The ladybird guides

Book II

Day 3 - demodulation

Day 3: De-modulation

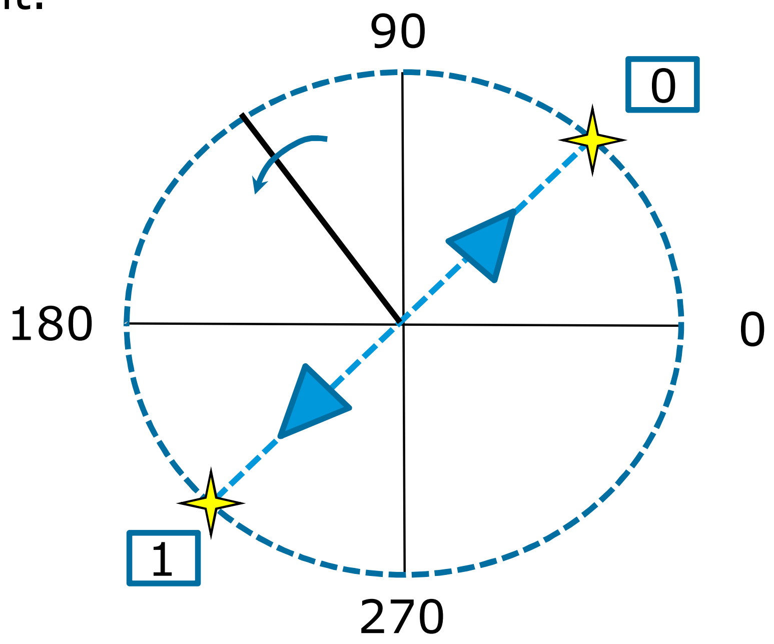
Objective:
To appreciate the difference between the modulation and demodulation
To introduce the concepts of signal, noise and error probabilities

1. To appreciate the difficultly in recovering a baseband signal
2. Recovering the baseband signal
3. Phase Locked Loops
4. Subcarriers
5. Costa’s Loops
6. Difference between uplink and downlink
7. Signal to noise ratio and the probability of error
8. The importance of timing and bit sync

The carrier recovery problem

The frequency and phase of the original wave (carrier) that was phase modulated to produce the IF signal will have been modified slightly by all the factors discussed earlier, especially Doppler and an imperfect spacecraft local oscillator ("jitter").

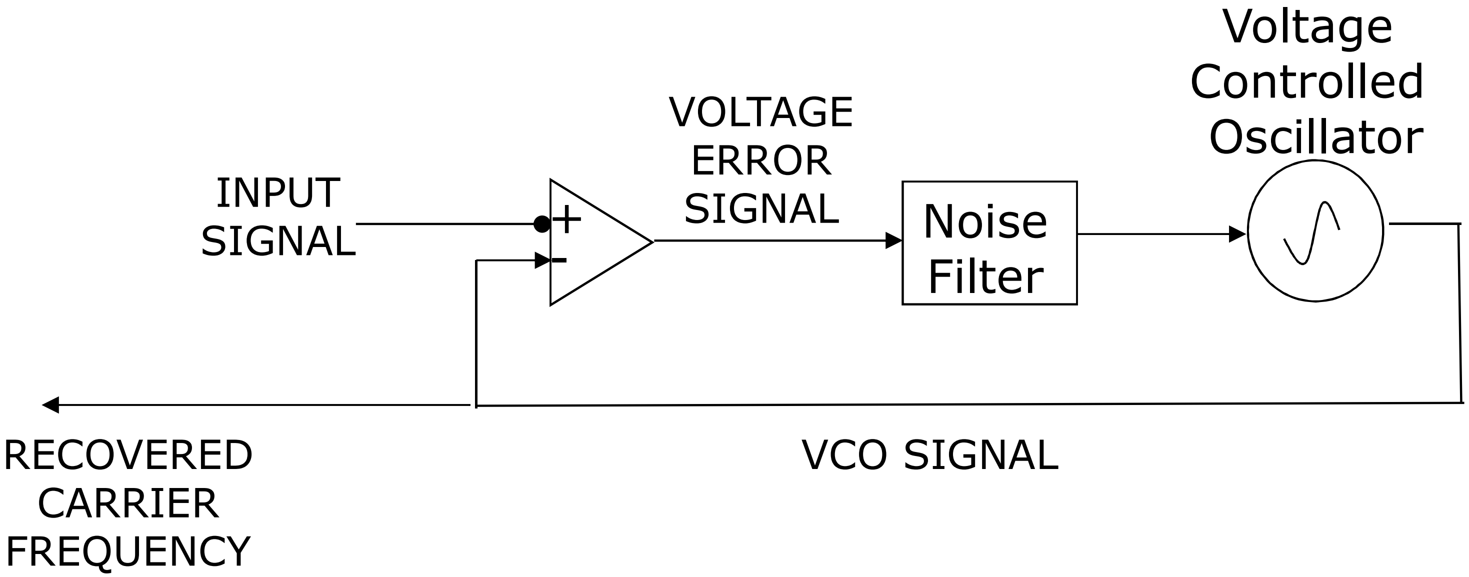
What happens if we just try to guess the carrier frequency and phase?
We will introduce a phase offset between the recovered baseband signal and the transmitted one. This shows up in the constellation diagram of the receiver and we will incorrectly demodulate it.



Note: There are schemes that do not need to recover the carrier (Asynchronous Down Conversion) e.g. DBPSK…

The Phase Locked Loop (PLL)

Recovering the carrier frequency can be done using a phase-locked loop (PLL) circuit. Phase and frequency are connected so if we keep two signals in phase they must be at the same frequency.



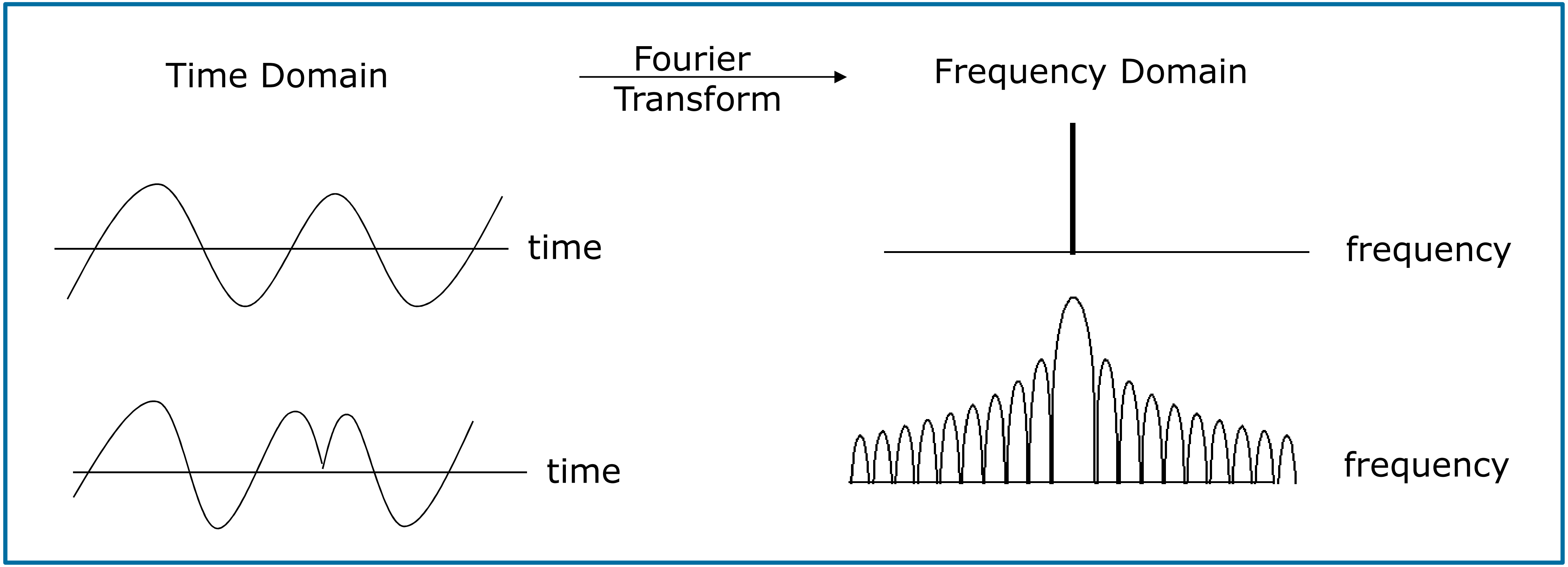
When the loop has converged and the voltage error signal is below a certain threshold we say that we have CARRIER LOCK.

QUIZ: What is the input signal?

How can the PLL detect the carrier phase?

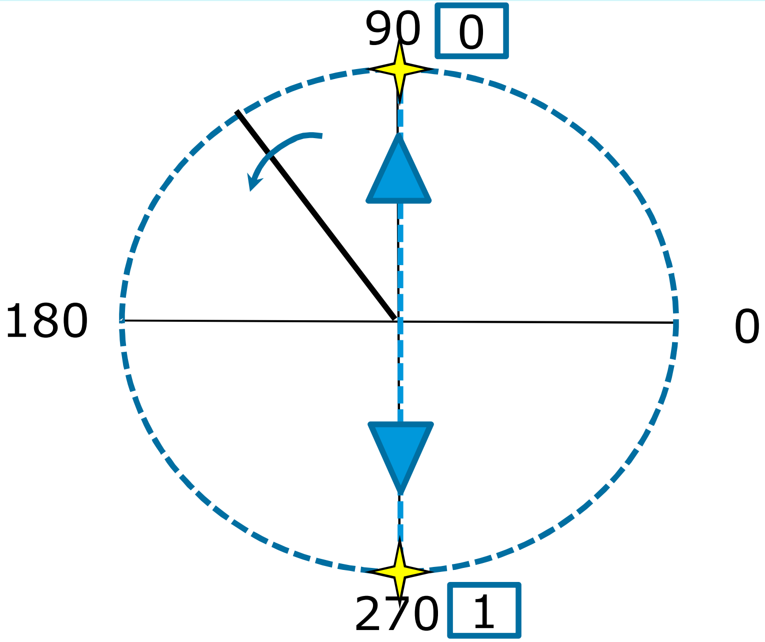
We need to make the carrier signal obvious compared to the rest of the signal so that it can be tracked by a PLL.

Ensure the carrier signal is stronger than the data signal and very sharp.

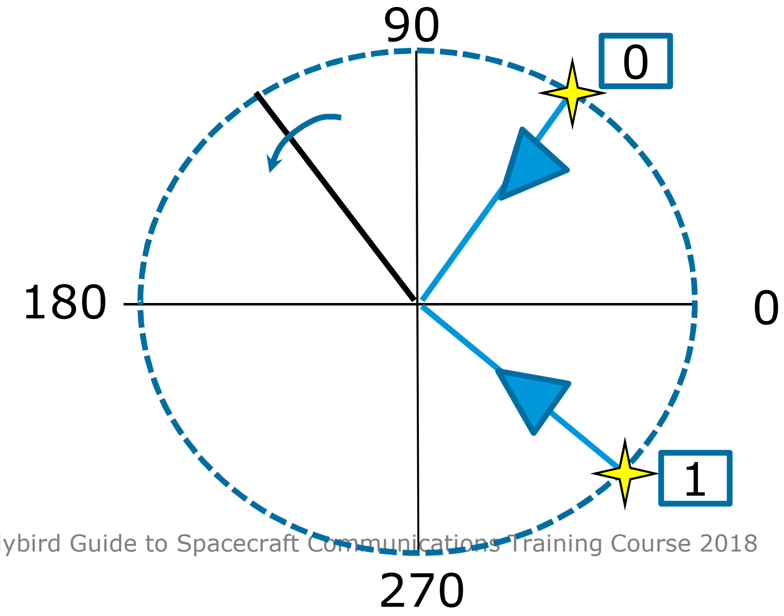


So we have a problem….
Remember the modulation index?
The amount of power in the carrier compared to the rest of the signal can be altered by changing the modulation index.

