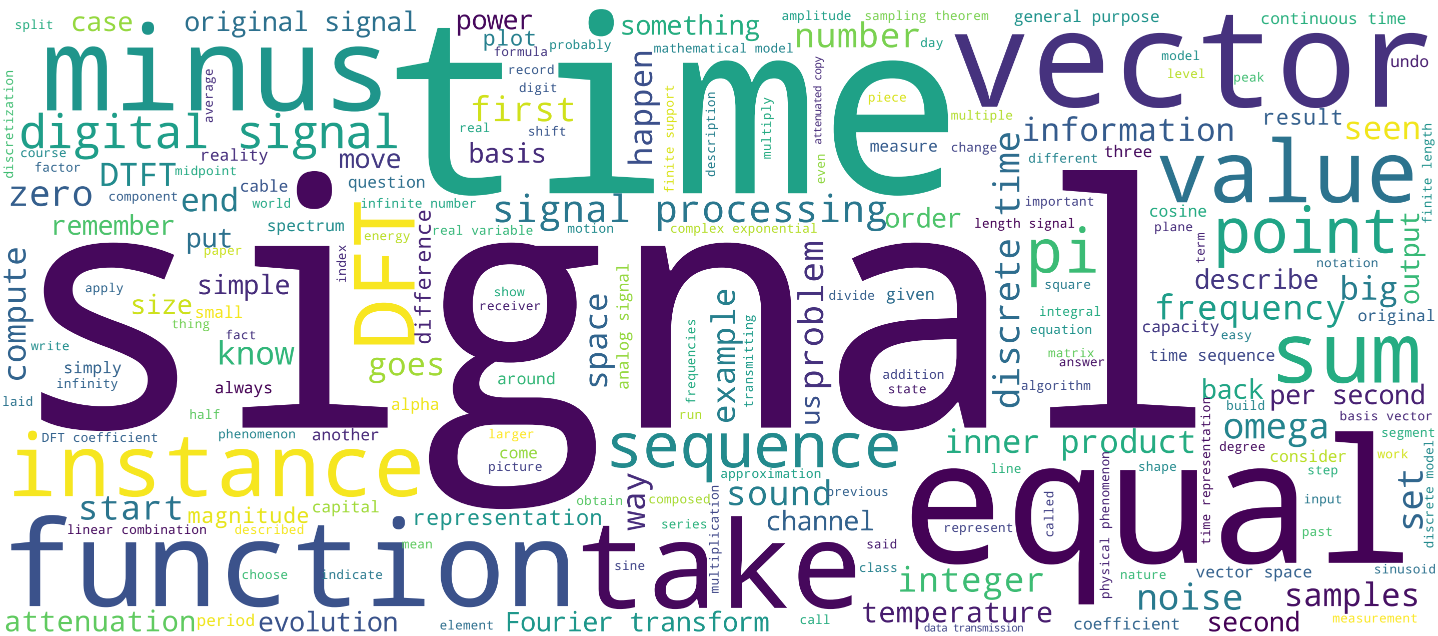
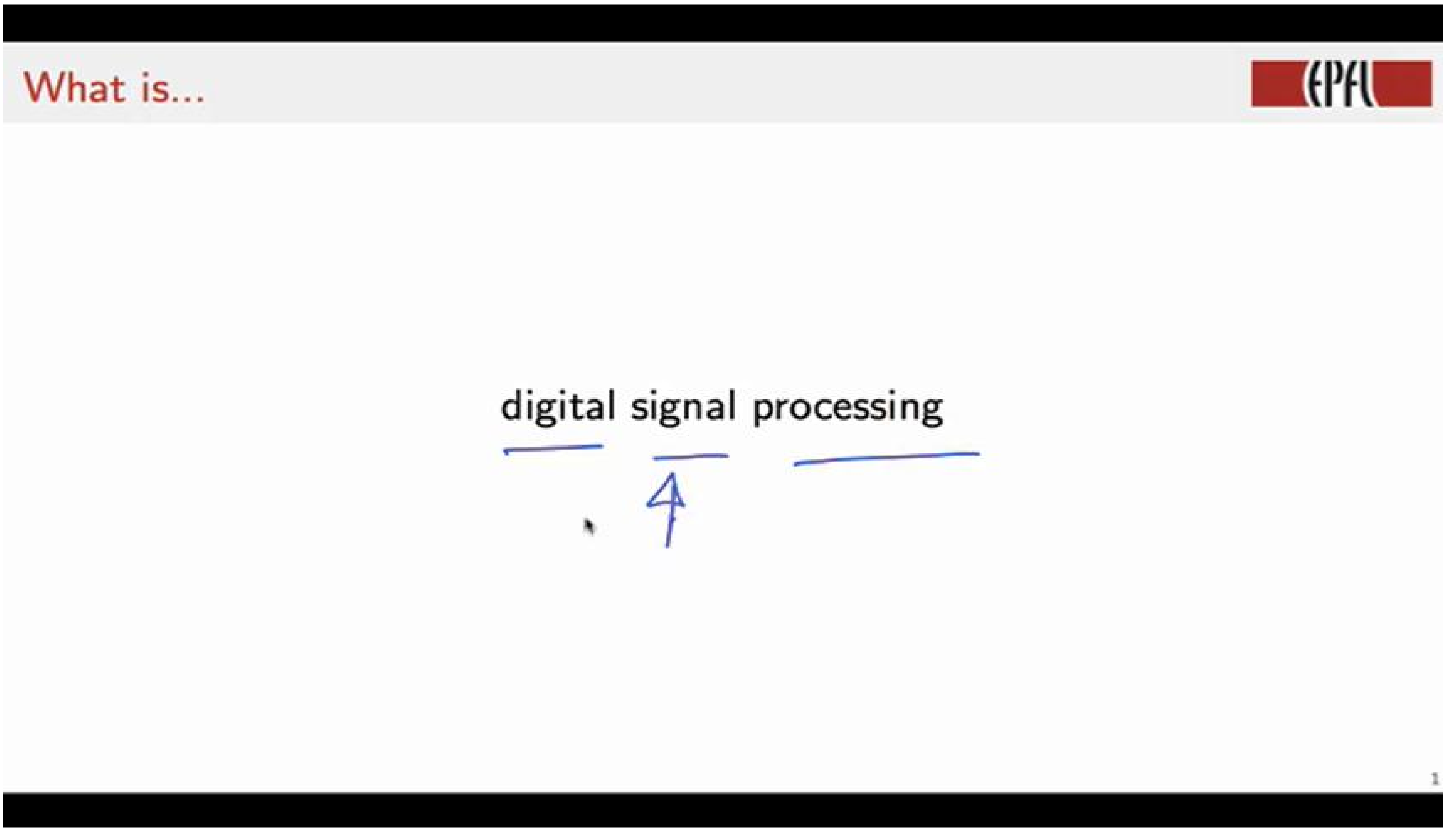
Digital Signal Processing

Introduction



Search MOOC

Video

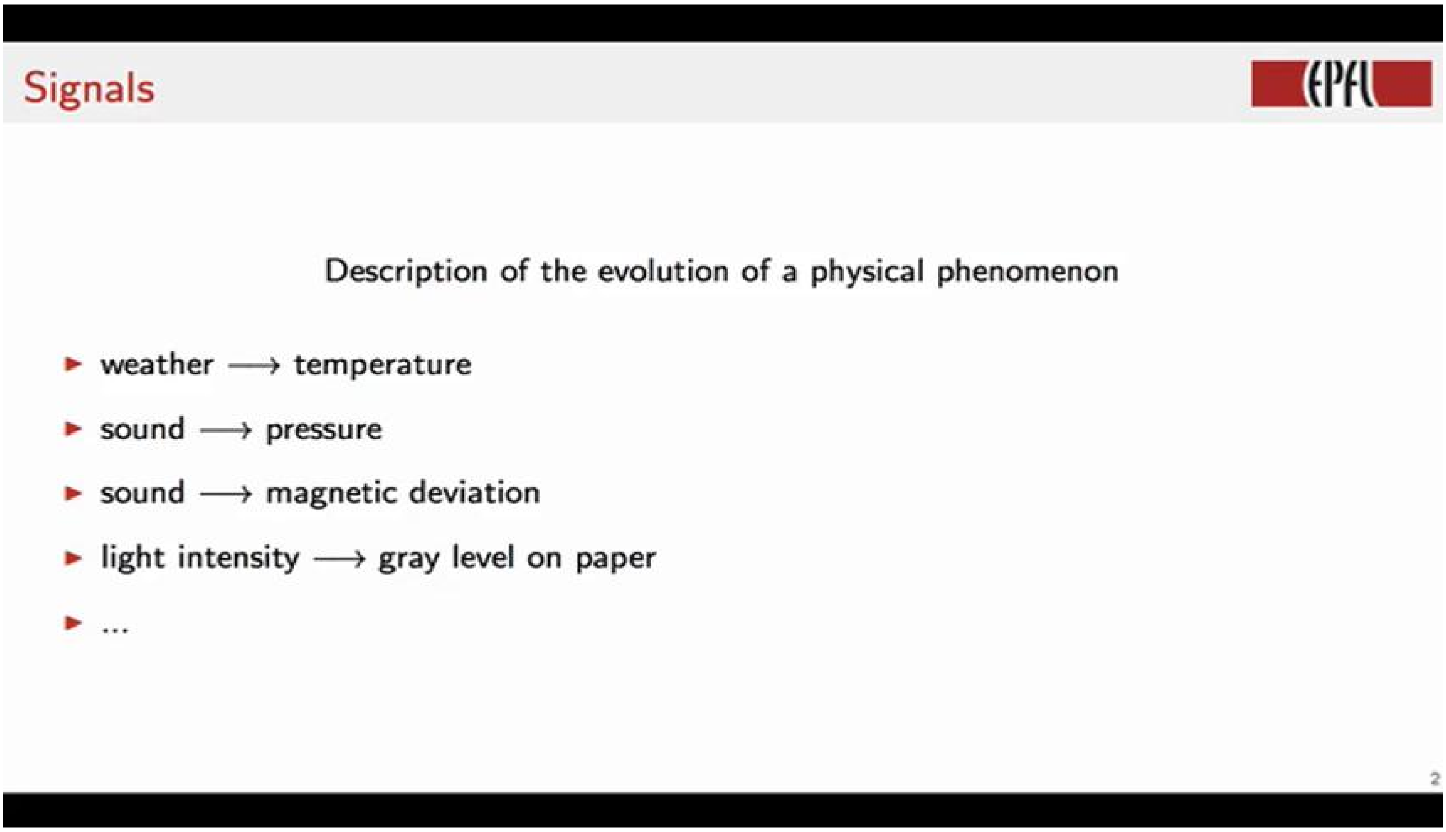


Hi, and welcome to the new edition of our Digital Signal Processing Class. In this introduction, we would like to talk about what signal processing is and what makes it so interesting and relevant today. So and we will start the usual way by picking apart the name of the class. So we have three words, digital signal processing. Let's start with a centerpiece signal, and try to define what a signal is.

Notes

Summary

0m 00s



The best definition that I can come up with is that a signal, is that the description of the evolution of a physical phenomenon. And that is best explained by example. So take the weather, for instance, if you measure the temperature over time, you have one possible description of the phenomenon weather, take a sound. A sound is generated by, say, my vocal tract by creating a pressure wave. If you measure this pressure at a point in space with a microphone, for instance, you have a description of the phenomenon sound, but of course, that is just one possible description. If the sound had been recorded on a tape recorder, for instance, the description of that phenomenon would take the form of the magnetic deviation that is recorded on the tape. In more dimensions, if you take the light intensity on the surface and you encode that as a gray level on paper, you have two dimensional black and white photograph, that is also a signal that varies in space rather than in time. So we have a signal.

Notes

Summary

0m 25s

Processing

Analysis: *understanding* the information carried by the signal

Synthesis: *creating* a signal to contain the given information

3

Now we have the processing part. And the processing part is where we make sense of this information that has been described by the signal. We can process a signal in two ways. We can analyse it, namely we want to understand the information carried by the signal and perhaps extract some more high level description, or we can synthesise a signal that's also signal processing. And that's when we create a physical phenomenon that contains a certain amount of information that we want to put out in the world. And this is really what we do when we transmit information, like when we use our cell phone or the radio or when we generate sounds with a music synthesiser.

Notes

Summary

1m 30s

Signals

Description of the evolution of a physical phenomenon

4

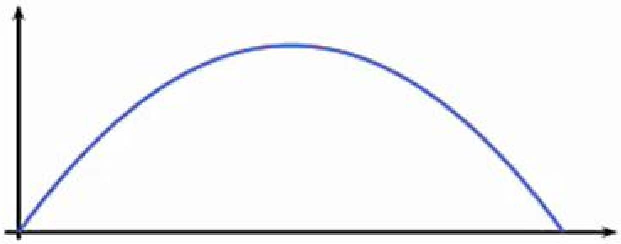
So we said that the signal is a description of the evolution of a physical phenomenon, but signals like so are not an exclusivity of digital signal processing.

