The ladybird guides

Book II

Day 2 - modulation

Day 2: Modulation

Objective:

To understand why information is modulated onto high frequency waves

To appreciate the choices available

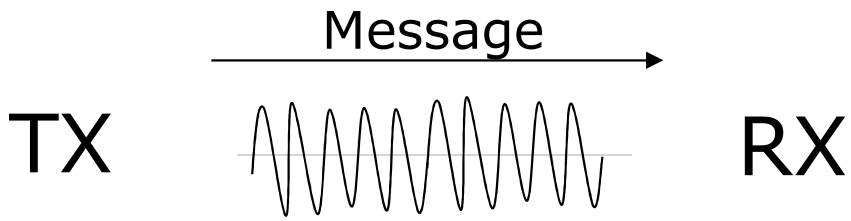
To understand the basic techniques and their advantages & disadvantages

To appreciate the importance of baseband signal choice

To describe real modulation systems

To describe a theoretical demodulation system

Design constraint

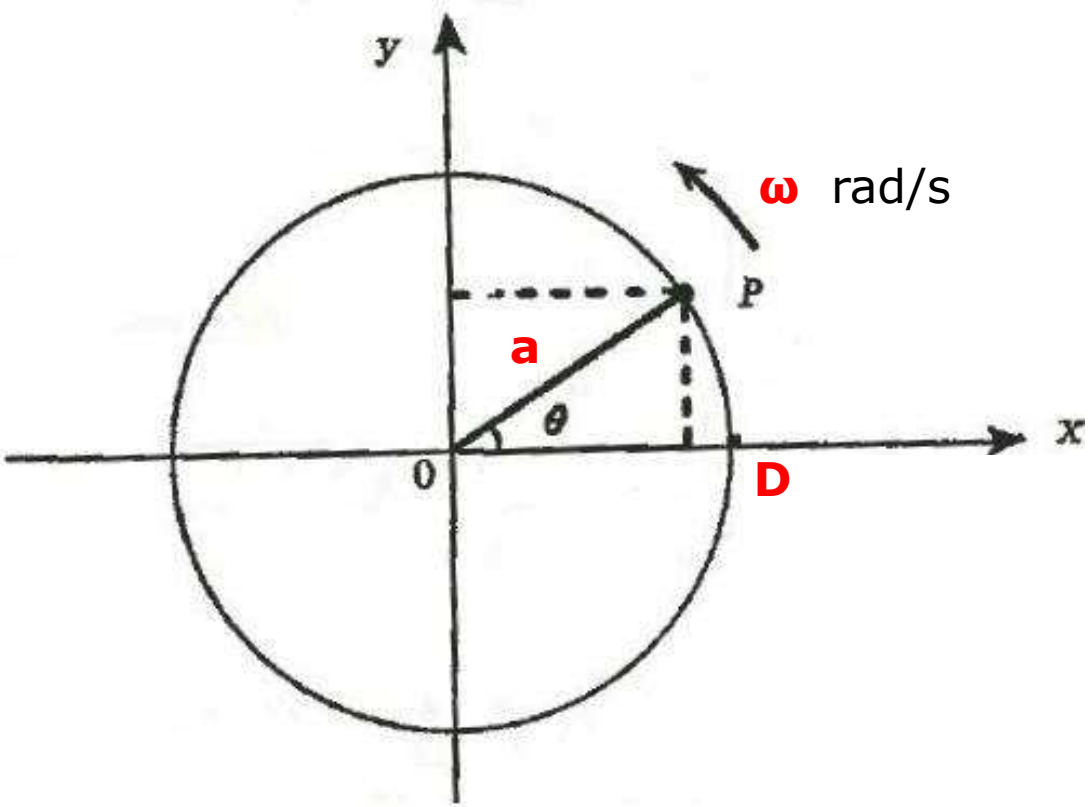


Our message exists as a binary number in a computer whereas an electromagnetic wave is a real thing with physical properties. We can touch it, it can even harm us…..

How do we imprint our “virtual” message onto an electromagnetic wave?

MODULATION

Modulation - Basics

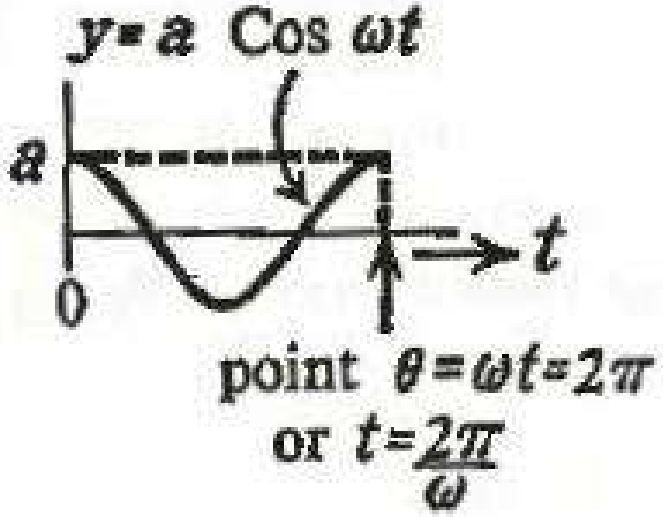


OX = OP cos θ
OY = OP sin θ
OP = OD = a

If we start at point D then
θ = ωt

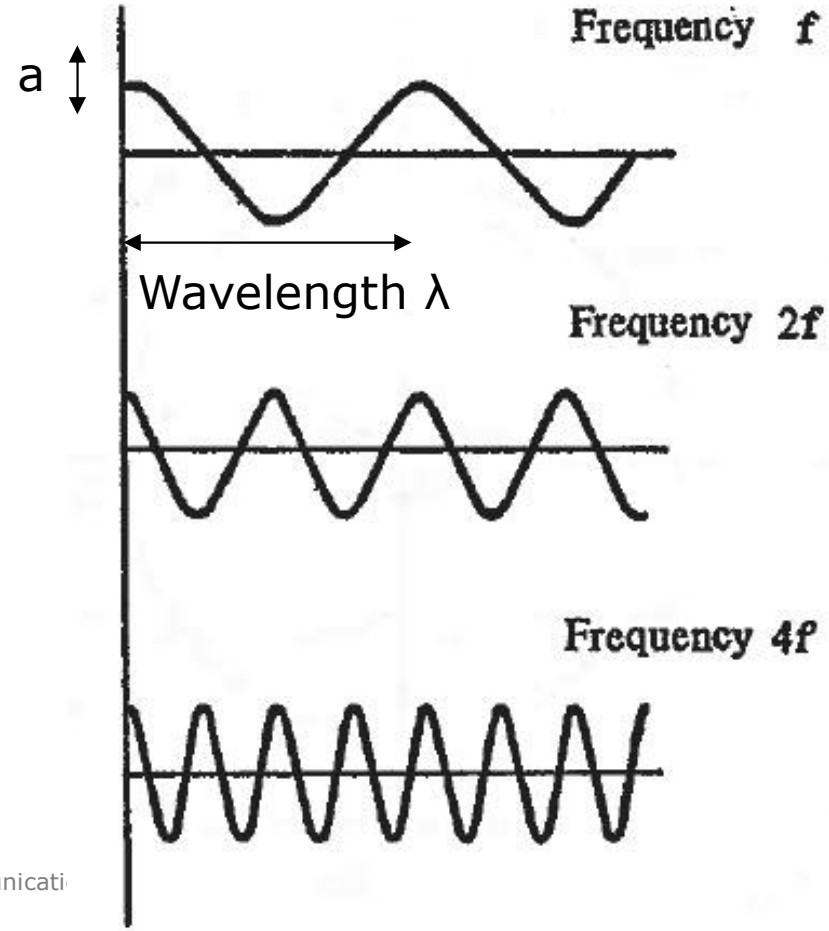
We have three defining proper...

1) Size of arm **(amplitude)**
2) Speed of rotation **(frequency)**
3) Starting point **(phase)**



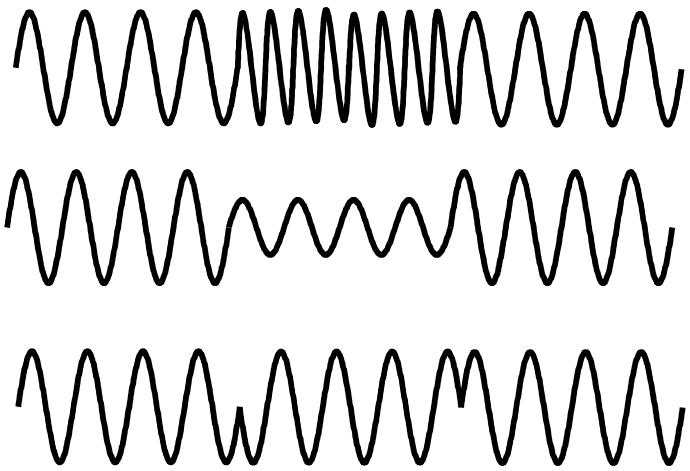
Modulation - Basics

Number of rotations per second is called the frequency (ω/2Π)
Distance from peak to peak is one wavelength λ
One period, T is the time it takes to complete one wavelength
Size of a peak is called the Amplitude, a



Modulation - Basics

- Frequency
- Amplitude
- Phase



Amplitude modulation is obviously not suitable for our design due to the massive variations in signal strength already discussed.

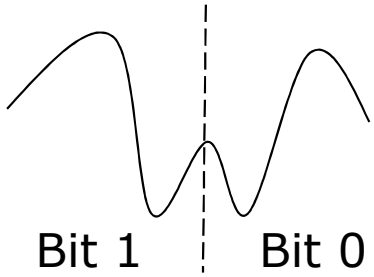
Frequency modulation could be used (and is by radio amateurs for example). It is much easier to implement than phase modulation but crucially requires more bandwidth.

Hence all ESA satellite communications use phase modulation.

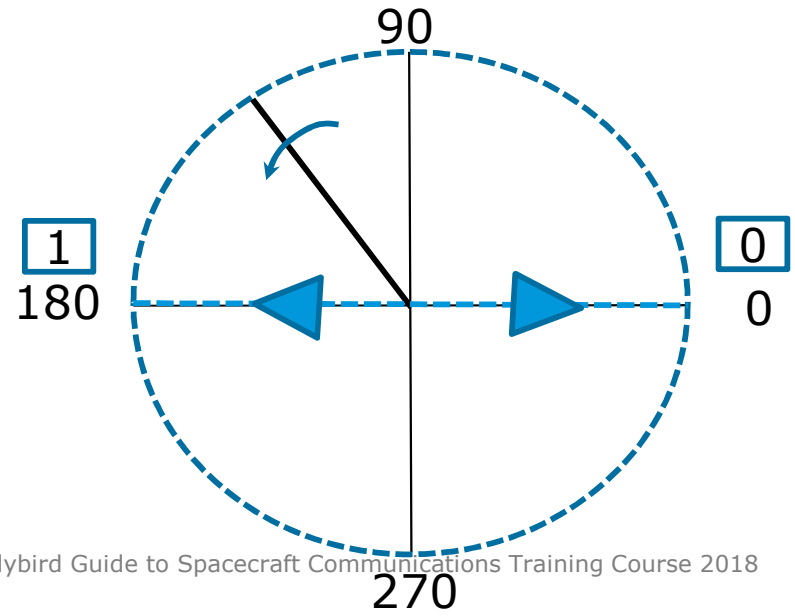
But how do we change the phase?

