Processing of Speech Signals

Acquisition of Speech signal

A/D conversion

Concept of convolution & LTI system

Correlation

DFT relations & graphical interpretation.

pole-zero Analysis

LP-analysis

Filter-bank analysis

Cepstral & MFCC

PLP - coefficients

Sinusoidal analysis

HNM analysis

group delay analysis

Processing Speech Signals

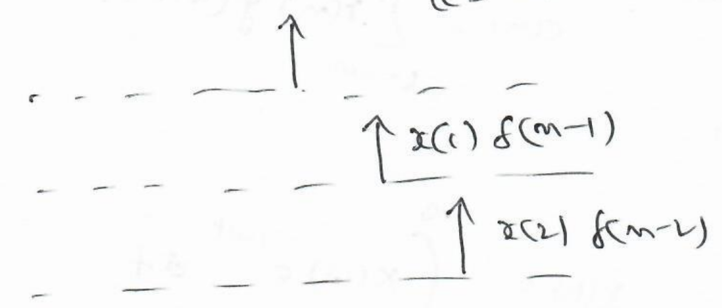
30/1/08

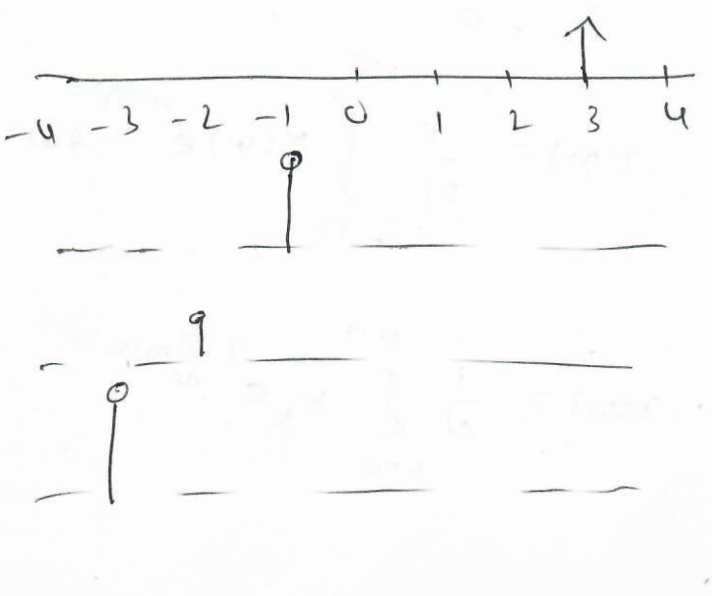
Speech (pressure variations) → [Microphone] → analog signal (voltage variations)

A/D conversion

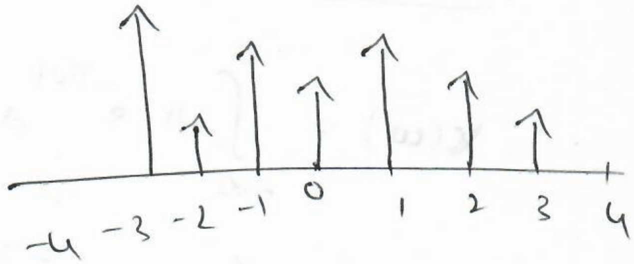
Analog I/p $x(t)$ → [Sampler] → [Quantizer] → [Encoder] → digital data $x(n)$

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\,\delta(n-k)$$

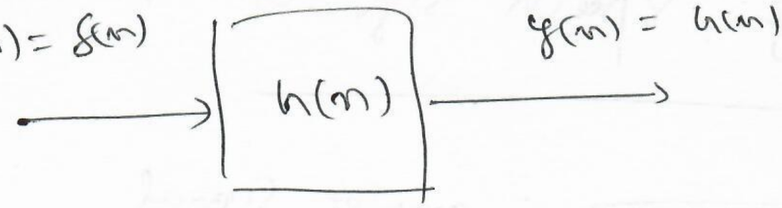




=)



$x(n)$



I/p $\quad$ O/p

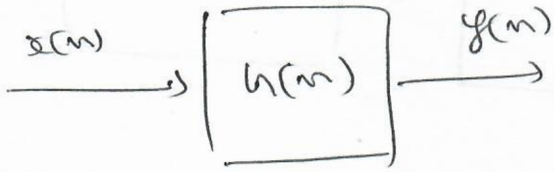
$\delta(n) \quad h(n)$

$\delta(n-k) \quad h(n-k)$

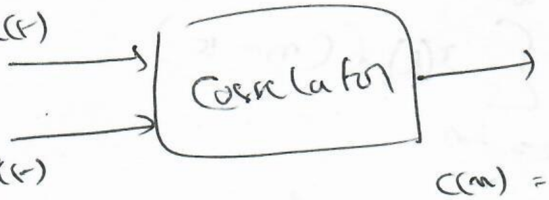
$x(k)\,\delta(n-k) \quad x(k)\,h(n-k)$

$\sum x(k)\,\delta(n-k) \quad \sum x(k)\,h(n-k) \Rightarrow x(n) * h(n)$

Convolution



Correlation



$\int x(t)\,y(t+\tau)$

$$c(n) = \sum_{k=-\infty}^{\infty} x(n)\,y(n+k)$$

Fourier Transform relations

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\,e^{-j\omega t}\,dt \qquad x(t) = \int_{-\infty}^{\infty} X(\omega)\,e^{j\omega t}\,dt$$

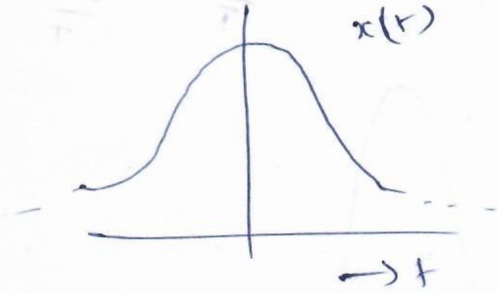
$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)\,e^{-j\omega n} \qquad x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)\,e^{j\omega n}\,d\omega$$

$$X(k) = \sum_{n=0}^{N-1} x(n)\,e^{-j\frac{2\pi}{N}nk} \qquad x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_k\,e^{j\frac{2\pi}{N}nk}$$