Notes

Summary

2m 12s

Signals (physics)



$$y(t) = v_y t - \frac{1}{2}\vec{g}\, t^2$$

(Galileo, 1638)

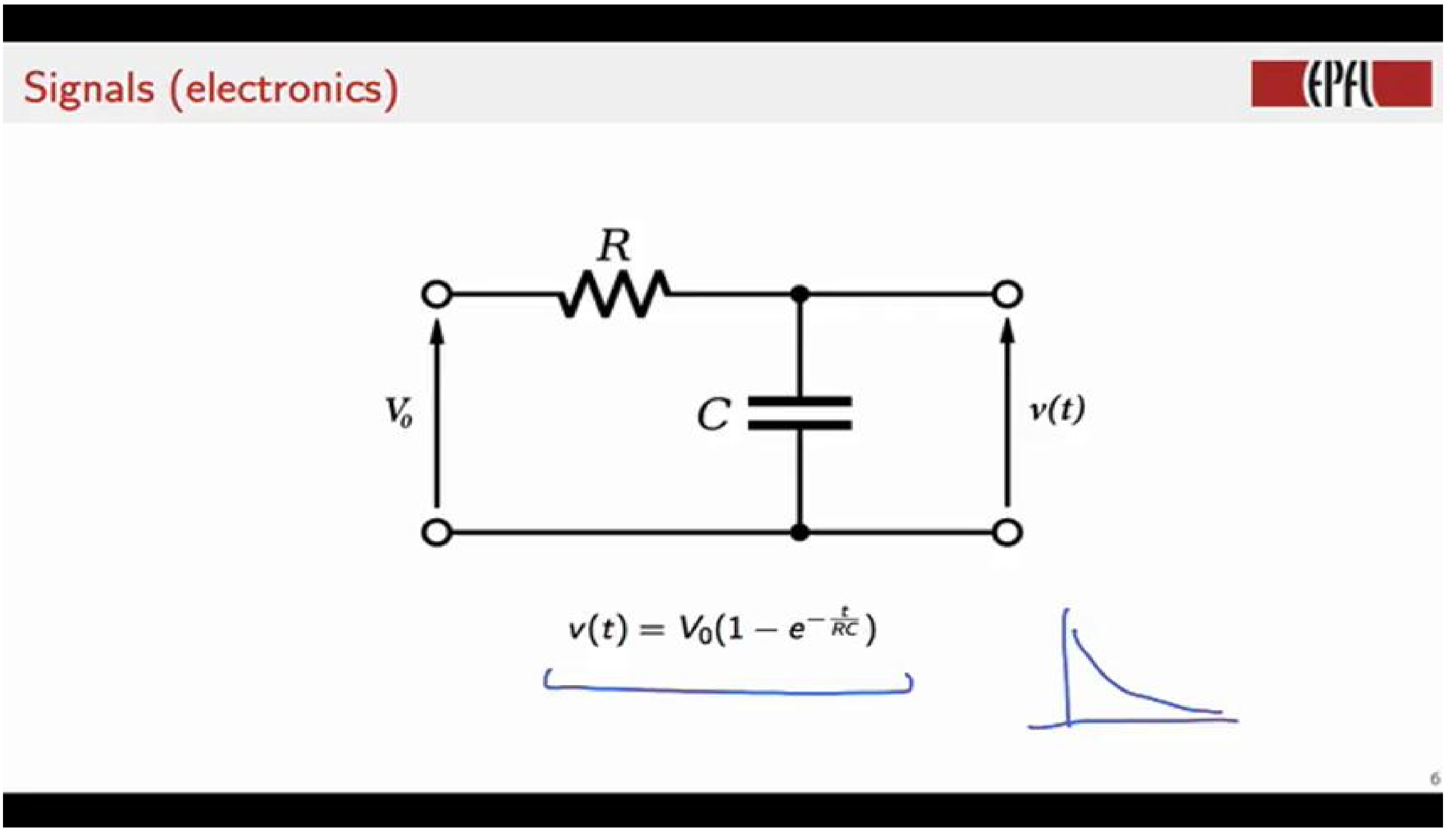
5

In physics, for instance, perhaps you remember the equation that describes the motion of a projectile and that was discovered by Galileo in 1338, and you have a simple equation that relates the vertical position to the initial velocity and to time. And you have this beautiful curve that describes the instantaneous position.

Notes

Summary

2m 22s



In electronics, for instance, you have a lot of equations that describe the inner workings of electronic circuits. Here you have an RC network and the voltage at the output is described by this law of charge or discharge of a capacitor that would be something like this. So what makes digital signal processing interesting?

Notes

Summary

2m 43s

Analog signals

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

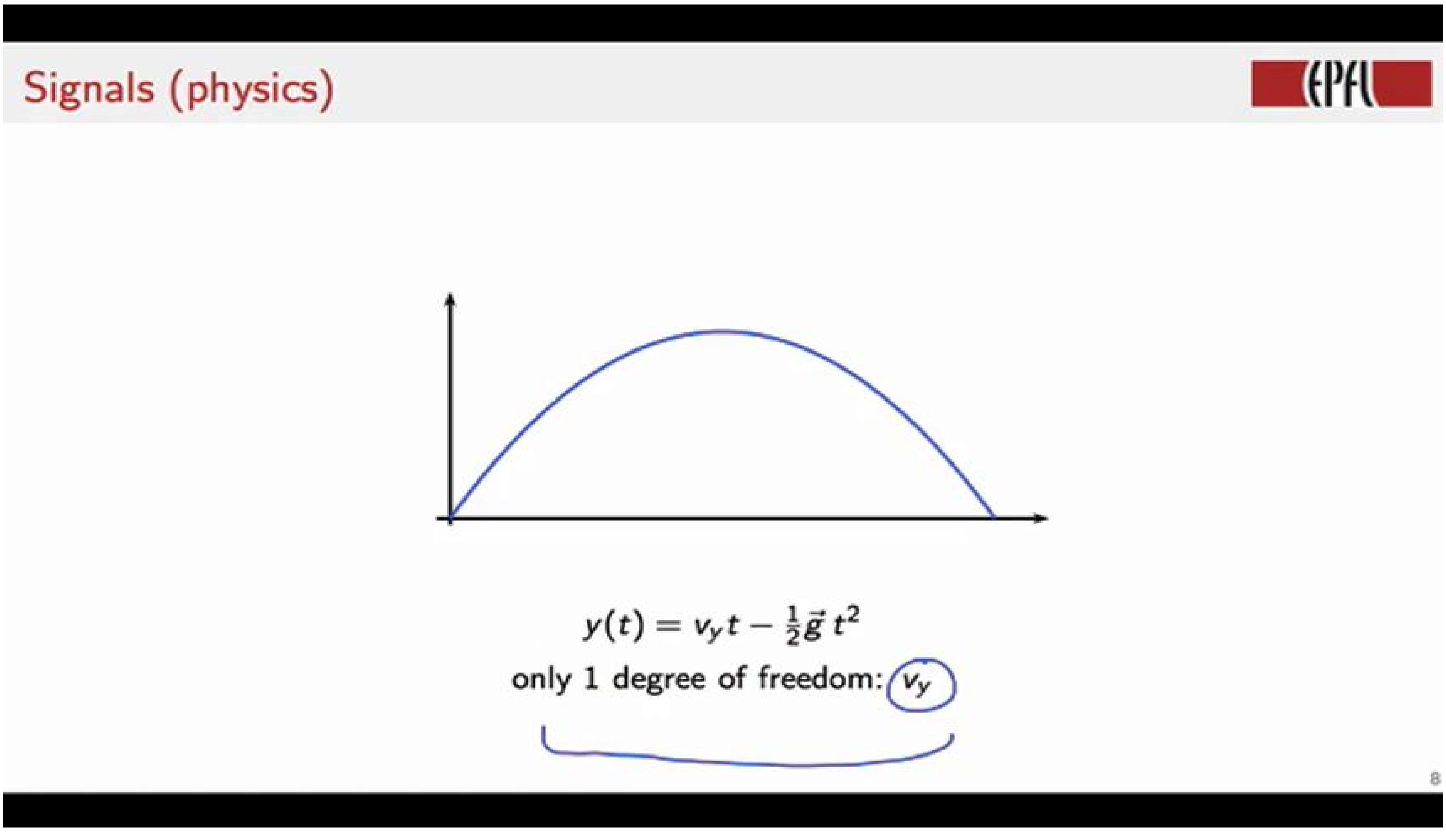
7

In these two examples, you can see that the mathematical model that we use to describe the signals is that of a function of a real variable time. This is a standard model for what we call analog signals.

Notes

Summary

3m 03s

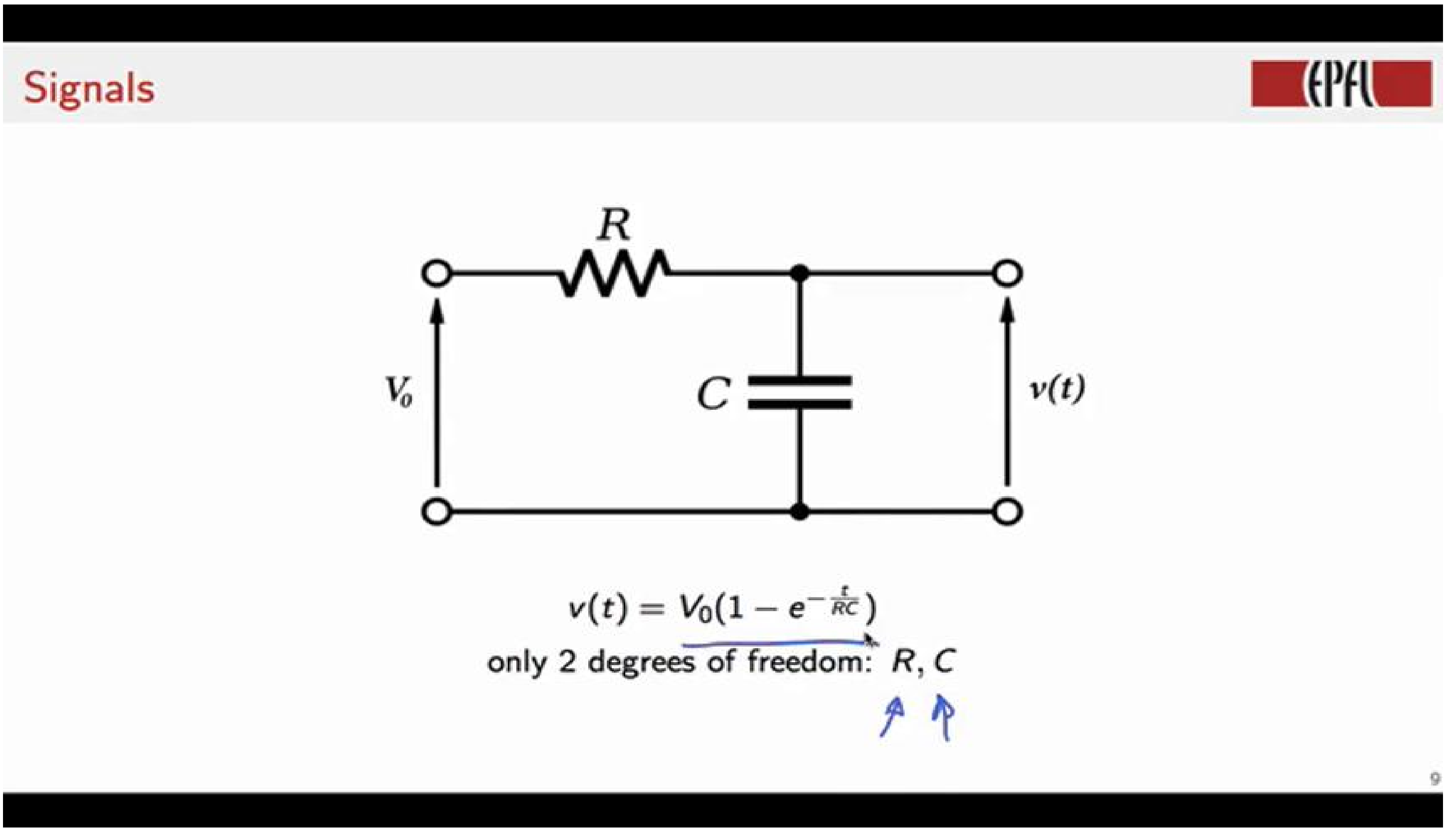


But if we go back to the motion of a projectile we see that this equation has only one degree of freedom, the only thing you can choose is the initial velocity, but then the shape of the signal will always be the same.

Notes

Summary

3m 15s

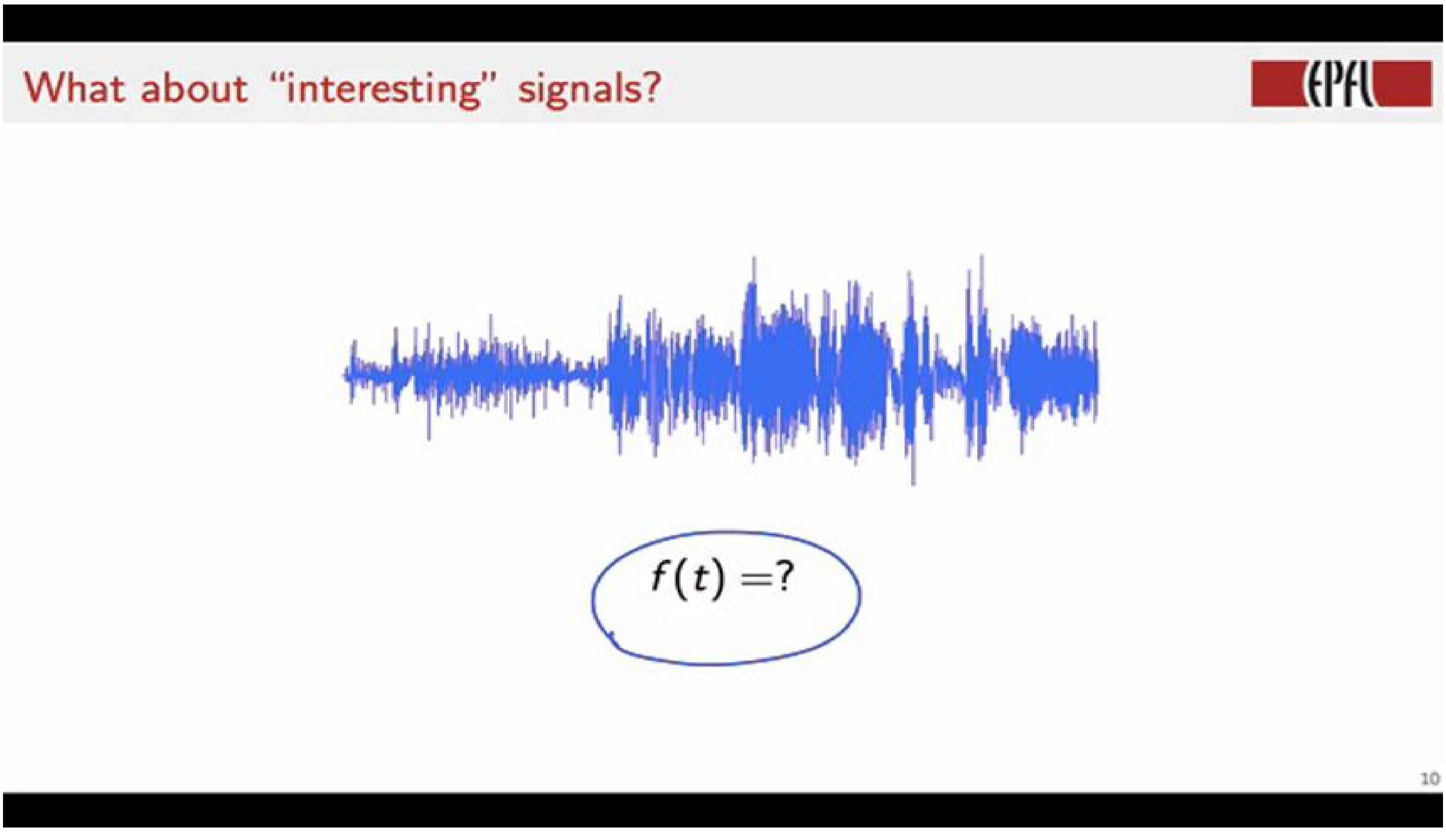


And similarly, in an RC network, you can choose the value for the resistor and for the capacitor, but the signal will always have the same exponentially decaying shape.