We use a system called phase shift keying (PSK) is used to modulate the a signal with digital data.

The simplest form of PSK is Binary Phase Shift Keying (BPSK). This is represented below

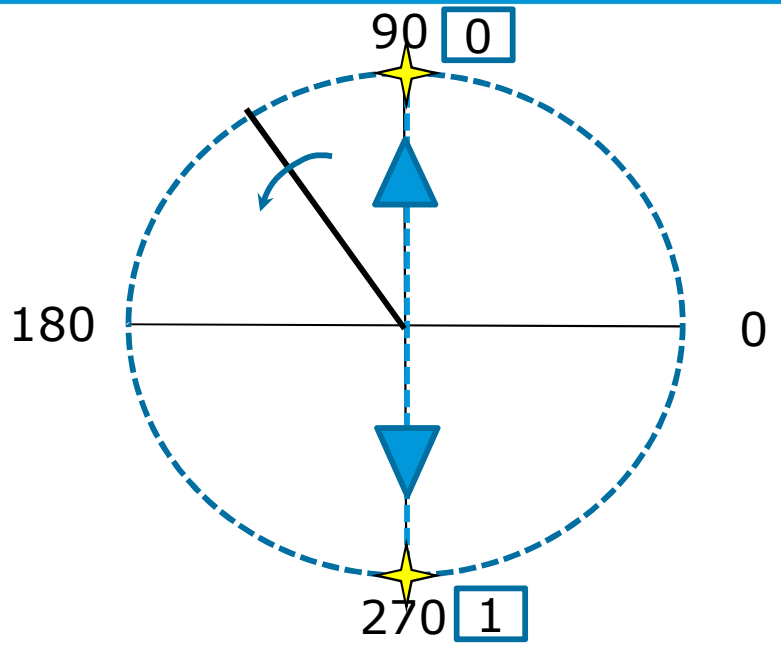


A one and a zero are represented by 180° changes in phase.

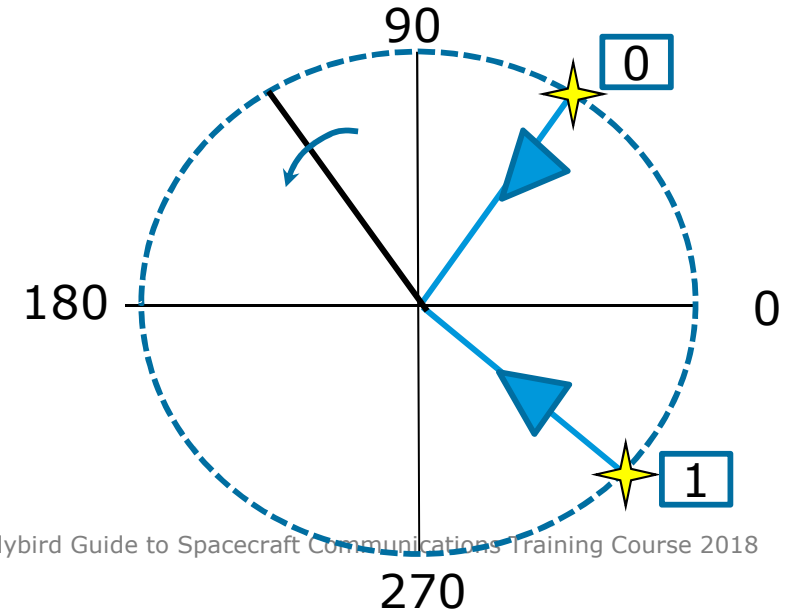


We can represent this on a constellation diagram as such. The difference in phases is called the modulation index. Here it is 0° and 180° i.e. 0 and ∏ rad.

BPSK and modulation index



We could also have chosen a modulation index of 90° and 270° i.e. +/- ∏/2 rad (this is the more common scheme).



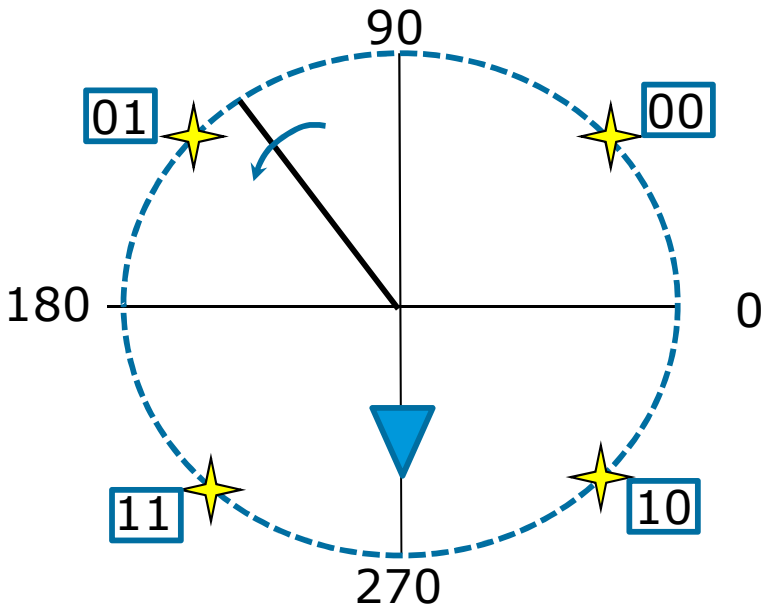
Note we can choose any two values of phase. This would result in a modulation index of less than +/- ∏/2 rad. This is often simply called PSK.

We send Channel Symbols on a microwave not bits

Each point on the constellation diagram is called a channel symbol.

For BPSK each channel symbol represents one bit. However things can get exciting if we increase the number of phase changes we use to four.

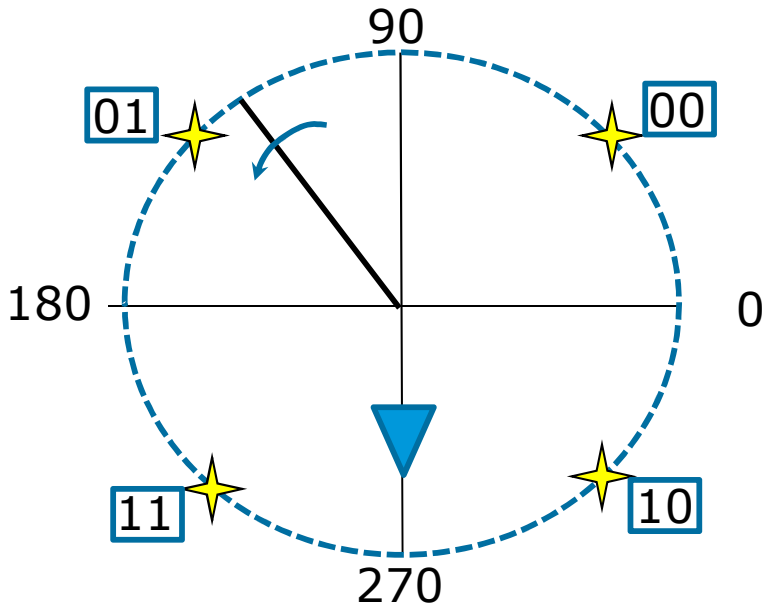
We can now assign two unique bit combinations to each of the four changes. This is called Quadrature Phase Shift Keying (QPSK).



In this case each channel symbol is carrying two bits now.

QPSK – a free lunch?

Now each channel symbol can be assigned two bits. We have doubled the rate at which we can send information for free!!!



As the number of points on the diagram increases there is less distance between each point. Hence any noise present will cause a greater error rate. The only way to avoid that is to increase the power of the signal.

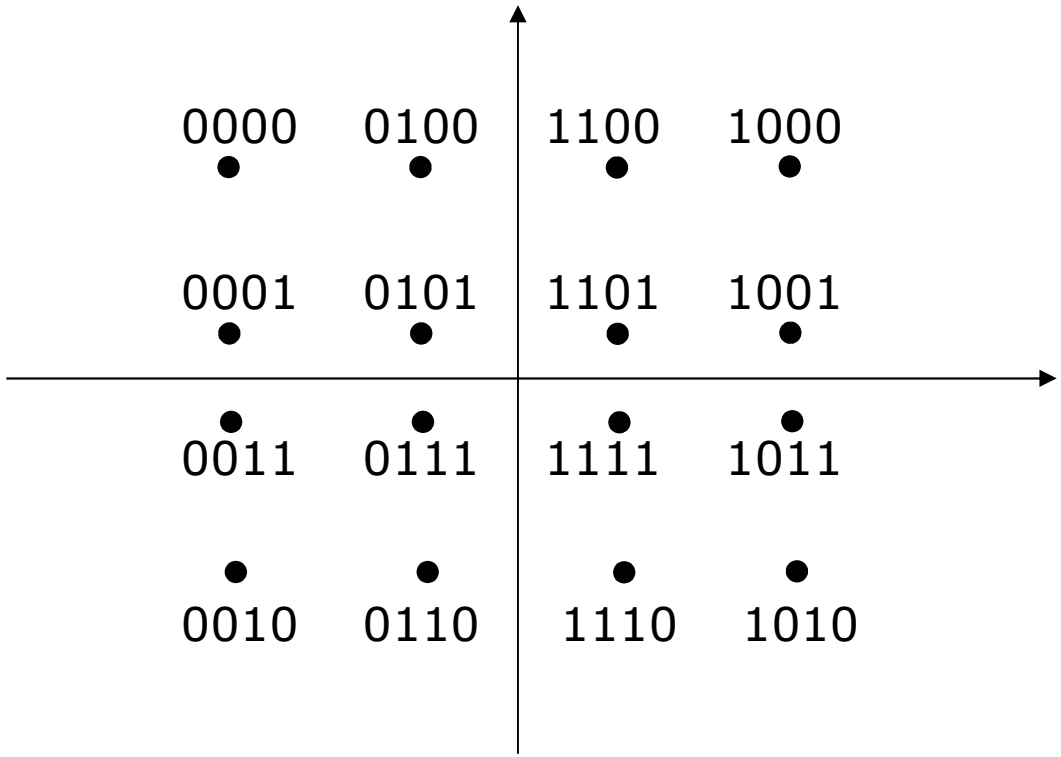
Higher order modulation schemes

Scheme	Bits in a symbol	Data rate increase	Relative Power Required (approx)	Data rate/Power efficiency
BPSK	1	100%	X 1	100%
QPSK	2	200%	X 2	100%
8PSK	3	300%	X 4	3/4 = 75%
16PSK	4	400%	X 8	4/8 = 50%
32PSK	5	500%	X 16	5/16 = 31%

We can use QPSK to double our data rate while keeping the same bandwidth usage but we have to double our power

