# Signals and Systems for Bioinformatics

*Sistemi e segnali per la bioinformatica*

Gloria Menegaz

AA 2009-2010

# Scheduling

- October to December
- Timetable
  - Monday 16:30-18:00 theory
  - Wed. 9:00-10:30 theory
  - Thursday/Friday 13:30-15:30 lab. (exercises and Matlab sessions)
- Objective
  - Acquisition of the theoretical and practical basis of digital signal processing focusing on aspects that are relevant to bioinformatics
    - Processing of medical signals (1D, 2D, 3D, 3D+time), DNA sequences, microscopy imaging
- Background
  - Mathematical and statistical basis
    - Functional analysis
    - Notions of mean, variance, and other basic statistical concepts
- Desirable: take also the “Image processing for bioinformatics” course

# Didactic material

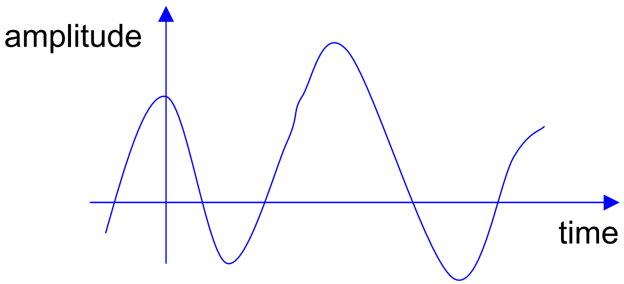
- Textbook
  - Signal Processing and Linear Systems, B.P. Lathi, CRC Press
- Other books
  - Signals and Systems, Richard Baraniuk’s lecture notes, available on line
  - Digital Signal Processing (4th Edition) (Hardcover), John G. Proakis, Dimitris K Manolakis
  - Teoria dei segnali analogici, M. Luise, G.M. Vitetta, A.A. D’Amico, McGraw-Hill
  - Signal processing and linear systems, Schaun's outline of digital signal processing
- All textbooks are available at the library
- Handwritten notes will be available on demand

# Signals&Systems

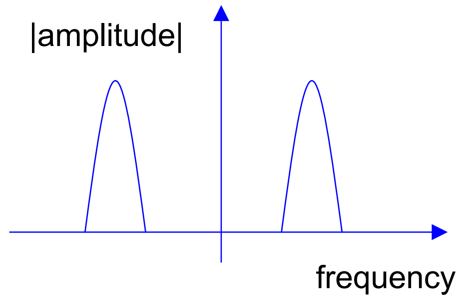
Input signal

Output signal

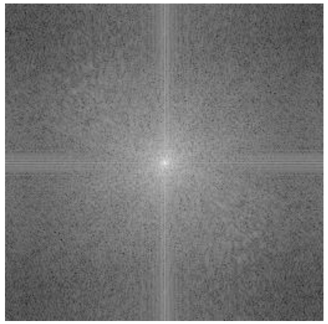
System




Linear time invariant systems (LTIS)




LTIS perform any kind of processing on the input data to generate output data



# Contents

## Signals

- Signal classification and representation
  - Types of signals
  - Sampling theory
  - Quantization
- Signal analysis
  - Fourier Transform
    - Continuous time, Fourier series, Discrete Time Fourier Transforms, Windowed FT
  - Spectral Analysis

## Systems

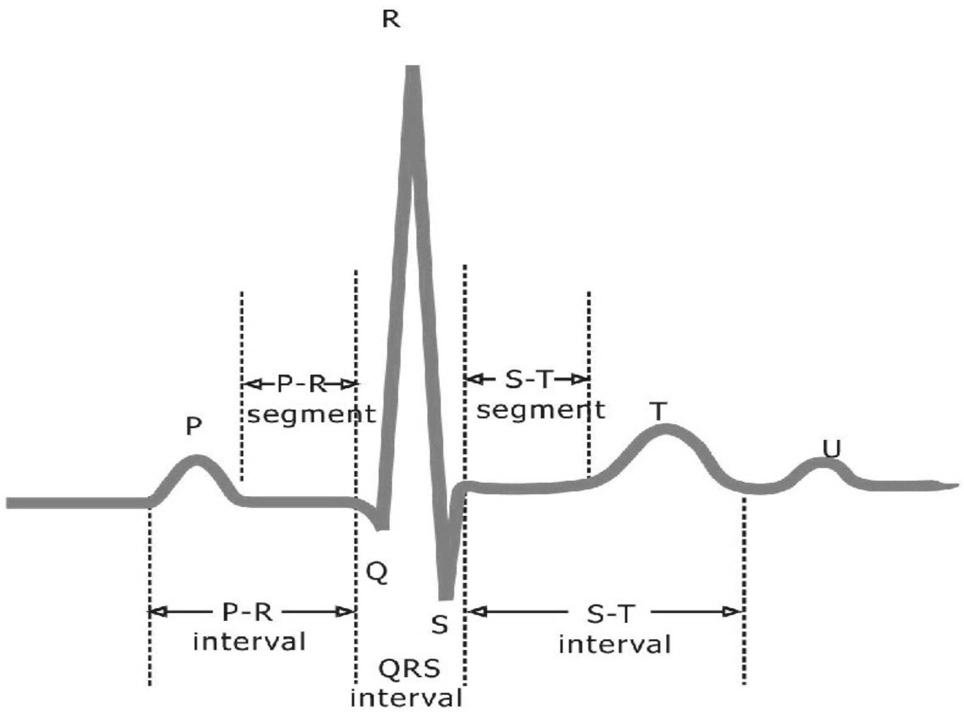
- Linear Time-Invariant Systems
  - Time and frequency domain analysis
  - Impulse response
  - Stability criteria
- Digital filters
  - Finite Impulse Response (FIR)
- Mathematical tools
  - Laplace Transform
    - Basics
  - Z-Transform
    - Basics

Applications in the domain of Bioinformatics

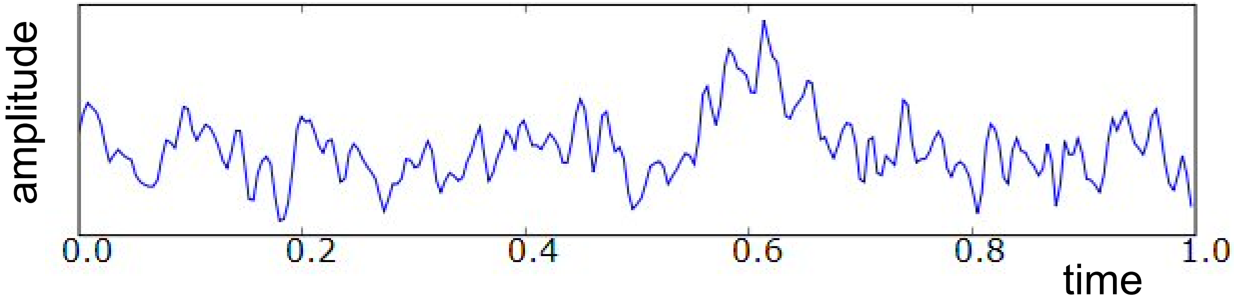
# What is a signal?

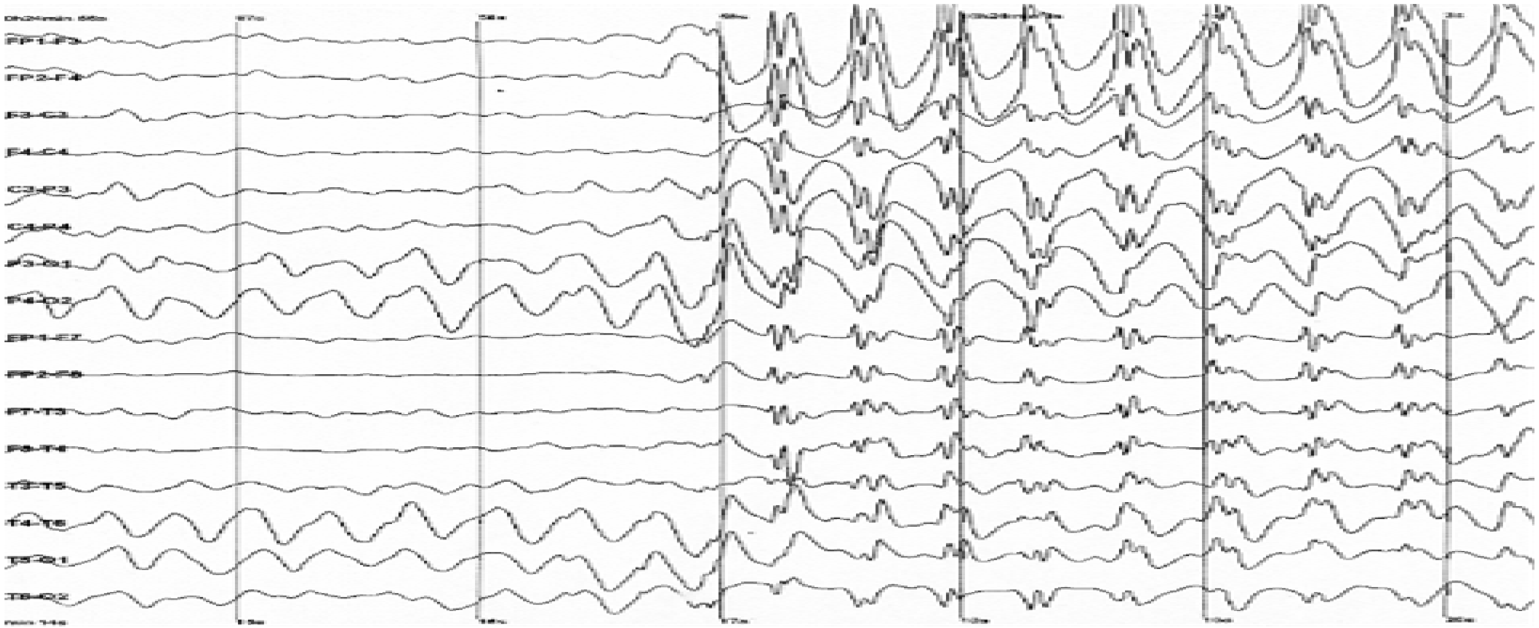
- A signal is a set of information of data
  - Any kind of physical variable subject to variations represents a signal
  - Both the independent variable and the physical variable can be either scalars or vectors
    - Independent variable: time (t), space (x, $\mathbf{x}=[x_1,x_2]$, $\mathbf{x}=[x_1,x_2,x_3]$)
    - Signal:
    - Electrochardiography signal (EEG) 1D, voice 1D, music 1D
    - Images (2D), video sequences (2D+time), volumetric data (3D)

