



More on MODEMs by WU3C

Intro and Synchronization

12 July 20

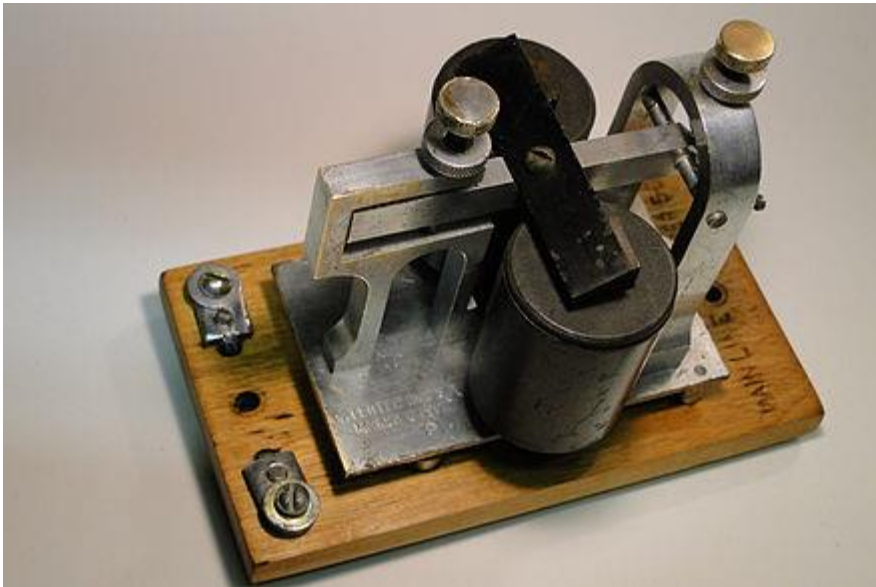
Caution: this is a fairly raw presentation!

MODEM old and new

A system that allows reliable transmission of information over a channel

The MODEMs job is ubiquitous, and taken completely for granted

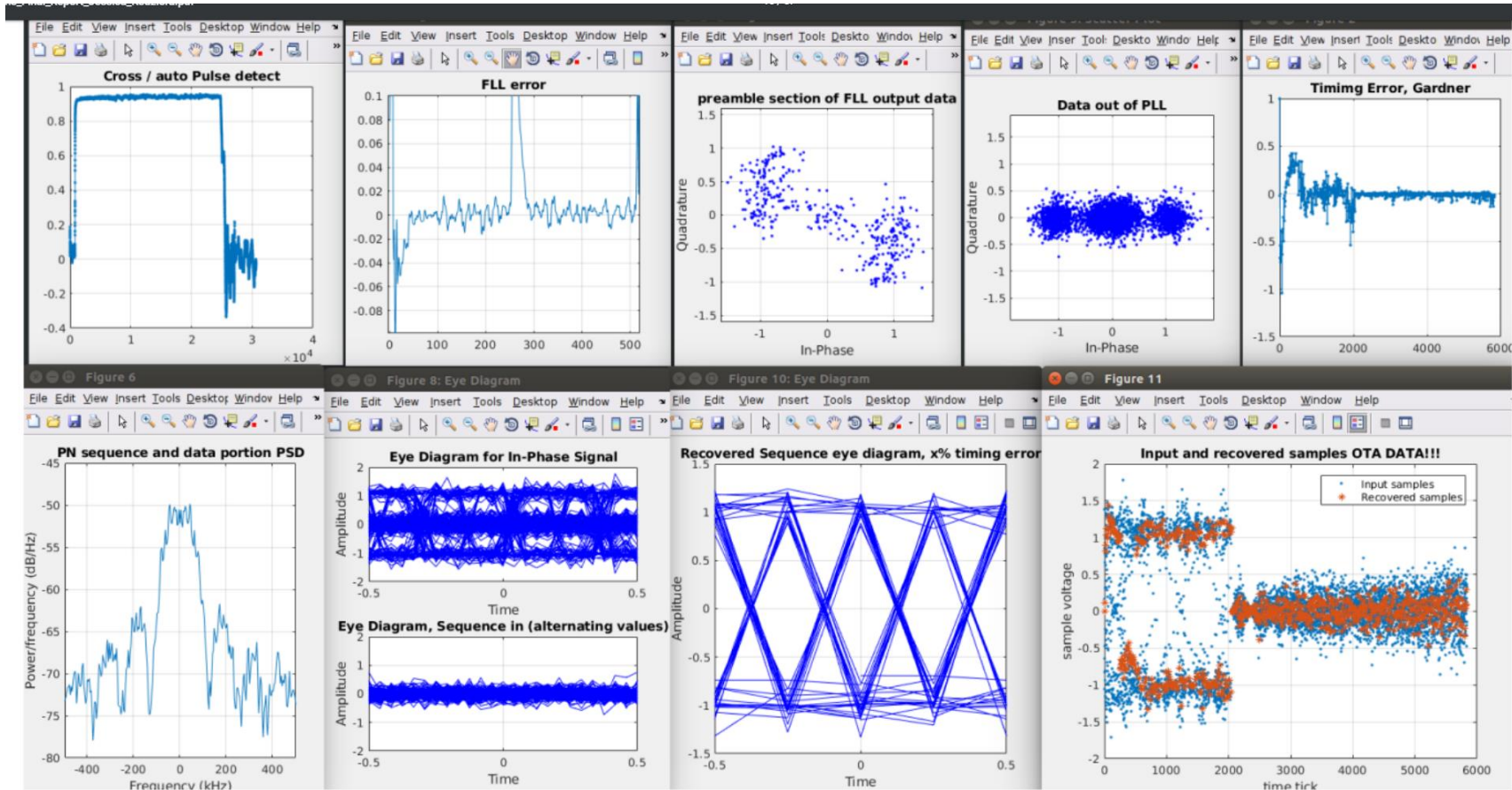
~45BPS



~100 000 000 000 bps



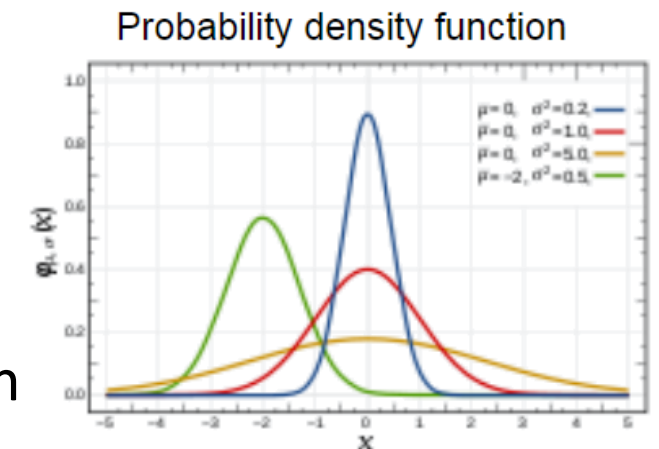
Goal: understand these images



Signal to noise ratio? REVISIT^T

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

- Also pretty easy for hams:
 - Ratio of signal power to noise power
 - Handy to normalize to 1Hz to avoid confusion (not common in HR)
 - AWGN: jargon, Additive White Gaussian Noise
- Confusion results when signal occupied bandwidth is much smaller than bandwidth used to compute noise power.
- “Boiled down” E_b/N_0 – energy per bit to the noise power in 1 hz BW (GAUSSIAN)
- “SNR” is then $C/N = E_b/N_0 * (R/B)$, where
 R = bit rate B = channel bandwidth



The red curve is the *standard normal distribution*

BER: Bit error rate

- The charts are ubiquitous in comms
 - Could spend hours just on this!
- They allow equal comparison of modulation schemes.
- Coherent modulation advantage is the basic motivation of this talk!
- Need to understand what it is and how to implement it

Coherent = better, but comes with burden

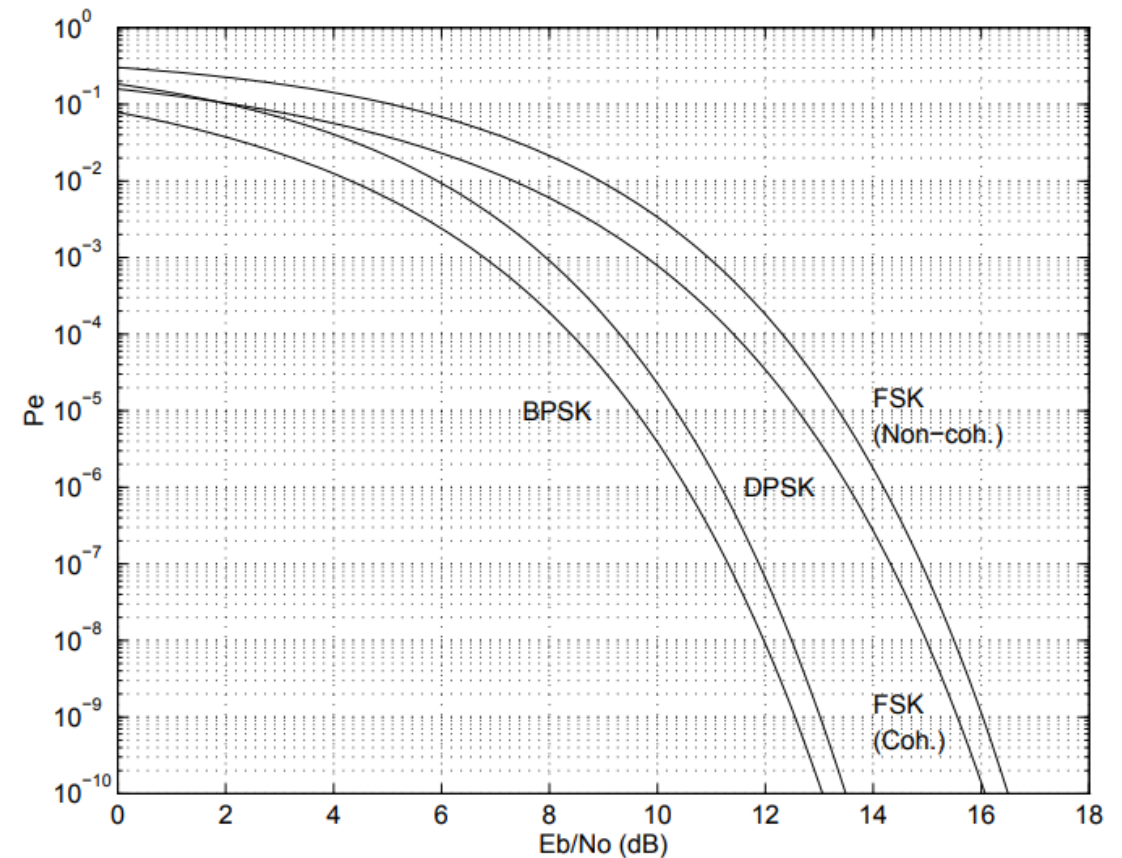


Figure 2.21: P_b for BPSK, DBPSK, and FSK

Symbol error rate vs. bit error rate

- Some confusion occurs in error rate graphs when symbol error rate is reported vs bit error rate
- Not super simple actually
 - Given N bits per symbol, Bit error rate is *approximately* symbol error rate / N
 - How bits are mapped and symbol probability change this though.