Notes

Summary

3m 26s

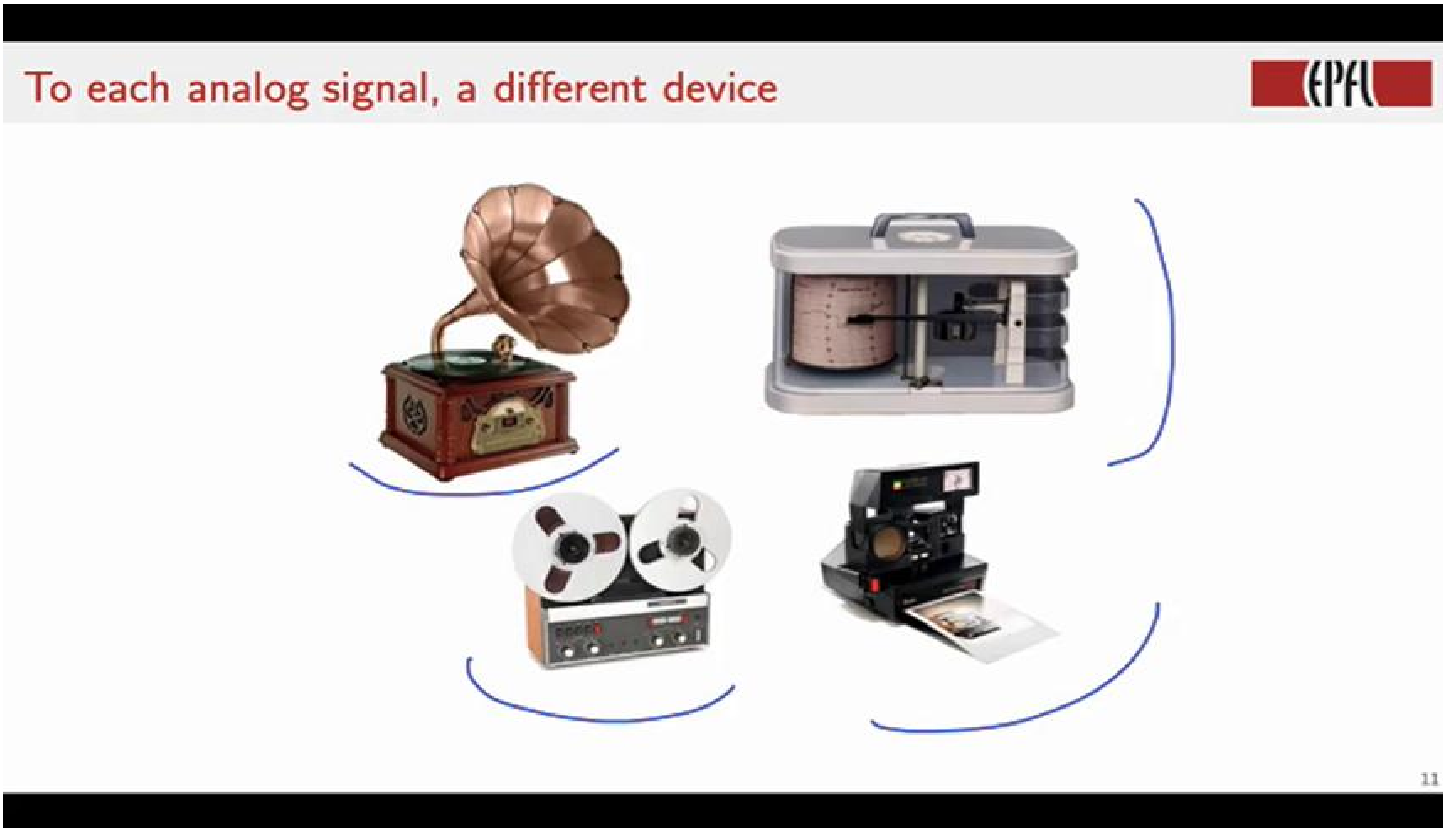


Now, if I show you a waveform like this, you probably have seen this before and you can guess that is some representation of a sound or speech. And if it were to apply this mathematical model to this information, the question is, what is the function that describes the sound? Well, there's no really an easy answer for that. What I can try to do is record this information and people have invented extremely sophisticated devices to do so.

Notes

Summary

3m 35s

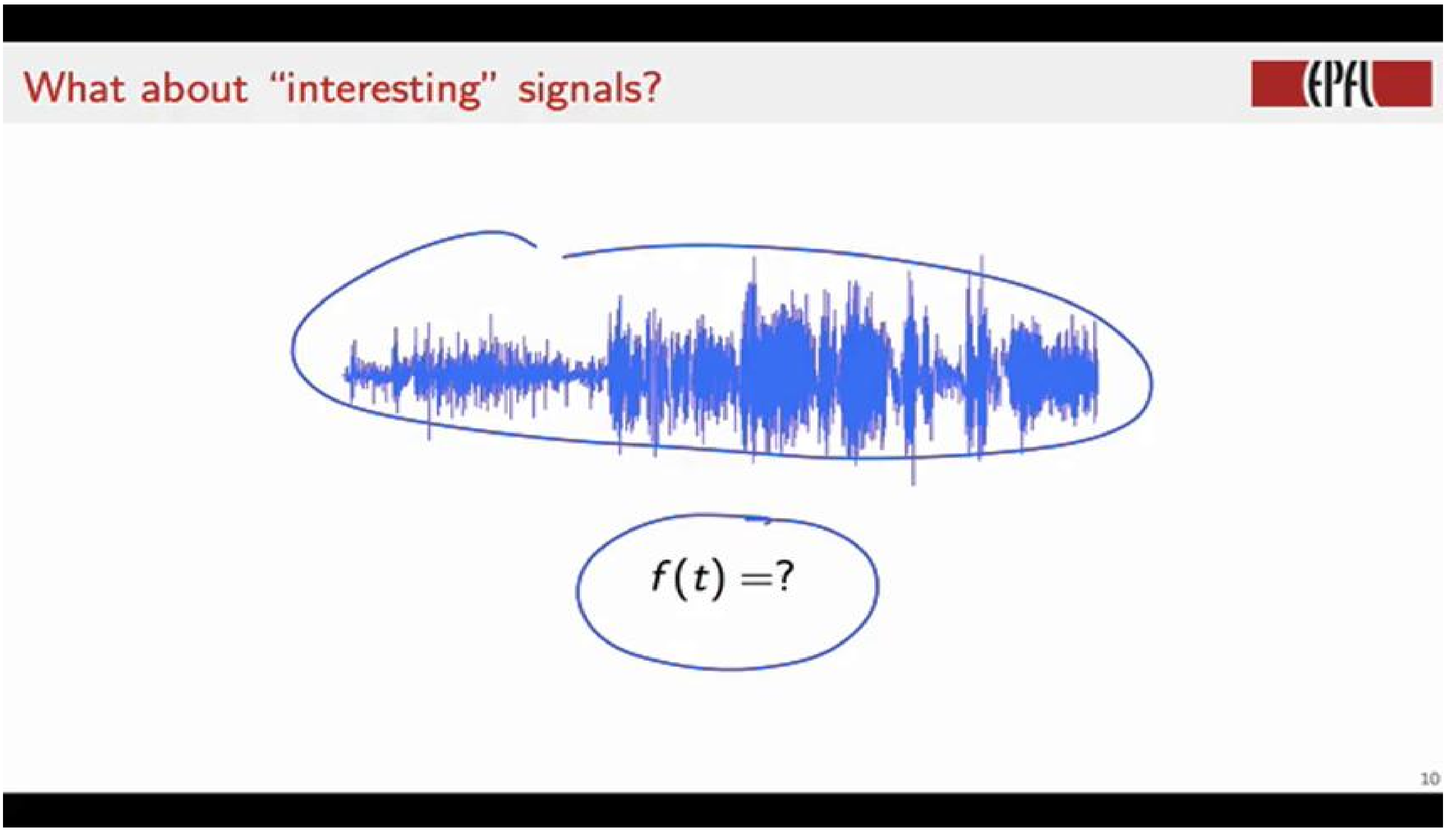


For sound, for instance, I could come up with a racket player I could come up with a tape recorder. And then if I want to measure a temperature signal, then I would have to come up with a mechanical system that drags a pencil on a piece of paper to record the evolution of temperature and to capture photographs, I will have to invent a camera. But you see, every device is specific to a certain signal.

Notes

Summary

4m 02s

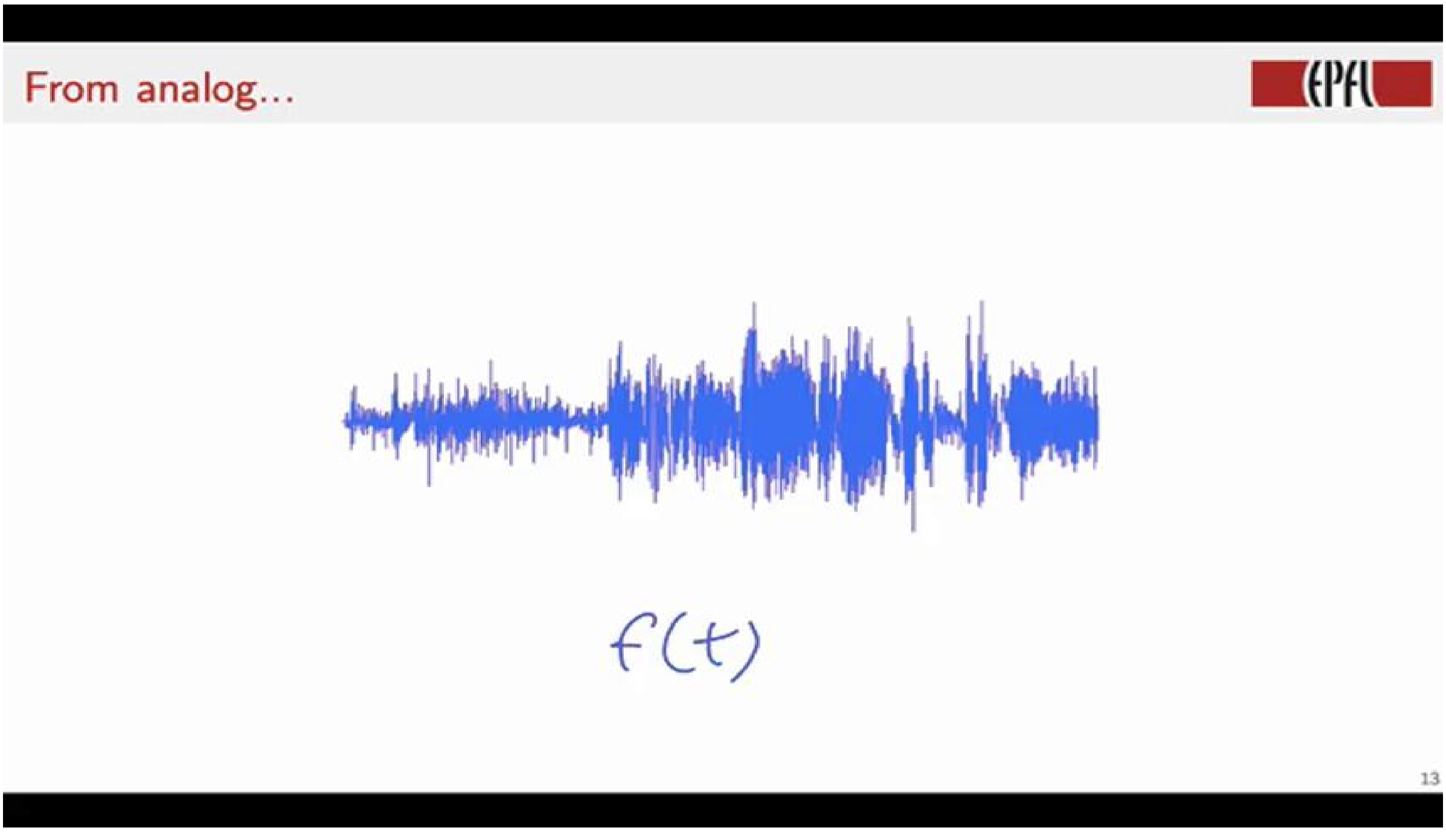


It will record information, but it will not let me manipulate this information easily and in a generic way in other words, the recording device will give me something like a picture or something like this, but it will not answer the question, what is the function that describes the phenomenon? This is the problem with analog signals.

Notes

Summary

4m 26s



The big paradigm shift and the power inherent to digital signal processing is that we're moving away from an analog model for signals like this. So we're not asking what is the function that represents a signal anymore, we just move into a recording of the values of this function and represent the phenomenon just as a series of numbers.

Notes

Summary

4m 45s

... to digital

... 74 31 -66 9 -123 33 159 -26 102 148 86
-136 -179 70 72 -84 -113 -42 -88 88 8 -180 -7
-133 8 164 -4 108 35 -82 74 -49 52 32 -31 ...

14

And in particular for digital signals, these numbers are integers.

