WAVE THEORY OF INFORMATION

This report is the continuation of the report submitted by ARUN DIWAN

LINK : https://github.com/WSU-Data-Science/report-on-assigned-topic-1-dasingh92/blob/main/wtoi_report_arun_19884240.ipynb (https://github.com/WSU-Data-Science/report-on-assigned-topic-1-dasingh92/blob/main/wtoi_report_arun_19884240.ipynb)

The core idea of the topic is to focus on information theory from the point of view of Wave Theory. As information theory is well defined for signals. Signals can be represented by mathematical function but there are few constraints like bandwidth limitation and frequency limitation one needs to overcome while representing wave as a signal.

We will start by understanding, what are signals and signal processing ?

1. Signals

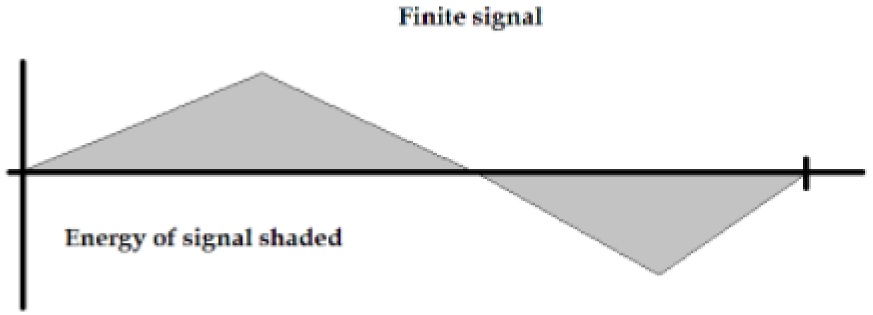
A class of real functions defined on a real line can be considered as a Signal. it could be the function of one or more than one independent variable.

A signal and its nature are defined by its characteristics.These characteristics are given below:

> 1. Amplitude
> 2. Frequency
> 3. Time Period
> 4. Phase

1.1 Signal Size : The size of a signal is a number that shows the strength or largeness of that signal. As we know, a signal’s amplitude varies with respect to time. Because of this variation, we cannot say that its amplitude can be its size. To measure the signal size, we have to take into account the area covered by the amplitude of the signal within the time duration.

1.2 Signal Energy: The energy of the signal is the area of the signal under its curve. But the signal can be in both positive and negative regions. Due to this, it will cancel each other's effect resulting in a smaller signal. To eradicate this problem, we take the square of the signal's amplitude which is always positive.



For a signal g(t), the area under the $\mathbf{g^2(t)}$ is known as the **Energy of the signal**.

$$E_g = \int_{-\infty}^{\infty} g^2(t)\, dt$$

1.3 Limitation : The energy of a signal can be measured only if the signal is finite. The infinite signal will have infinite energy, which is absurd. A finite signal's amplitude goes to 0 as the time (t) approaches **infinity (∞).**

Just like the energy of the signal, the measurement of the power of a signal also has some limitations that the signal must be of a periodic nature. An infinite and non-periodic signal neither has energy nor power.

So it is necessary that the signal is a finite signal if you want to measure its energy.

1.4 Signal Power : If the signal is an infinite signal i.e. its amplitude does not go to 0 as time t approaches to ∞, we cannot measure its energy. In such a case, we take the time average (Time period) of the energy of the signal as the power of the signal.

1.5 Deterministic And Random Signal

A signal which can be represented in mathematical or graphical form is called a deterministic signal. Deterministic signals have specified amplitude, frequency, etc. They are easy to process as they are defined over a long period of time and we can Evaluate their outcome if they are applied to a specific system based on their expression.

The random or non-deterministic signal is a signal which can only be represented in probabilistic expression rather than its full mathematical expression. Every signal that has some kind of uncertainty is a random signal. The noise signal is the best example of a random signal.

Generally, every message signal is a random signal because we are uncertain of the information to be conveyed to the other side.

Now, we have a basic understanding of the signals. The next step is to understand the concept of signal processing.

2. Singal Processing

Whether you're using your phone, turning on the radio, or strumming your electric guitar, you're sending and receiving signals all the time, and to get all that information, we need signal processing.

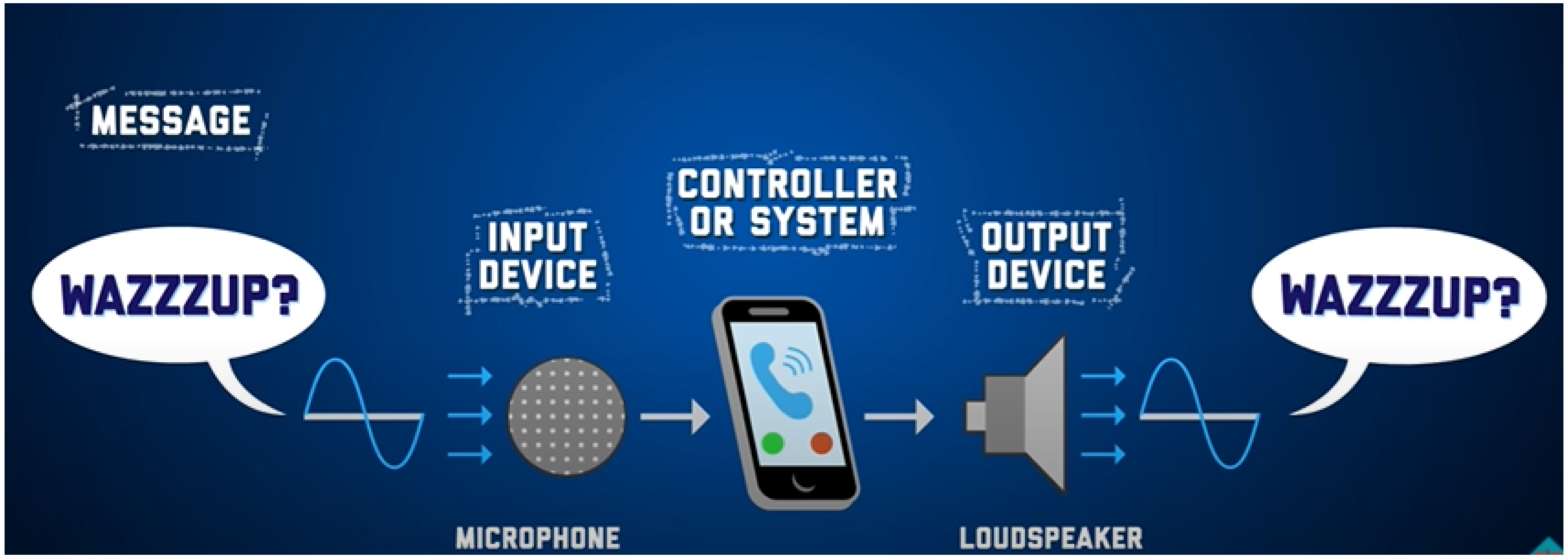
Testing 123

Can you hear me now?

Yeah!

Good!

As an engineer, communicating means more than having a chat in the break room. Whether you're watching YouTube videos, using satellite navigation, or just making a phone call, there's communication happening. Signals are representations of the information we're sending. Text, sounds, images, and even computer files will all be converted into a signal when we send them. communication is, sending stuff from one place to another to convey information. The basic task is to take content, turn it into a signal, transmit it, and then turn it all back into content on the other end. These steps are known as signal processing.