Notice the ever decreasing power efficiency per bit…

Mixed Modulation Schemes

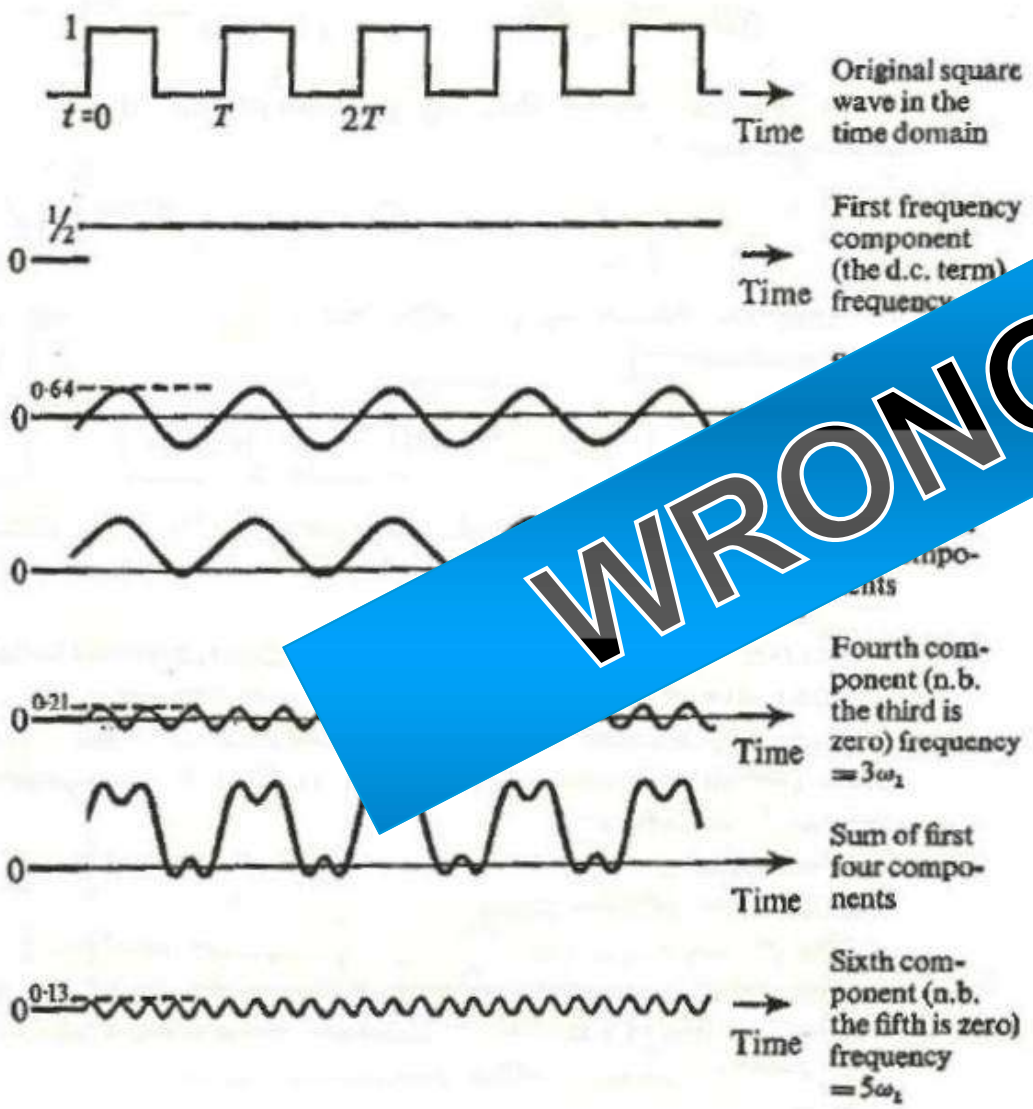


E.g. Rectangular QAM

Who is this man?

French revolutionary
Friend of Napoleon
Governor of Egypt after the battle of the Nile
Was famous for his mathematical description of heat flow
Discovered global warming

Fourier's brilliant insight



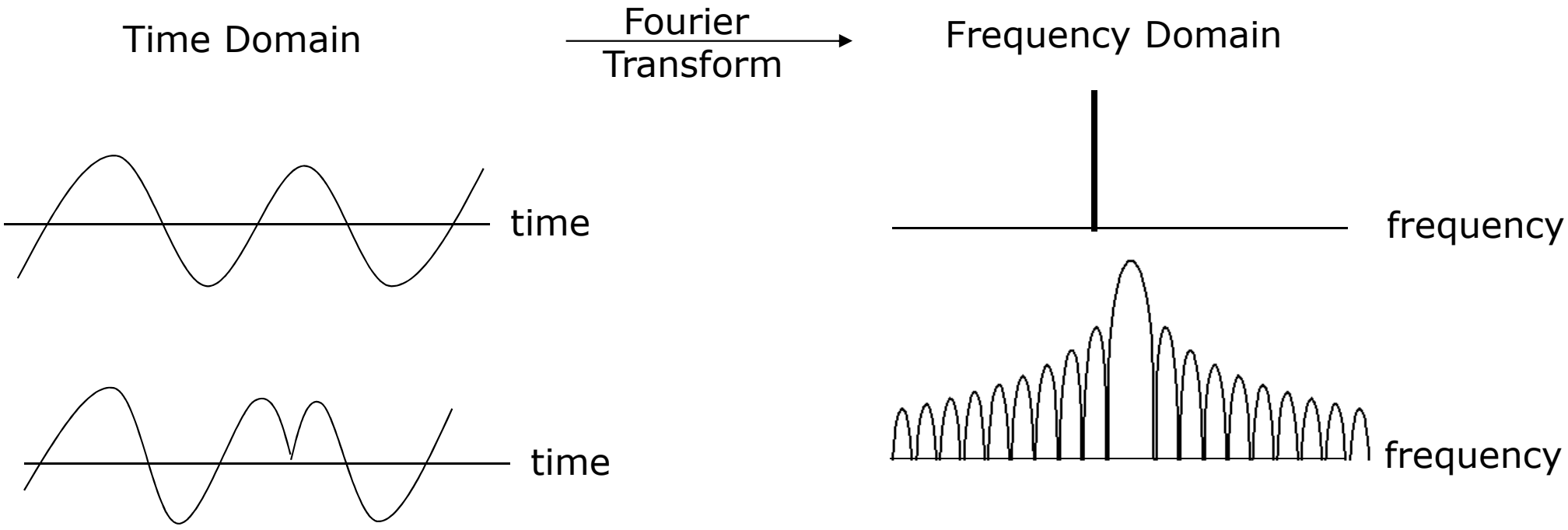
WRONG

Any periodic wave can be represented by the combination of a series of sine wave[illegible]f different amplitudes.

[illegible]acy of the representation [illegible]n how far we add "detail" by [illegible]g other waves.

The Fourier transform

Fourier gave us a way to mathematically manipulate a wave so that we can see it in terms of it's frequency components (i.e. all the sine waves that make it up) rather than the time domain.

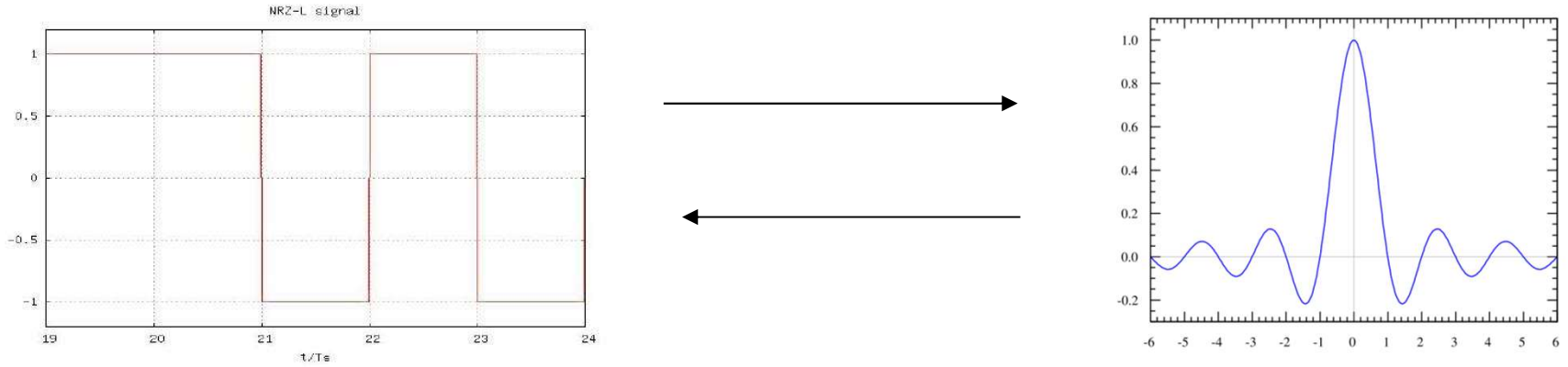


An amazing property of Fourier transforms is that for some waves (the ones we are interested in) if you do the exact same operation twice you arrive back at the same time series only reversed…..

Now this is weird and will be very useful later!

Square Wave Fourier Transform

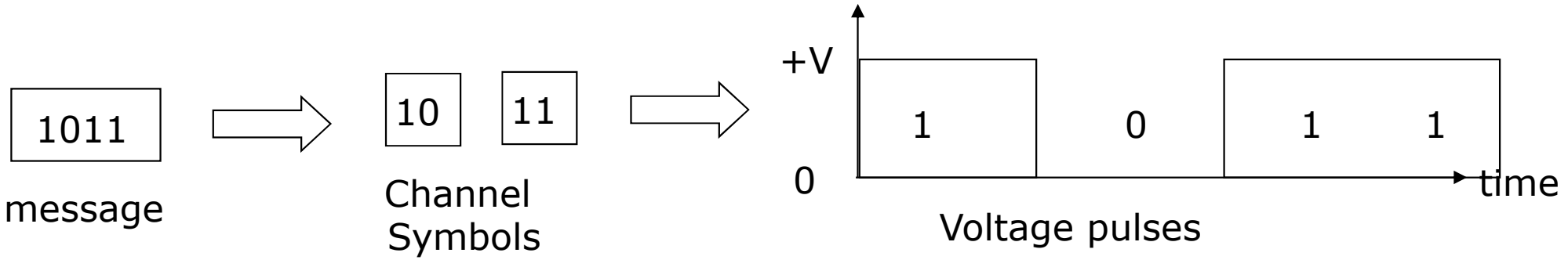
A very important Fourier transform….

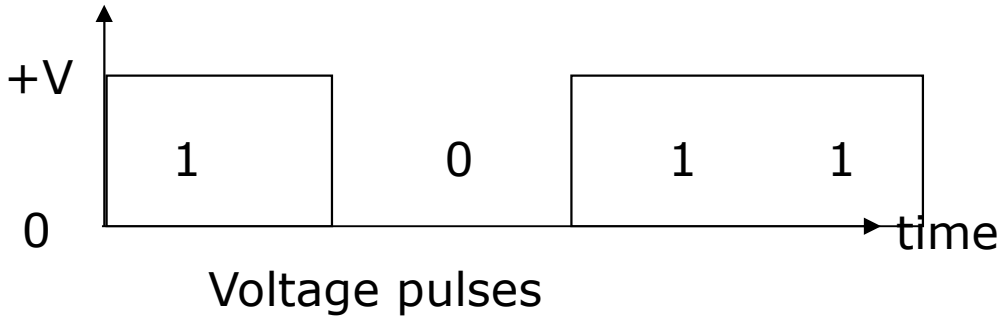


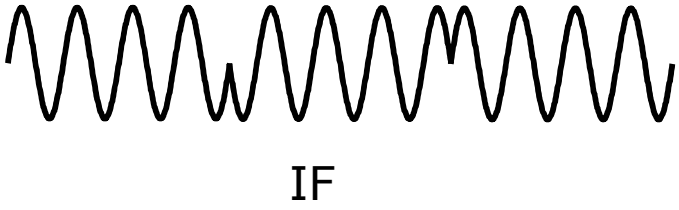
The Sinc function (sin (x)/x) above needs an infinite number of sine waves to represent it fully. That is it requires infinite bandwidth.