# Example: 1D biological signals: ECG



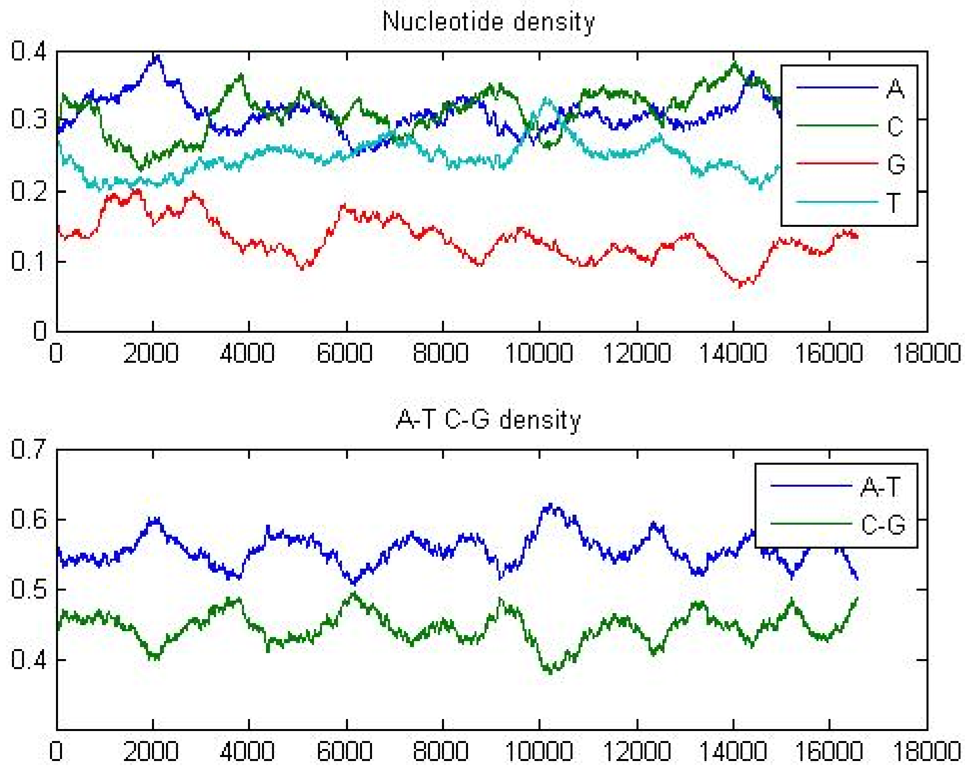
# Example: 1D biological signals: EEG



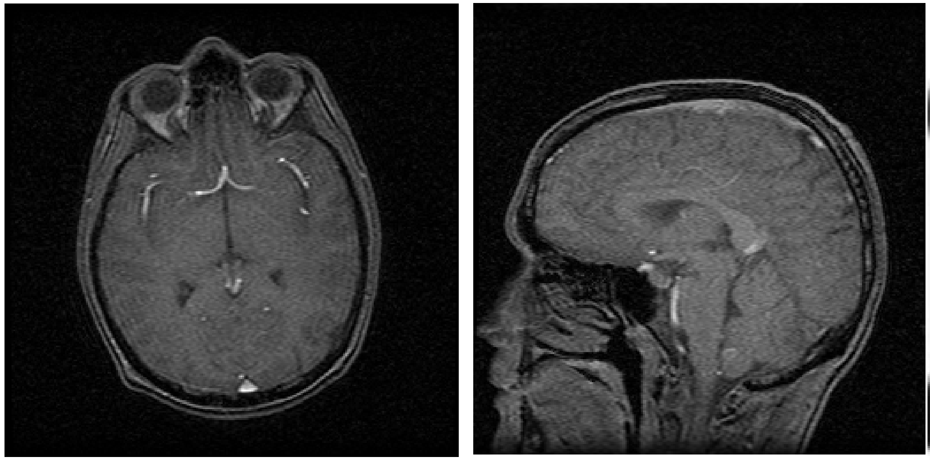
# 1D biological signals: DNA sequencing

GATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCATG……

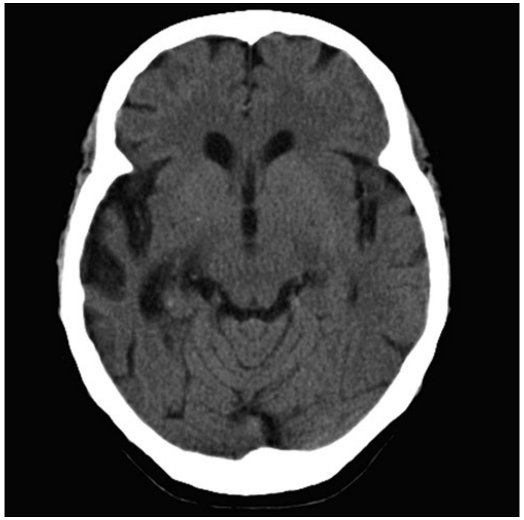


# Example: 2D biological signals: MI

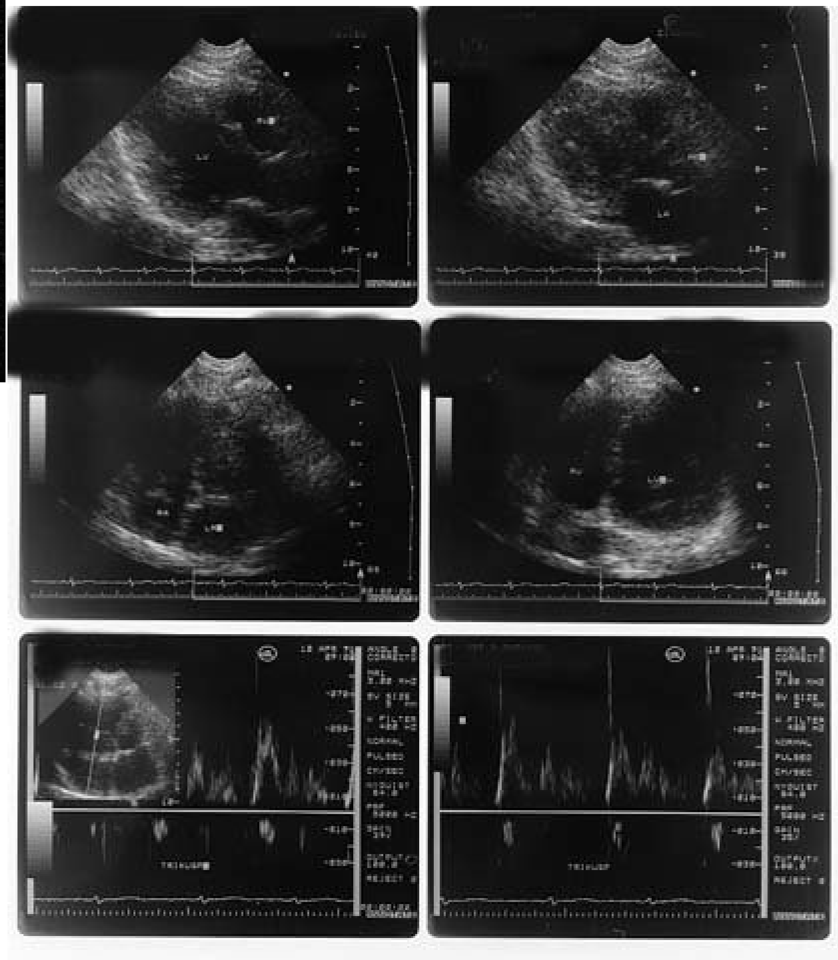
MRI



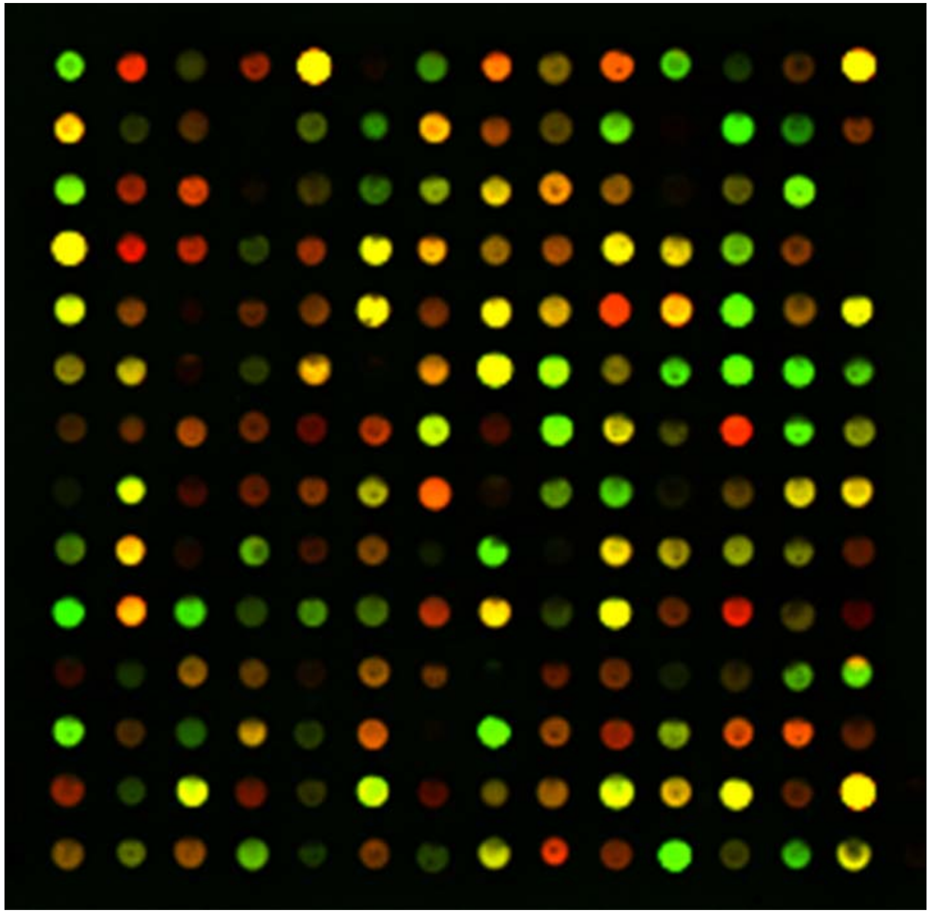
CT



US



# Example: 2D biological signals: microarrays



# Signals as functions

- Continuous functions of real independent variables
  - 1D: $f=f(x)$
  - 2D: $f=f(x,y)\ x,y$
  - Real world signals (audio, ECG, images)
- Real valued functions of discrete variables
  - 1D: $f=f[k]$
  - 2D: $f=f[i,j]$
  - *Sampled* signals
- Discrete functions of discrete variables
  - 1D: $f^d=f^d[k]$
  - 2D: $f^d=f^d[i,j]$
  - *Sampled and quantized* signals

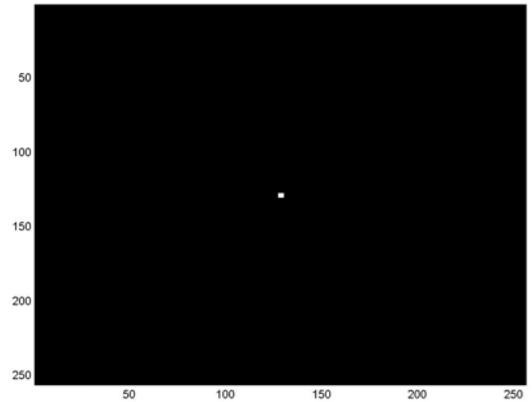
# Images as functions

- Gray scale images: 2D functions
  - Domain of the functions: set of $(x,y)$ values for which $f(x,y)$ is defined : 2D lattice $[i,j]$ defining the pixel locations
  - Set of values taken by the function : gray levels

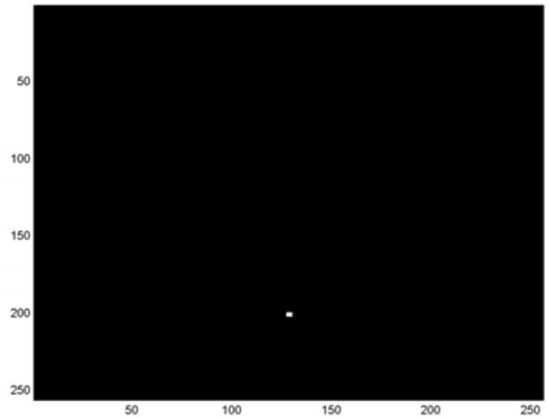
- Digital images can be seen as functions defined over a discrete domain $\{i,j: 0<i<I, 0<j<J\}$
  - $I,J$: number of rows (columns) of the matrix corresponding to the image
  - $f=f[i,j]$: gray level in position $[i,j]$

# Example 1: δ function

$$\delta[i,j]=\begin{cases}1 & i=j=0\\ 0 & i,j\neq 0; i\neq j\end{cases}$$



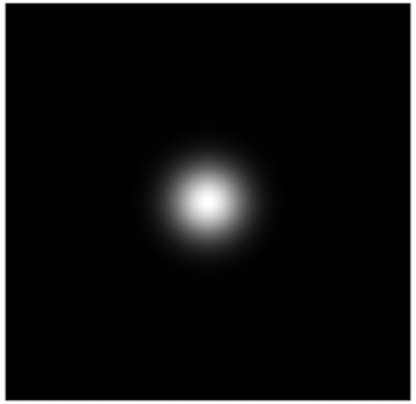

$$\delta[i,j-J]=\begin{cases}1 & i=0; j=J\\ 0 & otherwise\end{cases}$$



# Example 2: Gaussian

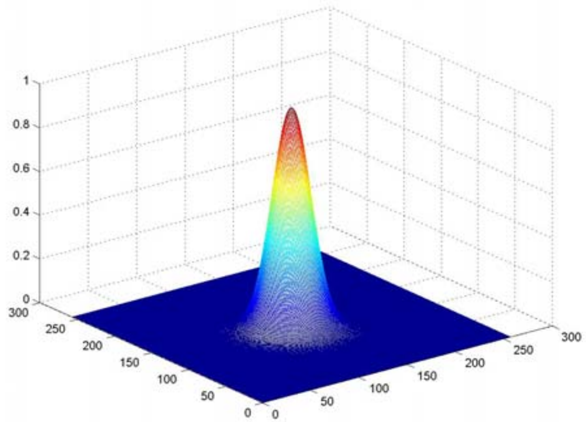
Continuous function

$$f(x,y)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{x^2+y^2}{2\sigma^2}}$$