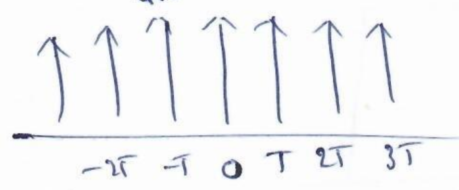
Processing of Speech signal

Continuous $\longrightarrow$ Discrete

$x(t) \longrightarrow x(t)\sum_m \delta(t - mT)$

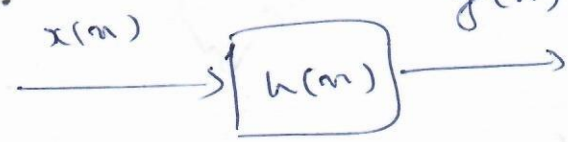


$\sum_m \delta(t - mT)$



$\Rightarrow$

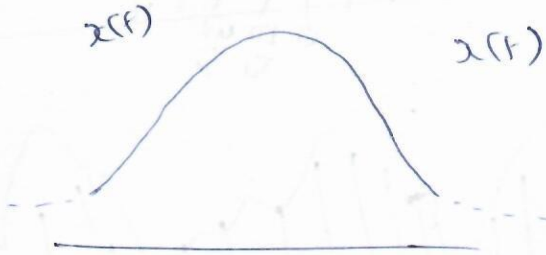
Prop of linear discrete time systems

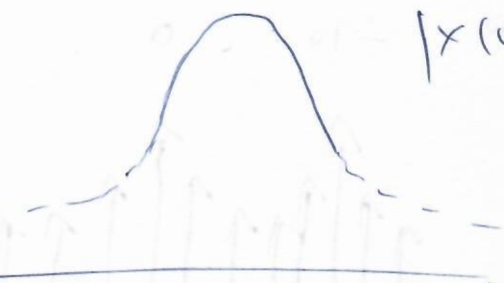


$y(n) = x(n) * h(n)$

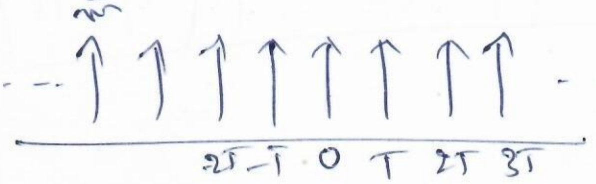
$= \sum h(k)\, x(n-k)$

$$\left.\begin{aligned} y(n) &= x(n) * h(n) \\ Y(\omega) &= X(\omega) H(\omega) \end{aligned}\right\} \quad x(n) * h(n) \Leftrightarrow X(\omega) H(\omega)$$

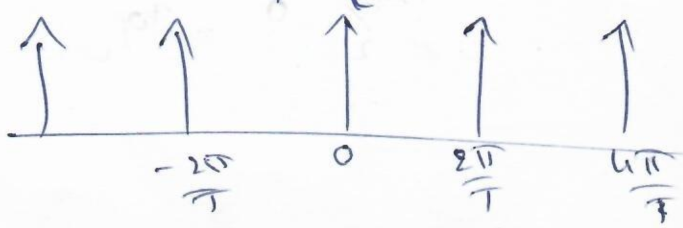




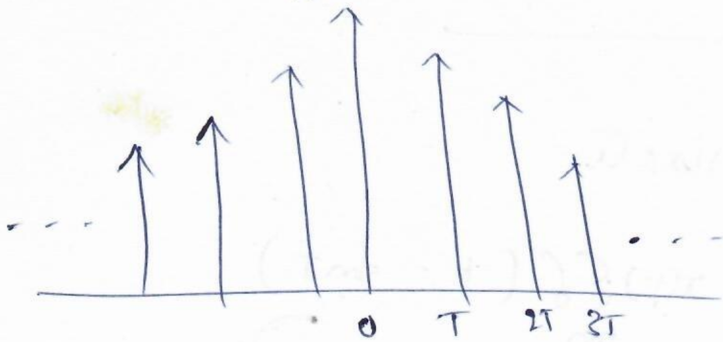
$\sum_m \delta(t + mT)$

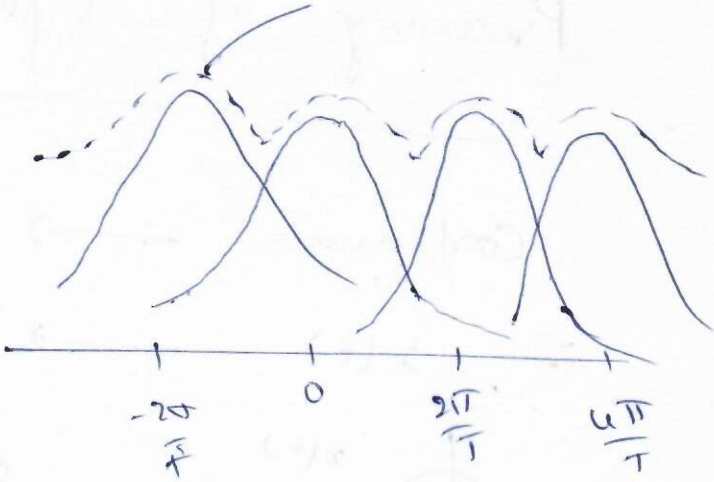


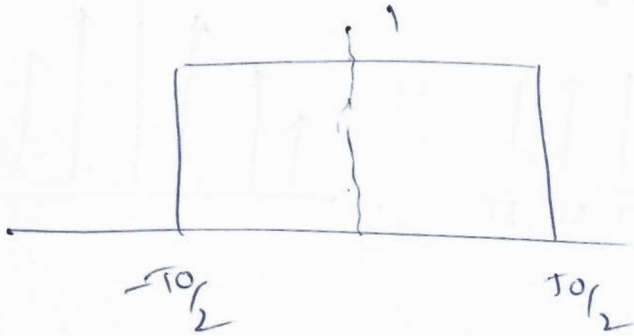
$\frac{2\pi}{T} \sum_k \delta\left(\omega - \frac{2\pi}{T} k\right)$