<https://dsp.stackexchange.com/questions/58124/why-is-bit-error-rate-symbol-error-rate-number-of-bits-per-symbol-in-qpsk>

What is a channel: a big question, but to chop it down:

- A medium with
 - Finite delay
 - AWGN noise (flat frequency noise)
 - Could contain impairments:
 - Multipath or echoes (memory and delay spread)
 - Fading not from multipath (variable loss)
 - Doppler
 - Interference
 - Distortion, dispersion
 - Where hardware is included, fixed and nonstationary
 - Frequency and phase error
 - Timing error
 - Distortion
 - Non AWGN noise

$$H(f, \tau, t)$$

What is a channel: a big question, but to chop it down:

- Wireless channels are generally linear (very helpful!) $H(f, \tau, t)$
 - Not always. EG:
 - when plasma's are present
 - Fiber under high power
- Can be represented as a linear frequency dependent transfer function with memory
- *Hams deal with VERY MILD channels: human ear can tolerate very little impairment.*

Shannon's Channel Capacity

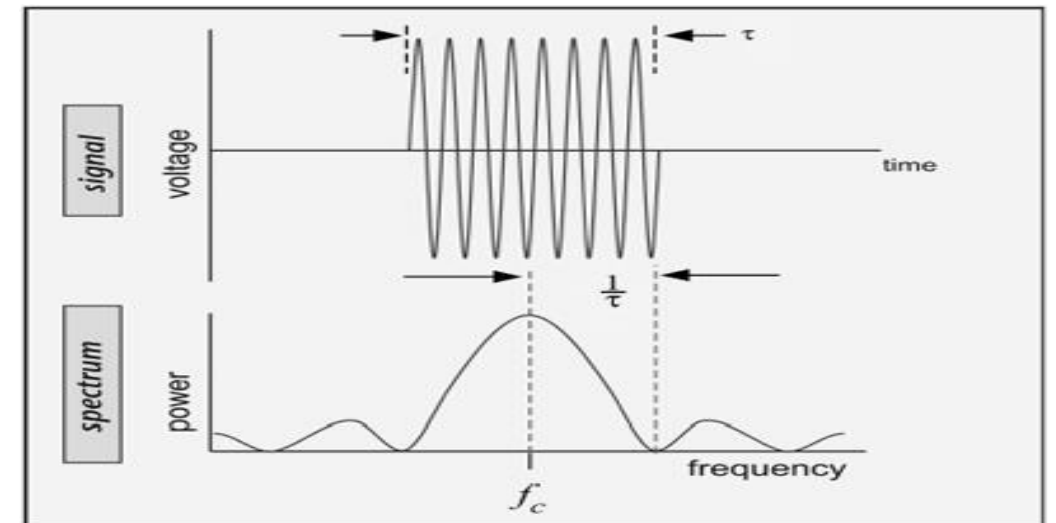
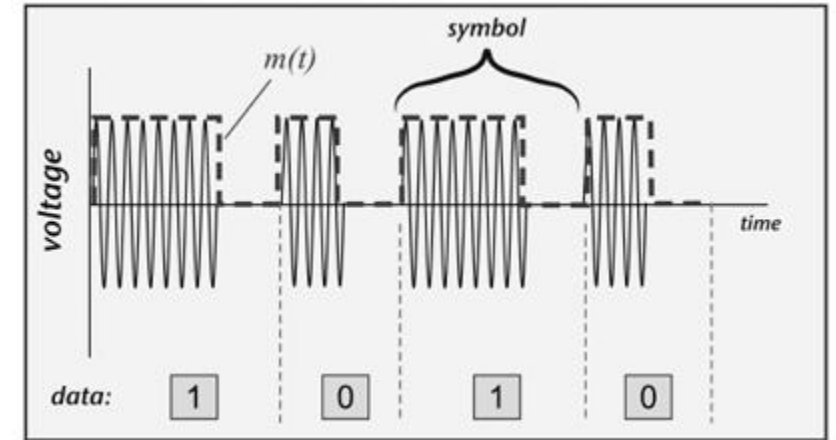
- Absolute lower limit computed using the Information matrix
 - Proof is difficult
- Channel Capacity = $C = Bw * \text{Log}_2 (1+ S/N)$ *for an arbitrarily low error rate*
- This is really a amazing thing akin to $E=MC^2$
- Simply because it gives the absolute limit of how good you can do
- There are not many things where that can be known.

Lets take a CW channel

- Using a 250Hz filter
- Assume a just barely detectable signal at 4dB SNR
- $C = 250 * \log_2 (1 + 2.55) = 250 * 1.82 \approx 457$ bits/ second!!
- Using PARIS as a standard word, 50WPM ≈ 41.66 bps (dit time)
- It is possible to send about 549 WPM inside this channel.
 - NOT possible with OOK!!!!
- Higher SNR = more capacity! This is not usually obvious.

What is possible with OOK?

- Nyquist criterion
 - BW must be $\geq$ larger than 2 sided BW
 - 250 HZ = 125BPS
 - 125BPS $\approx$ 150wpm
- Ok, so how does Shannon fit 549WPM in that channel?
- It is NOT just coding!
 - N bits/ symbol – modulation selection
 - Fix channel problems implied and necessary



<https://www.eetimes.com/tutorial-radio-basics-for-uhf-rfid-part-iii/#>

Back to Why's about MODEM's

- Two big points:
 - 1: You can put a lot of information through a channel, way more than CW, but complex things are required.
 - 2: Very severe channel impairment can be dealt with, way beyond what an “operator” could do
 - 3: in all cases synchronization is required

Now how do you get to the Shannon limit?

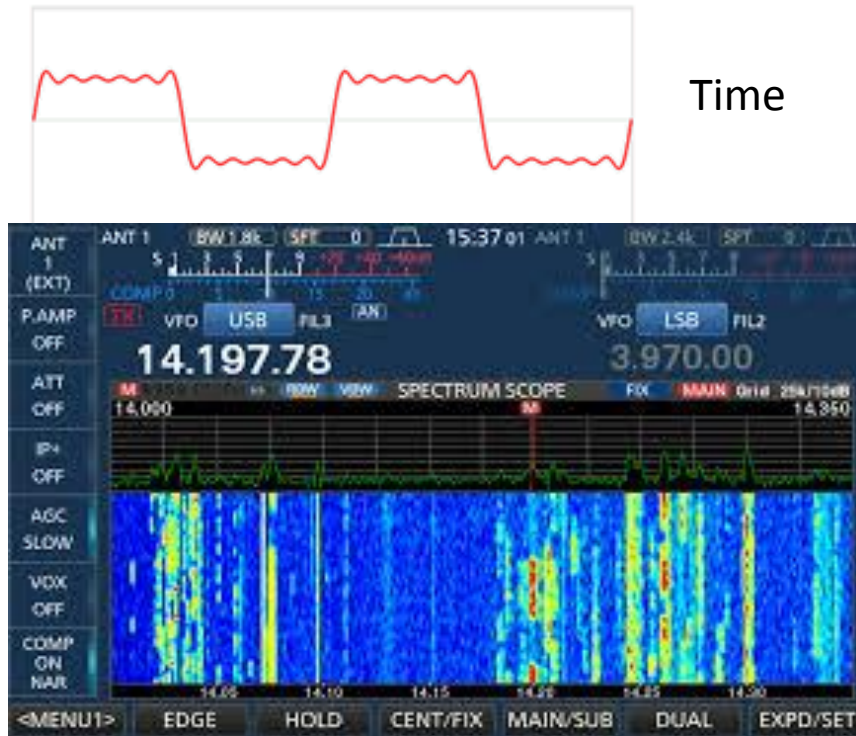
- As a broad brush 2 things:
 - Modulation degrees of freedom, methods of synchronization, impairment compensation
 - Coding : *assuming synchronization has been established*
- The second is where most people have some experience, but not the first. That is what this series of presentations will address.

Coherent modulation

- To understand synchronization, the concept of coherent modulation / demodulation is important.
- Broadly, coherent means - modulation where the signal frequency is aligned nearly perfectly and the phase has been adjusted (and the clock)
- Requires tracking as the channel changes
- To examine this problem, a little more background is useful on complex base band signals
- I'll start with another important concept for visualization and computation: the Modulation domain

The modulation domain

- Mostly we all know what the “Time domain” and “Frequency domain” are



Time-frequency

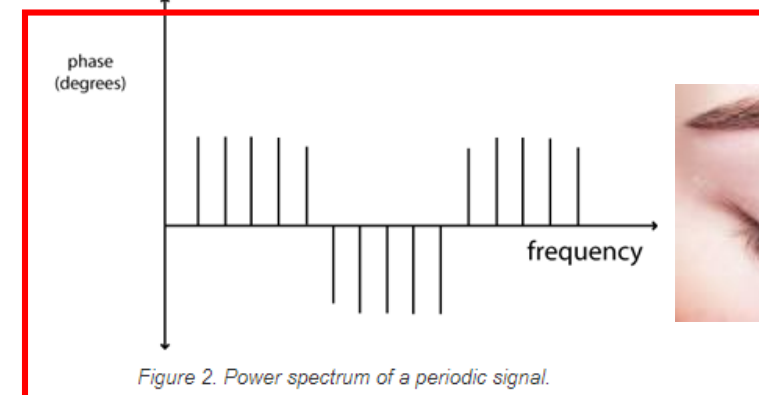
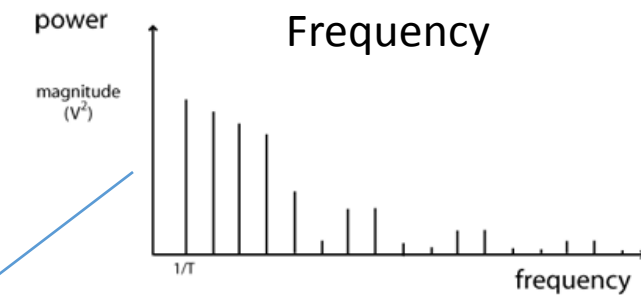
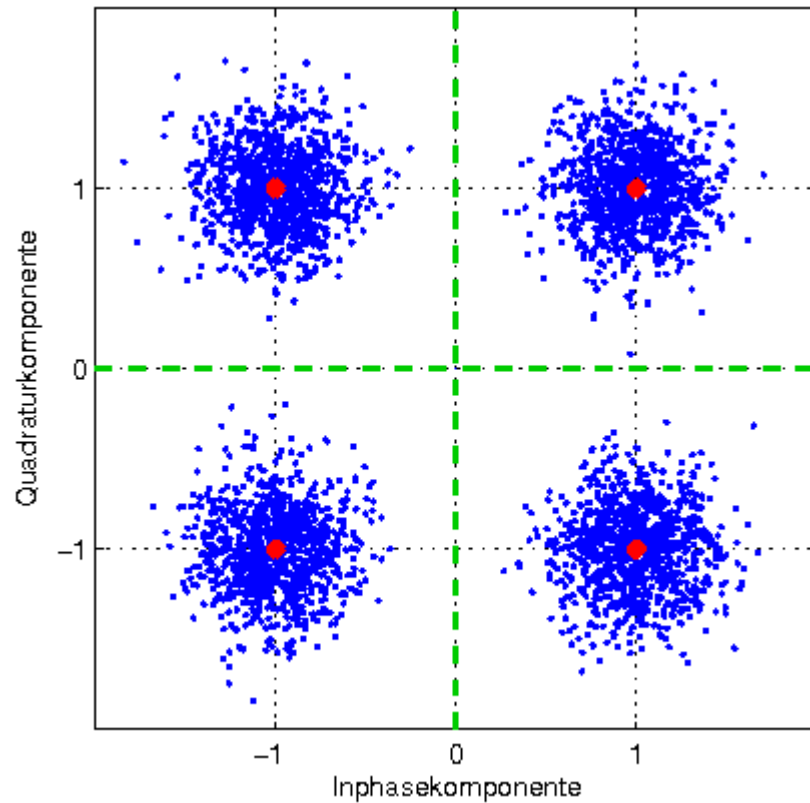


Figure 2. Power spectrum of a periodic signal.