Notes

Summary

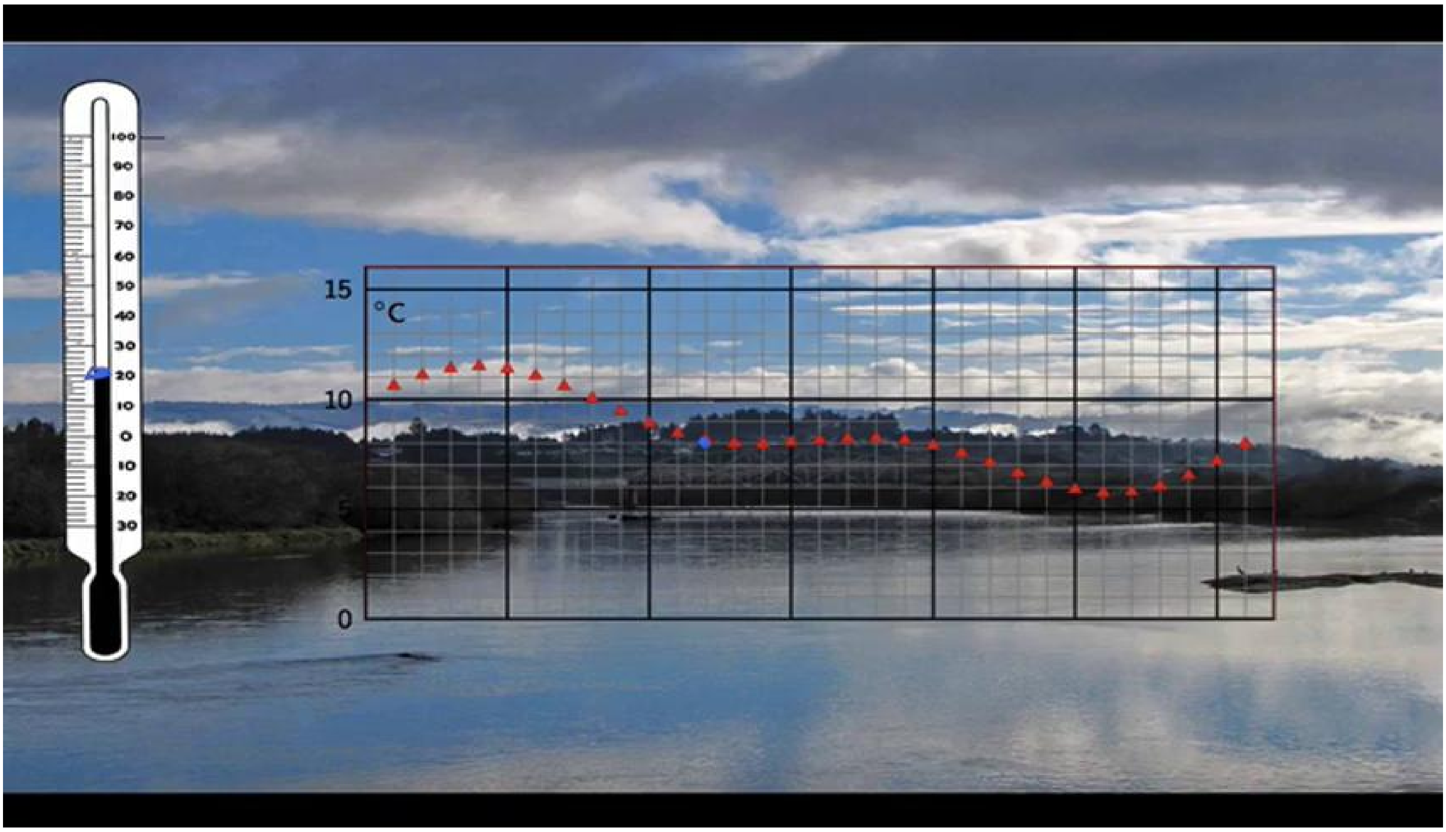
5m 05s

The so-called digital paradigm is composed of two fundamental ingredients. Discrete time and discrete amplitude. And let's start with discrete time because this is really the paradigm shift in the way we perceive the world.

Notes

Summary

5m 09s



When you were in elementary school, you were probably asked to run a little experiment. You record the temperature every morning and then plotted it on a piece of graph paper, and you obtained a plot that probably looked like this. So you look at a little point on the mercury scale and plot it in the right position on the graph paper. And when you do this for a number of days, the results in representation it seems pretty reasonable it makes sense to measure the temperature once or at most twice a day is a pretty descriptive representation of the weather. But can we really do that? Can we really splice up time into a series of discrete it instance and not lose information?

Notes

Summary

5m 22s

Discretizing time

What is time?

16

The question is really tied to the very nature of time, and this is something that philosophers have been grappling with since the beginning of intellectual speculation.

Notes

Summary

6m 03s

What is time?

What, then, is time? If no one asks of me, I know; if I wish to explain to him who asks, I know not.

Augustine of Hippo, Confessions, ca. 400CE

17

One of the most famous analysis of the problem of time was conducted by Augustine of Hippo in the fifth century. And his analysis led to fundamentally a negation of the existence of time because the past has already happened, the future is unknown, and the present instant cannot be pinpointed the moment I mention it, it's already become part of the past. And so, Augustine said time does not exist.

Notes

Summary

6m 14s

What is time?

[...] space and time are mere thought entities and creatures of the imagination [...] They precede the existence of objects of the senses a priori.

Immanuel Kant, 1787

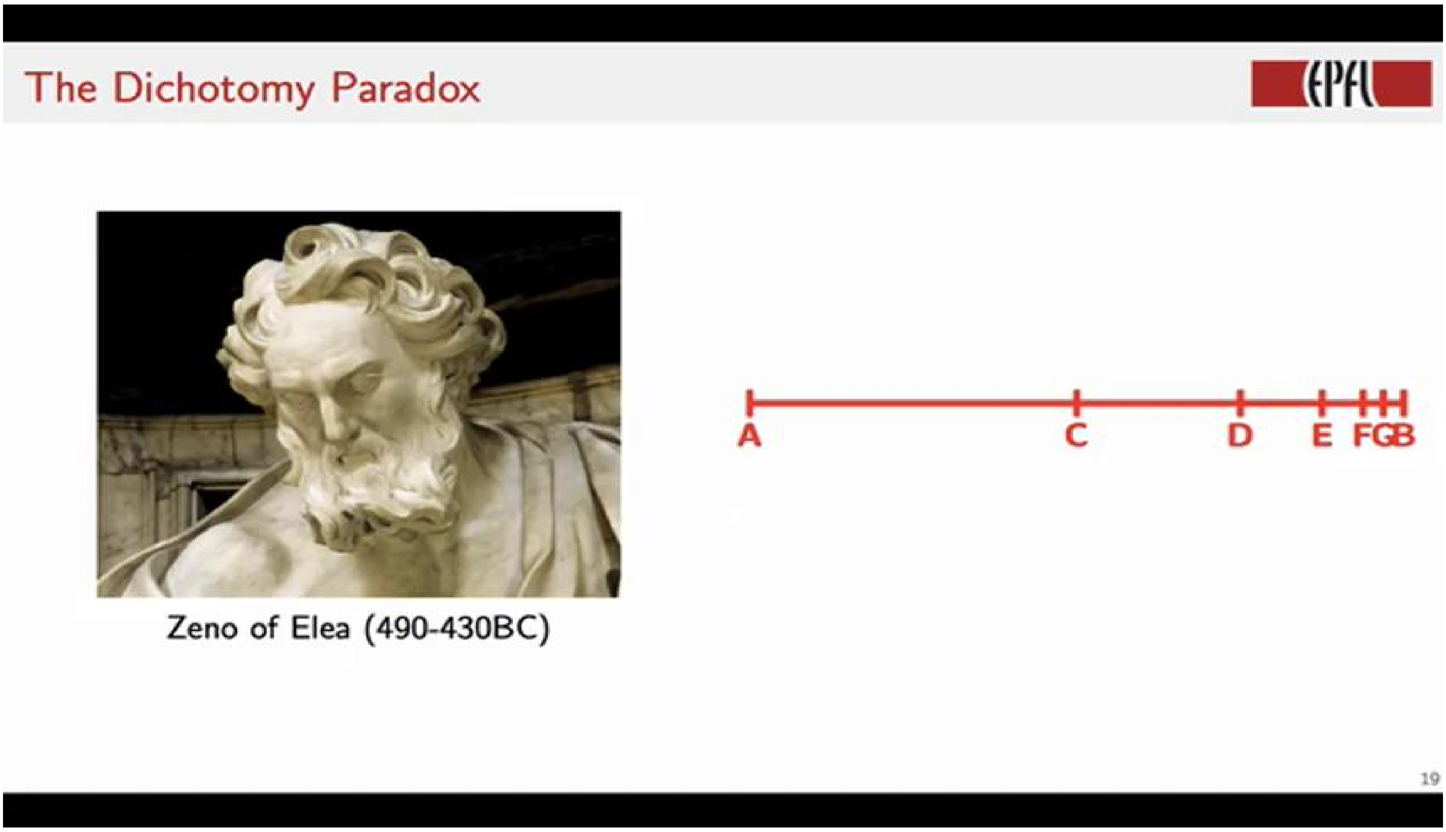
18

Centuries later, one of the last philosophers to try and systematise the world in philosophical terms, Emmanuel Kant solved the problem of time by saying that time and space are fundamental categories of the spirit fundamentally, we are hardwired to project a notion of time and space on everything that we observe.

Notes

Summary

6m 41s

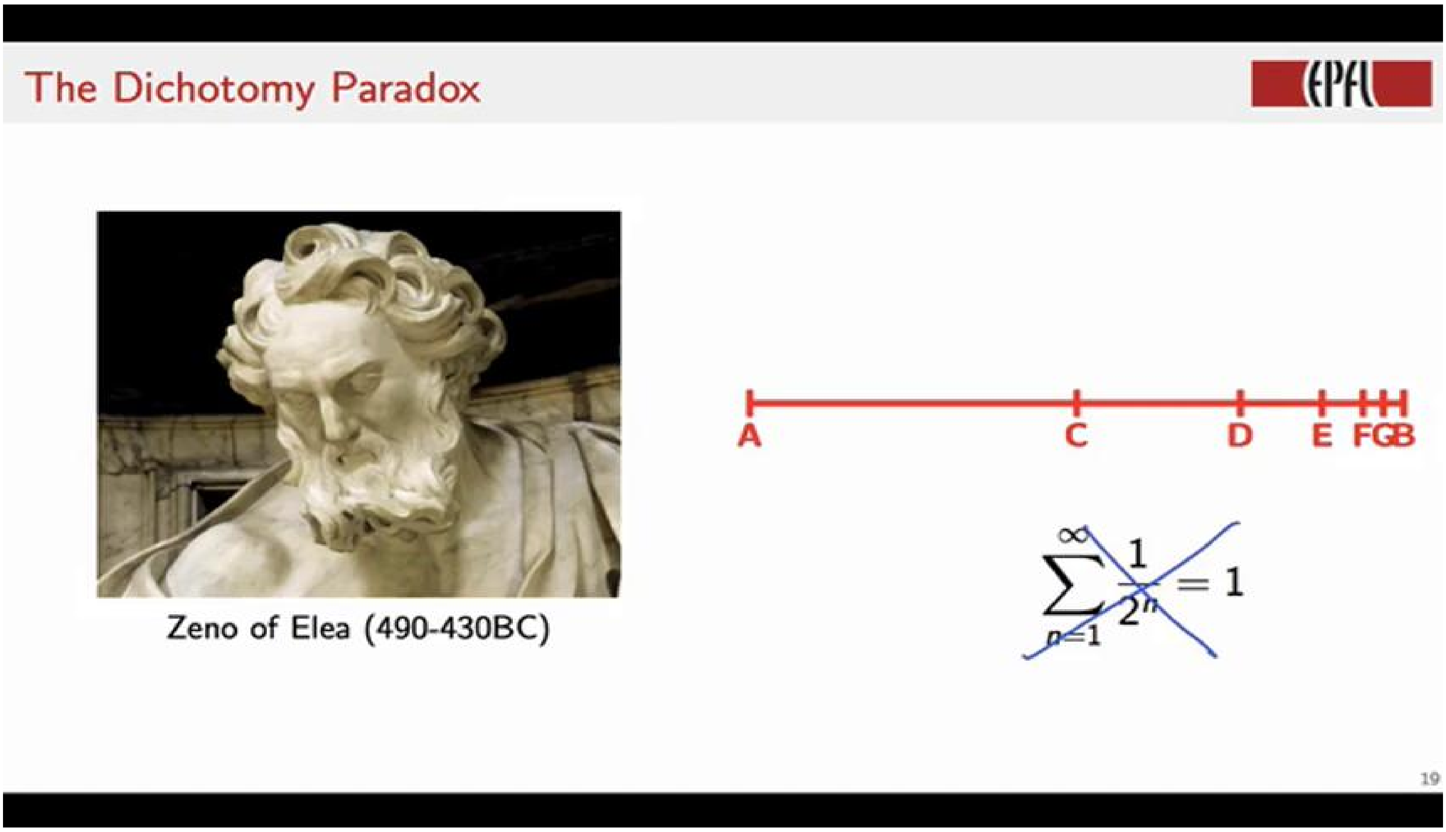


But perhaps the best known philosophical reflection on the concept of time was performed by Zeno of Elea in fifth century BC. Zeno came up with the famous paradoxes and the dichotomy paradox is one of the most celebrated. It states that, motion cannot really take place and the reasoning goes like so, if you want to move from point A to point B so cover some distance, well, in order to reach point B, you will have to go through A midpoint, let's call this midpoint C. And even if you go through that midpoint, then you will have another midpoint between C and B to go through, and then another midpoint and then another so on and so forth. Now Zeno said, no matter how small this subsegments that you can divide A and B into, you will need a finite amount of time across each one of them. So now you have an infinite number of segments and the sum of an infinite number of times however small, cannot be finite and therefore you will never reach point B. Of course, today, that any student with a little bit of calculus will tell you, "Oh, but no, because if you formalize this, you can see that you're actually taking an infinite sum of a geometric series with ratio one half, and we know that this sum is equal to one.

Notes

Summary

7m 03s



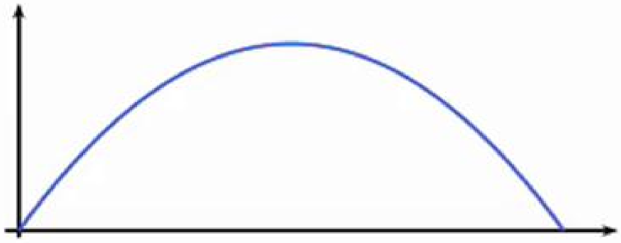
Well, this is not the solution to Zeno's paradox. The solution to Zeno's Paradox is represented by the last four centuries of mathematical research that have produced a mathematical model where infinite sums do not lead to contradictions. But Zeno's intuition was very profound. If we apply a simplistic mathematical model to reality, for instance, the possibility of dividing a segment into an infinite number of sub segments, then we run into contradiction. So we have to be careful to avoid these pitfalls.