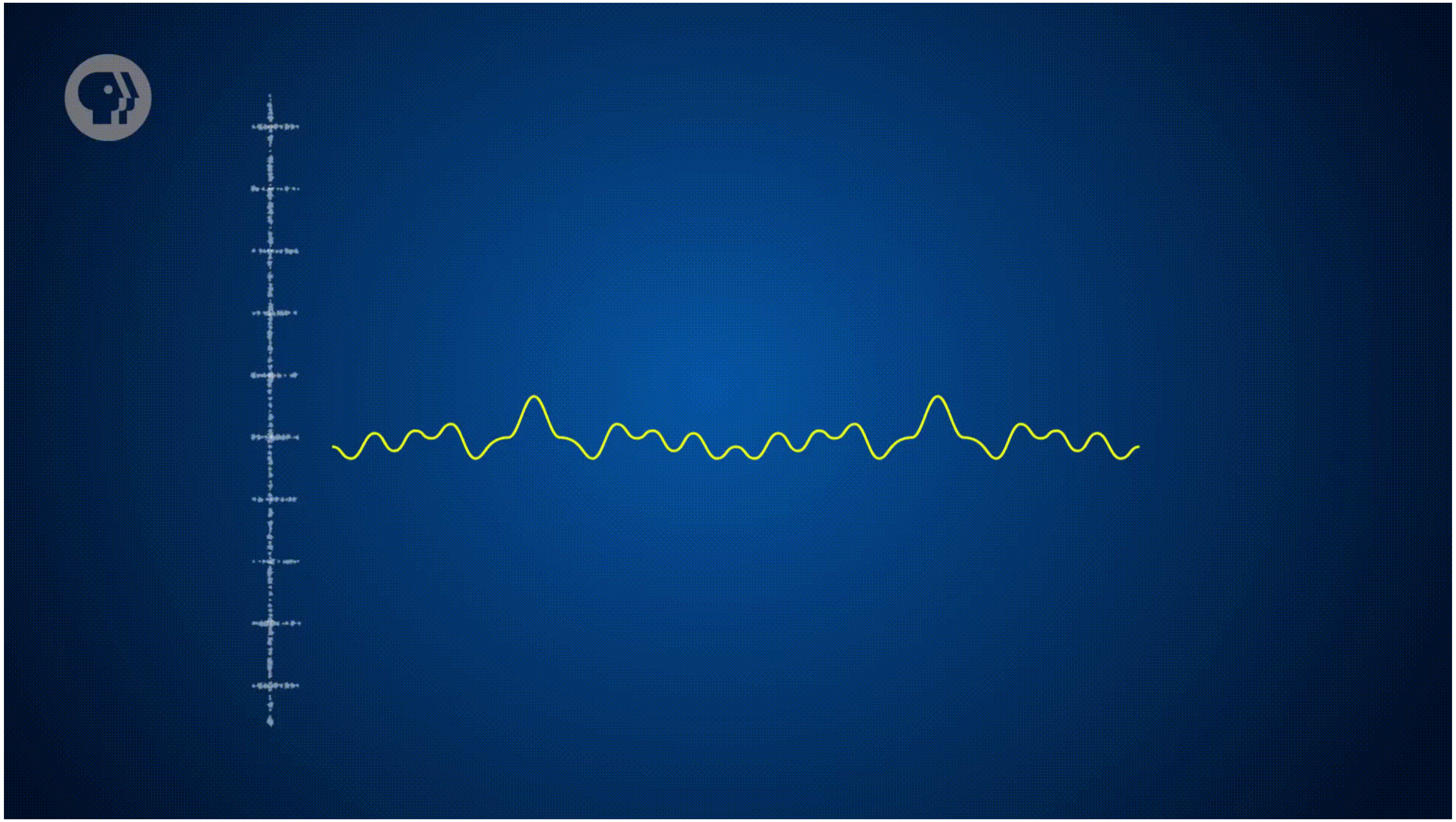
The signal itself will be a current running through a wire or an electromagnetic wave, like radio or light.

Consider radio waves, like the kind used to transmit signals between our phone and a cell tower. It's the wave nature of radio that lets our phone encode the information we need to make a call. Engineers design hardware that changes, or modulates, the behavior of that wave to encode information about the pressure of the air near the microphone, in other words, the physical effects of sound.

Two of the most common ways of doing this are **Amplitude Modulation** and **Frequency Modulation**, One adjusts the amplitude, or strength of the wave, while the other changes the frequency, or distance between one peak and the next. The transmitted wave carries the information you want, which is then decoded on the other side. Similar methods can even represent sounds and images.

But these methods have two pretty big limitations:

2.1.1 **Capacity :** The signal of a radio wave can be thought of as a combination of other, simpler waves put together. Specifically, we can represent a signal as the sum of radio waves with different frequencies. The range of different frequencies you can represent is called the **bandwidth**.

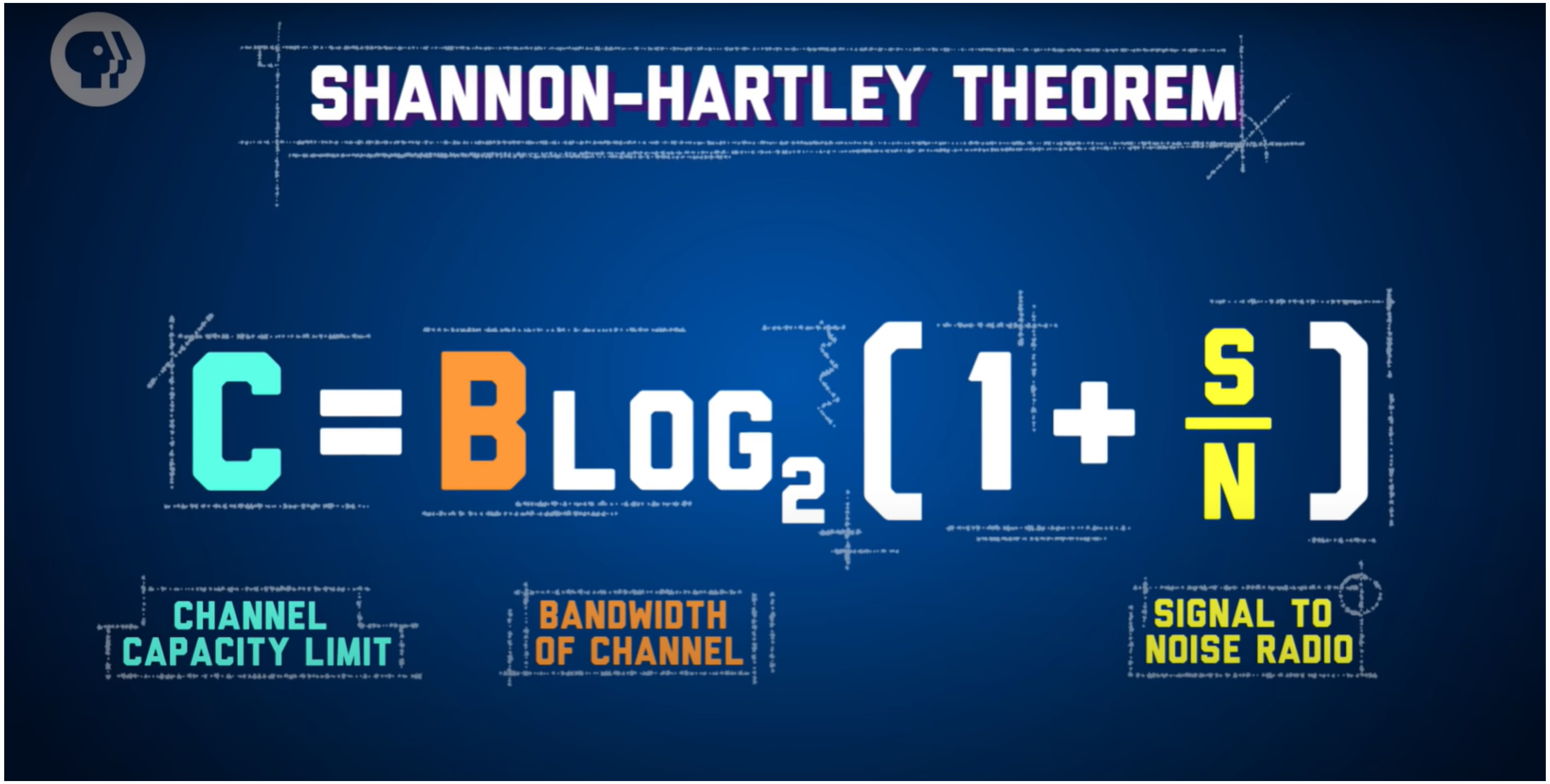


It limits how much information can be encoded by the signal, as well as how many of them can be sent at the same time.