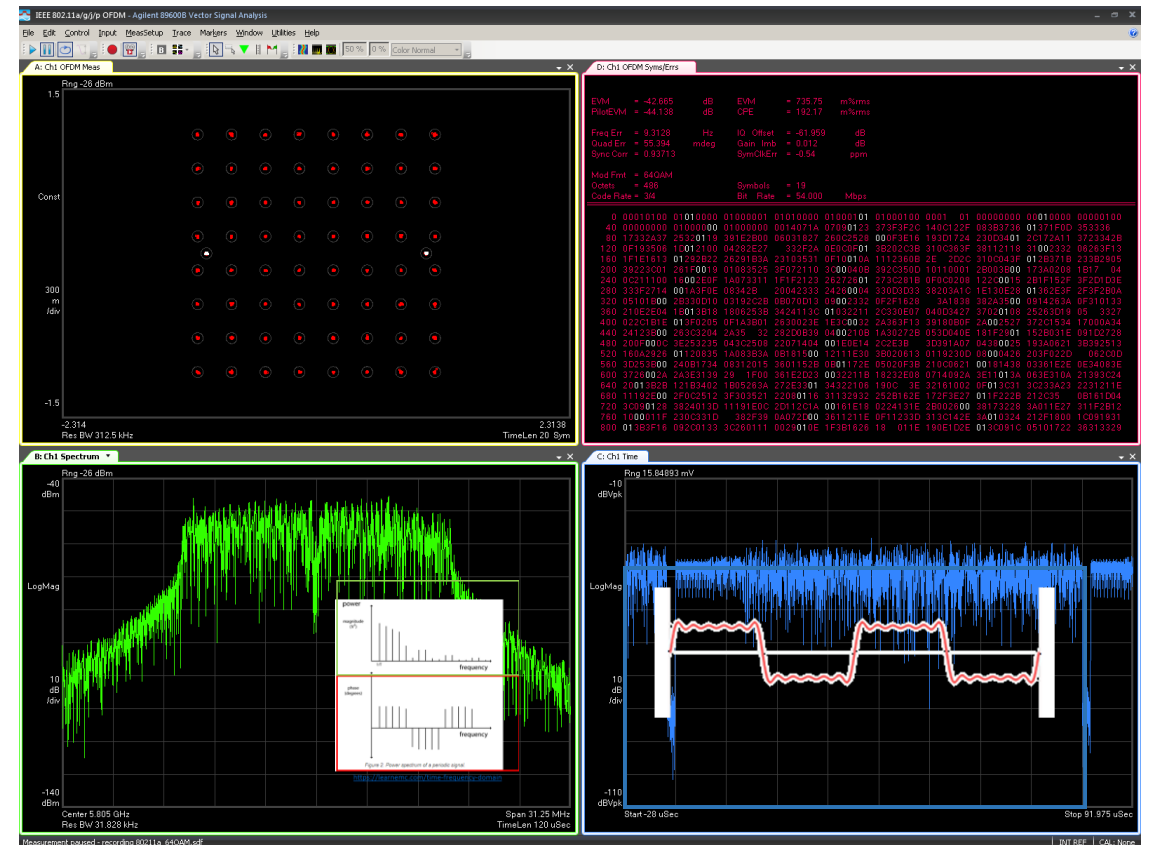
<https://learnemc.com/time-frequency-domain>

The modulation domain

- The modulation domain is where ALL the information is



Time is the series of points



Modulation domain: past OOK

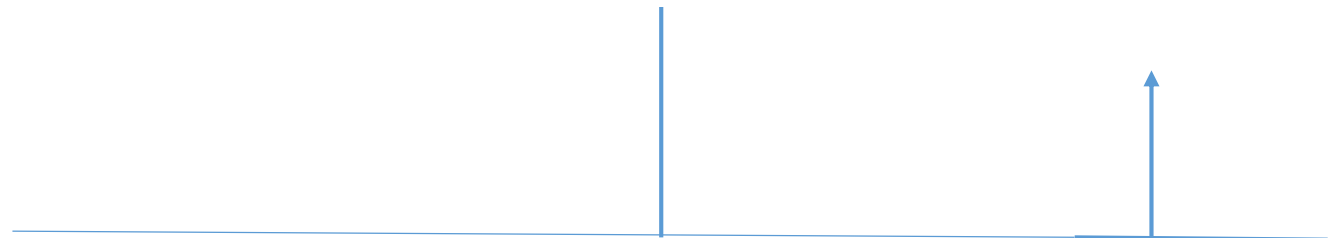
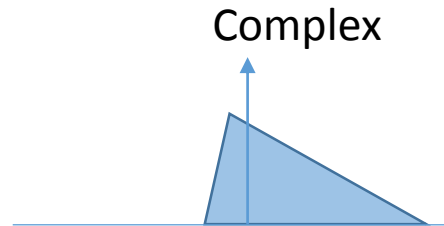
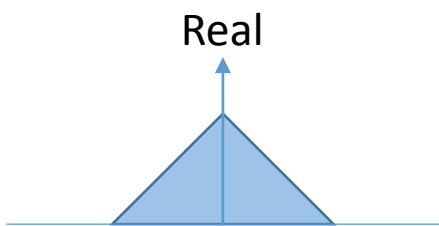
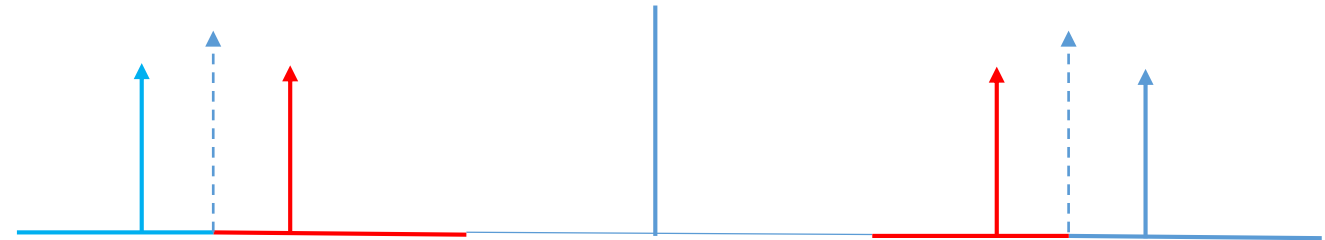
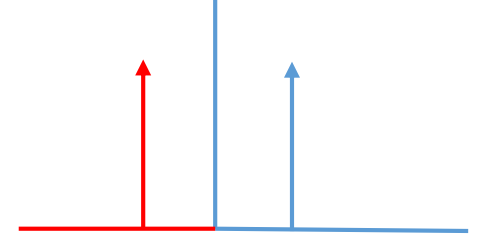
- The modulation domain is the complex (I and Q) envelope of any signal (can be frequency, polarization, etc)
- The “carrier frequency” is a stationary parameter, and contains no information
 - Though it must be determined!
- So called I and Q are orthogonal basis functions in the same way as the “X” and “Y” axis!
- These basis rotate at the carrier frequency, but their orthogonality is preserved always
 - Nonlinear distortion can mess it up!

Analytic signal, AKA special complex signal

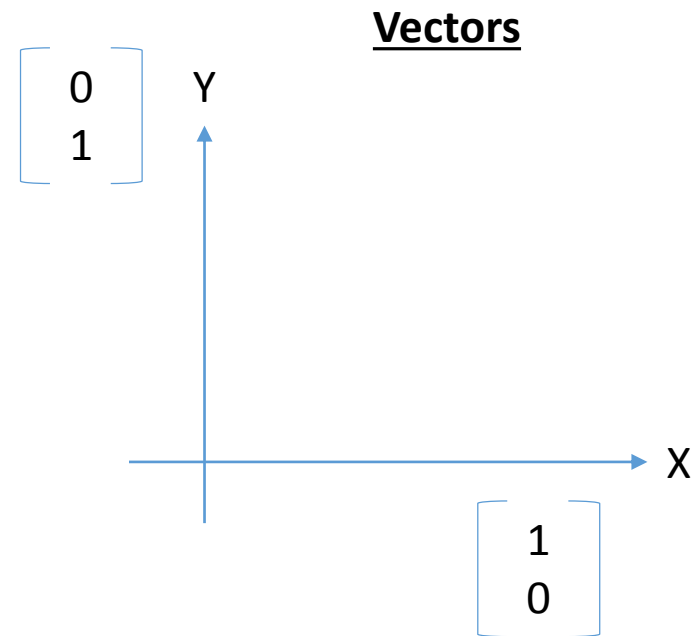
- A so called “real signal” is easy to understand
 - Cos, Sin Etc.
- An analytic signal is a special kind of signal that has a one sided spectrum
- $S'(t) = \cos(\omega t) + j \sin(\omega t) = Ae^{j\omega t}$

More generically (any signal)

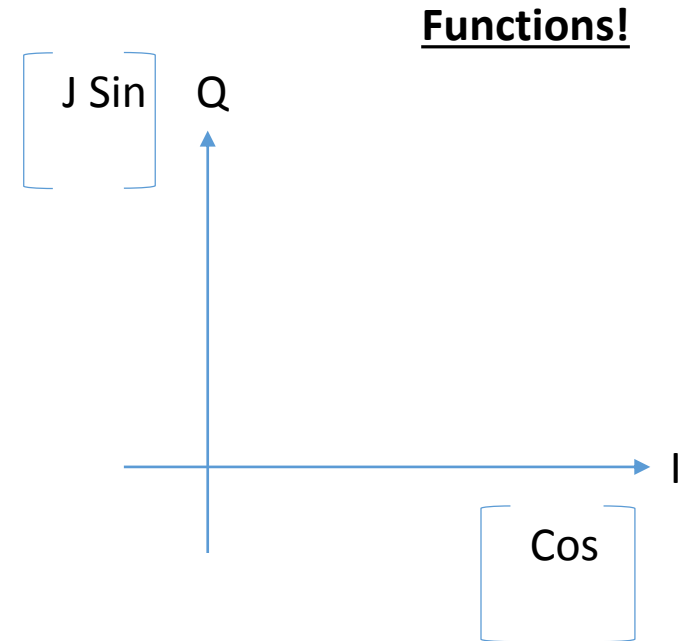
- $S'(t) = A_l(t) * \cos(\omega(\tau)t + k_1(\tau)\theta(\tau, t)) + j * A_q(t) * \sin(\omega(\tau)t + k_2(\tau)\theta(\tau, t))$



Basis functions for complex base band



$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0*1 + 1*0 = 0$$

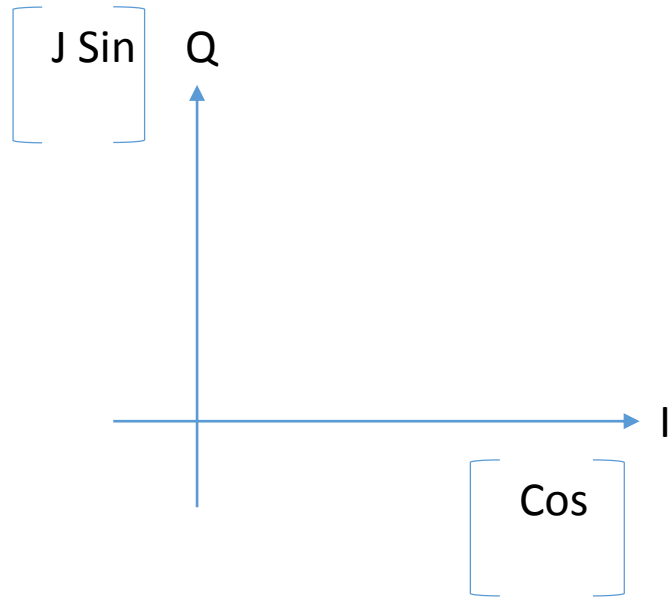


$$\begin{bmatrix} J \text{ Sin} \end{bmatrix} \bullet \begin{bmatrix} \text{Cos} \end{bmatrix} = 0*\text{cos} + \text{sin}*0 = 0$$

Complex dot product!

