Digital Signal Processing

Prof. S. C. Dutta Roy

Department of Electrical Engineering

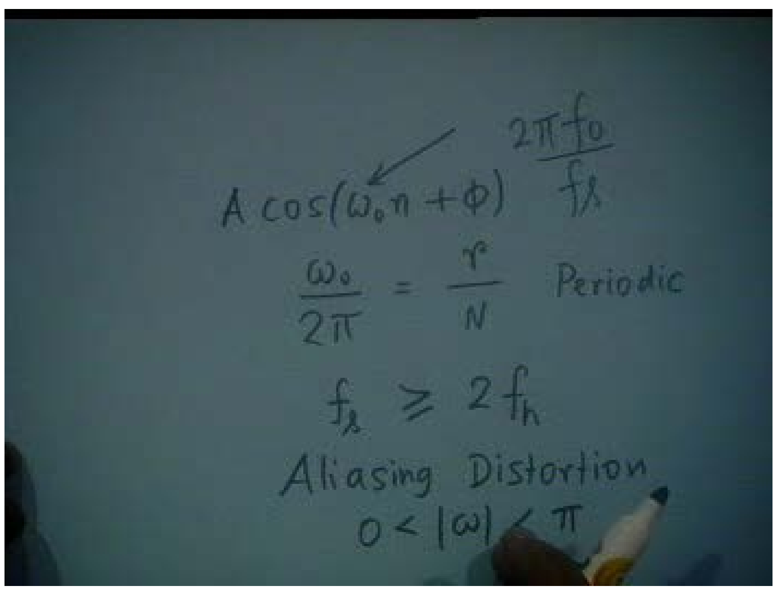
Indian Institute of Technology, Delhi

Lecture - 4

Characterization, Description and Testing of Digital Systems

This is the 4th lecture on Digital Signal Processing and our topic for today is Characterization, Description and Testing of Digital Systems. In the last lecture, we had introduced the exponential sequence and the sinusoidal sequence and we observed that a sinusoidal sequence which is A cos $(\omega_0 n + \varnothing)$ is not necessarily periodic.

(Refer Slide Time: 01:27 – 03:22)

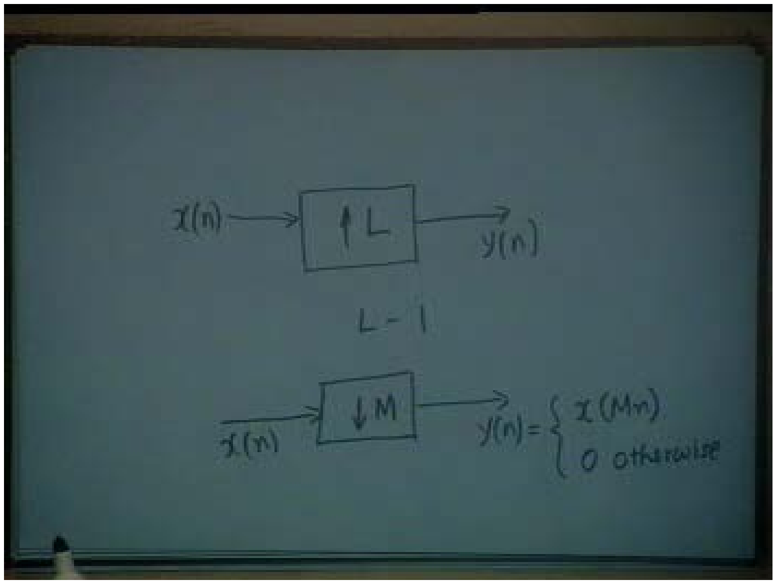


We examined the question of periodicity of sinusoidal sequences and we observed that it is periodic only if $\omega_0/(2\pi)$ is equal to a rational number, that is an integer divided by another integer. Then we looked at the sampling process and we showed by an example that the sampling frequency f_s must be greater than or equal to twice the highest frequency content of the signal. If this is not obeyed, then aliasing distortion occurs. Aliasing distortion means that high frequencies pose as low frequencies and that is the origin of the term aliasing. This has an effect on the band of frequencies that we should focus on.

I told you that ω_0 is a normalized digital frequency, it is actually $2 \pi f_0/f_s$ and the range of ω is restricted to $| \omega | \leq \pi$. The magnitude sign is indicated because actually the band is $-\pi$ to $+\pi$. Then we discussed operations on sequences, basically four in number, namely: multiplication, addition, time reversal and delay. Then we took an example of the schematic diagram of a digital system. We showed that you can write the difference equation from the given diagram and vice versa. That is, given the difference equation you can draw a schematic diagram. The schematic diagram is a description of the hardware as well as software; therefore it is a very useful form. Then we discussed some examples of digital systems.