2.1.2 **Noise:** As they travel through the atmosphere, radio waves interfere with each other and are warped by objects in their path, which both cause distortions.So, the signal the other person receives usually ends up pretty different from the one that we sent. Noise is anything that changes your signal from its original form, usually in a random way.The greater the noise, the more distorted and unrecognizable the received message.

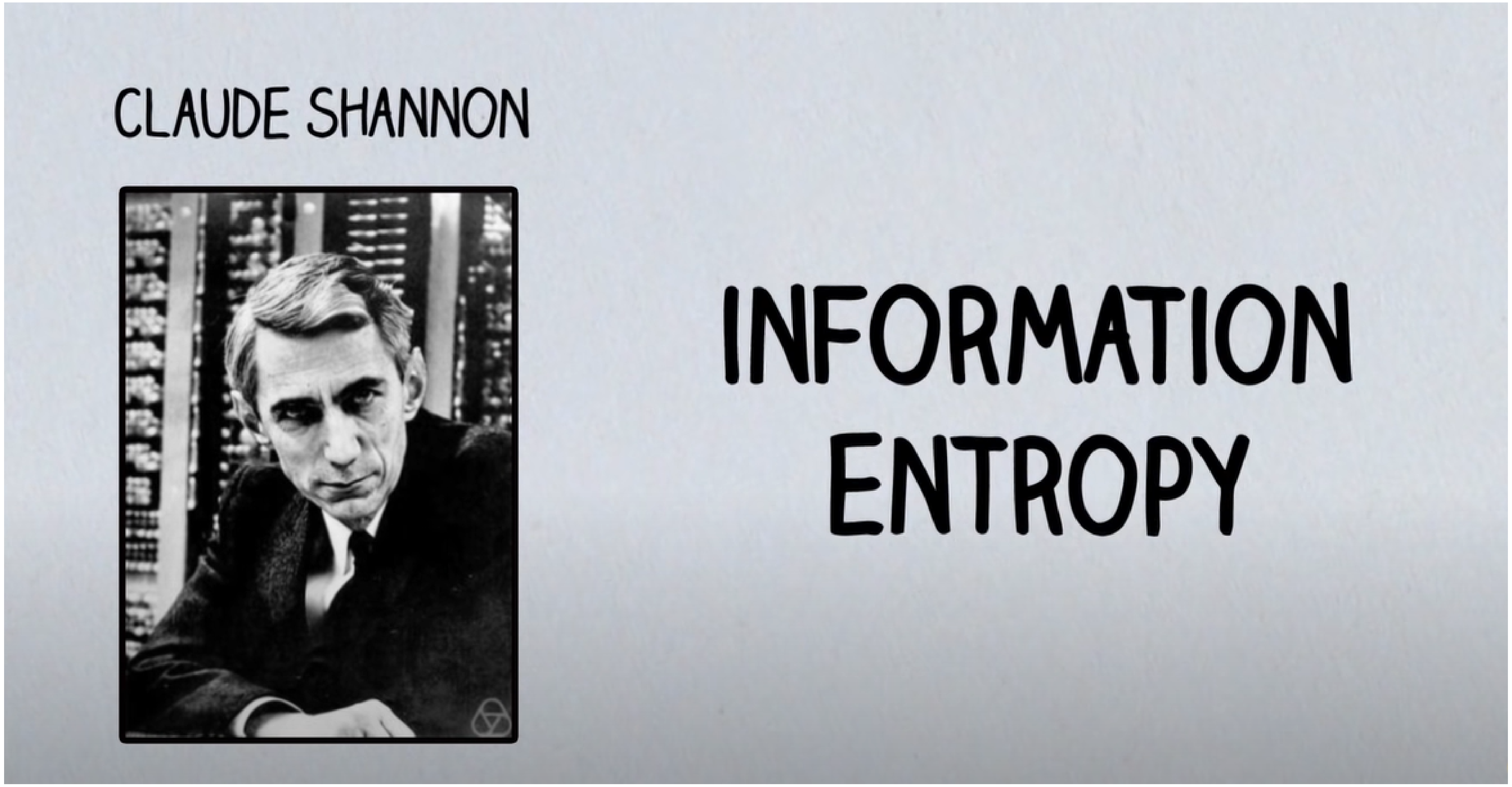
Both of these problems happen for wired communications as well. The signal traveling down a wire is also a wave, where the amplitude is represented by the power of the electric current at any given point in time.That’s how we modulate electric currents to carry signals, but it also means that those signals suffer from noise and capacity issues, too.

That’s how we modulate electric currents to carry signals, but it also means that those signals suffer from noise and capacity issues too, Which brought them to the attention of engineer and mathematician Claude Shannon. In 1948, he published **A Mathematical Theory of Communication** which revolutionized how engineers consider information itself, and what it takes to send information reliably. Among Shannon’s contributions was a mathematical formula for determining the conditions needed for sending a signal at a particular rate. In his paper, Shannon developed a formula that determines the number of bits you can transmit per second, or “bit-rate”. He figured out that it’s the ratio of the power of the signal to the power of the noise that determines the bit rate. So, either the signal needs to be strong enough, or the bandwidth needs to be large enough for there to be so many frequencies representing the signal that noise can’t affect them all at once.



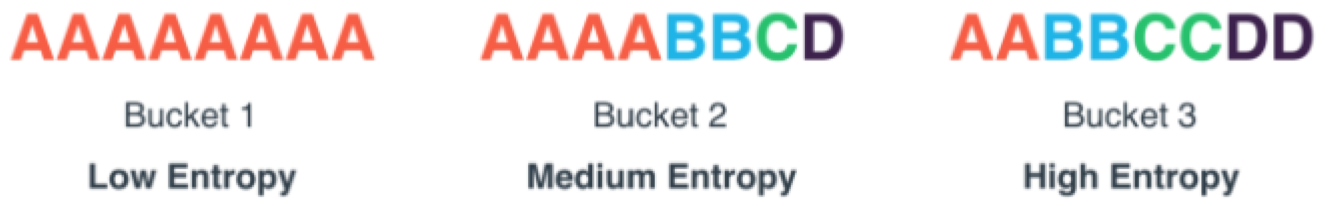
3. Information

For Understanding the core concept first we need to go understand, what is information ?



In 1948 Claude Shannon introduced the world to the notion of information entropy in his paper “A Mathematical Theory of Communication”

In information theory, we relate the concept of entropy with randomness. let's try to understand it with an example.



let's start by picking a random letter from any bucket and think that at average how many yes/no questions you need to ask for finding out what letter we picked.