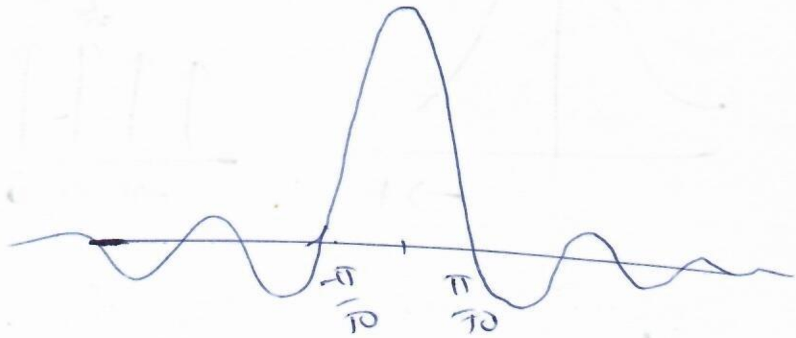


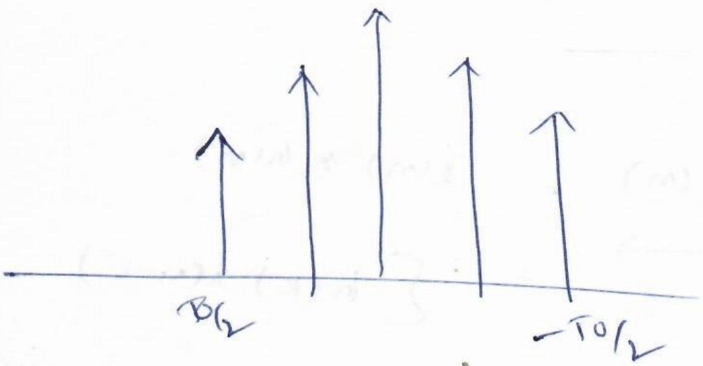
$$x(t) = \sum_m \delta(t - mT)$$

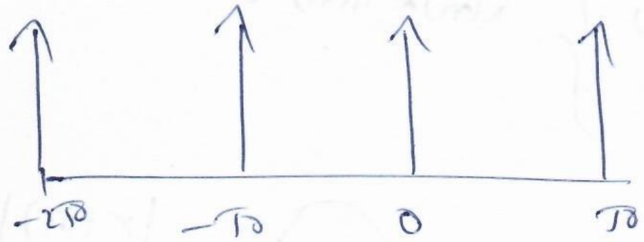


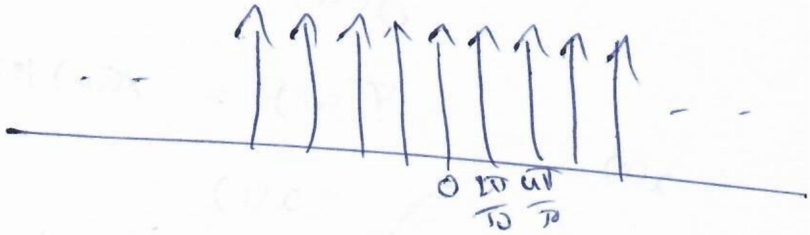


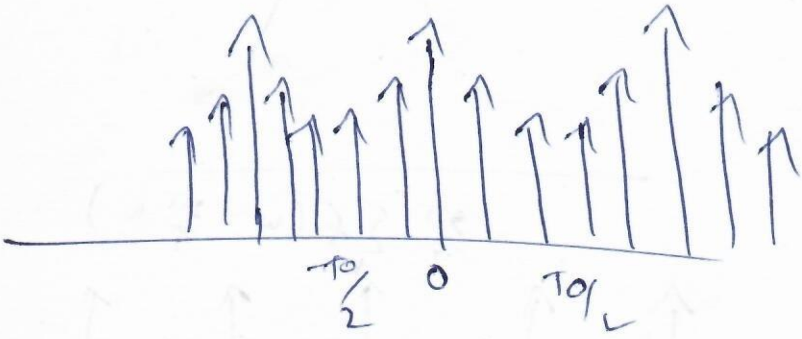


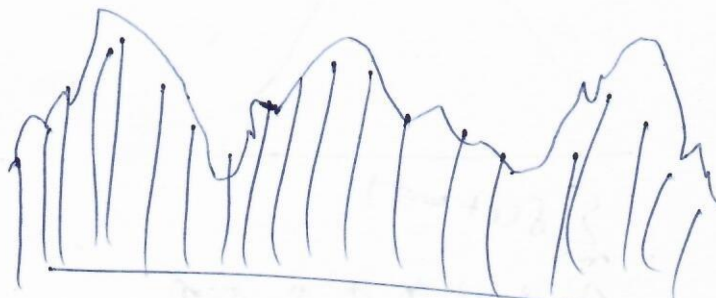












Processing of Speech Signals

Artifacts of DFT

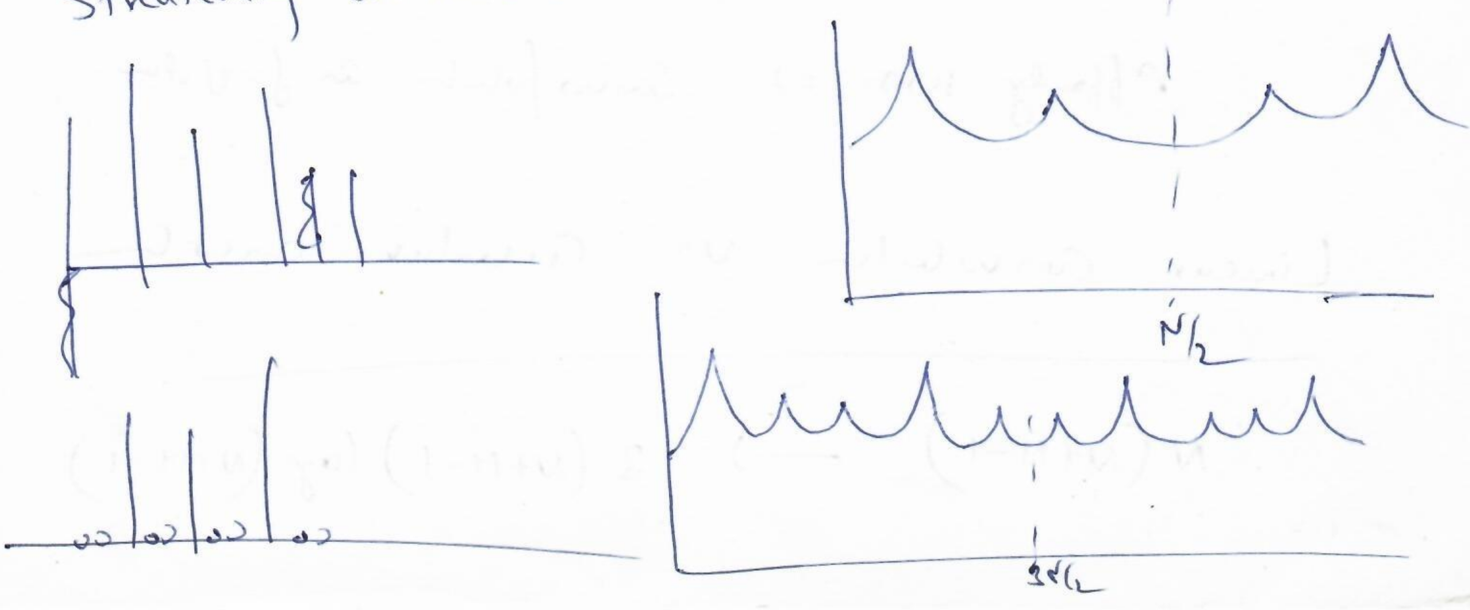
- Aliasing $\rightarrow$ Sampling in time domain
- Sidelobe leakage $\rightarrow$ Finite duration in time domain
- Picket fencing $\rightarrow$ Sampling in freq domain