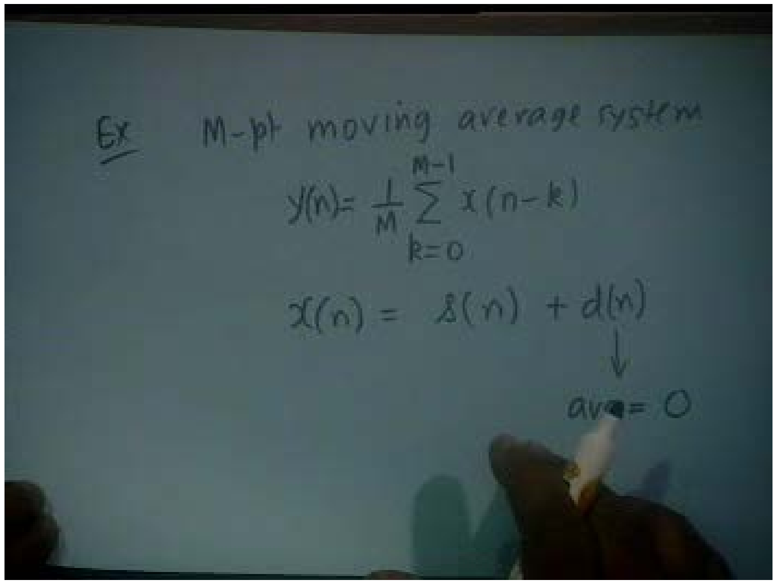
First, we took the example of an accumulator and then we took an example of an up sampler. Incidentally an up sampler is schematically represented like this. The input to an up sampler is x(n). If the sampling is increased by the factor L then the representation is an upward arrow. The factor by which the sampling rate is altered or increased is indicated beside the arrow. Basically, we showed, this amounts to padding L – 1 0's between two consecutive samples. Therefore the picture of the signal is not altered; only it is stretched with 0's in between two consecutive samples. In a similar manner, we can also think of a down sampler. Naturally the down sampler shall have an arrow pointing down. If the factor of down sampling is M, then the description of y(n) is simply x(Mn), and 0 otherwise. This simply means that the down sampler ignores M – 1 samples between 0^{th} and M^{th} also between M^{th} and twice M^{th} and so on. You see y(n) exist only for argument of x = 0, M, twice M, etc or on the other side at – M, – twice M and so on. In between 0 and M^{th} sample there will be M – 1 samples of x(n) which it simply ignores. This is the process of down sampling.

(Refer Slide Time: 04:28 – 06:26)



Next we take another example of a digital system, that of an M point moving average system. The description of this is that y(n) is taken as the average of the present value and the past M – 1 values. It sums up; it is an accumulator in that sense. Accumulation is over x(n), the present input, x(n – 1), x(n – 2) up to n – M + 1. That is, the present sample and the past M – 1 samples accumulate and then we divide by M to get an average. And it is moving average because as n changes the samples also change. So as you move with n or as you move with time, the average also changes. This is a very useful device for data smoothing. Suppose you have a given x(n) which is corrupted by some noise s(n) due to some reason. When you take an average, the noise component diminishes. In a process where the number of samples M is large, the average of s(n) would be equal to 0. Therefore moving average system is also called a data smoothing system. It is very much used in experimental observations.

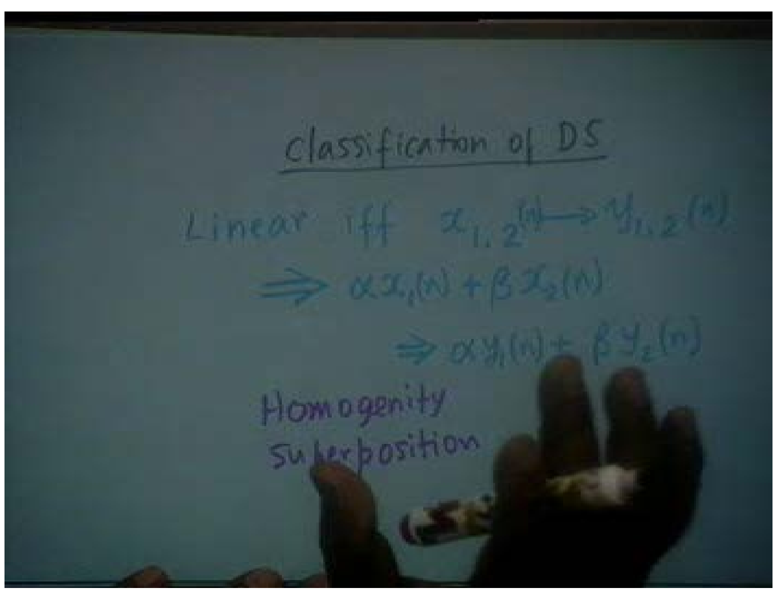
(Refer Slide Time: 06:51– 09:03)



And this brings us to classification of digital systems. There are various kinds of digital systems. We shall look at the classification and then concentrate on a particular class of digital systems for the purpose of this course. First, a digital system can be either linear or non linear. The formal definition is that a digital system is linear if and only if $x_{1,2}$ leads to $y_{1,2}$, then $\acute{\alpha} x_1 + \beta x_2$ leads to $\acute{\alpha} y_1 + \beta y_2$ (of course the argument is n). The symbolism means that if x_1 is the input, y_1 is the output; if x_2 is the input, then y_2 is the output. Also, "implies" is indicated by an arrow with two horizontal lines.