BPSK and modulation index recap

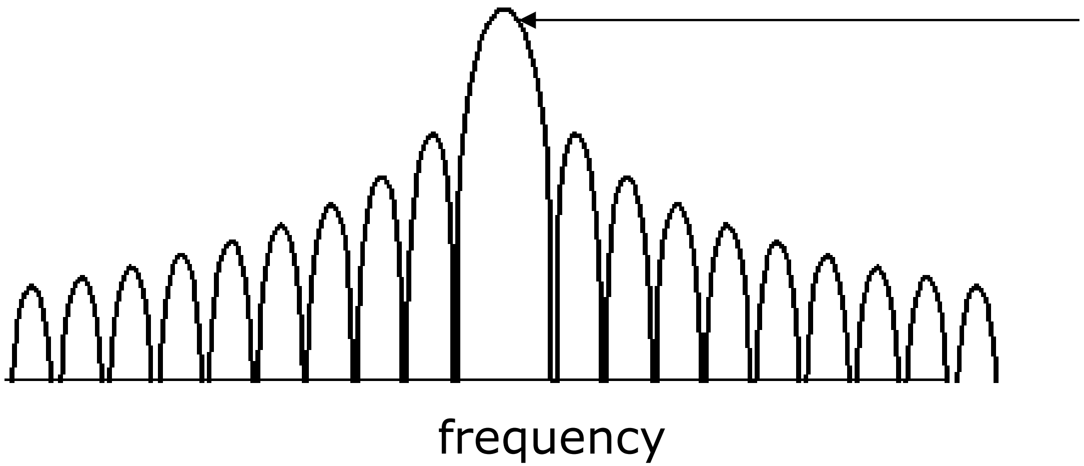


A modulation index of +/- ∏/2 rad

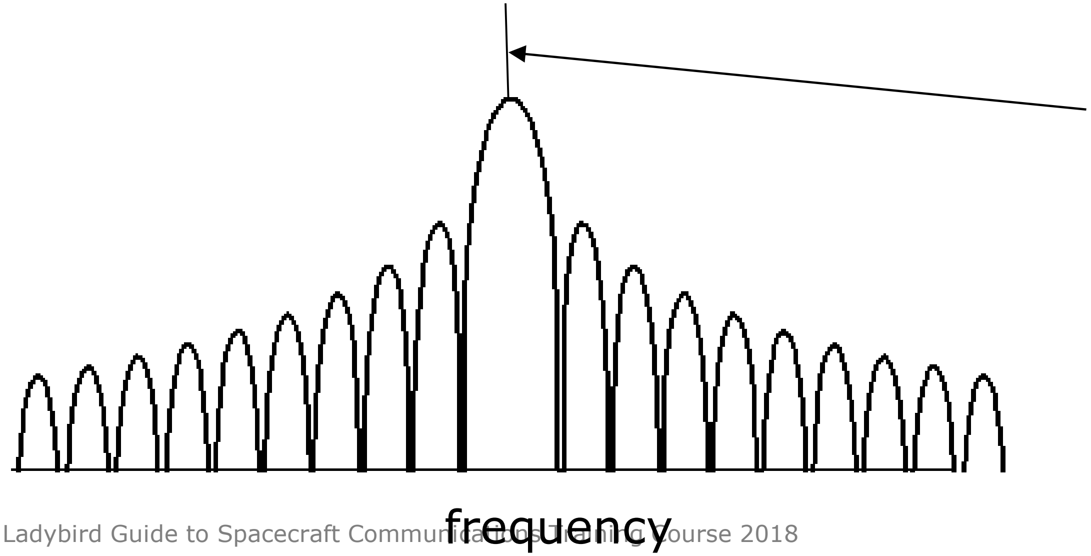


A modulation index of less than +/- ∏/2 rad.

BPSK, modulation index and carriers



Using a modulation index of +/- ∏/2 rad leaves no recognisable carrier signal.

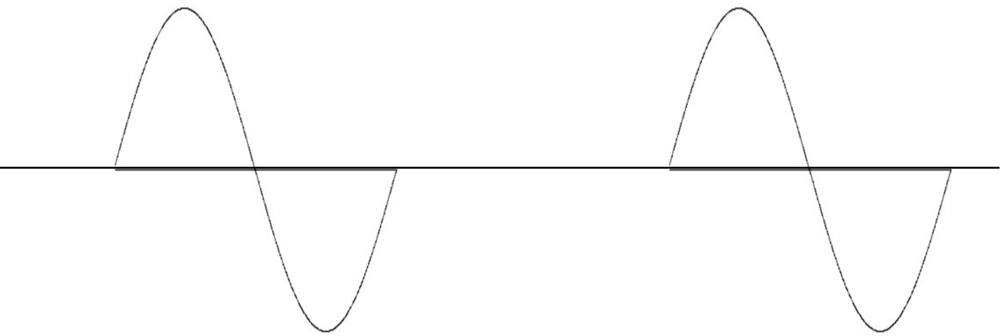


Reducing the modulation index below +/- ∏/2 rad has an important effect. It leaves a recognisable carrier signal.

A modulation index of +/- 1 rad is often used.

Square and Sine Baseband Signals

The amount of power in the carrier compared to the rest of the signal can also be altered due to the PULSE shape chosen.



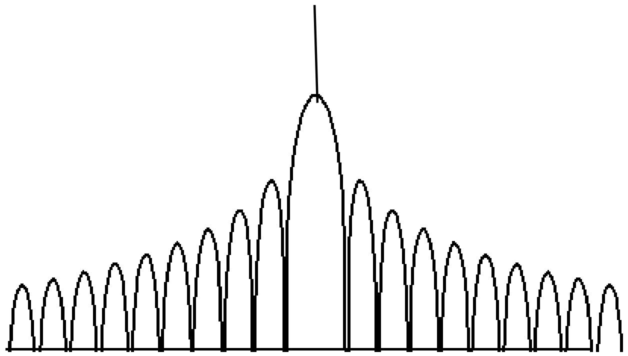
Pcarrier = total power x J^2_0 (Mod index)

Where J_0 is a Bessel function

Pcarrier = total power x $\cos^2$ (Mod index)

QUIZ: Which shape do you think leaves more power in the carrier compared to data?

BPSK, modulation index and carriers



Great, we now have a carrier signal that is stronger than the data signal and very sharp.

The PLL loop can recognise this and lock the local ground oscillator frequency to it.

This is called a RESIDUAL carrier scheme.

QUIZ: Have we answered the question about the PLL input signal?

PLL bandwidth

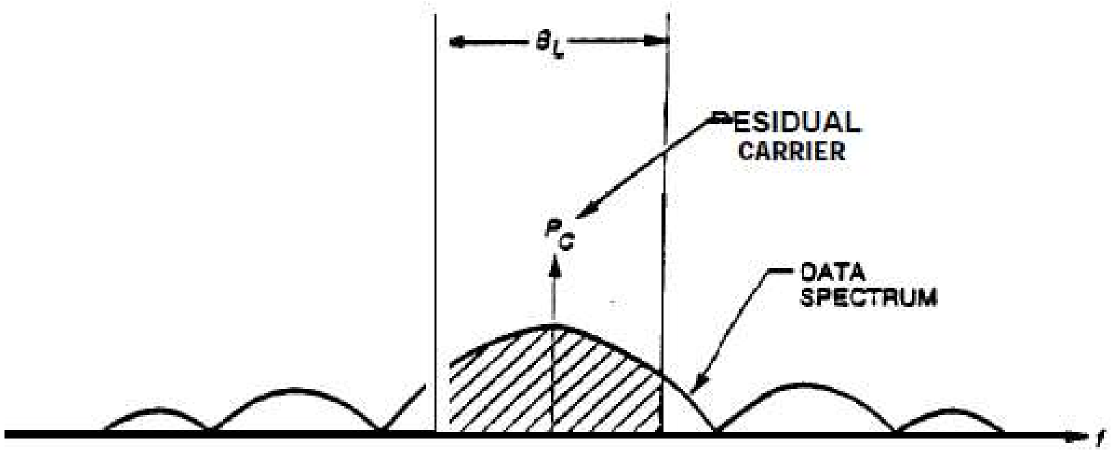
A very important parameter for the PLL loop is the loop bandwidth. This is defined as the frequency range in which the PLL can maintain lock.

As shown below in this frequency range, as well as the carrier signal there will be normal noise and noise caused by the data.

The carrier signal needs to be a certain amount of times stronger than the combined noise in the PLL bandwidth otherwise the PLL loop cannot track it.

The usual values are x 10 (10dB) for simple tracking purposes (e.g. Doppler) or TC recovery and **x 50 (i.e. 17dB)** if we want to use the signal for telemetry recovery.

PLL bandwidth



Never a free lunch - again

This leads to a very important trade-off.

The smaller the bandwidth, the less noise in the PLL loop BUT the less robust the loop. That is it is easy to lose lock if the Doppler change is too high, or the jitter is large etc.

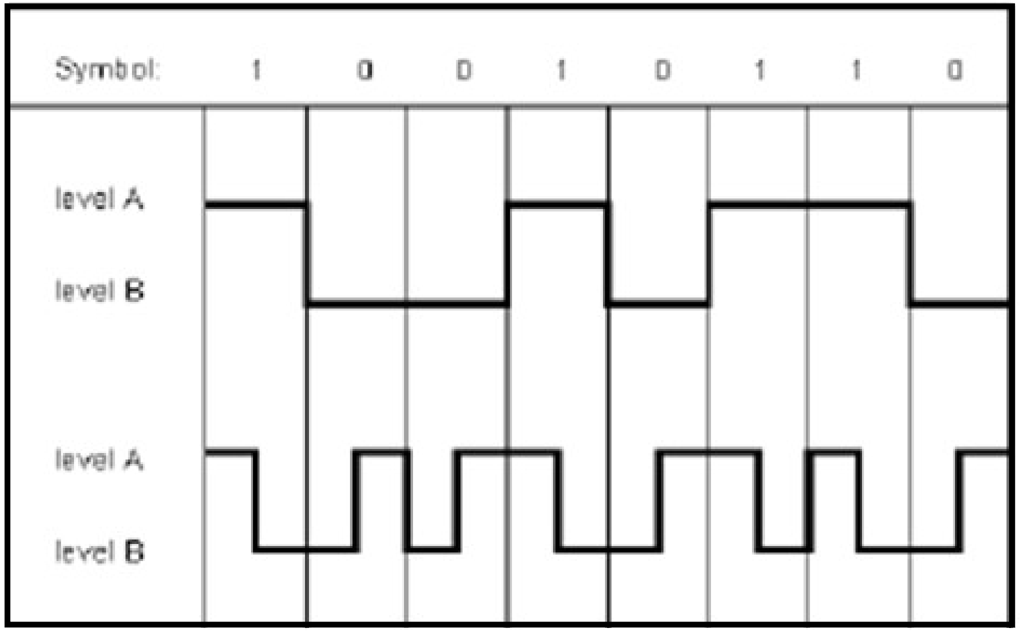
Remember: Make the PLL bandwidth small enough to be able to recover the carrier but do not make it any smaller than absolutely necessary.

It is possible to improve the robustness of the loop by feeding it with Doppler predictions (from Flight Dynamic calculations) or by using more complicated (second or third order loops). The advantage is that the size of the PLL bandwidth can then reduced and we recover the carrier frequency/phase with less signal power.

Also note that some spacecraft transponders allow it to be changed based on the received power and many ground stations have different settings for it that can be chosen

Split level techniques

Another technique that is used to reduce noise in the PLL bandwidth concerns the pulsing technique chosen.



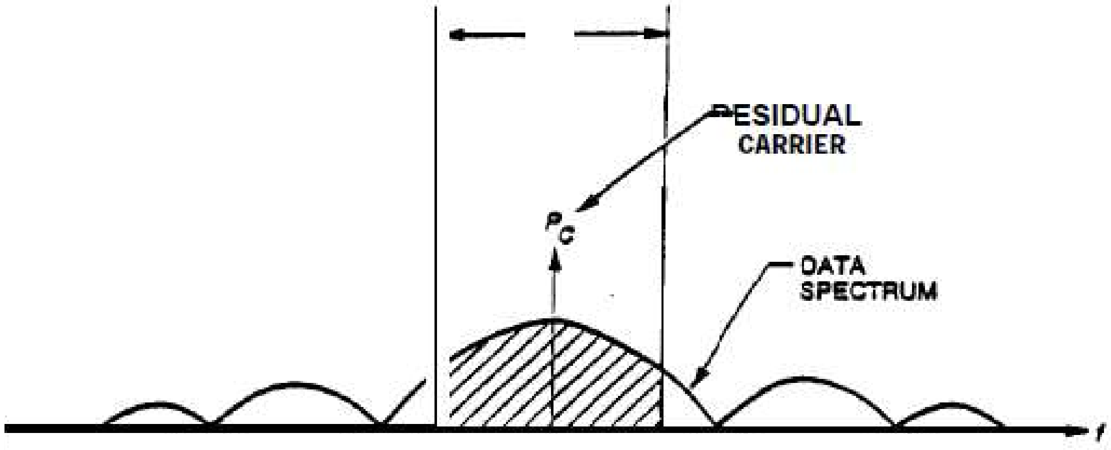
Non Return to Zero

Split Level

For the split level scheme we effectively double the channel symbols (i.e. cut the maximum possible data rate in halve) but something interesting happens when we perform a Fourier transform.

Split level techniques

NRZ-L

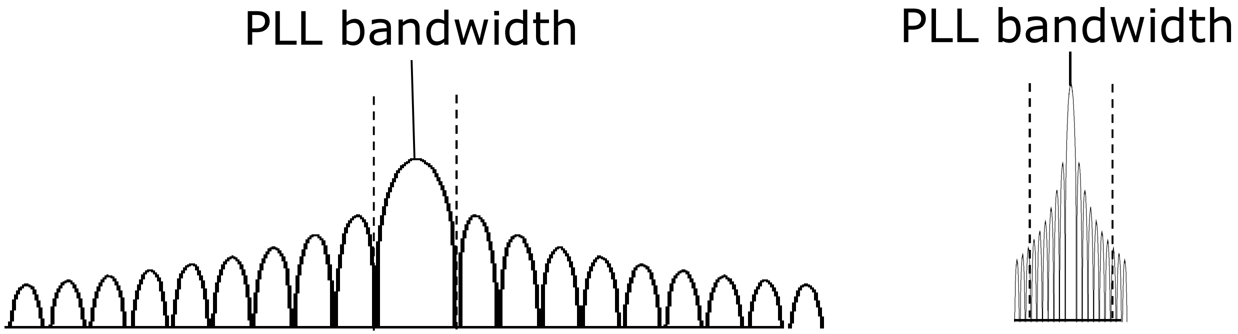


Subcarriers

Remember that if we have a very weak signal then we can only transmit a low data rate on it.

Also remember that the higher the data rate the wider the bandwidth used (spreads) and the lower the data rate the narrower the bandwidth used (bunches up).