Important prop of DFT

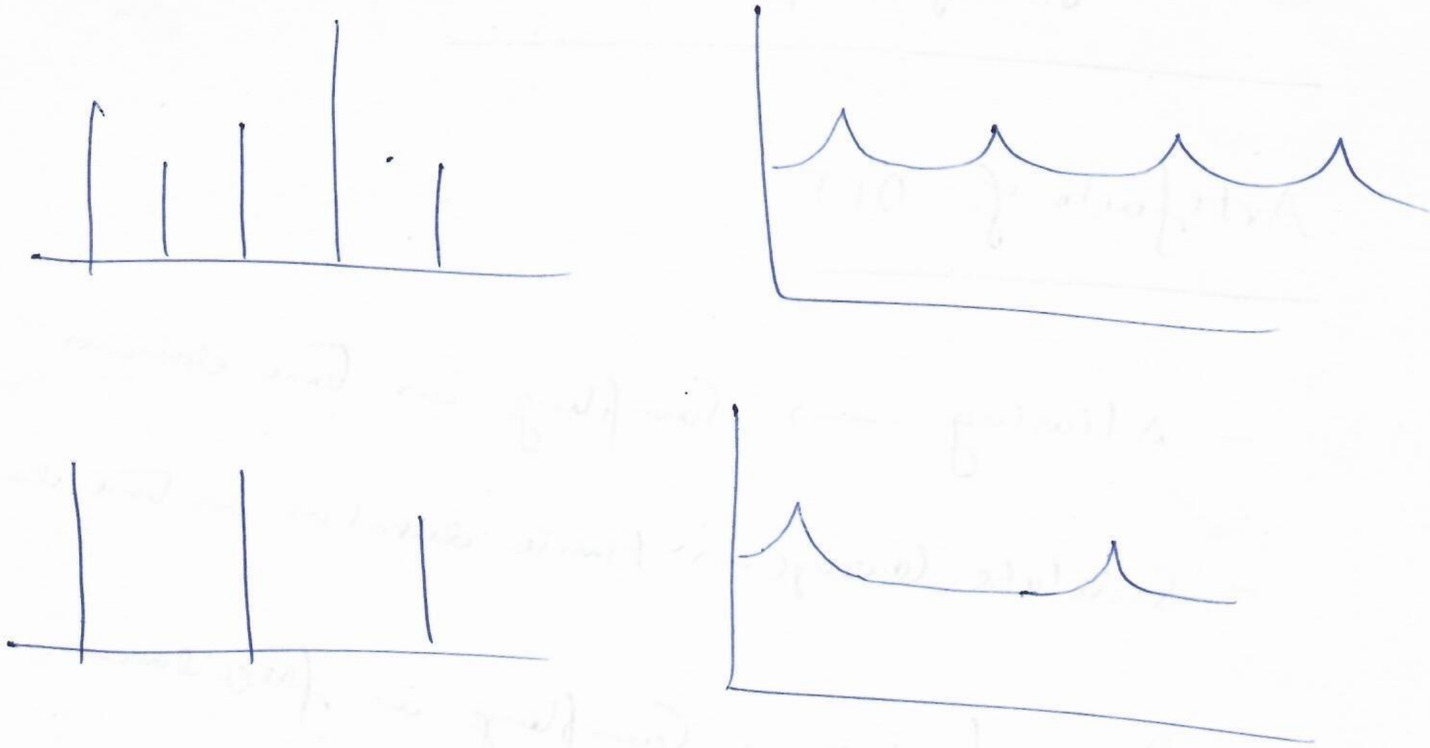
Symmetry & periodicity property

$$N^2 \longrightarrow N \log_2 N$$

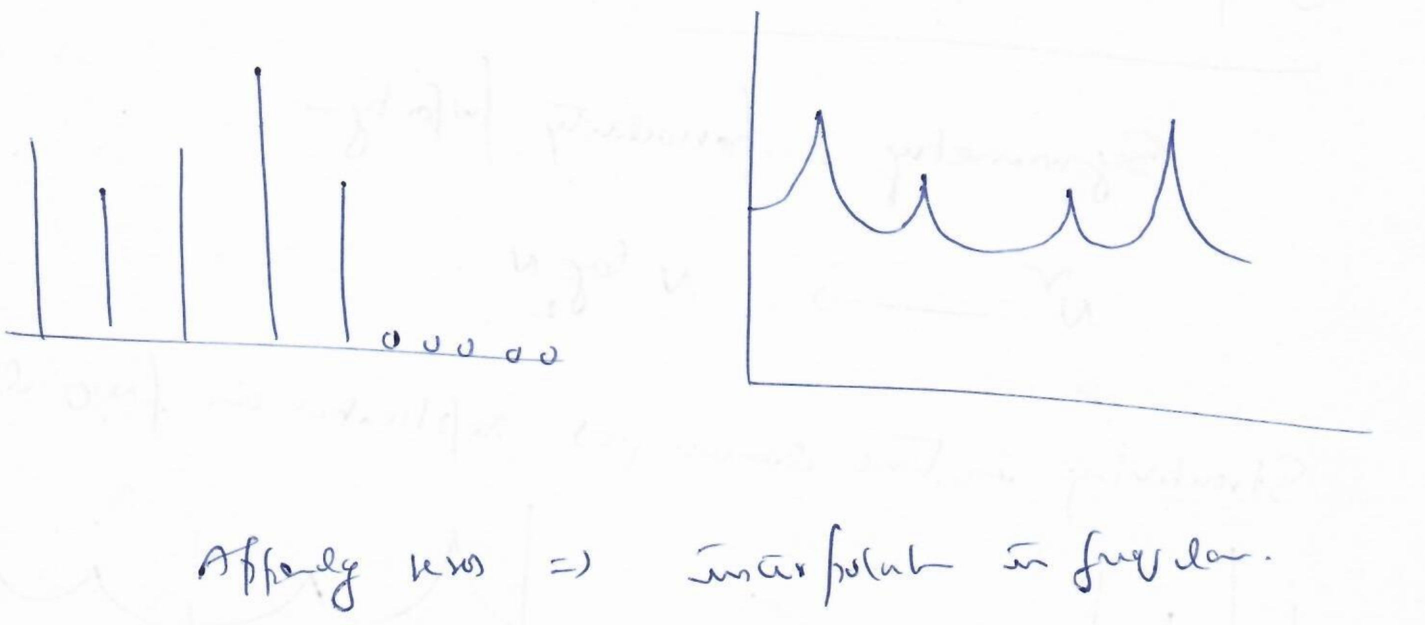
Stretching in time domain $\Rightarrow$ replication in freq domain