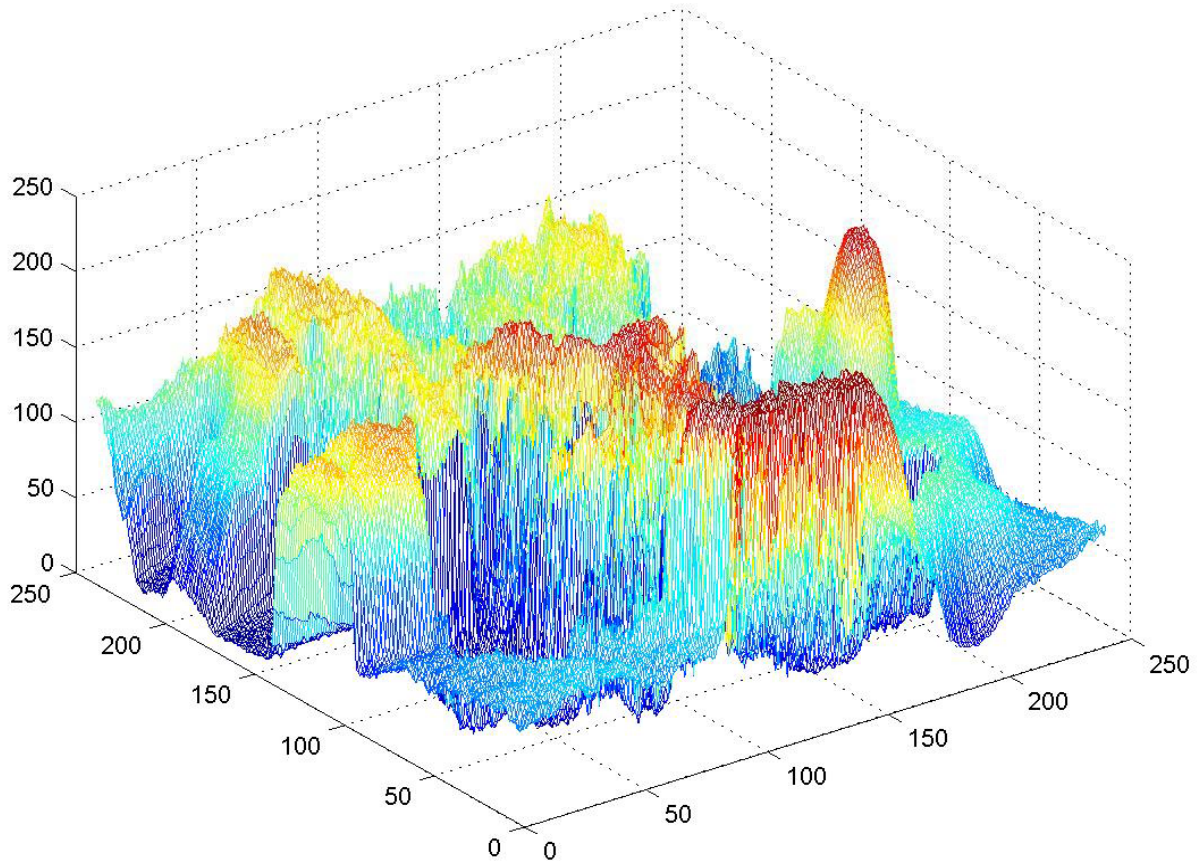


Discrete version

$$f[i,j]=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{i^2+j^2}{2\sigma^2}}$$



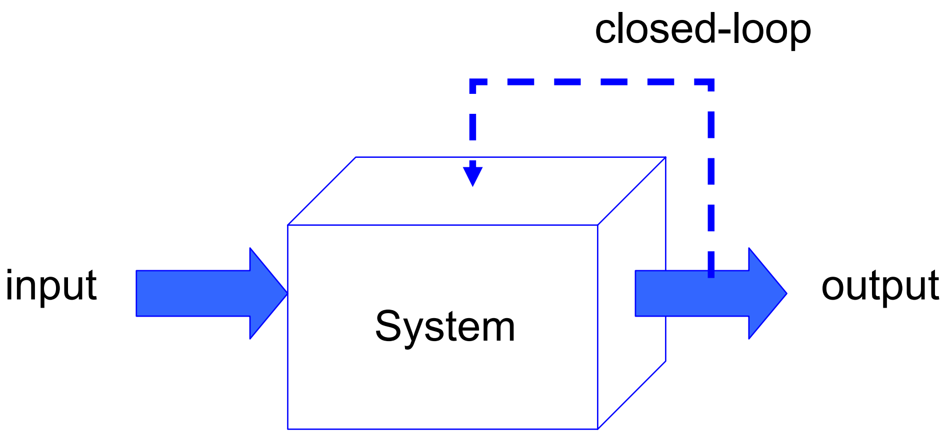
# Example 3: Natural image



# Example 3: Natural image

# What is a system?

- Systems process signals to
  - Extract information (DNA sequence analysis)
  - Enable transmission over channels with limited capacity (JPEG, JPEG2000, MPEG coding)
  - Improve security over networks (encryption, watermarking)
  - Support the formulation of diagnosis and treatment planning (medical imaging)
  - …….



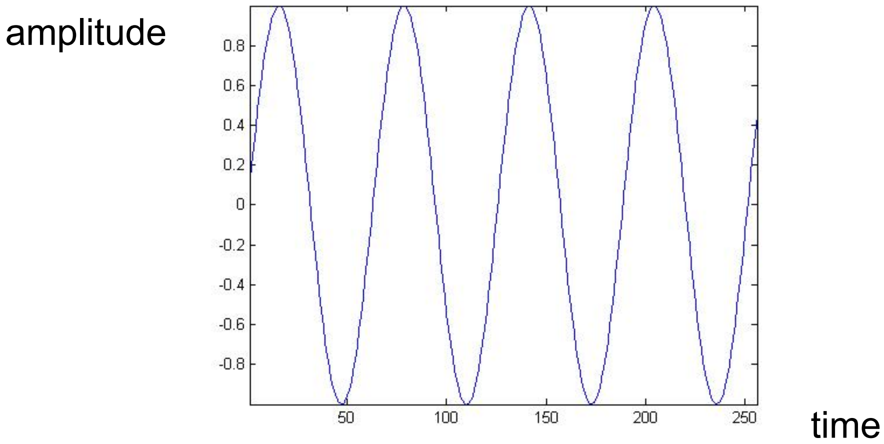

The function linking the output of the system with the input signal is called transfer function and it is typically indicated with the symbol $h(\bullet)$

# Classification of signals

- Continuous time – Discrete time
- Analog – Digital (numerical)
- Periodic – Aperiodic
- Energy – Power
- Deterministic – Random (probabilistic)
- Note
  - Such classes are not disjoint, so there are digital signals that are periodic of power type and others that are aperiodic of power type etc.
  - Any combination of single features from the different classes is possible

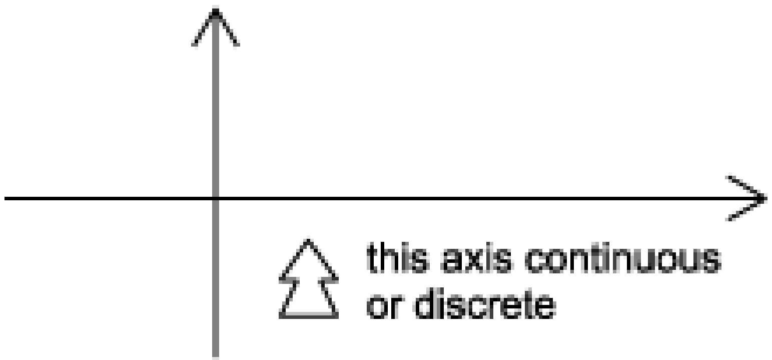
# Continuous time – discrete time

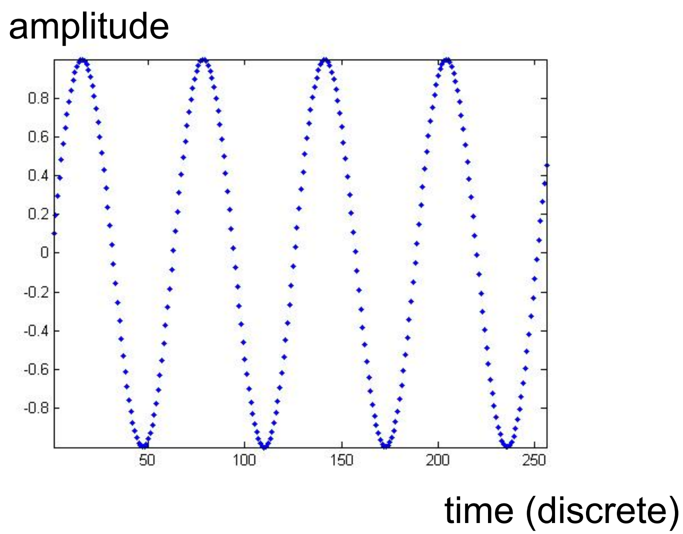
- Continuous time signal: a signal that is specified for every real value of the independent variable
  - The independent variable is continuous, that is it takes any value on the real axis
  - The domain of the function representing the signal has the cardinality of real numbers
    - Signal ↔ f=f(t)
    - Independent variable ↔ time (t), position (x)
    - For continuous-time signals: $t \in \mathbb{R}$



# Continuous time – discrete time

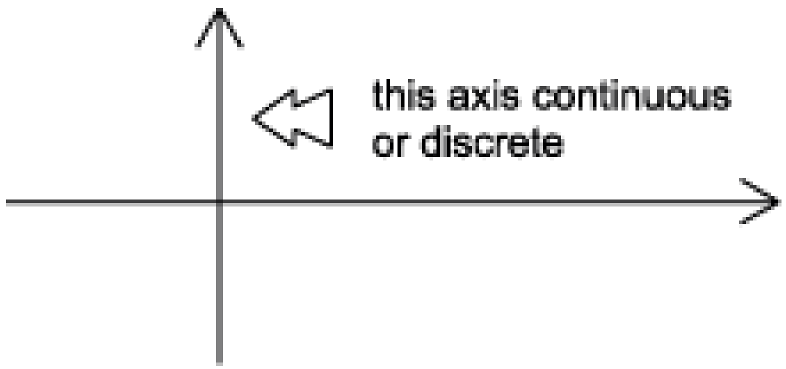
- Discrete time signal: a signal that is specified only for *discrete values* of the independent variable
  - It is usually generated by *sampling* so it will only have values at *equally spaced* intervals along the time axis
  - The domain of the function representing the signal has the cardinality of integer numbers
    - Signal ↔ f=f[n], also called “sequence”
    - Independent variable ↔ n
    - For discrete-time functions: $t \in \mathbb{Z}$



# Analog - Digital

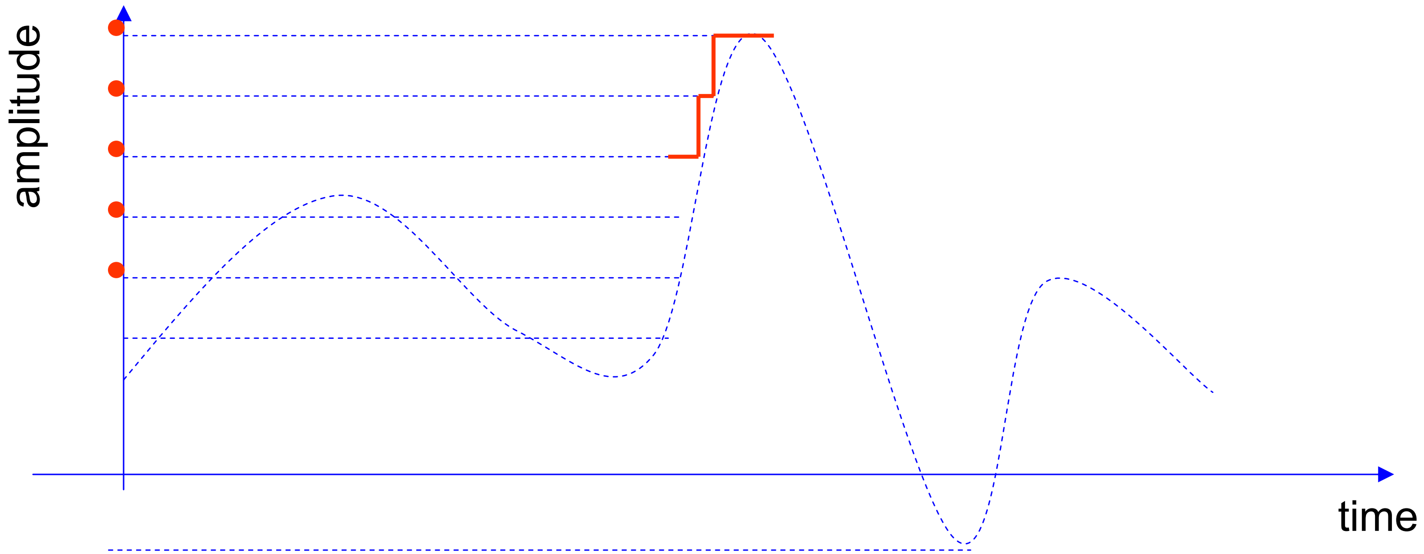
- **Analog signal**: signal whose amplitude can take on any value in a continuous range
  - The amplitude of the function f(t) (or f(x)) has the cardinality of real numbers
    - The dierence between analog and digital is similar to the dierence between continuous-time and discretetime. In this case, however, the difference is with respect to the value of the function (y-axis)
  - Analog corresponds to a continuous y-axis, while digital corresponds to a discrete y-axis.