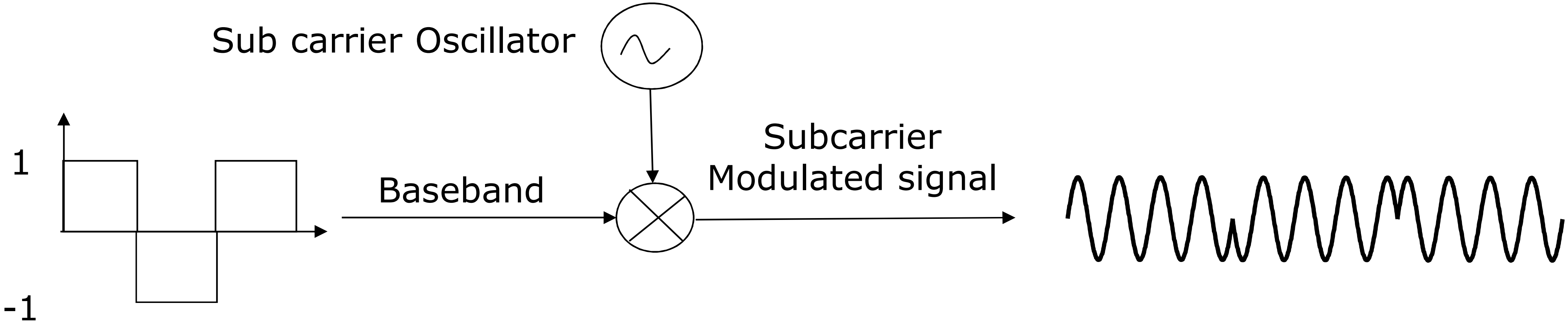
Hence at very low data rates most of the power of the data signal sits near the carrier right in the bandwidth of the PLL loop. This interferes with its operation.



There comes a point at which the PLL loop no longer works even with split level applied. We need to move the data signal away from the carrier another way.

Subcarriers

We do this using a scheme called subcarriers. First we mix the baseband signal with a sinusoidal signal that has a low frequency called the subcarrier.



Sub carrier Oscillator

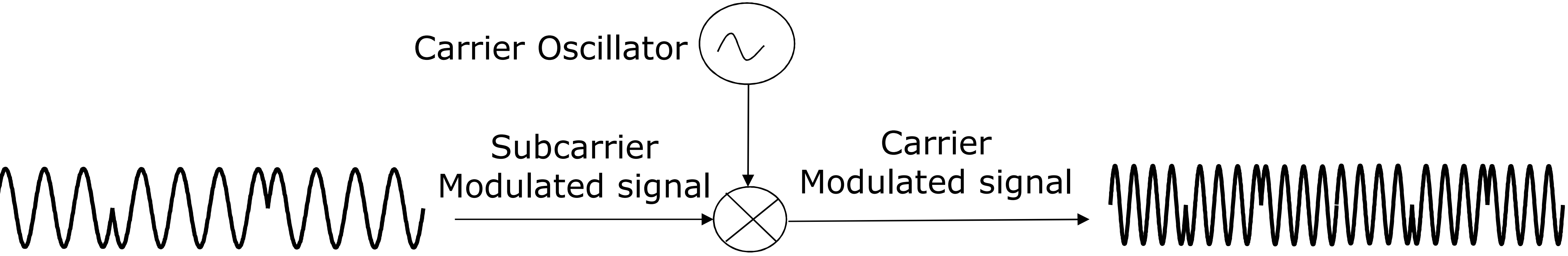
1

-1

Baseband

Subcarrier
Modulated signal

Then the resulting wave is phase modulated on to a carrier wave - as before.



Carrier Oscillator

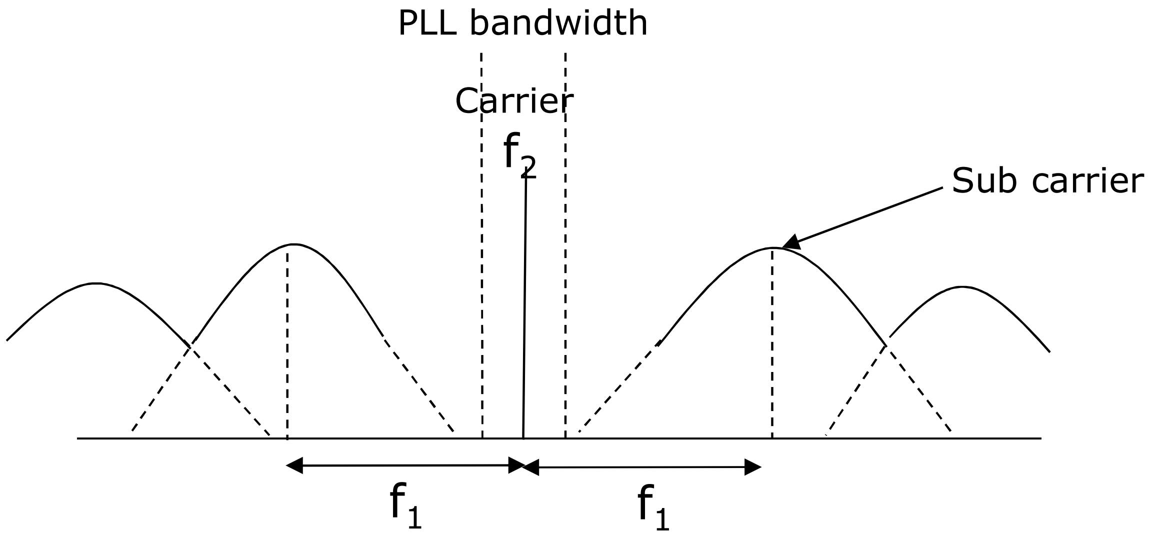
Subcarrier
Modulated signal

Carrier
Modulated signal

Subcarriers

If the baseband signal was modulated onto a subcarrier of frequency f_1 and then modulated onto a carrier with frequency f_2 then we end up with frequency plot as below.

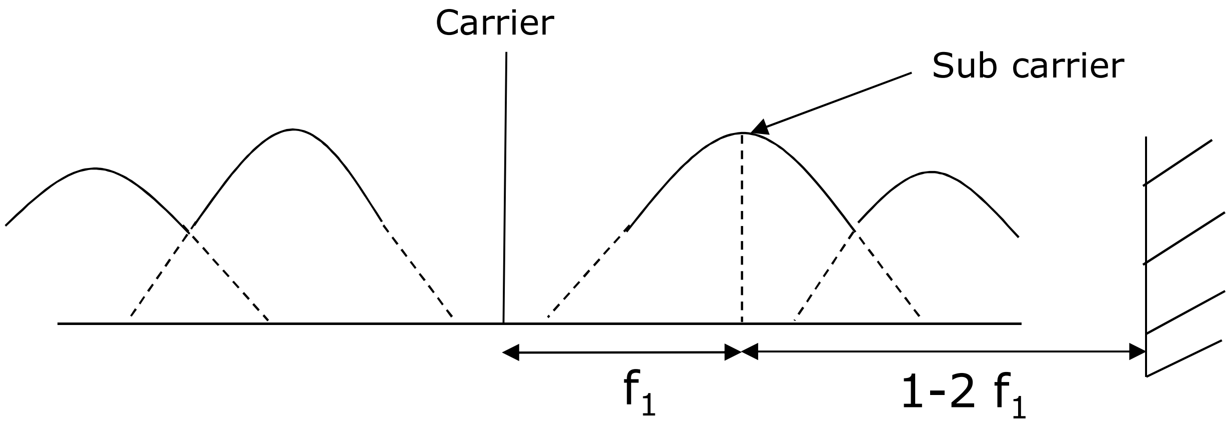
There are two identical peaks at a distance of f_1 from the carrier frequency as well as a sharp carrier signal.



f_1 is chosen so that the subcarrier data power spectrums have a null point where the original carrier appears.

Now the data and the carrier have been separated, the PLL will work.

Subcarriers are very, very bandwidth inefficient



Unfortunately using sub carriers is incredibly bandwidth inefficient.

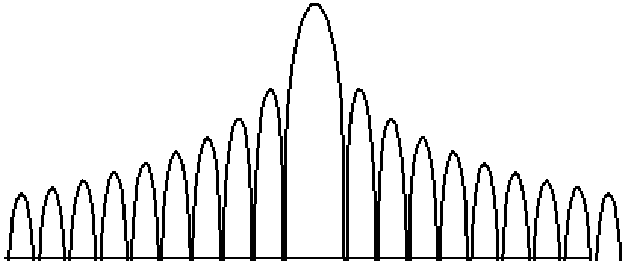
The distance f_1 must be 4 times the channel symbol rate (to be in the null point).

In order to ensure that 99% of the power in the signal is gone before the end of bandwidth allocation we need another 1-2 f_1 distance from the subcarrier frequency to the end of allocation.

Hence the use of sine wave pulses for near earth mission subcarriers where the data rate is typically high and the bandwidth is limited.

What about suppressed carrier schemes?

As noted earlier, if we use direct modulation with a modulation index of +/- ∏ rad for BPSK (or any other higher modulation scheme) we will totally suppress the carrier signal.



How can we recover the carrier signal when it has been totally suppressed?

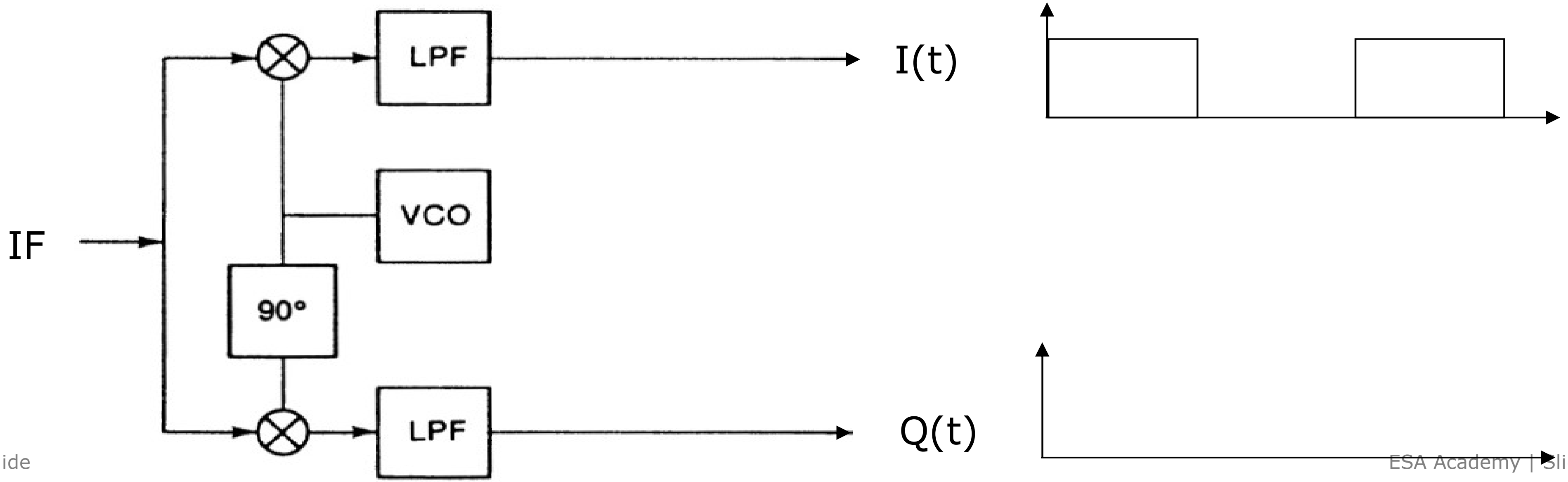
There are two poplar techniques

1) Costa’s loops
2) Squaring loops

Costa’s loop for BPSK

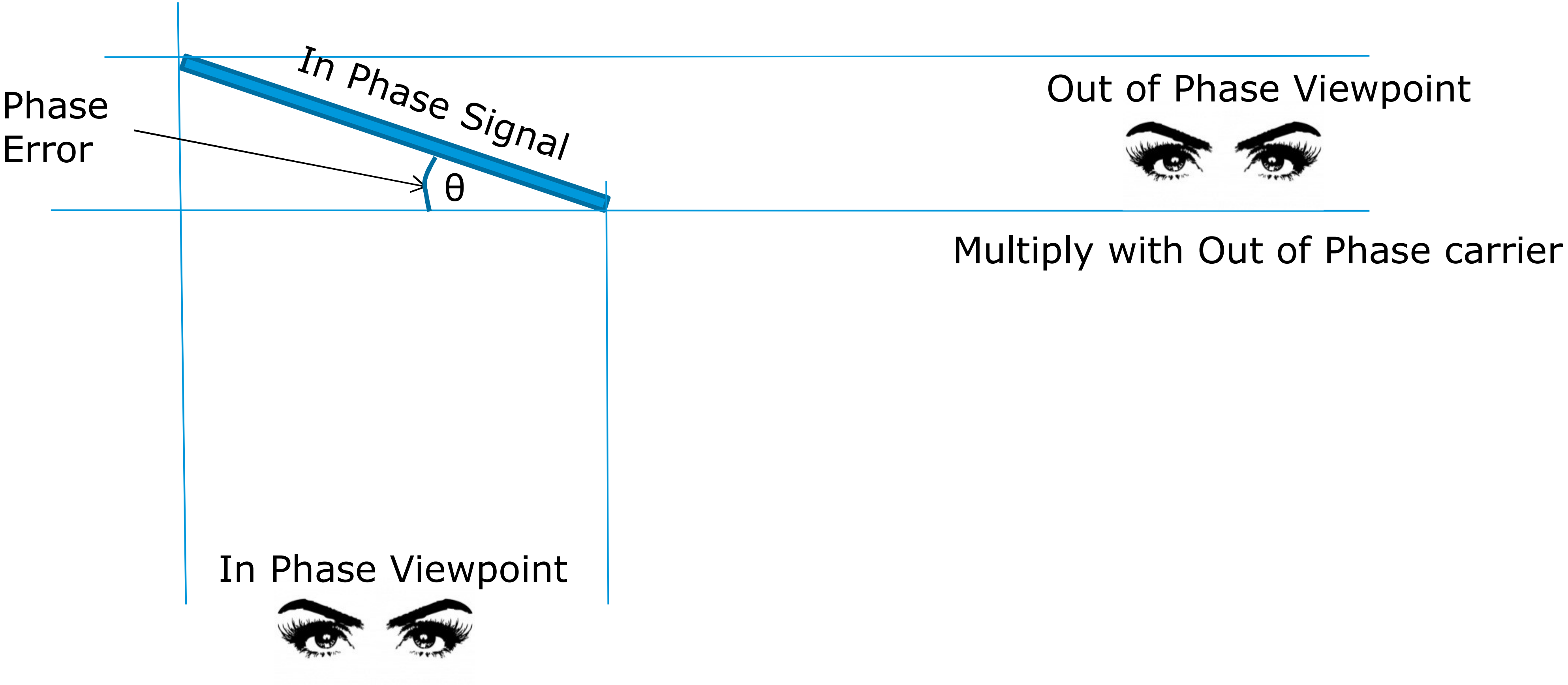
The more common way of determining the carrier frequency for a suppressed carrier signal is to use a Costa’s loop. These work by using the principle just discussed.