Decimation



Appending zeros



Appending zeros $\Rightarrow$ interpolation in frequency domain.

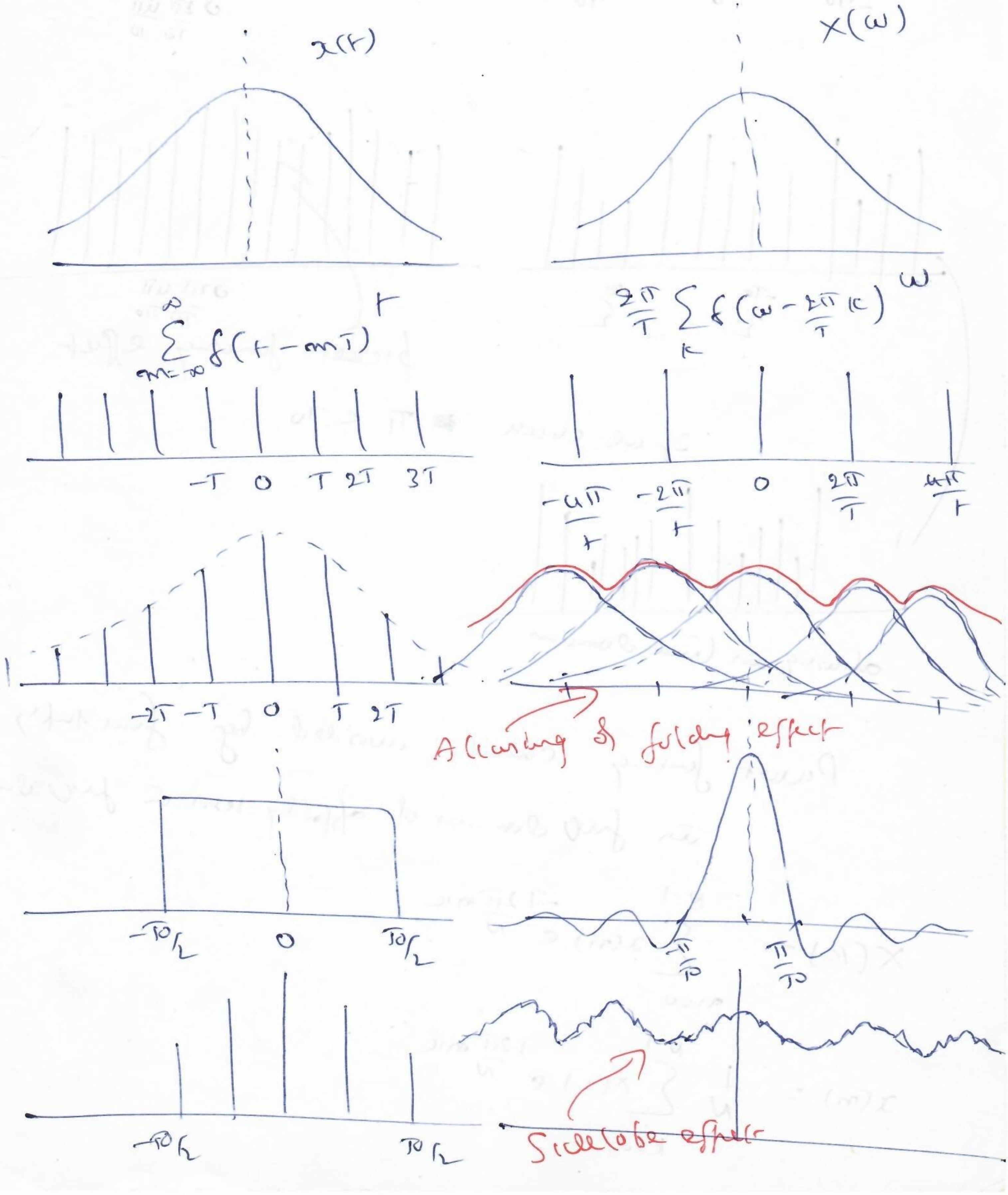
Linear convolution vs Circular convolution

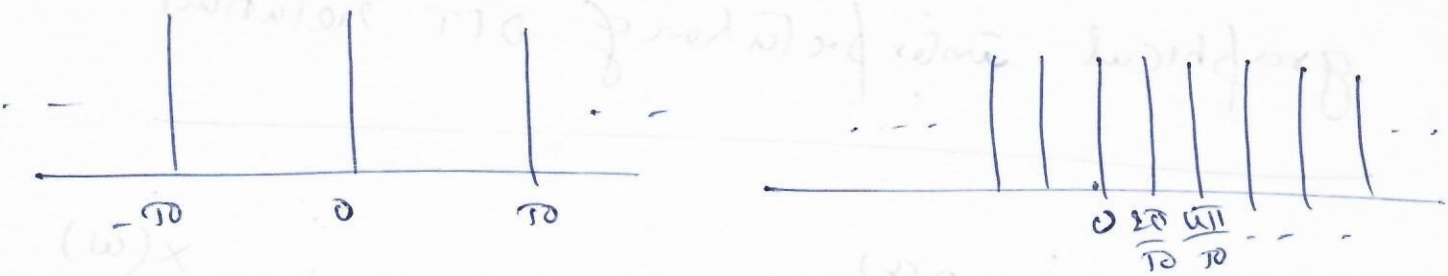
$$N(N+M-1) \longrightarrow 3(N+M-1)\log(N+M-1)$$