- An analog signal can be both continuous time and discrete time

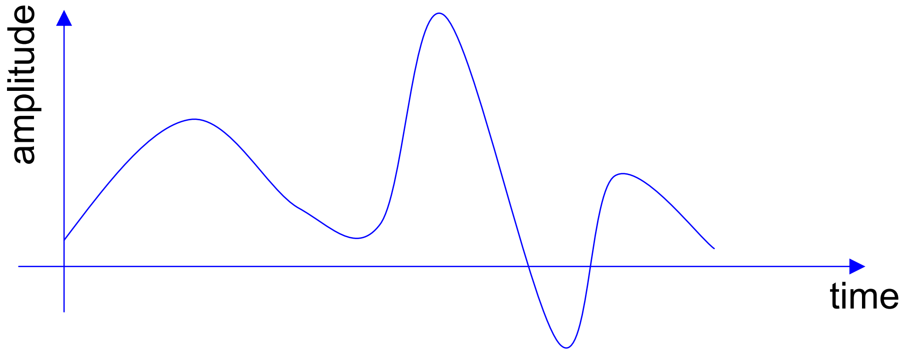
# Analog - Digital

- **Digital signal**: a signal is one whose amplitude can take on only a finite number of values
  - The amplitude of the function f() can take only a finite number of values
  - A digital signal whose amplitude can take only M different values is said to be M-ary
    - Binary signals are a special case for M=2

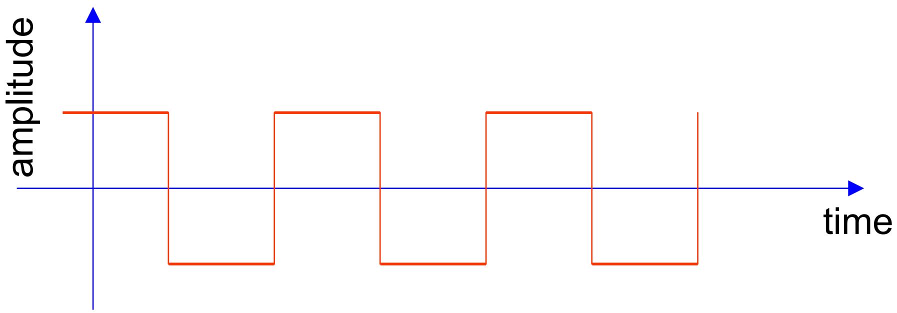


# Example

- Continuous time analog

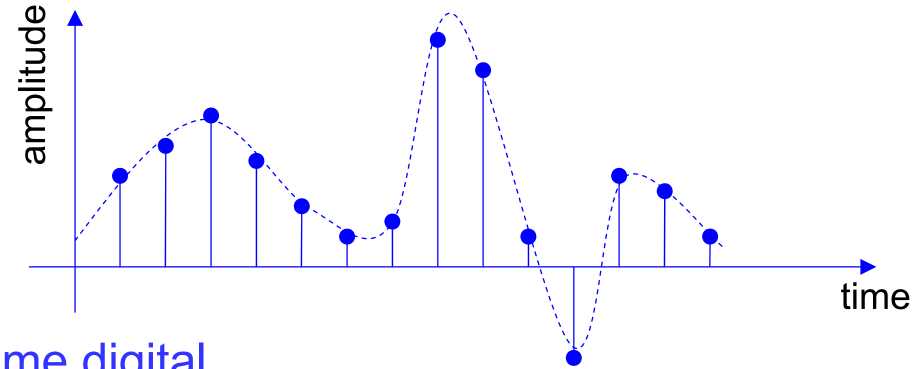



- Continuous time digital
  - binary sequence, where the values of the function can only be one or zero.

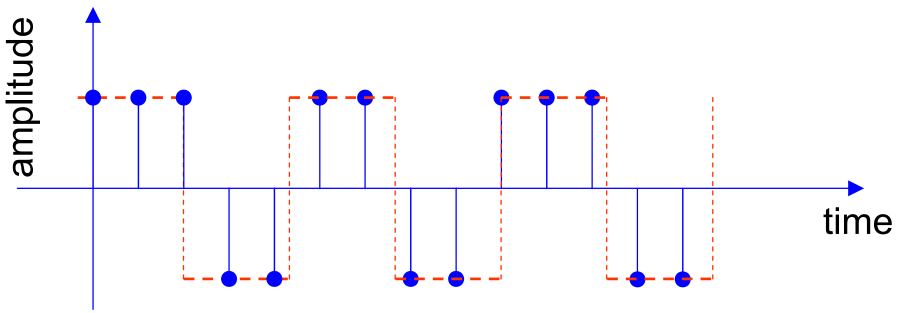


# Example

- Discrete time analog




- Discrete time digital
  - binary sequence, where the values of the function can only be one or zero.



# Summary

| Signal amplitude/ Time or space | Real | Integer |
|---|---|---|
| Real | Analog Continuous-time | Digital Continuous-time |
| Integer | Analog Discrete-time | Digital Discrete time |

# Periodic - Aperiodic

- A signal f(t) is *periodic* if there exists a positive constant $T_0$ such that

$$f(t+T_0) = f(t) \qquad \forall t$$

  - The *smallest* value of $T_0$ which satisfies such relation is said the *period* of the function f(t)
  - A periodic signal remains unchanged when *time-shifted* of integer multiples of the period
  - Therefore, by definition, it starts at minus infinity and lasts forever

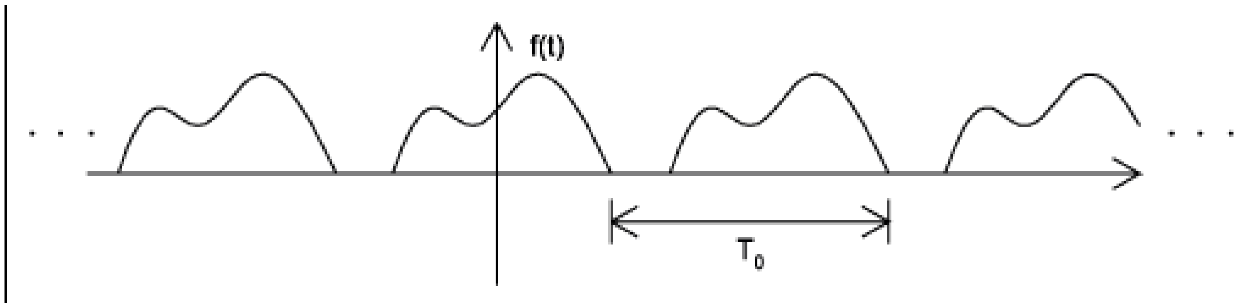
$$-\infty \leq t \leq +\infty \qquad t \in \mathbb{R}$$

$$-\infty \leq n \leq +\infty \qquad n \in \mathbb{Z}$$

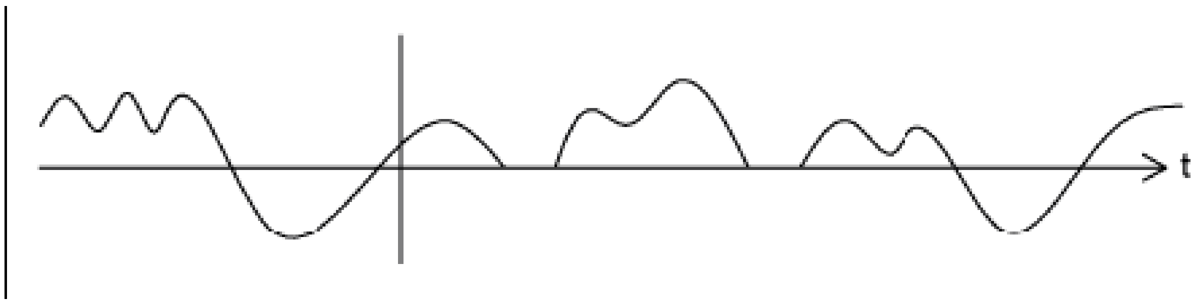
  - Periodic signals can be generated by *periodical extension*

# Examples

- Periodic signal with period $T_0$

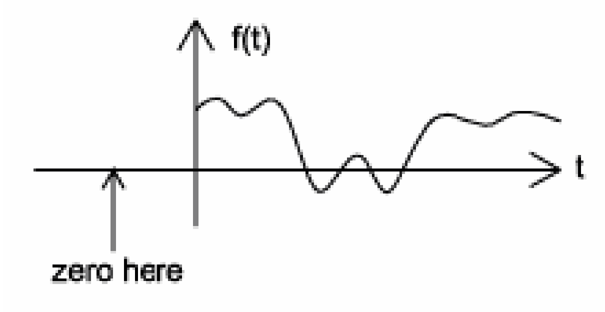



- Aperiodic signal

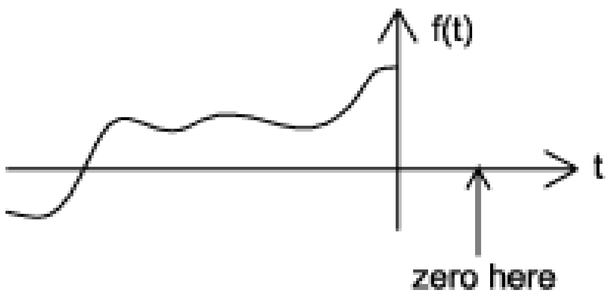


# Causal and non-Causal signals

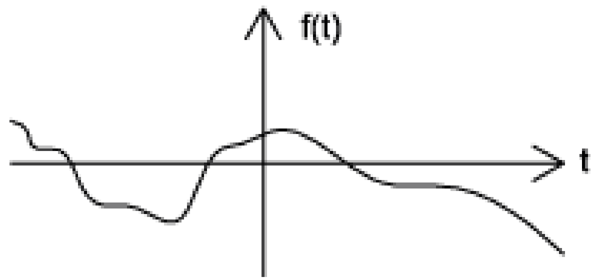
- *Causal* signals are signals that are *zero for all negative time (or spatial positions)*, while




- *Anticausal* are signals that are zero for all *positive time (or spatial positions)*.




- *Noncausal* signals are signals that have nonzero values in both positive and negative time



# Causal and non-causal signals

- Causal signals

$$f(t) = 0 \qquad t < 0$$

- Anticausals signals

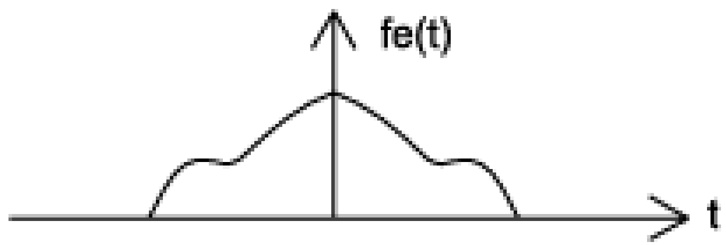
$$f(t) = 0 \qquad t \geq 0$$

- Non-causal signals

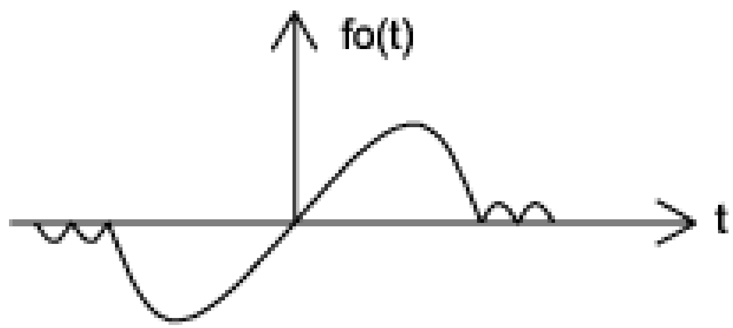
$$\exists t_1 < 0: \qquad f(t_1) = 0$$

# Even and Odd signals

- An even signal is any signal f such that f (t) = f (-t). Even signals can be easily spotted as they are symmetric around the vertical axis.