1. First “guess” the carrier frequency on the ground and mix the incoming stream with a cosine and sine wave carrier.
2. This effectively recreates the original I and Q streams from which the signal was made as low frequency output. This can be filtered off.
3. Note that for BPSK the Q stream will be zero and the I stream will be the original baseband signal



Costa’s loop for BPSK

In practise, we never guess perfectly, there will always be a phase error between the real carrier signal and the ground generated signal called θ.



Phase Error

Out of Phase Viewpoint

Multiply with Out of Phase carrier

In Phase Viewpoint

Recover with In-Phase carrier

Costa's loop for BPSK

In this case,

I(t) will have a data component x cos(θ)
Q(t) will have a data component x sin(θ)

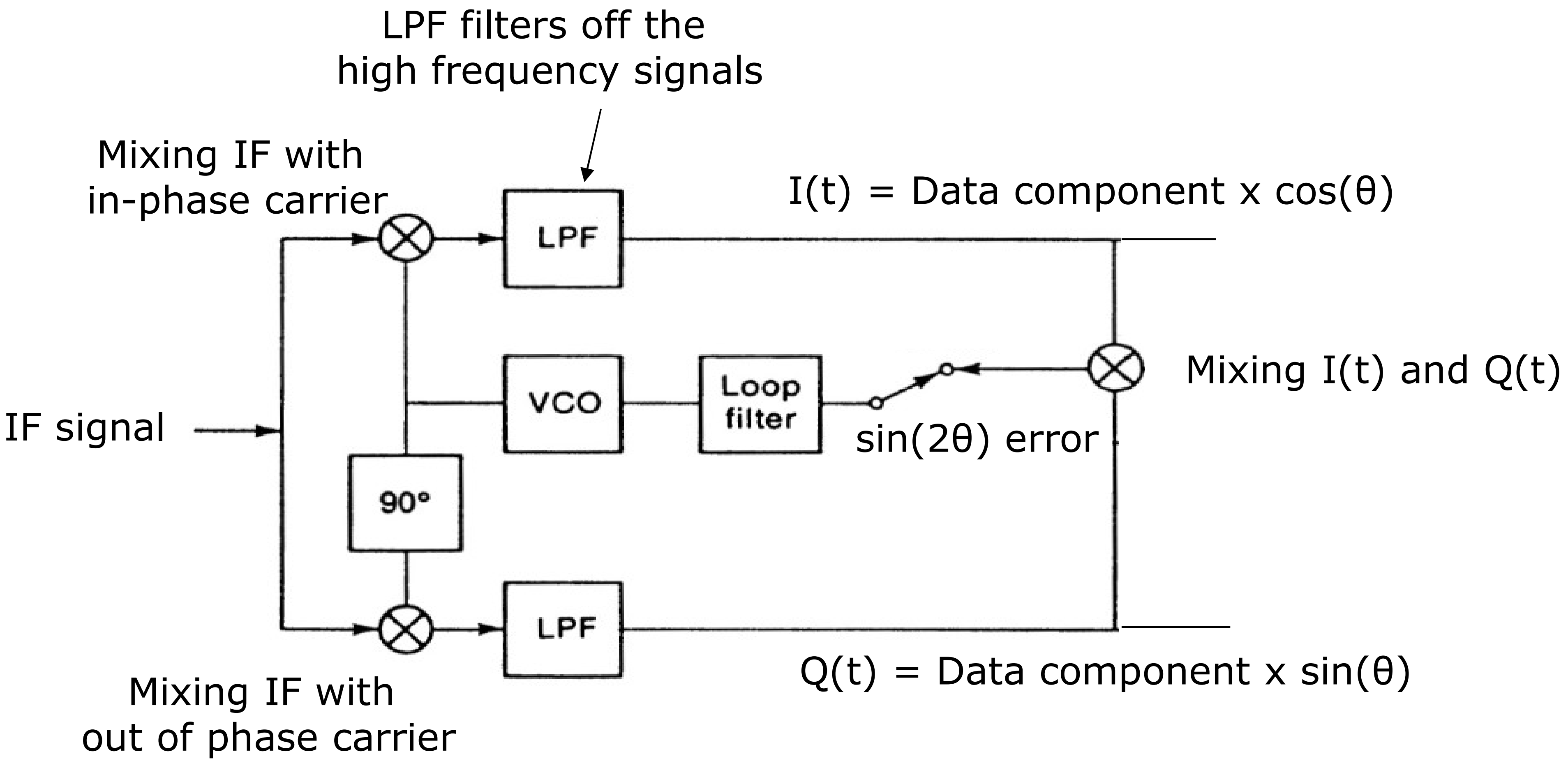
The I(t) and Q(t) signals are then multiplied together which becomes interesting..

2 x sin(θ) x cos(θ) is the same as sin(2θ). Therefore multiplying I(t) x Q(t) and then filtering produces a signal that is a function of sin(2θ)

Note the double sensitivity to the error (θ) - if θ is small - compared to PLL

This error signal can be used to drive the local oscillator through a classic PLL feedback loop until the phase error is zero and the oscillator is providing a perfect carrier frequency.

Costa’s loop for BPSK



LPF filters off the
high frequency signals

Mixing IF with
in-phase carrier

I(t) = Data component x cos(θ)

Mixing I(t) and Q(t)

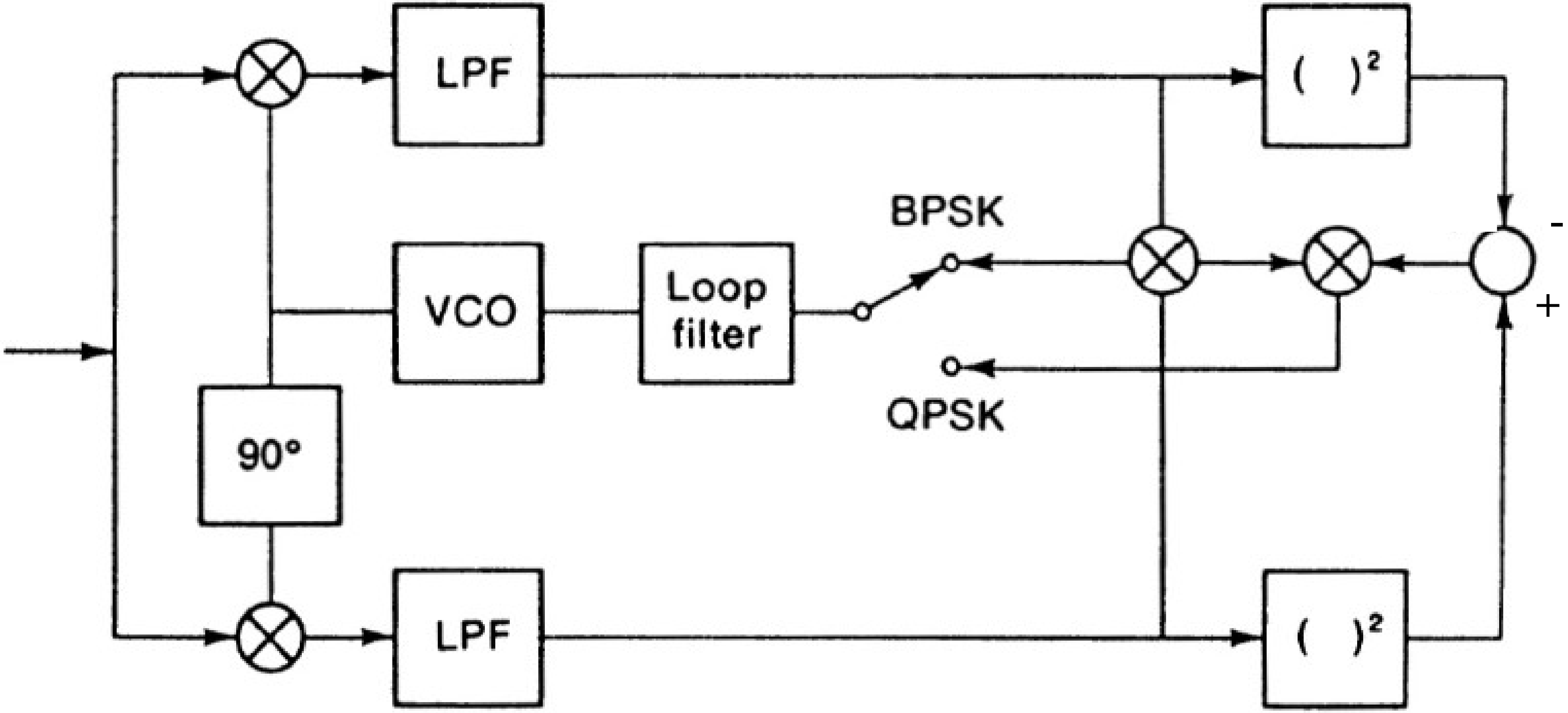
IF signal

sin(2θ) error

Mixing IF with
out of phase carrier

Q(t) = Data component x sin(θ)

Costa’s loop for QPSK



Costa’s loop for QPSK

There are some important implications of using Costa’s loops.

1. In the end it is a modified PLL so all the constraints on PLL use e.g. bandwidth still apply.

2. When we multiply the I and Q signals the total amount of noise in the PLL bandwidth increases due to noise-data and noise-noise multiplication products entering the frequency region. This is called the SQAURING loss.

3. Squaring loss depends on lots of factors and is difficult to calculate. Therefore in practise we simply say we need 6dB more signal power in the Costa’s PLL loop compared to that required for a simple PLL loop.

4. For BPSK there is always a 180 deg phase ambiguity on the result. This means the bits can effectively be inverted. This is usually dealt with by inserting a recognisable bit pattern in the data so that this can be resolved. For QPSK there are four possible combinations that have to be resolved.

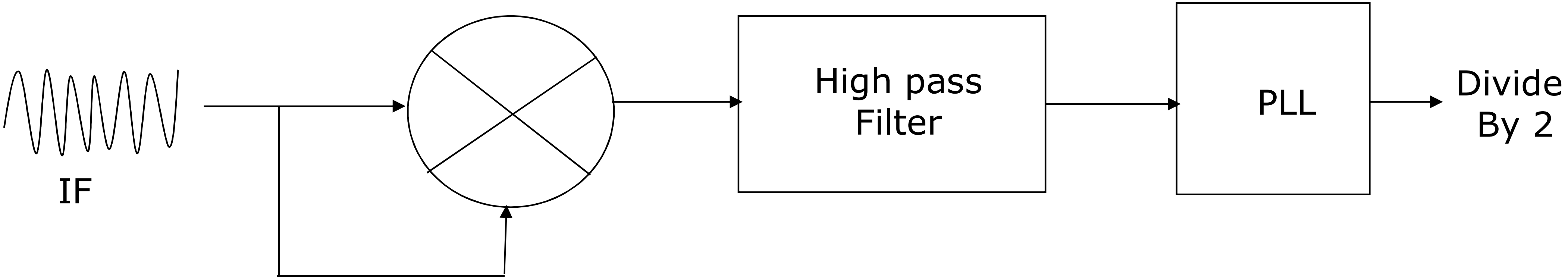
Squaring Loop

Another technique is to square the IF signal by mixing it with itself.

As we have seen before this will produce a signal centered at twice the carrier frequency.

This double frequency signal will contain a recognizable sharp signal at twice the original carrier frequency. This signal will be stronger than the noise after multiplication as the two waves that make up suppressed carrier BPSK are 180 degree phase shifted. Therefore doubling them will mean they will always reinforce each other. On the other hand the noise signals will add up randomly.

This type of loop is called a SQUARE and DIVIDE BY 2 loop.



Note it cannot work on QPSK as we would have to square the signals twice (90 x 4 = 360) in order to experience this reinforcement.