Processing of Speech Signals

5/2/08

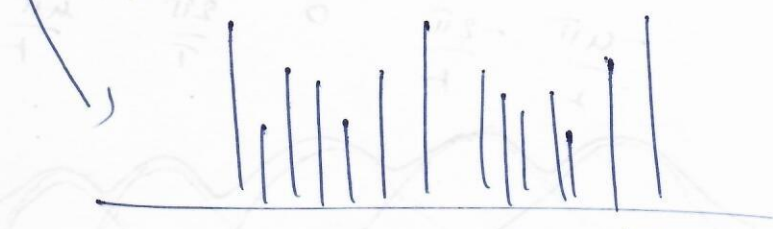
graphical interpretation of DFT relations





picket fencing effect

If we choose $T_1 < T_0$



aliasing in time domain

Picket fencing can be avoided by fine tuning in freq domain or appending zeros in freq dom

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-j\frac{2\pi}{N}nk}$$

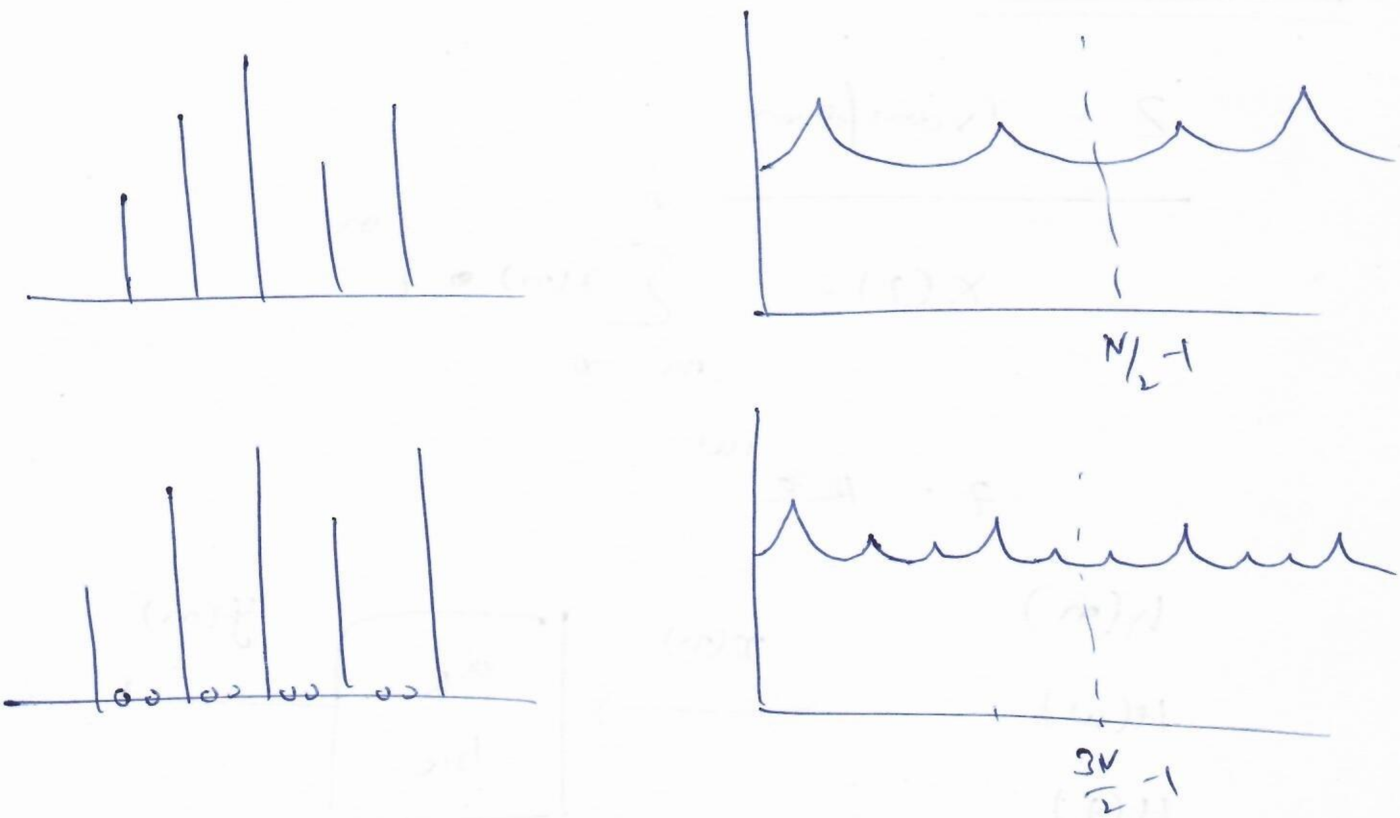
$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)\, e^{j\frac{2\pi}{N}nk}$$