Bucket 1: All the letters are A, so we know that letter picked is A. So, the Average number of questions is 0 for this case. Now for Bucket 2 and Bucket 3 :

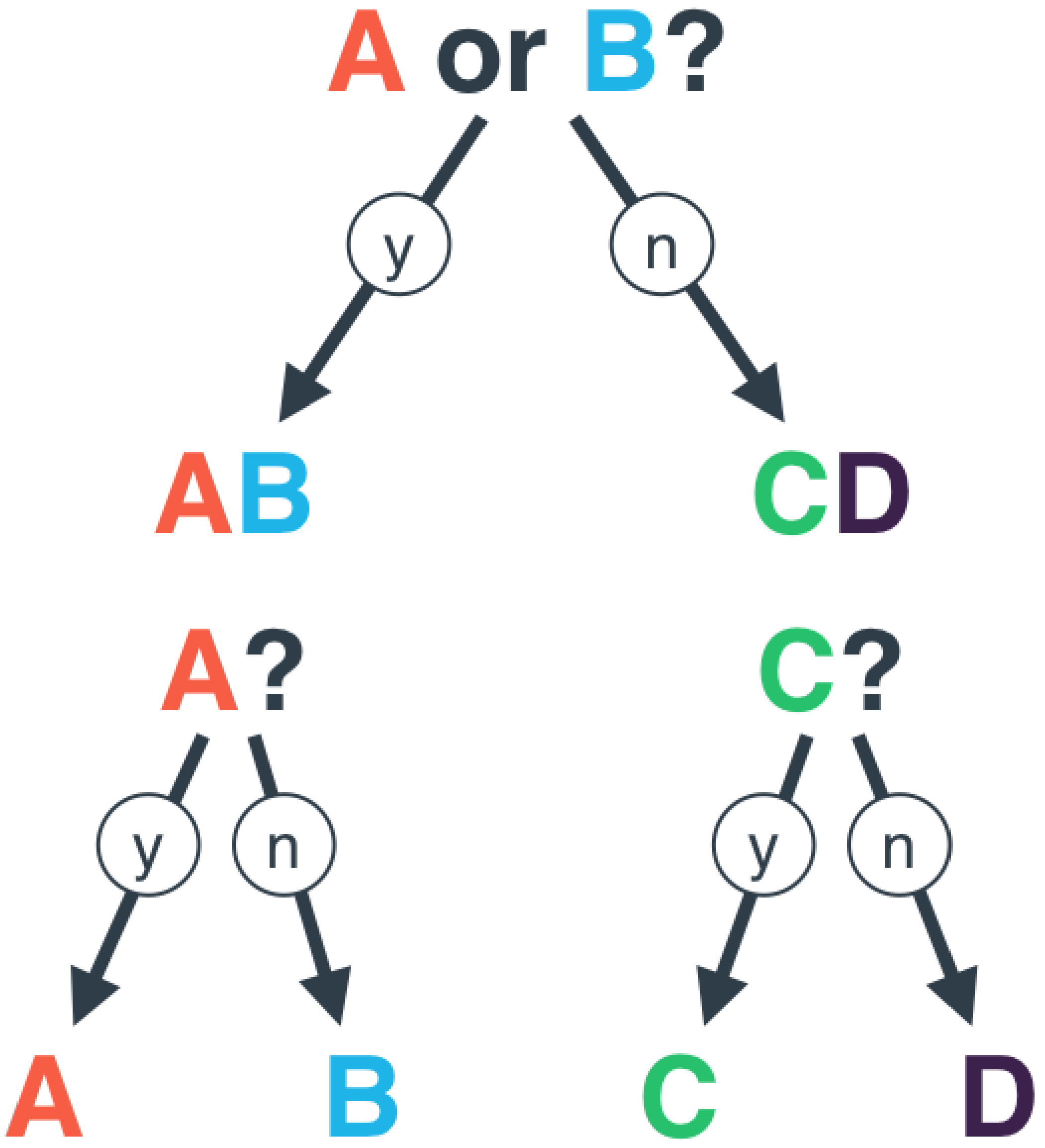
***Question 1:** Is the random letter A or B.

*If Yes *

Question 2: Is it A

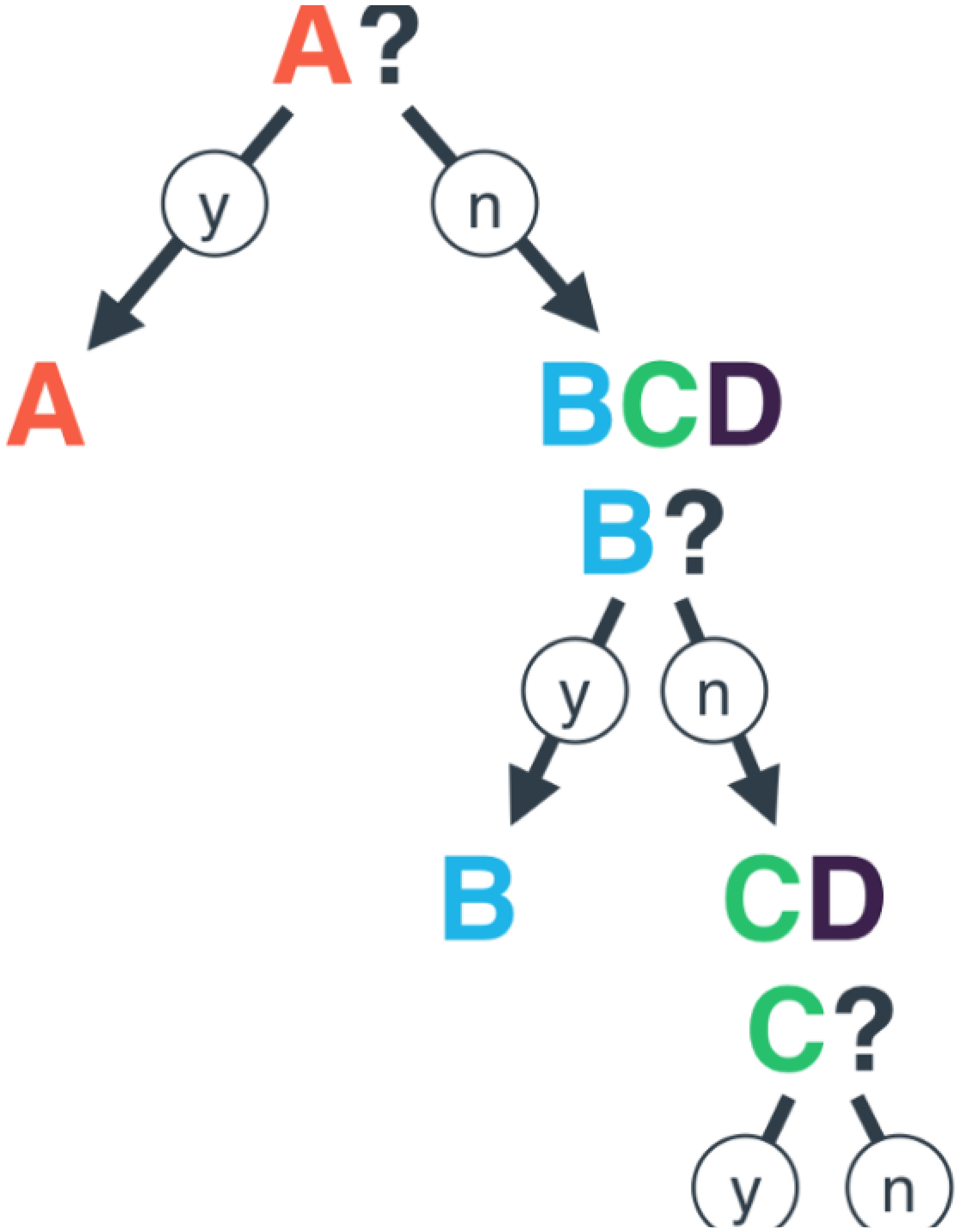
Similarly, we can create a tree of binary questions

Tree for Bucket 3


A or B?
y
n
AB
CD
A?
C?
y
n
y
n
A
B
C
D

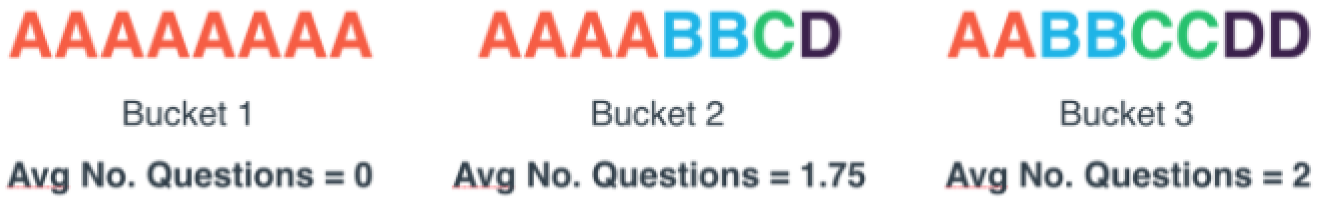
$$\text{Average Number of Questions} = \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 2$$

Tree for Bucket 2


A?
y
n
A
BCD
B?
y
n
B
CD
C?
y
n

$$\text{Average Number of Questions} = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = 1.75$$

So, for average number of questions for each bucket are as follows :



The Numbers of questions asked for getting the final state of the system gives the idea about randomness or information in the system.