Time for a quick summary

To extract the baseband signal from the intermediate frequency signal……

1. We need to mix it with an exact copy of the carrier that the baseband signal was modulated onto

2. This is not easy because it moves due to on-board oscillator temperature variations, Doppler and atmospheric effects

3. So we need to detect it in the signal and track it

4. We use an phase-locked loop to do this.

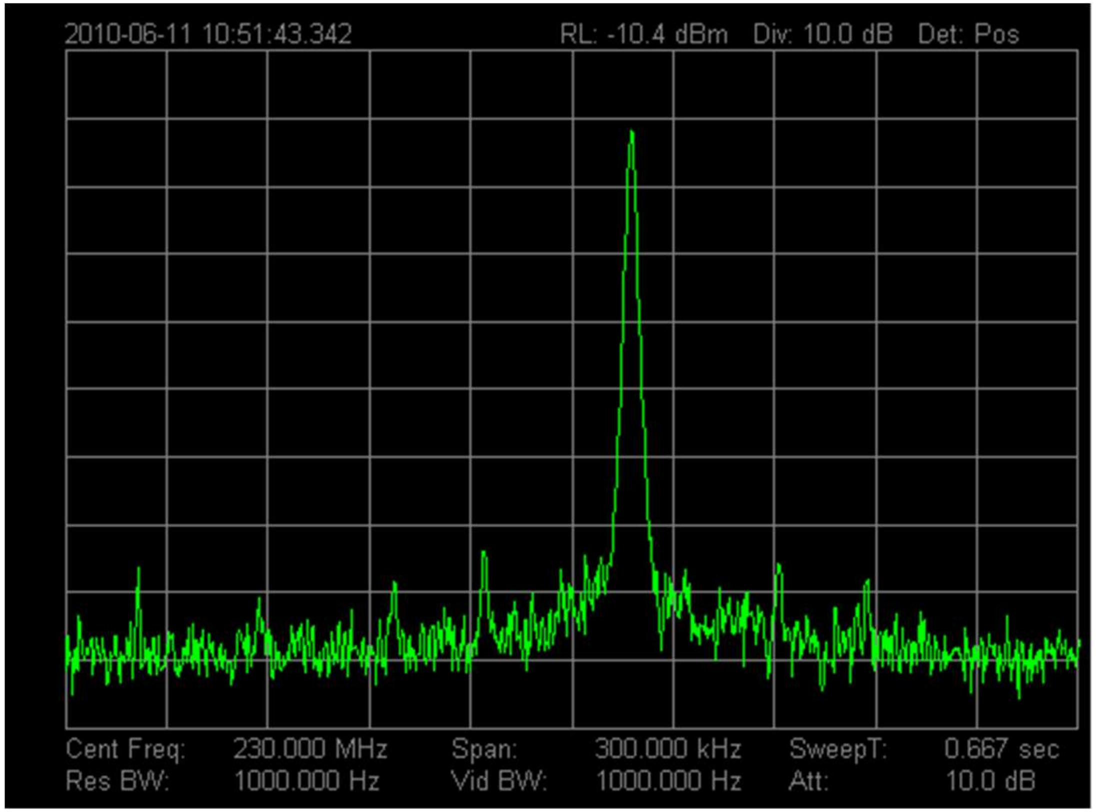
5. We have to present a clear, carrier signal for the PLL to track.

6. For this we can use residual carriers or subcarriers.

7. We have to make sure there is not too much noise in the loop. The relationship between PLL bandwidth and robustness is important.

8. If we use suppressed carrier (modulation index of pi) then we have to use a costa’s loop or a squaring loop to get the carrier frequency. This means losing 6dB.

What about uplink?



Spectrum Analyser of TC
BBE output at IF
(230 MHz in this case)

Sweep, Subcarrier Modulation – PLOP-2, Tone ranging on further subcarriers away from the data

QUIZ: Why do we sweep the uplink but not the downlink ?

Courtesy of Duncan Warren ESA/ESOC

Operational Experience: Voyager 2

Launched Aug 1977 (2 weeks earlier than Voyager 1) flybys Jupiter, Saturn, Uranus, Neptune

Now over 16 billion kilometers away. That is twice as far as Pluto!

The telemetry tape has traveled backwards and forwards about 4000 km!

Now close to the edge of the solar system

Will then take another 40,000 years to pass through the Oort Cloud

250,000 years after that will fly by Sirius 4.3 light years away

A demodulation story (1)

PROBLEM

8 months after launch ground forgot to command Voyager 2 for several days. The on-board software assumed that there was a problem and switched to the redundant receiver.

They then discovered that a capacitor in the PLL of the redundant receiver had broken.

So the ground had to get the frequency exactly right for the command to switch on the prime receiver to be accepted. It was sent many times with different frequencies and it worked!

………but the stress of so many “ON” commands blew the fuse on the prime receiver. Oops.

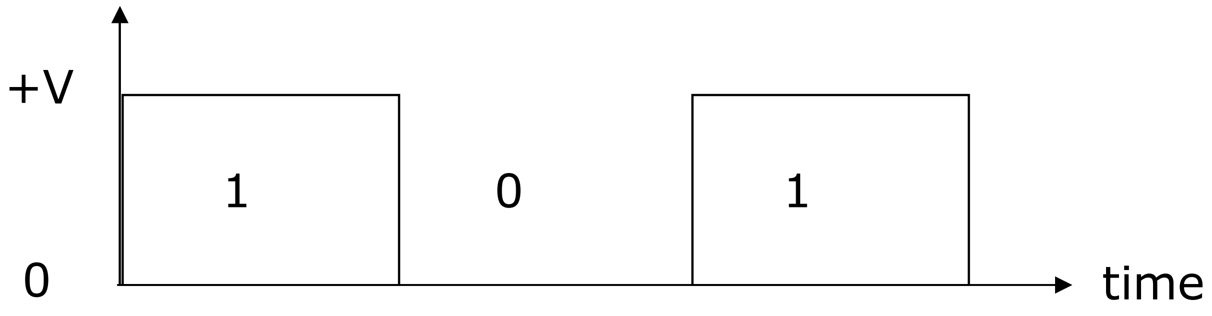
The redundant receiver was the only way in but now the exact acceptance frequency had to be calculated for each command. This depends on the Doppler shift plus the frequency of the local oscillator, which in turn depends on temperature. So when a system is switched off or on, the ground must wait until the temperature is stable before commanding.

This workaround has been in place for 35 years and the degraded receiver is still working.

QUIZ: What is the operational lesson?

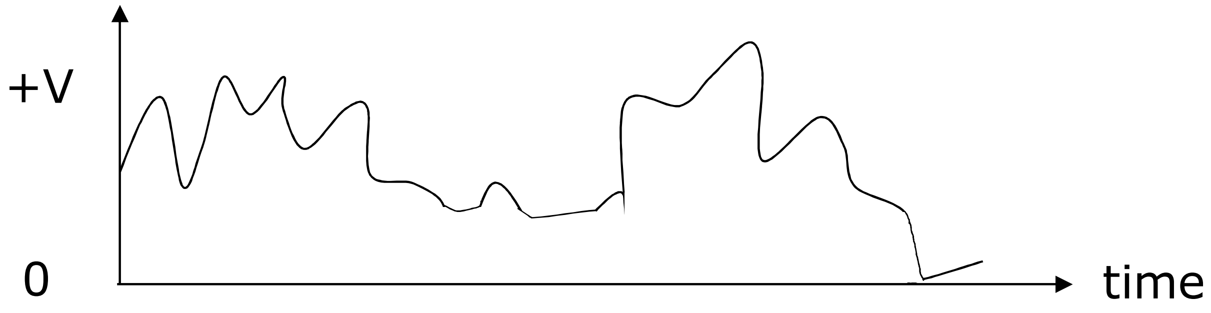
We have the baseband signal, now what?

We have recovered the baseband signal by multiplying the IF signal with the recovered carrier. However in the real world the recovered signal will not look like this..



Voltage pulses

Due to, inter Symbol Interference (ISI), Doppler, Jitter (jumping around of oscillator outputs) and good old fashioned noise it will look more like this..



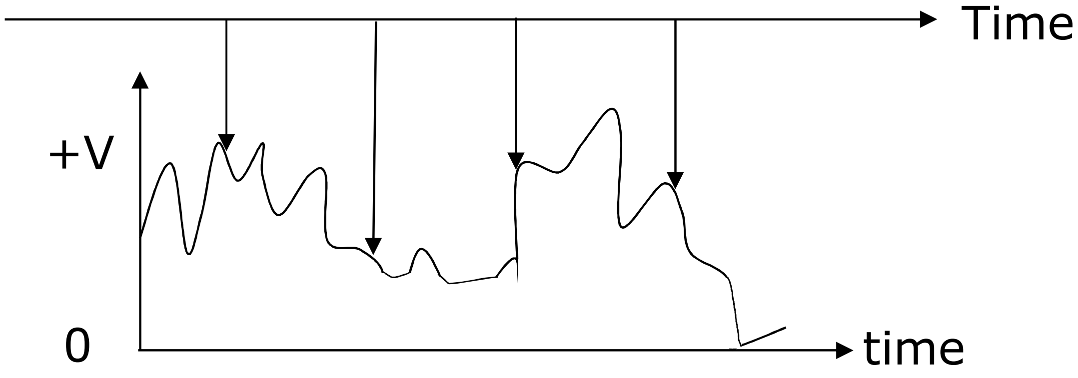
Voltage pulses

Strength and Timing

There are two important elements to being able to recover correct information from a weak, noisy signal like this:

1) Strength (making sure each transition has enough energy to be recognised above the noise)

1) Timing (knowing WHEN to measure)



Voltage pulses

Strength

Noise is like a stormy sea i.e. only predictable in a statistical sense

It is not a question of whether a symbol will always be completely submerged when you measure it but rather how many times that will happen over many measurements i.e. statistics

Eb/No

In order to simplify calculations we do not talk about symbol energy but bit energy E_b.

It is very easy to calculate

$$E_b = \frac{\text{Power of received signal}}{\text{Information rate}}$$

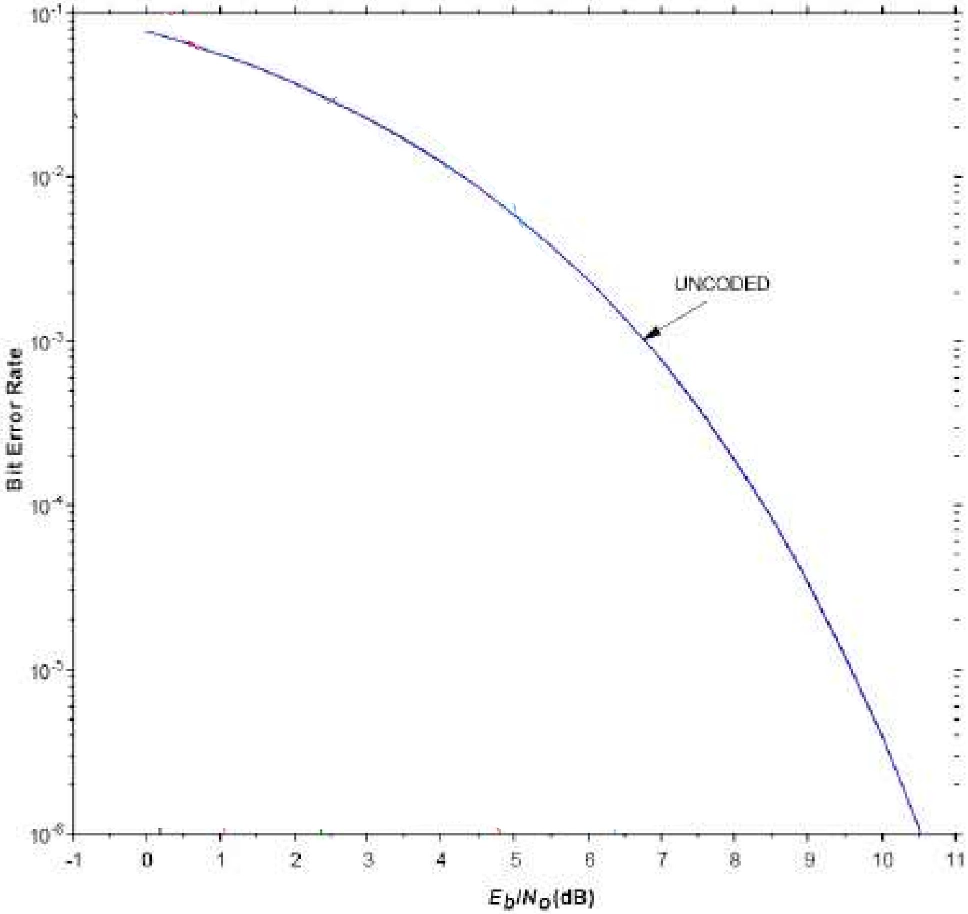
Going back to the rock in the sea example though, knowing the height of the rock is not enough information. We need to know how far it is **above the waves**.