- An odd signal, on the other hand, is a signal f such that f (t)= - (f (-t))



# Decomposition in even and odd components

- Any signal can be written as a combination of an even and an odd signals
  - Even and odd components

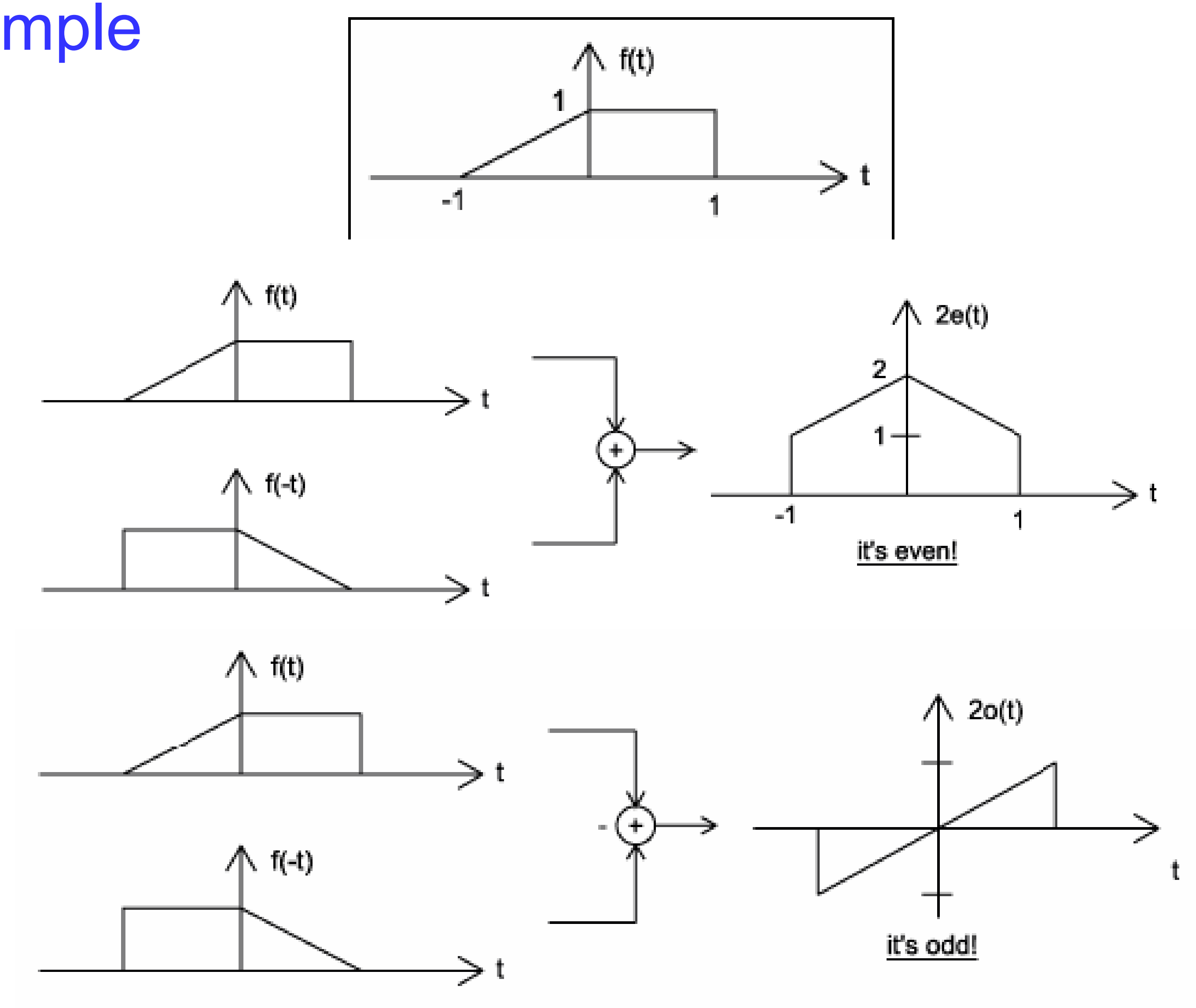
$$f(t) = \frac{1}{2}\left(f(t) + f(-t)\right) + \frac{1}{2}\left(f(t) - f(-t)\right)$$

$$f_e(t) = \frac{1}{2}\left(f(t) + f(-t)\right) \quad \text{even component}$$

$$f_o(t) = \frac{1}{2}\left(f(t) - f(-t)\right) \quad \text{odd component}$$

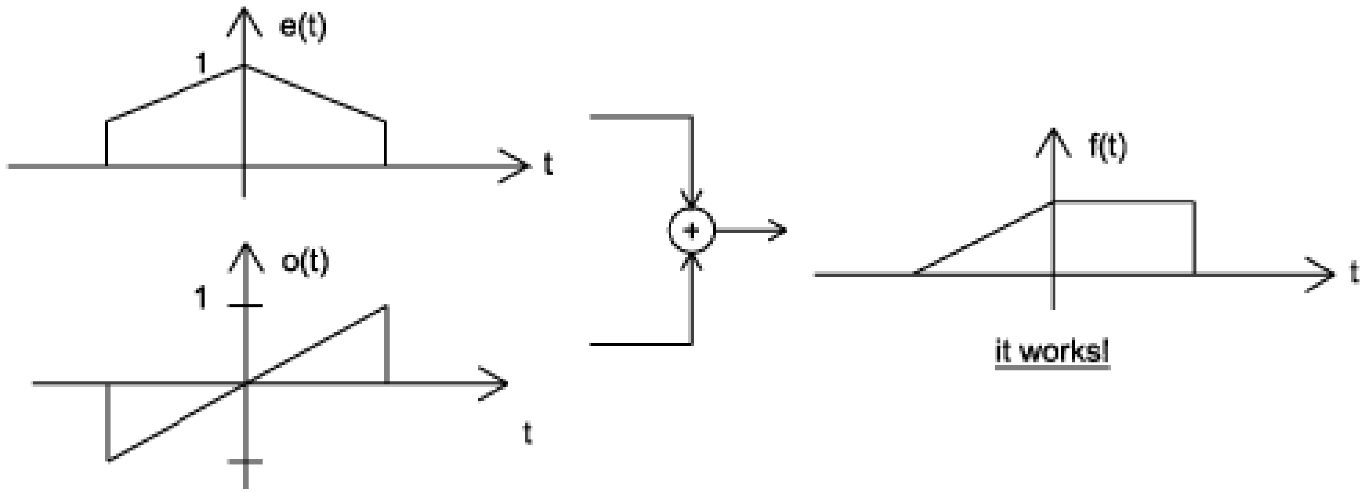
$$f(t) = f_e(t) + f_o(t)$$

# Example



# Example

- Proof



# Some properties of even and odd functions

- even function x odd function = odd function
- odd function x odd function = even function
- even function x even function = even function
- Area

$$\int_{-a}^{a} f_e(t)\,dt = 2\int_{0}^{a} f_e(t)\,dt$$

$$\int_{-a}^{a} f_e(t)\,dt = 0$$

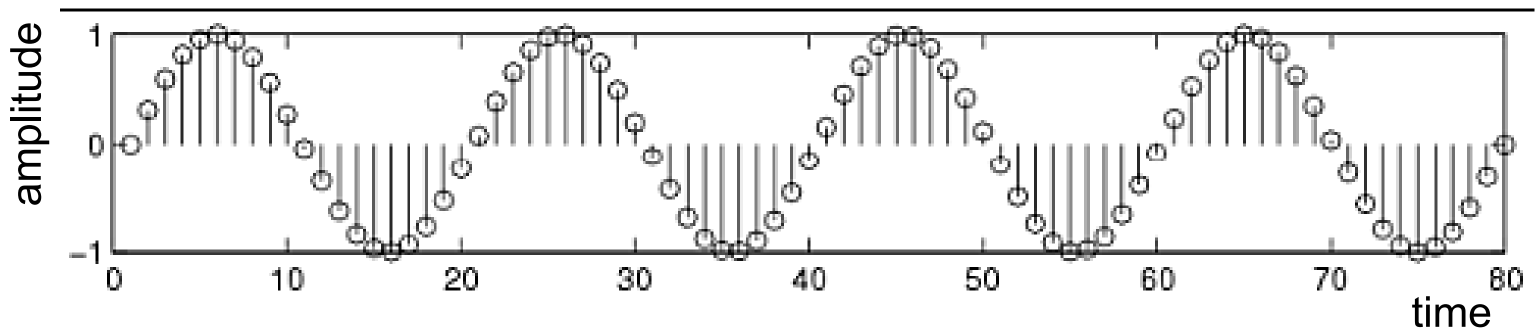
# Deterministic - Probabilistic

- Deterministic signal: a signal whose *physical description* in known completely
- A deterministic signal is a signal in which each value of the signal is fixed and can be determined by a mathematical expression, rule, or table.
- Because of this the future values of the signal can be calculated from past values with complete confidence.
  - There is *no uncertainty* about its amplitude values
  - Examples: signals defined through a mathematical function or graph

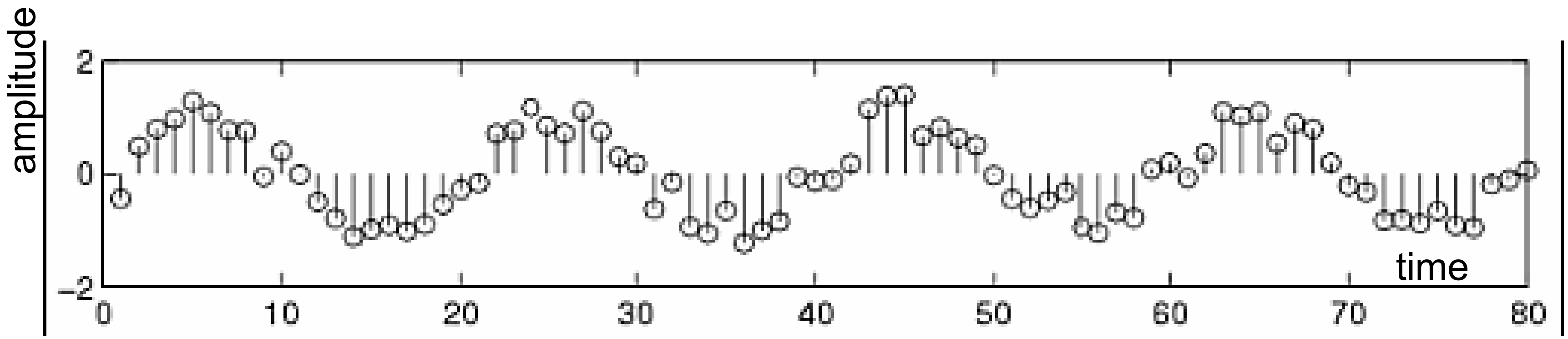
- Probabilistic (or random) signals: the amplitude values *cannot be predicted precisely* but are known only in terms of probabilistic descriptors
- The future values of a random signal cannot be accurately predicted and can usually only be guessed based on the averages of sets of signals
  - They are realization of a stochastic process for which a model could be available
  - Examples: EEG, evocated potentials, noise in CCD capture devices for digital cameras

# Example

- Deterministic signal




- Random signal



# Finite and Infinite length signals

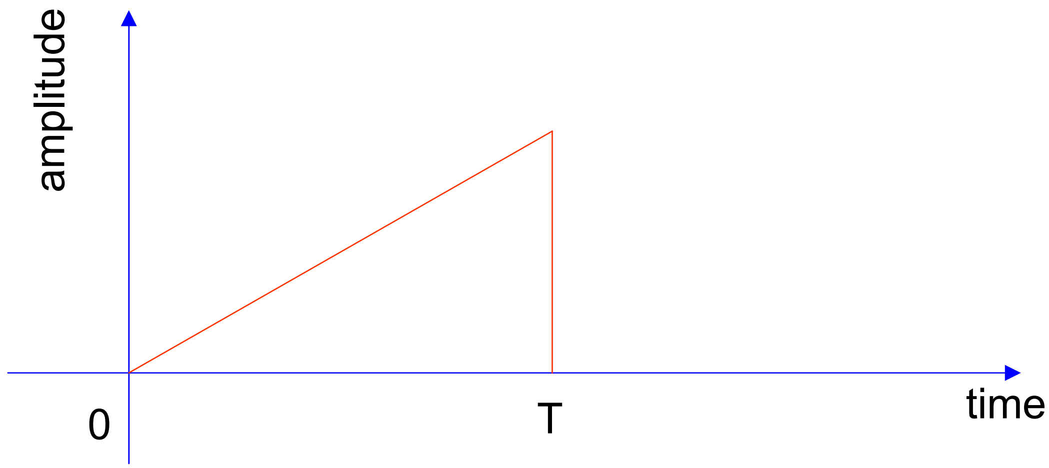
- A finite length signal is non-zero over a finite set of values of the independent variable

$$f = f(t), \forall t : t_1 \leq t \leq t_2$$
$$t_1 > -\infty, t_2 < +\infty$$

- An infinite length signal is non zero over an infinite set of values of the independent variable
  - For instance, a sinusoid $f(t)=\sin(\omega t)$ is an infinite length signal