Basis functions for complex base band

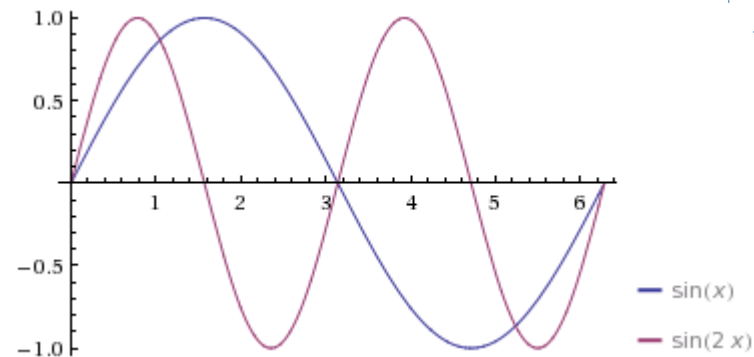
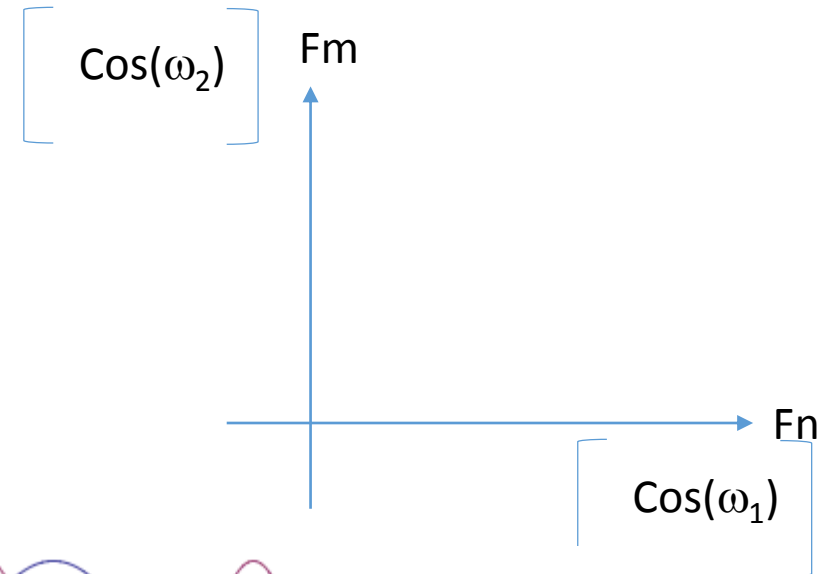
Functions of same frequency



$$\begin{bmatrix} J \sin \end{bmatrix} \bullet \begin{bmatrix} \cos \end{bmatrix} = 0 * \cos + \sin * 0 = 0$$

Complex dot product!

Functions of same different frequency!



<https://math.stackexchange.com/questions/474398/waves-of-differing-frequency-are-orthogonal-help-me-understand>

Complex base band

- Allows a two sided signal to be asymmetric
- Realization of any waveform
- Includes all AM, PM, FM
- Finite BW (Low pass) signal

$$S'(t) = \boxed{A_I(t)} * \boxed{\cos(\omega(\tau)t + k_1(\tau)\theta(\tau, t))} + j * \boxed{A_Q(t)} * \boxed{\sin(\omega(\tau)t + k_2(\tau)\theta(\tau, t))}$$

OOK

FSK

SSB

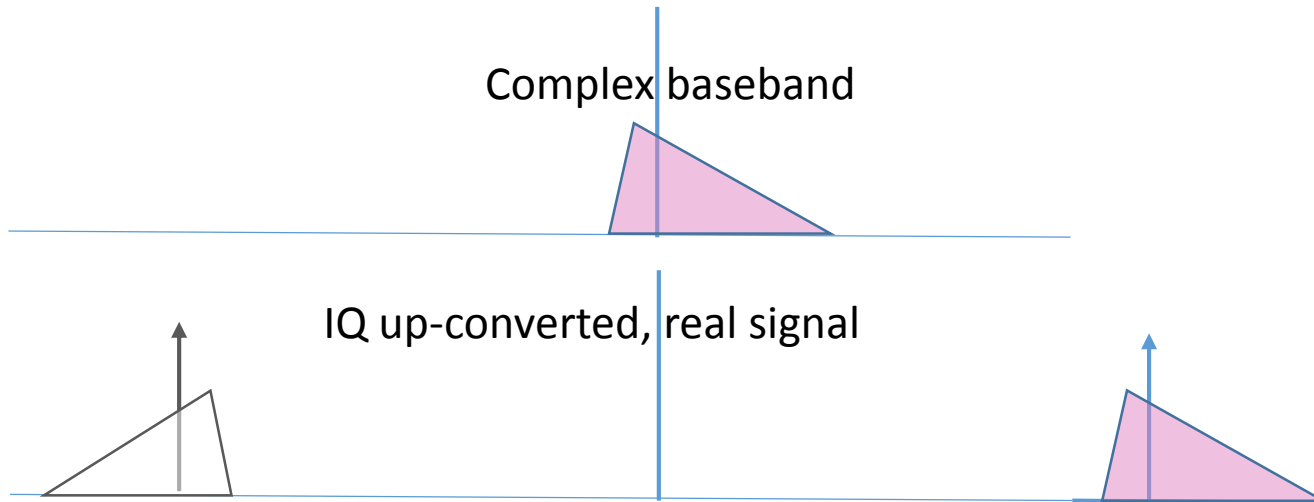
mPSK

MSK /GMSK

LFM

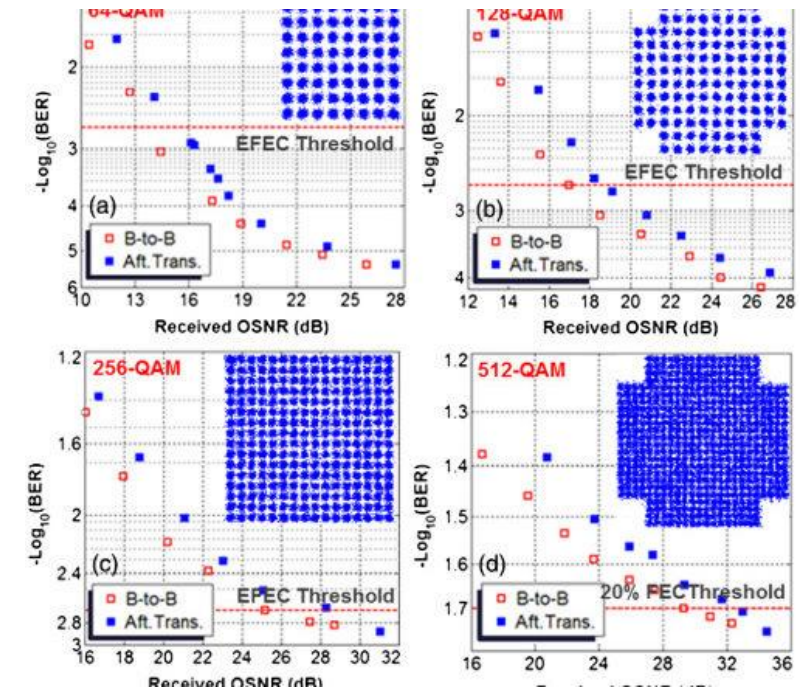
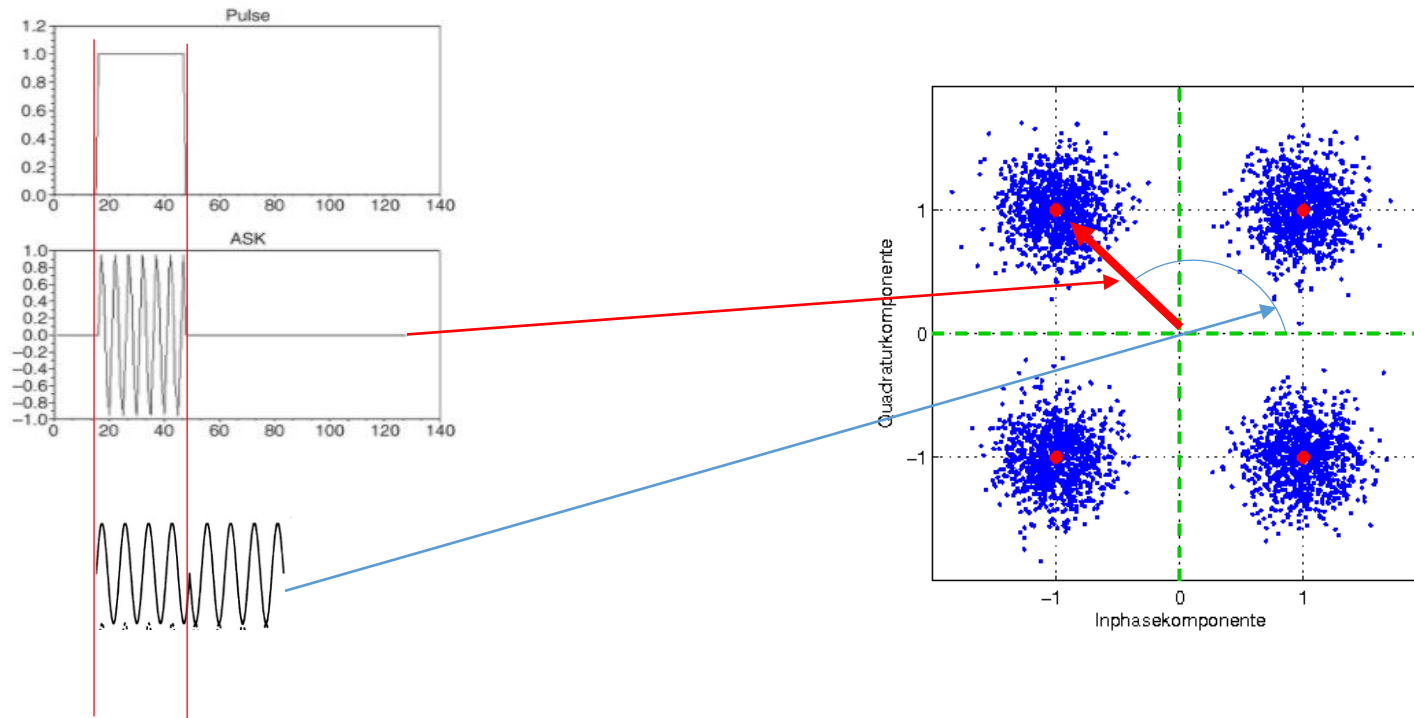
QAM

OFDM



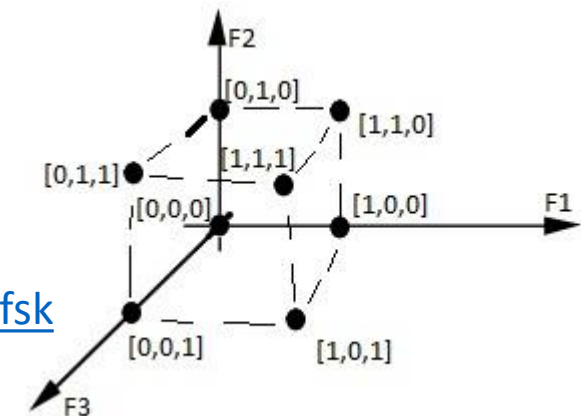
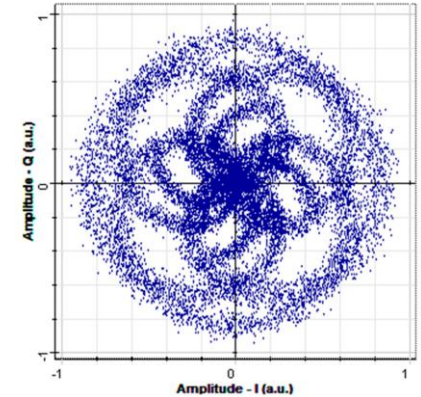
Bit to symbol mapping: define the states

- Starting with complex base band as our signal chalk board we can define “states” in the tuple of I and Q basis.
- Each state corresponds to a complex number.