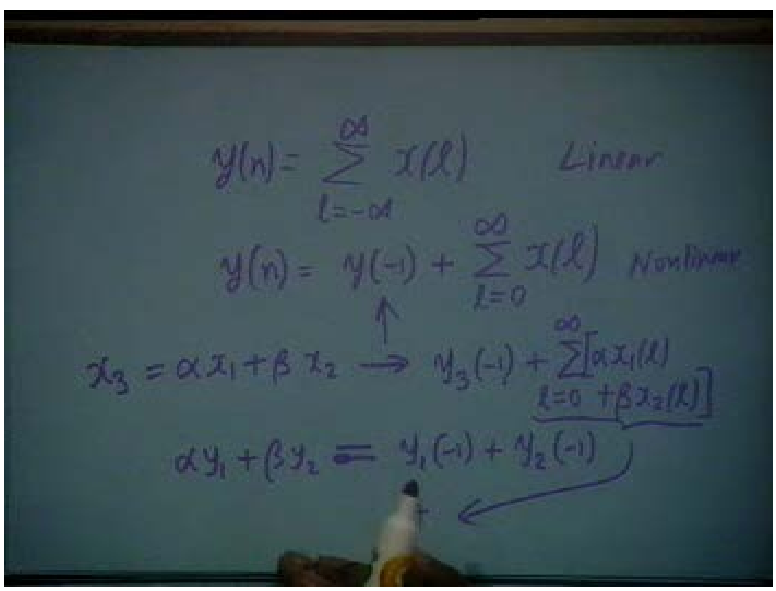
This implies that if ά and β are arbitrary constants, then an input $\acute{\alpha} x_1 + \beta x_2$, made by a linear combination of previous two inputs, should lead to $\acute{\alpha} y_1 + \beta y_2$. If this condition is obeyed then the system is said to be linear. You notice that linearity implies two principles; one is called homogeneity which says that if the input is multiplied by a constant the output will also be multiplied by the same constant. That is, with x_1 leads to y_1, if x_1 is multiplied by ά, then y_1 should also be multiplied by alpha. This has a very important corollary, that is, if ά is equal to 0 that is if you have a zero input, the output should also be zero. This is the principle of homogeneity. The other principle involved is superposition, that is we superimpose the 2 inputs άx_1 (n) and βx_2 (n), then the output should also be a superposition of the individual outputs. And the system is linear only if it obeys both of them: homogeneity as well as superposition. There are systems which obey homogeneity but not super position: they are nonlinear systems. Similarly there are systems which obey superposition but not homogeneity; they are also nonlinear systems. A linear system is one in which both homogeneity and superposition are valid. One of the important outcomes of this is that zero input for a linear system should lead to zero output. But this is only a necessary condition, it is not sufficient. Zero input leads to zero output only implies homogeneity, it does not imply superposition. Therefore it is a necessary condition but not sufficient. Homogeneity and superposition are necessary and sufficient conditions for linearity which means that if a system disobeys or violates one of these principles, it is non linear. A very simple test would be to ask: does zero input lead to zero output? If it does, then you proceed with further testing, not otherwise. If zero input does not lead to zero output, then a system is non linear. Then you can stop further testing; you do not have to test superposition.

(Refer Slide Time: 09:10 – 13:14)



Let us take several examples: one is y(n) = $\sum_{l=-\infty}^{\infty} x(l)$; it is an accumulator. This is a linear system because you find y_1 and y_2 due to x_1, x_2; multiply inputs by ά and β; superimpose to get άx_1 + βx_2. Observe that the output is άy_1 + βy_2. It is very simple; I do not have to carry out the steps. But look at this. The same system, as I told earlier in one of the lectures, can also be written as y (– 1) + $\sum_{l=0}^{\infty} x(l)$. This is not a linear system. Why not? It is because making x(ℓ) = 0 does not make y(n) = 0 because of the initial condition here. So the first system is linear but the second one is nonlinear. In the last characterization of the system y(– 1) is an initial condition that can be given any arbitrary value. But this non linear system shall become linear if y (– 1) = 0. If you want to test this system by rigorous means, which means that you take άx_1 + βx_2 as the input. What will be the output? Suppose this input we call x_3. Then the output y_3 will be y_3 (– 1) + $\sum_{l=0}^{\infty}[\alpha x_1(l) + \beta x_2(l)]$. On the other hand, ά$y_1$ + βy_2 shall be equal to y_1 (– 1) + y_2 (– 1) plus the same summation.

(Refer Slide Time: 13:38 – 16:41)

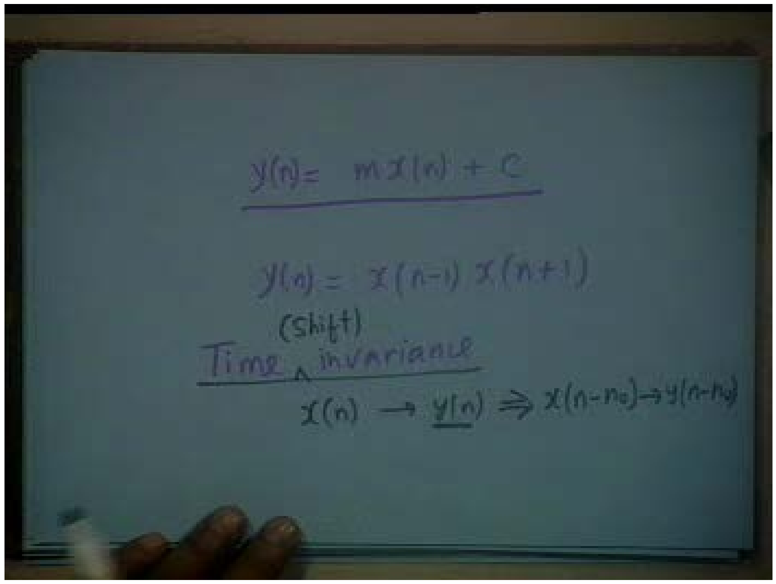


These two are not necessarily equal i.e. y_3 (– 1) is not guaranteed to be y_1 (– 1) + y_2 (– 1) because these initial conditions can be set arbitrarily. But our first test was sufficient i.e. x(n) = 0 does not lead to y(n) = 0. By the same token, if you have a system which is described by y(n) = M x(n)+ C, which is an equation to a straight line i.e. it is a linear equation, is not a linear system because C is an initial condition and x(n) = 0 does not leads to y(n) = 0. An initially charged capacitor is a nonlinear system. Initial flux in an inductor makes it a non linear inductor; this must be remembered. An equation to a straight line does not necessarily describe a linear digital system. By the same token, if you have y(n) = x(n – 1) × x(n + 1), the system is not linear because super position is not valid. If x(n) = 0 for all n, y(n) shall be = 0. Homogeneity is satisfied but not superposition. You can apply the test by taking input as $\alpha x_1 + \beta x_2$ and show that superposition is not valid.