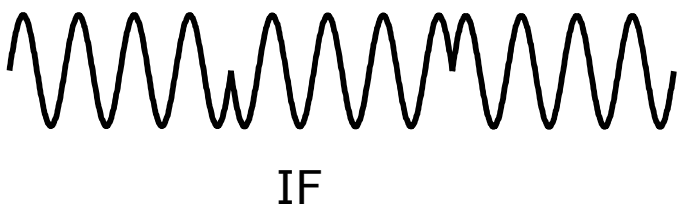
The amplitudes decay according to a cosine function, but it is important to think of a square wave as being bandwidth “hungry”.

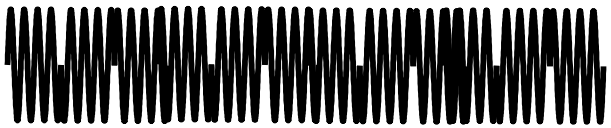
How do we imprint our message on to a microwave?

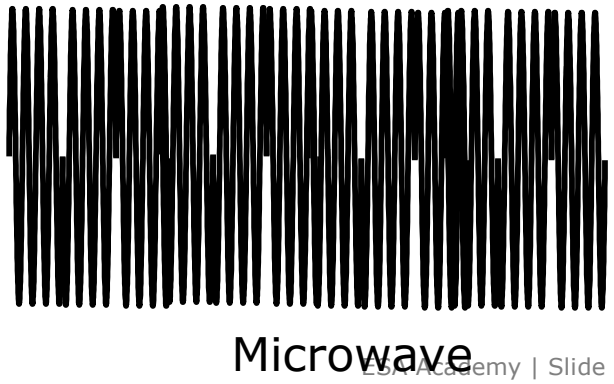








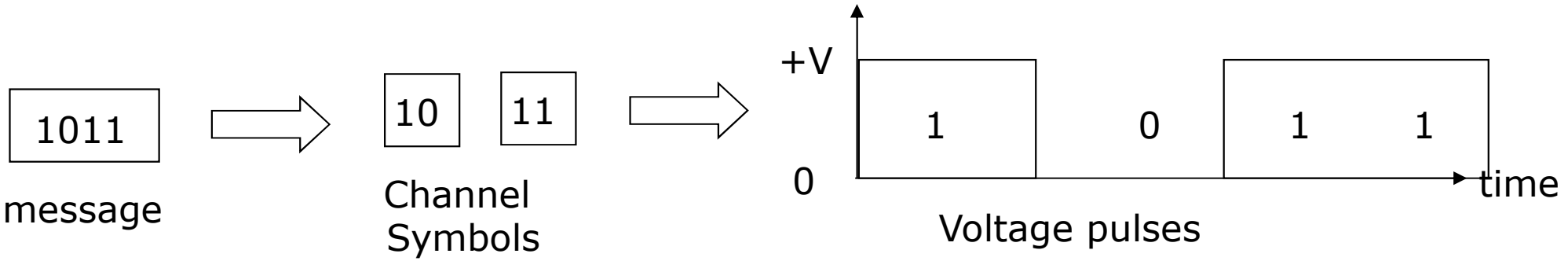




Pulse Code Modulation

Our message still exists as a set of zeros and ones in a computer's memory. So before we modulate a wave with it we have to turn it into a real physical signal.

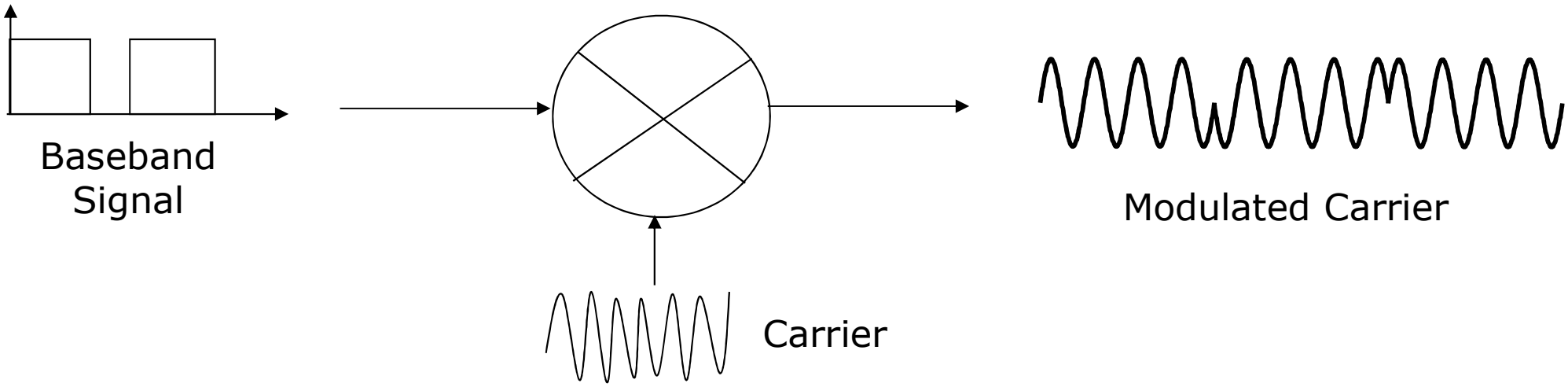
We do this by converting the digital channel symbols as a continuous string of voltage pulses. This process is called pulse code modulation (PCM)



The signal that is produced is called the **BASEBAND** signal

Carrier Modulation

We then take the baseband signal and mix it with a sinusoidal signal called the carrier to produce a modulated carrier signal.



STOP!

It is not that easy

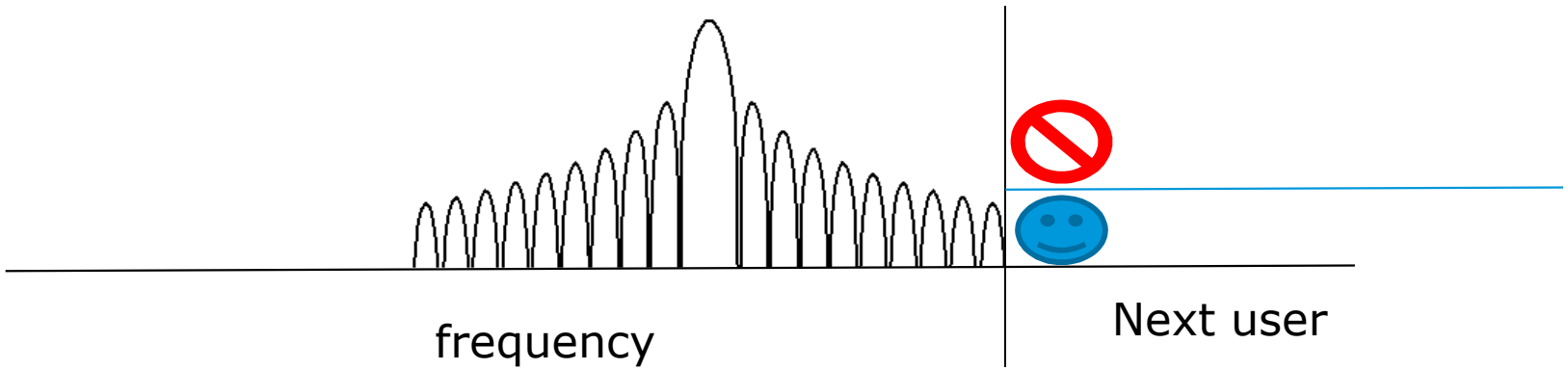
You are using too much bandwidth

Why is bandwidth so important?

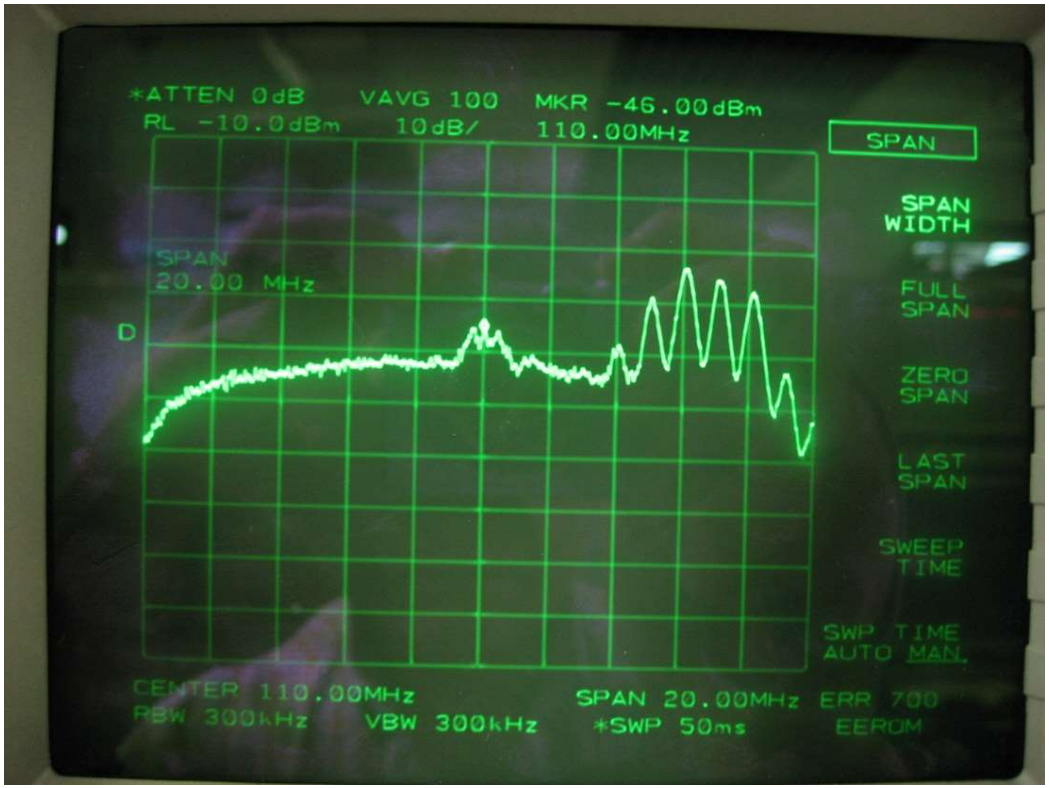
Remember we must "stay within our allocated bandwidth" to avoid interference with other users. Typically our signal must lose 99% of it's power (-20dB) before it enters an adjacent allocation.