Can do the same thing with frequency

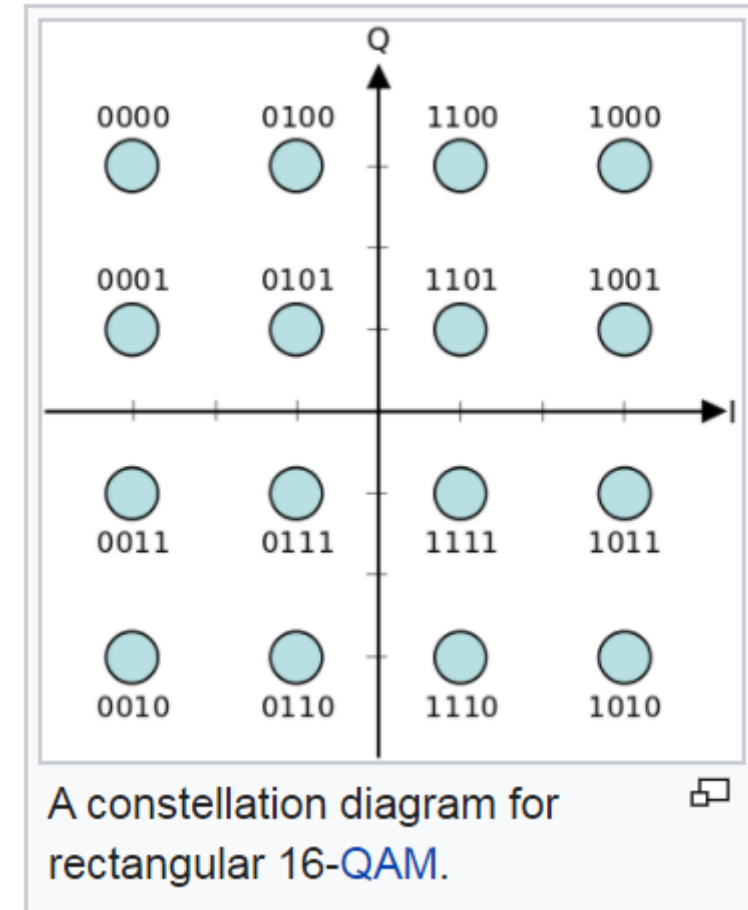
- It is more difficult to visualize, but with a set of orthogonal frequencies, there is a vector of complex coefficients that describe the amplitude of each component
- The amplitude in simple FSK systems is either 1 or 0
- OFDM- this is not true, each frequency can be modulated in phase and amplitude also!



<https://electronics.stackexchange.com/questions/22615/constellation-diagram-of-fsk>

Bit mapping

- For each valid modulation state a bit pattern is assigned according to the binary order of the “constellation”
- Good mapping ensures equal binary distance between elements (gray coding)
- Symbol probability is also valid (infrequency symbols get largest amplitudes – to reduce average power)



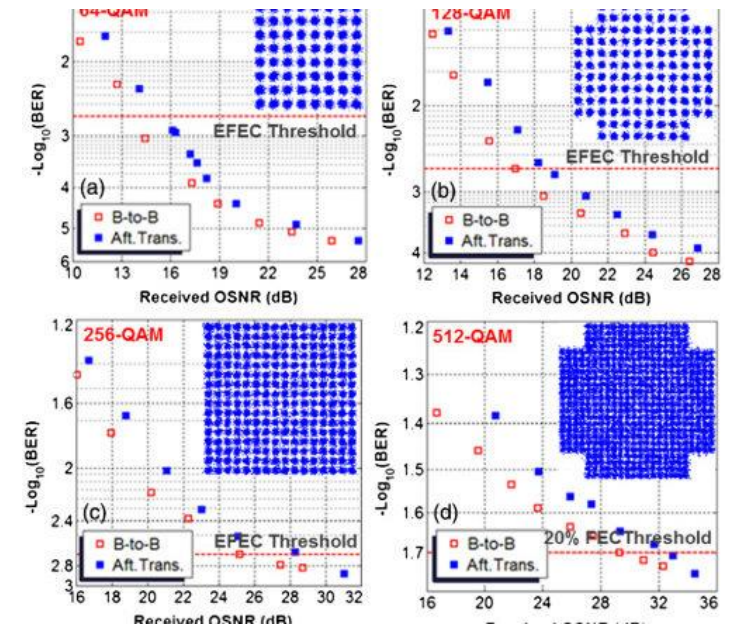
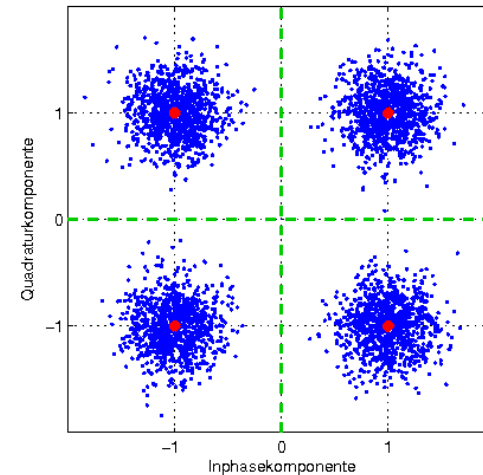
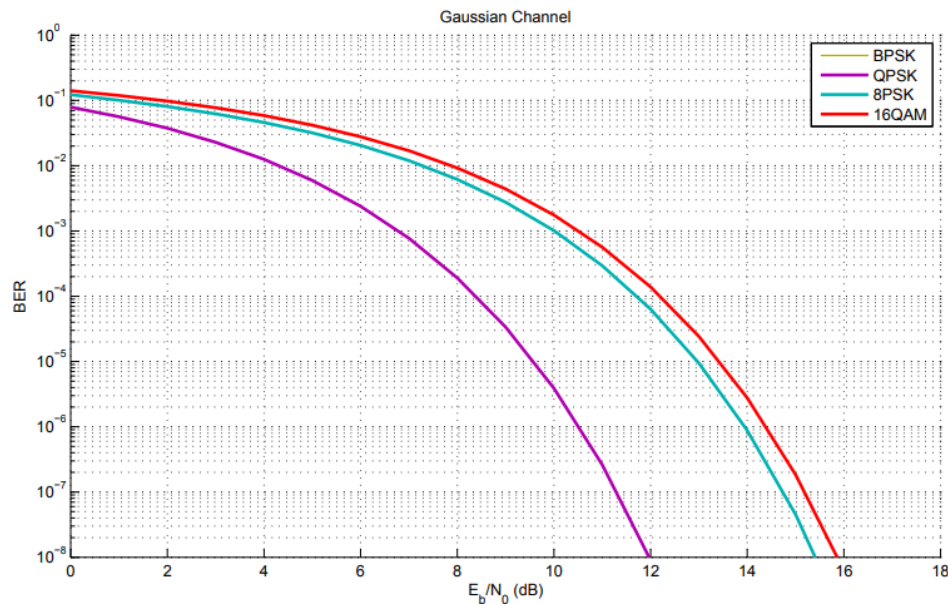
How did this increase channel capacity?

- For a given constellation, the required bandwidth is proportional to the rate of change.
- In modulation containing more than one bit per state, for a given rate, (BW) more bits per second are sent

CW = 1 bit per symbol $\leftrightarrow$ 128QAM = 7 bits per symbol
~Same bandwidth, 7x more information!

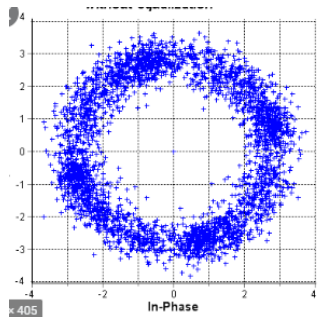
What is the cost?

- Because nothing is free.
- More bits per symbol means the Euclidean distance between the symbols is lower for identical power
- But this is also where CODEING comes in

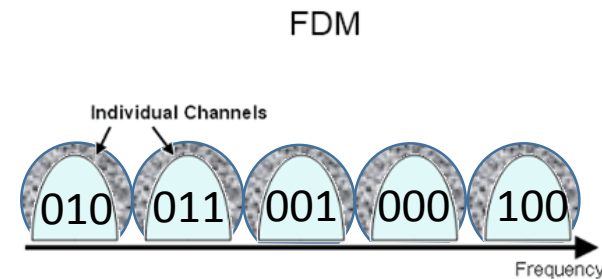
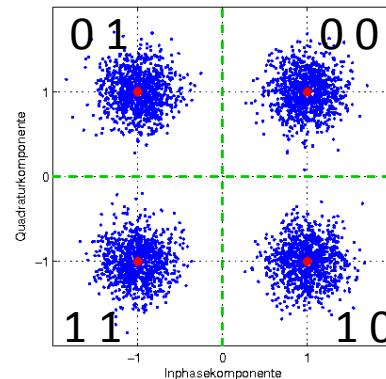
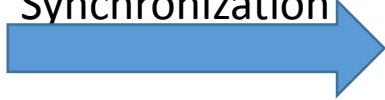


Assuming synchronization, how detect?

- This discussion is for hard decision detection only
 - Soft decision also possible, but more advanced topic
- For QAM variants (OOK, BPSK, QPSK, nPSK, nQAM) comparators in I and Q are used to “slice” the modulation domain
- Massively oversimplifying, a reverse look up of the symbol tuple that is closest to the value of the complex base band signal at that instant in time.



Synchronization



What encompasses synchronization?

- BER curves, link budgets, channel capacity are all reported “assuming synchronization has been established”
- If not stated, it is understood.
- Frequently performance is also after channel equalization and gain adjust
- This means transmitter and receiver deficiencies as well as the channel effects are removed.
- We need to fix: amplitude variation, frequency offset, phase offset, clock rate and sample position, channel impairments
- IMO, these are the cool things about modems!!!!

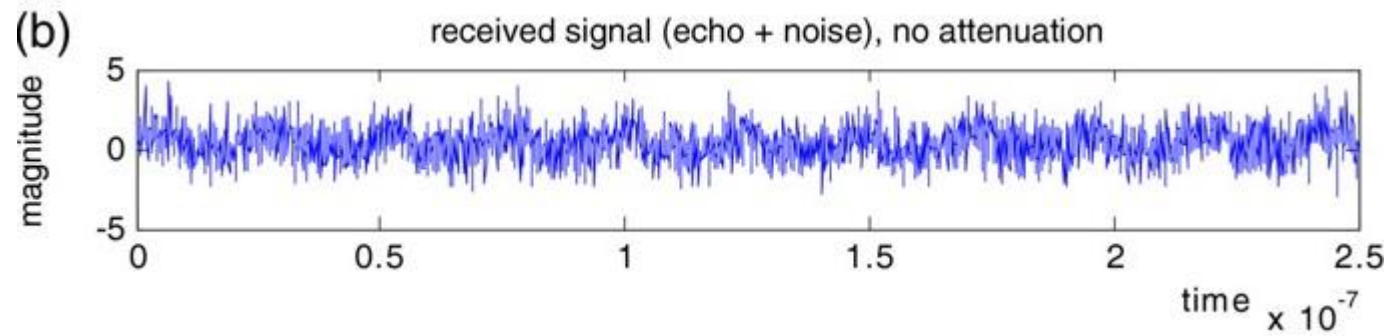
What goes wrong: why is synchronization needed?