We next go to time invariance. Because time has lost its significance in digital systems, it is also sometimes called as shift invariance because n – 1 simply means shift to the right by one sample; however we shall continue to use the nomenclature time varying or time invariant system. The definition is that if x(n) leads to y(n), then the system is time invariant if x(n – n_0) leads y (n – n_0). This looks very innocent. But there are systems in which it is a bit involved to be able to test whether the system is time variant or time invariant.

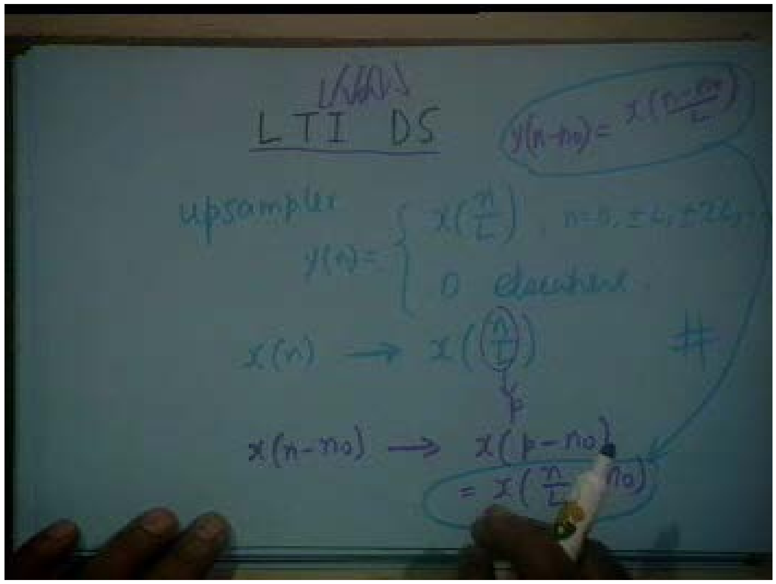
We shall take a few examples. The physical interpretation of time invariance is that the output waveform is preserved, with the same shift as in the input. Let the input be shifted by n_0 samples, and this shift can be to the left or to the right, i.e. n_0 is an integer either positive or negative. Then for a time – invariant system, the output shape shall remain the same, i.e. the shape of the plot of y(n) versus n, which is called the waveform remains the same, but shifted by the same amount as in the input. Therefore shift invariance or time invariance is also a waveform preserving transformation.

(Refer Slide Time: 16:54 – 20:01)



Now let us take one example at least for time invariance. Let me mention that if a system obeys both linearity and time invariance, it is called an LTI system and LTI Digital system shall be our focus in this course. Wherever a nonlinear or time varying system arises, we shall point it out specifically; if nothing is specified, then you take that it is an LTI system. Let me take my favorite example of an up sampler. The up sampler, as you have known, is described by y(n) = x(n/L), where L is the sampling rate increase factor. Y(n) exists for n equal to 0, ± L, ± 2L and so on, and it is 0 elsewhere. x(n) leads to x(n/L) which is non– zero only when n/L is an integer; so call n/L as small p which is an integer, then $x(n – n_0)$ should lead to $x(p – n_0)$ for time invariance. It is the argument of x which is shifted by n_0 and this is equal to $x((n/L) – n_0))$ whereas $y (n – n_0) = x((n – n_0)/L)$. This quantity and $x((n/L)– n_0)$ are not the same. One is $x((n– Ln_0)/L)$ and other is $x((n – n_0)/L)$; they are not identical. Therefore the shape of the waveform is not preserved when the input is delayed. So it is a time varying system. Similarly you can show that the down sampler is also a time varying system. Is the up sampler a linear system? Yes, it is a linear system, as you can easily show by taking an input $\alpha x_1 + \beta x_2$. To understand this concept to a little more depth, let us take a specific example.

(Refer Slide Time: 20: 32 – 23:03)