Notes

Summary

8m 26s

Two competing models for reality



ideal trajectory

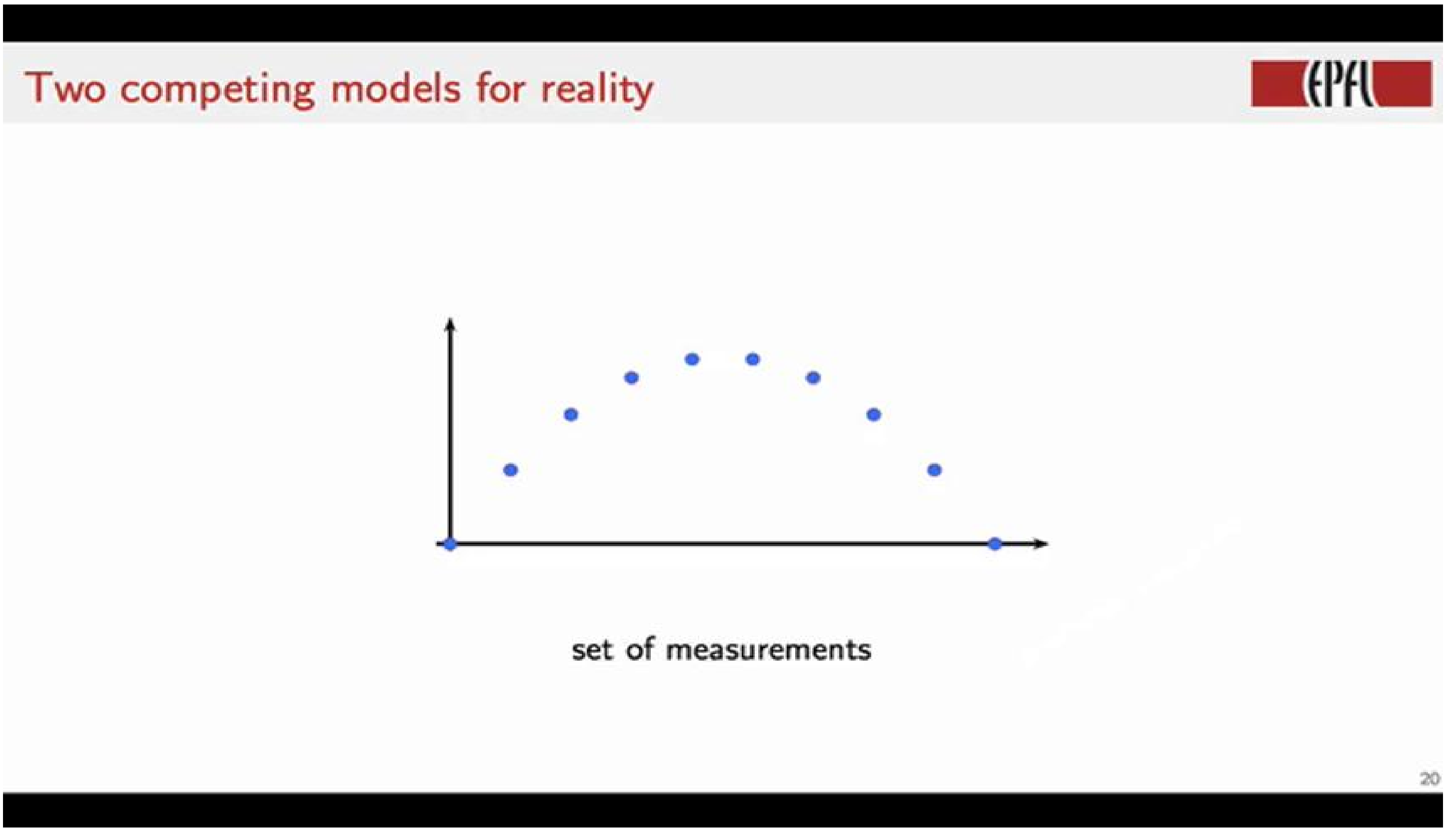
20

Back to our signal processing problem. If we start with a model of reality where we have for the projectile motion an ideal trajectory that can be modeled as a function of a real variable T, we are in a situation similar to the infinite divisibility of the segment we have a real number time that can be split in an infinite number of smaller intervals.

Notes

Summary

9m 00s



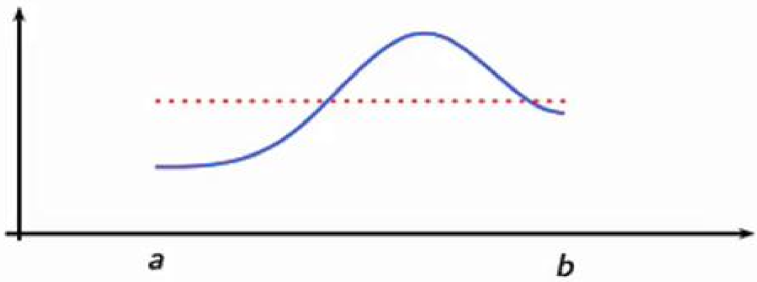
So how do we go from this to a model in which we have a finite number of, we can now call them with our name, samples of the original trajectory, a set of measurements which is finite in nature. Incidentally, having a discrete model of reality would be extremely practical for a lot of applications.

Notes

Summary

9m 25s

Discrete time is very practical! Computing the average...



$$\bar{x} = \underbrace{\frac{1}{b-a}}\,\underbrace{\int_a^b \underbrace{f(t)}dt}$$ (we need calculus!)

22

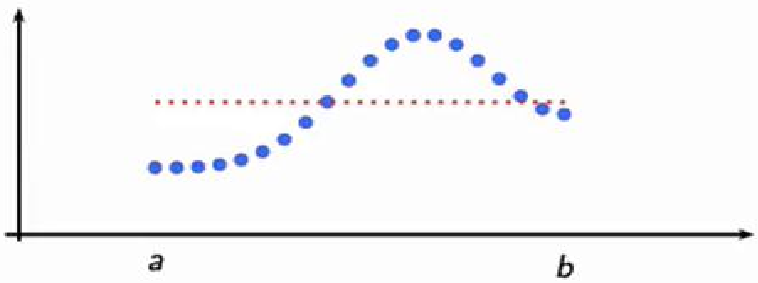
So, for instance, consider this temperature signal between A and B. If we use a continuous model of reality, an analog model and we want to compute the average temperature, then well we need to know the function F of t that represents the temperature and then we need calculus to compute the integral of this function over the interval interest and then divide by the length of the interval. Conversely, if we only have temperature measurements on a discrete set of times like so, for instance, well, the average is very simple.

Notes

Summary

9m 45s

Discrete time is very practical! Computing the average...



$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \text{ (here we don't!)}$$

22

We just sum these values and divide by the number of samples. This we know how to do intuitively. And this is valid in general, discrete models are extremely easy to use computationally speaking.

Notes

Summary

10m 17s

The discrete-time model

$$\cancel{f : \mathbb{R} \to \mathbb{V}}$$

$$x : \mathbb{Z} \to \mathbb{V}$$

23

So we want to move away from this platonic ideal functions of a real variable time and adopt a model of a reality where signals are described as sequences mathematically these are mappings from the set of integers to a set of values V, which could be the set of real numbers or as in the case of fully digital signals the set of integers as well.

Notes

Summary

10m 28s

The discrete-time model

$$x[n] = \ldots, 1.2390, -0.7372, 0.8987, 0.1798, -1.1501, -0.2642\ldots$$

24

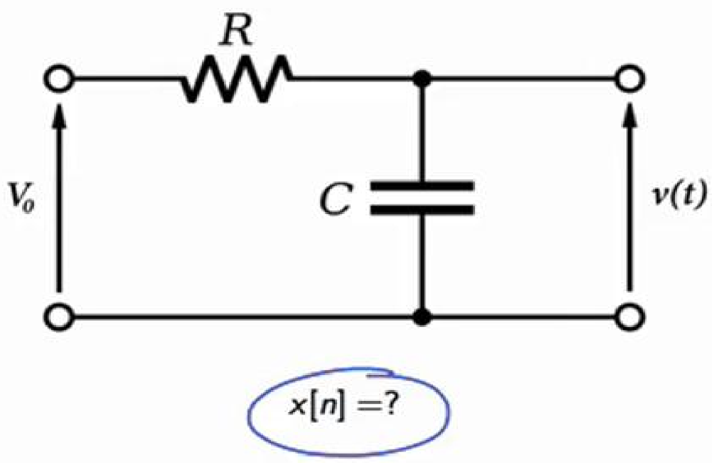
The notation is very simple, we will indicate a discrete time signal as X of N where X is a signal and N is the time, but we will see pretty soon that N does not have a physical dimension is just an ordinal number that orders the samples one after the other.

Notes

Summary

10m 54s

Are we burning bridges?



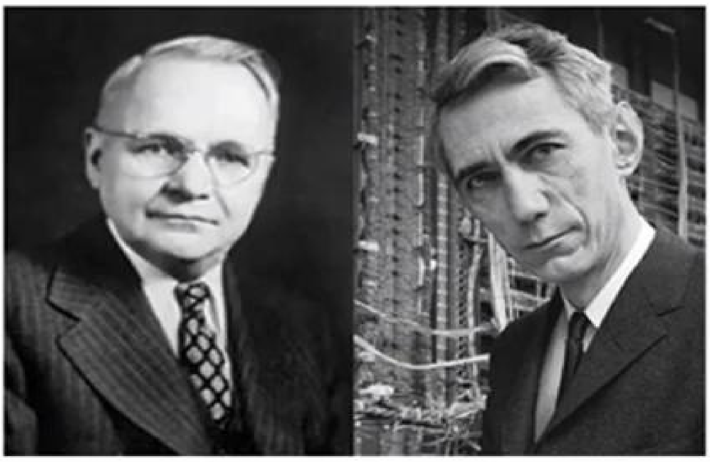
Are we burning bridges? Are we losing some power of representation when we move to this discrete model of reality. So what happens inside an RC circuit? Can we still describe this physical reality using a discrete time sequence?

Notes

Summary

11m 14s

Bridging the models: the founding fathers



Harry Nyquist and Claude Shannon

26

The question is of fundamental importance and the answer was given to us at the beginning of last century by Harry Nyquist and Claude Shannon. The answer is positive and it states that under very mild conditions, the continuous time representation and the discrete time representations are equivalent.

Notes

Summary

11m 28s

The Sampling Theorem (1920)

Under appropriate "slowness" conditions for $x(t)$ we have:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \operatorname{sinc}\left(\frac{t - nT_s}{T_s}\right)$$

27

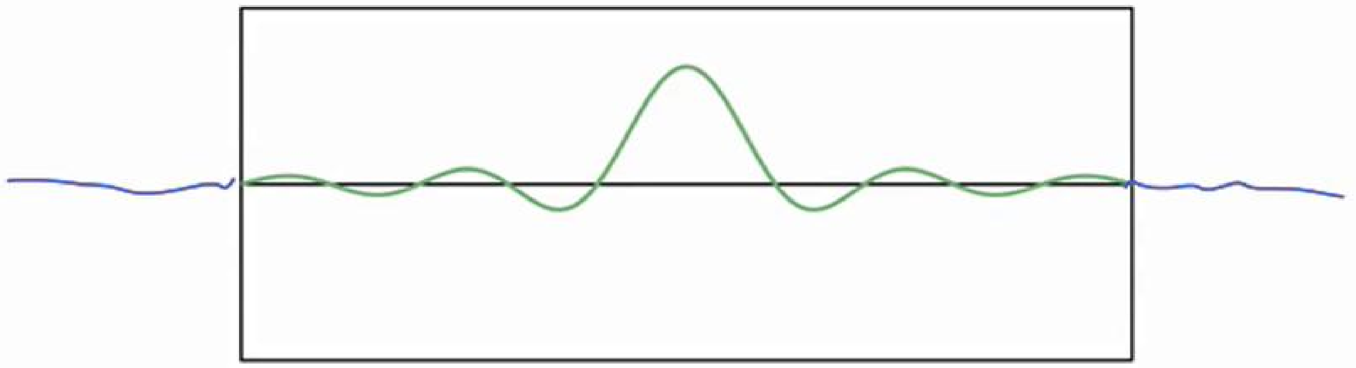
Mathematically, the result is known as a sampling theorem, and it has a very simple statement. The relationship between the continuous time representation of a signal and its discrete time counterpart is given by this formula, and you can see that we can build the continuous time representation as a linear combination of copies of a prototypical function or building block called the sync shifted and scaled by the values of the discrete time sequence.

Notes

Summary

11m 44s

Sinc: he building block



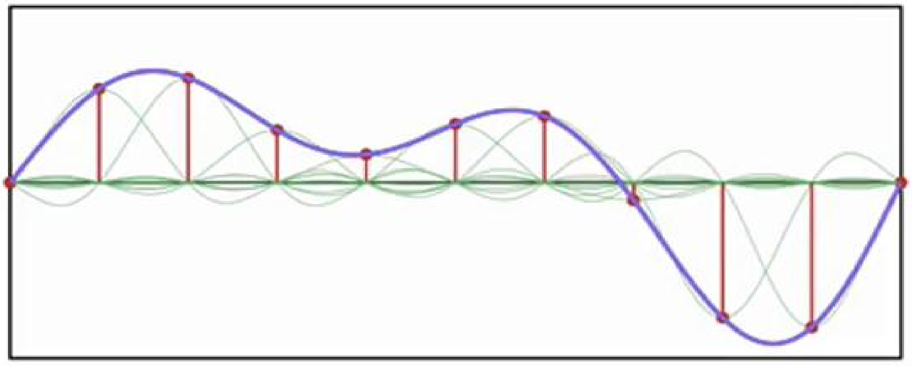
The sync looks like so you probably have seen it in countless signal processing logos, it's actually an infinite support function that keeps oscillating from minus infinity to plus Infinity, and the sampling theorem graphically looks like so.

Notes

Summary

12m 14s

From continuous to discrete time (and back)



29

You start with a continuous time signal and then you take measurements, so you convert this to a discrete time sequence. You take regular measurements of this function and then you throw away all the rest and you're just left with the discrete time sequence. To go back to continuous time, all you need to do is take copies of the sync function and place them at each sample location scaled by the amplitude of the sample. And when you do that, and then you sum all these copies of the sync together, you get back exactly the original function.

Notes

Summary

12m 29s

When can we do all this? Joseph Fourier will tell us.

30

The conditions under which you can do this are given by the statement of the sampling theorem and will require a tool called Fourier Analysis. The Fourier transform will give us a quantitative measure of how fast the signal moves. And once we know the speed, we will always be able to choose a sampling interval namely a space between measurements when we convert a function to a sequence that will satisfy the hypothesis of the sampling theorem.

Notes

Summary

13m 01s

Digital Signals

Key ingredients:

- discrete time
- **discrete amplitude**

31

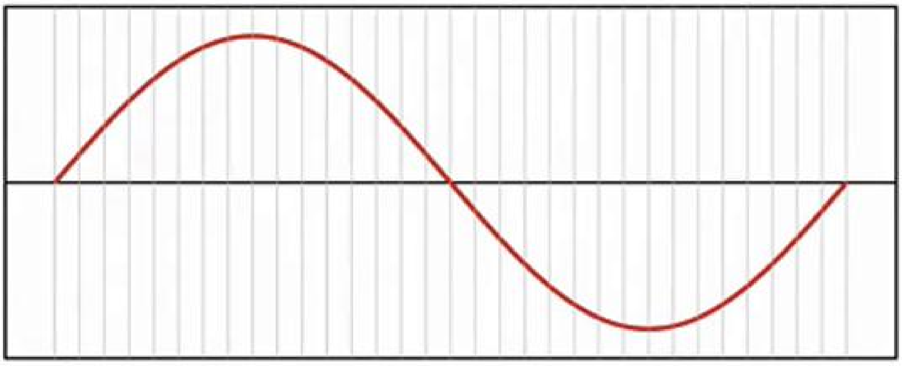
Okay, so we have said the digital signals are composed of two ingredients and we have talked a length about the discretization of time. Now let's look at the other aspect of discretization of amplitude.