Important DFT properties

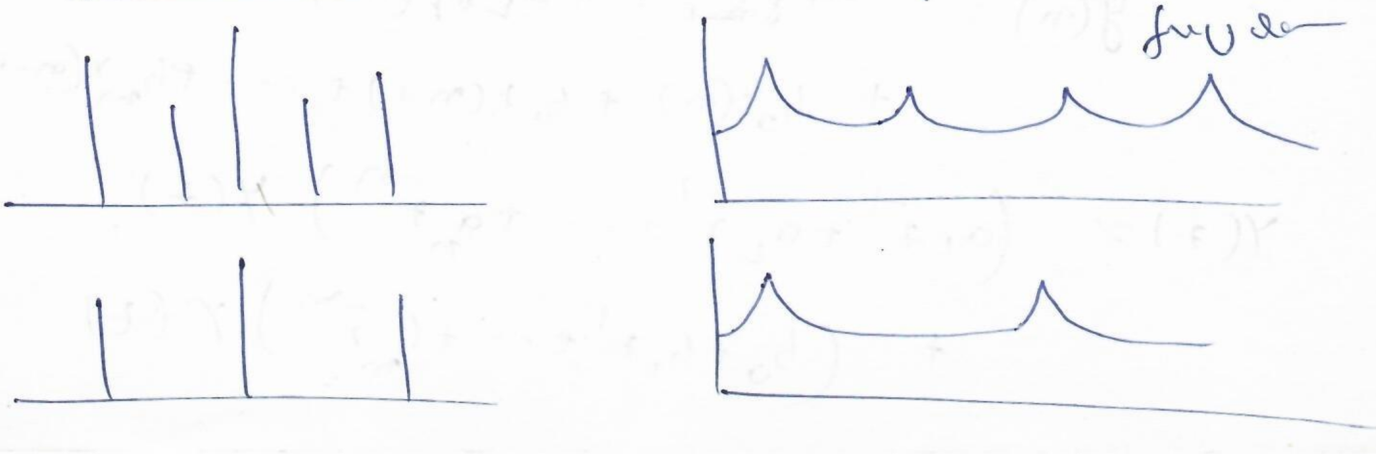
Symmetry & periodicity

$N^2 \longrightarrow N \log_2 N$ (FFT computation)

(1) Stretching in time domain => Replication in freq domain.

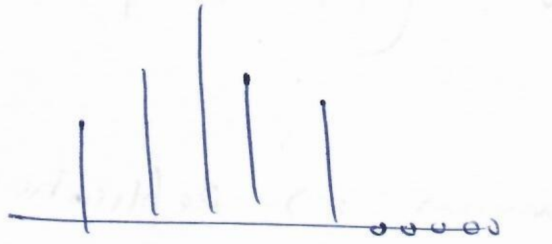


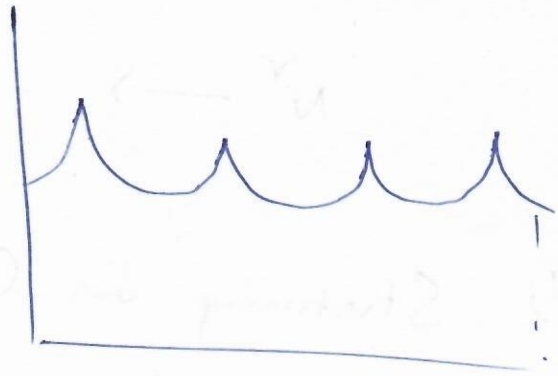
(2) Decimation => remove samples => aliasing in freq do



(3) Appending zeros in time domain

$\Rightarrow$ interpolation in freq domain





Z - Transform

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) \cdot z^{-n}$$

$$z = Re^{j\omega}$$

$h(n)$

$H(\omega)$

$H(z)$

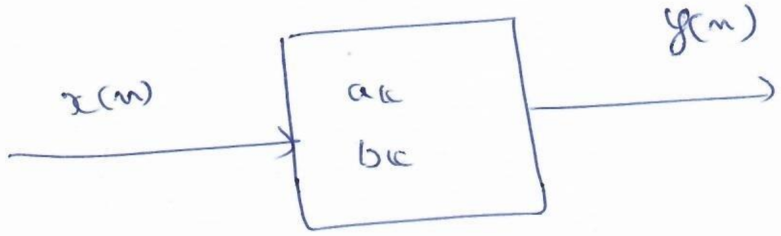
$$x(n) \longrightarrow \boxed{\begin{matrix} a_k \\ b_k \end{matrix}} \longrightarrow y(n)$$

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \cdots + a_N y(N-1)$$
$$+ b_0 x(n) + b_1 x(n-1) + \cdots + b_m x(n-m)$$

$$Y(z) = \left(a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}\right) Y(z)$$
$$+ \left(b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}\right) X(z)$$

Equivalent Representation of Signals & Systems

1. $h(n)$ — Basic rep of system (time domain)
2. $H(\omega)$ — Visualize some important info, which is not seen in $h(n)$
3. $H(z)$ — Design, compact rep, easy to implement convolution
4. (a_k, b_k) — linear const coeff difference eq
5. Pole-zero rep



$$y(n) = -\sum a_k y_{n-k} + \sum b_k x_{n-k}$$

$$= -\sum_{k=1}^{M} a_k y(n-k) + \sum_{k=0}^{N} b_k x(n-k)$$

$$Y(z) = -\sum_{k=1}^{M} a_k z^{-k} Y(z) + \sum b_k z^{-k} X(z)$$

$$Y(z)\left[1 + \sum_{k=1}^{M} a_k z^{-k}\right] = \sum b_k z^{-k} \cdot X(z)$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{N} b_k z^{-k}}{\sum_{k=0}^{M} a_k z^{-k}}$$

$$Y(z)\left[1-\sum_{k=1}^{N} a_k z^{-k}\right] = \sum_{k=0}^{m} b_k z^{-k} X(z)$$

$$H(z)=\frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{m} b_k z^{-k}}{1-\sum_{k=1}^{N} a_k z^{-k}} = \frac{\sum_{k=0}^{m} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}$$

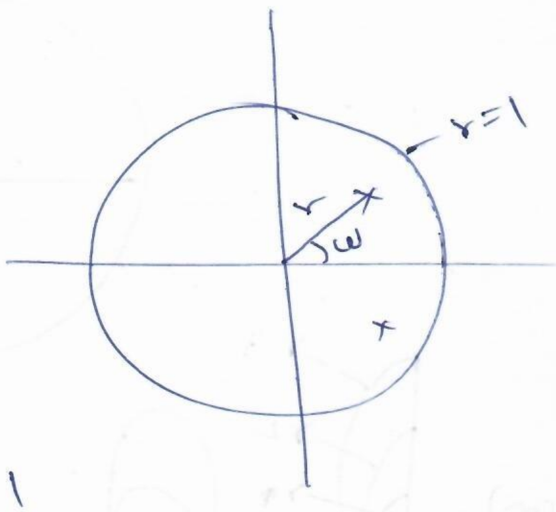
If $a_k = 0$ $\Rightarrow$ All zero sys, FIR sys