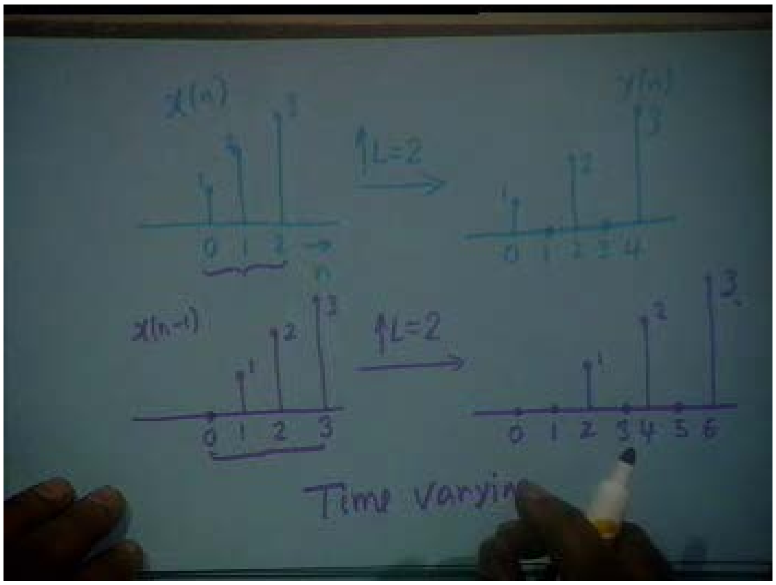


Suppose we have a x(n) which is 1 at n = 0, 2 at n = 1 and 3 at n = 2. If I up sample it by a factor 2 then what will happen? In between two samples, we shall have a 0. Therefore up sampling by a factor 2 will mean that the 0^{th} sample shall be preserved and then at n = 1 there shall be a 0, then at n = 2, there shall be a sample of amplitude 2, then another 0 at n = 3, followed by a sample of amplitude 3 at n = 4. So this is the up sampled version of the input waveform, which is my y(n). Now let us delay x(n) by 1 sample. Let us see what is x(n – 1). This will be x(n) shifted by 1 sample. So I will start with 1 at n = 1, then 2 at n = 2 and 3 at n = 3. The samples starting from n = 0 shall be 0, 1, 2, 3. If this is up sampled by the same factor 2, what will the waveform become? I shall have 0 at n = 0, then at n = 1, I shall have a 0 because between any two samples there is a 0. At n = 2, I shall have an amplitude 1, then at n = 3, I shall have a 0, at n = 4, I shall have an amplitude 2, at n = 5, there shall be a 0, and at n = 6, I shall have an amplitude 3. Is this one sample delayed version of y(n)? No, it is changed, the waveform has not been preserved. Thus the up sampler is a time varying system. Whenever you are in doubt for an analytical proof, you take an example, a very simple example, and that will clearly show whether the system is time variant or time invariant.

<A student asks a question:>

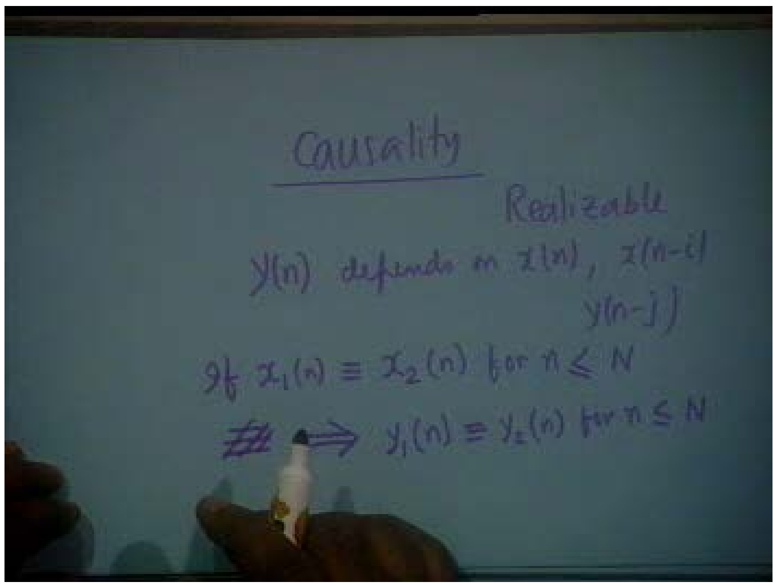
The question was whether the new output is not a delayed version of the previous output. Yes, it is, but the shift is by 2 samples and not one, as in the input.

(Refer Slide Time: 23:48 – 29:12)



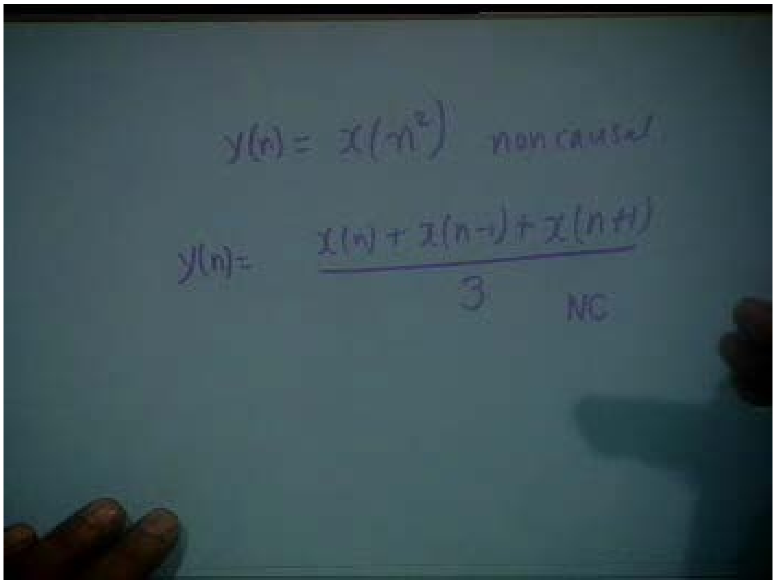
The next characterization we talk of is causality. First, let us understand conceptually what it means. A causal system is one which cannot predict the future. There are human beings who claim to be non causal; they can predict and are called Astrologers. However in electrical engineering systems, one cannot predict what will come in the future; they are called causal. So another name for causal system is also a realizable system; a casual system is realizable. Formally the definition is that y(n), the present output, depends on the present input x(n), its past values x(n – i), i being a positive integer, and perhaps also the past values of output y (n – j) where j is also a positive integer. Suppose y(n) depends on x(n + 1); that means the system at the present time predicts what will come one sample later. No, that is not permitted, we do not have a physical device which can advance time. But then let see what the formal definition is. Formal definition is in terms of formal mathematical language. It says that a digital system is causal if and only if $x_1(n) \equiv x_2(n)$ (three parallel lines means identically equal) for n less than or equal to some integer N implies that the output during this interval must also be identical for n less than or equal to N. What does it mean? It means that n exceeds capital N, x_1 and x_2 may be different, then y_1 and y_2 will also be different. But so long as the two inputs are identical what comes in future in x_1 or x_2 is not anticipated by the system. This is the formal definition and this is the interpretation.