- Start with ignoring what the channel does (primarily multipath) other than attenuate the signal
 - Look at synchronization, leave equalization for another time!
- The things we need to figure out:
 - What should the RX gain be
 - Is there a signal, where does the frame / packet / message start?
 - What is its frequency – adjust and track
 - If coherent: what is its phase – adjust and track
 - Where are the symbol boundaries – clock recovery

Synchronization: fixing the things that go wrong

- Each of these can easily occupy many hours of slides: methods, performance, algorithms, specifics for each modulation type etc.
- Instead just an overview here now that we have enough background
- Detection
- Frequency adjust
- Phase adjust
- Clock recovery

Detection: Is there a pulse?

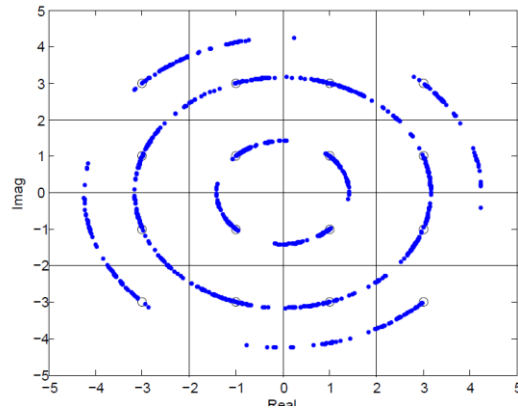


Yes, but where?

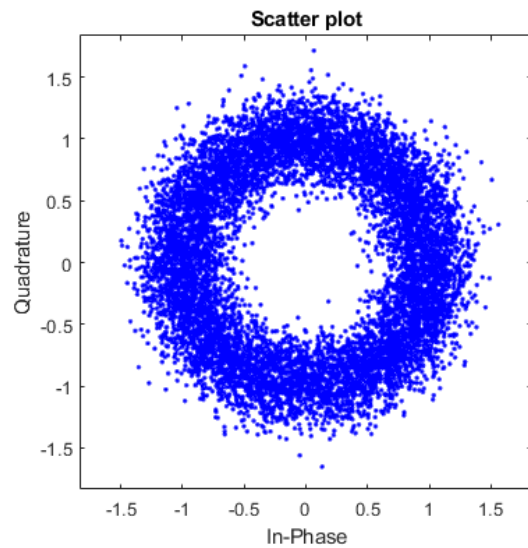
Detection of the signal

- Why do we need to do this?
 - In packet based systems (mostly everything) all of tasks of the modem dealing with equalization and coding need to know precisely where the start of message is.
 - Error of a fraction of a clock period can cause everything to fail.
- Thresholding: the obvious way
 - The trick is, in low SNR conditions it is very difficult to set a threshold
 - Properly implemented, modulation looks like noise
 - In the presence of interference, impossible to distinguish start of signal
 - The signal strength varies orders of magnitude even within a packet
- Typically ratios of auto to cross correlation of the preamble portion of the signal are used to produce a “normalized” threshold method

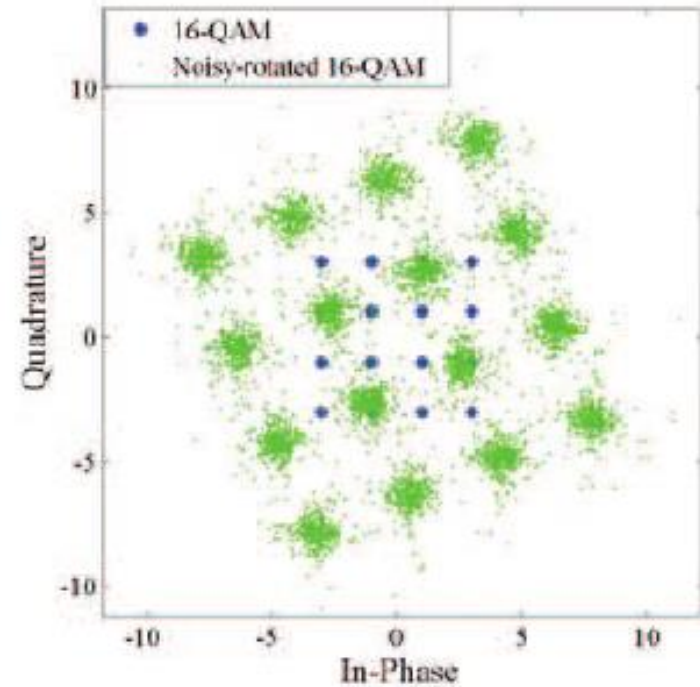
Frequency adjust



Mild 16QAM



Pretty bad, could be
BPSK, QPSK, mPSK etc.



Frequency adjust

$$\omega(t) = \frac{d\varphi(t)}{dt}$$

- Frequency is the rate of change of phase
- If we are interested in the phase, any frequency offset at all will cause a ramping phase – hard to track, not good!
- All transmitters and receivers have small frequency errors that vary with time. Simply not possible to have them both exactly the same.
- Frequency always needs to be adjusted where coherent modulation is used. PLL's can track small offsets.
- It may not be obvious this is being done.
- Eg. WSJT reports error – because it is correcting for the frequency.

FLL

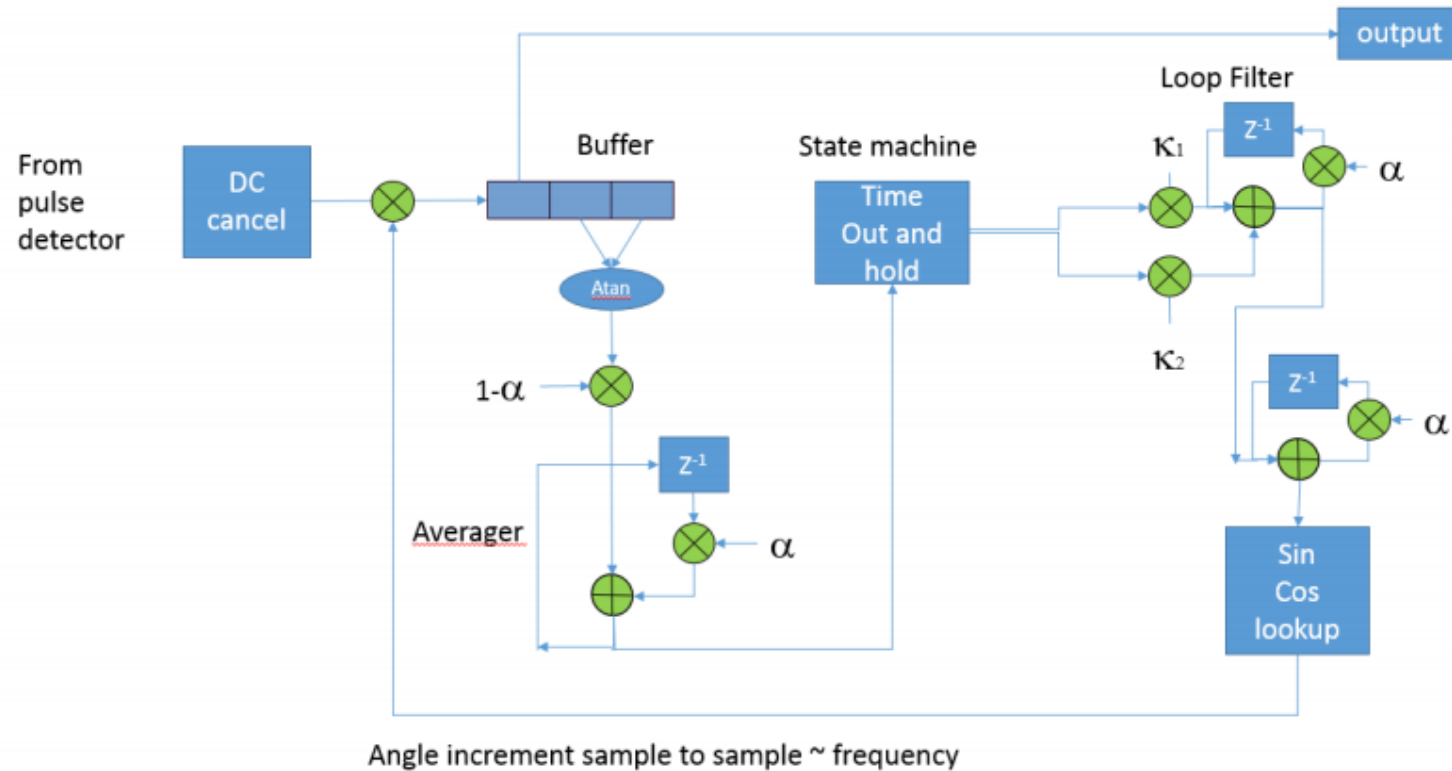
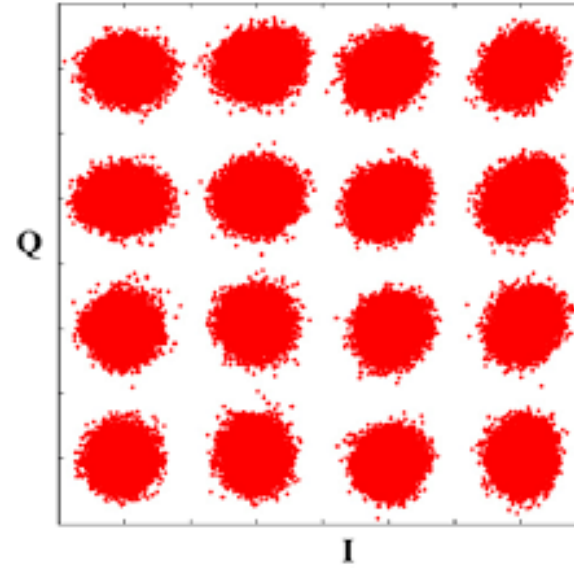
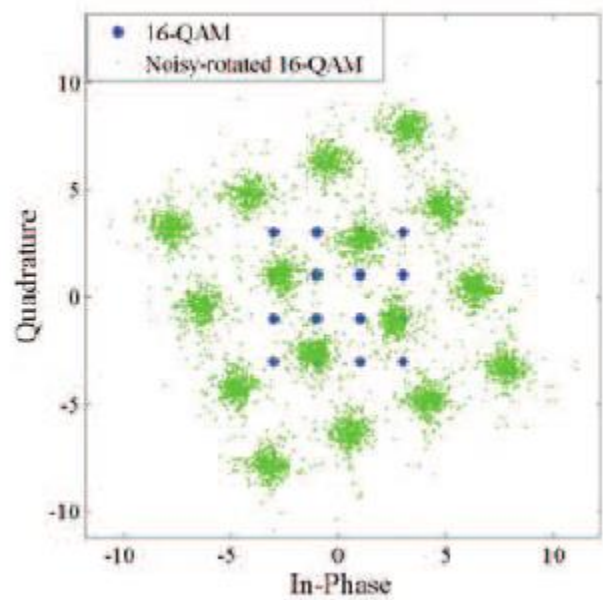
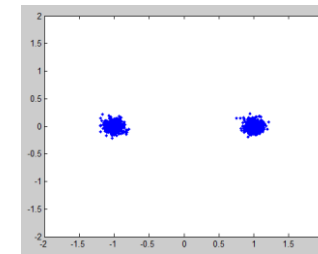
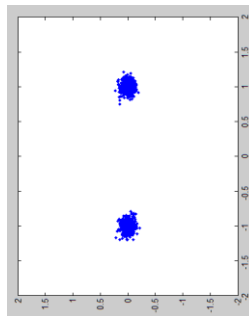
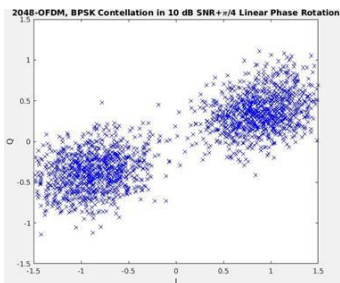