Luckily, the power spectral density of a given function of time is the Fourier transform of its autocorrelation function.

This provides us with a method of calculating the bandwidth impact of our signal.



Fourier transforms and the associated tools (e.g. spectrum analysers) give us the tools to design communication links taking into account their bandwidth impact. Without them it would be impossible.

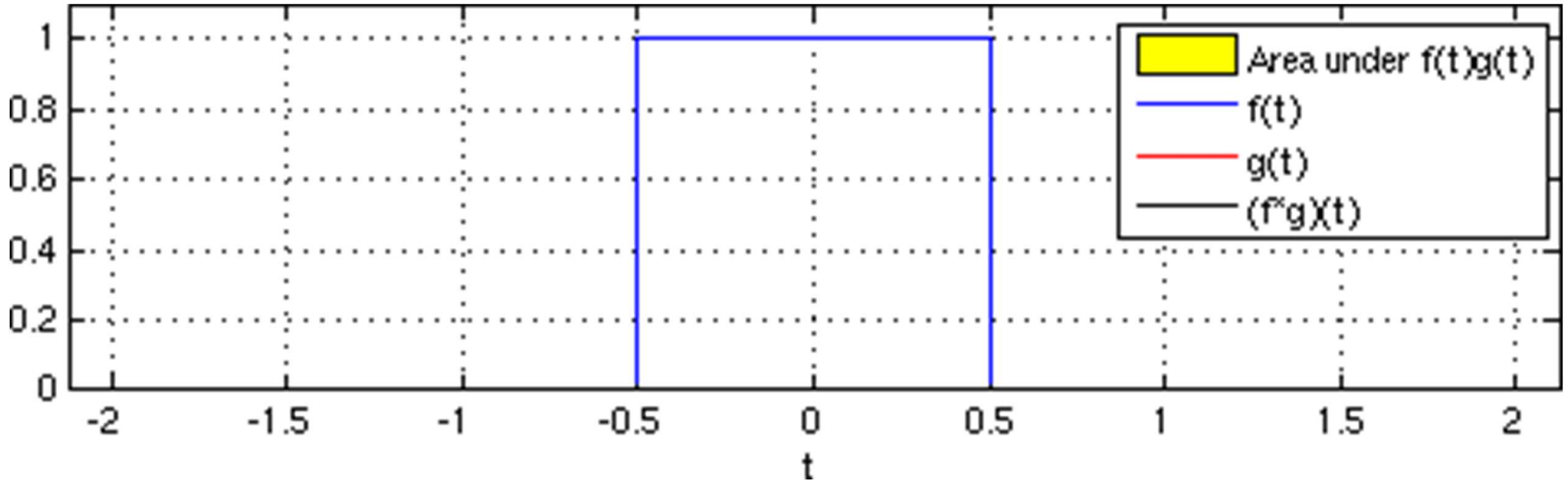


Courtesy of Enrico Vassalo ESA/ESOC

Managing the bandwidth

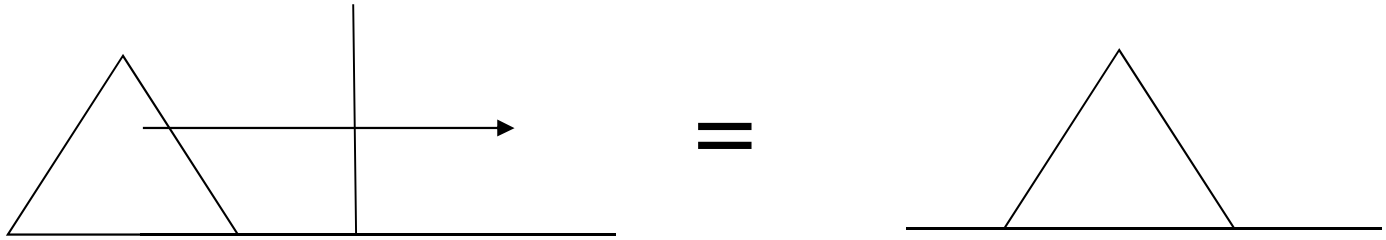
A useful property of Fourier Transforms is that multiplying two signals in the time domain (mixing) is equivalent to convolving them in the frequency domain.

Convolution can be seen as "integral of the product of the two functions after one is reversed and shifted"



Managing the bandwidth

As the carrier is a sine wave (a straight line in the frequency domain), the modulated signal will have the same shape in the frequency domain as the baseband signal

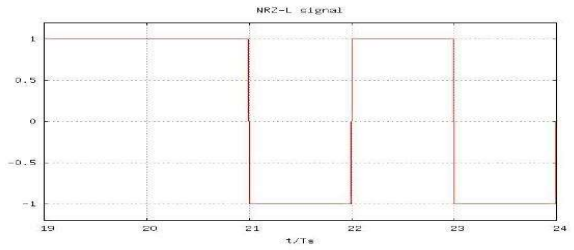


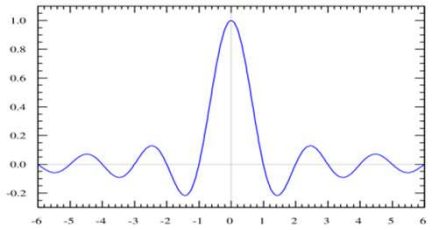
Therefore the amount of bandwidth the baseband signal spectrum is crucial for the overall bandwidth usage.

We can influence this by controlling the shape of the voltage pulses.

Square is not ideal for bandwidth usage

The most obvious choice of pulse shape to choose is SQUARE. However as shown previously it produces a sinc frequency profile which requires infinite bandwidth.



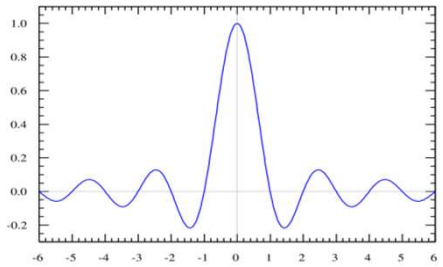


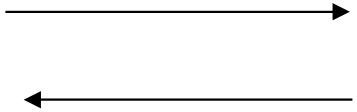
Now we are getting somewhere

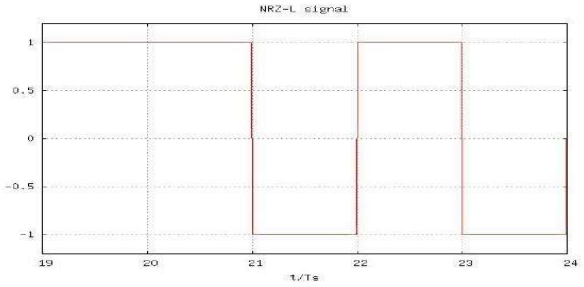
Quiz: What is the ideal pulse shape then?

As we mentioned, Fourier transforms work in both directions. Therefore if we use a sinc pulse shape we would have a nice square profile in the frequency domain!

Nyquist showed that ideal pulse shapes like this would require only half the information rate as bandwidth requirement.







Unfortunately they are impossible to make using present filters.

Raised cosines and Roll-off

Special pulse shapes such as raised cosines are made by using filters, see later.

Schemes that use them are sometimes called Filtered PSK.

The measure of their performance compared to the ideal sinc shape is given by a parameter called roll-off measured between one and zero. Zero roll off it is the ideal shape. Typical values would be 0.2 – 0.5.

QUIZ: why don't we use raised cosines or zero roll-off all the time?

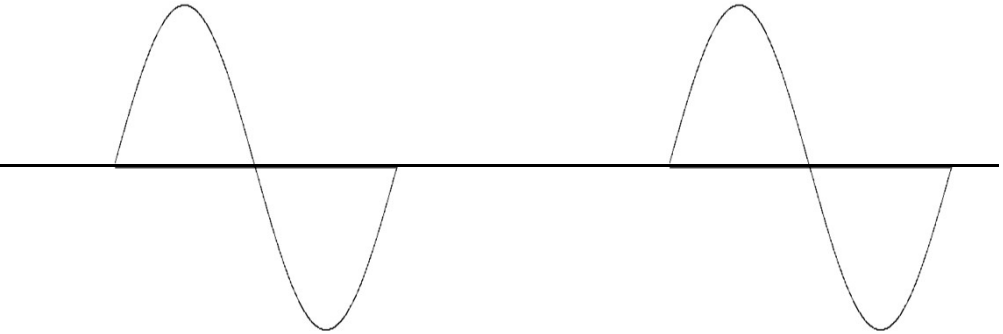
Because it is a trade-off between ease of detection of the symbol changes and bandwidth usage.

With raised cosine pulse shaping and very low roll-off it is very difficult to detect the symbol boundaries as they are smooth. For square waves it is much easier.

There is no free lunch!

Square and Sine Baseband Signals

So, practical waveforms are squares, sines or special shapes such as raised cosines.



Power decays according to **Bessel functions (J_0, J_1 etc)** i.e. uses less bandwidth than square waves.

Less power in the signal and more in the carrier. Harder to detect transitions.

Used on subcarriers for near earth space missions (see later)

Power decays according to **cosine functions** i.e. uses more bandwidth than sine waves.

More power in the signal and less in the carrier. Easier to detect transitions.

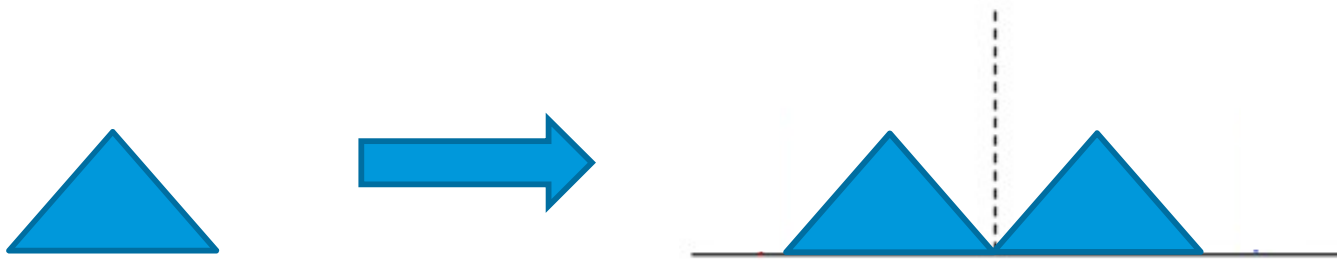
Used on subcarriers for deep space missions and some direct carrier modulation (see later).

Bandwidth calculations

(Nyquist) Ideal bandwidth = ½ x Channel Symbol rate

but this is for the baseband signal only.