For digital communication systems the equivalent sea level, is N_0 or noise density. Again it is easy to calculate.

$$N_0 = \frac{\text{Power of noise}}{\text{Bandwidth of signal}}$$

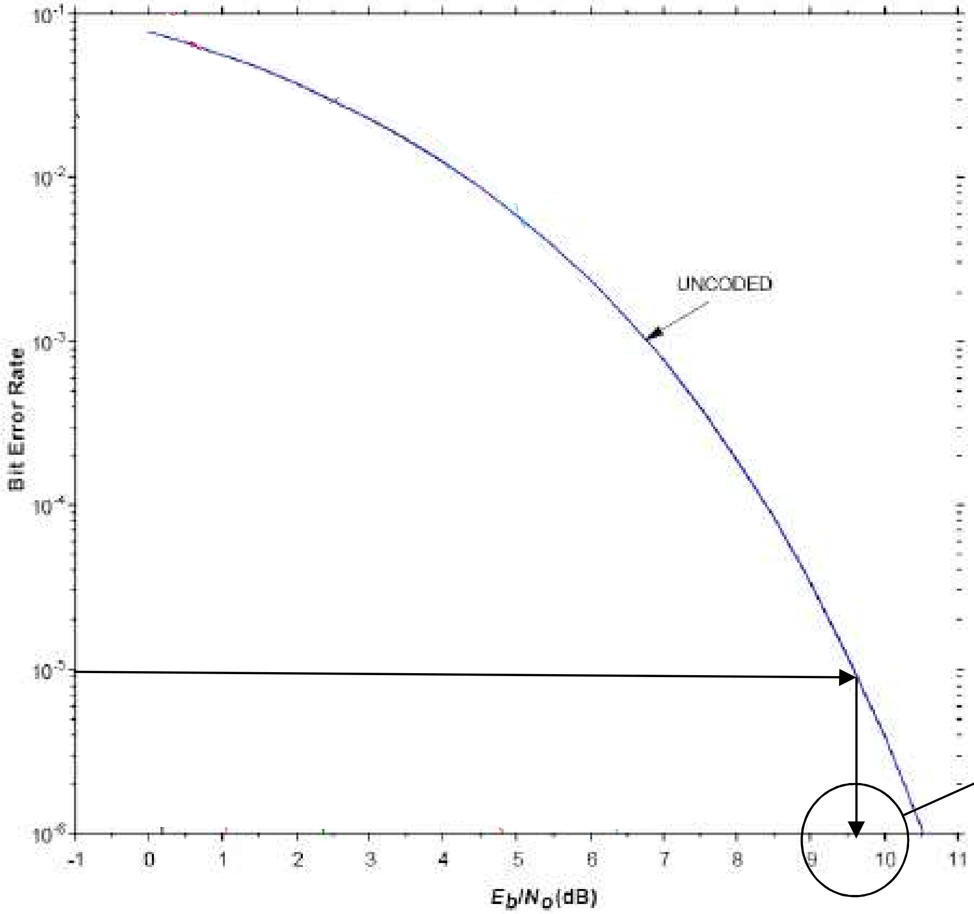
Eb/No and probability of error

There is a function that relates the probability of an error to Eb/No for certain modulation schemes. They are usually presented as graphs as below for coherent BPSK demodulation. A scaling factor can be applied to use these graphs with other modulation scheme.



Eb/No and probability of error

The communications link designer must decide how much error he can accept on the link and find the corresponding Eb/No.



This becomes an important DESIGN goal for the link

A word of warning

Bit Error Rate is not the practical error rate.

QUIZ: Why not?

Beware of the BER

Timing

In order to give ourselves the best chance of recovering the original information the phase of the IF signal (of amplitude of I and Q) must be sampled at the right time while the channel symbol is being transmitted.

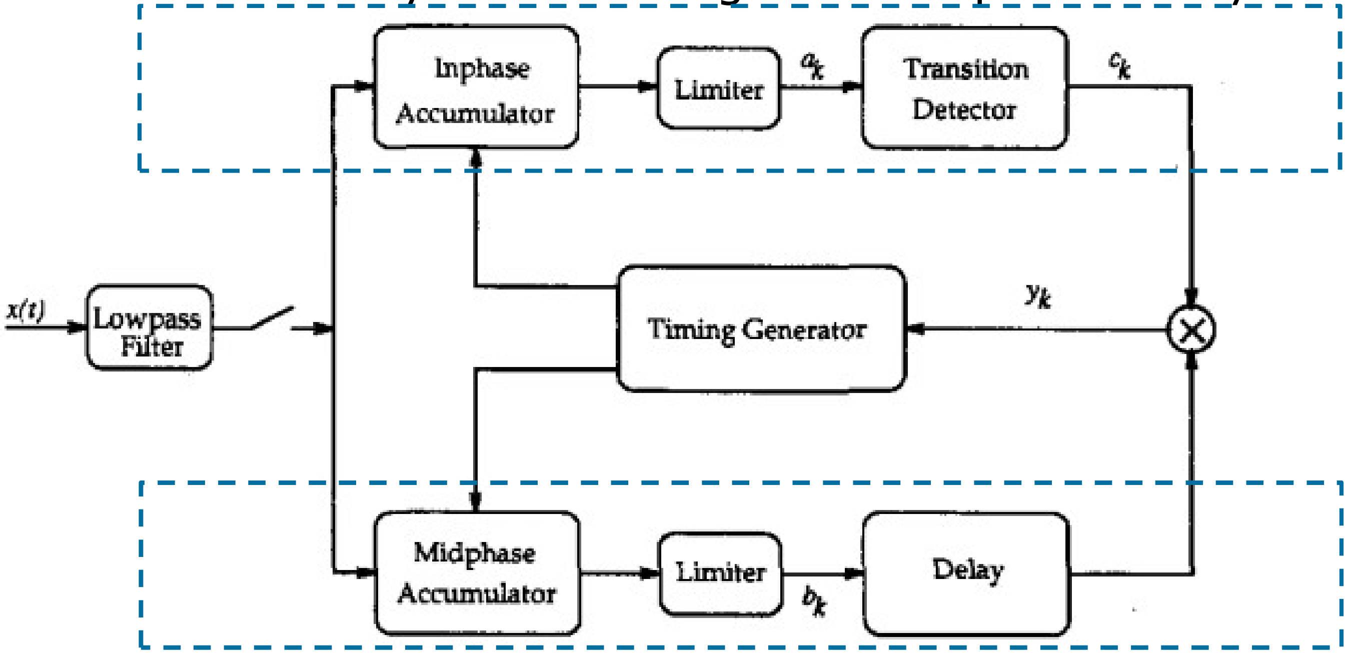
Otherwise we will sample it when it transitioning to another symbol and make errors.

So the timing of the measurements are also crucial

Timing

We need a control loop to predict and track the symbol timing we use to make the samples.

This is often called a **bit synchroniser** and the control loop is called a **Data Transition Tracking Loop (DTTL).**



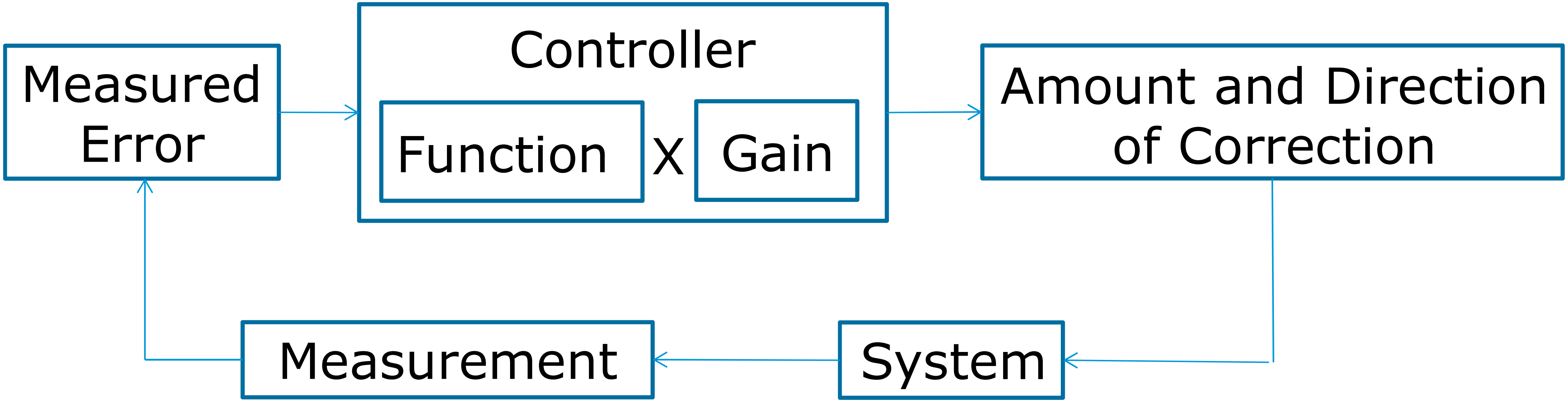
Measures Symbols – integrates samples over symbol start to end

The Loop outputs Symbols and a Clock

Measures transitions – integrates samples over sample mid point to mid point

The DTTL is a classic control loop

Note the DTTL is a control loop which tracks the transition timings.
Measured Error is the time offset between the expected data transition and the measured one.



We control the “strength” of the loop via the Gain parameter.

Too strong and we overshoot and oscillate while tracking.
Too weak and we never converge (or even diverge) while tracking.

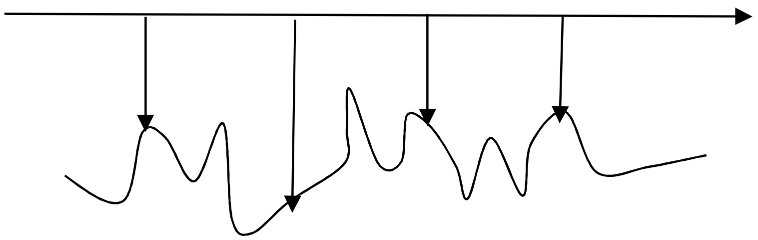
Gain is closely related to the loop bandwidth i.e. the amount of frequency shift required to cause a symbol slip

Timing

When the loop is working and symbol transitions are being correctly predicted and found by the loop we call it BIT LOCK

When the loop fails to find the transitions in the predicted window we refer to symbol or bit slips…

Clock



The DTTL measures the amplitude of the I and Q signals and these are then combined and used to estimate the most likely transmitted symbol based on the measurement.

This is then converted into a set of bits according to the constellation diagram and passed further down the chain for frame synchronisation (see coding session).

A demodulation story (2)

1st August 1992, 10 hours after shuttle launch:
It was discovered that the power connection between EURECA and the robot arm did not work and the batteries would only last 3 hours.

So the FCT had 3 hours to deploy the solar arrays or the mission would end.

During main computer activation the TM (which was routed via the shuttle) dropped.

FCT execute CRPs to recover but each time TM drops. They decided to take ESTRACK passes Maspalomas & Kourou. The Shuttle had to be rotated to allow this.

But suddenly the Maspalomas station was declared red, now only Kourou was available
The activation procedure had to be stripped bare to fit the 5 minute pass

The procedure was run during the Kourou pass and afterwards the TM to shuttle was OK
The arrays were deployed just in time (in sunlight against recommendations).
The mission was saved.

EURECA deployment problems

The problem:

During computer activation a TM packet from the ADCS was enabled which contained too many zeros

QUIZ: Why was this a problem?

Too much time between transitions for the shuttle **DTTL** to remain locked.

Luckily the stripped down procedure skipped the activation of this packet. So in a sense, the timely loss of Maspalomas saved the mission!

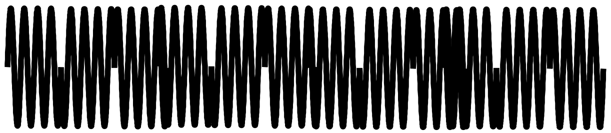
The lesson:

When you have a really serious problem it can help to strip everything that is not essential to the task out - or switch it off. This sometimes means lowering safety levels (like deactivating FDIR) so only do it as a last resort.