(Refer Slide Time: 29: 45 – 33:49)



Let us take a very simple example and see whether the system is causal or non causal. Suppose you have a system in which $y(n) = x(n)^2$; is this causal? No, it is non casual, put n = 2 then y (2) = x(4), x(4) is yet to come, therefore this system is non causal. Suppose you have y(n) = [x(n) + x(n – 1) + x(n + 1)]/3; this is also a non casual system because y(n) depends on x(n + 1).

(Refer Slide Time: 33:59 – 34:42)

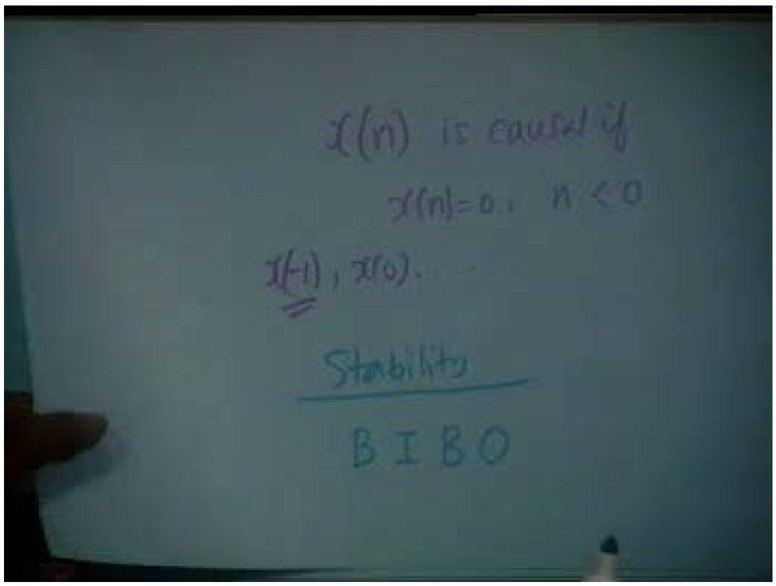


But then, if we cannot realize a non causal system and if all our systems are causal, why are we talking of it? Why are we making a distinction? Distinction arises because non causal signal processing is possible from recorded data. The geophysicist, for example, goes to the field and records the vibrations by a seismograph, gets back to the laboratory and then analyzes it. Now, when he is analyzing a particular output y(n), he knows what is going to come because this is a recorded data. So on recorded information non causal signal processing is possible. And in geophysics, in particular, and also in weather prediction, non causal processing is mostly possible. But you cannot realize this in hardware. It is data processing basically. Hardware realization always has to be causal, you cannot anticipate what is going to come. What about signals? Formally a signal x(n) is said to be causal if x(n) = 0 for n less than 0. You have to start the signal somewhere and you call that n = 0 and therefore the signal is causal. But then if you have values x(– 1), x(– 2) and so on, this is a non causal signal. So signals can be non causal. What is x(– 1)? You start your processing at n = 0; x(– 1) is therefore the initial value. If you combine this then it is a non causal signal. Mostly we shall talk of causal signals.

Causal system cannot predict what will happen in future. A signal is causal if it is 0 for n less than 0.

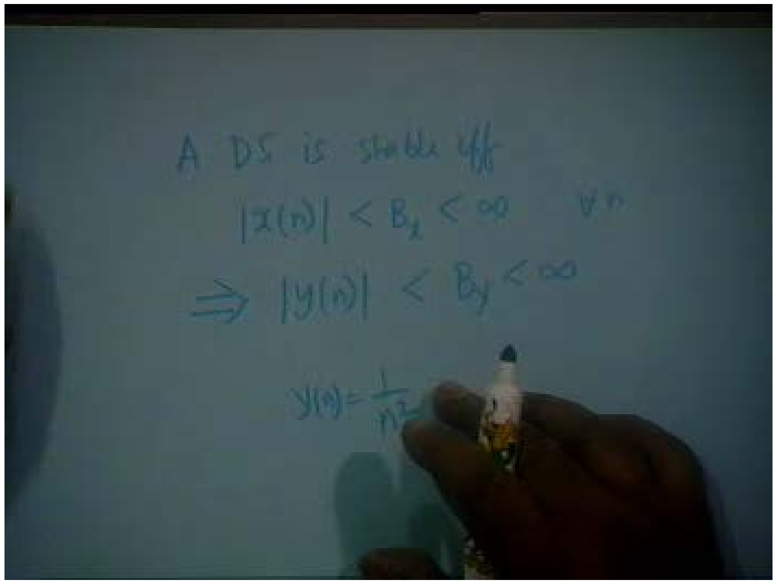
Our next concern is stability. Unless you want a digital oscillator, all digital systems, to be useful, have to be stable; there is no other way. And stability that we bother about in DSP is the so called BIBO stability, where BIBO stands for Bounded Input Bounded Output.