3.1 Probabilistic Information Measures

For understanding the entropy it is relevant to understand the concept of bits can be either 0 or 1 which makes it similar to the task we were performing above of representing 2 facts (aka information), that is either true or false.

$$outcome = 1/p$$

p = probability of that outcome

$$bounces = log_2(1/p)$$

$$\begin{aligned} H(X) &= -\sum_{x} P(x)\cdot\log P(x) \\ &= \sum_{x} \underbrace{P(x)}_{\text{The prob. of event X}} \cdot \underbrace{\log\left(\frac{1}{P(x)}\right)}_{\text{WHAT IS THIS ?}} \end{aligned}$$

Because -log P(x) = log (1/P(x))

H(X) = total amount of information in a probability distribution.

In above example think each level as a bounce. So, number of bounces depends on outcome at each level. The number of outcome at every bounce is based on probability.

Why 1/p(x) : In above example, the probability of all 4 events in Bucket 1 is equally likely which is (p = 1/4)

We are using waves as a information carrier , which make it important to defiend the notion of information in term of probability. This function theory of information based on capacity and entropy is given by the school of Kolmogorov which was inspired by Shannon's probabilistic theory of information. The probabilistic point of view is also based on the surprise element in the highly unpredictable signal. The amount of information a signal transport identifies the surprise. The coefficients representing the signal are random variables described by a probability distribution. Considering the noise affecting the signal the Shannon capacity is defined in terms of the largest number of bits that can be transported from transmitter to receiver by anyone signal in the space, with a negligible probability of error. This can be described in terms of another probabilistic quantity, the mutual information, which depends only on the joint distribution of the observed and transmitted signals

Apart from all theory, it will be great if we can see the amount of information in any sentence

3.2 The code below calculate the amount of information in a sentence you are passing into it

In [19]:
```
import math
import random

def H(sentence):

    entropy = 0
    # 256 possible ASCII characters
    for character_i in range(256):
        Px = sentence.count(chr(character_i))/len(sentence)
        if Px > 0:
            entropy += - Px * math.log(Px, 2)
    return entropy
```

In [20]:
```
short_msg ="".join([chr(random.randint(0,32)) for i in range(10000)])
```

In [21]:
```
H(short_msg)
```

Out[21]:
```
5.041705973790044
```

In [22]:
```
# using all 255 chars
All_chars ="".join([chr(random.randint(0,255)) for i in range(10000)])
```

In [23]:
```
H(All_chars)
```

Out[23]:
```
7.981877345616825
```

Isn't it cool to check amount of information in your name

In [38]:
```
H("Vishal")
```

Out[38]:
```
2.584962500721156
```

Every wave we use as an information carrier can essentially be defined by N real numbers. Getting the accurate representation of waves is impossible due to storage limitation issues. However, the result does not limit the information amount a signal is carrying. The N real numbers identifying the waveform can be specified up to arbitrary precision. Waveforms that are distinct as mathematical objects cannot be resolved as distinct physical quantities at a scale smaller than the noise affecting the measurement process. For this reason, we may consider two waveforms distinguishable only if, for a given ϵ ϵ

$$\|f_1 - f_2\| > \epsilon.$$

3.3 Kolmogorov Entropy

Continuing the investigation of Shannon theory in 1956 Andrey Nikolaevich introduced the notion of ϵ - entropy Kolmogorov Entropy which is the minimum number **H** of bits that can represent any signal in the space with accuracy ϵ

$$\left(\int_{-\infty}^{\infty} f^2(t)dt\right)^{1/2} = \left(\sum_{n=1}^{N_0} a_n^2\right)^{1/2} \leq \sqrt{E},$$

The entropy geometrically corresponds to the logarithm base two of the covering number Q(E), which is the minimum number of balls of radius such that their union contains the whole space which helps in identifying any point in the space.

$$H_\epsilon = \log Q_\epsilon(E) \text{ bits}.$$

The volume of ϵ ball divides the volume of the space which gives the reasonable guess about the entropy of the system

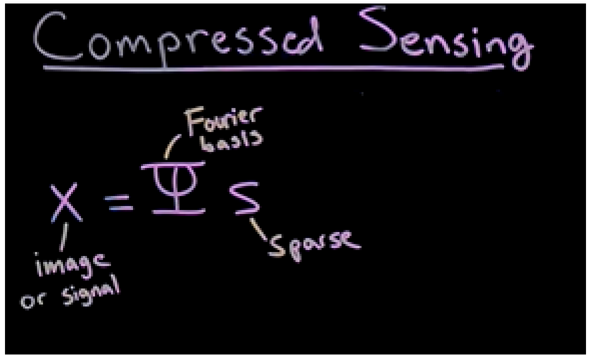
$$H_\epsilon = \log(\sqrt{E}/\epsilon)^{N_0} = N_0 \log(\sqrt{E}/\epsilon).$$

4. Functional Approximation

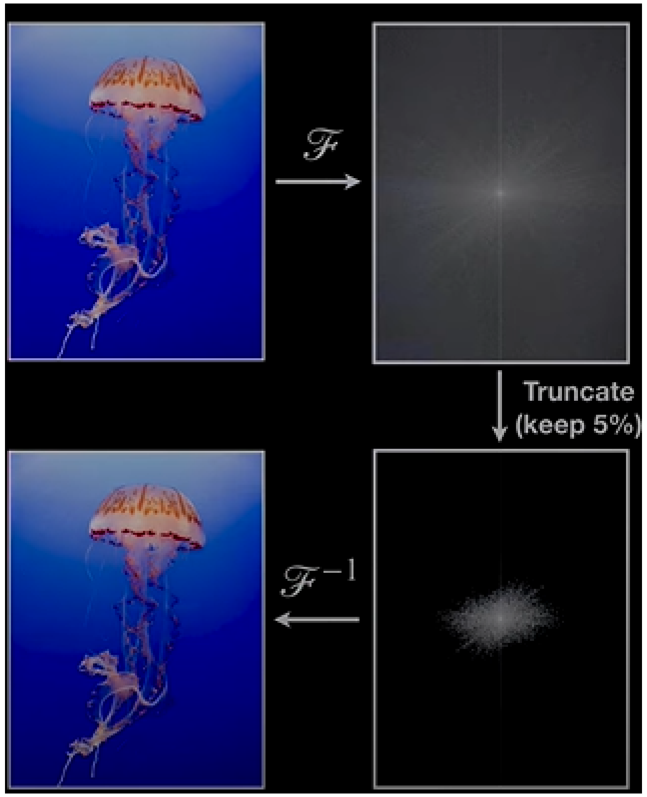
The set of signals for the communication engineer corresponds to an infinite-dimensional functional space for the mathematician. This can be viewed as a vector space on which norm (i.e., length), inner product (i.e., angle), and limits can be defined. The engineering problem of determining the affective dimension of the signals' space then falls in the mathematical framework of approximation theory that is concerned with finding a sequence of functions with desirable properties whose linear combination best approximates a given limit function. The approximation is a wellstudied problem in analysis. In terms of abstract Hilbert spaces, the problem is to determine what functions can asymptotically generate a given Hilbert space in the sense that the closure of their span (i.e., finite linear combinations and limits of those) is the whole space. Such a set of functions is called a basis, and its cardinality, taken in a suitable limiting sense, is the Hilbert dimension of the space. Various differential equations arising in physics have orthogonal solutions that can be interpreted as bases of Hilbert spaces. One example is the solution of the wave equation leading to the prolate spheroidal wave functions examined in the previous chapter. Another notable example arises in quantum mechanics in the context of the Schrödinger differential equation