http://www.esa.int/esaMI/ESOC_40_Anniversary/SEMQIU19R9F_0.html

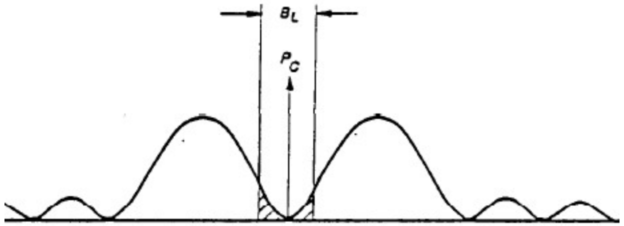
Demodulation - recap

Low Noise Amplification



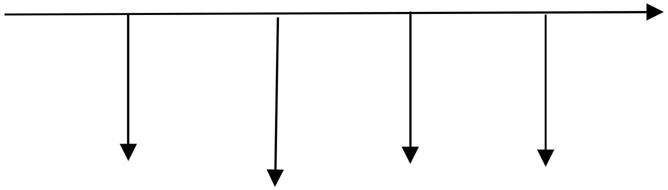
Down Conversion to IF

Carrier recover



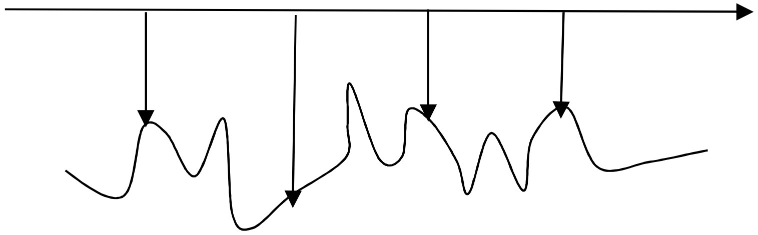
CARRIER LOCK

DTTL Symbol timing recover



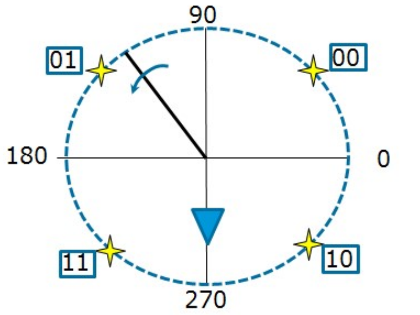
BIT LOCK

Amplitude sampling of I+Q



IQ IQ

Channel Symbols



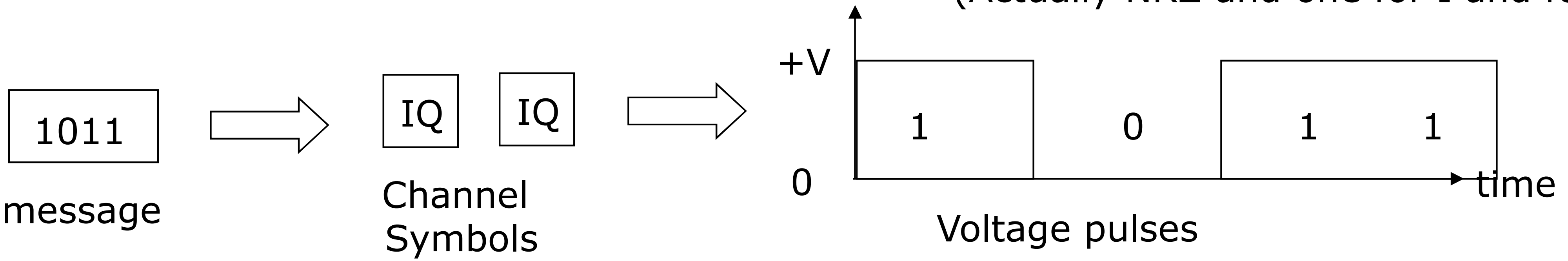
Channel Symbol to Bit conversion

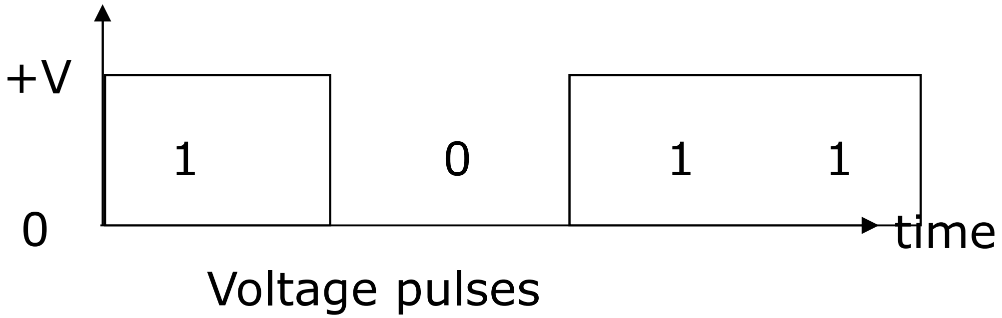
1011

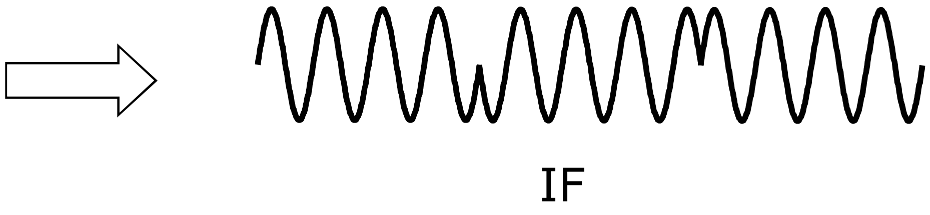
message

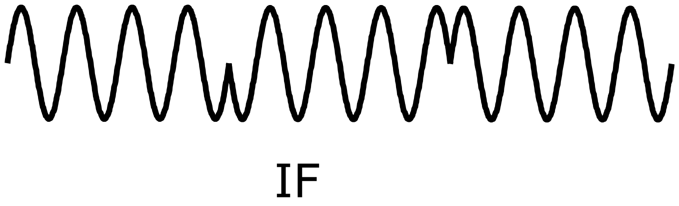
Modulation - recap

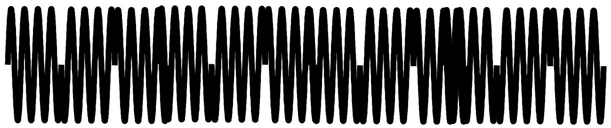
(Actually NRZ and one for I and for Q)

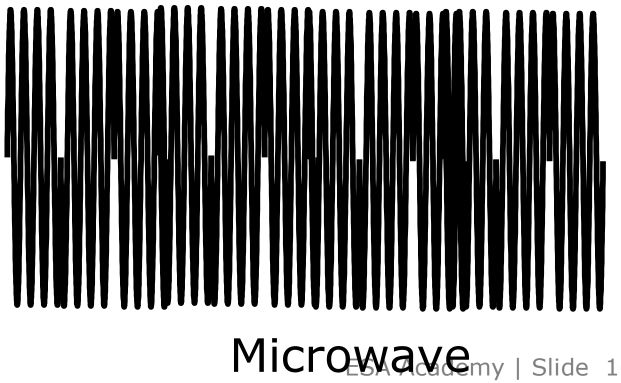












A demodulation story (3)

ESA still holds the record for the furthest landing in the solar system

http://www.esa.int/SPECIALS/Operations/SEMHJFZ7QQE_0.html

Huygens-Cassini Story

In 2000, Boris Smeds, an ESOC engineer championed changing a standard test of the communications link between Huygens and Cassini into a full blown test with data. No-one thought it was worth it.

He made a data pattern on his PC and performed a test using NASA’s deep space network antenna. Simulated fixed Doppler shift, varying power and data recovered.

He found that the more power he had, the less data could be recovered. No-one believed him.

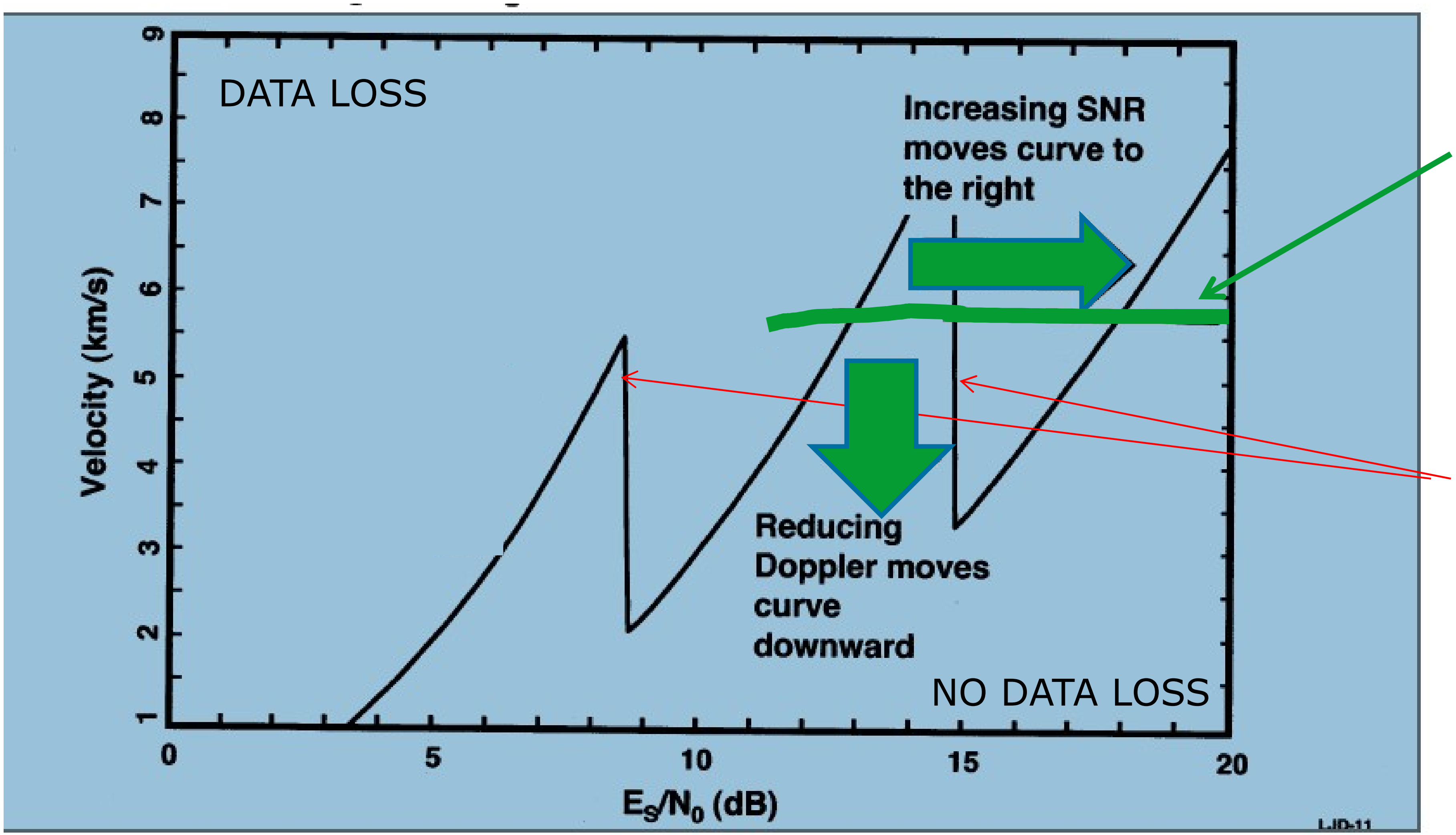
After preserving and doing tests on the engineering model at ESOC, Kevin Kewin and himself convinced the management there was a core problem.

The on-board receiver DTTL (bit synchroniser) loop gain was too weak given the expected Doppler shift. To make matters worst the loop gain reduced in response to increasing signal power.

If Huygens had flown like this there would have been very little data return from the mission.

http://spectrum.ieee.org/aerospace/space-flight/titan-calling/1

Trying to solve the problem

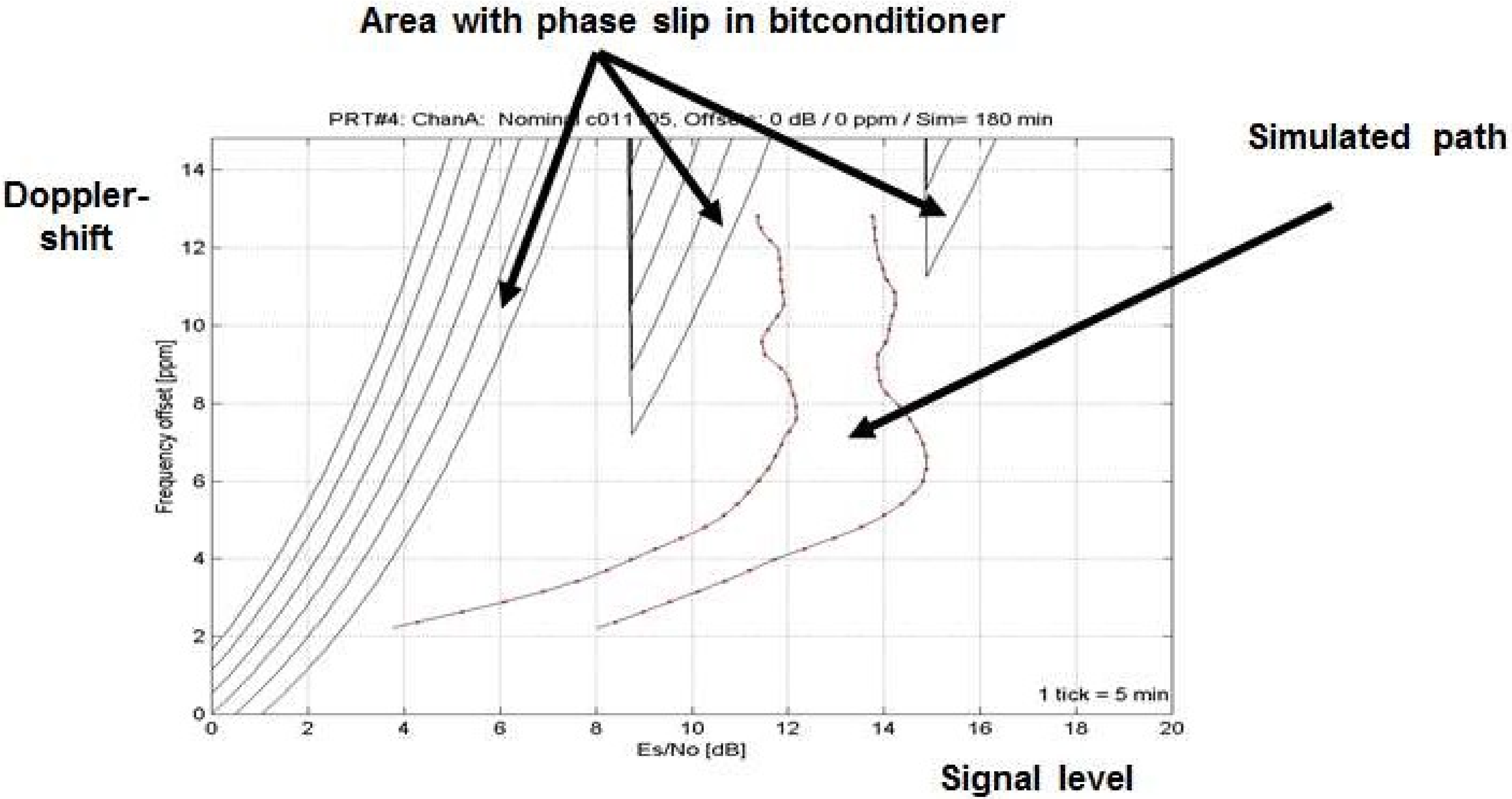


Original mission Profile Means Massive Data Loss

Loop Gain halves when these Eb/No Values are reached

Quiz: Why does increasing the number of bit transitions push the contours upwards?