If $b_k = 0$ $\Rightarrow$ All pole sys, IIR sys

Poles & zeros of the system

$$r = e^{-\pi BT}$$

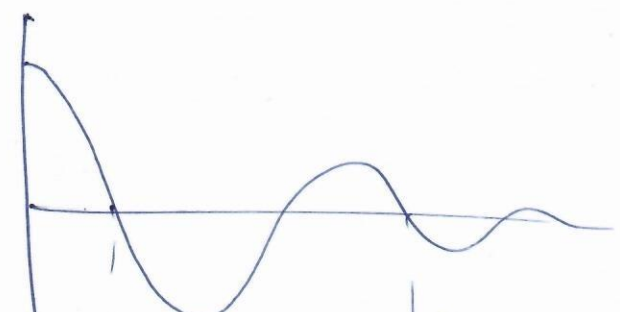
$$\omega = 2\pi FT$$



Complex conjugate poles

$r < 1$

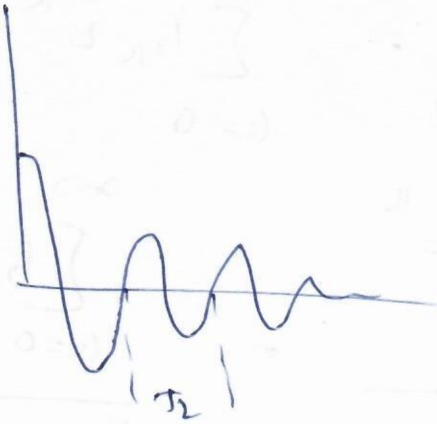
$h(n)$

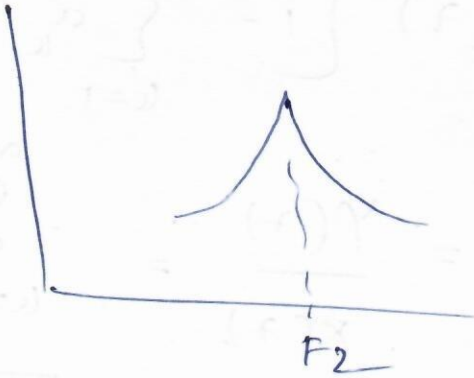


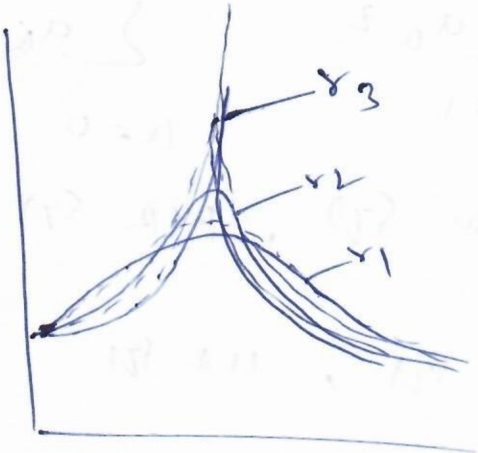
$T_1 = \frac{1}{F_1}$

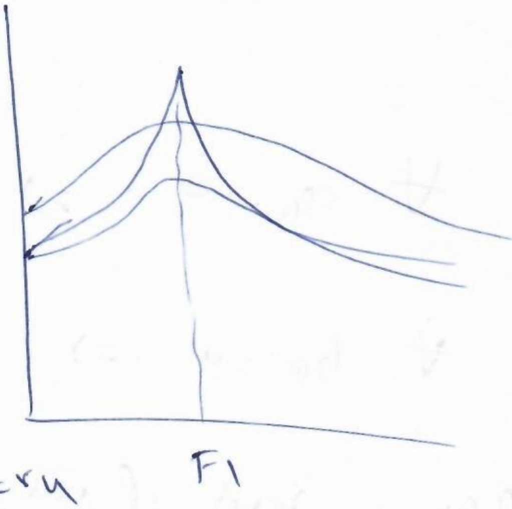
$H(\omega)$

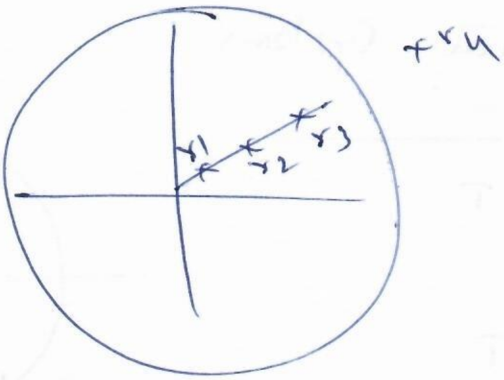
F_1

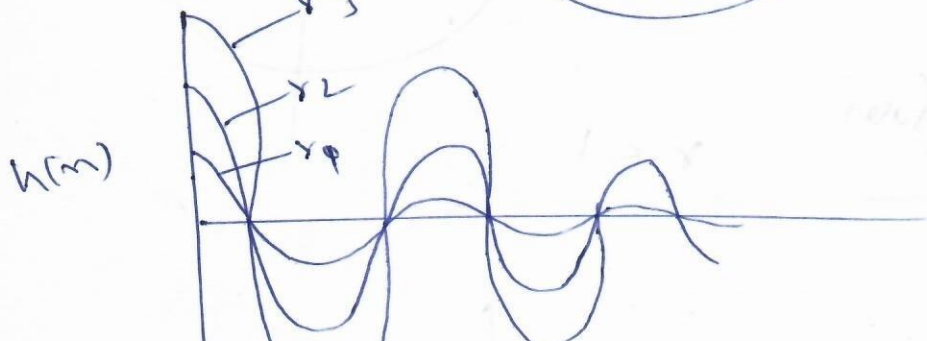

r2

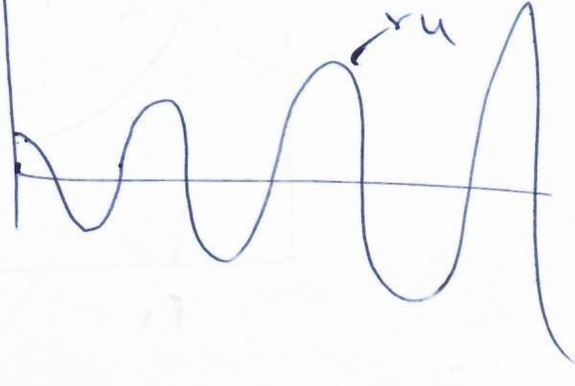

F2


r3
r2
r1


F1


+r4
r1
r2
r3


h(m)
r3
r2
rφ


r4

Pole on the real axis