Notes

Summary

13m 31s

Digital Signals



$x(t)$

33

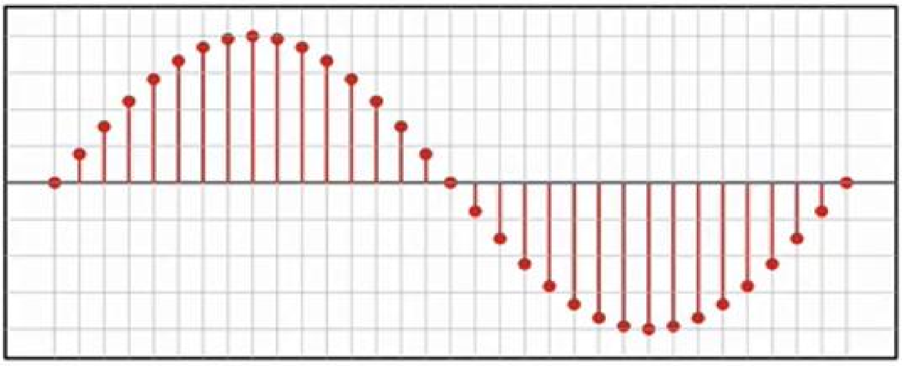
Take, for instance, the sign away. The first discretization happens in time, and we get a discrete set of samples.

Notes

Summary

13m 44s

Digital Signals



33

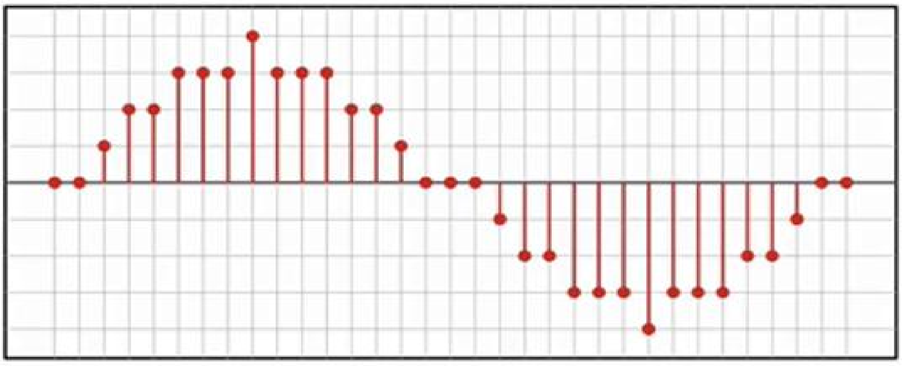
And then the second discretization happens in amplitude, where each sample can take values now only amongst a predetermined set of possible levels.

Notes

Summary

13m 51s

Digital Signals



33

And the very important consequence of this discretization is that, independently on the number of levels the set of levels is countable. So we can always map the level of a sample onto an integer. Now, if our data is just a set of integers, it means that its representation is completely abstract and completely general.

Notes

Summary

14m 00s

Digital amplitude

Why it is important:

- storage
- processing
- transmission

34

And this has very important consequences in three domains. Storage becomes very easy because any memory support that can store integers can store signals now. And computer memory comes to mind as a first candidate for storage medium. Processing now becomes completely independent on the nature of the signal, because all we need is a processor that can deal with integers. And again, CPUs, the heart of computers, are general purpose processors that can deal with integers very well. And finally, transmission with digital signals, we can deploy extremely effective ways to combat noise and maximize the capacity of a communication channel and we will see an example in a second.

Notes

Summary

14m 21s

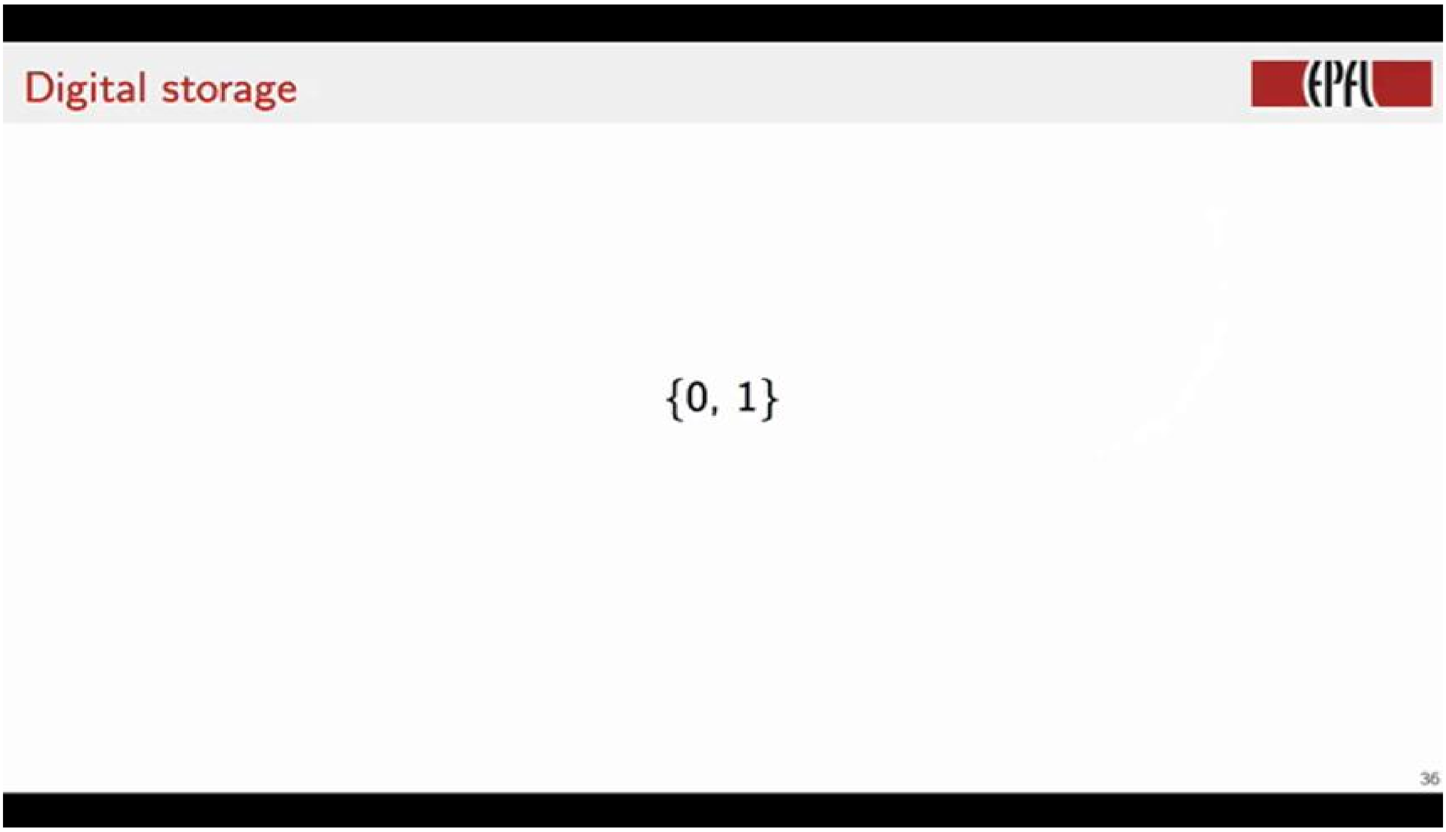
Analog storage

As far as storage is concerned, just consider the difference between attempting to store an analog signal, which required a medium, that was dependent on each signal and on each application and compare that to what we do today, which is using general purpose computer memory for all kinds of signals. So here in this picture, for instance, you can see what we used to do for sound, we started with phonograph, perhaps wax cylinders and records and then moved on to tape records and each medium was incompatible with the previous one. Same goes with temperature, no way of transmitting the temperature other than physically delivering of cylinder of paper to weather station and so on, so forth.

Notes

Summary

15m 07s



Today, all we do is store zeros and ones, and this is a consequence of the fact that the possible values of a discrete time sequence are mapped onto accountable set of integers and therefore onto binary digits.

Notes

Summary

15m 52s

Storage

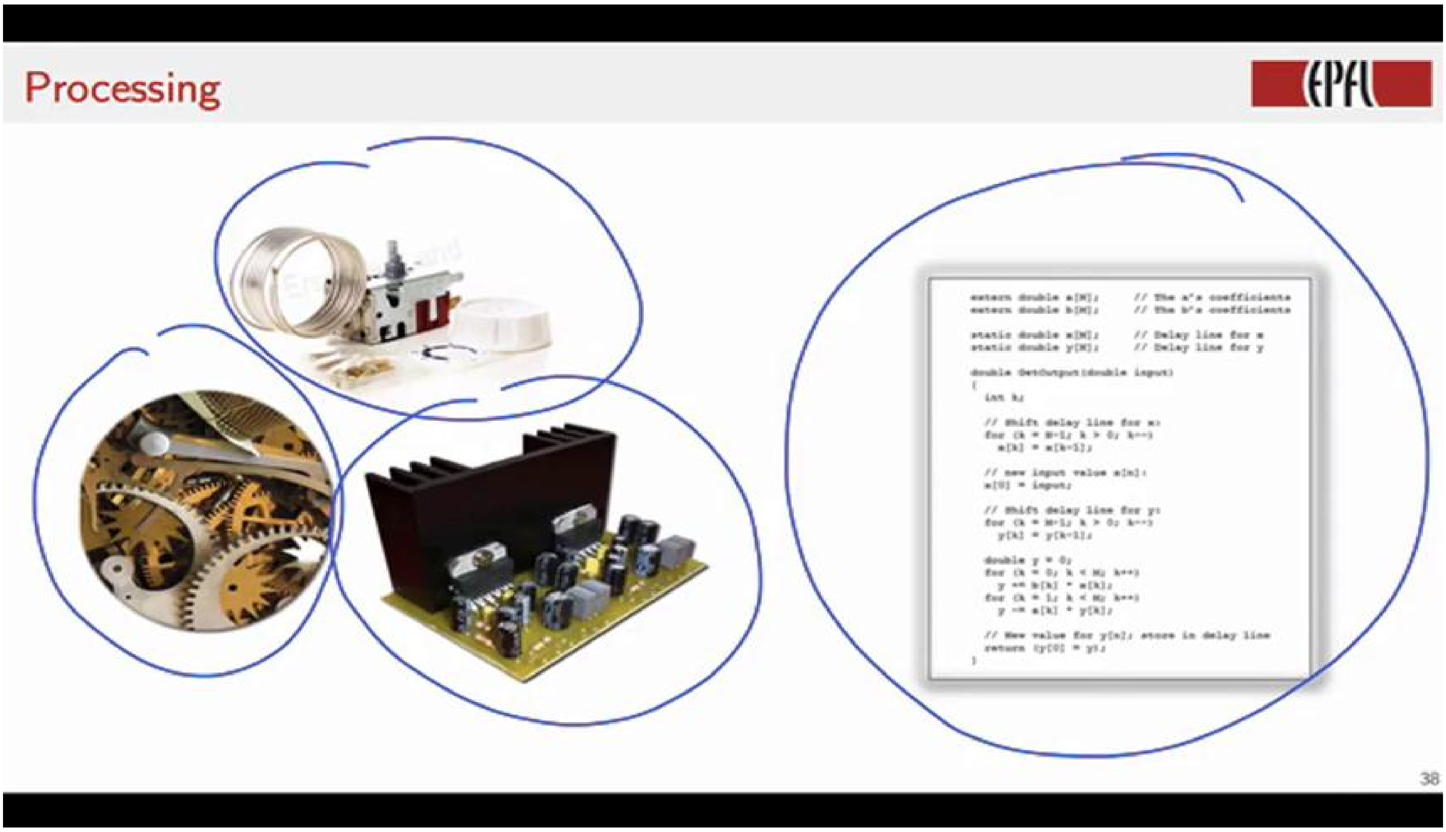
37

There is a famous picture going around the Internet, that shows how much information you can store today on a micro SD card compared to the storage abilities of just a few years back. And this fast evolution of technology is what you get when you can pull resources. You don't have to specialize into different devices for different purposes anymore and all the research goes into perfecting a single technology that is shared across a variety of applications.

Notes

Summary

16m 06s



Processing was also a problem in the past with analog signals you had to devise specific devices that would be able to react to the physical phenomena that had to interact with. So, for instance, here you see a thermostat that had special coil that expanded and triggers switches in order to regulate the temperature. Mechanical systems require the design of very complex gears, and sound equalization system required discrete electronics. Today with digital signals, all you need to do is write a piece of computer code that will run on a general purpose architecture and perform the same tasks that required hardware in the past.

Notes

Summary

16m 34s

Data transmission

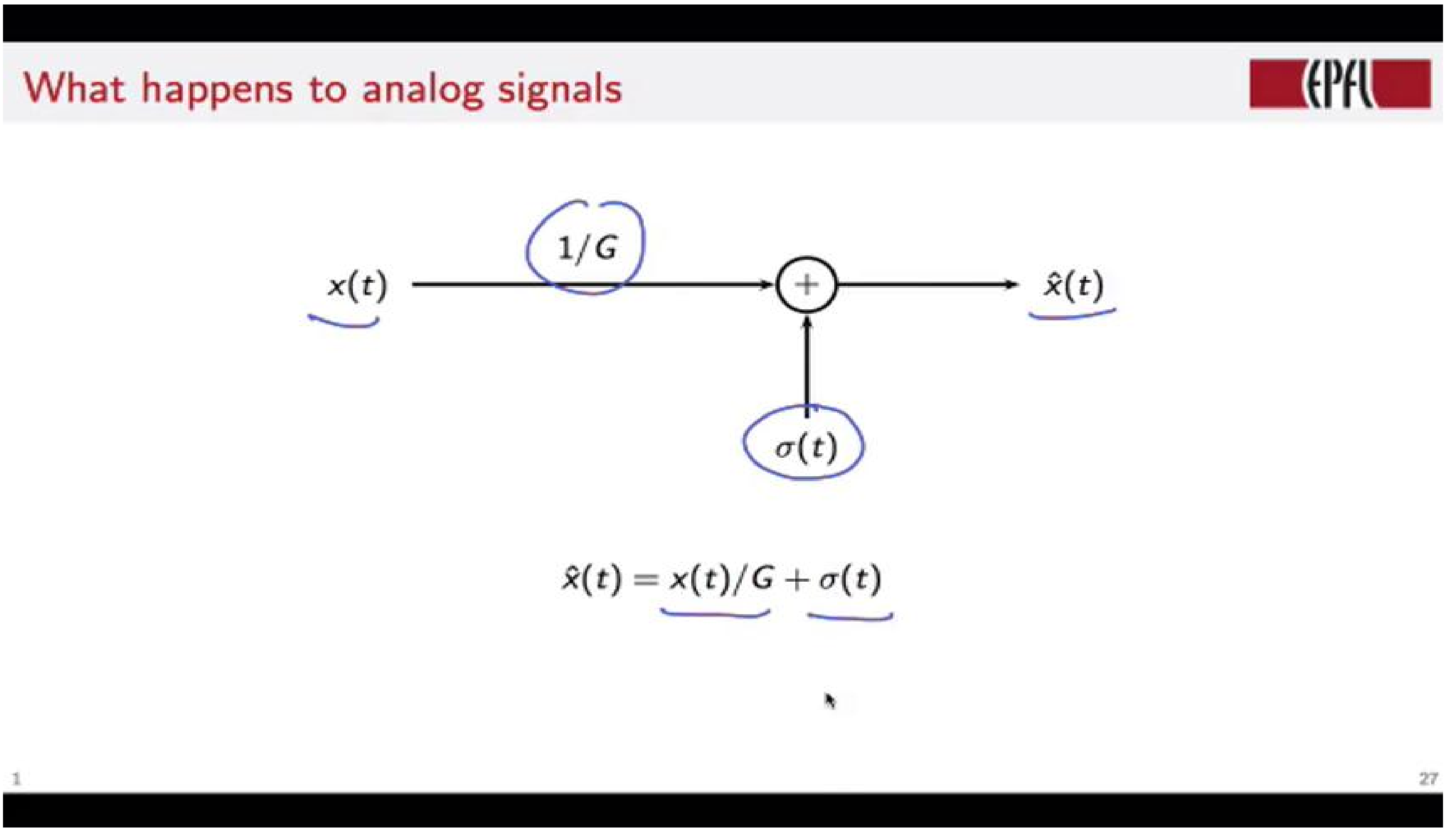
1 26

Finally, let's consider the problem of data transmission, which is probably the domain where digital signal processing has made the most difference in our day to day life. So if you have a communication channel and you try to send information from a transmitter to a receiver, you are faced with a fundamental problem of noise. So let's see what happens inside the channel.