4.1. Compressed Sensing

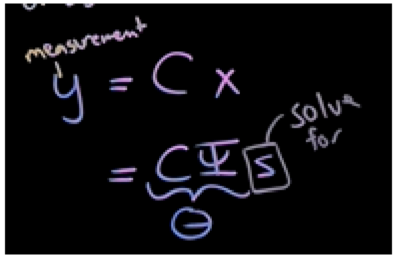
It is a way of taking a high dimensional signal like an image or an audio signal and reconstructing that signal even if you have massively down sampled measurements. it is based on advances in sparse optimization.



Let's say we have high dimensional signal X(image in this example), the idea here is that if get the Fourier to transform this image we can get away with throwing most of the Fourier coefficients maybe we only keep 5% or 1%, and when we inverse Fourier transform that sparse signal we can recover a pretty faithful representation of the original image the original signal X



This is standard compression happens on your computers and smartphones. This is a relatively recent advance in the last 15 or so years is begging the question what if you actually only had a very few measurements of your systems. so remember we're starting with million pixels of this kind of megapixel image and what we're going to do is, we take this megapixel image and its Fourier transform and throw most of it away we're going to throw 95 or 99% of this information away why did I collect all of that information in the first place. So, that's the whole idea of compressed sensing is what if we start with a massively downsampled or sub-sampled version of the image. From those measurements can we infer what the nonzero coefficients are what the sparse vector had to be consistent with the measurements and then from that sparse vector **s** we can again inverse Fourier transform and reconstruct the full image.



the idea here is that Y is our measurement and ideally we're gonna measure many fewer entries than the earlier. If we are going to keep one percent of my Fourier coefficients then course we have to measure more than one percent of pixels. These measurements have to to be random. we're solving for this sparse vector **s** that is consistent with these measurements. Once we have **s** since we know that this was sparse in Fourier we just inverse Fourier transform and we recover our full image that's compressed sensing.

5. APPLICATION

One of the application of wave theory is fourier transform. It can be used for classfication of signals as well. One of such example is given below.

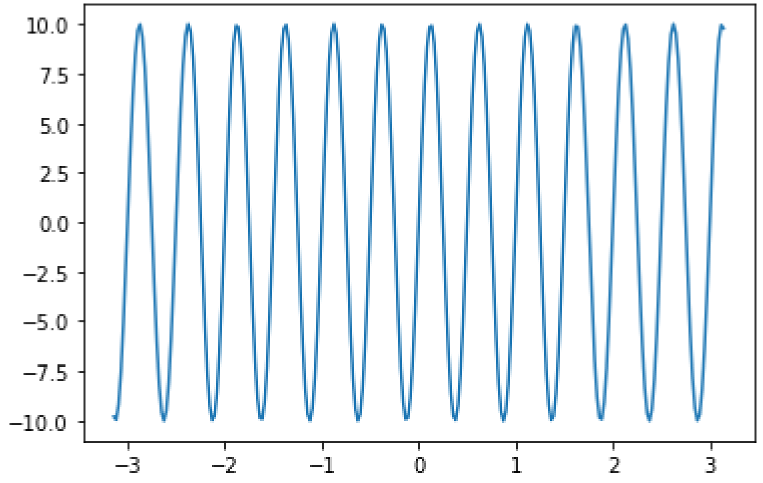
In [1]:

```
import matplotlib.pyplot as plt
import numpy as np
from scipy import signal
```

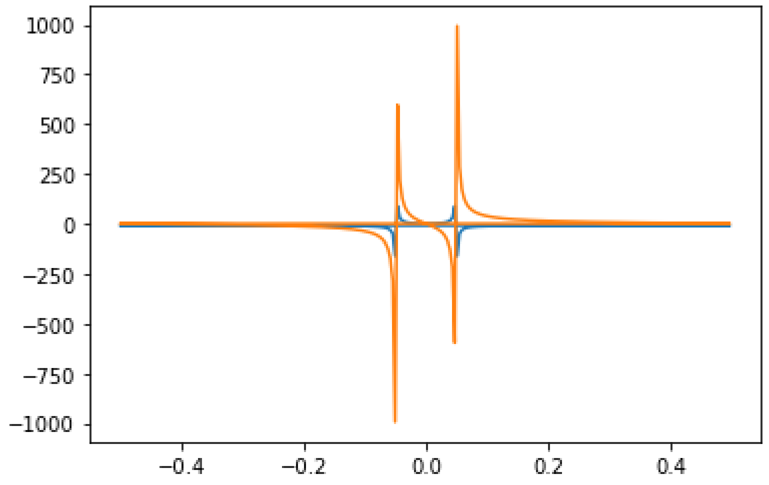
In [3]:

```
A=float(input('Enter amplitude\t:'))
f=int(input('Enter frequency\t:'))
t=np.linspace(-np.pi, np.pi,256)
y=A*np.sin(2*np.pi*f*t)
plt.plot(t,y)
plt.show()
#print(t)
```

```
Enter amplitude :10
Enter frequency :2
```



In [4]:
```
sp = np.fft.fft(y)
freq = np.fft.fftfreq(y.shape[-1])
plt.plot(freq, sp.real, freq, sp.imag)
plt.show()
```

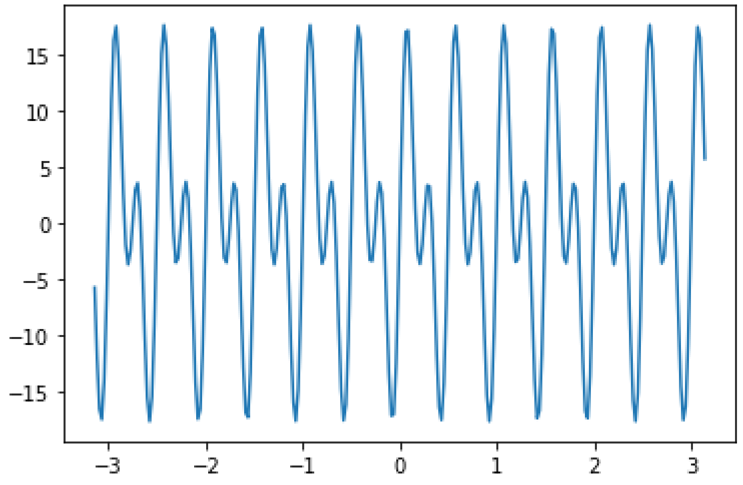


It shows says that there is one distinct frequency in the signal.