When we modulate this onto a higher frequency carrier signal the range of frequencies it occupies doubles (reflected about the carrier frequency)



Ideal bandwidth after modulation = ½ x Channel Symbol rate X 2

Ideal bandwidth after modulation = Channel Symbol rate

Bandwidth calculations without coding

(Nyquist) Ideal bandwidth = Channel Symbol rate

but this bandwidth is needed to carry the symbols only… in order to reduce to -20dB power by the next user we have to apply the roll off…

Occupied bandwidth = (1 + roll off) x Channel Symbol rate

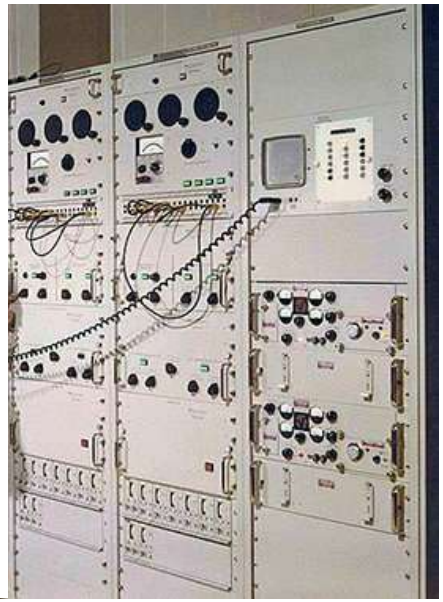
Without channel coding we can say

$$\text{Channel Symbol rate} = \frac{\text{information rate}}{\text{number of bits per symbol}}$$

QUIZ: Information rate is 1 MBps, QPSK modulation, no channel coding and rolloff is 35%
What is my Occupied Bandwidth approximately?

Analogue to Digital at baseband/modulation level

THEN



NOW

Usually the modulated wave that is produced after filtering and modulation is not at the final microwave frequency. The higher the frequency the harder it is to deal with..

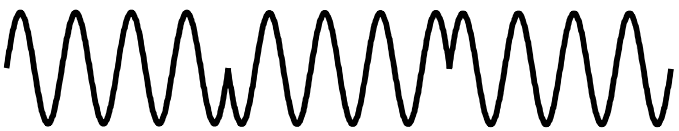
However with faster and faster digital machines it can now be done especially at S band – URSPs and SDRs in SMILE.

Intermediate to Microwave Frequency

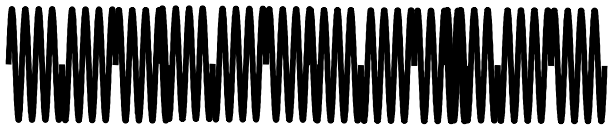
The frequency chosen is called the **Intermediate Frequency** (it is usually 70MHz). ESTRACK standard.

Once the Intermediate Frequency has been created it needs to be up-converted (simple frequency multiplication) to the desired frequency, amplified and then released through the antenna.

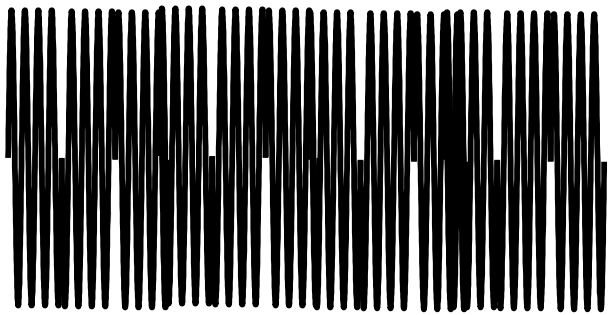
IF



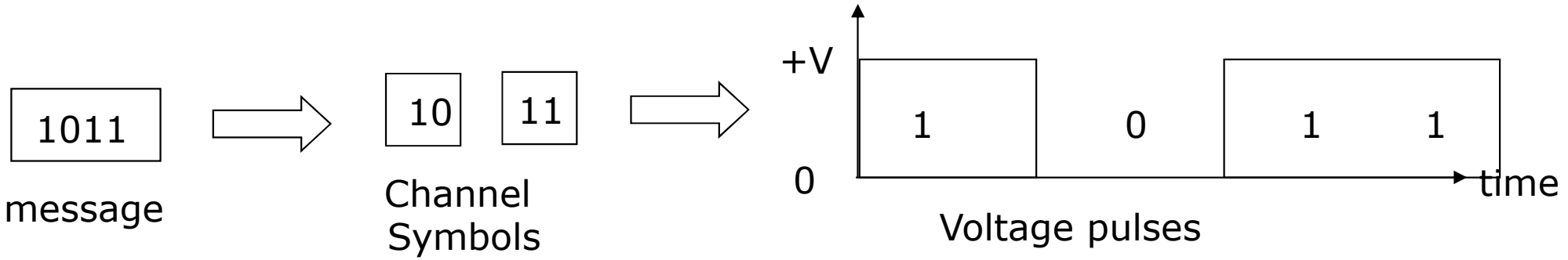
Up conversion

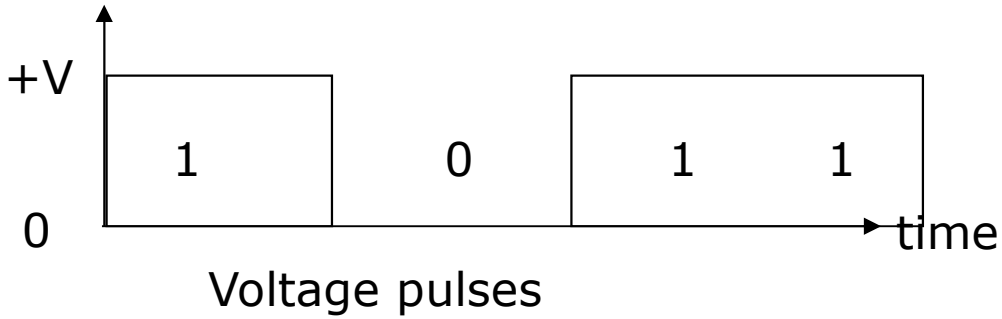


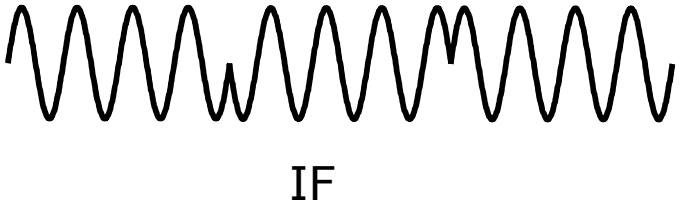
Amplification

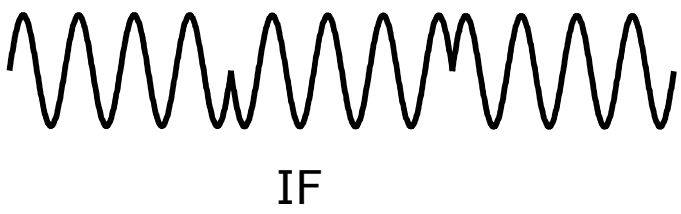


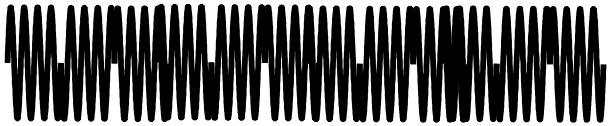
How do we imprint our message on to a microwave - recap

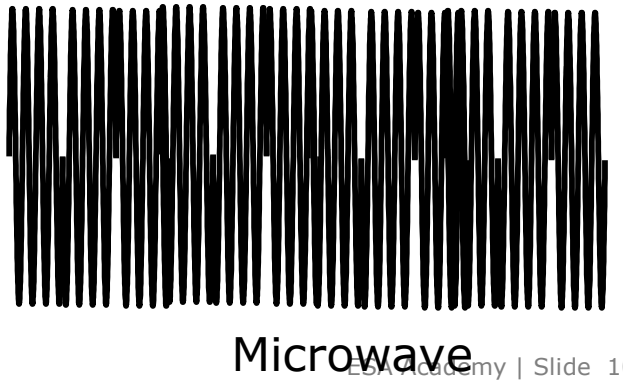






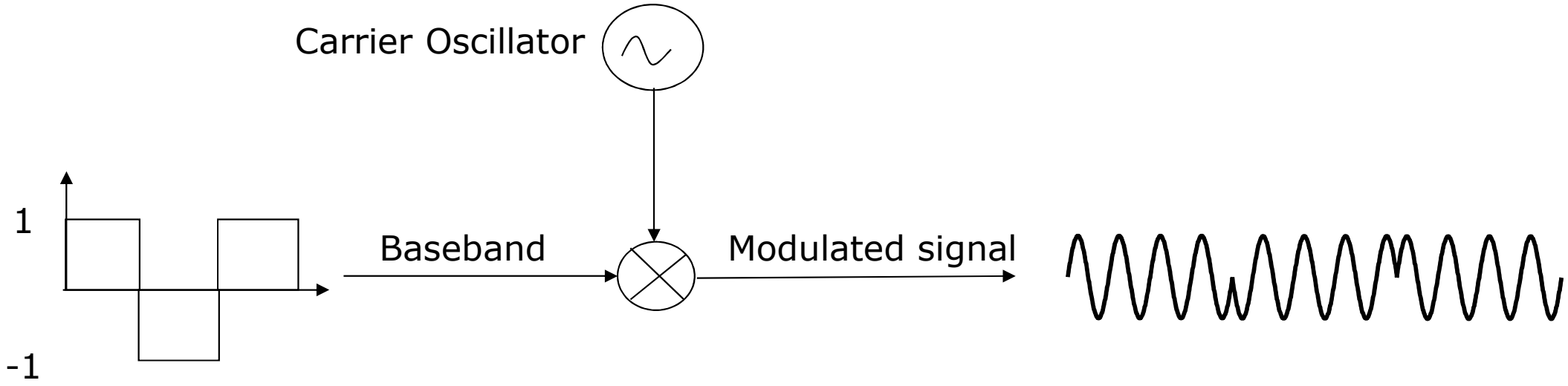






The real world: BPSK

If we want to produce a BPSK signal with a phase difference of 180 deg i.e. modulation index of +/- ∏/2 rad it is simple. We mix a bipolar (i.e. +/- 1) baseband signal with the carrier.



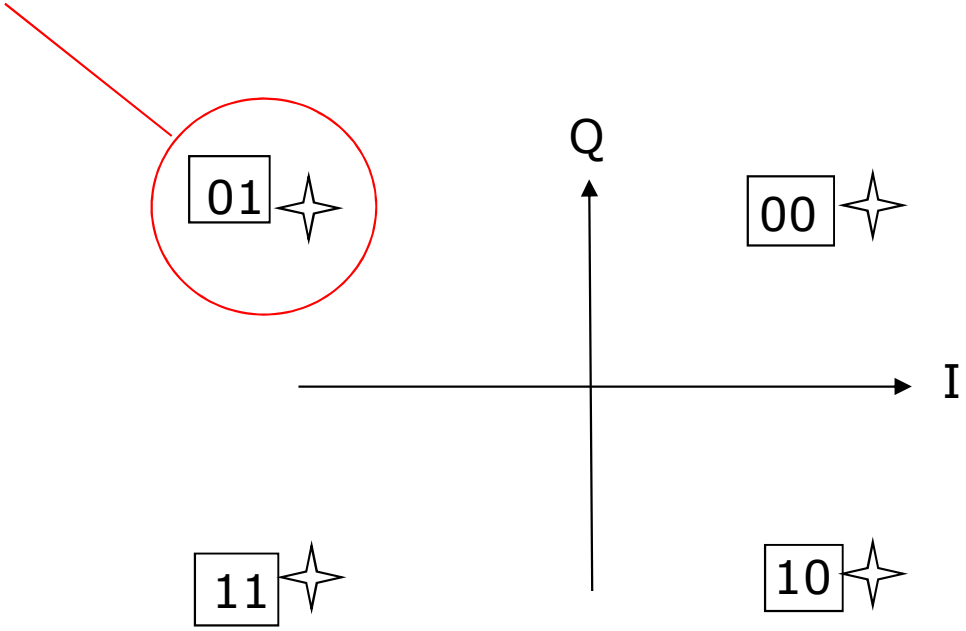
If we want to produce a BPSK signal with a phase difference of less that 180 deg i.e. modulation index of less than +/- ∏/2 rad, the mixer has to be replaced by a phase modulator which is set up with the required modulation index.