Notes

Summary

17m 13s

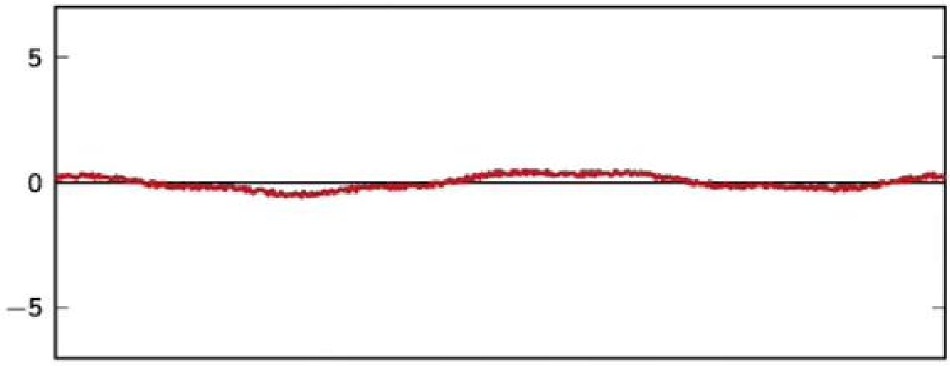


You have a signal. So that will be put into the channel. Will introduce an attenuation. It will lower the volume of the signal, so to speak, but it will also introduce some noise indicated here as sigma of T. And what you will receive at the end is an attenuated copy of your original signal plus noise. These are just facts of nature that you cannot escape.

Notes

Summary

17m 34s



$x(t)/G + \sigma(t)$

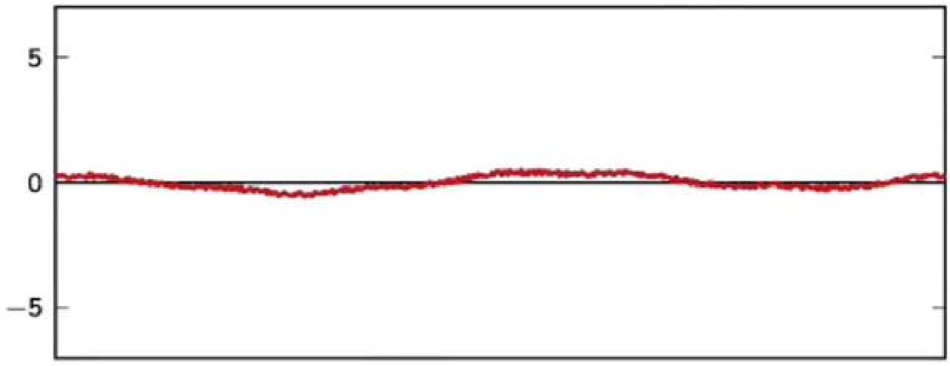
So, if this is your original signal, what you will get at the end is an attenuated copy scaled by a factor of G, plus noise. So how do you recover the original information? Well, you try to undo the effects introduced by the channel, but the only thing you can undo is the attenuation. So you can try and multiply the received signal by a gain factor that is the reciprocal of the attenuation introduced by the channel.

Notes

Summary

17m 55s

Transmission of analog signals



$x(t)/G + \sigma(t)$

1 30

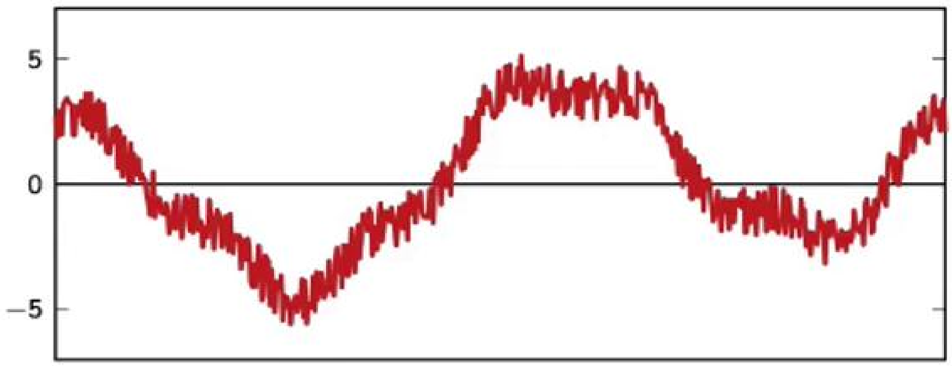
So if you do that, you introduced again here at the receiver, and what you get is let's start again with the original signal, attenuated copy, some noise added and then let's undo the attenuation.

Notes

Summary

18m 25s

Transmission of analog signals



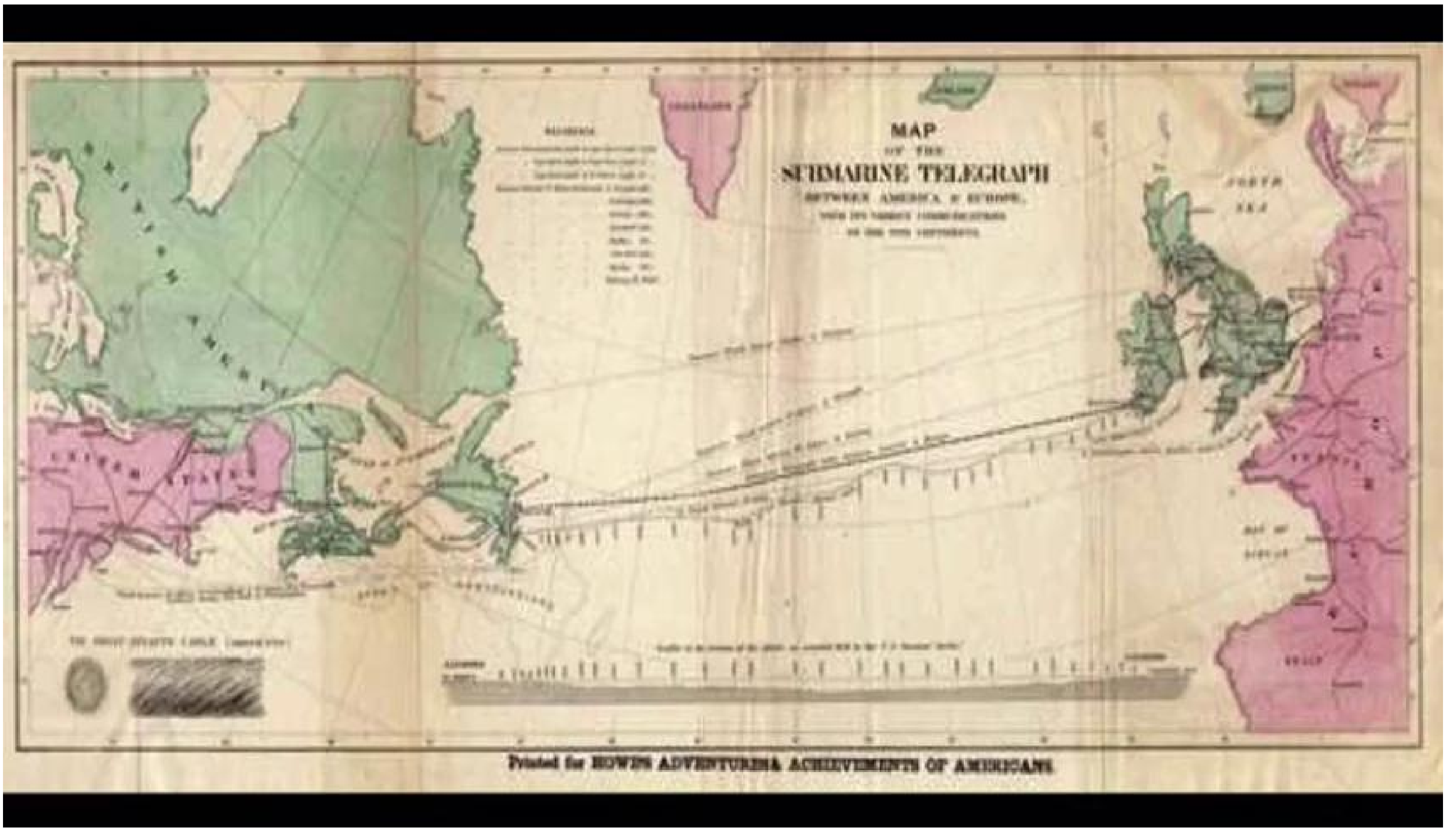
$$\hat{x}_1(t) = G[x(t)/G + \sigma(t)] = x(t) + \underbrace{G\sigma(t)}$$

Well, what happens, and surprisingly, is that the gain factor has also amplified the noise that was introduced by the channel. So you get a copy of the signal that is, yes, of a comparable amplitude to the original signal, but in which the noise is much larger as well. This is a typical situation that you get in second generation or third generation copies of, say, a tape or if you try and do a photocopy of a photocopy, just to give you an idea of what happens with this noise amplification problem.

Notes

Summary

18m 40s



Now, why is this very important? This is important because if you have a very long cable, so for instance, if you have a cable that goes from Europe to the United States and you try to send a telephone conversation over there, what happens is that you have to split the channel into several chunks and try to undo the attenuation of a chunk in sequence.

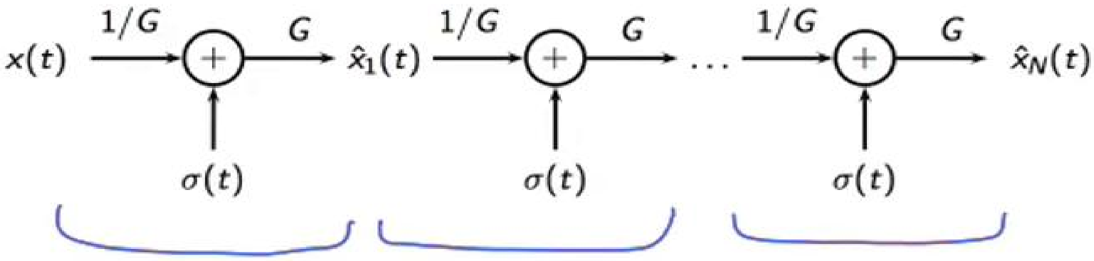
Notes

Summary

19m 11s

Transmitting a signal overseas

For a long, long channel we need repeaters



$$\hat{x}_N(t) = x(t) + NG\sigma(t)$$

1 33

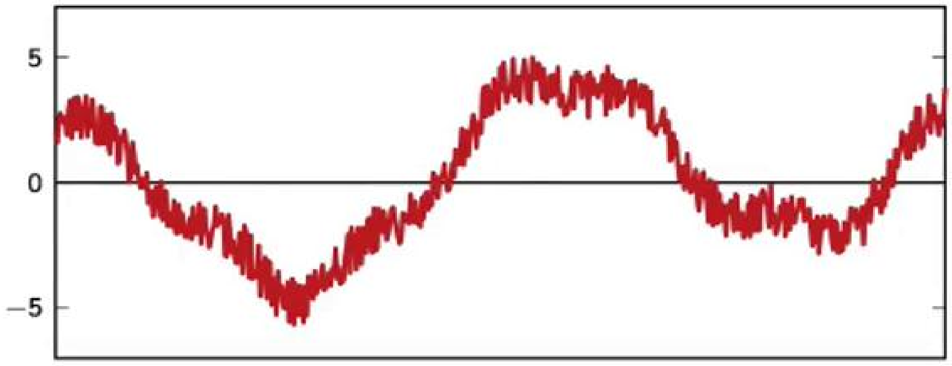
So you actually put what are called repeaters along the line that regenerate the signal to the original level every say, 10 kilometers of cable or so. But unfortunately, the cumulative effect of this chain of receiver is that some noise gets introduced at its stage and gets amplified over and over again.

Notes

Summary

19m 33s

Transmission of analog signals



$$\hat{x}_1(t) = G[x(t)/G + \sigma(t)] = x(t) + G\sigma(t)$$

1

34

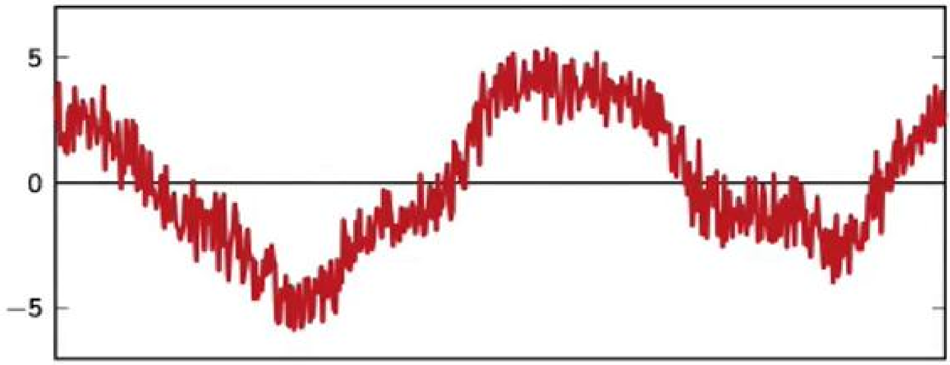
So, for instance, if this is our original signal, which again gets attenuated and gets corrupted by noise in the first segment of the cable after amplification, you would get this, which is seen that before.