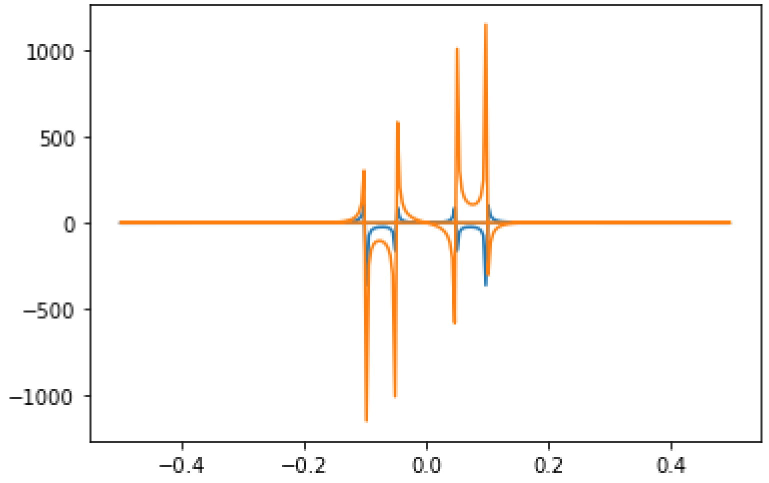
In [5]:

```
A1=float(input('Enter amplitude\t:'))
f1=int(input('Enter frequency\t:'))
t=np.linspace(-np.pi,np.pi,256)
y1=A1*np.sin(2*np.pi*f1*t)
A2=float(input('Enter amplitude\t:'))
f2=int(input('Enter frequency\t:'))
y2=A2*np.sin(2*np.pi*f2*t)
y=y1+y2
plt.plot(t,y)
plt.show()
#print(t)
```

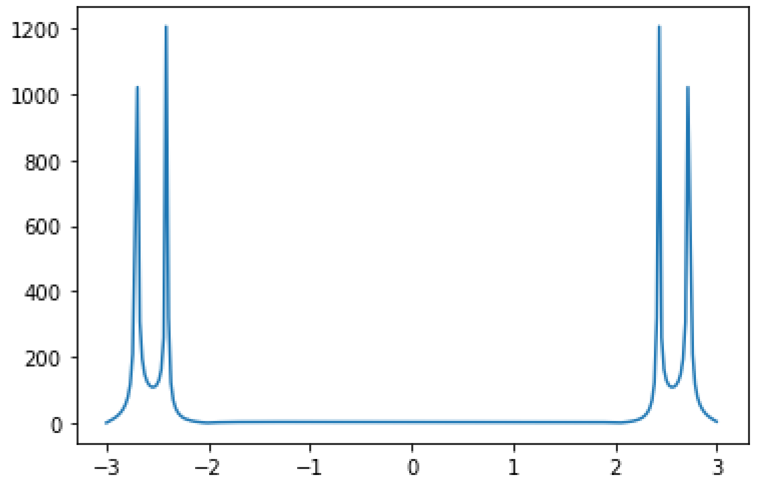
```
Enter amplitude :10
Enter frequency :2
Enter amplitude :10
Enter frequency :4
```



In [6]:
```
sp = np.fft.fft(y)
freq = np.fft.fftfreq(y.shape[-1])
plt.plot(freq, sp.real, freq, sp.imag)
plt.show()
f=np.linspace(-3,3,256)
mod1=[np.sqrt((i.real**2+i.imag**2)) for i in sp]
mod2=np.abs(sp)
plt.plot(f,mod2)
```



Out[6]:
```
[<matplotlib.lines.Line2D at 0x2069f48c1c0>]
```



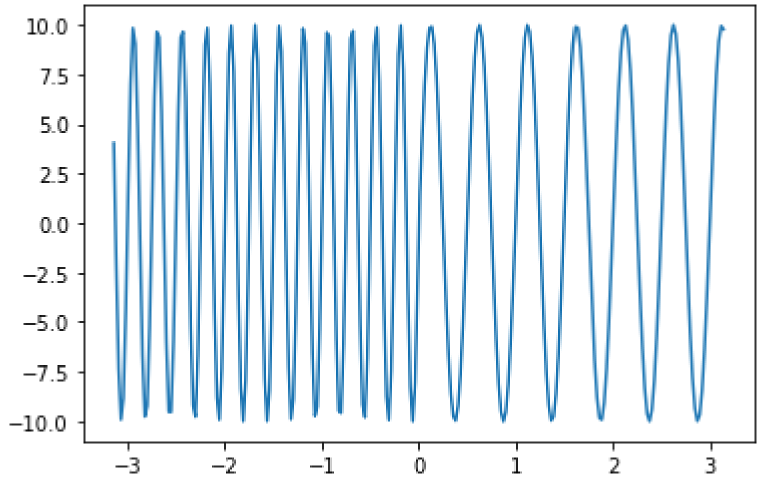
It shows there are 2 different frequencies in the signal

Limitation : It doesn't tell the order of the frequencies.

Solution to this is STFT : Take a window and tell the frequencies in that window.

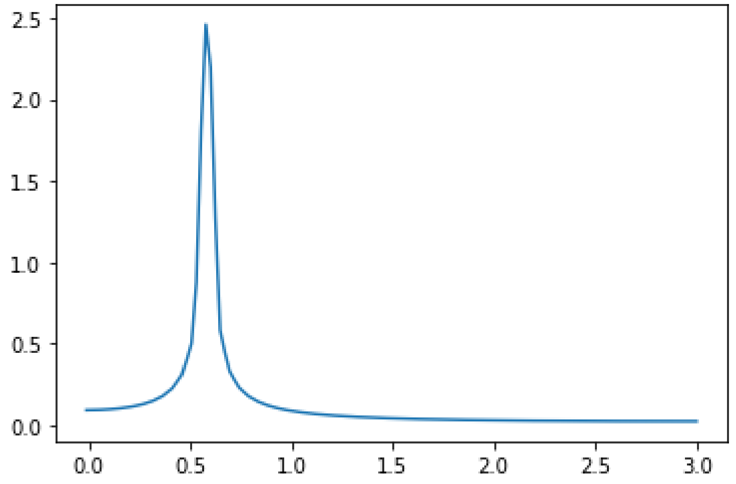
In [7]:
```
A=float(input('Enter amplitude\t:'))
f1=int(input('Enter frequency 1\t:'))
f2=int(input('Enter frequency 2\t:'))
t=np.linspace(-np.pi, np.pi,256)
y=np.zeros(256)
for i in range(128):
    y[i]=A*np.sin(2*np.pi*f1*t[i])
for i in range(128,256):
    y[i]=A*np.sin(2*np.pi*f2*t[i])
plt.plot(t,y)
plt.show()
#print(t)
```

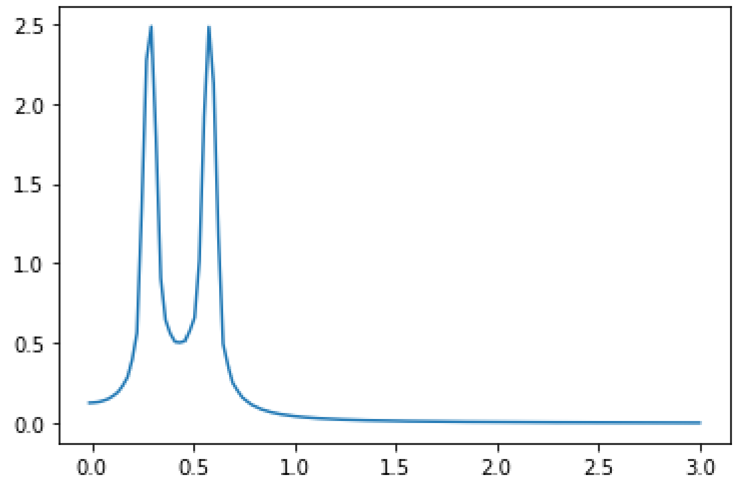
```
Enter amplitude :10
Enter frequency 1       :4
Enter frequency 2       :2
```