# Size of a signal: Norms

- "Size" indicates largeness or strength.
- We will use the mathematical concept of the norm to quantify this notion for both continuous-time and discrete-time signals.
- The energy is represented by the area under the curve (of the squared signal)



# Energy

- Signal energy

$$E_f = \int_{-\infty}^{+\infty} f^2(t)dt$$

$$E_f = \int_{-\infty}^{+\infty} |f(t)|^2 dt$$

- Generalized energy : $L_p$ norm
  - For p=2 we get the energy ($L_2$ norm)

$$\|f(t)\| = \left(\int \left(|f(t)|\right)^p dt\right)^{1/p}$$

$$1 \leq p < +\infty$$

# Power

- Power
  - The power is the time average (mean) of the squared signal amplitude, that is the *mean-squared* value of f(t)

$$P_f = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} f^2(t)dt$$

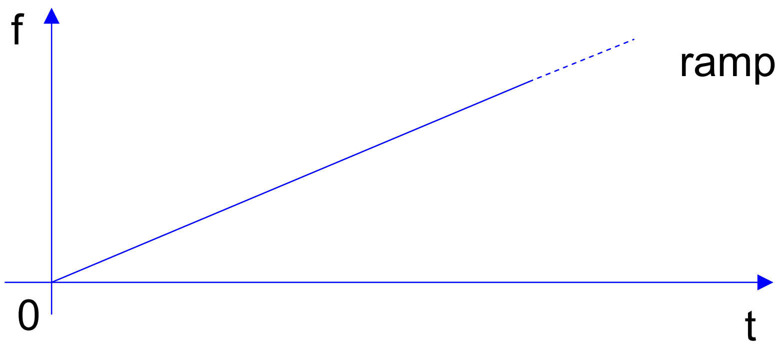
$$P_f = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} |f(t)|^2 \, dt$$

# Power - Energy

- The square root of the power is the root mean square (*rms*) value
  - This is a very important quantity as it is the most widespread measure of similarity/dissimilarity among signals
  - It is the basis for the definition of the Signal to Noise Ratio (SNR)

$$SNR = 20\log_{10}\left(\sqrt{\frac{P_{signal}}{P_{noise}}}\right)$$

  - It is such that a constant signal whose amplitude is =rms holds the same power content of the signal itself
- There exists signals for which neither the energy nor the power are finite

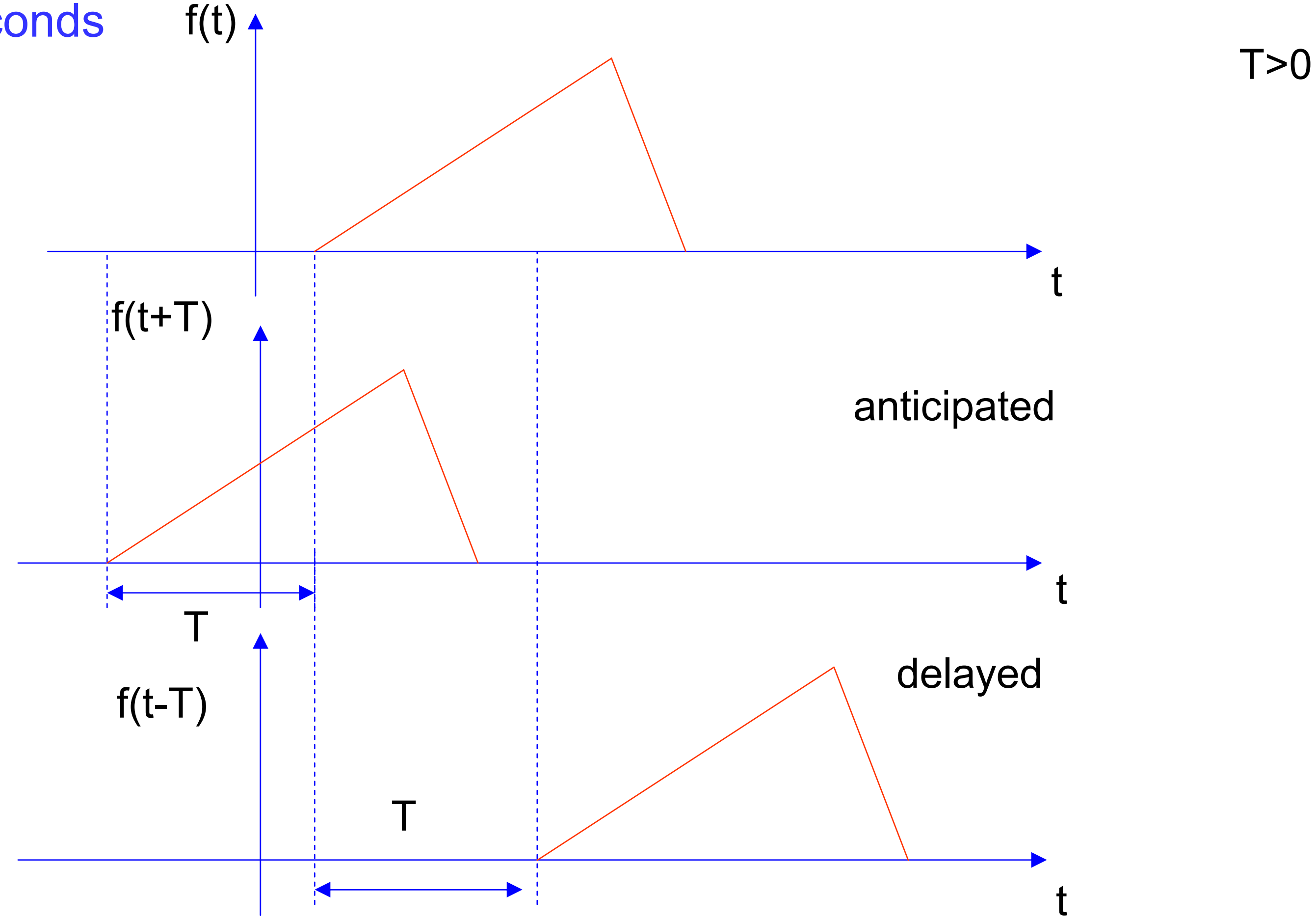


# Energy and Power signals

- A signal with finite energy is an energy signal
  - Necessary condition for a signal to be of energy type is that the amplitude goes to zero as the independent variable tends to infinity
- A signal with finite and different from zero power is a power signal
  - The mean of an entity averaged over an infinite interval exists if either the entity is periodic or it has some statistical regularity
  - A power signal has infinite energy and an energy signal has zero power
  - There exist signals that are neither power nor energy, such as the ramp
- All practical signals have finite energy and thus are energy signals
  - It is impossible to generate a real power signal because this would have infinite duration and infinite energy, which is not doable.

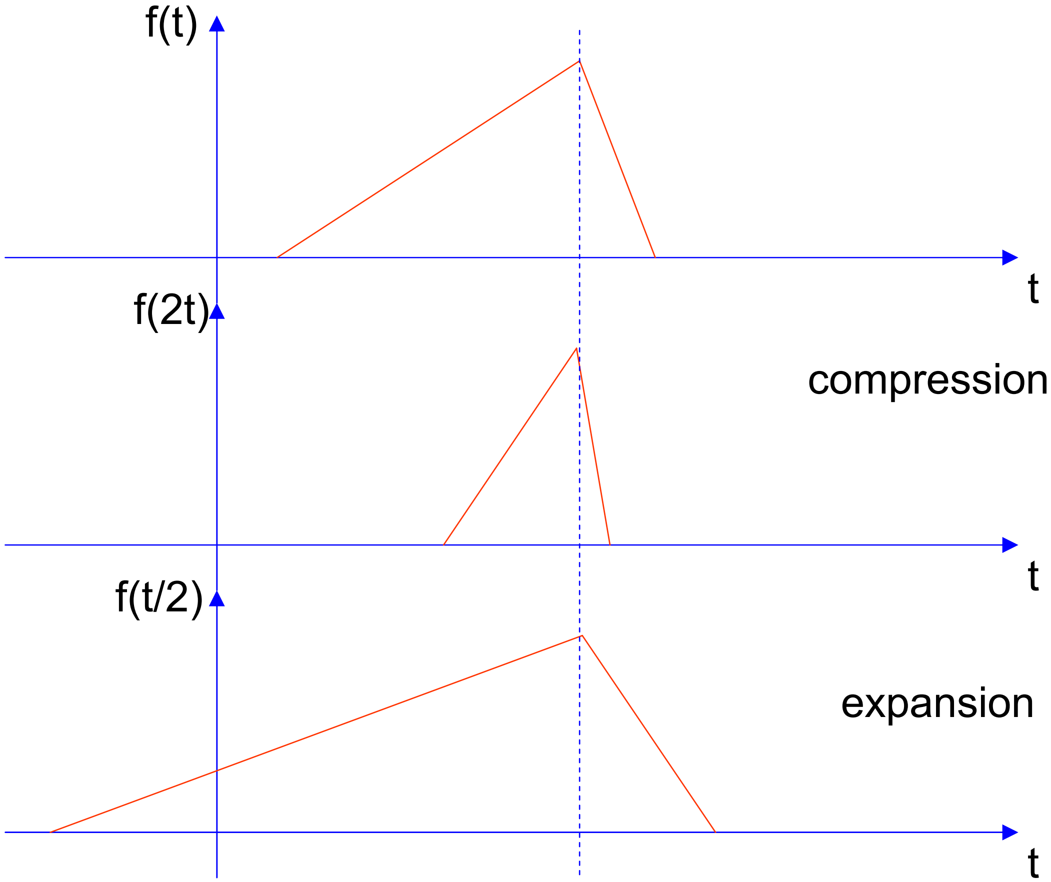
# Useful signal operations: shifting, scaling, inversion

- **Shifting**: consider a signal f(t) and the same signal delayed/anticipated by T seconds



# Useful signal operations: shifting, scaling, inversion

- (Time) Scaling: compression or expansion of a signal in time




compression $\varphi(t) = f(2t)$

expansion $\varphi(t) = f(t/2)$

# Useful signal operations: shifting, scaling, inversion

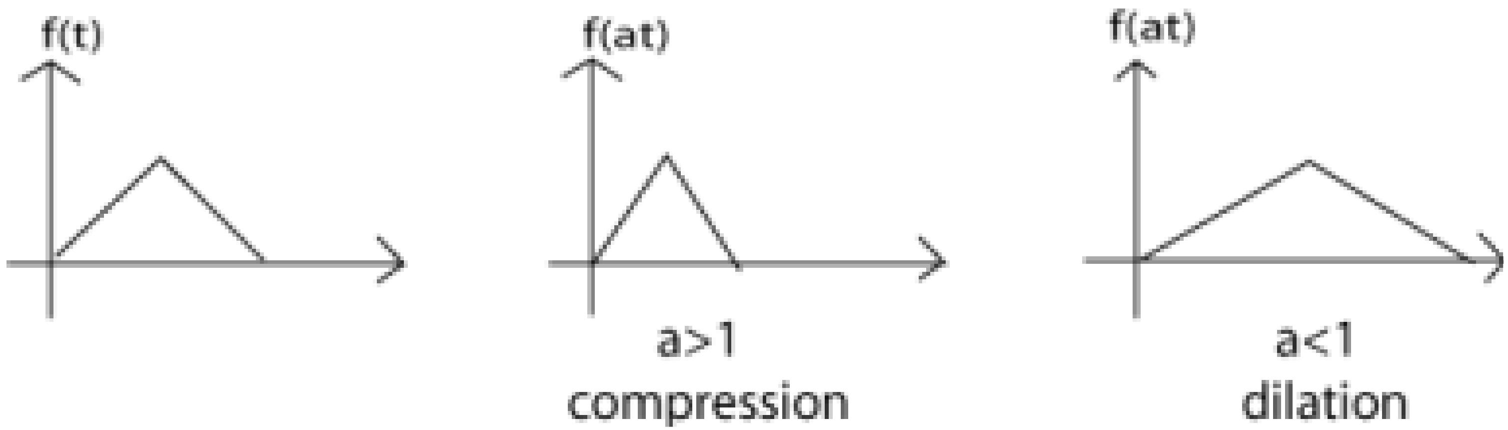
- Scaling: generalization

$$a > 1$$

$$\varphi(t) = f(at) \rightarrow \text{compressed version}$$

$$\varphi(t) = f\left(\frac{t}{a}\right) \rightarrow \text{dilated (or expanded) version}$$