Notes

Summary

19m 53s

Transmission of analog signals



$$\hat{x}_2(t) = G[\hat{x}_1(t)/G + \sigma(t)] = x(t) + 2G\sigma(t)$$

1 34

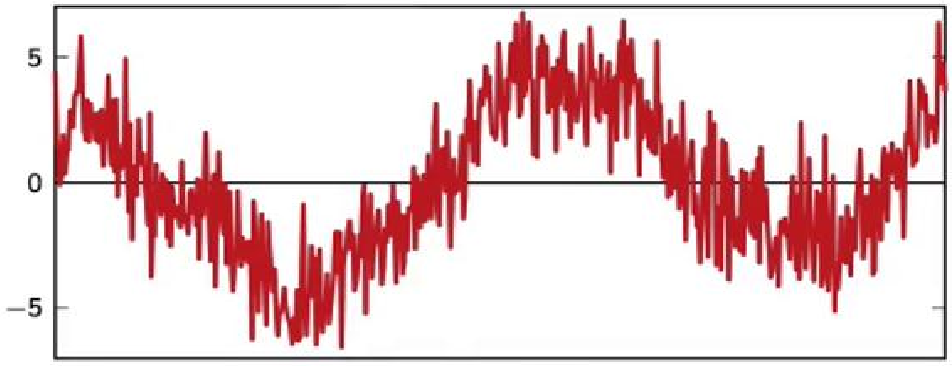
Then this signal is injected into the second section of the cable, it gets attenuated new noise gets added to it, and when you amplified it, you get double the amplified noise.

Notes

Summary

20m 07s

Transmission of analog signals



$$\hat{x}_N(t) = x(t) + NG\sigma(t)$$

1

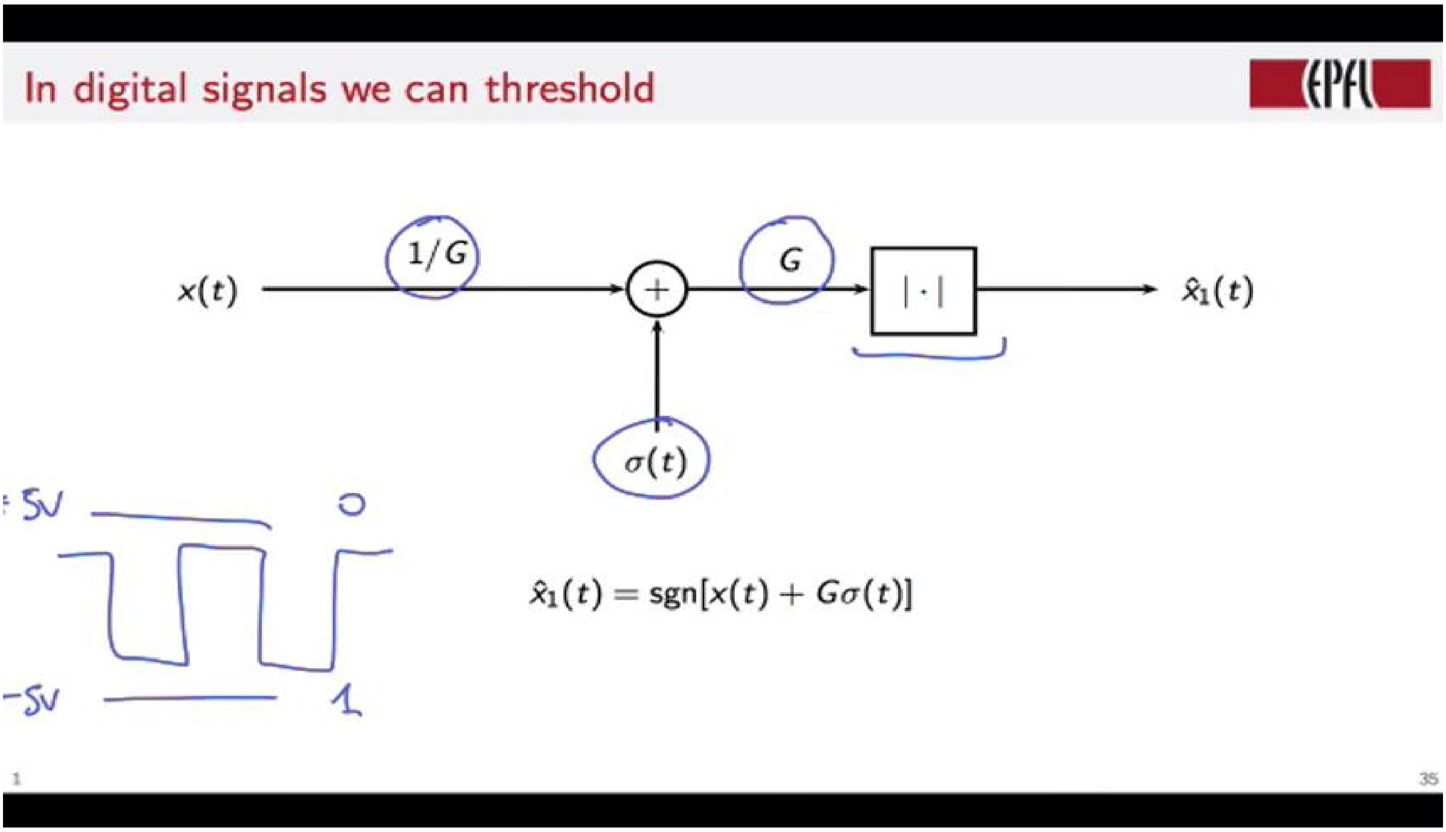
34

And after N sections of the cable, you have N times the amplified noise. This can lead very quickly to a complete loss of intelligibility, in a phone conversation.

Notes

Summary

20m 17s



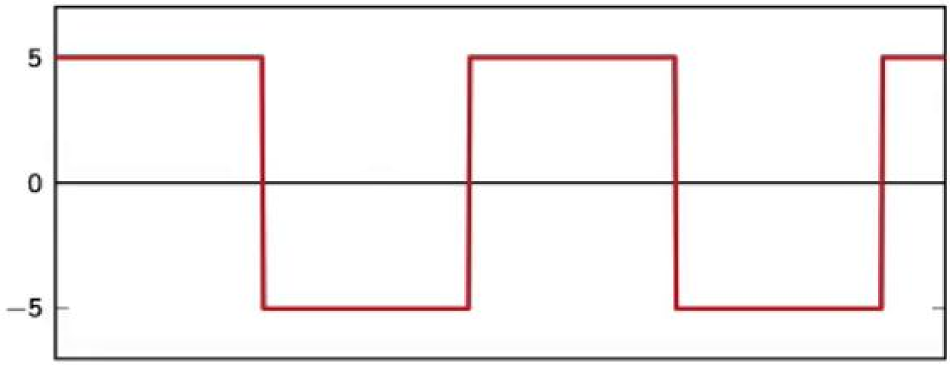
Let's now consider the problem of transmitting a digital signal over the same transotianic cable. Now, a digital signal, as we said before, is composed of samples whose values belong to a accountable finite set of levels. And so their values can be mapped to a set of integers. Now, transmitting a set of integers means that we can encode this integers in binary format and therefore we end up transmitting basically just a sequence of zeros and ones binary digits. We can build an analog signal associating, say, the level 5 volt to the digit zero and minus five volt to the digit one. And we will have a signal, and we will have a signal that will oscillate between these two levels as the digits are transmitted. What happens on the channel is the same as before, we will have an attenuation, we will have the addition of noise, and we will have an amplifier at each repeater that will try to undo the attenuation. But on top of it all, we will have what is called a threshold operator that will try to reconstitute the original signal as best as possible. Let's see how that works.

Notes

Summary

20m 28s

Transmission of quantized signals



$x(t)$

1

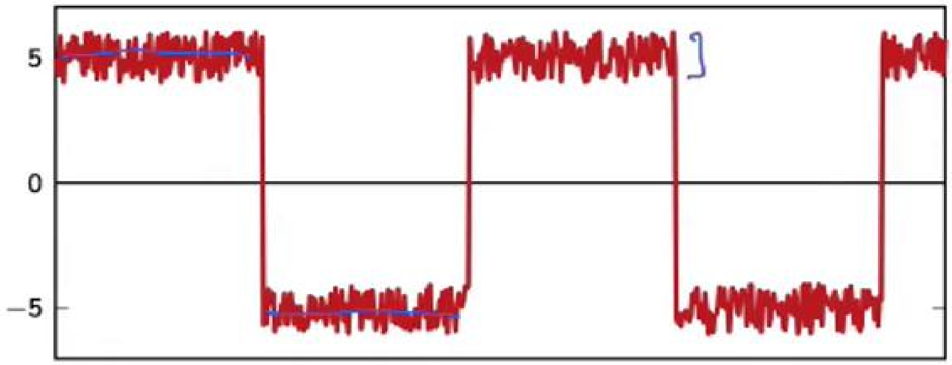
36

Notes

Summary

21m 42s

Transmission of quantized signals



$$G[x(t)/G + \sigma(t)] = x(t) + G\sigma(t)$$

1

36

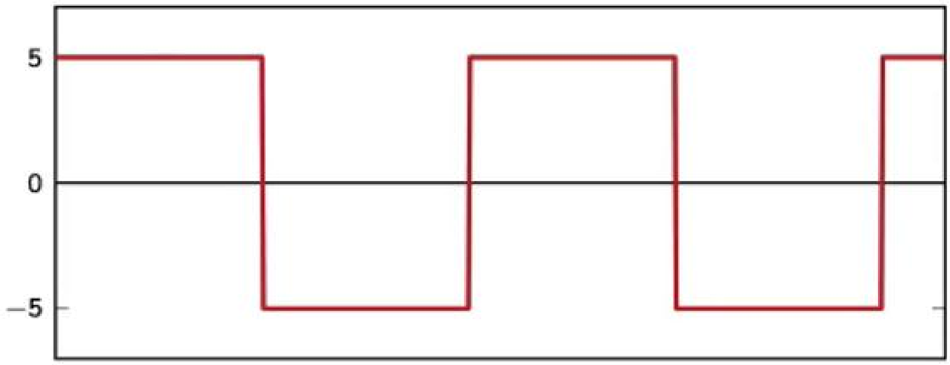
If this is what we transmit, say an alternation of zero and one mapped to this two voltage levels, the attenuation and the noise will reduce the signal to this state, the amplification will regenerate the levels and will amplify the noise. So the noise is much larger than before. But now we can just threshold and say, if the signal value is above zero, we just output five volts. And vice versa, if it's below zero, we will output minus five volts.

Notes

Summary

21m 43s

Transmission of quantized signals



$$\hat{x}_1(t) = G\text{sgn}[x(t) + G\sigma(t)]$$

1

36

So the thresholding operator will reconstruct a signal like so. You can see that at the end of the first repeater, we actually have an exact copy of the transmitted signal and not a noise corrupted copy.

Notes

Summary

22m 12s

Digital data throughputs

- Transatlantic cable:
  - 1866: 8 words per minute (≈5 bps)
  - 1956: AT&T, coax, 48 voice channels (≈3Mbps)
  - 2005: Alcatel Tera10, fiber, 8.4 Tbps (8.4×10^{12} bps)
  - 2012: fiber, 60 Tbps
- Voiceband modems
  - 1950s: Bell 202, 1200 bps
  - 1990s: V90, 56Kbps
  - 2008: ADSL2+, 24Mbps

1 37

The effectiveness of the digital transmission schemes can be appreciated by looking at the evolution of the throughput, the amount of information that can be put on a transatlantic cable. In 1866, the first cable was laid down and it had a capacity of eight words per minute, which corresponded to approximately five bits per second. In 1956 when the first digital cable was laid down on the ocean floor, the capacity all of a sudden skyrocketed to three megabits per second. So ten to the power of six orders of magnet is larger than the animal cable. And in 2005, when a fiber cable was laid down, another six orders of magnitude were added for a capacity of 8.4 terabits per second. Similarly, and literally closer to home, we can look at the evolution of the throughput for in-home data transmission. In the 50s, the first voiceband modems came out of Bad Labs voiceband, voiceband meaning that there were devices designed to operate over a standard telephone channel. Their capacity was very low, 1200 bits per second and there were analog devices. With a digital revolution in the 90s, digital modems started to appear, and very quickly reached basically the ultimate limit of data transmission over the voiceband channel, which was 56 kilobits per second at the end of the 90s.

Notes

Summary

22m 25s

Digital data throughputs

- Transatlantic cable:
  - 1866: 8 words per minute ($\approx$5 bps)
  - 1956: AT&T, coax, 48 voice channels ($\approx$3Mbps)
  - 2005: Alcatel Tera10, fiber, 8.4 Tbps (8.4×10^{12} bps)
  - 2012: fiber, 60 Tbps
- Voiceband modems
  - 1950s: Bell 202, 1200 bps
  - 1990s: V90, 56Kbps
  - 2008: ADSL2+, 24Mbps

1 37

The transition to ADSL pushed that limit up to over 24 megabits per second in 2008. Now, this evolution is, of course, partly due to improvements in electronics and to better phone lines. But fundamentally, its success and its affordability are due to the use of digital signal processing. We can use small yet very powerful and cheap general-purpose processors to bring the power of error correcting codes and data recovery, even in small home consumer devices. In the next few weeks, we will study signal processing starting from the ground up. And by the end of the class, we will have enough tricks in our bag to fully understand how an ADSL modem works.

Notes

Summary

23m 52s

Digital Signal Processing: key ideas

- discretization of time:
  - samples replace idealized models
  - simple math replaces calculus
- discretization of values:
  - general-purpose storage
  - general-purpose processing (CPU)
  - noise can be controlled

Let the digital revolution begin!

51

For now, let's resume the key ideas that we have seen in this, perhaps slightly rambling introduction. The discretization of time allows us to replace the idealized models of physics and electronics with sequences of measurements of samples. And in so doing, we will be able to replace calculus with much simpler math in our processing. The discretization of values on the other hand, is of extreme practical importance because it will allow us to use general purpose storage for our signals and general-purpose processing units for processing. Discrete values also allow us to control the noise very effectively and build very efficient communication systems. So, let's go on with the digital revolution.

Notes

Summary

24m 39s