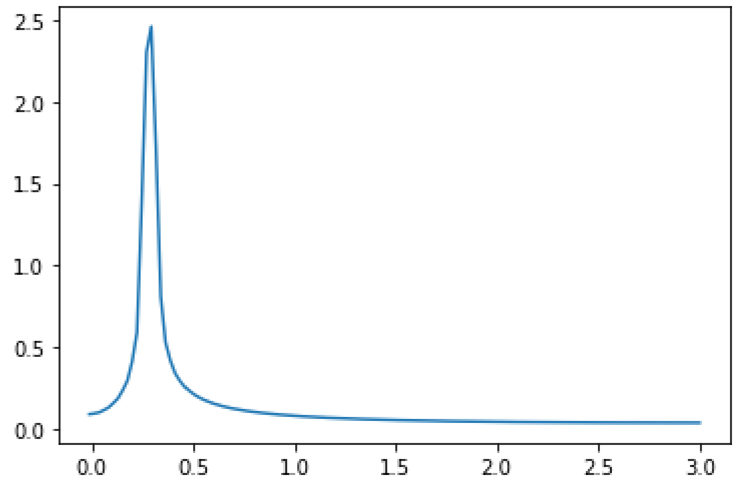


In [8]:
```
f, t,sp = signal.stft(y, 1000, nperseg=1000)
#plt.plot(freq, sp.real, freq, sp.imag)
plt.show()
f=np.linspace(-3,3,256)
mod1=[np.sqrt((i.real**2+i.imag**2)) for i in sp]
mod2=np.abs(sp)
plt.plot(f[127:],mod2[:,0])
plt.show()
plt.plot(f[127:],mod2[:,1])
plt.show()
plt.plot(f[127:],mod2[:,2])
plt.show()
```

```
C:\Users\visha\anaconda3\lib\site-packages\scipy\signal\spectral.py:1961: UserWarning: nperseg = 1000 is greater th
an input length  = 256, using nperseg = 256
  warnings.warn('nperseg = {0:d} is greater than input length '
```




2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

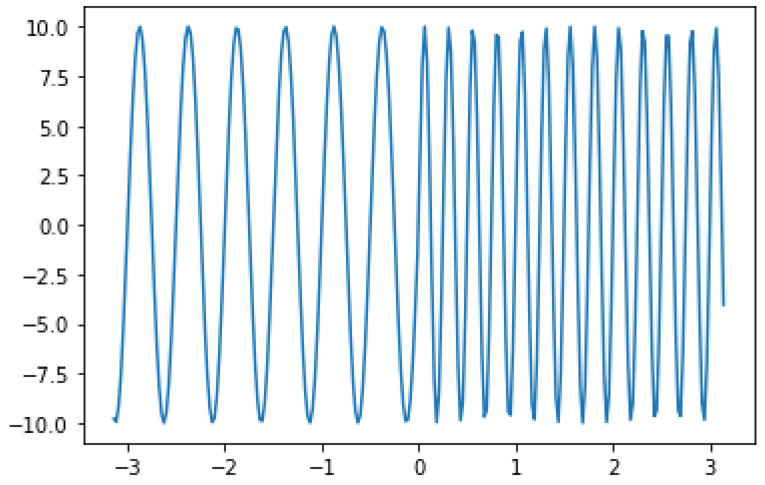

2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

1st is 1st frequency , 2nd part is combination of both and 3rd is giving the third frequency

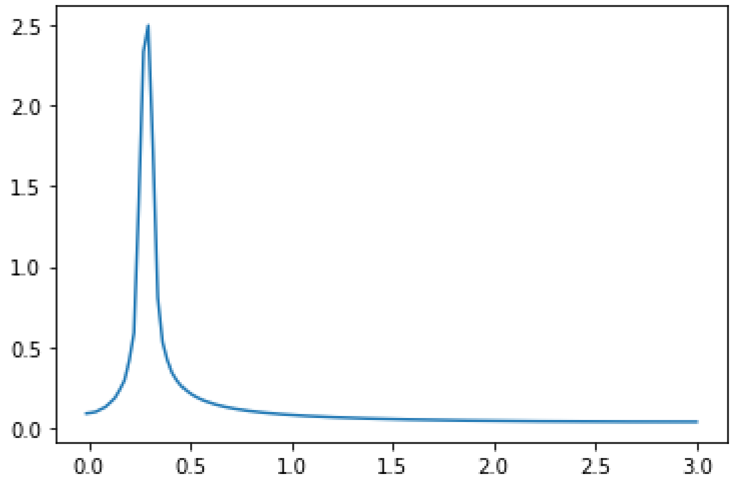
In [10]:

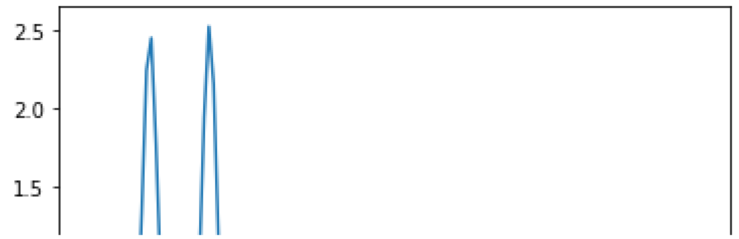
```
A=float(input('Enter amplitude\t:'))
f1=int(input('Enter frequency 1\t:'))
f2=int(input('Enter frequency 2\t:'))
t=np.linspace(-np.pi, np.pi,256)
y=np.zeros(256)
for i in range(128):
    y[i]=A*np.sin(2*np.pi*f1*t[i])
for i in range(128,256):
    y[i]=A*np.sin(2*np.pi*f2*t[i])
plt.plot(t,y)
plt.show()
#print(t)
f, t,sp = signal.stft(y, 1000, nperseg=1000)
#plt.plot(freq, sp.real, freq, sp.imag)
plt.show()
f=np.linspace(-3,3,256)
mod1=[np.sqrt((i.real**2+i.imag**2)) for i in sp]
mod2=np.abs(sp)
plt.plot(f[127:],mod2[:,0])
plt.show()
plt.plot(f[127:],mod2[:,1])
plt.show()
plt.plot(f[127:],mod2[:,2])
plt.show()
```

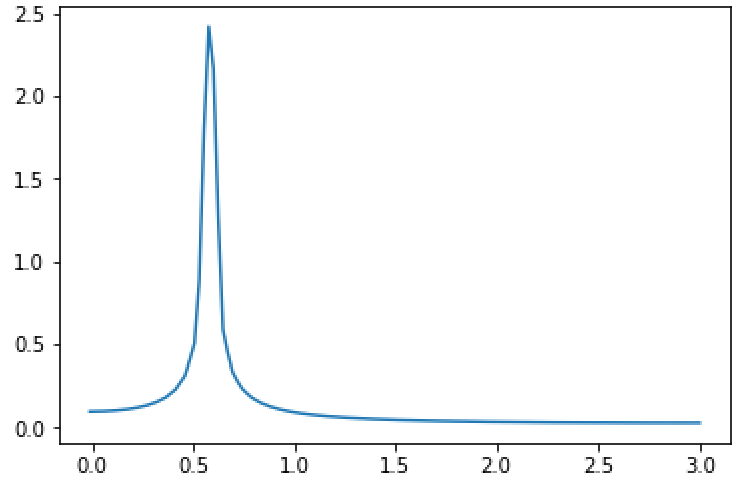
```
Enter amplitude :10
Enter frequency 1       :2
Enter frequency 2       :4
```



```
C:\Users\visha\anaconda3\lib\site-packages\scipy\signal\spectral.py:1961: UserWarning: nperseg = 1000 is greater th
an input length  = 256, using nperseg = 256
  warnings.warn('nperseg = {0:d} is greater than input length '
```







Summary

Shannon’s model of communication using bandlimited signals was introduced in Shannon (1949). The rigorous formulation of his capacity result for continuous signals is given by Wyner (1965). Close examination of these results shows that refined noise models are required to account for additional quantum constraints and super-resolution effects occurring at high frequencies.

```
In [ ]:
```