Viceversa for $a < 1$

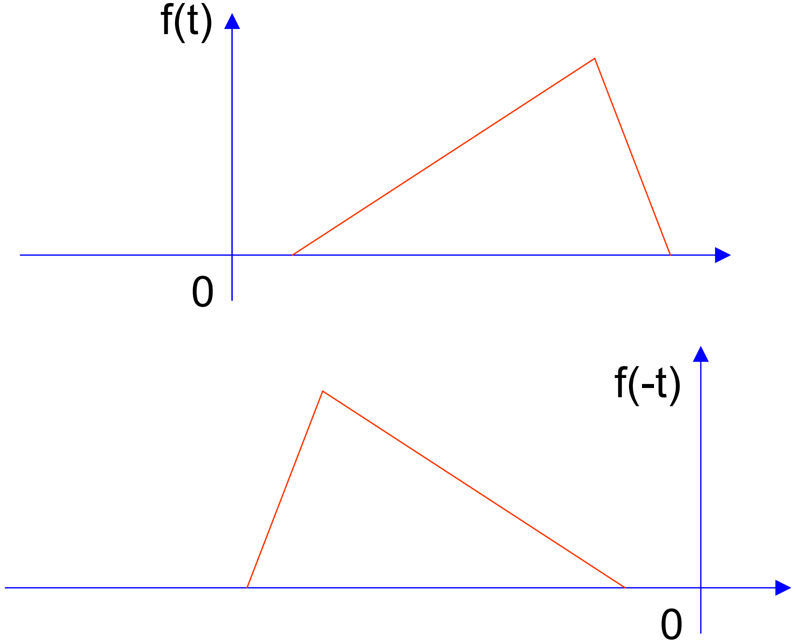


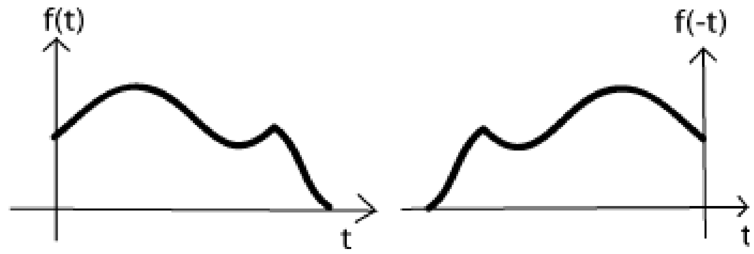

Gloria Menegaz

# Useful signal operations: shifting, scaling, inversion

- (Time) inversion: mirror image of f(t) about the vertical axis

$$\varphi(t) = f(-t)$$



# Useful signal operations: shifting, scaling, inversion

- Combined operations: $f(t) \rightarrow f(at-b)$
- Two possible sequences of operations

1. *Time shift f(t) by to obtain f(t-b). Now time scale the shifted signal f(t-b) by a to obtain f(at-b).*
2. *Time scale f(t) by a to obtain f(at). Now time shift f(at) by b/a to obtain f(at-b).*
   - *Note that you have to replace t by (t-b/a) to obtain f(at-b) from f(at) when replacing t by the translated argument (namely t-b/a))*

# Useful functions

- Unit step function
  - Useful for representing causal signals

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

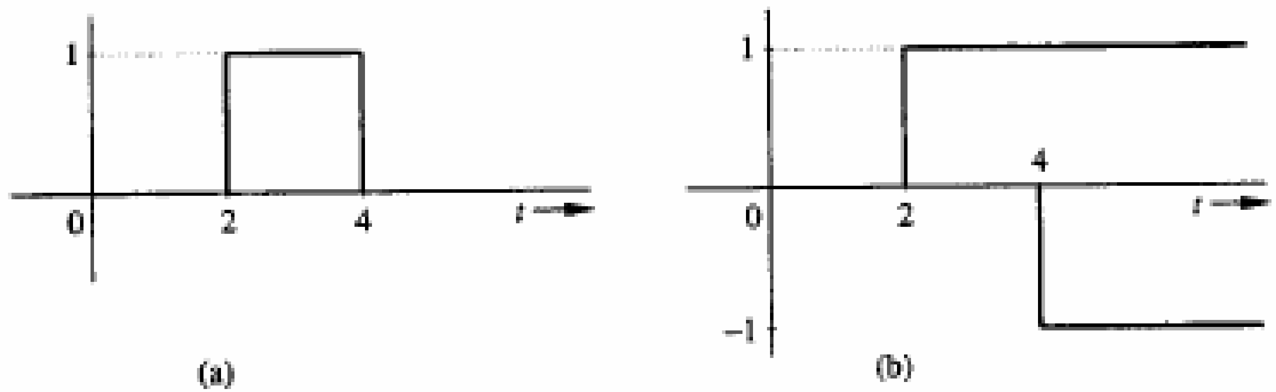
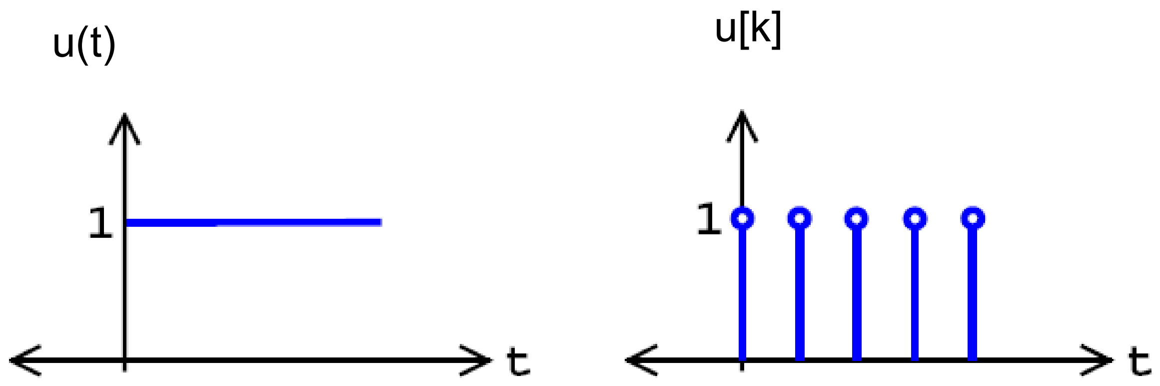



**Fig. 1.15** Representation of a rectangular pulse by step functions.

$$f(t) = u(t-2) - u(t-4)$$

# Useful functions

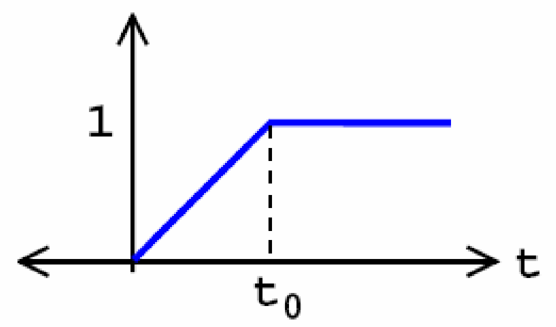
- Continuous and discrete time unit step functions



# Useful functions

- Ramp function (continuous time)

$$r(t) = \begin{cases} 0 \text{ if } t < 0 \\ \frac{t}{t_0} \text{ if } 0 \leq t \leq t_0 \\ 1 \text{ if } t > t_0 \end{cases}$$

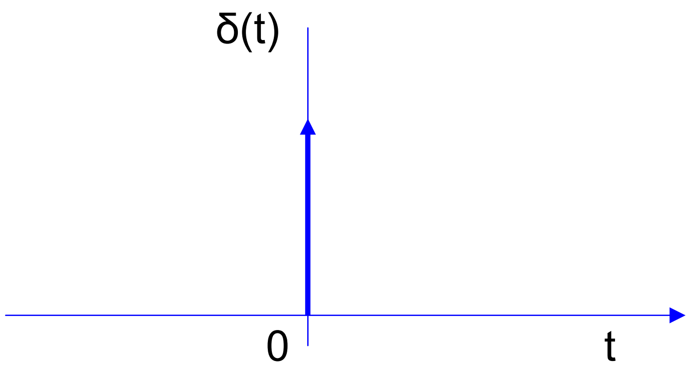


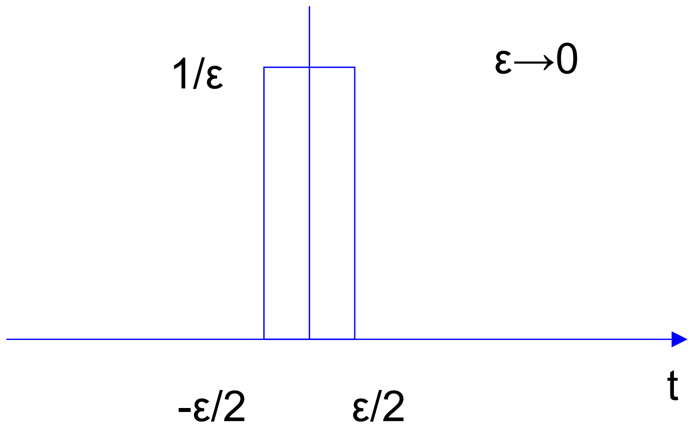
# Useful functions

- Unit impulse function

$$\delta(t) = 0 \quad t \neq 0$$

$$\int_{-\infty}^{+\infty} \delta(t)\, dt = 1$$



# Properties of the unit impulse function

- Multiplication of a function by impulse

$$\phi(t)\delta(t) = \phi(0)\delta(t)$$
$$\phi(t)\delta(t-T) = \phi(T)\delta(t-T)$$

- Sampling property of the unit function

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

  – The area under the curve obtained by the product of the unit impulse function shifted by T and $\varphi(t)$ is the value of the function $\varphi(t)$ for $t=T$

# Properties of the unit impulse function

- The unit step function is the integral of the unit impulse function

$$\frac{du}{dt} = \delta(t)$$

$$\int_{-\infty}^{t} \delta(t)\,dt = u(t)$$

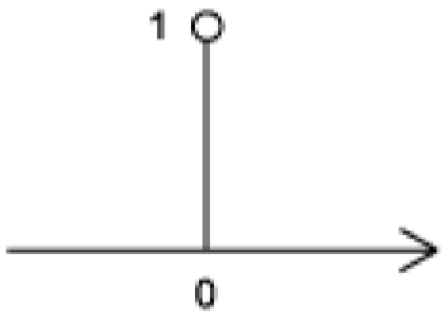
  - Thus

$$\int_{-\infty}^{t} \delta(t)\,dt = u(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$$

# Properties of the unit impulse function

- Discrete time impulse function

$$\delta[n] = \begin{cases} 1 \text{ if } n = 0 \\ 0 \text{ otherwise} \end{cases}$$



# Useful functions

- Continuous time complex exponential

$$f(t) = Ae^{j\omega t}$$

- Euler's relations

$$Ae^{jwt} = Acos(\omega t) + j(Asin(\omega t))$$

$$cos(\omega t) = \frac{e^{jwt} + e^{-(jwt)}}{2}$$

$$sin(\omega t) = \frac{e^{jwt} - e^{-(jwt)}}{2j}$$

$$e^{jwt} = cos(\omega t) + jsin(\omega t)$$

- Discrete time complex exponential
  - k=nT

$$\begin{aligned} f[n] &= Be^{snT} \\ &= Be^{j\omega nT} \end{aligned}$$

# Useful functions

- Exponential function $e^{st}$
  - Generalization of the function $e^{j\omega t}$

$$s = \sigma + j\omega$$

Therefore

$$e^{st} = e^{(\sigma + j\omega)t} = e^{\sigma t} e^{j\omega t} = e^{\sigma t}(\cos \omega t + j \sin \omega t) \tag{1.30a}$$