Huygens-Cassini Story



Area with phase slip in bitconditioner

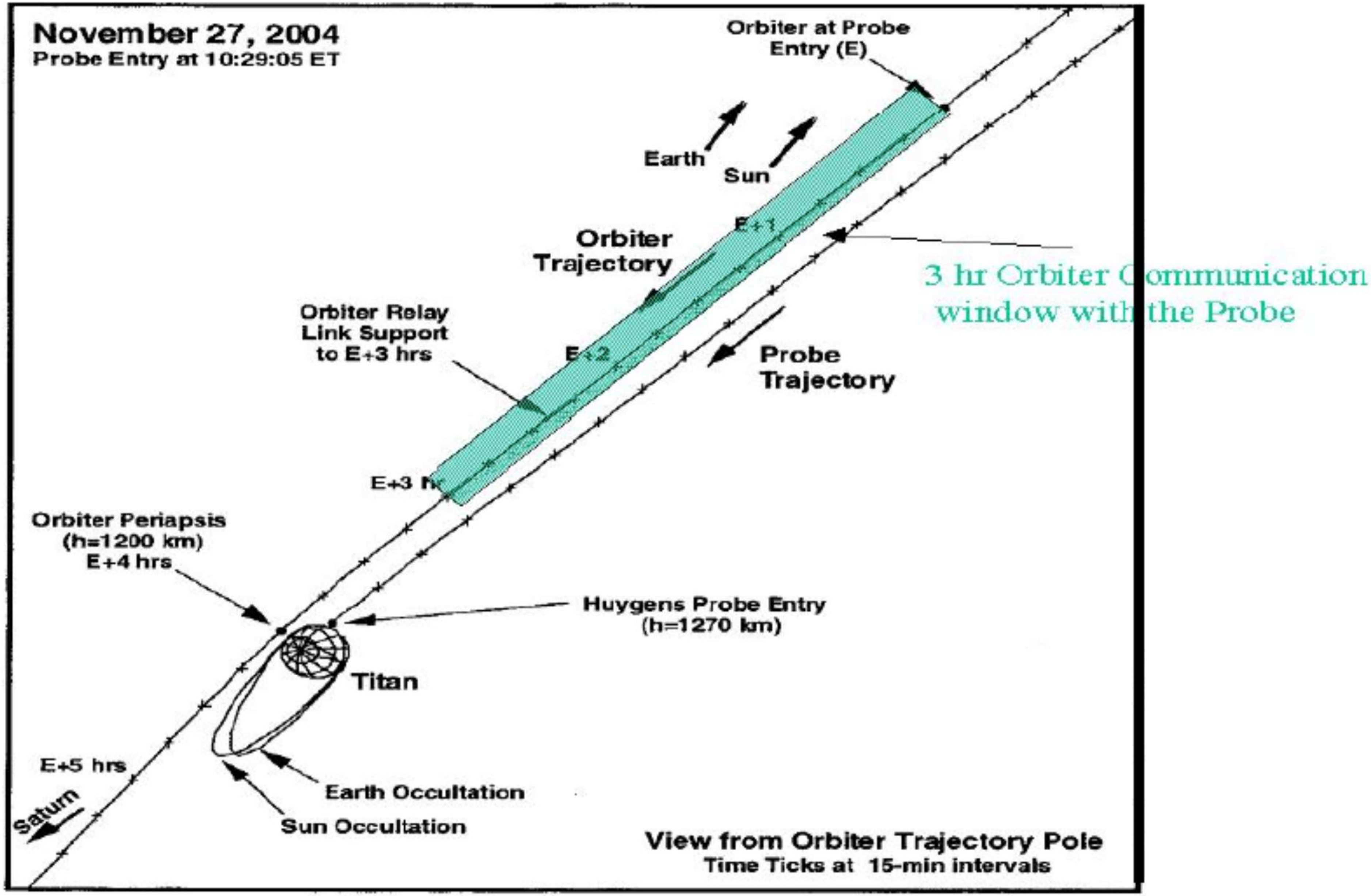
Doppler-shift

Simulated path

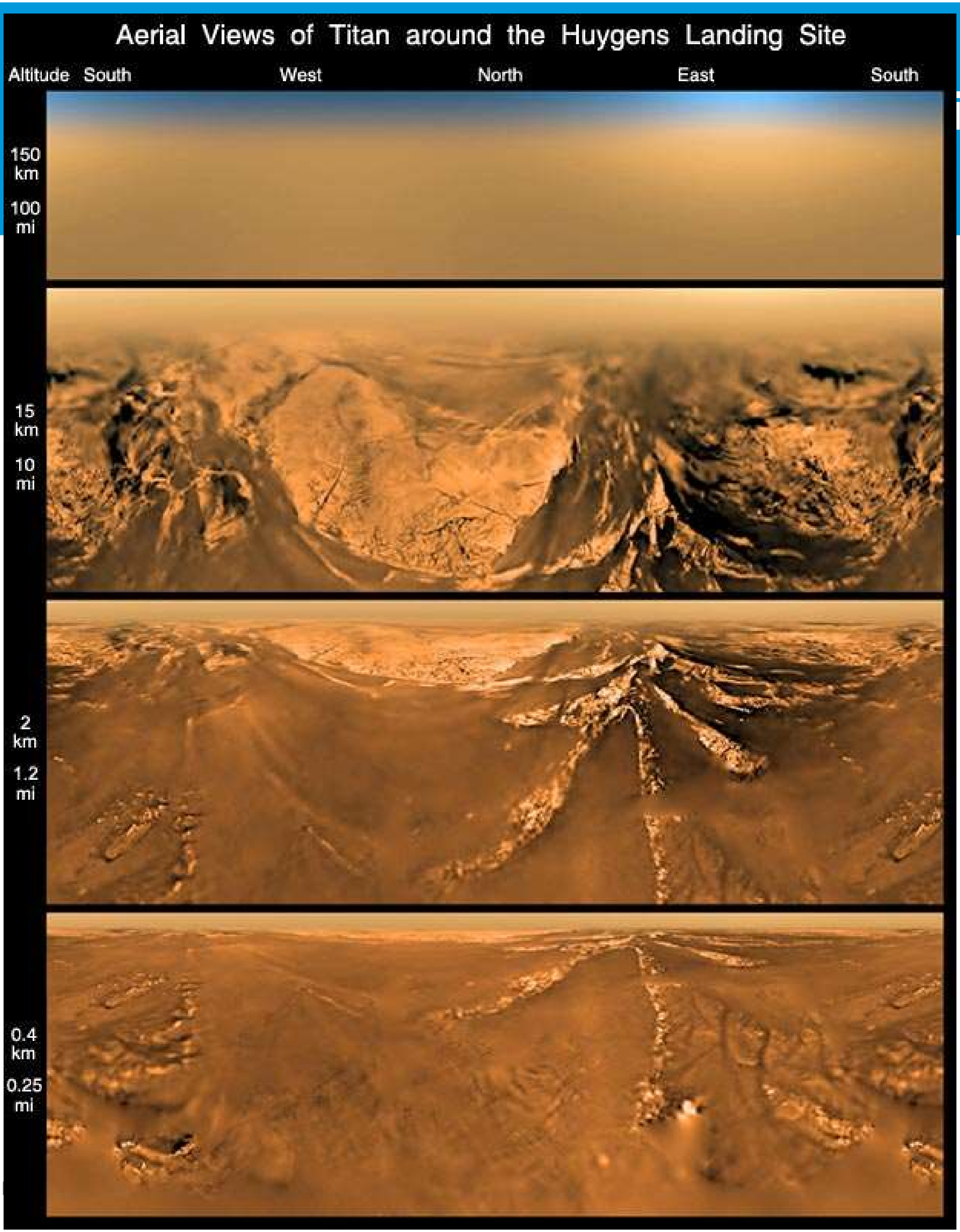
Signal level

Mission approach redesigned so that instead of following Huygens, Cassini approached parallel with the probe. This kept the Doppler, bit transition probability and S/N ratio in reasonable limits.

Huygens-Cassini Problem



i Problem



On Nov 2004, ESA successfully landed on Titan

Huygens-Cassini and Boris Smeds

TOMORROW'S TECHNOLOGY TODAY

IEEE SPECTRUM

OCTOBER 2004

www.spectrum.ieee.org

HUYGENS HERO

HOW ONE ENGINEER SAVED THE MISSION TO SATURN'S LARGEST MOON

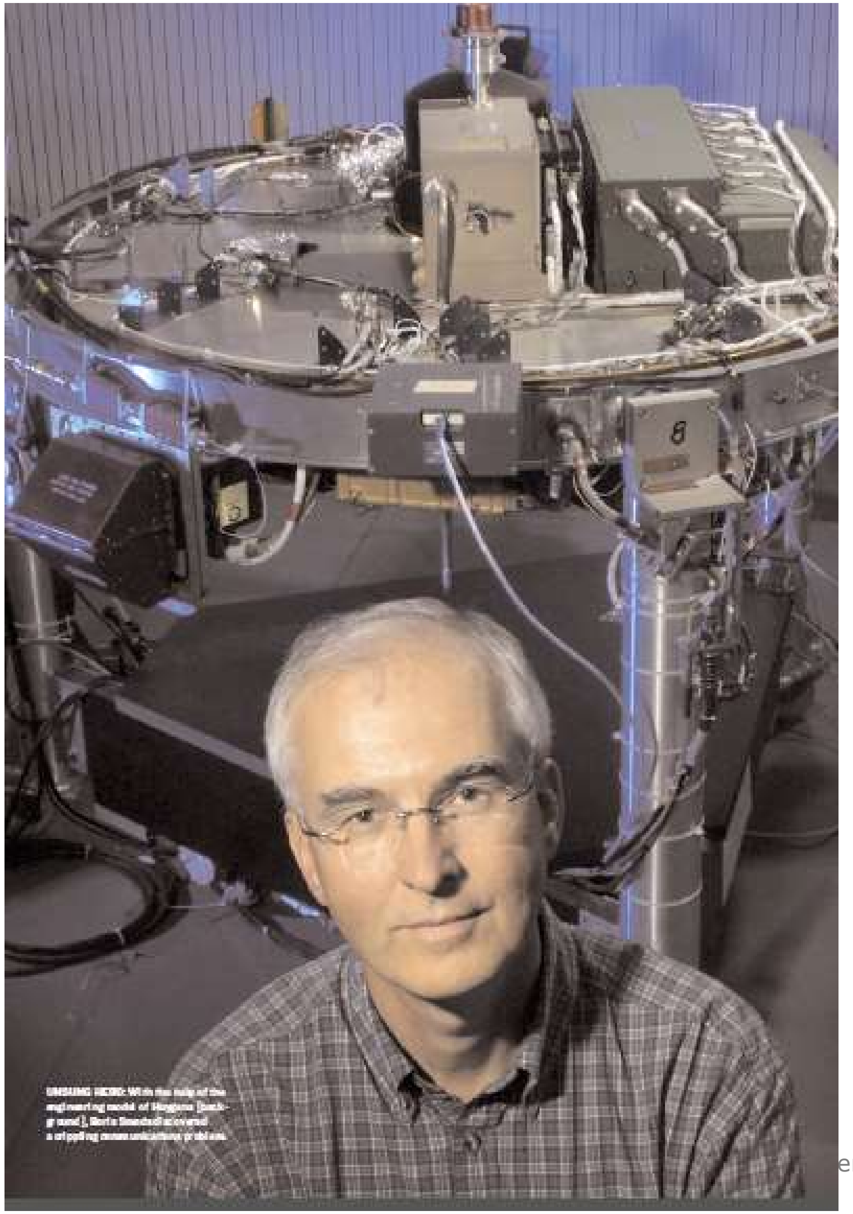
\+ IMPLANTED MICROCHIPS DELIVER LIFE-SAVING DRUGS P. 36

WHY ELECTRONIC VOTING STILL DOESN'T MAKE THE GRADE P. 22

WILL EAST ASIA'S TECH BOOM GO BUST? P. 30

DISPLAY UNTIL: 3 November 2004
www.ieee.org $4.99 US/$8.50 CAN

IEEE



But the Huygens-Cassini Story was not quite over

The link between Huygens and Cassini used two redundant S-band systems (Channel A and B)

But not all the data was carried on each channel.

Channel A was the only path for measuring wind speeds via Doppler changes and the only path for 350 of the 700 descent pictures.

A command was missed in the software sequence to turn on the receiver on Cassini for Channel A. Therefore all the data from channel A was transmitted by Huygens but never received.

It was an ESA error and the cause was human.

http://www.spaceflightnow.com/cassini/050115science.html

A final quiz

What pattern was on the spectrum analyser in the XMM 2008 anomaly?

If the TM carrier does not lock, will going to a lower data rate help?

What might help then?

- Commanding a different modulation scheme e.g. split level
- Reducing the PLL bandwidth
- Commanding a better on-board attitude
- Pointing the ground station antenna better
- Lowering the LNA temperature

If the carrier is locked but there is no bit lock, where is the problem?