Figure 28 Improved Frequency Lock Loop block diagram

Align Phase

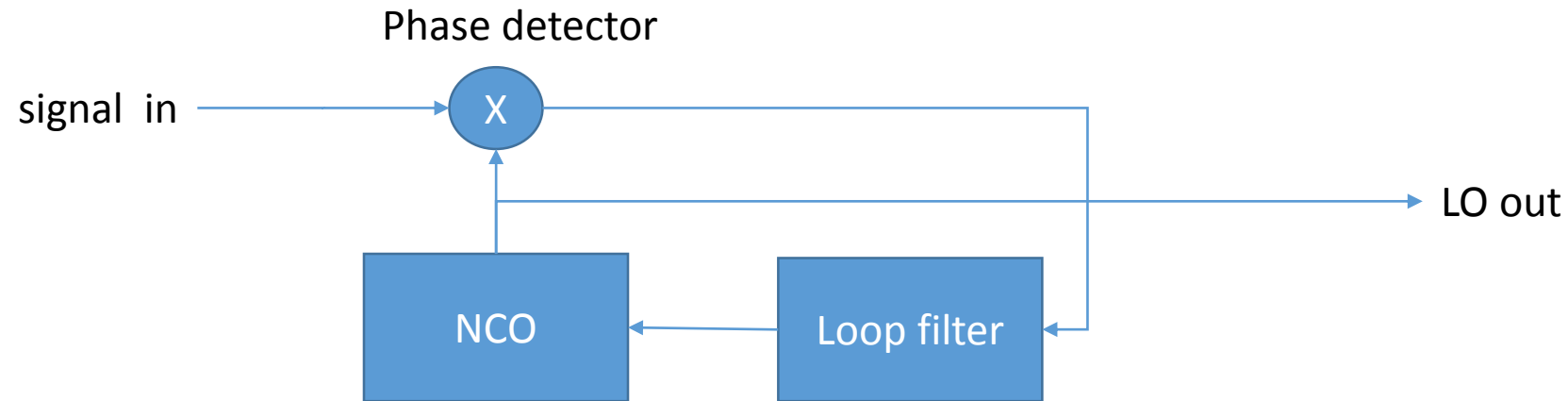


This could be moving in time also!

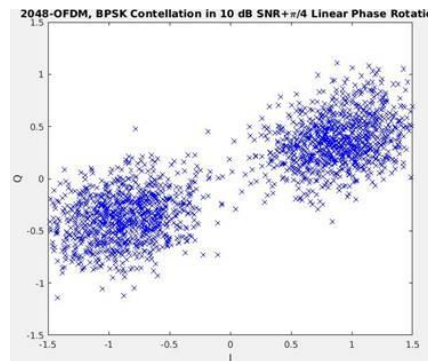


<https://dsp.stackexchange.com/questions/8456/how-to-perform-carrier-phase-recovery-in-software>

Phase locked loop



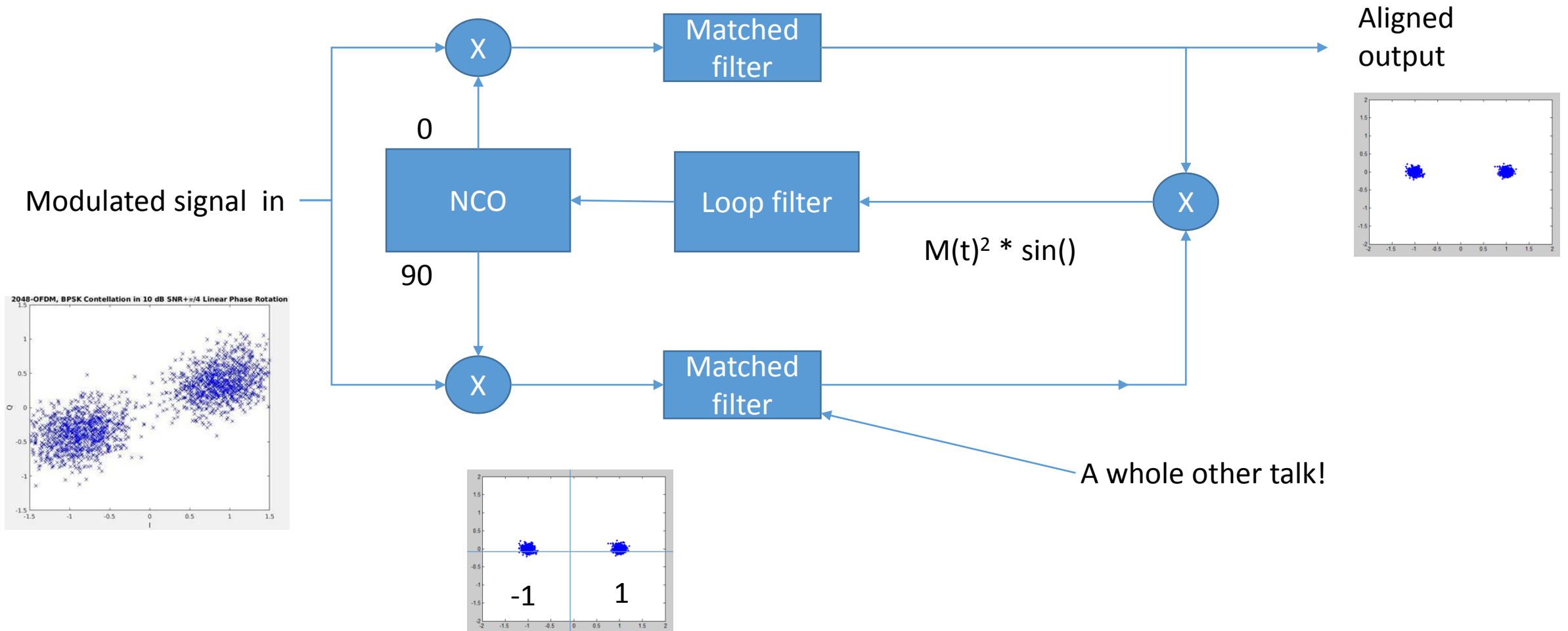
- What if modulation is present?
 - Slow modulation, tracked out
 - Fast modulation, loop fails



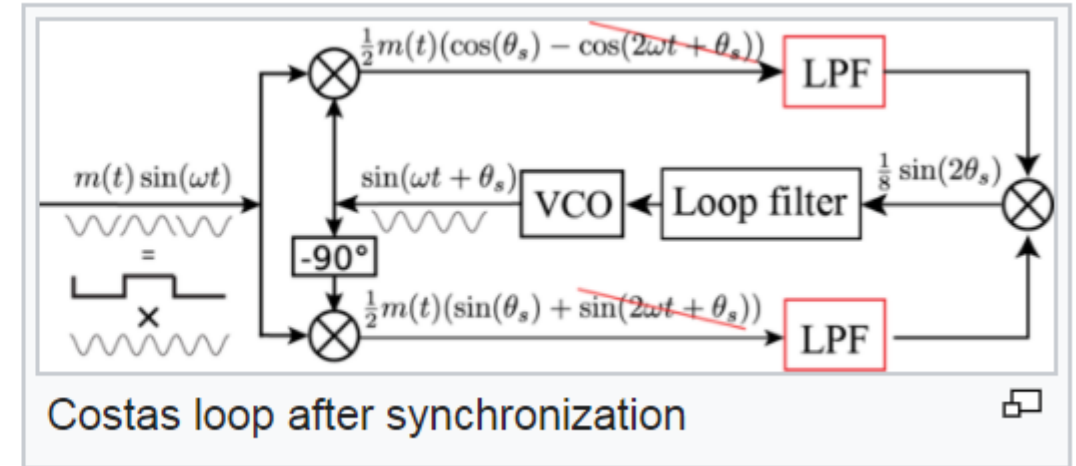
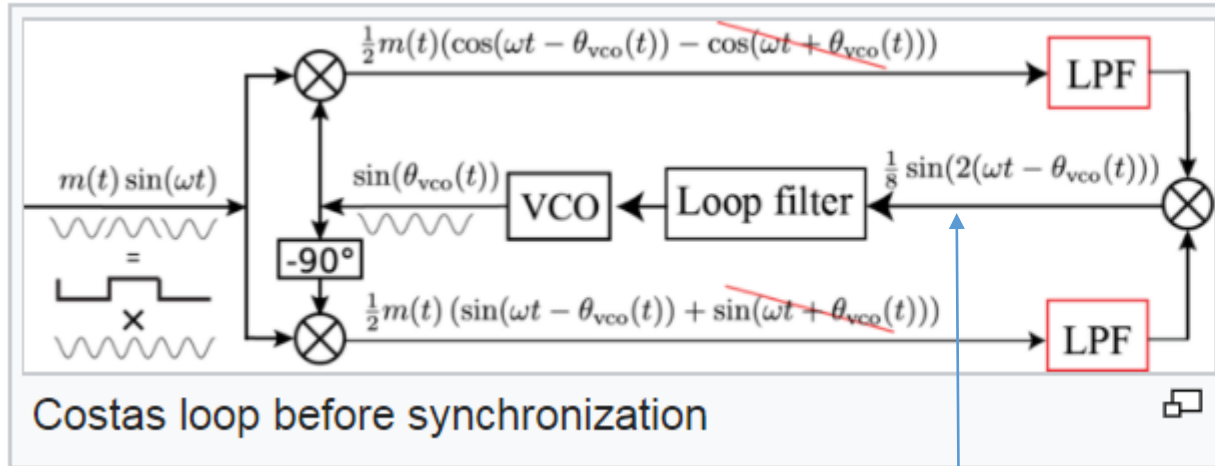
Phase adjust

- When modulation is present, things are a bit harder
 - Modulation must be “removed” from error signal
 - Many methods exist, a common one is the Costas loop