The real world: QPSK

QPSK is more difficult. We have to put several elements together to produce a QPSK wave. Basically we split the phase diagram into two axes. The In-Phase (I) axis and the Out of Phase (Q) axis.

When two bits arrive, we look up the corresponding channel symbol and then work out the corresponding values of I and Q.

E.g. "01" is I = -1, Q = 1.

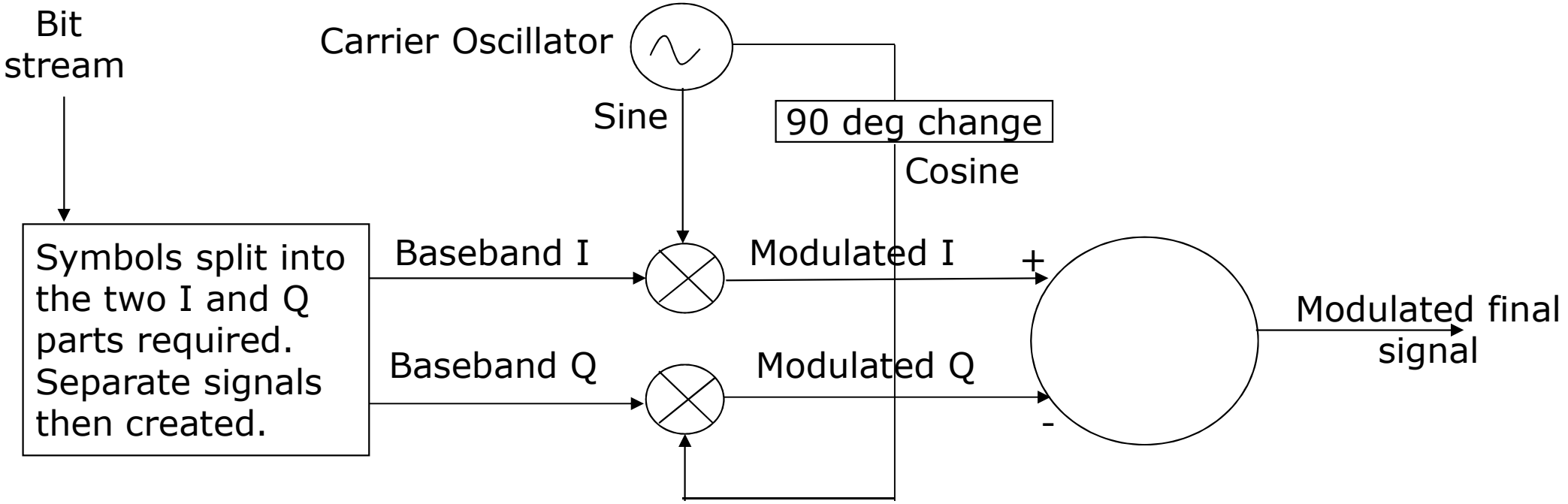


The real world: QPSK

1. The I and Q values are used to produce different baseband signals (pulses).

2. The I and Q baseband signals are then **mixed** with a carrier sine or cosine carrier wave (respectively)

3. They are then **added together** to produce the final signal.

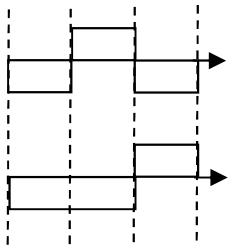


Note QPSK production uses mixers only as the modulation index is always +/- Π/2 rad

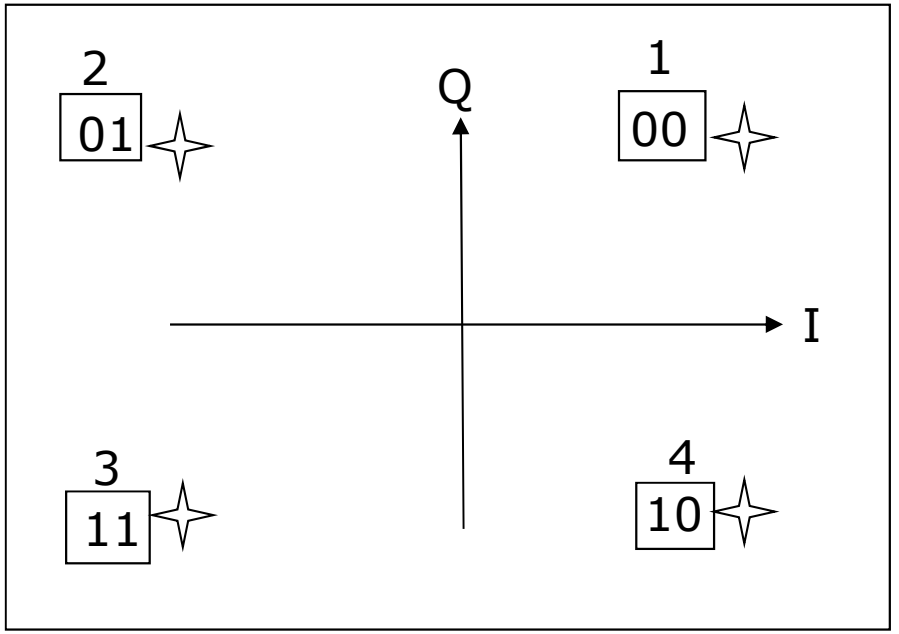
QPSK production

Bits = 111001
Channel Symbols = 3(11),4(10),2(01)

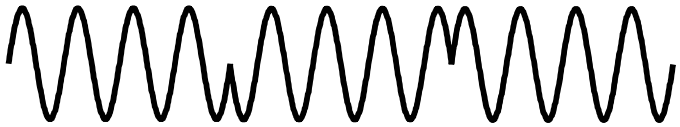
I baseband = -1,1,-1 =



Q baseband = -1,-1,1 =

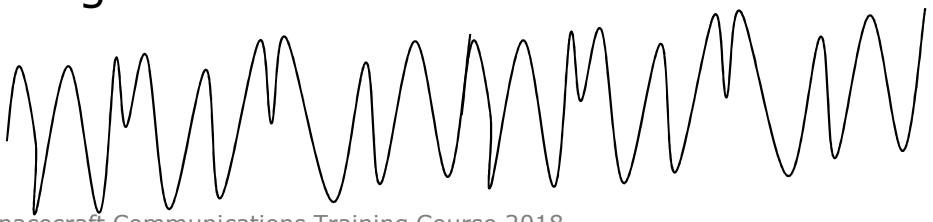


Modulated I stream =



Modulated Q stream =

Final signal after addition



Demodulation

Now we have modulated our message onto a microwave and sent it on its journey across space,

When it arrives we must reverse the process to retrieve the message.

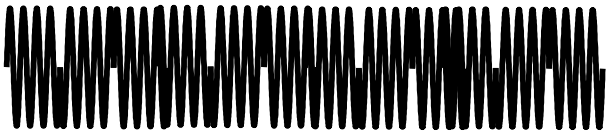
This process is called

DEMODULATION

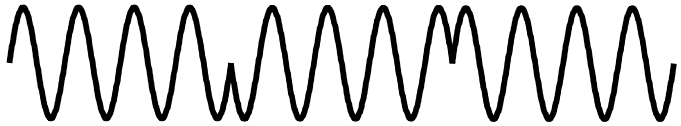
At first glance it looks easy….

Demodulation in principle

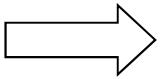
Low Noise Amplification

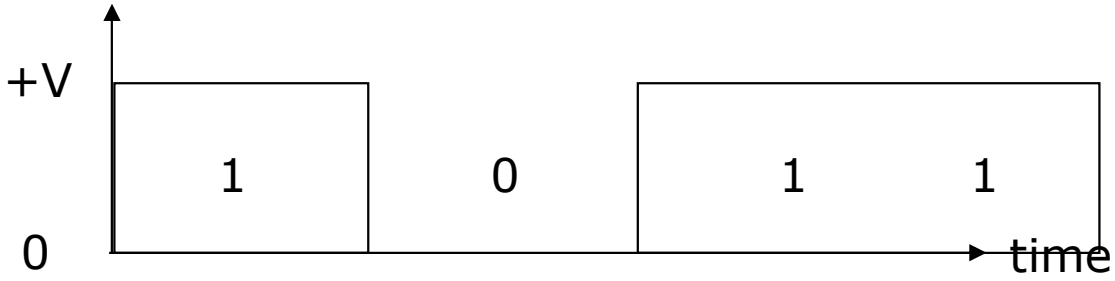


Down conversion



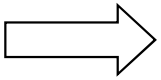
IF

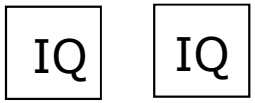




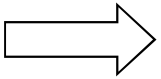
Voltage pulses

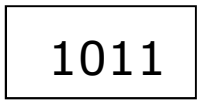
(Actually NRZ and one for I and for Q)





Channel Symbols





message

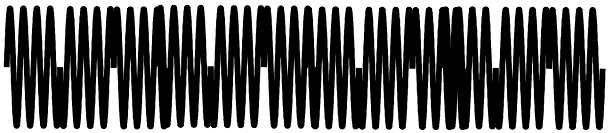
LNA and Down conversion

The low noise amplification of the incoming signal and it's down conversion to from microwave frequency to the intermediate frequency (IF) is performed by the front-end systems of the ground station, see ground station session.

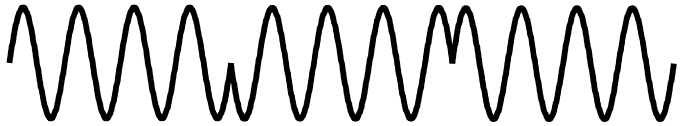
The IF signal produced must have sufficient bandwidth to cover any deviations from the original transmitted signal due to Doppler shift etc.