(Refer Slide Time: 36:21 – 38:24)



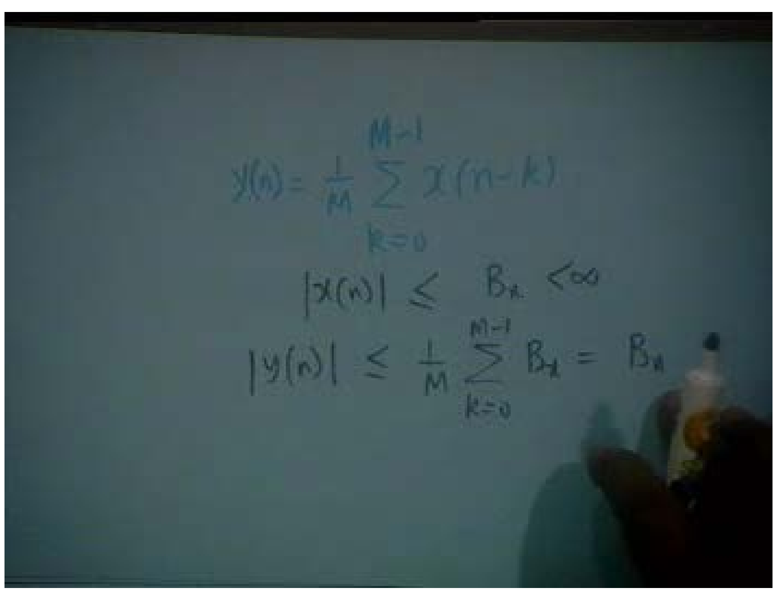
The formal definition is that a digital system is stable if and only if a bounded input x(n), i.e. $|x(n)| \le B_x < \infty$, for all n, implies that the output y(n) is also bounded i.e. $|y(n)| \le B_y < \infty$ for all n. If the input sequence is bounded the output sequence should also be bounded. For example, $y(n) = x(n)/n^2$ is an unstable system.

(Refer Slide Time: 38:29 – 39:30)



Let us take some examples. For the moving average system, $y(n) = \frac{1}{M}\sum_{k=0}^{M-1} x(n-k)$. Now, if x(n) is bounded by B_x, then obviously y(n) shall be less than or equal to B_x. Therefore this is a stable system.

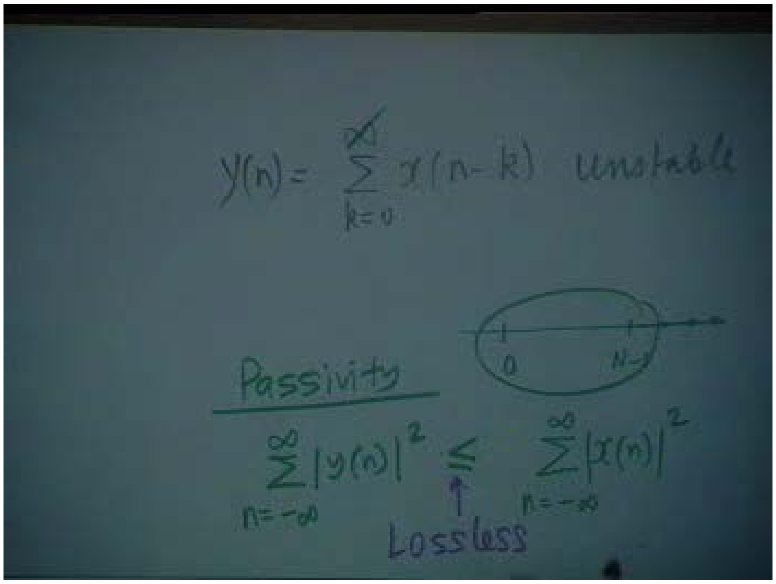
(Refer Slide Time: 39:41 – 40:40)



Suppose my y(n) is simply an accumulator, i.e. $y(n) = \sum_{k=0}^{\infty} x(n-k)$; is this a stable system? No, because there are infinite number of samples, so even if x(n) is bounded, the output is bound to go to infinity as n goes to infinity; so this is an unstable system. Suppose x(n) exists between 0 and N – 1, all the rest are 0; then obviously it is a stable system. Is stability inherently related to feedback? Not necessarily. In y(n) is $x(n)/n^2$ there is no feedback, yet the system is unstable.

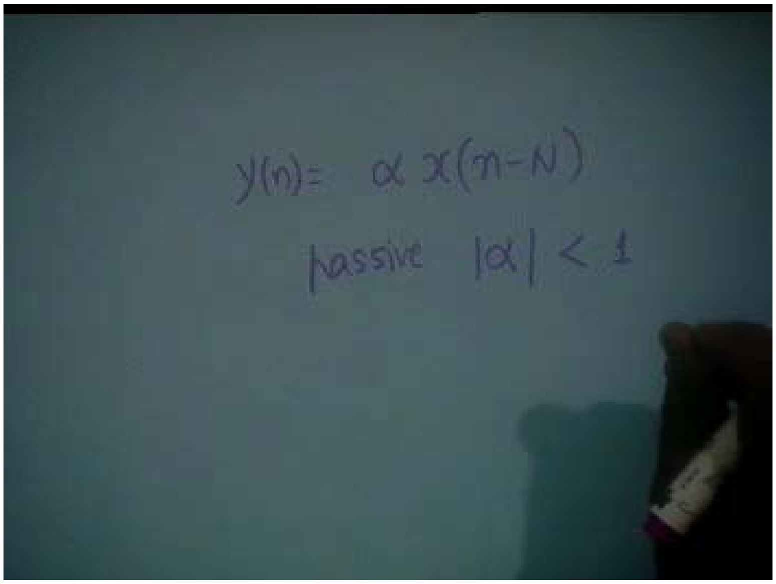
Next, we talk of passivity; the concept of passivity is very familiar in analog system. That is, the energy of the output cannot exceed the energy of the input. A digital system is said to be passive if this criterion is satisfied namely the energy of the output which by definition is $\Sigma|y(n)|^2$ where n can be from – infinity to + infinity, does not exceed the energy of the input i.e. $\sum_{n=-\infty}^{\infty} |x(n)|^2$ .The system is passive if the output energy is less than or equal to the input energy. In other words the system cannot generate energy. RLC networks are passive. On the other hand if you have RLC network combined with op amps or transistors it can become active. Where does the extra power come from? It comes from the power supply. A passive system is defined the same way as in analog systems. If the output energy is the same as the input energy, then you call this system as lossless. This is an artificial concept; there is no concept of energy in numbers, but we introduce this concept because it facilitates mathematical treatment.