Costas loop for BPSK

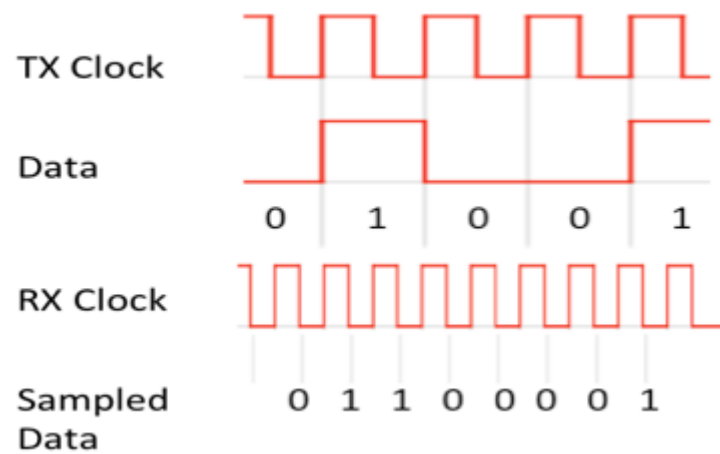


Costas loop for BPSK

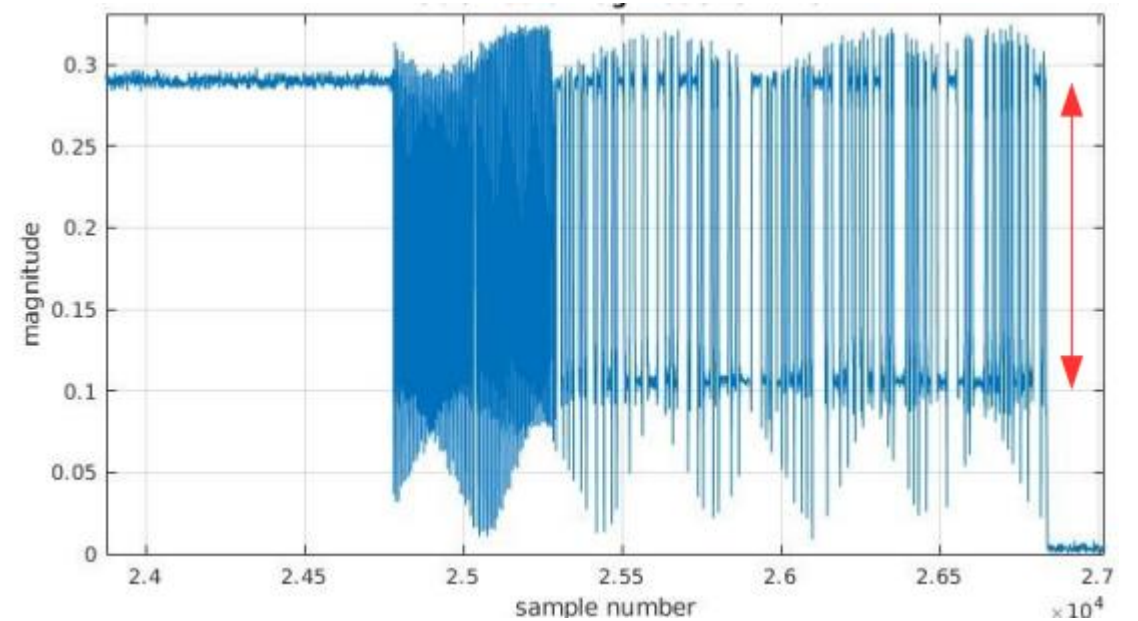


Clock recovery: when to sample?

Obvious

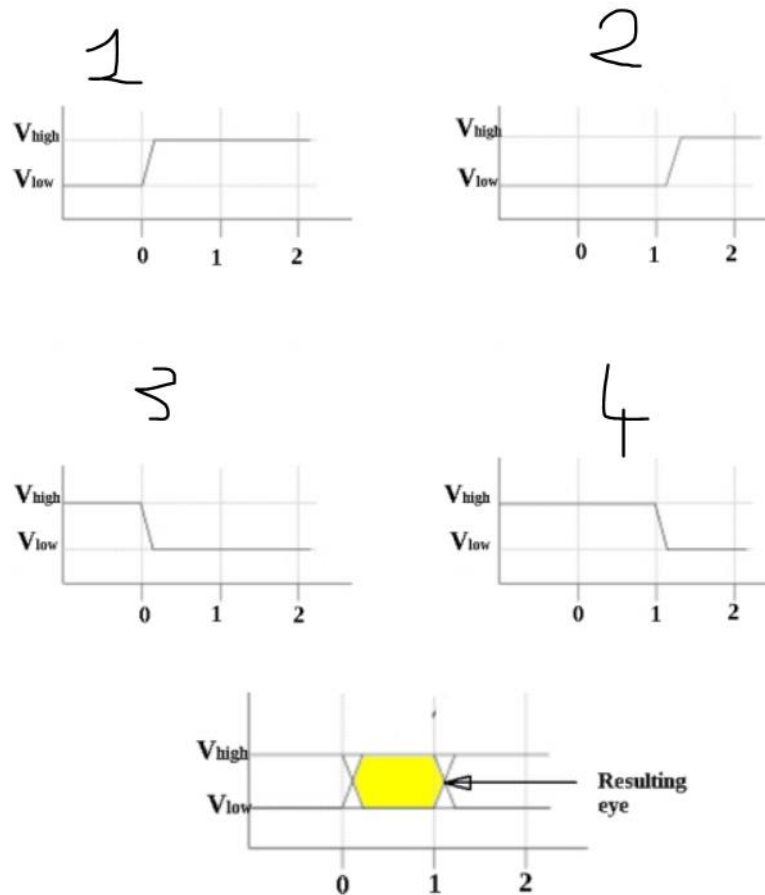


NOT So Obvious

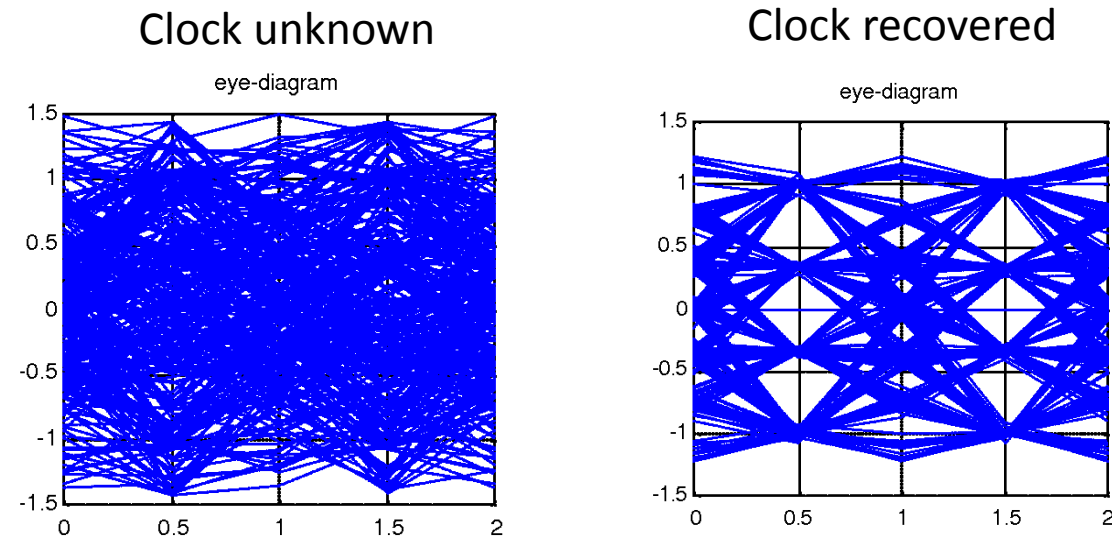


Eye diagrams

<https://dsp.stackexchange.com/questions/64613/eye-pattern-construction-and-interpretation>



Eye diagrams: clock recovery



<https://www.semanticscholar.org/paper/Teaching-MODEM-Concepts-and-Design-Procedure-with-Harris/0d8f7db7882739c7b49b634f40cde3f569d99bd2/figure/3>

Clock recovery

- Once the frequency is adjusted so the phase does not ramp
- Then the phase is adjusted to align the constellation
- Then the slicer can make a decision on what the symbol is – but it is changing at the clock rate, so how do you know when to sample?
- In low SNR case, it is not at all obvious
- Methods include
 - Delay locked loop
 - Early late method
 - Garner algorithm
 - Polyphase interpolation
 - Etc etc
- Just a taste here

Some examples

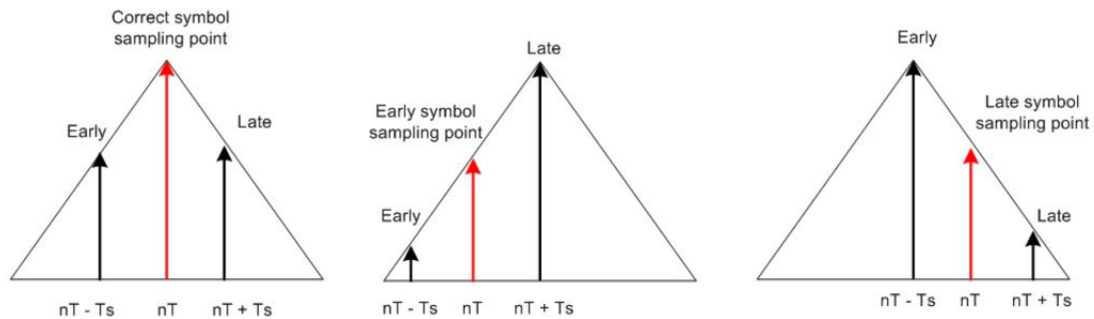
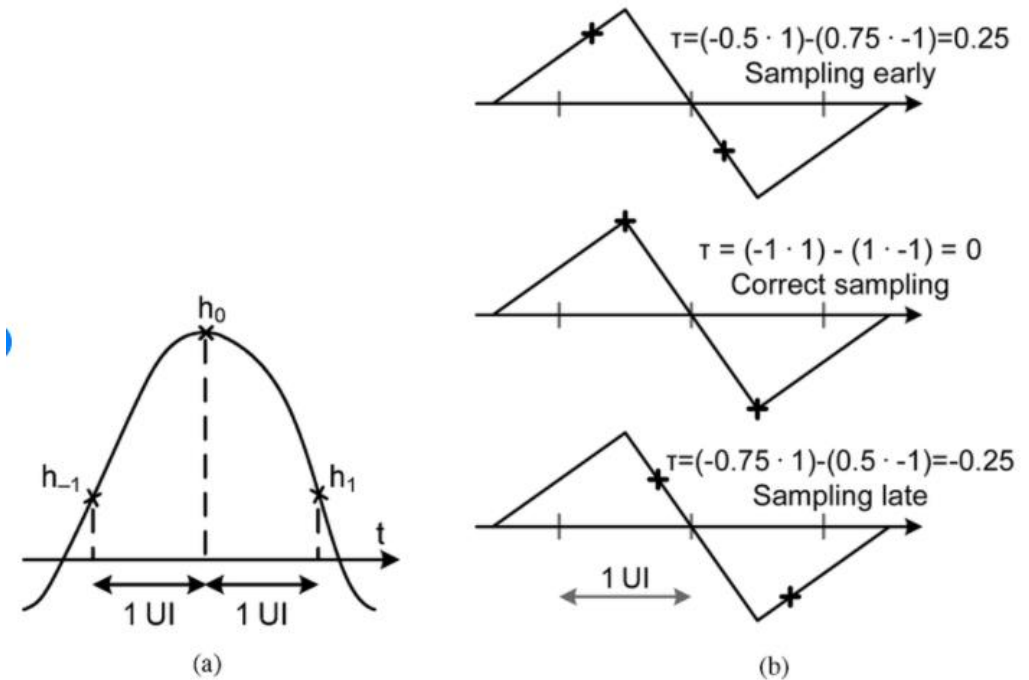


Figure 1: Late-early timing error computation



Mueller-Müller timing recovery scheme. (a) Impulse response. (b) Continuous data.

<https://www.nutag.com/blog/symbol-timing-recovery-methods-digital-iq-demodulator>

https://www.researchgate.net/figure/Mueller-Mueller-timing-recovery-scheme-a-Impulse-response-b-Continuous-data_fig1_224144594

Equalization: last taste

- Pretty advanced topic, but needs mention
- The frequency response of a channel is typically not “flat” and can be rapidly changing
- Primarily caused by multipath.
- These echos act as delay taps to form a filter function that is not flat in frequency, and introduce inter-symbol interference
- Not seen so much at narrow channel on HF bands since wavelength is very long, but still present.
- Extreme at microwave frequencies with Wi-Fi bandwidths
- All modern wireless standards must include equalization

Wrap up

- This presentation was a fast gloss over a huge subject
 - At least six graduate level classes
- I hoped to only get enough info out that it would be easier to penetrate the subject if interested
- I added links to the image sources to help in that case
- I would like to spin a talk about each one of the things touched on here if interest exists
- Thank you!