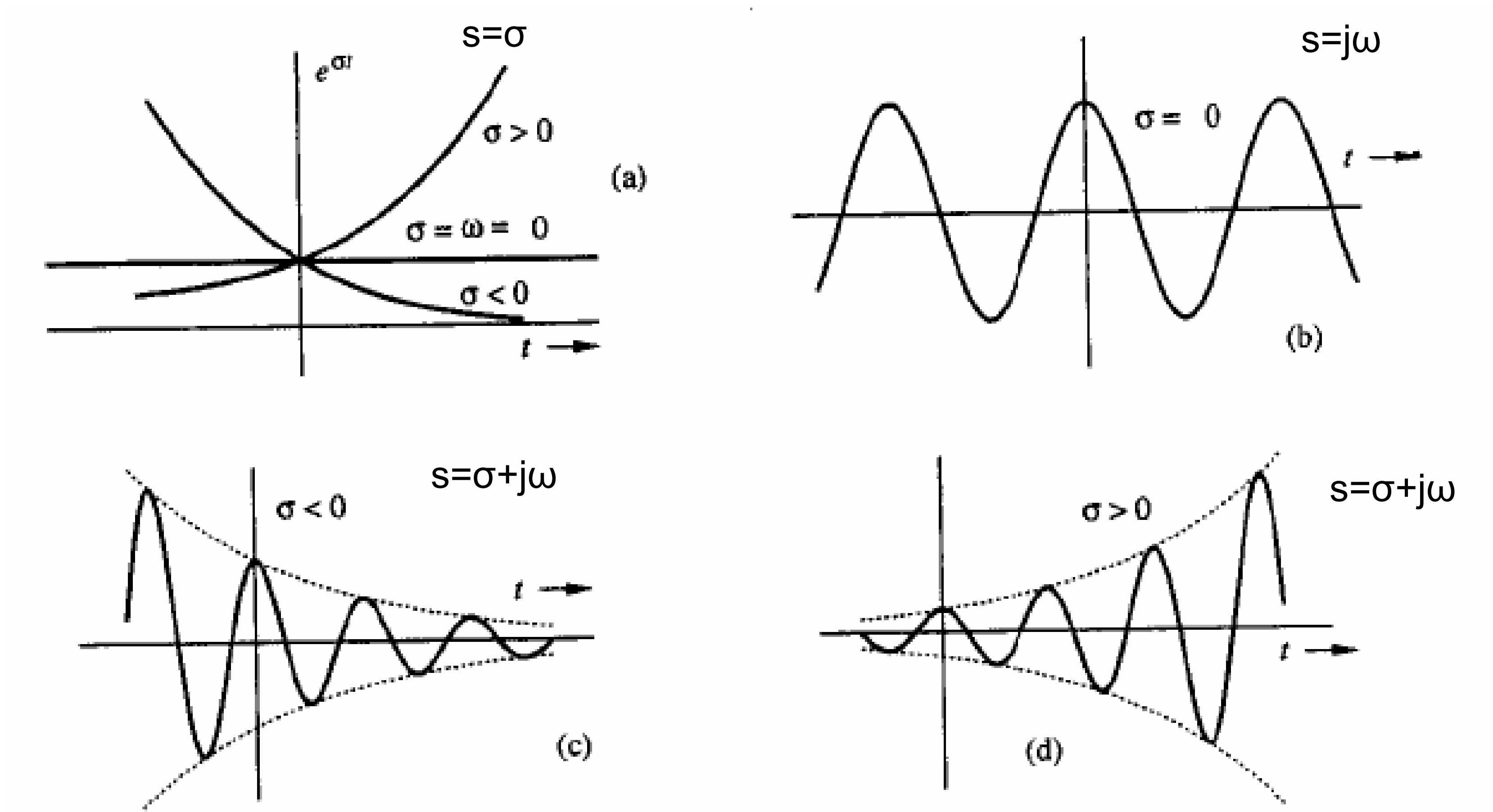
If $s^* = \sigma - j\omega$ (the conjugate of $s$), then

$$e^{s^* t} = e^{\sigma - j\omega} = e^{\sigma t} e^{-j\omega t} = e^{\sigma t}(\cos \omega t - j \sin \omega t) \tag{1.30b}$$

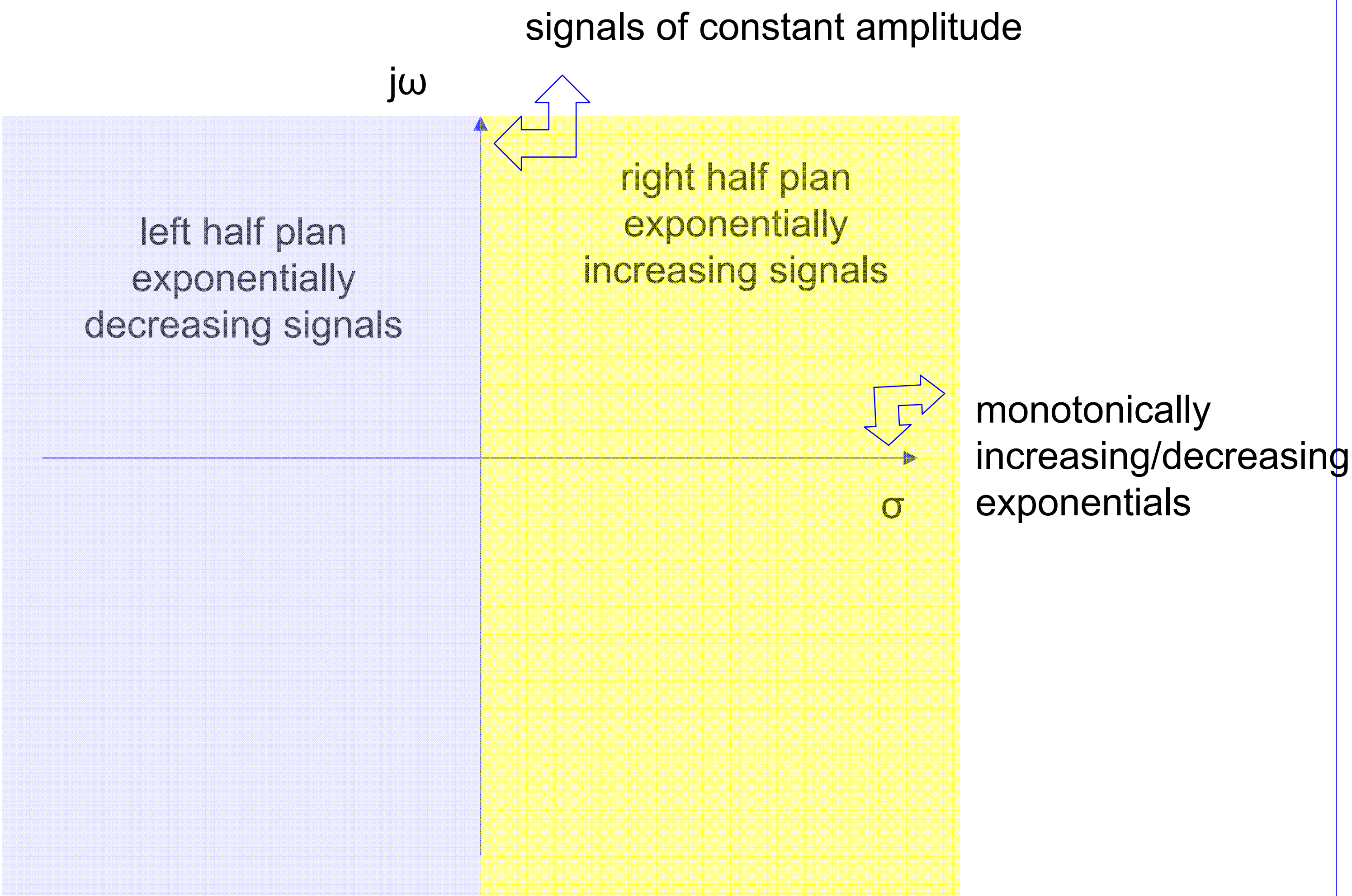
and

$$e^{\sigma t} \cos \omega t = \frac{1}{2}(e^{st} + e^{s^* t}) \tag{1.30c}$$

# The exponential function

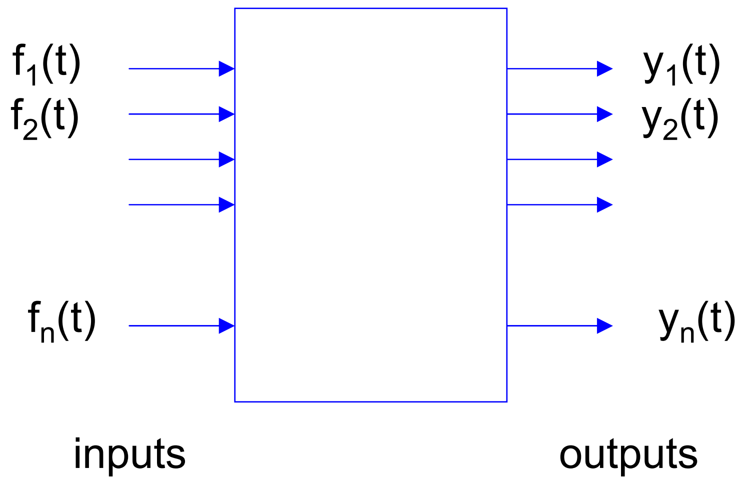


# Complex frequency plan



# Systems

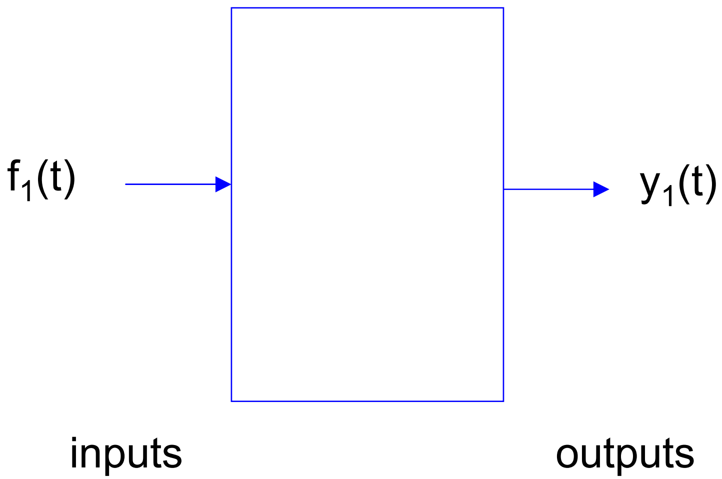
- A system is characterized by
  - inputs
  - outputs
  - rules of operation (mathematical model of the system)

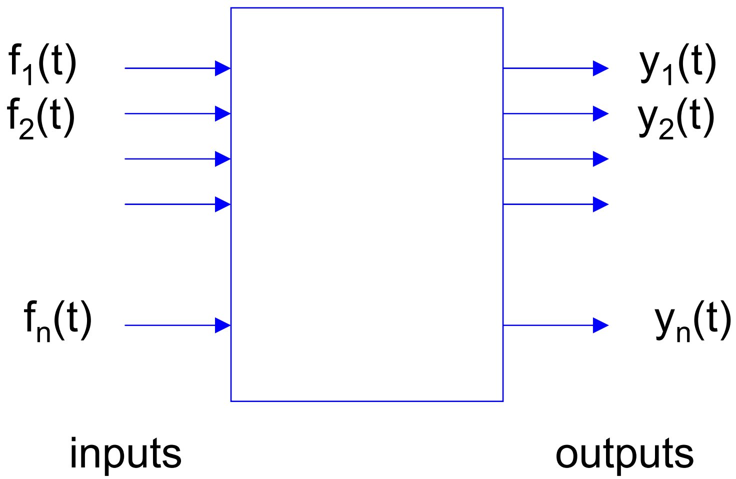


# Systems

- Study of systems: mathematical modeling, analysis, design
  - Analysis: how to determine the system output given the input and the system mathematical model
  - design or synthesis: how to design a system that will produce the desired set of outputs for given inputs

- SISO: single input single output - MIMO: multiple input multiple output

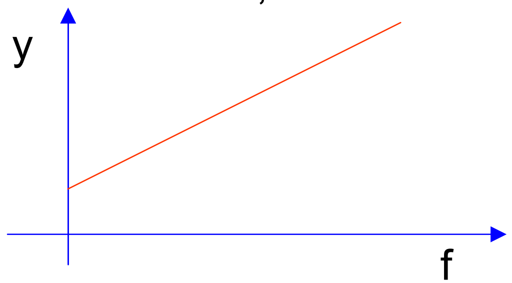


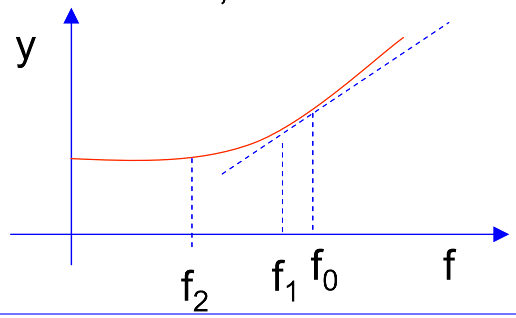
# Response of a linear system

- Total response = Zero-input response + Zero-state response
  - The output of a system for t≥0 is the result of two independent causes: the initial conditions of the system (or system state) at t=0 and the input f(t) for t≥0.
  - Because of linearity, the total response is the sum of the responses due to those two causes
  - The zero-input response is only due to the initial conditions and the zero-state response is only due to the input signal
  - This is called decomposition property
- Real systems are *locally* linear
  - Respond linearly to small signals and non-linearly to large signals

causal, linear




causal, non linear




locally linear around $f_0$