One small problem…

Low Noise Amplification

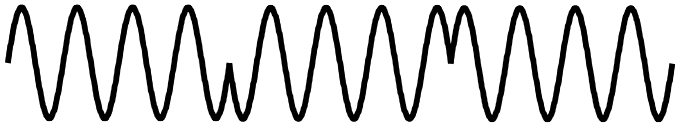


Down conversion



IF

The recovered IF signal does not look like this



It looks more like this

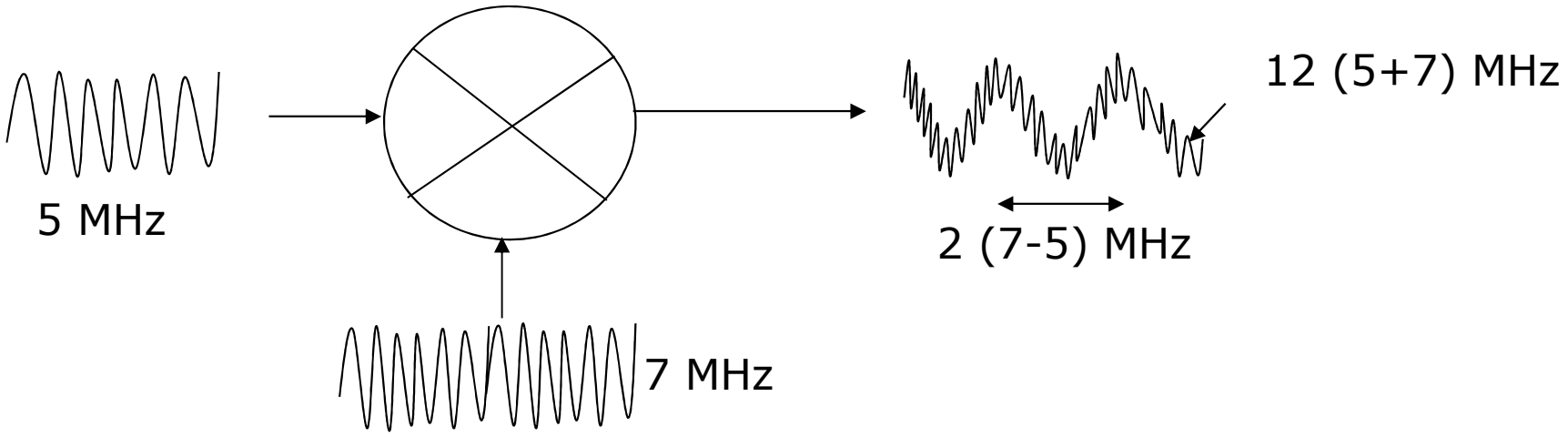
So how do we retrieve the information from that!

By multiplying signals

When we multiply signals together the resulting signal can be expressed as the sum of two distinct signals.

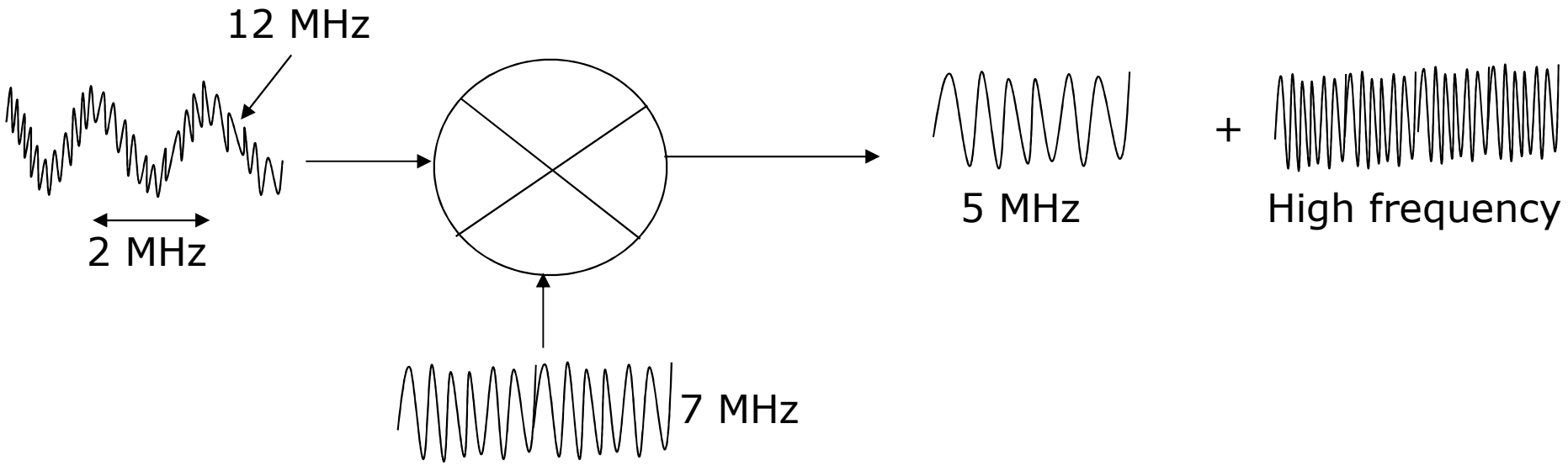
A high frequency signal (sum of the input frequencies)

A low frequency signal (the difference in the frequencies).



Multiplying the signal again by the same carrier

Now let's see what happens in that last example when we take the output signal and multiply again by the same 7Mhz input.



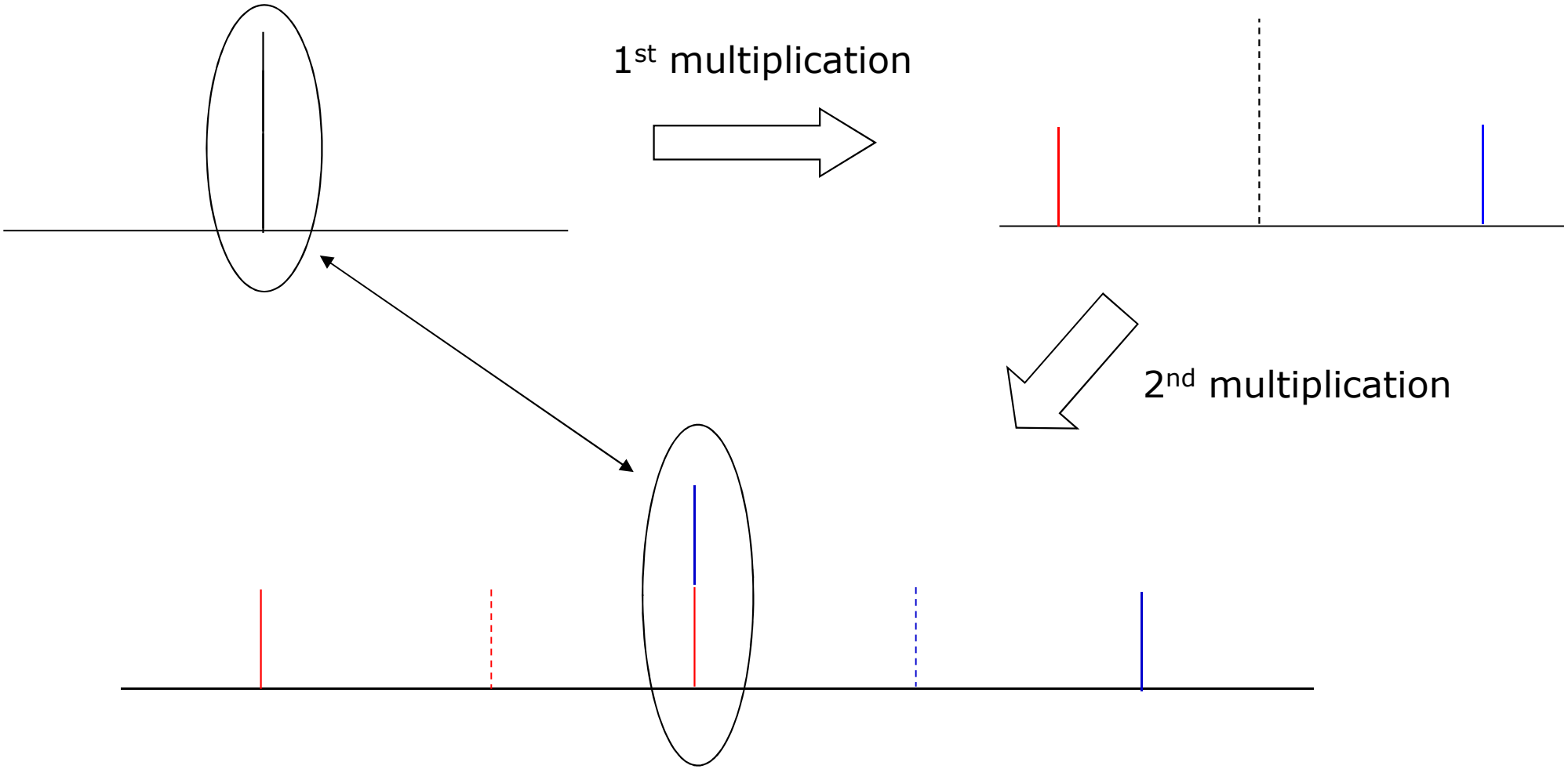
Note there are

1. High frequency outputs at 17 MHz (7+12) and 9 KHz (7+2)
2. And a low frequency output at 5 MHz (7-2)

The low frequency output is actually the original input wave. Isn't that amazing?

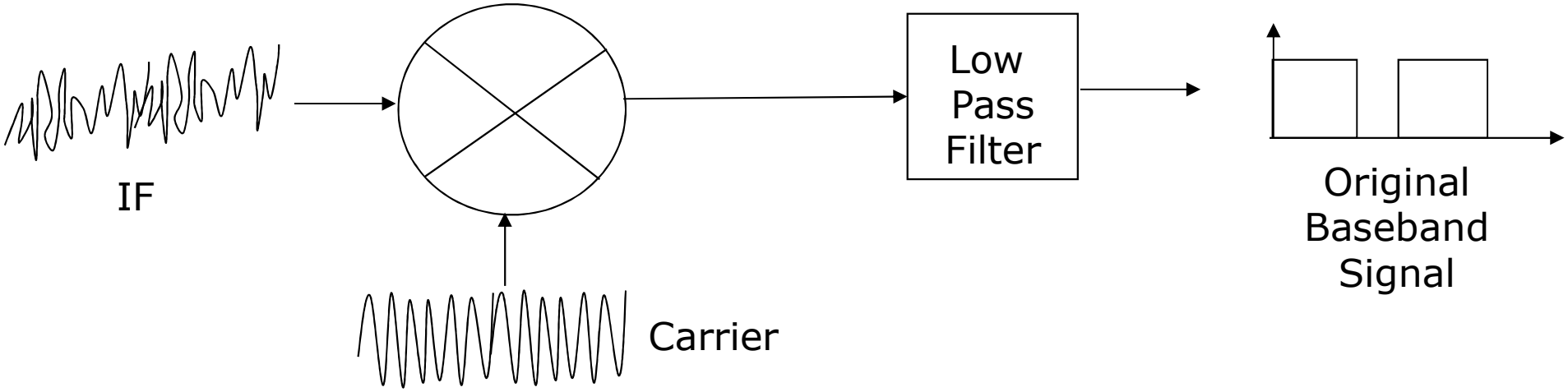
Multiplying the signal again by the same carrier

This is clearer to see in the frequency domain



Recovering the baseband signal

So it follows that if we multiply the IF input by the a signal that is identical to the carrier that produced it and low pass filter the output, we get the original baseband signal again.



If we can regenerate the carrier signal we can use it to extract our baseband signal from the IF signal

BUT we need to exactly regenerate the original carrier somehow

The process of finding the real carrier frequency and phase is called carrier recovery.