciding correspondingly:

- $> \Leftrightarrow s_1(t)$  has been transmitted
- $= \Leftrightarrow s_1(t)$  or  $s_2(t)$  has been transmitted
- $< \Leftrightarrow s_2(t)$  has been transmitted.

## Performance of optimum detection for binary equal-energy transmission

- We find  $\mu_V = \mathcal{E}_b(1 - \rho_r)$  and

$$\begin{aligned}
 \sigma_V^2 &= E[(\langle n(t), s_1(t) - s_2(t) \rangle)^2] \\
 &= E\left[\int_0^T \int_0^T n(t_1)n(t_2)(s_1(t_1) - s_2(t_1))(s_1(t_2) - s_2(t_2)) dt_1 dt_2\right] \\
 &= \frac{N_0}{2} \int_0^T \int_0^T \delta(t_1 - t_2)(s_1(t_1) - s_2(t_1))(s_1(t_2) - s_2(t_2)) dt_1 dt_2 \\
 &= \frac{N_0}{2} \int_0^T (s_1(t) - s_2(t))^2 dt = \frac{N_0}{2} (\mathcal{E}_b - 2\rho_r \mathcal{E}_b + \mathcal{E}_b) \\
 &= N_0 \mathcal{E}_b (1 - \rho_r).
 \end{aligned}$$

- Since we assume that  $s_1(t)$  has been transmitted,  $P(V < 0)$  is the conditional probability of error (conditioned on  $s_1(t)$  being transmitted).

## Performance of optimum detection for binary equal-energy transmission

- The conditional error probability thus results to

$$\begin{aligned}
 P(V < 0) &= \int_{-\infty}^0 p(v) dv = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sigma_V} \exp\left(-\frac{1}{2\sigma_V^2}(v - \mu_V)^2\right) dv \\
 &\stackrel{x=\frac{v-\mu_V}{\sigma_V}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\mu_V/\sigma_V} \exp\left(-\frac{x^2}{2}\right) dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{\mathcal{E}_b(1-\rho_r)}{N_0}}}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = Q\left(\sqrt{\frac{\mathcal{E}_b(1-\rho_r)}{N_0}}\right).
 \end{aligned}$$

- Here,  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-t^2/2) dt$  denotes the ***Q-function*** describing the **tail probability of a standard  $\mathcal{N}(0; 1)$  distribution.**

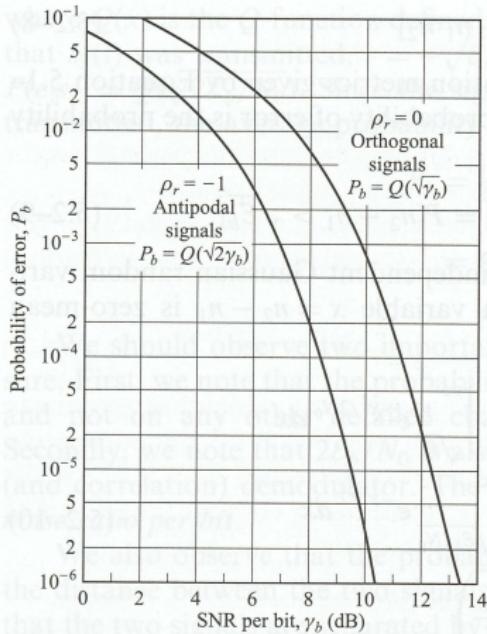
## Performance of optimum detection for binary equal-energy transmission

- Due to symmetry conditions, the **average probability of error** for equally likely signals  $s_1(t)$  and  $s_2(t)$  is given by

$$P_b = \frac{1}{2}P(V < 0 | s_1(t)) + \frac{1}{2}P(V > 0 | s_2(t)) = Q\left(\sqrt{\frac{\mathcal{E}_b(1 - \rho_r)}{N_0}}\right).$$

- Obviously, the average error probability depends only on
  - ① the **signal-to-noise ratio (SNR)** per bit  $\gamma_b$  defined as  $\gamma_b = \mathcal{E}_b/N_0$
  - ② the **signal correlation coefficient**  $\rho_r$ .
- Since the tail probability decreases monotonically for increasing arguments of the  $Q$ -function, **the optimum binary signals for a given  $\mathcal{E}_b$  are antipodal with  $\rho_r = -1$ .**
- Thus, for BPSK signal transmission, the average probability of error is  
 $P_b = Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)$ , while for binary orthogonal transmission with  $\rho_r = 0$ , we have  
 $P_b = Q\left(\sqrt{\frac{\mathcal{E}_b}{N_0}}\right)$ , i.e. a loss of  $10 \log 2 \approx 3$  dB in SNR as compared to BPSK transmission.

## Performance of optimum detection for binary equal-energy transmission



**FIGURE 5.2-4**  
Probability of error for binary signals.

## Performance of optimum detection for binary equal-energy transmission

- Finally, we want to express the average probability of error by the distance properties of the involved signals. The distance of the two equal-energy signals  $s_1(t)$  and  $s_2(t)$  is given by

$$\begin{aligned}d_{12}^2 &= \int_0^T [s_1(t) - s_2(t)]^2 dt \\&= \mathcal{E}_b - 2\rho_r \mathcal{E}_b + \mathcal{E}_b = 2\mathcal{E}_b(1 - \rho_r).\end{aligned}$$

- Upon inserting the distance expression into the average probability of error, we obtain

$$P_b = Q\left(\sqrt{\frac{d_{12}^2}{2N_0}}\right).$$

We conclude that the performance improves for an increasing distance between  $s_1(t)$  and  $s_2(t)$ .