(Refer Slide Time: 40:46 – 45:10)



We take an example; suppose y(n) = αx(n– N); then this would be a passive system if |α|, where α can be complex, is less than 1.On the other hand, if |α| is greater than 1 then it is an active system.

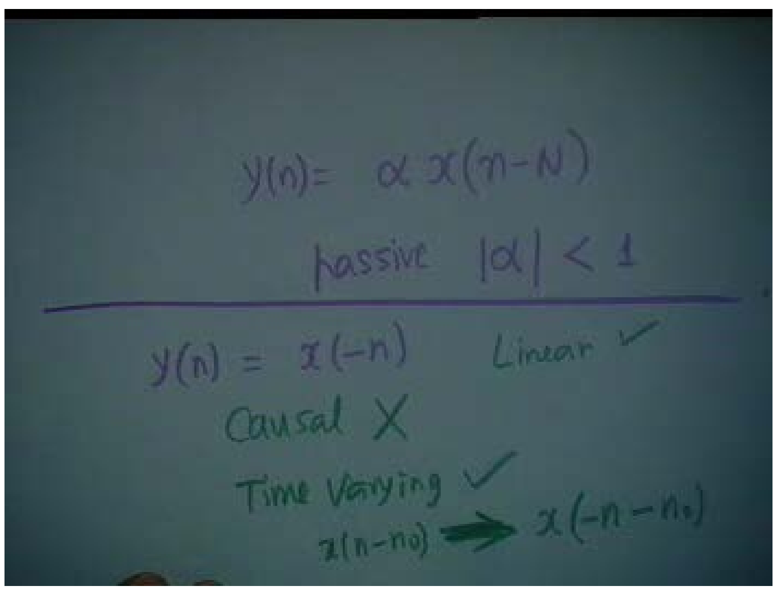
(Refer Slide Time: 45:15 – 45:45)



We conclude this lecture in a couple of minutes with another example. Suppose you have a time reversal system, y(n) = x(– n), a digital system which simply reverses the time. Is it a linear system? Yes, it is linear. As you can easily test, it obeys super position as well as homogeneity. Therefore it is a linear system. Is it causal? No, because put n = – 1 then your output is x(1), which will come in future, so it is non causal. Is it time varying? It is time varying, because $x(n - n_0)$ leads to $x(-n - n_0)$, not $x(-n + n_0)$.

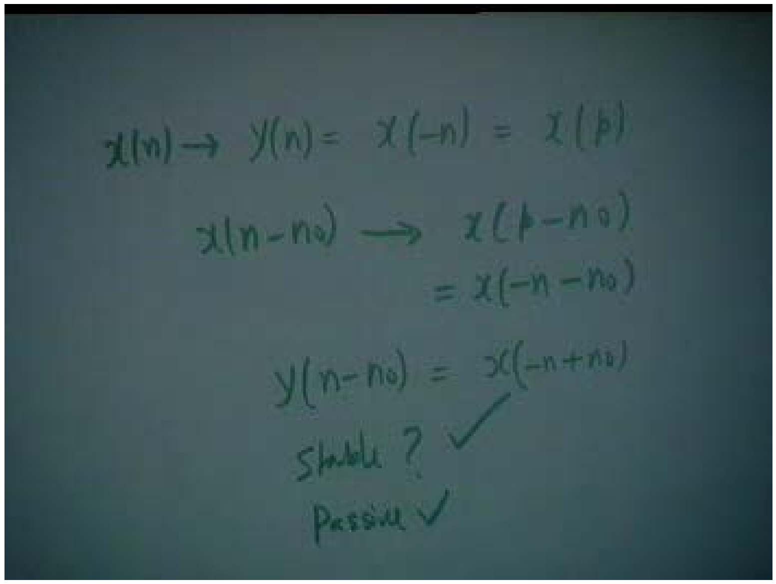
In such cases where there can exist a confusion, what you do is to write y(n) = x(– n) = x(p), then $x(n - n_0)$ shall lead to $x(p - n_o) = x(-n - n_0)$; it is the variable which is delayed and y(n) is equal to $x(-n - n_0)$, not $x(-(n - n_0))$.

(Refer Slide Time: 45:53 – 47:12)



Therefore it is a time varying system. Is it a stable system? Yes, of course. If x(n) is bounded, y(n) has to be bounded, so it is stable. Finally, is it passive? Yes it is passive.

(Refer Slide Time: 47:14 – 48:18)



In the next lecture, we shall talk about impulse and step response of a digital system and the correlation between the two.